



Eidgenössische Technische Hochschule Zürich  
Swiss Federal Institute of Technology Zurich

# **Study of Anisotropic Antiferromagnets on a Triangular Lattice through Neural Quantum States**

Master's Thesis

Sami Zweidler  
`samiz@student.ethz.ch`

Institute of Theoretical Physics  
Department of Physics, D-PHYS  
ETH Zürich

## **Supervisors:**

Prof. Dr. Juan Carrasquilla  
Dr. Jannes Nys  
Dr. Matija Medvidović

May 23, 2025

## **Acknowledgments**

I express my gratitude to Prof. Juan Carrasquilla for giving me the opportunity to work collaboratively within his group. I am also thankful for his inspiring lecture on computational quantum physics, which laid the foundation for this thesis. Special thanks go to my supervisors, Dr. Jannes Nys and Dr. Matija Medvedovic, for their continuous guidance, insightful discussions, and patience throughout this project. Without their support, this work would not have been possible. I also thank Vinícius Mohr for the engaging and helpful discussions that contributed to shaping the direction of this work. Last but not least, I want to thank my friends and family, especially my mother, who supported me both financially and emotionally throughout this journey.

# Abstract

Analyzing the physics of two-dimensional magnetic systems, particularly those that demonstrate frustration, is a prominent subject in current condensed matter physics research. When frustration occurs, it disrupts the magnetic order, potentially causing its suppression to absolute zero temperature. Consequently, accurately characterizing the ground state becomes a challenge. This thesis explores the quantum antiferromagnet  $\text{Cs}_2\text{CoBr}_4$ , which demonstrates two competing mechanisms of frustration due to its anisotropic and geometric properties.

We employ neural quantum states combined with variational Monte Carlo methods to examine its ground state at zero temperature and absent magnetic field. Initially, we introduce an extended version of the Marshall-Peierls sign rule that accounts for potential anisotropies in the Hamiltonian model, aiding in the search for ground states. Using the architectures of a restricted Boltzmann machine and a Vision Transformer as a variational ansatz, we evaluate the ground state across finite lattice sizes. Our results show an excellent match for small linear dimensions, although refinement is needed for larger linear sizes.

Moreover, we notice a uni-dimensional pattern in the spin structure factor, suggesting *quasi-1D ordering* in the ground state that persists up to the thermodynamic limit. Our findings indicate that employing machine learning methods for quantum state representation is a promising approach to tackle significant challenges in quantum many-body physics.

# Contents

<b>Acknowledgements</b>	<b>ii</b>
<b>Abstract</b>	<b>iii</b>
<b>1. Introduction</b>	<b>1</b>
<b>2. Theory</b>	<b>3</b>
2.1. Background to Quantum Many Body Theory for Spin Systems . . . . .	3
2.2. Physical System . . . . .	4
2.3. Variational Monte Carlo . . . . .	6
2.3.1. Markov Chain Monte Carlo Sampling . . . . .	7
2.3.2. Stochastic Gradient Descent . . . . .	8
2.3.3. Stochastic Reconfiguration . . . . .	8
2.3.4. Accuracy of Variational States . . . . .	10
2.4. Symmetries . . . . .	10
2.5. Sign Structure . . . . .	11
2.5.1. Marshall Peierls Sign Rule . . . . .	11
2.5.2. Generalized Sign Rule . . . . .	12
2.6. Variational States . . . . .	14
2.6.1. Neural Quantum States . . . . .	15
2.6.2. Restricted Boltzmann Machine . . . . .	16
2.6.3. Vision Transformer . . . . .	17
<b>3. Results &amp; Discussion</b>	<b>20</b>
3.1. One Dimensional Case . . . . .	20
3.2. Two Dimensional Case $\text{Cs}_2\text{CoBr}_4$ . . . . .	21
3.2.1. Perturbative Approach . . . . .	21
3.2.2. Benchmarks using RBM & ViT . . . . .	22
3.2.3. Finite Size Scaling . . . . .	28
<b>4. Conclusion</b>	<b>30</b>
<b>A. Derivations</b>	<b>32</b>
A.1. Stochastic Reconfiguration . . . . .	32
A.1.1. Fidelity Optimization . . . . .	32
A.1.2. Residual Norm Optimization . . . . .	33
A.2. Marshall Peierls Sign Rule . . . . .	34
A.3. Symmetries . . . . .	34
A.4. Data & Code . . . . .	35

# Introduction

---

Quantum many-body systems exhibit some of the most fascinating and unexpected phenomena in nature, arising from the collective behavior of interacting particles.

One striking example is magnetism, a phenomenon rooted in the collective alignment of electron spins in a material. Although at a microscopic level individual electron spins follow the rules of quantum mechanics, it is their coordinated alignment that gives rise to large-scale magnetic fields. These macroscopic effects are exploited in technologies ranging from magnetic resonance imaging machines in medical imaging to data storage devices in modern computers.

In addition to magnetism, many-body systems also underpin exotic states of matter. One example is superconductivity, where materials conduct electricity without resistance. Another example are quantum spin liquids, which remain disordered at even absolute zero temperature and display fractionalized excitations and topological order [1]. Exploring these systems remains a rich and challenging field that promises both profound theoretical insights and technological advances.

However, studying these systems is far from trivial. As the number of particles increases, we encounter the curse of dimensionality: an exponential growth of the Hilbert space that makes analytical solutions exceedingly rare and computational simulations immensely challenging.

One of the most powerful tools for studying many body physics is the quantum Monte Carlo method (QMC) [2]. This method allows for the calculation of multi-dimensional integrals via the Monte Carlo approach, yielding numerically precise results for different observables like energy and correlation functions. Despite this, Quantum Monte Carlo (QMC) techniques can be particularly challenging when dealing with non-stoquastic Hamiltonians, such as those found in frustrated magnetic systems or involving fermionic interactions, due to appearing the sign problem [3].

One way to avoid the notorious sign problem is through variational methods, where the goal is to approximate the exact quantum state using a set of adjustable parameters. A major challenge in this approach lies in designing a variational ansatz that can faithfully capture the true nature of the system. Popular methods are variational Monte Carlo (VMC) [2] and density renormalization group (DMRG) [4]. An interesting approach to overcome this difficulty is employing neural networks to parameterize variational wave functions [5], so-called neural quantum states (NQS).

While NQS-based VMC methods can be computationally intensive due to the need for repeated sampling and optimization, they are well-suited for acceleration on modern hardware such as GPUs. Libraries such as *NetKet* [6] enable efficient implementation and scalability, making these approaches practical and competitive for simulating complex quantum many-body systems.

Up to date, a whole zoo of neural quantum states have been introduced [5, 7–11], enabling us to overcome various issues in many-body quantum systems. These advancements include the ability to effectively represent volume-law entangled states [12], and, in general, demonstrate particular suitability for two-dimensional systems where matrix product states (MPS) encounter limitations. Within the realm of spin systems, significant research has been conducted to explore and characterize two-dimensional systems, including the Ising model, Heisenberg model, and  $J_1 - J_2$  model across various lattice configurations.

This thesis examines the underlying model of the quantum antiferromagnet  $\text{Cs}_2\text{CoBr}_4$ , represented by an XYZ-model on a triangular lattice. This model is intriguing because it incorporates geomet-

## *1. Introduction*

ric frustration not only through the design of the lattice but also through the anisotropy in coupling strengths. While the material's existence has been known for a long time [13], its properties within the realm of quantum magnetism were not thoroughly explored. Only in recent years has there been significant experimental research [14–16], yet a theoretical investigation into how these frustration mechanisms interact remains limited [17]. This thesis aims to establish a foundation and strives for an accurate description of the system's ground state using NQS.

# Theory

---

This chapter starts with an introduction to the antiferromagnet being analyzed and outlines its importance. We delve into its physical properties and clarify why using exact diagonalization (ED) methods is impractical for identifying ground state properties. Subsequently, we present the numerical techniques that replace ED, alongside certain physics-inspired strategies to enhance the numerical methods.

## 2.1. Background to Quantum Many Body Theory for Spin Systems

States describing a many-body system are represented by a vector in a Hilbert space  $\mathcal{H}$ . This vector is a linear combination of observable system configurations  $\{c_i\}$ , known as the computational basis.

In the context of spin 1/2 systems, each spin can be measured in two states, spin-up or spin-down which can be represented as  $\uparrow, \downarrow$  respectively  $\pm 1$ . For  $N$  spins there are  $2^N$  configurations in the computational basis. In particular, a state  $|\psi\rangle$  can be written as

$$|\psi\rangle = \sum_{i=1}^{2^N} \psi(c_i) |c_i\rangle$$

where  $|c_i\rangle$  represents an array of spin configuration with  $N$  spins, e.g.  $\uparrow\uparrow\downarrow\uparrow\dots\downarrow$  and  $\psi(c_i)$  is the wave function component which is in general a complex number.

Here we can already explicitly see the curse of dimensionality, namely the exponential scaling of the size of the Hilbert space with the number of spins, which is a numerical challenge. For a system of  $N = 40$  spins, we already need  $2^{40} \approx 10^{12}$  bits to store a single state. This will hardly fit on the memory of a computer.

The interactions between particles in a many-body quantum system are described by the Hamiltonian  $\hat{H}$ , a Hermitian operator acting on the corresponding Hilbert space  $\mathcal{H}$ . The energy levels and corresponding quantum states are governed by the time-independent Schrödinger equation,

$$\hat{H} |\psi\rangle = E |\psi\rangle$$

where  $E$  is an eigenvalue of  $\hat{H}$  and  $|\psi\rangle$  the corresponding eigenvector. In principle, the eigenvectors can be determined through the eigenvalue decomposition of the Hamiltonian. However, for a system of spins  $N$ , the Hamiltonian matrix increases exponentially in size to  $2^N \times 2^N$ . This exponential scaling limits the full decomposition of eigenvalues to relatively small systems, typically  $N \leq 20$ .

When the focus is on the lowest-lying eigen states, iterative solvers such as the Lanczos method [18] can be employed. This approach makes it feasible to study systems of size  $N \approx 40$  [19]. The Lanczos method, often referred to as exact diagonalization (ED), is a powerful tool for addressing such systems. For larger system size, one has to use numerical tools such as quantum Monte Carlo (QMC) and variational Monte Carlo (VMC) which are based on Monte Carlo sampling or variational methods such as density functional theory (DFT) density matrix renormalization group (DMRG). This thesis primarily employs variational Monte Carlo methods to explore the physical system outlined in the subsequent section. In particular, our goal is to identify the system's state with the smallest energy, known as the ground state. This state characterizes the system's properties at absolute zero temperature.

## 2. Theory

### 2.2. Physical System

In certain quantum magnets, the nature of spin ordering is not immediately clear because of the interplay of quantum fluctuations and magnetic frustration. A well-established picture for frustration in a quantum magnet is the competing Heisenberg terms in a spin- $\frac{1}{2}$  Hamiltonian. The standard example is the antiferromagnetic Heisenberg model on a generic triangular lattice [20]. When anisotropy occurs, it is typically considered to be a minor disturbance linked to spin-orbit coupling. Anisotropy in spin models refers to the direction-dependent interaction strengths between spin components. For instance, the XXZ model on a triangular lattice shows physics similar to the Heisenberg scenario [21]. Nevertheless, studies have shown that anisotropy can be the main factor leading to frustration, thereby rendering our initial assumptions inaccurate. A well-known example is the Kitaev model, which is exactly solvable. In this model, the three bond-dependent Ising couplings create a spin liquid in the ground state with fractional excitations [22, 23].

In 2D magnets, the interaction of quantum fluctuations, frustrations, and anisotropy results in a diverse range of intriguing phenomena, some of which mirror the features of one-dimensional systems. Frustrated quasi-2D magnets can show some characteristics of 1D magnets even in regions with long-range magnetic order [24]. A popular candidate for this is the crystal  $\text{Cs}_2\text{CuCl}_4$ . The material exhibits both the spinon continuum of spin-1/2 chains and quasi-two-dimensional magnons [25, 26]. The Hamiltonian addressed in this work is present in the quantum antiferromagnet  $\text{Cs}_2\text{CoBr}_4$ . This material belongs to the family of quantum magnets characterized by a distorted triangular lattice  $\text{Cs}_2\text{RX}_4$ , where X is chlorine or bromine and R is copper or cobalt. Other materials in this family, such as the afore-mentioned  $\text{Cs}_2\text{CuCl}_4$ , have been thoroughly explored both theoretically [27–29], and experimentally [25, 26].

Recently, the authors of the article [14] have investigated  $\text{Cs}_2\text{CoBr}_4$  experimentally and discovered intriguing physical phenomena. Notably, this quantum magnet presents an interesting phase diagram when subjected to external magnetic fields. Their research demonstrates that, at low temperatures, the material transitions through five distinct magnetic ordered phases as the magnetic field is increased. These stages have not been completely comprehended or physically characterized so far. To provide

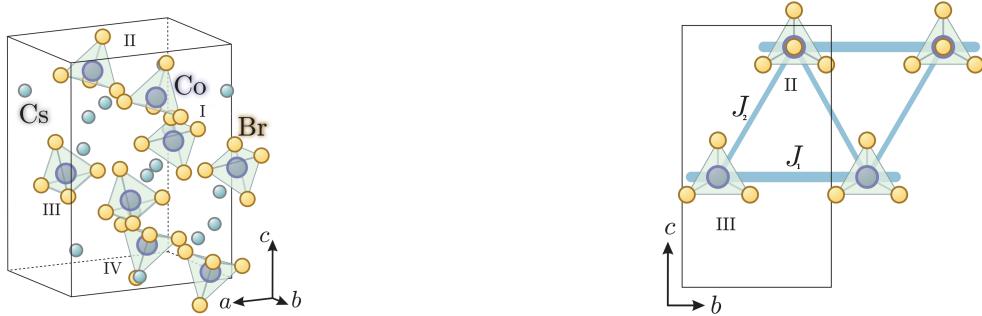


Figure 2.1.: Depiction of the quantum magnet  $\text{Cs}_2\text{CoBr}_4$ . The image on the left presents a snapshot of its crystal lattice, while the right illustrates the projection onto the  $bc$  plane, revealing the distorted triangular bonding arrangement. Both pictures were adapted from the work [14].

a clearer understanding, see figure 2.1 illustrates the crystal structure of  $\text{Cs}_2\text{CoBr}_4$ . Specifically, examining a cross section in the  $bc$  plane reveals its effective two-dimensional structure that resembles a triangular lattice. The primary contributors to the energy characterization are the spin- $\frac{3}{2}$  cobalt ions, which introduce anisotropy to the system because of their large spin-orbit couplings. The spin- $\frac{3}{2}$  Hamiltonian for the cobalt spins can be written as

$$\hat{H}_{3/2} = \sum_{i,j} D \left[ \left( \hat{S}_{2i,j}^z \right)^2 + \left( \hat{S}_{2i+1,j}^x \right)^2 \right] + J_1 (\hat{\vec{S}}_{i,j} \cdot \hat{\vec{S}}_{i,j+1}) + J_2 (\hat{\vec{S}}_{i,j} \cdot \hat{\vec{S}}_{i+1,j}) + J_3 (\hat{\vec{S}}_{i,j} \cdot \hat{\vec{S}}_{i+1,j+1}). \quad (2.1)$$

Here,  $D$  represents an anisotropy constant,  $J_1$  denotes the dominant interaction within chains aligned

## 2. Theory

along the  $b$ -axis, and  $J_2$  indicates the weaker interaction strength that forms a zig-zag connection, arranging the chains into a triangular configuration. Due to the fact that  $D \gg J_1, J_2$  the lower doublet  $|\pm 1/2\rangle$  of each spin is significantly distanced from the higher doublet  $|\pm 3/2\rangle$ . This makes it possible to construct an effective low-energy Hamiltonian to discuss dynamics at low temperatures. Basically, we introduce pseudospins with  $s = \frac{1}{2}$  at every site to represent the lower doublet  $|\pm 1/2\rangle$  and project out the high-spin doublet  $|\pm 3/2\rangle$ . This is achieved by the Schrieffer-Wolff transformation [30] where to zeroth order replace the spin  $\frac{3}{2}$  operators through spin- $\frac{1}{2}$  ones in the following way:  $\hat{S}^{x,y} \rightarrow 2\hat{S}^{x,y}$ ,  $\hat{S}^z \rightarrow \hat{S}^z$  for the even chains and  $\hat{S}^{z,y} \rightarrow 2\hat{S}^{z,y}$ ,  $\hat{S}^x \rightarrow \hat{S}^x$  for the odd chains. The resulting Hamiltonian has the form

$$\hat{H} = \sum_{i,j} \sum_{\alpha=x,y,z} J_e^{\alpha\alpha} \hat{S}_{i,2j}^{\alpha} \hat{S}_{i+1,2j}^{\alpha} + J_o^{\alpha\alpha} \hat{S}_{i,2j+1}^{\alpha} \hat{S}_{i+1,2j+1}^{\alpha} + J'^{\alpha\alpha} \left( \hat{S}_{i,2j}^{\alpha} [\hat{S}_{i,2j+1}^{\alpha} + \hat{S}_{i,2j-1}^{\alpha}] + \hat{S}_{i,2j}^{\alpha} [\hat{S}_{i+1,2j+1}^{\alpha} + \hat{S}_{i-1,2j-1}^{\alpha}] \right), \quad (2.2)$$

with  $i, j$  enumerating the lattice sites within. Here the  $J$ 's are diagonal interaction tensors of the form

$$J_e^{\alpha\alpha} = J_1 \cdot (1, 1 + \delta, 1 - \Delta), \quad J_o^{\alpha\alpha} = J_1 \cdot (1 - \Delta, 1 + \delta, 1), \quad J'^{\alpha\alpha} = J_2 \cdot (1 - \delta', 1, 1 - \delta') \quad (2.3)$$

From experiment [14], [15], those parameters were determined to be  $\Delta_{xy} = 0.75$ ,  $\delta = 0.1$ ,  $\delta' = 0.5$ . Furthermore in the supplementary material of the work [15] they estimate the interaction couplings to be  $J_1 \approx 0.8$  meV and  $J_2 \approx 0.35$  meV through linear spin wave calculations. In other words, the interchain coupling is almost half as strong as the in-chain one. In the calculations carried out within the scope of this thesis, we set  $J_1 = 0.8$  and  $J_2 = 0.35$ . Figure 2.2 displays a graphical illustration of the Hamiltonian, highlighting the easy plane/axis, which indicates the area where the spins are inclined to align.

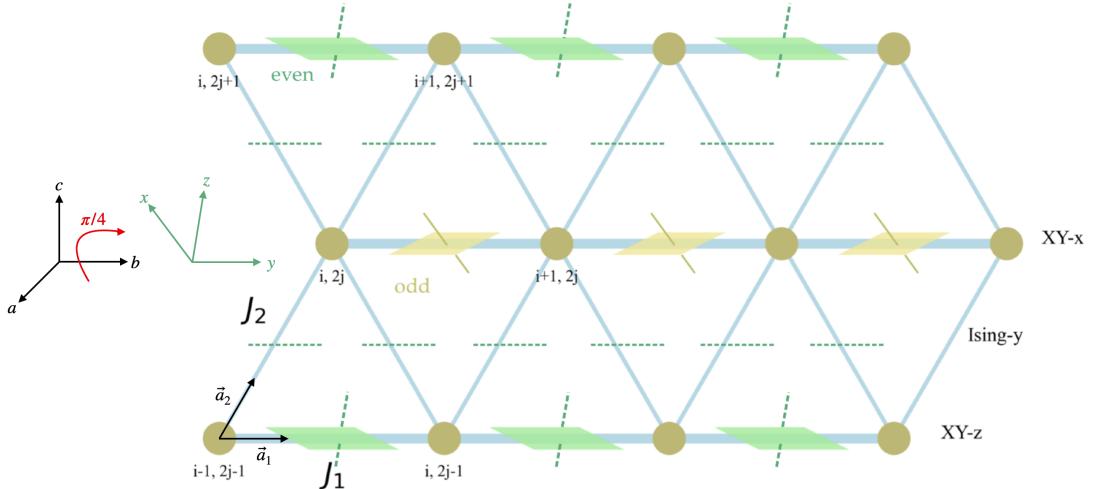


Figure 2.2.: Illustration of the effective Hamiltonian on a triangular lattice of  $\text{Cs}_2\text{CoBr}_4$ . The locations of the cobalt ions are marked by beige dots. Dashed lines and planes denote the easy axes and planes corresponding to the anisotropies in the Hamiltonian in Eq (2.2). The  $abc$  system refers to the structural coordinate framework used to describe atom positions, while the  $xyz$  coordinate framework denotes the axes related to the spin operators, rotated by  $\pi/4$  relative to each other. Vectors  $\vec{a}_1$  and  $\vec{a}_2$  act as basis vectors that span the lattice. Furthermore, the variation in line thicknesses represents the different coupling strengths, with  $J_2/J_1 \approx 0.5$  being highlighted.

The presented Hamiltonian in Eq. (2.2) has a Heisenberg-like structure on a triangular lattice with present anisotropies. However, unlike the XXZ model or the XYZ model, our model does not have a unique anisotropy axis. Along the chains, we have couplings with strong XY-nature, with easy plane

## 2. Theory

alternating between the chains. This part is described by the first two terms in Equation (2.2). The alternating pattern between the chains is evident from the crystal structure 2.1: within the  $bc$ -plane, the tetrahedral substructures align in chains, yet exhibit a different orientation in three-dimensional space. These chains are also referred to as even and odd chains.

In contrast, the interchain interactions, also called zig-zag bonds, are Ising-like with an easy axis collinear to the structural b-axis, which coincides with the y-axis. These interactions are described by the third term in Equation (2.2).

In other words, this system has two ongoing frustration mechanisms: one deriving from not having a unique anisotropy axis and the other coming due to the non-bipartite nature of the triangular lattice (geometrical frustration). These two frustrations result in a complicated interplay within the material. Such frustration mechanisms are relatively well explored for a perfect triangular lattice [31], as well as for less symmetric cases, such as the anisotropic Heisenberg model on a triangular lattice described by a set of chains with couplings  $J_1$  ( $J_2$ ) along (in-between) chains. However, within the community, opinions vary regarding numerical computations. For example, at  $J_2/J_1 \approx 0.5$ , certain studies suggest the presence of a spin-liquid ground state [28], whereas others advocate a coplanar spiral ground state [32].

For the case of  $\text{Cs}_2\text{CoBr}_4$ , the authors of [14, 15] found interesting physics such as the coexistence of two magnetic plateaus [14] and the observation of an incommensurate longitudinal SDW phase and its locking into an UUD plateau state [15]. This unique frustration mechanism has yet to be theoretically explained, and further investigation could possibly result in discovering new magnetic excitations. The primary focus of our forthcoming work will be the quantum magnet, especially the Hamiltonian represented in Eq. (2.2). We aim to evaluate its ground state in the absence of a magnetic field and at zero temperature. The methodology for achieving this is explained in the following section.

### 2.3. Variational Monte Carlo

In variational Monte Carlo (VMC) a trial wave function  $\psi_\theta$  is optimized to minimize the energy expectation value. Using the Rayleigh-Ritz variational principle,

$$E_\theta = E[\psi_\theta] = \frac{\langle \psi_\theta | \hat{H} | \psi_\theta \rangle}{\langle \psi_\theta | \psi_\theta \rangle} \geq E_0 \quad (2.4)$$

the exact ground state energy  $E_0$  can be seen as a lower boundary and the time-independent Schrödinger equation can be reformulated as a stochastic optimization problem. By varying the wave function  $\psi_\theta$  we want to find the parameters  $\tilde{\theta}$  which minimize the energy functional

$$E_{\tilde{\theta}} = \min_{\theta} \frac{\langle \psi_\theta | \hat{H} | \psi_\theta \rangle}{\langle \psi_\theta | \psi_\theta \rangle} \geq E_0. \quad (2.5)$$

In the case where the dimension of the Hilbert space is too large, the expectation value in (2.4) can be evaluated using Monte Carlo. Consider an observable  $A$  and inserting identities  $\mathbb{1} = \sum_\sigma |\sigma\rangle \langle \sigma|$  we can rewrite the expectation value to

$$\langle \hat{A} \rangle = \frac{\langle \psi_\theta | \hat{A} | \psi_\theta \rangle}{\langle \psi_\theta | \psi_\theta \rangle} = \frac{\sum_\sigma |\psi_\theta(\sigma)|^2 A_L(\sigma)}{\sum_{\sigma'} |\psi_\theta(\sigma')|^2} = \sum_\sigma p_\theta(\sigma) A_L(\sigma) \approx \frac{1}{N_s} \sum_{i=1}^{N_s} A_L(\sigma_i), \quad (2.6)$$

where we have defined the so-called local observables  $A_L(\sigma) = \langle \sigma | \hat{A} | \psi_\theta \rangle / \psi_\theta(\sigma)$  and the probability distribution  $p_\theta(\sigma) = |\psi_\theta(\sigma)|^2 / \sum_{\sigma'} |\psi_\theta(\sigma')|^2$ .

The last equality estimates the sum over exponentially many basis states using  $N_s$  Monte Carlo samples. With a sufficiently large number of samples, the estimate will converge to the mean, as a direct consequence of the central limit theorem [33].

For the average in Eq.(2.6) we need to compute the overlaps  $\langle \sigma | A | \psi_\theta \rangle$ . We can rewrite the local ob-

## 2. Theory

servable to the following

$$A_L(\sigma) = \sum_{\sigma'} \langle \sigma | \hat{A} | \sigma' \rangle \frac{\psi_\theta(\sigma')}{\psi_\theta(\sigma)}. \quad (2.7)$$

Now we see that the Monte Carlo estimation is only useful if the number of connected elements  $\langle \sigma | \hat{A} | \sigma' \rangle$  that are nonzero, grows at most polynomial with the number of degrees of freedom[34]. An operator of special interest for us is the local Hamiltonian  $H$  of a system. The associated local observables  $E_L(\sigma) = \langle \sigma | \hat{H} | \psi_\theta \rangle / \psi_\theta(\sigma)$  are called local energies.

In physics, many Hamiltonians of interest feature local interactions, including the tight-binding model and Hubbard model for fermionic systems, as well as the Ising or Heisenberg models for spin systems. In these scenarios, the number of connected components in the Hamiltonian increases linearly with the degrees of freedom, making variational Monte Carlo a suitable method for these cases.

We use the package NetKet [6] to perform VMC simulations, focusing on variational ground state searches. NetKet is written in Python and JAX, rendering it extremely fast and GPU-accelerated. It is also highly flexible, as variational ansätze can be defined easily using Flax's linen module.

### 2.3.1. Markov Chain Monte Carlo Sampling

To evaluate expressions for the expectation values such as in Eq. (2.6),  $N_s$  configurations are drawn according to some probability distribution  $p(\sigma)$  using Markov chain Monte Carlo (MCMC) methods. In the case of VMC,  $p(\sigma)$  is the Born probability distribution  $|\psi_\theta(\sigma)|^2$ . A popular choice for an MCMC method is the Metropolis-Hastings algorithm [35]. Belonging to the family of Markov chain Monte Carlo methods, the algorithm creates a sequence of configurations where each sample only depends on the previous one.

Assuming ergodicity and detailed balance we start from an initial configuration  $\sigma$ , a new configuration  $\sigma'$  is proposed with transition probability  $g(\sigma \rightarrow \sigma')$  and accepted as the next sample with the acceptance probability [35]

$$A(\sigma \rightarrow \sigma') = \min \left( 1, \frac{|\psi_\theta(\sigma')|^2 g(\sigma' \rightarrow \sigma)}{|\psi_\theta(\sigma)|^2 g(\sigma \rightarrow \sigma')} \right). \quad (2.8)$$

If it gets rejected, the previous sample is taken again. In practice, the acceptance or rejection of a newly proposed sample is determined by generating a random number  $u \in [0, 1]$ . The sample is accepted if  $u \leq A(\sigma \rightarrow \sigma')$  and rejected if  $u > A(\sigma \rightarrow \sigma')$ . Since we consider the quotient of the wave functions in the acceptance probability, no normalization is needed.

The initial sample of a Markov chain is typically drawn from a random or arbitrary configuration, which does not follow the target probability distribution  $p(\sigma)$ . As a result, the early stages of the sampling, commonly referred to as the burn-in or warm-up phase, do not produce statistically reliable samples. During this transient period, the Markov chain has not yet reached equilibrium, and consequently, estimates of expectation values based on these early samples may be significantly biased. To counteract this bias, it is standard practice to discard a number of initial samples from each chain before collecting data for statistical analysis. This ensures that only samples drawn from the stationary distribution contribute to the final estimates.

There are many ways in which we can propose a new configuration. In the context of discrete degrees of freedom such as spins, we can simply flip a random spin to go from the initial configuration  $\sigma$  to  $\sigma'$ . Moreover, if we consider a system which conserves total magnetization, e.g. the Heisenberg model, we enforce that symmetry by flipping a randomly chosen pair of up-down spins.

Even in the case of an XYZ model (where we still have pairwise products of spin operators and no anisotropic coupling limit) we can sample by flipping two randomly chosen spins due to the Hamiltonian's nature of two-spin interactions. It is also possible to combine such update rules, which can lead to a better exploration of the configuration space.

## 2. Theory

### 2.3.2. Stochastic Gradient Descent

In order to minimize the energy in Eq. (2.5) we can use a gradient descent scheme to update the parameters  $\theta$  accordingly.

$$\theta^\mu \rightarrow \theta^\mu - \lambda \nabla_{\theta^\mu} E_\theta. \quad (2.9)$$

Here,  $\lambda$  is the so-called learning rate. Usually, it is chosen to be a constant, but it can also be decreased during the optimization to further improve the results. We now want to rewrite the energy gradient in order to be able to estimate it stochastically through Monte Carlo. Given an ansatz for the wave function  $\langle x|\psi_\theta\rangle = \psi_\theta(x)$ , which is described by the parameters  $\theta$ , we first rewrite the derivative with respect to  $\theta$  of the variational state into a more convenient form,

$$\hat{O}_\mu |\psi_\theta\rangle = \frac{\partial}{\partial \theta^\mu} |\psi_\theta\rangle = \partial_\mu |\psi_\theta\rangle = \left( \sum_x \partial_\mu \log \psi_\theta(x) |x\rangle \langle x| \right) |\psi_\theta\rangle. \quad (2.10)$$

The diagonal operator  $\hat{O}_\mu = \sum_x |x\rangle O_\mu(x) \langle x|$  consists of the local logarithmic derivatives  $O_\mu(x) = \partial_\mu \log \psi_\theta(x)$  of the wavefunction. Using the ansatz for the wave function  $\langle x|\psi_\theta\rangle = \psi_\theta(x)$  and the diagonal operator  $\hat{O}_\mu$ , the gradient of the energy with respect to the parameters  $\theta$  can be rewritten as

$$\begin{aligned} (\nabla_\theta E_\theta)_\mu &= \partial_\mu E_\theta = \partial_\mu \frac{\langle \psi_\theta | \hat{H} | \psi_\theta \rangle}{\langle \psi_\theta | \psi_\theta \rangle} \\ &= \frac{\langle \partial_\mu \psi_\theta | \hat{H} | \psi_\theta \rangle + \langle \psi_\theta | \hat{H} | \partial_\mu \psi_\theta \rangle}{\langle \psi_\theta | \psi_\theta \rangle} + \langle \psi_\theta | \hat{H} | \psi_\theta \rangle (-1) \frac{\langle \partial_\mu \psi_\theta | \psi_\theta \rangle + \langle \psi_\theta | \partial_\mu \psi_\theta \rangle}{\langle \psi_\theta | \psi_\theta \rangle^2} \\ &= \frac{\langle \hat{O}_\mu^\dagger \hat{H} \rangle_\theta + \langle \hat{H}^\dagger \hat{O}_\mu \rangle_\theta}{\langle \psi_\theta | \psi_\theta \rangle} - \frac{\langle \hat{H} \rangle_\theta \left( \langle \hat{O}_\mu^\dagger \rangle_\theta + \langle \hat{O}_\mu \rangle_\theta \right)}{\langle \psi_\theta | \psi_\theta \rangle^2} = 2 \operatorname{Re} \left\{ \langle \hat{O}_\mu^\dagger \hat{H} \rangle - \langle \hat{O}_\mu^\dagger \rangle \langle \hat{H} \rangle \right\} \end{aligned} \quad (2.11)$$

where  $\langle \dots \rangle$  in the last equality is the expectation value with respect to the normalized state  $\frac{|\psi_\theta\rangle}{\sqrt{\langle \psi_\theta | \psi_\theta \rangle}}$ .

In addition, we make use of the hermiticity of  $\hat{H}$ , i.e.  $\hat{H}^\dagger = \hat{H}$ .

The expression for the energy gradient in (2.11) can be estimated again by Monte Carlo sampling, and this gradient descent scheme is often referred to as a stochastic gradient descent (SGD).

If  $|\psi_\theta\rangle$  is an eigenstate of  $\hat{H}$ , i.e.  $\hat{H} |\psi_\theta\rangle = E |\psi_\theta\rangle$ , one quickly sees that the gradient is zero. This statement is true not only for the ground state with the lowest eigenvalue but also for any eigenstate. A stochastic gradient descent approach might converge to an excited state if it corresponds to a local minimum or stationary point in the energy landscape.

### 2.3.3. Stochastic Reconfiguration

Although stochastic gradient descent is widely employed, it has certain limitations. One challenge arises from the potentially highly nonlinear nature of variational ansatz wave functions, where small changes in some parameters can significantly alter the wave function, whereas modifications to others may have minimal impact. To address these issues, we explore an alternative approach proposed by Sorella [2, 36], based on the imaginary time evolution of the variational wave function, which avoids these drawbacks. The idea of imaginary time evolution is to evolve a generic state  $|\psi\rangle$  long enough to eventually converge to the ground state  $|0\rangle$ .

$$|0\rangle \propto \lim_{\tau \rightarrow \infty} e^{-\tau \hat{H}} |\psi\rangle. \quad (2.12)$$

We briefly demonstrate how this can be seen. First, we assume that our trial wavefunction is not orthogonal to the ground state, i.e.,  $\langle \psi | 0 \rangle \neq 0$  and the ground state is non-degenerate. Then we expand

## 2. Theory

our trial state in a basis of energy eigenstates of  $H$ ,  $|\psi\rangle = \sum_n c_n |n\rangle$ , leading to

$$e^{-\tau\hat{H}} |\psi\rangle = \sum_n e^{-\tau E_n} c_n |n\rangle \propto |0\rangle + \sum_{n>0} \frac{c_n}{c_0} e^{-\tau(E_n - E_0)} |n\rangle \quad (2.13)$$

where  $E_0 < E_n$ . Taking the limit  $\tau \rightarrow \infty$ , we see that the second term in Eq. (2.13) vanishes.

For the numerical implementation, we evolve the trial state in several short time steps  $\delta\tau$  and update the parameters  $\theta' = \theta + \delta\theta$  such that the overlap is maximal  $|\psi_{\theta'}\rangle \approx e^{-\delta\tau\hat{H}} |\psi_\theta\rangle$ .

We do this by maximizing their fidelity. The fidelity of two states  $\phi, \varphi$  is defined as

$$\mathcal{F}(\phi, \varphi) = \frac{\langle \phi | \varphi \rangle \langle \varphi | \phi \rangle}{\langle \phi | \phi \rangle \langle \varphi | \varphi \rangle}. \quad (2.14)$$

Expanding the expression of  $\mathcal{F}$  up to second order in  $\delta\tau$  and  $\delta\theta$  and maximizing it, we obtain the update rule

$$\delta\theta = -\delta\tau S^{-1} \nabla_\theta E_\theta. \quad (2.15)$$

where the matrix  $S_{\mu\nu} = 2 \operatorname{Re} \left\{ \langle \hat{O}_\mu^\dagger \hat{O}_\nu \rangle - \langle \hat{O}_\mu^\dagger \rangle \langle \hat{O}_\nu \rangle \right\}$  is the so-called quantum geometric tensor (QGT). For the derivation, we assumed the wavefunction to be complex, i.e.  $\psi : \mathcal{H} \rightarrow \mathbb{C}$  and real valued parameters  $\theta \in \mathbb{R}^n$ . A detailed derivation can be found in the Appendix A.1.

If we approximate the QGT by  $S \approx \mathbb{1}$ , we recover the stochastic gradient descent scheme identifying the learning rate  $\lambda$  as  $\delta\tau$ .

Inversion of the QGT is done by solving a linear system with methods such as conjugate gradient or singular value decomposition. The entire framework of variational Monte Carlo is incorporated into *NetKet* [6], a Python package developed on the JAX library.

According to the stochastic reconfiguration update, the variational parameters get shifted by

$$\delta\theta = -\tau(S + \lambda \mathbb{1}_{P \times P})^{-1} \mathbf{F} \quad (2.16)$$

where  $S$  denotes a  $P \times P$  matrix,  $\tau$  represents the learning rate, and  $\lambda$  serves as a regularization parameter ensuring the invertibility of  $S$ . Furthermore,  $\mathbf{F} = \nabla_\theta E_\theta$  indicates the energy gradient concerning the variational parameters, with  $P$  being the number of these parameters. The computational cost of this matrix inversion is expensive when the number of parameters is increased, and therefore the method is limited to a regime with a relatively small number of parameters.

However, the works [37, 38] proposed an adaptation of SR such that we can still work in a regime where the number of parameters  $P$  exceeds the number of samples  $M$ . The idea is to stochastically estimate  $F$  and  $S$ . That approach leads to the following update rule

$$\delta\theta = -\tau X(X^T X + \lambda \mathbb{1}_{2M})^{-1} \mathbf{f} \quad (2.17)$$

where  $\tau$  is the learning rate and  $\lambda$  is the regularization number. The matrix  $X = (Y_R, Y_I) \in \mathbb{R}^{P \times 2M}$  is derived through the concatenation of quantities  $Y_{\alpha i} = (O_{\alpha i} - \bar{O}_\alpha)/\sqrt{M}$  and  $\epsilon_i = -2(E_{Li} - \bar{E}_L)^*/\sqrt{M}$ , where  $O_{\alpha i}$  represent the logarithmic derivatives of the variational wave function. Additionally, the vector  $\mathbf{f} = (\epsilon_R, -\epsilon_I) \in \mathbb{R}^{2M}$  is derived from and described by the rescaled local energies  $\epsilon_i = -2(E_{Li} - \bar{E}_L)^*/\sqrt{M}$

A detailed derivation of equation (2.17) can be found in Ref. [38]. In terms of computational complexity and memory usage, we notice that in this rewritten form we have to invert a  $2M \times 2M$  matrix which is very useful in the case of  $P \gg M$ . This adaptation of SR is implemented in NetKet [6] with the name VMC\_SRt.

## 2. Theory

### 2.3.4. Accuracy of Variational States

To evaluate the precision of different variational approaches, we compare their ground-state energy expectation values. According to imaginary time evolution, the method that produces a lower energy is considered to be more accurate.

An absolute metric is advantageous as it allows for the accuracy of a method to be assessed without relying on comparisons with others. Therefore, we use it to further validate our results. For rather small systems we can evaluate the relative error [5]

$$\epsilon_{\text{rel}} = \left| \frac{E_{\text{VS}}(\theta) - E_{\text{exact}}}{E_{\text{exact}}} \right| \quad (2.18)$$

where  $E_{\text{VS}}(\theta)$  is the energy of the variational states of the current configuration of parameters and  $E_{\text{exact}}$  is the actual groundstate energy obtained through exact diagonalization or numerically exact QMC. Due to the exponential scaling of the Hilbert space dimension, this metric is only applicable to small systems. During most of the variational methods we also keep track of the variance of the energy

$$\text{Var}(E) = \langle \hat{H}^2 \rangle - \langle \hat{H} \rangle^2. \quad (2.19)$$

The variance of the Hamiltonian's expectation value becomes zero if we reach an eigenstate  $\psi$  with corresponding energy  $E$ .

$$\text{Var}(E) = \langle \hat{H}^2 \rangle - \langle \hat{H} \rangle^2 = \langle \psi | \hat{H}^2 | \psi \rangle - \langle \psi | \hat{H} | \psi \rangle^2 = \langle \psi | \psi \rangle E^2 - (\langle \psi | \psi \rangle E)^2 = 0 \quad (2.20)$$

Therefore, the variance serves as a well-suited measure to evaluate how closely the variational state approximates the a priori unknown exact ground state and, consequently, its ground state energy.

Recently, several authors proposed a new measure for variational accuracy, the so-called V-score [39]. It is a rescaled energy variance in order to make it a dimensionless quantity

$$\text{V-score} = \frac{N \text{Var}(E)}{(E - E_{\infty})^2}. \quad (2.21)$$

Here  $N$  is the number of degrees of freedom, i.e., the number of spins in our model and  $E_{\infty}$  is some reference energy.

$E_{\infty}$  is set to be the energy expectation value of a random state of the Hilbert space  $\mathcal{H}$  which can be calculated using  $\text{tr} \hat{H} / \dim \mathcal{H}$ . For quantum spin models, we often have  $E_{\infty} = 0$  since the Hamiltonian, consisting only of Pauli operators, is traceless.

The V-score is a useful metric for evaluating how well a variational wave function approximates the exact ground state. Since an ideal eigenstate has zero energy variance, a lower V-score indicates a more accurate and reliable ansatz. Furthermore, since it is normalized by the energy, it allows for fair comparisons across different system sizes and system sizes. In [39], the authors demonstrate that the V-score serves as an effective metric across a broad range of Hamiltonians and system sizes.

## 2.4. Symmetries

Exploiting the symmetries of the system and incorporating them into the variational ansatz of the wave function helps to reduce the size of the Hilbert space for sampling and also improves the accuracy [40, 41]. For instance, when conducting a ground state search for the antiferromagnetic Heisenberg model, we can take advantage of the SU(2) symmetry, which indicates that the total magnetization in the system is conserved. Moreover, for the ground state, it is known that the total magnetization equals zero [42]. This allows us to limit our sampling to the subspace where the total magnetization is zero, enhancing the quality of samples and speeding up convergence. In the case of our model, we can exploit the following symmetries.

## 2. Theory

Looking at the model in Eq. (2.2) we observe a translation symmetry with periodicity of one site length along the direction  $a_1 = (1, 0)$  and with periodicity of two site lengths along the direction  $a_2 = (1, \sqrt{3})/2$ .

Our Hamiltonian retains reflection symmetry, where the mirror axis is orthogonal to the even/odd chains. Furthermore, our Hamiltonian remains unchanged under a global spin reversal. Lastly, we have magnetization-parity symmetry, that is, the Hamiltonian in Eq. (2.2) commutes with the operator  $\mathcal{P}_M = e^{i\pi M}$  where  $M = \sum_i \hat{S}_i^z$ . You can refer to the appendix A.3 for the proof of the conservation of these quantities under the considered symmetries. For the case of magnetization parity, we can restrict the sampling to the subspace of states with even/odd total magnetization. For the case of translation and spin flip we implement the invariance by taking the following linear combination

$$\Psi_G(\sigma) = \frac{1}{|G|} \sum_{\mathcal{T} \in G} \Psi(\mathcal{T}\sigma) \quad (2.22)$$

where we take the sum over all transformations  $\mathcal{T}$  of the symmetry group  $G$ .  $\mathcal{T}\sigma$  is the spin configuration transformed by symmetry. On the one hand, it has been shown that including symmetries helps to improve the variational accuracy [40], but on the other hand, we increase the computational cost by a factor  $|G|$ .

In programming terms, these symmetries are expressed as permutations of the lattice nodes. A clever method to determine these permutations is by setting up a tailored graph using the class [6] `netket.graph.Graph`, which allows us to employ a colored graph to account for the Hamiltonian's anisotropies. Then the graphs method `graph.automorphisms()` provides the symmetry group of the graph.

## 2.5. Sign Structure

The sign structure of a quantum many-body system refers to the pattern of signs in the coefficients of the wave function when expressed in a chosen basis, such as the occupation number or the spin basis. In mathematical terms, the sign structure can be described as a unitary transformation which acts onto the wave function or the Hamiltonian. Assuming we are capable of performing exact diagonalization on any matrix, regardless of size, it follows that any Hamiltonian can be made stoquastic, as ensured by its hermitian nature. In principle, variational states can represent every sign structure, but actually learning them appears to be a big challenge, especially in frustrated systems. Moreover, imposing the sign structure for the unfrustrated limit of a system helps stabilize optimization for the frustrated case [38]. In the following subsections, we illustrate the implementation of the sign structure, known as the Marshall Peierls sign rule [43], in non-frustrated antiferromagnetic systems. Our initial focus will be on the Heisenberg model with isotropic interactions. Subsequently, we extend the principle to establish a sign configuration for an XYZ model exhibiting anisotropic interactions. Finally, we explore the application of these principles to systems that exhibit frustration.

### 2.5.1. Marshall Peierls Sign Rule

For the antiferromagnetic Heisenberg model, there is an exact result [44] for the ground state which is useful from a theoretical perspective but is also often used in numerical simulations. Let us consider the antiferromagnetic Heisenberg Hamiltonian,

$$\hat{H} = \sum_{\langle i,j \rangle} J_{ij} \hat{\mathbf{S}}_i \cdot \hat{\mathbf{S}}_j \quad (2.23)$$

where the sum runs over all sites of a d-dimensional bipartite lattice (e.g. square lattice) and  $J_{ij} > 0$  and  $\langle \dots \rangle$  denote that the sum runs only over nearest neighbors.

The Marshall-Peierls sign rule makes a statement about the signs of the coefficients of expansion of

## 2. Theory

the ground state of the Hamiltonian (2.23) in the so-called Ising basis. The Ising basis or computational basis is specified by assigning the value of  $\hat{S}_i^z$  at each lattice site, that is,  $|\sigma\rangle = \bigotimes_i |\sigma_i\rangle$  with  $\hat{S}_i^z |\sigma_i\rangle = m_i |\sigma_i\rangle$  and  $|m_i| \leq s$ , where  $s$  is the spin quantum number. We first perform a unitary transformation on the Hamiltonian,

$$\hat{\mathcal{U}}^\dagger = \exp\left(-i\pi \sum_{i \in \mathcal{B}} (s + \hat{S}_i^z)\right) \quad (2.24)$$

The physical meaning of the above unitary is to flip the z-axis on the sublattice  $\mathcal{B}$ . The transformed Hamiltonian is

$$\hat{\mathcal{U}}^\dagger \hat{H} \hat{\mathcal{U}} = - \sum_{\langle i,j \rangle} J_{ij} \left( \hat{S}_i^x \hat{S}_j^x + \hat{S}_i^y \hat{S}_j^y \right) + \sum_{\langle i,j \rangle} J_{ij} \hat{S}_i^z \hat{S}_j^z. \quad (2.25)$$

We notice that the off-diagonal part gets an overall minus sign. We conclude by the Perron-Frobenius theorem [45] that the ground state expansion over the Ising basis has real and non-negative amplitudes, i.e.,

$$|\tilde{\psi}_0\rangle = \sum_{\sigma} f_{\sigma} |\sigma\rangle \quad (2.26)$$

where  $f_{\sigma} \geq 0$ . A more comprehensive explanation for the applicability of the Perron-Frobenius theorem is available in the appendix A.2. If we return to the original representation, the ground state obeys the Marshall-Peierls sign rule, that is

$$|\psi_0\rangle = \hat{\mathcal{U}} |\tilde{\psi}_0\rangle = \sum_{\sigma} (-1)^{N(\sigma)} f_{\sigma} |\sigma\rangle \quad (2.27)$$

where  $N(\sigma) = \sum_{i \in \mathcal{B}} (s + m_i)$ . In the case of spin  $s = \frac{1}{2}$ ,  $N(\sigma)$  corresponds to the number of spins up on the sublattice  $\mathcal{B}$ . Hamiltonians like the one previously mentioned, characterized by non-positive off-diagonal matrix elements, are termed *stoquastic* [46], since their associated Gibbs density matrix remains non-negative. On a separate note: this is why QMC is particularly effective with stoquastic Hamiltonians, as we calculate the off-diagonal matrix elements  $\langle x | e^{-\beta \hat{H}} | x' \rangle$ , which must be positive because we view them as probabilities.

In contrast, for frustrated systems such as the  $J_1 - J_2$  model and the Heisenberg model on a triangular grid, this rule does not hold. In fact, no straightforward unitary transformation is known to render the off-diagonal elements of the corresponding Hamiltonian non-positive.

### 2.5.2. Generalized Sign Rule

Now let us consider the antiferromagnetic Heisenberg model with the addition of with anisotropies in the interactions for different spin directions such that

$$\hat{H}_{XYZ} = \sum_{\langle i,j \rangle} J_x \hat{S}_i^x \hat{S}_j^x + J_y \hat{S}_i^y \hat{S}_j^y + J_z \hat{S}_i^z \hat{S}_j^z \quad (2.28)$$

with  $J_x \neq J_y \neq J_z$  and  $J_{\alpha} > 0, \forall \alpha \in \{x, y, z\}$ . We again assume the model to live on a bipartite lattice. The XYZ-model can still be turned stoquastic with the previously defined transformation (2.24) if we assume  $J_x > J_y$ . To explicitly see this, we write one exchange term in matrix form

$$J_x \hat{S}_i^x \hat{S}_j^x + J_y \hat{S}_i^y \hat{S}_j^y = \begin{pmatrix} 0 & 0 & 0 & J_x - J_y \\ 0 & 0 & J_x + J_y & 0 \\ 0 & J_x + J_y & 0 & 0 \\ J_x - J_y & 0 & 0 & 0 \end{pmatrix}. \quad (2.29)$$

We see that all off-diagonal entries are positive and under the transformation in Eq. (2.24) the above matrix becomes stoquastic.

## 2. Theory

Now what about the case where  $J_x$  is smaller than  $J_y$ ? Transforming the Hamiltonian into a stoquastic form is no longer straightforward because the signs of the off-diagonal elements differ. However, if we first map  $J_x \leftrightarrow J_y$ , the Marshall sign rule still holds for the Hamiltonian. This extension can be realized through the following unitary

$$\hat{\mathcal{F}} = \bigotimes_{j \in \mathcal{L}} \hat{\mathcal{S}}_j, \quad \text{where } \hat{\mathcal{S}}_j = \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix} \quad (2.30)$$

where  $\mathcal{S}$  is the so-called S-gate which maps the operators  $\hat{S}_i^x \rightarrow \hat{S}_i^y$  and  $\hat{S}_i^y \rightarrow -\hat{S}_i^x$ . The action on the spin operators is given again through  $\hat{\mathcal{F}}^\dagger(\dots)\hat{\mathcal{F}}$ . Thus the unitary  $\hat{\mathcal{U}}_{tot} = \hat{\mathcal{U}} \cdot \hat{\mathcal{F}}$  makes the XYZ model stoquastic in the case of  $J_x < J_y$ . The generalized sign rule, respectively phase rule, is given by

$$|\psi_0\rangle = \hat{\mathcal{U}}_{tot} |\tilde{\psi}_0\rangle = \sum_{\sigma} (-1)^{N_{\mathcal{B}}(\sigma)} (i)^{N_{\downarrow}(\sigma)} f_{\sigma} |\sigma\rangle \quad (2.31)$$

where  $N_{\mathcal{B}}(\sigma)$  corresponds to the number of spins up on the sub-lattice  $\mathcal{B}$  and  $N_{\downarrow}(\sigma)$  is the number of spins down on the whole lattice  $\mathcal{L}$ .

Let's examine the generalized version of the XYZ model, which allows for varying anisotropies along different edges,

$$\hat{H} = \sum_{\langle i,j \rangle} J_x^{ij} \hat{S}_i^x \hat{S}_j^x + J_y^{ij} \hat{S}_i^y \hat{S}_j^y + J_z^{ij} \hat{S}_i^z \hat{S}_j^z, \quad (2.32)$$

with a set of  $N$  different nearest-neighbor interaction triplets  $\{(J_x^1, J_y^1, J_z^1), \dots, (J_x^N, J_y^N, J_z^N)\}$ . The question we ask, is it still possible to turn our Hamiltonian stoquastic?

Consider, for instance, an XYZ model formulated on a 2x2 lattice characterized by four distinct edges. Each edge exhibits unique coupling parameters for the three spin interaction terms. In order to discuss this we introduce the following unitary operators

$$\hat{\mathcal{H}} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \quad \hat{\mathcal{S}} = \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix}, \quad \text{and} \quad \hat{\mathcal{Q}} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix}. \quad (2.33)$$

The three operators have the following effect on the single site spin-1/2 operators

$$\begin{aligned} \hat{\mathcal{H}}^\dagger \hat{S}^{x/z} \hat{\mathcal{H}} &= \hat{S}^{z/x}, & \hat{\mathcal{H}}^\dagger \hat{S}^y \hat{\mathcal{H}} &= -\hat{S}^y \\ \hat{\mathcal{S}}^\dagger \hat{S}^{x/y} \hat{\mathcal{S}} &= \mp \hat{S}^{y/x}, & \hat{\mathcal{S}}^\dagger \hat{S}^z \hat{\mathcal{S}} &= \hat{S}^z \\ \hat{\mathcal{Q}}^\dagger \hat{S}^{y/z} \hat{\mathcal{Q}} &= \pm \hat{S}^{z/y}, & \hat{\mathcal{Q}}^\dagger \hat{S}^x \hat{\mathcal{Q}} &= \hat{S}^x. \end{aligned} \quad (2.34)$$

With them, we can map two Pauli operators onto each other while the other one gets mapped onto itself. By constructing their tensor product over all sites, that is,  $\prod_{i \in \mathcal{L}} \hat{\mathcal{U}}_i$ , for  $\hat{\mathcal{U}}_i \in \{\hat{\mathcal{S}}, \hat{\mathcal{H}}, \hat{\mathcal{Q}}\}$  we can achieve all permutations along interaction triplets simultaneously.

Therefore, *if two couplings have the same order in all triplets*, e.g.  $J_z^\alpha > J_x^\alpha$  for  $\alpha = 1, \dots, N$ , with the help of the aforementioned, we can swap them in the triplet sequence such that the first coupling is larger than the second one. In other words, we are able to make our Hamiltonian stoquastic again.

By imposing the sign rule, QMC simulations can sample configurations much more efficiently, avoiding the sign problem entirely in these systems. The Marshall sign rule is also useful in variational Monte Carlo. For antiferromagnetic systems, we can fully determine the sign structure for the wave functions, which reduces the parameter space and improves convergence.

The Marshall-Peierls sign rule cannot be directly applied to frustrated systems, as their lattice geometry or interaction patterns prevent a straightforward bipartite division. However, a clever approach can still be employed:

## 2. Theory

We can initially tune specific interaction strengths to zero to temporarily reduce the system to a bipartite lattice, where the sign structure can be reliably imposed on the wave function. From this simplified starting point, the interaction strength is gradually restored to its original value through a perturbative process. As an example, consider the system of this thesis depicted in Figure 2.2. There are two ways to recover a bipartite lattice: we set the interactions along the zig-zag bonds to zero or we nullify the interactions along the diagonal from top left to bottom right. The limits of the bipartite lattice are depicted in Figure 2.3. We abbreviate those bipartite lattice limits as  $J_2$ -, respectively,  $J_3$ -limit. Regarding the  $J_2$  limit, we have some flexibility in selecting the sublattices, as the entire lattice is made up of independent one-dimensional spin chains. For greater clarity, consider the image on the left in Fig. 2.3: we can select the sequence of nodes [0, 1, 2, 3, 8, 9, 10, 11] and its complement as sublattices, or alternatively choose [0, 1, 2, 7, 8, 9, 10, 15] and its complement. We refer to the former as the  $J_2$  limit, while the latter is termed the  $J_2$  Vers2 limit.



Figure 2.3.: Diagrams depict the resulting bipartite lattice derived from the triangular lattice. To the left, the interaction along zig-zag bonds is nullified, yielding independent 1D spin chains ( $J_2$  limit). On the right, interactions are present along the diagonal extending from the top left to the bottom right, resulting in a 2D rhomboidal lattice ( $J_3$  limit).

To ensure that the wave function can adapt to the evolving interaction landscape, it is essential to allow complex parameters in the variational state. These complex parameters provide the necessary flexibility for the initial imposed sign structure to evolve smoothly, ultimately accommodating the frustration effects present in the original system.

## 2.6. Variational States

In Section 2.3, we discussed approaches for obtaining precise estimates of ground state wave functions in quantum many-body systems. However, we did not describe the explicit form of a variational state  $|\psi_\theta\rangle$ . Since VMC does not utilize the complete Hilbert space, the success of optimization is heavily dependent on the expressivity of the variational ansatz. Therefore, selecting an appropriate variational ansatz is crucial to accurately representing the core physical characteristics of the system being considered.

To get more familiar, we first introduce the simplest and most fundamental variational states used in VMC: the mean-field product state and the Jastrow state.

### Mean-Field State

The simplest class of variational wavefunctions assumes that each degree of freedom in the system behaves independently. For spin- $\frac{1}{2}$  systems, this leads to a product state of single site wave functions,

$$|\psi_{\text{MF},\theta}\rangle = \bigotimes_i |\phi_i\rangle \quad (2.35)$$

## 2. Theory

where the local state  $|\phi_i\rangle$  can be written as  $|\phi_i\rangle = \cos \varphi_i |\uparrow\rangle + e^{i\chi_i} \sin \varphi_i |\downarrow\rangle$ . For a given spin configuration  $\sigma = (\sigma_1, \sigma_2, \dots, \sigma_N)$  the corresponding wave function amplitude reads

$$\psi_{\text{MF},\theta}(\sigma) = \phi_1(\sigma_1)\phi_2(\sigma_2)\dots\phi_N(\sigma_N). \quad (2.36)$$

The ansatz involves  $2N$  variational parameters, represented by the set  $\theta = \{\chi_1, \varphi_1, \dots, \chi_N, \varphi_N\}$ , with  $N$  indicating the number of spins in the system. Instances where the mean-field method was employed are available in the following reference [47] or the tutorial [48].

### Jastrow State

One of the most intuitive and well-established approaches in variational methods is the Jastrow factor. Initially designed for systems of continuous particles [49], this method introduces pairwise correlations among degrees of freedom on top of a non-interacting reference state. Within the context of a spin system containing  $N$  spins, the Jastrow state is formulated as

$$\psi_\theta(\sigma) = \exp \left( \sum_{i \neq j} J_{ij} \sigma_i \sigma_j \right), \quad (2.37)$$

where the variational parameters  $\theta$  are represented by the symmetric matrix  $J_{ij}$ . An illustration of Jastrow factors can be found in the work [7], where the trial wavefunction consists of a product involving a Jastrow correlator and a Gutzwiller-projected fermionic ground state wavefunction, aimed at investigating the frustrated XY model on the Kagome lattice.

In the following section, we explore a class of variational states constructed not from physical principles but inspired by machine learning techniques.

#### 2.6.1. Neural Quantum States

In order to numerically address strongly correlated quantum many-body issues, it is essential to identify compact representations of quantum states. These representations should encapsulate their key characteristics while substantially decreasing the degrees of freedom, which are otherwise exponentially increasing in Hilbert space. Achieving this involves dimensional reduction and feature extraction, and in recent years, artificial neural networks have proven to be a formidable tool for this purpose. Machine learning methods, particularly artificial neural networks, have shown remarkable success in tackling complex problems, such as image and speech recognition [50], by efficiently optimizing hundreds of thousands of parameters [51].

The authors in [5, 52] were pioneers in proposing the application of neural networks for solving quantum many-body systems, especially in parameterizing variational states.

Further studies on a wide variety of many-body systems in one- and two-dimensional dimensions have shown that highly accurate results can be obtained for the description of the ground state [9, 53–55]. Moreover, NQS are well-suited candidates for strongly correlated systems since they can represent highly entangled quantum states [56].

Alternatively, matrix product states (MPS) along with DMRG techniques can accurately capture states in one-dimensional systems. On the other hand, extending these methods to two dimensions presents challenges. Using high-rank tensor structures such as PEPS is required or a quasi-one-dimensional approach is adopted employing low-rank tensors arranged in a snake-like pattern [57]. However, recent research has successfully achieved accurate results for the Hubbard model of two dimensions employing tensor networks paired with VMC [58].

In this thesis, we will focus on variational ansätze which are based on neural network architectures.

We start by introducing the concept of a feedforward neural network, i.e. a function that has a layered structure. The input is a vector  $\vec{x}$  and the output is some vector  $\vec{y}$ . To establish the connection with

## 2. Theory

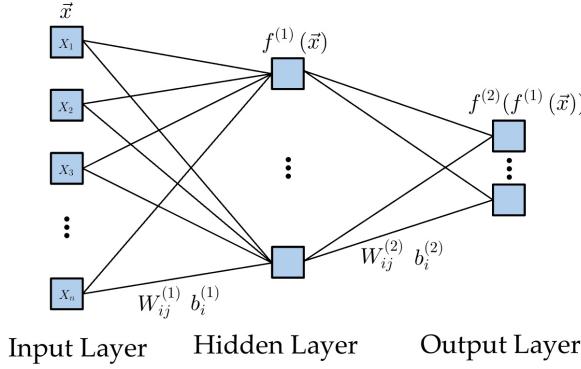


Figure 2.4.: Neural network with one hidden layer. In general every layer can have an arbitrary number of features. In the specific case of a neural quantum state  $\psi_\theta$ , the output layer consists only of a single feature, the amplitude of the wavefunction  $\psi_\theta(\sigma)$ .

physics, the input vector  $\vec{x}$  is defined by the system's degrees of freedom, such as the spin configuration. A standard layer  $s$  is given as a combination of a linear projection and a non-linear activation function.

$$f_i^{(l)}(\vec{x}) = g\left(W_{ij}^{(l)}x_j + b_i^{(l)}\right), \quad i = 1, \dots, n_l \quad (2.38)$$

where  $W_{ij}^{(l)}$  and  $b_i^{(l)}$  are called the weight matrix and the bias vector and  $g$  is the activation function. The value  $n_l$  is the number of neurons in the  $l$ -th layer, which can differ from layer to layer.

An important fact is that the activation function must be non-linear, as otherwise the whole network would collapse into a linear function and would not gain anything from adding additional layers. Typical choices for activation functions are  $\tanh x$ , the sigmoid function, or the rectified linear unit (ReLU) function [59].

The neural network is then the sequential concatenation of several layers; an example is depicted in Fig. 2.4, where  $\vec{y} = f^{(2)}(f^{(1)}(\vec{x}))$ . The depicted network consists of one input layer, one hidden layer, and an output layer.

In our case, a neural quantum state is a network architecture which takes as input the degrees-of-freedom configuration of the system (e.g. spins or particle positions) and the output layer is a single complex number, i.e. the amplitude of the wavefunction.

As anticipated, there exists a whole zoo of neural quantum states, each built upon distinct architectures. In the subsequent sections, we will introduce and examine two specific neural quantum state ansätze that were utilized throughout the thesis.

### 2.6.2. Restricted Boltzmann Machine

Another very famous shallow architecture of a neural quantum state is the restricted Boltzmann machine (RBM) introduced by Troyer and Carleo [5]. In their work, they have achieved exceptional accuracy for representing the ground states of the Ising and Heisenberg models in one- and two-dimensional dimensions as well as accurate dynamical descriptions of the transverse-field Ising model. Furthermore, it has been shown that RBMs can accurately describe the excited states of quantum spin Hamiltonians [40, 41, 60], where existing numerical tools often encounter numerical difficulties. Other studies, including the work of Nomura [40], have achieved high accuracy in computing the ground state for the frustrated  $J_1$ - $J_2$  model. The RBM wavefunction not only provides a flexible representation for spin systems, but also serves as a versatile framework for both bosonic and fermionic systems [61, 62], as well as states with volume entanglement entropy [12]. The RBM, depicted in Fig. 2.5 is a neural network consisting of two layers (hidden and visible). For the case of a spin  $\frac{1}{2}$  system with  $N_s$  spins the

## 2. Theory

RBM wave function is given by

$$\psi(\sigma) = \sum_{\{h_k\}} \exp \left( \sum_j a_j \sigma_j + \sum_{j,k} W_{jk} \sigma_j h_k + \sum_k b_k h_k \right) \quad (2.39)$$

$$= \exp \left( \sum_j a_j \sigma_j \right) \times \prod_k 2 \cosh \left( b_k + \sum_j W_{jk} \sigma_j \right) \quad (2.40)$$

where  $\sigma = (\sigma_1, \dots, \sigma_{N_s})$  is the spin configuration of the systems (the visible units) and  $h_k = \pm 1 (k = 1, \dots, M)$  are the values of the hidden units. Typically, we determine  $M$  by providing the density of the layer  $\alpha := \frac{M}{N}$ .

The network parameters are  $a_j$ ,  $b_k$  and  $W_{jk}$ .  $a_j$  and  $b_k$  are called visible and hidden biases and  $W_{jk}$  is the kernel, describing the interaction between the visible and hidden layer. The network parameters are, in general, complex values that must be used in order to describe both the amplitude and phase of the wave function. Moreover, the RBM is holomorphic, meaning that it is complex differentiable throughout its domain. This property allows us to compute the complex derivative within the framework of VMC.

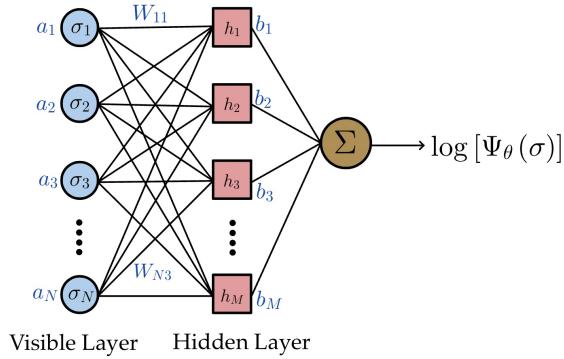


Figure 2.5.: Structure of a restricted Boltzmann machine (RBM) with  $N$  visible units  $\sigma_i = \pm 1$  and  $M$  hidden units  $h_k = \pm 1$ . The networks parameters  $\theta$  are given by the set of weights  $\{a_i, b_j, W_{ij}\}$ .

### 2.6.3. Vision Transformer

A notable architecture in machine learning is the Transformer [63], originally developed for natural language processing. Subsequently, this concept was applied to computer vision, achieving state-of-the-art results in image classification [64], thus overshadowing the previously dominant performance of CNNs. The main feature of the vision transformer (ViT) is the so-called multi-head attention mechanism. This layer creates all-to-all connections among a batch of input vectors, meaning, for spin systems, it establishes interactions between all spins within the system. These adjustable connections have the potential to replicate strong correlations or links, regardless of where they are. ViT quantum states can learn the properties of the ground state of several physical systems, such as the 1D transverse Ising model and the 1D Heisenberg  $J_1 - J_2$  model [65]. The vision transformer belongs to the family of deep neural networks. Recent works have shown that depth is essential to obtain high accuracy in 2 dimensional systems [8, 37]. However, training deep neural networks is a difficult task, and it requires the involvement of methods such as Layer Normalization [66] and Residual Connections [67]. The problem is that all of these techniques were developed for real-valued architectures and are not easily extendable to complex parameters.

The authors in Ref. [10, 38, 68] proposed different approaches for integrating complex-valued parameters. Moreover, they have achieved precise results in ground state studies, particularly in comparison to DMRG calculations for frustrated spin systems such as the 1D  $J_1 - J_2$  [10, 38] and the 2D

## 2. Theory

Shastry-Sutherland model [68]. Adjusting the architecture's hyperparameters has also led to successful outcomes on large clusters without an a priori encoding of the sign structure.

In this study, we utilize the framework presented in [68] and modify it slightly to align with the characteristics of the XYZ model on the triangular lattice. The variational ansatz consists of a composition of two functions

$$z = V(\sigma, \phi) \quad (2.41)$$

$$\log \Psi_\theta(\sigma) = f(z, \varphi) \quad (2.42)$$

where we compose the variational parameters into two sets  $\theta = \{\phi, \varphi\}$ . The function  $V(\cdot, \phi)$  denotes a real-valued deep neural network that translates the physical spin configuration  $\sigma$  into a  $d$ -dimensional vector  $z$ , which is referred to as the hidden representation. The latter is then mapped by  $f(\cdot, \varphi)$  to a single scalar. A shallow neural network [5] is used to represent the function  $f(\cdot, \varphi)$ . The single scalar output is used to estimate the amplitude of the input spin configuration. To detail both the absolute value and phase of the variational state—crucial when the sign structure is absent prior—we utilize the parameters  $\varphi$  as complex numbers.

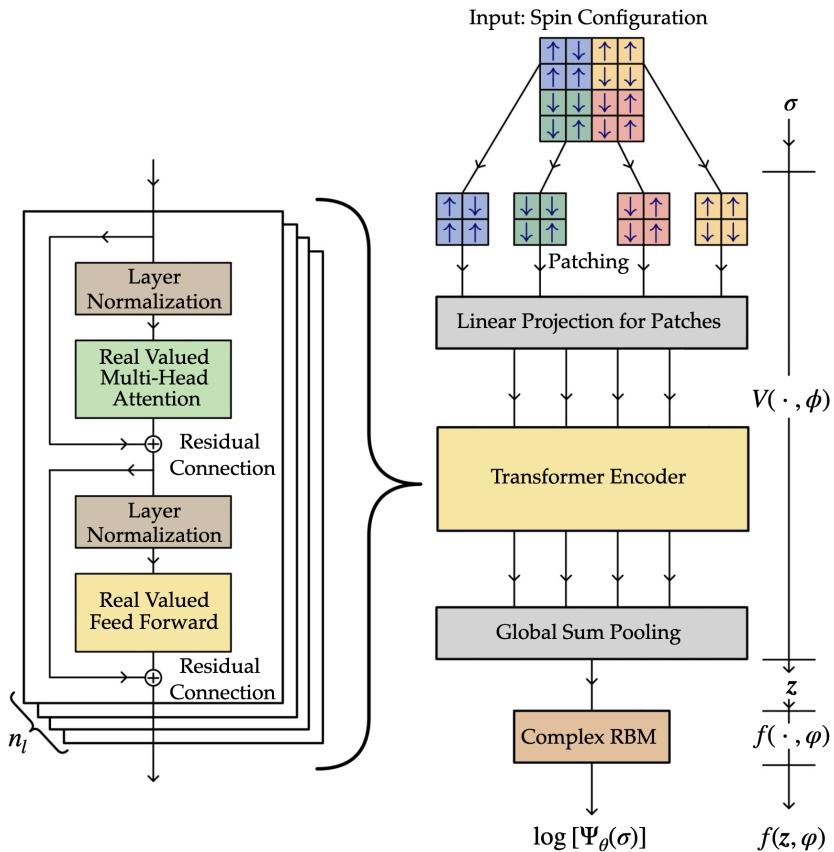


Figure 2.6.: Overview of the architecture for the vision transformer network. The network begins by dividing an input spin configuration into patches, which are then linearly embedded into a  $d$ -dimensional space. These embeddings are subsequently processed by a Transformer encoder. Finally, the output from the Transformers is aggregated and processed by a complex restricted Boltzmann machine, resulting in the computation of the wave function's logarithm.

The framework of our transformer network designed to generate the log amplitude from an input spin configuration is described as follows

1. The input spin configuration  $\sigma$  (size  $L$ ) is divided into  $n$  patches(groups) of size  $L/n$ .
2. The patches are linearly mapped to a  $d$ -dimensional space, resulting in a list of vectors  $(x_1, \dots, x_n)$

## 2. Theory

3. A transformer processing the input  $(\mathbf{x}_1, \dots, \mathbf{x}_n)$  and returning a sequence  $(\mathbf{y}_1, \dots, \mathbf{y}_n)$  where  $\mathbf{y}_i \in \mathbb{R}^d$ .
4. We sum up the sequence (global sum pooling)  $\mathbf{z} = \sum_{i=1}^n \mathbf{y}_i$ .
5. Finally, the hidden representation is passed to a complex-valued restricted Boltzmann machine, which generates the log amplitudes  $\log \Psi(\sigma)$  for the input configuration  $\sigma$ . Specifically, we choose  $\alpha = 1$ .

The core part of the Vision transformer is the transformer encoder. More generally, ViT consists of an applied sequence of  $n_l$  encoder blocks. The input of the  $l$ -th encoder is the list  $(\mathbf{x}_1, \dots, \mathbf{x}_n)$  that is normalized before being passed through the multi-head-attention mechanism.

Each  $\mu$ -th attention vector, denoted as  $\mathbf{A}_i^\mu \in \mathbb{R}^{d/h}$ , is generated by initially applying a linear transformation to  $\mathbf{x}_i$  via a local mapping given by  $V^\mu(\mathbf{x}_i) = M^\mu \mathbf{x}_i + b^\mu$  (with  $M^\mu$  representing a matrix and  $b^\mu$  as a bias). Subsequently, it is combined with the entire sequence of vectors using the attention mechanism. [63]

$$\mathbf{A}_i^\mu = \sum_{j=1}^n \alpha_{i-j}^\mu V^\mu(\mathbf{x}_j) \quad (2.43)$$

where  $\mu = 1, \dots, h$  are the number of heads. The parameters  $\alpha_{i-j}^\mu \in \mathbb{R}$  are the attention weights. In general, there would be  $n^2$  distinct  $\alpha_{ij}^\mu$  attention weights for each head  $\mu$ . However, by employing relative positional encoding for periodic systems, we reduce this number to  $n$ , focusing on attention weights  $\alpha_{i-j}^\mu$ . Thus, we establish translational symmetry between the patches.

Subsequently, a linear transformation is applied across the head dimension. Specifically, this process involves a linear combination of the attention vectors corresponding to the different heads. Then we concatenate the attention vectors along the head dimension to obtain a list of vectors  $(\mathbf{A}_1, \dots, \mathbf{A}_n)$ , where  $\mathbf{A}_i \in \mathbb{R}^d$ .

Afterwards, each attention vector  $\mathbf{A}_i$  is individually passed through a non-linearity. The nonlinear transformation is implemented as a fully connected two-layer feedforward network. Within this network, a linear projection is made in a  $2d$ -dimensional representation space, and the Gaussian Error Linear Unit (GELU) is employed as the activation function.

The output is a sequence of vectors  $(\mathbf{y}_1, \dots, \mathbf{y}_n)$  with  $\mathbf{y}_i \in \mathbb{R}^d$ . Moreover, we employ layer normalization and incorporate residual connections both prior to and following the multi-head attention mechanism and the feed-forward network. Finally, we obtain the hidden representation  $\mathbf{z} = \sum_{i=1}^n \mathbf{y}_i$  by Global Sum Pooling. We note that up to this point, all network parameters should be regarded as real-valued.

Finally, the hidden representation is input into a customized variant of a restricted Boltzmann machine with complex parameters as described in Subsection 2.6.2, where the density  $\alpha$  is set to 1. In Figure 2.6 we show a graphical representation of the ViT architecture. The system design we employ varies from the architecture described in work [68] by incorporating a layer normalization on the input  $\mathbf{z}$  and by adding a visible bias based on the configurations  $\mathbf{z}$ . It is well known that arbitrary initializations can slow down or even completely stall the convergence process [69]. Thus, for initializing the parameters  $\theta$ , we employ the method known as Xavier or Glorot initialization [70] for the transformer's parameters and a variance scaling with scale = 0.01 for the parameters of the RBM. In particular, the biases are set to 0, and the weights of the kernel are uniformly sampled from an interval that depends on the dimensions of the input and output layers.

# Results & Discussion

---

## 3.1. One Dimensional Case

We start by studying and numerically confirming the generalized sign rule described in Subsection 2.5.2. We examine a general XYZ model applied to a spin chain that has periodic boundary conditions, where the anisotropy is consistent across all edges. The anisotropies are selected based on the interactions of the even chains in our physical system, denoted as  $J_e^{\alpha\alpha}$ , as referenced in Eq.(2.3). Additionally,  $J_y > J_x > J_z$ , which means that the Hamiltonian in its current form does not become stoquastic using the Marshall-Peierls sign rule. However, by initially swapping the coupling constants  $J_x$  and  $J_y$ , followed by applying Marshall's sign rule, the Hamiltonian can be converted to a stoquastic form. We perform ground state search on the Hamiltonian with and without prior imposed sign structure. The results are shown in Figure 3.1. We use a Vision Transformer architecture [10] as an ansatz for the variational state , similar to the one described in Subsection 2.6.3.

We employ the identical multi-head attention mechanism with translation invariant attention weights. In addition, all ViT parameters are represented as complex values in order to eventually learn the sign structure.

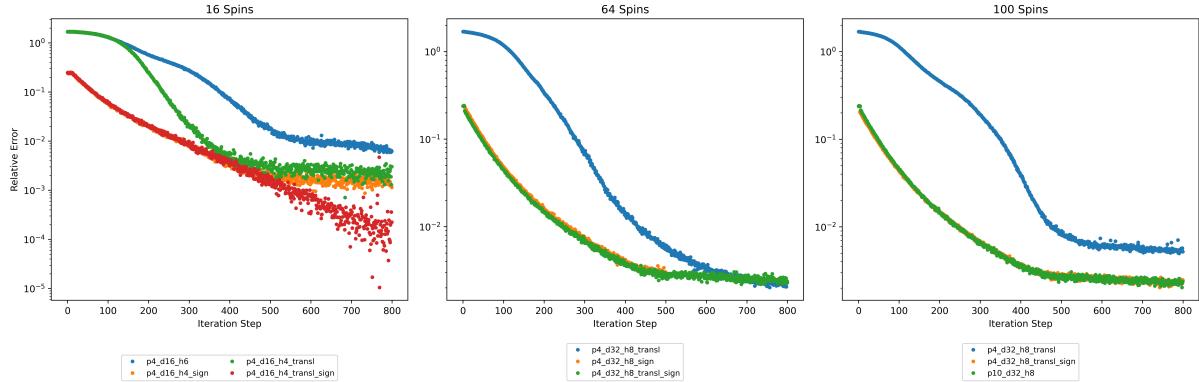


Figure 3.1.: We present plots of the ground state optimization results for the XYZ model across various system sizes. We attain a relative error approximately  $10^{-3}$  and variances all under  $2 \times 10^{-1}$ . The legend abbreviations correspond to: p for the batch size, d for the embedding dimension of the batches, h for the count of heads in the multi-head attention mechanism. Additionally, 'transl' indicates the application of full translational symmetry to the variational ansatz, and 'sign' refers to the imposed sign structure.

As a first observation, we see that the concept of a generalized structure works. However, we also observe that a sufficiently expressive ansatz, like the ViT architecture, can independently learn the sign structure. In addition, enforcing symmetries such as translation invariance and sign structure can greatly improve optimization accuracy. Furthermore, it reduces computation time, since optimizations without the sign structure took more than twice as long for any system size compared to those with the imposed sign structure. However, for one-dimensional systems, DMRG remains unmatched in terms of both precision and computational efficiency [4]. Hence, we employ it as a reference to gauge the effectiveness of our optimization results. The DMRG computations are conducted using the TeNPy

### 3. Results & Discussion

library [71]. For all calculations a max bond dimension of 200 and energy tolerance of  $10^{-10}$  were used.

## 3.2. Two Dimensional Case $\text{Cs}_2\text{CoBr}_4$

In this section we focus on the investigation on the Hamiltonian described in Section 2.2. In the initial subsection, our aim is to explore how accurately we can approximate the ground state by employing only the principles of the sign structure along with a simple network design. Then, we utilize the complete set of tools, incorporating symmetries, improved sampling techniques, and more well-established ansätze for the variational state.

As a side note, all numerical computations were performed on lattices with an even linear size, e.g. 4, since an odd linear size together with periodic boundary conditions would introduce additional artificial frustration for antiferromagnetic models. The code used for the simulations can be found in the appendix A.4.

### 3.2.1. Perturbative Approach

The sign rule cannot be directly imposed on the Hamiltonian since it is defined on a triangular lattice, which is not of bipartite nature. By setting certain edge interactions to zero, we can recover a bipartite lattice. We consider two different sign structure limits which are depicted in Figure 2.3. Under this constraint, we can enforce the sign structure and determine the ground state using VMC. Subsequently, we slightly reintroduce the interactions, using the previously acquired parameters as initial values, and proceed with the optimization again. This iterative process is repeated until we retrieve the original values of the interaction strengths. We perform this investigation on the 4x4 lattice, where we also have access to the exact ground state. The results of the optimization are shown in Figure 3.2.

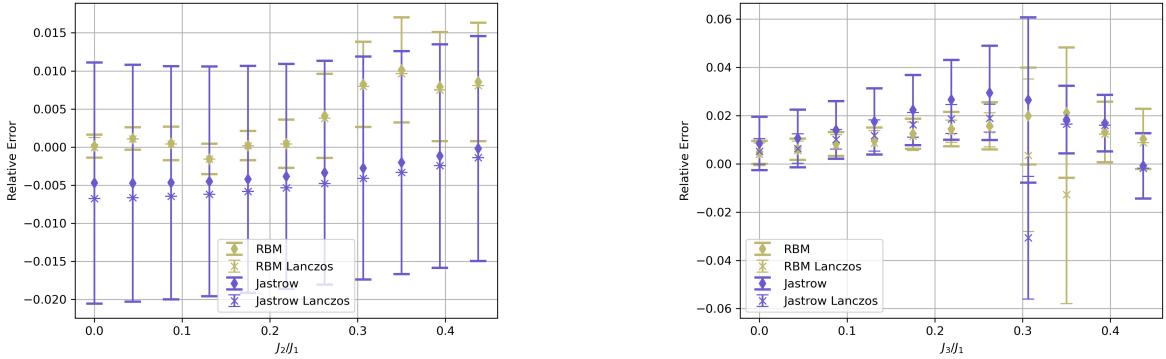


Figure 3.2.: Results from the ground state optimization via the perturbative method for the physical model on a 4x4 lattice with  $J_3/J_1 = J_2/J_1 = 0.4375$ . The relative error beginning from the  $J_2$ -Limit is depicted on the left, while starting from the  $J_3$ -Limit is shown on the right, as illustrated in Figure 2.3. Additionally, we employed a Lanczos step post-optimization [37] to enhance the ground state energies.

In the optimization process, we utilized two shallow networks as ansätze for the variational state: a Jastrow factor 2.6 and a restricted Boltzmann machine 2.6.2 with  $\alpha = 4$ . Additionally, we have not applied any symmetry constraints yet. We observe that the relative error for the physical system lies around 1%, which is already quite close considering the fact that we did not use any other physical information about the system. In general, the results are not as precise as desired because the standard error of the mean in the illustration is comparatively significant with samples  $M = 4000$ . Additionally, the variance in both directions within the physical configuration is approximately 1.0, further indicating that the variational state does not fully capture the true ground state. Despite this, it demonstrates

### 3. Results & Discussion

that the perturbative method, along with the sign structure, is highly effective in achieving a description of the ground state that is close to the true one.

#### 3.2.2. Benchmarks using RBM & ViT

In the following, we employ the full toolbox of symmetries in the optimization. First of all, we include the magnetization parity, and thus effectively reduce the space for sampling to states with even magnetization. Examining the magnetization distribution of the spin configurations within the exact ground state for the 4x4 scenario revealed that the majority of states exhibit zero magnetization. We therefore have chosen a uniform probabilistic approach for the transition rule: either swapping a pair of opposite spins or locally flipping two spins connected by an edge, each with equal probability. This combination of rules can be executed using NetKet's class `netket.sampler.rules.MultipleRules` [6].

We proceed by incorporating additional symmetries in this manner: first, we employ SR/minSR to train the variational ansatz without considering symmetries, persisting until the energy levels show stability. Next, we apply full translational symmetry and perform a global spin inversion. In the context of the vision transformer, there is significant flexibility in choosing the hyperparameters. We have set them to the following values:  $d = 24$ ,  $h = 6$ ,  $n_l = 1$ , and  $\alpha = 1$ . Furthermore, we introduce a method to incrementally enhance the diagonal shift in the QGT across several iterations, transitioning eventually to optimization via stochastic gradient descent.

#### Lattice with linear Size $L = 4$

For the  $4 \times 4$  lattice scenario, we compare the optimization results with exact diagonalizations, as outlined in Table 3.1.

ansatz	# parameters	energy	relative error	fidelity	magnetization M
rbm $\alpha = 1$	288	-25.999(4)	$4.5 \times 10^{-6}$	0.99946	$2(3) \times 10^{-4}$
rbm $\alpha = 2$	560	-25.994(2)	$1.87 \times 10^{-4}$	0.99953	$1(7) \times 10^{-4}$
rbm $\alpha = 4$	1104	-25.996(2)	$1.0 \times 10^{-4}$	0.99958	$2(8) \times 10^{-4}$
vit 12	4482	-25.91(1)	$3.3 \times 10^{-3}$	0.9916	$8(7) \times 10^{-4}$
vit 12, SS	4482	-25.966(2)	$1.3 \times 10^{-3}$	0.9942	$8(6) \times 10^{-4}$
vit 22	4506	-25.965(2)	$1.3 \times 10^{-3}$	0.9942	$6(5) \times 10^{-4}$
vit 22, SS	4506	-25.965(2)	$1.3 \times 10^{-3}$	0.9940	$8(8) \times 10^{-4}$

Table 3.1.: Overview of the precision of various variational state ansätze. The abbreviations listed in the left column have the following definitions: SS: prior imposed sign structure, xy refers to clusters of the 2D lattice; for instance, 12 signifies that the lattice is divided into batches containing one node in the x-direction and two nodes in the y-direction. The relative error is determined using Eq.(2.18). Moreover the fidelity is defined as  $\mathcal{F} = \frac{\langle \chi | \psi \rangle \langle \psi | \chi \rangle}{\langle \chi | \chi \rangle \langle \psi | \psi \rangle}$ .

To facilitate a comparison, we calculate the relative error and the fidelity. From the table, we infer that our variational ansätze effectively capture the ground state of 16 spins with high accuracy. Additionally, the RBM architecture, although utilizing fewer parameters, achieves lower and more precise energy values, as well as a greater overlap with the exact ground state compared to the ViT architecture. Moreover, we see that the variational state is able to learn the ground state with and without imposing a sign structure. Furthermore, we observe that with the smallest feasible batch size, the transformer architecture demonstrates the poorest convergence. This occurs because excessive symmetry is imposed, leading the state into a suboptimal parameter region, thereby hindering accurate results. Therefore, we can infer that initiating with fewer symmetries and subsequently applying all available symmetries during optimization tends to enhance convergence. Subsequent to optimization, we further incorporate reflection by calculating the symmetric sum.

We also note that avoiding patching (essentially using one large patch equivalent to the entire spin configuration) results in poor convergence. Therefore, we typically implement a patch size that does not

### 3. Results & Discussion

exceed a quarter of the spin configuration. Furthermore, we seek to illustrate that our improved wave function not only yields accurate energy measurements but also effectively captures other observables. To this end, we calculate the overall magnetization, the spin-spin correlation across all distances, and the corresponding Fourier transform: the static spin structure factor,

$$\hat{\mathbf{S}}_i \cdot \hat{\mathbf{S}}_j = (\hat{S}_i^x \hat{S}_j^x + \hat{S}_i^y \hat{S}_j^y + \hat{S}_i^z \hat{S}_j^z), \quad \text{and} \quad S(\mathbf{k}) = \frac{1}{L^2} \sum_{i,j} e^{i\mathbf{k} \cdot (\mathbf{r}_i - \mathbf{r}_j)} \langle \hat{\mathbf{S}}_i \cdot \hat{\mathbf{S}}_j \rangle. \quad (3.1)$$

In particular, we compute  $\hat{\mathbf{S}}_0 \cdot \hat{\mathbf{S}}_r$  where  $r \neq 0$  is an arbitrary node in the lattice. From symmetry considerations 2.4, we know that the expected total magnetization is zero. The estimates for the variational states shown in table 3.1 are all on the order of  $10^{-4}$ , which is consistent with the expectation of zero magnetization. In figure 3.3 we show the results of the spin-spin correlation. We note that the variational states accurately represent the physics across all distances. Moreover, in Figure 3.4 we show the results of the spin structure factor. From a visual perspective, it is evident that the estimates of the structure factor align closely with the trend of the exact result. In summary, we attain a precise characterization of the ground state for the  $L = 4$  case.

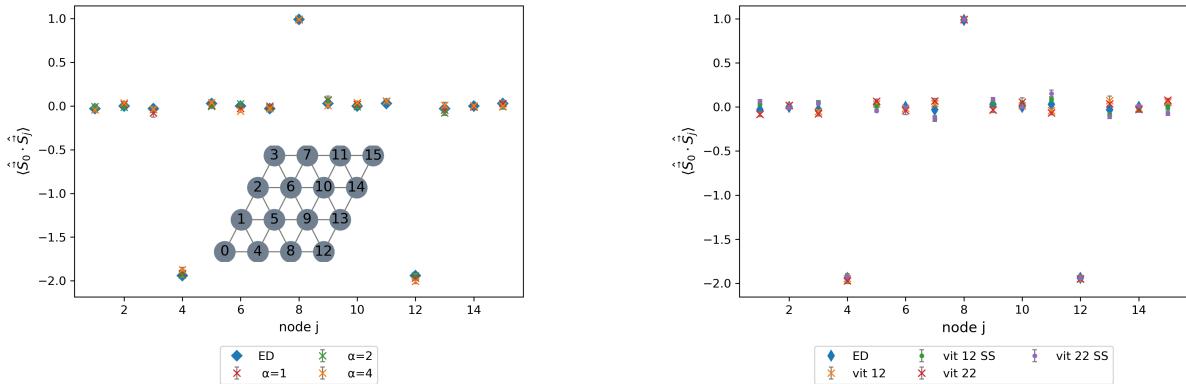


Figure 3.3.: Illustrated here are the spin-spin correlations on a  $4 \times 4$  lattice. The left panel presents the correlations for several RBM ansätze, while the right panel displays those for different ViT ansätze. The nodes are labeled based on their numbering in the inset located on the left side. It is evident that our variational states successfully replicate the correlations across all distances. We observe that we have an alternating correlation along the  $\vec{a}_1$  direction while almost no correlation along the  $\vec{a}_2$  direction.

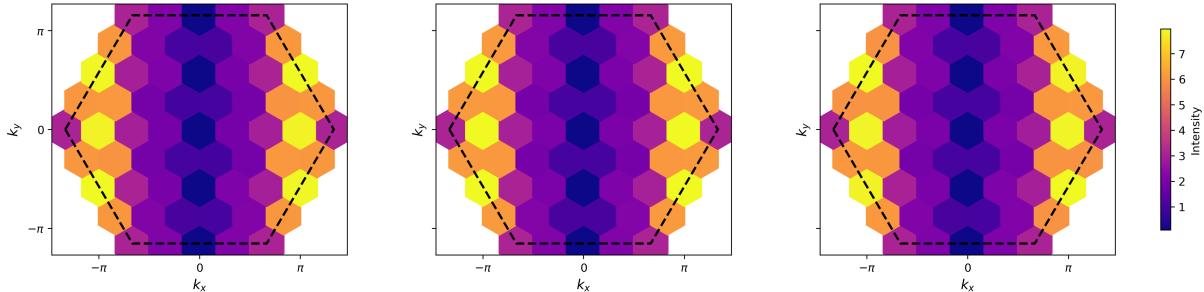


Figure 3.4.: Plots of the Spin Structure factor in the first Brillouin zone (spacing of  $\mathbf{k}$  point scales with  $\mathcal{O}(L^{-1})$ ). On the left, an RBM with  $\alpha = 1$  is utilized. In the center, the exact outcome is shown, and on the right, a ViT is employed with partitioning into cells of size  $2 \times 2$ . The black hexagonal structure represents the first Brillouin zone of the underlying lattice. We observe that  $S(\mathbf{k})$  exhibits a one-dimensional character, depending solely on  $k_x$  while being flat with respect to  $k_y$ .

### 3. Results & Discussion

#### Lattice with $L = 6, 8$

We now present the results for large system sizes, where exact solutions are no longer achievable. For architecture specifications, we again use a depth of  $n_l = 1$ ,  $d = 24$ ,  $h = 6$  and a total number of samples of 4000. In table 3.2, we summarize the results in terms of achieving the lowest energies for the  $6 \times 6$  lattice.

It is noted that the energies, variance, and V-score derived using the ViT architecture are marginally lower than those from the RBM, suggesting that ViT may yield slightly more precise results. Moreover, it is noted that starting with reduced symmetry (in this case, by selecting a larger batch size) facilitates achieving a lower energy estimate. In particular, choosing a batch size of  $1 \times 2$ , which corresponds to imposing full translation symmetry, does not yield proper convergence, which means that the energy does not approach the energies achieved by other optimizations and instead fluctuates significantly, sometimes yielding positive values. After optimization we then finally use a projector to also impose symmetry of reflection. Furthermore, we note that we get similar results with and without imposing a prior sign structure.

ansatz	# parameters	energy	variance	V-score	magnetization $M$
rbm $\alpha = 1$	1368	-54.98(3)	5.50	0.065	$6(7) \times 10^{-4}$
rbm $\alpha = 4$	5364	-54.91(2)	2.38	0.028	$1(5) \times 10^{-4}$
vit 22, $n_l = 1$	4536	-55.03(1)	0.60	$7.1 \times 10^{-3}$	$3(1) \times 10^{-4}$
vit 22, SS, $n_l = 1$	4536	-55.04(1)	0.79	$7.1 \times 10^{-3}$	$3(4) \times 10^{-4}$
vit 22, $n_l = 2$	7704	-55.05(1)	0.28	$3.5 \times 10^{-3}$	$7(1) \times 10^{-4}$

Table 3.2.: Overview of the precision of various variational state ansätze for the  $6 \times 6$  Lattice. The abbreviation *SS* refers to the sign structure imposed in the  $J_2$  limit see subsection 2.5.2. Moreover, the V-score can be determined using Eq.(2.21).

We again want to check the other spin observables. In figure 3.5 we plot the spin spin correlations and spin structure factor. We again observe strong alternating correlations along the  $\vec{a}_1$  direction, indicating antiferromagnetic order. Furthermore, the correlation  $\langle \hat{S}_i \cdot \hat{S}_j \rangle$ , for any  $i$  referring to any of the remaining nodes on the lattice, demonstrates a consistent pattern that is not illustrated here due to space limitations. Looking at the structure factor, we again observe Bragg peaks in the form of straight lines at  $k_x = \pm\pi$  indicating the a *quasi-one-dimensional order* along the  $\vec{a}_1$  direction. Moreover, we see the same behavior for the spin-spin correlation and structure factor across all variational states.

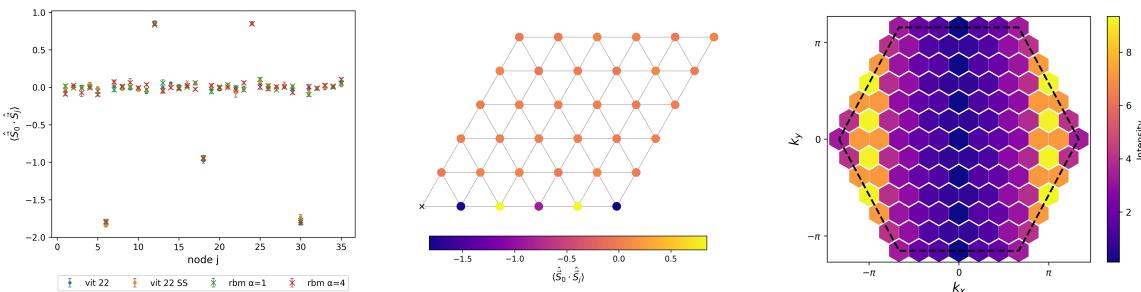


Figure 3.5.: An overview of the  $6 \times 6$  Lattice Spin Observables is provided. The left panel displays the real-space spin-spin correlation  $\langle \hat{S}_0 \cdot \hat{S}_j \rangle$ , which describes the interaction between the spin at node 0 and each other node  $j$  across different variational states. In the middle and right sections, the spin-spin correlation and spin structure factor for the vision transformer with  $2 \times 2$  patches on the lattice are displayed. A significant correlation is also observed along the direction of  $\vec{a}_1$ .

Observing this one-dimensional ordering in the two-dimensional system allows us to compare it with the one-dimensional spin chain. Figure 3.6 demonstrates a comparison of the one-dimensional

### 3. Results & Discussion

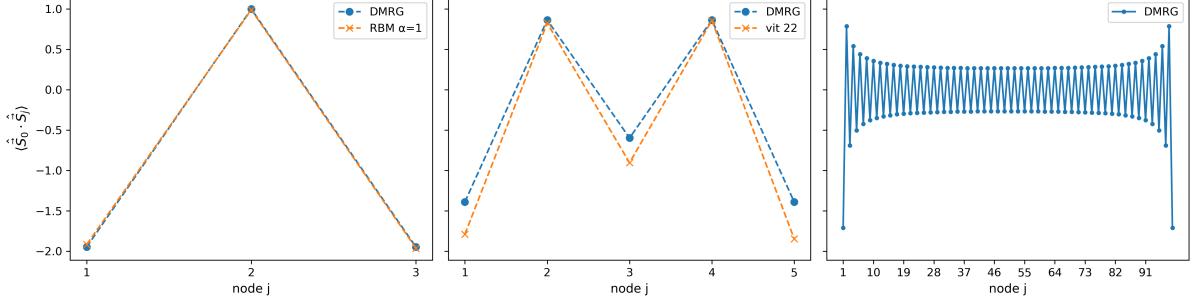


Figure 3.6.: The left and middle plots illustrate the real-space spin-spin correlations,  $\langle \hat{S}_0 \cdot \hat{S}_j \rangle$ , for the two-dimensional model's bottom chain with  $L = 4$  and  $6$ , compared alongside DMRG results that treat it as a one-dimensional chain. On the right, the correlations  $\langle \hat{S}_0 \cdot \hat{S}_j \rangle$  for a  $100$ -node spin chain are displayed. This calculation indicates the presence of quasi long-range antiferromagnetic order (algebraic decay). Our DMRG calculations are performed utilizing TeNPy [71].

features observed in the Spin structure factor by analyzing the DMRG results in a single dimension. For smaller system sizes, the spin-spin correlations exhibit a notable alignment with those observed in the one-dimensional case, thus supporting our claim that the correlations demonstrate one-dimensional behavior in the two-dimensional model.

Next, consider the lattice dimension of  $8 \times 8$ . The results of the optimization are summarized in Table 3.3. We first notice that the ViT architecture outperforms the RBM architecture in terms of having a lower energy and a smaller variance/V-score. We successfully acquired an estimate of the ground state both with and without the imposition of a prior sign structure. Additionally, it is noted that starting with reduced symmetry (in this case, by selecting a larger batch size) facilitates achieving lower energy levels. This can be seen when comparing to the data from vit 44 and vit 22. A similar behavior was observed in the case  $6 \times 6$ .

ansatz	# parameters	energy	variance	V-score	magnetization $M$
rbm $\alpha = 1$ , SS $J_2$	4224	-93.91(7)	15.5	0.11	$5(5) \times 10^{-4}$
vit 22	4578	-94.53(3)	3.77	0.027	$1(1) \times 10^{-3}$
vit 44	4794	-95.74(1)	1.47	0.0099	$3(1) \times 10^{-3}$
vit 44, SS $J_2$	4794	-95.54(3)	1.98	0.0138	$4(1) \times 10^{-3}$
vit 44, SS $J_3$	4794	-95.56(2)	1.97	0.0138	$1(2) \times 10^{-3}$
vit 44, SS $J_2$ Vers2	4794	-95.61(4)	2.45	0.0172	$2(1) \times 10^{-3}$

Table 3.3.: Summary of the accuracy of different variational state ansätze applied to the  $8 \times 8$  Lattice. The term SS indicates the use of a sign structure, while  $J_i$  denotes the specific limit being considered. For further information, refer to subsection 2.5.2. Additionally, the V-score can be determined using Eq.(2.21).

Although we obtain similar variational estimates for the ground state energy, the situation is a bit different for the spin observables. First of all, we observe the same strong alternating spin-spin correlation along the  $\vec{a}_1$ -direction, indicating the quasi-one-dimensional order. Looking at the structure factor depicted in Figure 3.7, we observe something strange: using the same variational ansatz but imposing different sign structures leads to different Bragg peaks inside the first Brillouin zone. Taking observations from smaller lattice sizes into considerations, we think that this different pattern in the structure factor is an artifact of the different imposed sign structures and the fact that the estimate of the ground state is not accurate enough. In other words, the observed peaks correspond to those when the model is considered in the bipartite lattice limit. The estimate of the spin structure factor of the ViT architecture without sign structure is closest to the results for smaller lattice sizes.

### 3. Results & Discussion

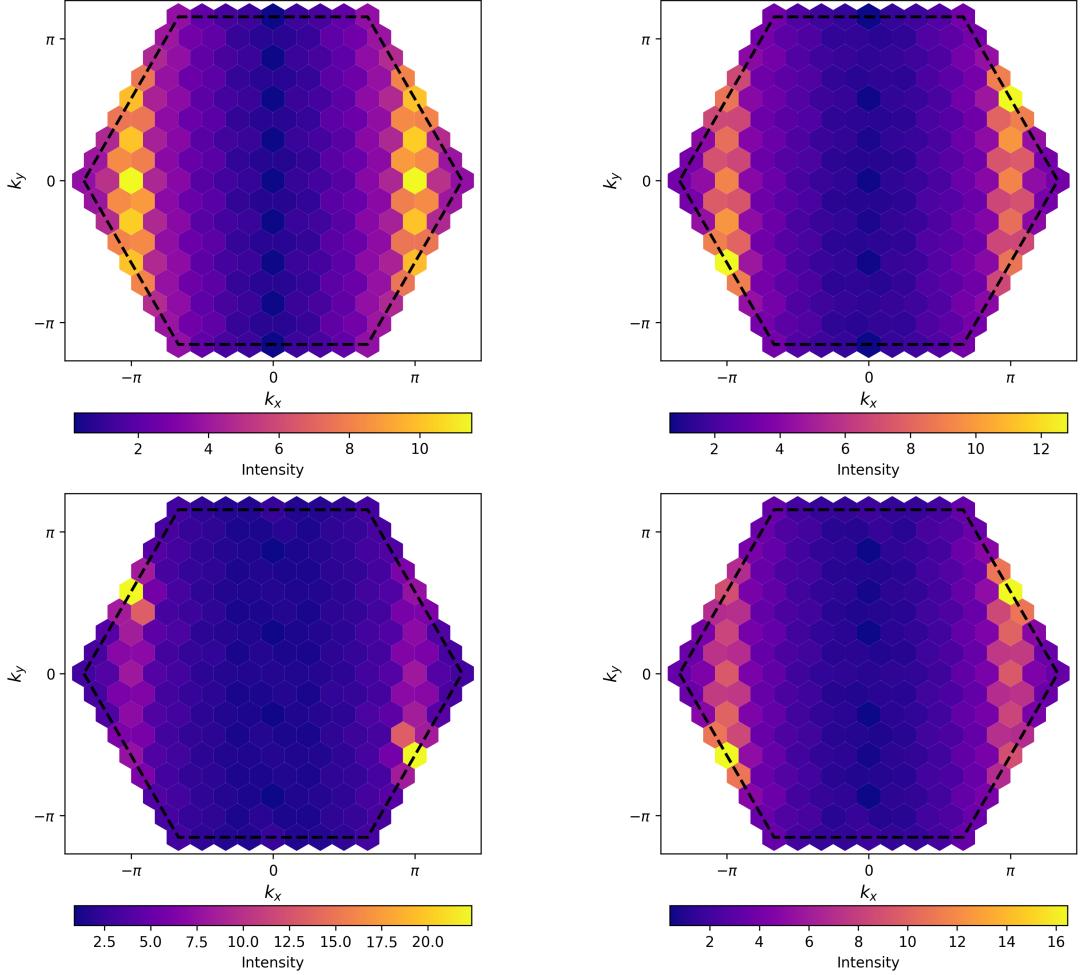


Figure 3.7.: Visualizations of the structure factor employing the ViT architecture characterized by a patch size of  $4 \times 4$ . In the top-right plot, the sign structure of the  $J_3$  limit has been applied, whereas in the bottom left and right plots, the sign structure corresponding to the  $J_2/\text{Vers 2}$  limit has been enforced. For further details, refer to subsection 2.5.2. In contrast, no sign structure was imposed on the top-left plot. We observe that the top left plot has Bragg peaks along two lines at  $k_x = \pm\pi$  where as the others show peaks at the  $M$  points of the Brillouin zone.

#### Larger Lattice Sizes

We move on to finite lattices of linear size  $L = 10, 12, 14$ . The optimization results are summarized in Table 3.4.

We have slightly adapted the architecture of vit by changing the embedding dimension to  $d = 32$  and increasing the number of heads to  $h = 8$ . Firstly, it is important to note that for every result, we enforced a prior sign structure. Attempts to proceed without such a structure failed to converge or stagnated at a value marked by extremely high variance.

We again see that the ViT architecture provides energies lower than those of the RBM. Moreover, we obtain similar variational energies for different imposed sign structures for linear sizes  $L = 10, 12$ . Ultimately, we note that the outcome for the  $L = 14$  scenario exhibits a large variance, and despite our efforts, we were unable to decrease it; therefore, this result is excluded from further analysis.

Moving on to the spin observables we have a similar situation as in the  $L = 8$  case: we still have a clear alternating pattern along the  $\vec{a}_1$ -direction for the real space correlation, but in addition we also have some non-vanishing correlation in  $\vec{a}_2$  which does not decrease over distance. The real space correlations are depicted in Figure 3.8. In the picture of the structure factor (depicted in Figure 3.9) we observe

### 3. Results & Discussion

$L$	ansatz	# parameters	energy	variance	V-score	magnetization $M$
10	rbm $\alpha = 1$ , SS	10200	-144.41(6)	11.9	0.06	$5(6) \times 10^{-4}$
10	vit 52, SS $J_2$	8056	-146.48(3)	3.24	0.02	$1.3(7) \times 10^{-3}$
10	vit 52, SS $J_3$	8056	-146.44(4)	3.21	0.02	$2(2) \times 10^{-3}$
12	vit 44, SS $J_2$	8240	-210.12(3)	5.4	0.02	$2(2) \times 10^{-3}$
12	vit 44, SS $J_3$	8240	-210.05(4)	6.1	0.02	$2(1) \times 10^{-3}$
14	vit 72, SS $J_2$	8216	-274.9(1)	40.0	0.10	$4(4) \times 10^{-4}$

Table 3.4.: Overview of precision of various ansätze of variational states for large system sizes. The term SS indicates the use of a sign structure, while  $J_i$  denotes the specific limit being considered. For further information, refer to subsection 2.5.2.

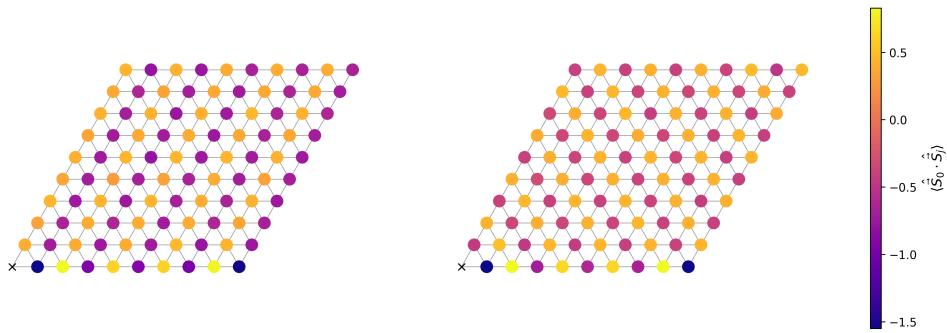


Figure 3.8.: The images depict the real-space spin-spin correlation  $\langle \hat{S}_0 \cdot \hat{S}_j \rangle$  on a  $10 \times 10$  lattice. The left image illustrates the imposed sign structure in the  $J_2$  limit, while the right image shows the imposed sign structure in the  $J_3$  limit. The bottom row clearly demonstrates the antiferromagnetic correlation diminishing with increasing distance.

again Bragg peaks at the M-points of Brillouin zone instead of this peaked lines. This is again the effect of imposing a specific sign structure and the fact that the variational state lacks the capacity to capture the true ground state, instead converging on the state within the bipartite lattice limit. The exact same behavior is observed for the variational ansätze on the  $12 \times 12$  lattice.

Finally, we aim to compare the ground state estimates across all linear dimensions. To achieve this, we graph the V-scores of the lowest achieved energies, indicating whether they were obtained using an RBM (denoted by dot markers) or a ViT (indicated by cross markers) ansatz. These results are depicted in Figure 3.10. We find that, aside from the  $4 \times 4$  scenario, the ViT architecture produces a lower V-score compared to the RBM ansatz. This result is interesting because the ViT architecture builds upon the RBM structure by integrating the transformer. We hypothesize that the transformer component effectively encodes the spin input configuration into a more informative hidden representation, thereby enabling the RBM-based final layer of the ViT to better approximate the ground state. Therefore, we speculate that the transformer efficiently translates the spin input configuration into a more manageable hidden representation, facilitating the last layer of the ViT (an RBM) in more easily learning the ground state.

Moreover, we show the available results of the V-score for the Heisenberg model on the triangular lattice. We see that these V-scores are significantly higher than for our physical model under investigation, which might suggest increased frustration in the Heisenberg model resulting from the uniformity of coupling strengths.

### 3. Results & Discussion

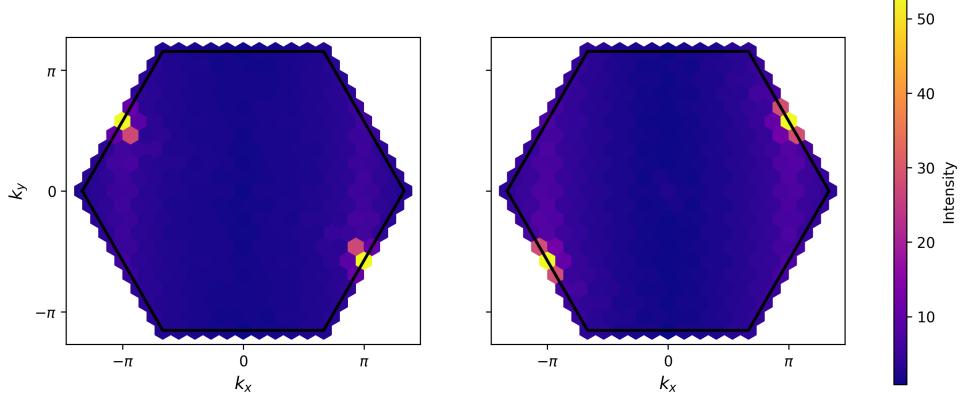


Figure 3.9.: The figures depict the spin structure factor within the first Brillouin zone for a linear size of  $L = 10$ . The left figure applies the sign structure within the  $J_2$  limit, whereas the right figure presents it in the  $J_3$  limit on the variational wave function. The variational state consists of a ViT featuring a patch size of  $5 \times 2$ .

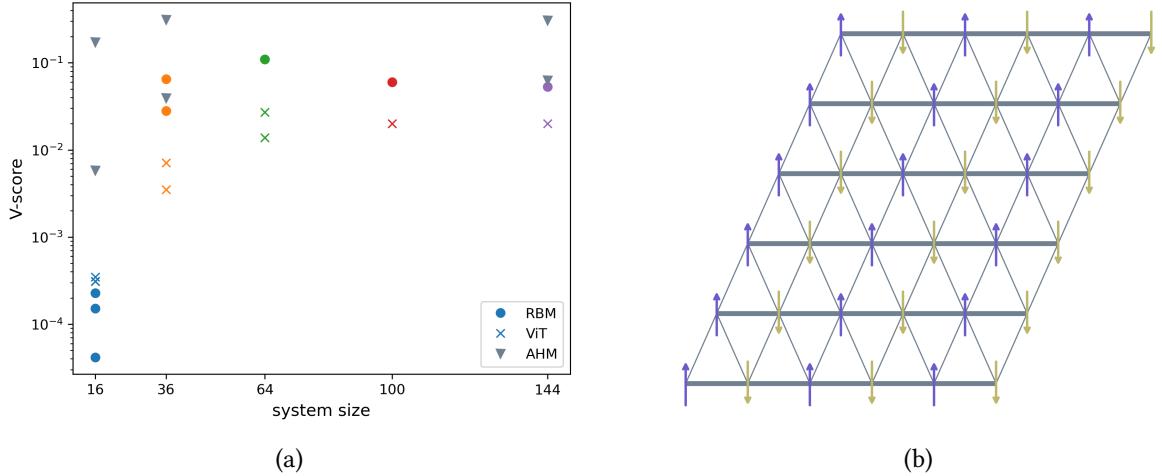


Figure 3.10.: (a) Graph displaying V-scores for the smallest energies achieved over various lattice sizes, utilizing either the RBM architecture (dot marker) or the ViT model (cross marker) as the variational ansatz. Triangular markers indicate the V-scores associated with the triangular Heisenberg model when employing various variational ansätze, as reported in the study [39]. (b) Sketch of possible order arrangement on the triangular lattice emerging from the real space spin-spin correlations of our determined ground state representation.

#### 3.2.3. Finite Size Scaling

The variational energy provides us with insight into the accuracy of the simulation on finite lattice sizes. However, it does not say much about the actual ground state of the material. We estimate the ground state energy per spin in the thermodynamic limit using the finite size behavior of the antiferromagnetic Heisenberg model with periodic boundary conditions [72]

$$e(L) = e_\infty + \frac{a_1}{L^3} + \mathcal{O}(1/L^4). \quad (3.2)$$

We perform a linear fit for the energies with the smallest variance as a function of  $1/L^3$ . The fit is depicted in Figure 3.11. The thermodynamic limit estimate is given by:  $e_\infty = -0.363(1)$ .

To investigate the presence of one-dimensional Néel order in the thermodynamic limit, we carry out a finite size scaling for the real space correlation over longest separation in the  $\vec{a}_1$ -direction, specifically

### 3. Results & Discussion

$M_{C_{L/2}}^2 = \langle \hat{\mathbf{S}}_r \cdot \hat{\mathbf{S}}_{r+L/2 \cdot a_1} \rangle$ , and this quantity can be approximated using a linear or quadratic fit [72]

$$M_{C_{L/2}}^2 = M_\infty^2 + \frac{b_1}{L} + \frac{b_2}{L^2} + \mathcal{O}(1/L^3). \quad (3.3)$$

as a function of  $1/L$ . The fits are depicted in figure 3.11. Despite the significant uncertainties in the estimated values, they remain positive, suggesting the presence of long-range quasi-one-dimensional Néel order. A possible order is sketched in Figure 3.10.

We again draw a comparison with the one-dimensional spin chain. We analyze the scenario using a large spin chain consisting of 100 spins, as illustrated in Figure 3.6: the DMRG analysis indicates the presence of antiferromagnetic spin-spin correlations over extended distances. The comparison within the thermodynamic limit is more nuanced, as our one-dimensional model closely resembles an XXZ model (Luttinger liquid limit), which does not exhibit long-range order [73]. This is further suggested by the algebraic decay observed in the correlations in Figure 3.6.

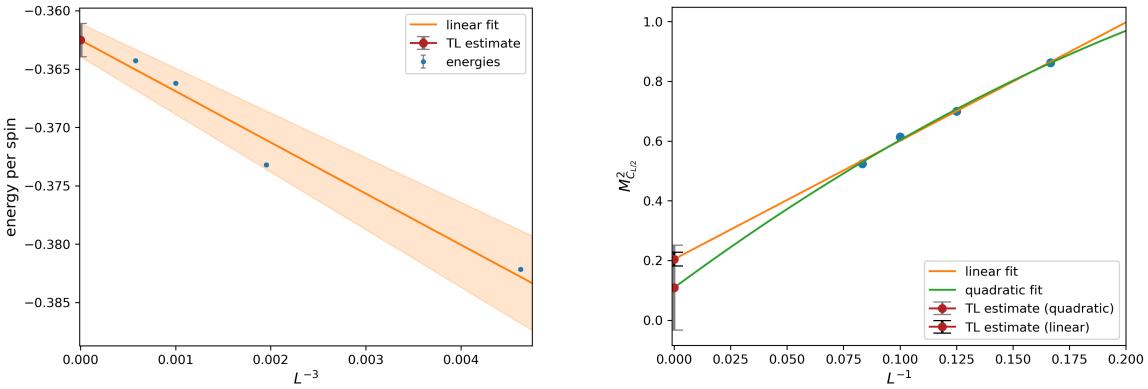


Figure 3.11.: On the left panel the size scaling of energy per spin as a function of  $1/L^3$  is shown. The energy estimate from the linear fit is given by  $e_\infty = -0.363(1)$ . On the right panel we show the size scaling of the real space correlation as a function of  $1/L$  is shown. For the linear fit we obtain  $M_{C_{L/2}} = 0.20(2)$  and for the quadratic one  $M_{C_{L/2}} = 0.1(1)$ . For both fit we use the data for linear sizes  $L = 6, 8, 10, 12$ .

# Conclusion

---

In this study, we investigate the effective spin Hamiltonian characterizing the quantum antiferromagnet  $\text{Cs}_2\text{CoBr}_4$  at low temperatures. Our approach involved neural quantum states (NQS), incorporating both restricted Boltzmann machines (RBMs) and vision transformer (ViT) architectures in order to describe the ground state of the system. It is well established that understanding the sign structure of frustrated quantum systems poses a significant challenge, which in turn complicates the variational optimization of their ground states.

To address this, we introduce a *generalized version of the Marshall-Peierls sign rule*, extending its applicability to spin Hamiltonians with anisotropic couplings. Enforcing an appropriate sign structure not only facilitates convergence during the variational optimization but can also substantially accelerate it.

For smaller system sizes, we successfully characterize the ground state with high fidelity. Interestingly, postponing the enforcement of Hamiltonian symmetries until advanced stages of optimization results in better convergence than applying full symmetry right at the beginning.

Our variational states reveal a clear *antiferromagnetic order* along the  $\vec{a}_1$  direction, while correlations along the  $\vec{a}_2$  direction are weak or near zero. This anisotropy points to a *quasi-one-dimensional ordering* of the ground state. Although the pattern resembles a stripe-like arrangement, it differs from conventional stripe order (as illustrated in Figure 3.10), due to the absence of significant correlations in the  $\vec{a}_2$  direction. This behavior is supported by the spin structure factor, which exhibits two distinct peaks at  $k_x = \pm\pi$  for system sizes  $L = 4, 6, 8$ , further indicating a quasi-1D character.

As we scale up system size to approach the thermodynamic limit, the accuracy of our variational states decreases, which is associated by increased variance and higher V-scores. Additionally, the structure factor evolves: instead of linear peak features, we observe discrete peaks at the  $M$ -point in the Brillouin zone when imposing a sign structure. We attribute this shift to a lack of expressiveness in the variational ansatz, which fails to capture the correct sign structure and instead defaults to one corresponding to the bipartite limit. The resulting Bragg peaks match those expected from the imposed sign structure, rather than from the true ground state of the model.

Despite these limitations, our finite-size scaling analysis supports the presence of a *quasi-1D* order in the thermodynamic limit. Overall, our findings indicate that neural quantum states, effectively directed by sign structure and symmetry principles, serve as a powerful tool for exploring the ground state characteristics of frustrated quantum magnets like  $\text{Cs}_2\text{CoBr}_4$ .

## Outlook

With regard to the prospects for this research area, two primary avenues can be explored. One approach involves the machine learning domain, where an unresolved question is whether employing deeper and/or broader architectures enhances the ground state description results. In addition, it would be intriguing to determine whether transitioning from a global batch correlation to a localized attention mechanism improves performance in systems of larger size.

Physically, it would be interesting to investigate phase diagrams characterized by the interaction ratio  $J_2/J_1$ , to determine the presence of distinct phases and orderings. Specifically, it would be of interest to examine the possible occurrence of a phase transition in which the material exhibits a shift from one-dimensional to two-dimensional behavior.

#### *4. Conclusion*

Furthermore, incorporating both a magnetic field and temperature would allow for a comparison with neutron scattering results [14, 15], thereby facilitating a theoretical analysis of these five experimentally predicted phase transitions.

# Derivations

---

## A.1. Stochastic Reconfiguration

### A.1.1. Fidelity Optimization

In this section we show how to derive the update rule for the parameters in stochastic reconfiguration (SR)

$$S\dot{\theta} = -\nabla_{\theta}E_{\theta}. \quad (\text{A.1})$$

We recall that stochastic reconfiguration translates the imaginary time evolution into updates of variational parameters, accounting for the geometry of the wave function manifold via the Fisher Information matrix  $S$ . We start by evolving the wave function for a short time step  $\delta\tau$ ,

$$e^{-\delta\tau\hat{H}}|\psi_{\theta}\rangle \approx |\psi_{\theta+\delta\theta}\rangle$$

and ask what small changes  $\theta + \delta\theta$  would best approximate this evolved wave function. Since the time step is small, we expand both expression above up to first order,

$$e^{-\delta\tau\hat{H}}|\psi_{\theta}\rangle = \left(\mathbb{1} - \delta\tau\hat{H}\right)|\psi_{\theta}\rangle = |\psi_{\theta'}\rangle \quad (\text{A.2})$$

$$|\psi_{\theta+\delta\theta}\rangle = \left(\mathbb{1} + \delta\theta^{\mu}\frac{\partial}{\partial\theta^{\mu}}\right)|\psi_{\theta}\rangle = \left(\mathbb{1} + \delta\theta^{\mu}\hat{O}_{\mu}\right)|\psi_{\theta}\rangle \quad (\text{A.3})$$

where we have introduced the derivative operator  $\hat{O}_{\mu} = \sum_x \partial_{\mu} \ln \psi_{\theta}(x) |x\rangle\langle x|$  and we used Einsteins summation convention. There are two approaches to determine how to update the parameters by  $\delta\theta$ . First, we determine  $\delta\theta$  by maximizing the fidelity of the states  $|\psi_{\theta'}\rangle$  and  $|\psi_{\theta+\delta\theta}\rangle$ . The fidelity of two states  $\phi, \chi$  is defined as

$$F(\phi, \chi) = \frac{\langle\phi|\chi\rangle\langle\chi|\phi\rangle}{\langle\phi|\phi\rangle\langle\chi|\chi\rangle}.$$

Keeping only terms up to quadratic order in  $\delta\theta$  and  $\delta\tau$  we obtain

$$\begin{aligned} F(\delta\theta) &= \frac{|\langle\psi_{\theta'}|\psi_{\theta+\delta\theta}\rangle|^2}{\langle\psi_{\theta'}|\psi_{\theta'}\rangle\langle\psi_{\theta+\delta\theta}|\psi_{\theta+\delta\theta}\rangle} \\ &= 1 + \delta\theta^{\mu} \left( \langle O_{\mu} \rangle + \langle O_{\mu}^{\dagger} \rangle \right) + \delta\tau \left( \langle \hat{H} \rangle + \langle \hat{H}^{\dagger} \rangle \right) - \delta\tau^2 \langle \hat{H} \rangle^2 \\ &\quad - \delta\tau\delta\theta^{\mu} \left( \langle \hat{H}^{\dagger}\hat{O}_{\mu} \rangle - \langle \hat{H}^{\dagger} \rangle \langle \hat{O}_{\mu}^{\dagger} \rangle - \langle \hat{O}_{\mu} \rangle \langle \hat{H} \rangle + \langle \hat{O}_{\mu}^{\dagger}\hat{H} \rangle \right) \\ &\quad + \delta\theta^{\mu}\delta\theta^{\nu} \left( \frac{1}{2} \langle \hat{O}_{\mu}\hat{O}_{\nu} \rangle + \langle \hat{O}_{\mu} \rangle \langle \hat{O}_{\nu}^{\dagger} \rangle + \frac{1}{2} \langle (\hat{O}_{\mu}\hat{O}_{\nu})^{\dagger} \rangle \right). \end{aligned} \quad (\text{A.4})$$

Now we need some intermediate results to simplify the above expression. Assuming the normalization of the wave function and taking derivatives

$$\text{i.) } \partial_{\mu} \langle\psi_{\theta}|\psi_{\theta}\rangle = 0 \iff \langle\partial_{\mu}\psi_{\theta}|\psi_{\theta}\rangle + \langle\psi_{\theta}|\partial_{\mu}\psi_{\theta}\rangle = 0, \text{ i.e., } \text{Re} \langle\hat{O}_{\mu}\rangle = 0.$$

$$\text{ii.) } 0 = \partial_{\mu}\partial_{\nu} \langle\psi_{\theta}|\psi_{\theta}\rangle \implies \text{Re} \langle\hat{O}_{\mu}\hat{O}_{\nu}\rangle = -\text{Re} \langle\hat{O}_{\mu}^{\dagger}\hat{O}_{\nu}\rangle.$$

### A. Derivations

$$\text{iii.) } \delta\theta^\mu\delta\theta^\nu\langle\hat{O}_\mu\rangle\langle\hat{O}_\nu^\dagger\rangle = \frac{1}{2}\delta\theta^\mu\delta\theta^\nu\left(\langle\hat{O}_\mu\rangle\langle\hat{O}_\nu^\dagger\rangle + \langle\hat{O}_\nu\rangle\langle\hat{O}_\mu^\dagger\rangle\right) = 2\operatorname{Re}\left[\langle\hat{O}_\mu^\dagger\rangle\langle\hat{O}_\nu\rangle\right]$$

In addition, taking the derivative of the energy  $E_\theta = \frac{\langle\psi_\theta|\hat{H}|\psi_\theta\rangle}{\langle\psi_\theta|\psi_\theta\rangle}$  we arrive at

$$\partial_\mu E_\theta = 2\operatorname{Re}\left\{\langle\hat{O}_\mu\hat{H}\rangle - \langle\hat{O}_\mu^\dagger\rangle\langle\hat{H}\rangle\right\} \quad (\text{A.5})$$

where the averages denote  $\langle\dots\rangle = \langle\psi_\theta|\dots|\psi_\theta\rangle / \langle\psi_\theta|\psi_\theta\rangle$ . Using those identities, we can rewrite the fidelity

$$\begin{aligned} F(\delta\theta) &= 1 - \delta\tau\delta\theta^\mu 2\operatorname{Re}\left[\langle\hat{O}_\mu^\dagger\hat{H}\rangle - \langle\hat{O}_\mu^\dagger\rangle\langle\hat{H}\rangle\right] + \delta\theta^\mu\delta\theta^\nu\underbrace{\left[-\operatorname{Re}\langle\hat{O}_\mu^\dagger\hat{O}_\nu\rangle + \operatorname{Re}\langle\hat{O}_\mu^\dagger\rangle\langle\hat{O}_\nu\rangle\right]}_{=\tilde{S}_{\mu\nu}} + 2\delta\tau\langle\hat{H}\rangle + \delta\tau^2\langle\hat{H}\rangle^2 \\ &= 1 - \delta\tau\delta\theta^\mu\partial_\mu E_\theta - \delta\theta^\mu\delta\theta^\nu\tilde{S}_{\mu\nu} + 2\delta\tau\langle\hat{H}\rangle + \delta\tau^2\langle\hat{H}\rangle^2. \end{aligned} \quad (\text{A.6})$$

Maximizing the fidelity with respect to  $\delta\theta$  and writing the quantum geometric tensor  $S_{\mu\nu} = 2\tilde{S}_{\mu\nu}$  we obtain the update rule

$$\delta\theta = -\delta\tau S^{-1}\nabla_\theta E_\theta \quad (\text{A.7})$$

and in the limit  $\delta\tau \rightarrow 0$  we retrieve equation (A.1).

#### A.1.2. Residual Norm Optimization

The second way to derive the SR equation is by rewriting the Schrödinger equation as a least square problem. We start with the imaginary time Schrödinger equation

$$\frac{\partial}{\partial\tau}|\Psi_\theta\rangle = -\hat{H}|\Psi_\theta\rangle. \quad (\text{A.8})$$

We rewrite the equation to a residual norm

$$R = \left\|\partial_\tau|\Psi_\theta\rangle + \hat{H}|\Psi_\theta\rangle\right\|^2 = \langle\partial_\tau\Psi_\theta|\partial_\tau\Psi_\theta\rangle + \left\langle\partial_\tau\Psi_\theta\middle|\hat{H}\Psi_\theta\right\rangle + \left\langle\hat{H}\Psi_\theta\middle|\partial_\tau\Psi_\theta\right\rangle + \left\langle\hat{H}\Psi_\theta\middle|\hat{H}\Psi_\theta\right\rangle. \quad (\text{A.9})$$

Now we have to be careful since the state has a gauge freedom and thus the following form

$$|\Psi_\theta\rangle = e^{i\chi}\frac{|\psi_\theta\rangle}{\sqrt{\langle\psi_\theta|\psi_\theta\rangle}} \quad (\text{A.10})$$

where  $\chi \in \mathbb{R}$  is the phase and the unnormalized  $|\psi_\theta\rangle$  will be effectively represented through a neural network. Using the chain rule for derivatives, we can rewrite the time derivative as  $\partial_\tau|\Psi_\theta\rangle = \sum_\mu \frac{\partial}{\partial\theta^\mu}|\Psi_\theta\rangle \dot{\theta}^\mu + i\dot{\chi}|\Psi_\theta\rangle$ . Now minimizing equation (A.9) with respect to  $\dot{\chi}$  yields a constraint condition for the derivative of the phase:

$$\dot{\chi} = i\dot{\theta}^\mu \operatorname{Im}\langle\hat{D}_\mu\rangle. \quad (\text{A.11})$$

where  $\langle D_\mu\rangle = \langle\Psi_\mu|\partial_\mu\Psi_\mu\rangle$ . Using the chain rule the residual can be written in the following form

$$\begin{aligned} R &= \dot{\chi}^2 - i\dot{\chi}\dot{\theta}^\mu\langle\hat{D}_\mu\rangle_\Psi - i\dot{\chi}\langle\hat{H}\rangle_\Psi + i\dot{\chi}\dot{\theta}^\mu\langle\hat{D}_\mu^\dagger\rangle_\Psi + \dot{\theta}^\mu\dot{\theta}^\nu\langle\hat{D}_\mu^\dagger\hat{D}_\nu\rangle_\Psi + \dot{\theta}^\mu\langle\hat{D}_\mu^\dagger\hat{H}\rangle_\Psi \\ &\quad + i\dot{\chi}\langle\hat{H}\rangle_\Psi + \dot{\theta}^\mu\langle\hat{H}\hat{D}_\mu\rangle_\Psi + \langle\hat{H}^2\rangle_\Psi. \end{aligned} \quad (\text{A.12})$$

Furthermore we want to express the above in terms of the normalized state  $|\psi_\theta\rangle$ . Thus the derivative operator  $\hat{D}_\mu$  can be expressed as  $\hat{D}_\mu = \hat{O}_\mu - \operatorname{Re}\langle\hat{O}_\mu\rangle$  where  $\hat{O}_\mu$  is the partial derivative of the state  $|\psi_\theta\rangle$  and the average is to be understood as  $\langle\dots\rangle = \langle\psi_\theta|\dots|\psi_\theta\rangle / \langle\psi_\theta|\psi_\theta\rangle$ . Inserting the constraint for  $\dot{\chi}$  and using the rewritten form of the derivative, the residual norm can be simplified to the following

### A. Derivations

expression

$$R = \dot{\theta}^\mu \dot{\theta}^\nu \left[ \text{Re}\langle \hat{O}_\mu^\dagger \hat{O}_\nu \rangle - \underbrace{\text{Re}\langle \hat{O}_\mu^\dagger \rangle \text{Re}\langle \hat{O}_\nu \rangle + \text{Im}\langle \hat{O}_\mu^\dagger \rangle \text{Im}\langle \hat{O}_\nu \rangle}_{\text{Re}\langle \hat{O}_\mu^\dagger \rangle \langle \hat{O}_\nu \rangle} \right] + \dot{\theta}^\mu \underbrace{2 \text{Re} \left[ \langle \hat{O}_\mu^\dagger \hat{H} \rangle - \langle \hat{O}_\mu^\dagger \rangle \langle \hat{H} \rangle \right]}_{\partial_\mu E(\theta)}. \quad (\text{A.13})$$

Now minimizing the above equation with respect to  $\dot{\theta}^\mu$  we obtain

$$\frac{\partial R}{\partial \dot{\theta}^\mu} = 0 = S_{\mu\nu} \dot{\theta}^\nu + \partial_\mu E(\theta) \quad (\text{A.14})$$

which is exactly equation (A.1).

## A.2. Marshall Peierls Sign Rule

In this section, we aim to provide an informal proof to explain why the Marshall-Peierls sign rule is effective. Assume a stoquastic Hamiltonian where  $H_{ij} \leq 0$  for every  $i \neq j$ . The diagonal elements may be any real numbers. To apply the Perron-Frobenius theorem, it is important to verify that the matrix  $A$  being analyzed is both non-negative and irreducible. To ensure irreducibility, it is often adequate to confine our analysis to a symmetry sector, such as the total magnetization. To achieve non-negativity, we define the matrix as  $A = c\mathbb{1} - H$ , where  $c > 0$  is chosen to be sufficiently large so that  $A$  becomes non-negative. We subsequently observe that the eigenvector corresponding to the smallest eigenvalue of  $H$  coincides with the eigenvector corresponding to the largest eigenvalue of  $A$ . According to the Perron-Frobenius theorem [45], the eigenvector associated with the largest eigenvalue of  $A$  exhibits real and positive components. Consequently, the ground state, represented by the eigenvector corresponding to the smallest eigenvalue of the stoquastic Hamiltonian, similarly possesses real and positive components.

## A.3. Symmetries

In here we want to give proof that global spin flip and magnetization parity are symmetries of the XYZ Hamiltonian in eq. (2.2). Moreover, we show that the expectation value of total magnetization is 0 for eigenstates of the Hamiltonian, in particular the ground state  $|0\rangle$ .

The operator that describes a global spin flip is given by  $\mathcal{P} = \prod_i \sigma_i^x$ , where the product runs at all lattice sites  $i$ . The magnetization parity is given by  $\mathcal{P}_M = e^{i\pi M}$  where  $M = \sum_i \sigma_i^z$  is the total magnetization of the system. To check if those two operators are symmetries of the system, we verify that the following holds,  $O\hat{H}O^{-1} = \hat{H}$ , where  $O \in \{\mathcal{P}, \mathcal{P}_M\}$ .

The Hamiltonian in Eq. (2.2) has the generic form

$$\hat{H} = \sum_{i,j} J_{ij}^x S_i^x S_j^x + J_{ij}^y S_i^y S_j^y + J_{ij}^z S_i^z S_j^z. \quad (\text{A.15})$$

The spin flip operator  $\mathcal{P}$  acts on the spin operators at a single site as follows:

$$\sigma_i^x S_i^x (\sigma_i^x)^{-1} = S_i^x, \quad \sigma_i^x S_i^{y,z} (\sigma_i^x)^{-1} = -S_i^{y,z}. \quad (\text{A.16})$$

Given that each term in the Hamiltonian is composed of a product of two spin operators, the negative signs are negated:

$$\mathcal{P} \hat{H} \mathcal{P}^{-1} = \hat{H} = \sum_{i,j} J_{ij}^x S_i^x S_j^x + J_{ij}^y (-S_i^y)(-S_j^y) + J_{ij}^z (-S_i^z)(-S_j^z) = \hat{H} \quad (\text{A.17})$$

### A. Derivations

indicating that the spin flip operator is a symmetry of the Hamiltonian.

Now, let us go over the magnetization parity. We again check how the magnetization parity acts on the single-site spin operators,

$$e^{i\pi\sigma_j^z/2} S_j^z e^{-i\pi\sigma_j^z/2} = S_j^z, \quad e^{i\pi\sigma_j^z/2} S_j^{x,y} e^{-i\pi\sigma_j^z/2} = -S_j^{x,y} \quad (\text{A.18})$$

where we used that  $e^{a\sigma^z} = \text{diag}(e^a, e^{-a})$ . Now again, since every term in the Hamiltonian consists of two spatially equal spin operators the additional signs get canceled, and thus magnetization parity is a symmetry of the Hamiltonian as well.

Finally, let us consider the expectation value for the total magnetization. We know that the ground state  $|0\rangle$  is an eigen state of the parity operator  $\mathcal{P}|0\rangle = \pm|0\rangle$  since the parity commutes with the Hamiltonian.

$$\langle M \rangle_0 = \langle 0 | \sum_i \sigma_i^z | 0 \rangle = (\pm 1)^2 \langle 0 | \mathcal{P}^{-1} \sum_i \sigma_i^z \mathcal{P} | 0 \rangle = -\langle 0 | \mathcal{P}^{-1} \mathcal{P} M | 0 \rangle = -\langle M \rangle_0. \quad (\text{A.19})$$

where we utilized the condition that for each  $i$ , the relation  $\sigma_i^z \sigma_j^x = -\delta_{ij} \sigma_i^x \sigma_j^z$  is satisfied. Therefore, we can conclude that the expected value of the total magnetization is 0. The above argument is valid for every eigenstate of the Hamiltonian.

## A.4. Data & Code

The code and certain data utilized in this thesis can be accessed on the following GitHub repository.

# Bibliography

1. Savary, L. & Balents, L. Quantum spin liquids: a review. *Reports on Progress in Physics* **80**, 016502. ISSN: 0034-4885, 1361-6633. <https://iopscience.iop.org/article/10.1088/0034-4885/80/1/016502> (2025) (Jan. 2017).
2. Becca, F. & Sorella, S. *Quantum Monte Carlo Approaches for Correlated Systems* 1st ed. ISBN: 978-1-107-12993-1 978-1-316-41704-1. <https://www.cambridge.org/core/product/identifier/9781316417041/type/book> (2025) (Cambridge University Press, Nov. 2017).
3. Troyer, M. & Wiese, U.-J. Computational Complexity and Fundamental Limitations to Fermionic Quantum Monte Carlo Simulations. en. *Physical Review Letters* **94**, 170201. ISSN: 0031-9007, 1079-7114. <https://link.aps.org/doi/10.1103/PhysRevLett.94.170201> (2025) (May 2005).
4. Schollwöck, U. The density-matrix renormalization group. en. *Reviews of Modern Physics* **77**, 259–315. ISSN: 0034-6861, 1539-0756. <https://link.aps.org/doi/10.1103/RevModPhys.77.259> (2025) (Apr. 2005).
5. Carleo, G. & Troyer, M. Solving the Quantum Many-Body Problem with Artificial Neural Networks. *Science* **355**. arXiv:1606.02318 [cond-mat], 602–606. ISSN: 0036-8075, 1095-9203. <http://arxiv.org/abs/1606.02318> (2025) (Feb. 2017).
6. Vicentini, F. et al. NetKet 3: Machine Learning Toolbox for Many-Body Quantum Systems. *SciPost Physics Codebases*, 7. <https://scipost.org/10.21468/SciPostPhysCodeb.7> (2025) (Aug. 2022).
7. Carrasquilla, J., Ciolo, A. D., Becca, F., Galitski, V. & Rigol, M. Nature of the phases in the frustrated X Y model on the honeycomb lattice. en. *Physical Review B* **88**, 241109. ISSN: 1098-0121, 1550-235X. <https://link.aps.org/doi/10.1103/PhysRevB.88.241109> (2025) (Dec. 2013).
8. Roth, C., Szabó, A. & MacDonald, A. H. High-accuracy variational Monte Carlo for frustrated magnets with deep neural networks. en. *Physical Review B* **108**, 054410. ISSN: 2469-9950, 2469-9969. <https://link.aps.org/doi/10.1103/PhysRevB.108.054410> (2025) (Aug. 2023).
9. Hibat-Allah, M., Ganahl, M., Hayward, L. E., Melko, R. G. & Carrasquilla, J. Recurrent neural network wave functions. en. *Physical Review Research* **2**, 023358. ISSN: 2643-1564. <https://link.aps.org/doi/10.1103/PhysRevResearch.2.023358> (2025) (June 2020).
10. Viteritti, L. L., Rende, R. & Becca, F. Transformer Variational Wave Functions for Frustrated Quantum Spin Systems. en. *Physical Review Letters* **130**, 236401. ISSN: 0031-9007, 1079-7114. <https://link.aps.org/doi/10.1103/PhysRevLett.130.236401> (2025) (June 2023).
11. Kim, J. et al. Neural-network quantum states for ultra-cold Fermi gases. en. *Communications Physics* **7**, 148. ISSN: 2399-3650. <https://www.nature.com/articles/s42005-024-01613-w> (2025) (May 2024).
12. Deng, D.-L., Li, X. & Das Sarma, S. Quantum Entanglement in Neural Network States. en. *Physical Review X* **7**, 021021. ISSN: 2160-3308. <http://link.aps.org/doi/10.1103/PhysRevX.7.021021> (2025) (May 2017).
13. Seifert, H. & Al-Khudair, I. Über die systeme alkaliemetallbromid/kobalt(II)-bromid. de. *Journal of Inorganic and Nuclear Chemistry* **37**, 1625–1628. ISSN: 00221902. <https://linkinghub.elsevier.com/retrieve/pii/0022190275802879> (2025) (July 1975).
14. Povarov, K. Y. et al. Magnetization plateaux cascade in the frustrated quantum antiferromagnet Cs<sub>2</sub>CoBr<sub>4</sub>. en. *Physical Review Research* **2**, 043384. ISSN: 2643-1564. <https://link.aps.org/doi/10.1103/PhysRevResearch.2.043384> (2025) (Dec. 2020).

## Bibliography

15. Facheris, L. *et al.* Spin Density Wave versus Fractional Magnetization Plateau in a Triangular Antiferromagnet. en. *Physical Review Letters* **129**, 087201. ISSN: 0031-9007, 1079-7114. <https://link.aps.org/doi/10.1103/PhysRevLett.129.087201> (Aug. 2022).
16. Soldatov, T. A., Smirnov, A. I. & Syromyatnikov, A. V. *Dynamics of anisotropic frustrated antiferromagnet Cs<sub>2</sub>CoBr<sub>4</sub> in a spin-liquid regime* arXiv:2309.02266 [cond-mat]. Sept. 2023. <http://arxiv.org/abs/2309.02266> (2025).
17. Soldatov, T. A., Smirnov, A. I. & Syromyatnikov, A. V. *Spin dynamics in ordered phases of anisotropic triangular-lattice antiferromagnet Cs<sub>2</sub>CoBr<sub>4</sub>* arXiv:2307.16251 [cond-mat]. July 2023. <http://arxiv.org/abs/2307.16251> (2025).
18. Golub, G. H. & Van Loan, C. F. *Matrix computations* 3rd ed. ISBN: 978-0-8018-5413-2 978-0-8018-5414-9 (Johns Hopkins University Press, Baltimore, 1996).
19. Sandvik, A. W. Computational Studies of Quantum Spin Systems. Publisher: arXiv Version Number: 1. <https://arxiv.org/abs/1101.3281> (2025) (2011).
20. Starykh, O. A. Unusual ordered phases of highly frustrated magnets: a review. *Reports on Progress in Physics* **78**, 052502. ISSN: 0034-4885, 1361-6633. <https://iopscience.iop.org/article/10.1088/0034-4885/78/5/052502> (2025) (May 2015).
21. Yamamoto, D., Marmorini, G. & Danshita, I. Quantum Phase Diagram of the Triangular-Lattice X X Z Model in a Magnetic Field. en. *Physical Review Letters* **112**, 127203. ISSN: 0031-9007, 1079-7114. <https://link.aps.org/doi/10.1103/PhysRevLett.112.127203> (2014).
22. Kitaev, A. Anyons in an exactly solved model and beyond. en. *Annals of Physics* **321**, 2–111. ISSN: 00034916. <https://linkinghub.elsevier.com/retrieve/pii/S0003491605002381> (2006).
23. Takagi, H., Takayama, T., Jackeli, G., Khaliullin, G. & Nagler, S. E. Concept and realization of Kitaev quantum spin liquids. en. *Nature Reviews Physics* **1**, 264–280. ISSN: 2522-5820. <https://www.nature.com/articles/s42254-019-0038-2> (2019).
24. Zvyagin, S. A. *et al.* Dimensional reduction and incommensurate dynamic correlations in the  $S=\frac{1}{2}$  triangular-lattice antiferromagnet Ca<sub>3</sub>ReO<sub>5</sub>Cl<sub>2</sub>. en. *Nature Communications* **13**, 6310. ISSN: 2041-1723. <https://www.nature.com/articles/s41467-022-33992-5> (2022).
25. Smirnov, A. I., Povarov, K. Y., Petrov, S. V. & Shapiro, A. Y. Magnetic resonance in the ordered phases of the two-dimensional frustrated quantum magnet Cs<sub>2</sub>CuCl<sub>4</sub>. en. *Physical Review B* **85**, 184423. ISSN: 1098-0121, 1550-235X. <https://link.aps.org/doi/10.1103/PhysRevB.85.184423> (2012).
26. Veillette, M. Y., James, A. J. A. & Essler, F. H. L. Spin dynamics of the quasi-two-dimensional spin-1/2 quantum magnet Cs<sub>2</sub>CuCl<sub>4</sub>. en. *Physical Review B* **72**, 134429. ISSN: 1098-0121, 1550-235X. <https://link.aps.org/doi/10.1103/PhysRevB.72.134429> (2005).
27. Yunoki, S. & Sorella, S. Two spin liquid phases in the spatially anisotropic triangular Heisenberg model. en. *Physical Review B* **74**, 014408. ISSN: 1098-0121, 1550-235X. <https://link.aps.org/doi/10.1103/PhysRevB.74.014408> (2006).
28. Heidarian, D., Sorella, S. & Becca, F. Spin-1/2 Heisenberg model on the anisotropic triangular lattice: From magnetism to a one-dimensional spin liquid. en. *Physical Review B* **80**, 012404. ISSN: 1098-0121, 1550-235X. <https://link.aps.org/doi/10.1103/PhysRevB.80.012404> (2009).
29. Ghorbani, E., Tocchio, L. F. & Becca, F. Variational wave functions for the S = 1/2 Heisenberg model on the anisotropic triangular lattice: Spin liquids and spiral orders. en. *Physical Review B* **93**, 085111. ISSN: 2469-9950,

## Bibliography

- 2469-9969. <https://link.aps.org/doi/10.1103/PhysRevB.93.085111> (Feb. 2016).
30. Breunig, O. *et al.* Spin-1/2 XXZ Chain System Cs<sub>2</sub>CoCl<sub>4</sub> in a Transverse Magnetic Field. en. *Physical Review Letters* **111**, 187202. ISSN: 0031-9007, 1079-7114. <https://link.aps.org/doi/10.1103/PhysRevLett.111.187202> (2025) (Oct. 2013).
31. Zhu, Z., Maksimov, P. A., White, S. R. & Chernyshev, A. L. Topography of Spin Liquids on a Triangular Lattice. en. *Physical Review Letters* **120**, 207203. ISSN: 0031-9007, 1079-7114. <https://link.aps.org/doi/10.1103/PhysRevLett.120.207203> (2025) (May 2018).
32. Weichselbaum, A. & White, S. R. Incommensurate correlations in the anisotropic triangular Heisenberg lattice. en. *Physical Review B* **84**, 245130. ISSN: 1098-0121, 1550-235X. <https://link.aps.org/doi/10.1103/PhysRevB.84.245130> (2025) (Dec. 2011).
33. Robert, C. P. & Casella, G. *Monte Carlo Statistical Methods* ISBN: 978-1-4419-1939-7 978-1-4757-4145-2. <http://link.springer.com/10.1007/978-1-4757-4145-2> (2025) (Springer New York, New York, NY, 2004).
34. Medvidović, M. & Moreno, J. R. Neural-network quantum states for many-body physics. *The European Physical Journal Plus* **139**. arXiv:2402.11014 [cond-mat], 631. ISSN: 2190-5444. <http://arxiv.org/abs/2402.11014> (2025) (July 2024).
35. Hastings, W. K. Monte Carlo sampling methods using Markov chains and their applications. en. *Biometrika* **57**, 97–109. ISSN: 1464-3510, 0006-3444. <https://academic.oup.com/biomet/article/57/1/97/284580> (2025) (Apr. 1970).
36. Sorella, S. Wave function optimization in the variational Monte Carlo method. en. *Physical Review B* **71**, 241103. ISSN: 1098-0121, 1550-235X. <https://link.aps.org/doi/10.1103/PhysRevB.71.241103> (2025) (June 2005).
37. Chen, A. & Heyl, M. Empowering deep neural quantum states through efficient optimization. en. *Nature Physics* **20**, 1476–1481. ISSN: 1745-2473, 1745-2481. <https://www.nature.com/articles/s41567-024-02566-1> (2025) (Sept. 2024).
38. Rende, R., Viteritti, L. L., Bardone, L., Becca, F. & Goldt, S. A simple linear algebra identity to optimize Large-Scale Neural Network Quantum States. *Communications Physics* **7**. arXiv:2310.05715 [cond-mat], 260. ISSN: 2399-3650. <http://arxiv.org/abs/2310.05715> (2025) (Aug. 2024).
39. Wu, D. *et al.* Variational Benchmarks for Quantum Many-Body Problems. *Science* **386**. arXiv:2302.04919 [quant-ph], 296–301. ISSN: 0036-8075, 1095-9203. <http://arxiv.org/abs/2302.04919> (2025) (Oct. 2024).
40. Nomura, Y. Helping restricted Boltzmann machines with quantum-state representation by restoring symmetry. *Journal of Physics: Condensed Matter* **33**. arXiv:2009.14777 [cond-mat], 174003. ISSN: 0953-8984, 1361-648X. <http://arxiv.org/abs/2009.14777> (2025) (Apr. 2021).
41. Choo, K., Carleo, G., Regnault, N. & Neupert, T. Symmetries and Many-Body Excitations with Neural-Network Quantum States. en. *Physical Review Letters* **121**, 167204. ISSN: 0031-9007, 1079-7114. <https://link.aps.org/doi/10.1103/PhysRevLett.121.167204> (2025) (Oct. 2018).
42. Lieb, E. & Mattis, D. Ordering Energy Levels of Interacting Spin Systems. en. *Journal of Mathematical Physics* **3**, 749–751. ISSN: 0022-2488, 1089-7658. <https://pubs.aip.org/jmp/article/3/4/749/227952/Ordering-Energy-Levels-of-Interacting-Spin-Systems> (2025) (July 1962).
43. Marshall, W. Antiferromagnetism. en. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* **232**, 48–68. ISSN: 0080-4630, 2053-9169. <https://royalsocietypublishing.org/doi/10.1098/rspa.1955.0200> (2025) (Oct. 1955).

## Bibliography

44. Capriotti, L. Quantum Effects and Broken Symmetries in Frustrated Antiferromagnets. *International Journal of Modern Physics B* **15**. arXiv:cond-mat/0112207, 1799–1842. ISSN: 0217-9792, 1793-6578. <http://arxiv.org/abs/cond-mat/0112207> (May 2001).
45. Perron, O. Zur Theorie der Matrices. de. *Mathematische Annalen* **64**, 248–263. ISSN: 0025-5831, 1432-1807. <http://link.springer.com/10.1007/BF01449896> (2025) (June 1907).
46. Bravyi, S., DiVincenzo, D. P., Oliveira, R. I. & Terhal, B. M. The Complexity of Stoquastic Local Hamiltonian Problems. Publisher: arXiv Version Number: 4. <https://arxiv.org/abs/quant-ph/0606140> (2025) (2006).
47. Huse, D. A. & Elser, V. Simple Variational Wave Functions for Two-Dimensional Heisenberg Spin-½ Antiferromagnets. en. *Physical Review Letters* **60**, 2531–2534. ISSN: 0031-9007. <https://link.aps.org/doi/10.1103/PhysRevLett.60.2531> (2025) (June 1988).
48. Vicentini, F. *Ground-State: Ising model – NetKet* Nov. 2021. <https://netket.readthedocs.io/en/stable/tutorials/gs-isng.html> (2025).
49. Jastrow, R. Many-Body Problem with Strong Forces. en. *Physical Review* **98**, 1479–1484. ISSN: 0031-899X. <https://link.aps.org/doi/10.1103/PhysRev.98.1479> (2025) (June 1955).
50. Lippmann, R. P. Review of Neural Networks for Speech Recognition. en. *Neural Computation* **1**, 1–38. ISSN: 0899-7667, 1530-888X. <https://direct.mit.edu/neco/article/1/1/1-38/5468> (2025) (Mar. 1989).
51. Hinton, G. E. & Salakhutdinov, R. R. Reducing the Dimensionality of Data with Neural Networks. en. *Science* **313**, 504–507. ISSN: 0036-8075, 1095-9203. <https://www.science.org/doi/10.1126/science.1127647> (2025) (July 2006).
52. Broecker, P., Carrasquilla, J., Melko, R. G. & Trebst, S. Machine learning quantum phases of matter beyond the fermion sign problem. en. *Scientific Reports* **7**, 8823. ISSN: 2045-2322. <https://www.nature.com/articles/s41598-017-09098-0> (2025) (Aug. 2017).
53. Sharir, O., Levine, Y., Wies, N., Carleo, G. & Shashua, A. Deep Autoregressive Models for the Efficient Variational Simulation of Many-Body Quantum Systems. en. *Physical Review Letters* **124**, 020503. ISSN: 0031-9007, 1079-7114. <https://link.aps.org/doi/10.1103/PhysRevLett.124.020503> (2025) (Jan. 2020).
54. Liang, X. et al. Solving frustrated quantum many-particle models with convolutional neural networks. en. *Physical Review B* **98**, 104426. ISSN: 2469-9950, 2469-9969. <https://link.aps.org/doi/10.1103/PhysRevB.98.104426> (2025) (Sept. 2018).
55. Szabó, A. & Castelnovo, C. Neural network wave functions and the sign problem. en. *Physical Review Research* **2**, 033075. ISSN: 2643-1564. <https://link.aps.org/doi/10.1103/PhysRevResearch.2.033075> (2025) (July 2020).
56. Levine, Y., Sharir, O., Cohen, N. & Shashua, A. Quantum Entanglement in Deep Learning Architectures. en. *Physical Review Letters* **122**, 065301. ISSN: 0031-9007, 1079-7114. <https://link.aps.org/doi/10.1103/PhysRevLett.122.065301> (2025) (Feb. 2019).
57. Stoudenmire, E. & White, S. R. Studying Two-Dimensional Systems with the Density Matrix Renormalization Group. en. *Annual Review of Condensed Matter Physics* **3**, 111–128. ISSN: 1947-5454, 1947-5462. <https://www.annualreviews.org/doi/10.1146/annurev-conmatphys-020911-125018> (2025) (Mar. 2012).
58. Liu, W.-Y., Zhai, H., Peng, R., Gu, Z.-C. & Chan, G. K.-L. *Accurate Simulation of the Hubbard Model with Finite Fermionic Projected Entangled Pair States* arXiv:2502.13454 [cond-mat]. Feb. 2025. [http://arxiv.org/abs/2502.13454](https://arxiv.org/abs/2502.13454) (2025).

## Bibliography

59. Dubey, S. R., Singh, S. K. & Chaudhuri, B. B. *Activation Functions in Deep Learning: A Comprehensive Survey and Benchmark* arXiv:2109.14545 [cs]. June 2022. <http://arxiv.org/abs/2109.14545> (2025).
60. Nomura, Y. & Imada, M. Dirac-Type Nodal Spin Liquid Revealed by Refined Quantum Many-Body Solver Using Neural-Network Wave Function, Correlation Ratio, and Level Spectroscopy. en. *Physical Review X* **11**, 031034. ISSN: 2160-3308. <https://link.aps.org/doi/10.1103/PhysRevX.11.031034> (2025) (Aug. 2021).
61. Saito, H. Solving the Bose–Hubbard Model with Machine Learning. en. *Journal of the Physical Society of Japan* **86**, 093001. ISSN: 0031-9015, 1347-4073. <https://journals.jps.jp/doi/10.7566/JPSJ.86.093001> (2025) (Sept. 2017).
62. Nomura, Y., Darmawan, A. S., Yamaji, Y. & Imada, M. Restricted Boltzmann machine learning for solving strongly correlated quantum systems. en. *Physical Review B* **96**, 205152. ISSN: 2469-9950, 2469-9969. <https://link.aps.org/doi/10.1103/PhysRevB.96.205152> (2025) (Nov. 2017).
63. Vaswani, A. et al. *Attention Is All You Need* Version Number: 7. 2017. <https://arxiv.org/abs/1706.03762> (2025).
64. Dosovitskiy, A. et al. *An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale* Version Number: 2. 2020. <https://arxiv.org/abs/2010.11929> (2025).
65. Zhang, Y.-H. & Di Ventra, M. Transformer Quantum State: A Multi-Purpose Model for Quantum Many-Body Problems. Publisher: arXiv Version Number: 4. <https://arxiv.org/abs/2208.01758> (2025) (2022).
66. Xiong, R. et al. *On Layer Normalization in the Transformer Architecture* arXiv:2002.04745 [cs]. June 2020. <http://arxiv.org/abs/2002.04745> (2025).
67. He, K., Zhang, X., Ren, S. & Sun, J. *Deep Residual Learning for Image Recognition* arXiv:1512.03385 [cs]. Dec. 2015. <http://arxiv.org/abs/1512.03385> (2025).
68. Viteritti, L. L., Rende, R., Parola, A., Goldt, S. & Becca, F. *Transformer Wave Function for two dimensional frustrated magnets: emergence of a Spin-Liquid Phase in the Shastry-Sutherland Model* Version Number: 3. 2023. <https://arxiv.org/abs/2311.16889> (2025).
69. Mishkin, D. & Matas, J. *All you need is a good init* Version Number: 7. 2015. <https://arxiv.org/abs/1511.06422> (2025).
70. Glorot, X. & Bengio, Y. *Understanding the difficulty of training deep feedforward neural networks* in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics* (eds Teh, Y. W. & Titterington, M.) **9** (PMLR, Chia Laguna Resort, Sardinia, Italy, 13–15 May 2010), 249–256. <https://proceedings.mlr.press/v9/glorot10a.html>.
71. Hauschild, J. et al. Tensor network Python (TeNPy) version 1. *SciPost Physics Codebases*, 41. ISSN: 2949-804X. <https://scipost.org/10.21468/SciPostPhysCodeb.41> (2025) (Nov. 2024).
72. Sandvik, A. W. Finite-size scaling of the ground-state parameters of the two-dimensional Heisenberg model. en. *Physical Review B* **56**, 11678–11690. ISSN: 0163-1829, 1095-3795. <https://link.aps.org/doi/10.1103/PhysRevB.56.11678> (2025) (Nov. 1997).
73. Lukyanov, S. Correlation amplitude for the XXZ spin chain in the disordered regime. en. *Physical Review B* **59**. Publisher: American Physical Society (APS), 11163–11164. ISSN: 0163-1829, 1095-3795. <https://link.aps.org/doi/10.1103/PhysRevB.59.11163> (2025) (May 1999).