The Use of Restricted Boltzmann Machines for Modeling a Many-body Quantum Systems

Alev Orfi

Supervisor: Bill Coish

Collaborator: Felix Fehse

(Dated: December, 2020)

A complete description of many-body quantum systems must encompass the correlations between each element of the system. Large systems become computationally challenging to simulate since the number of degrees-of-freedom grows exponentially with the number of particles considered. A restricted Boltzmann machine (RBM), a machine learning technique, has been shown to reparameterize a many-body wavefunction. With the use of this method, large many-body systems have been more accurately simulated compared to the leading variational techniques [1][2].

Here we apply an RBM to a specific many-body problem, the Gaudin magnet. This model describes the interaction between one central-spin and many environmental spins. The transverse-field Ising model and Heisenberg model on which the RBM was previously tested are known to be tractable [1]. Here we extend this technique to the Gaudin magnet in a case with no known simple analytical solution.

Specifically, an RBM is used to perform ground-state determination of the Gaudin magnet with two different sets of Hamiltonian parameters. Since the RBM gives an ansatz for the many-body wavefunction, a variational energy is minimized to find a ground-state description. With the implementation of stochastic sampling and an efficient minimization routine, systems of up to ten spins were considered. These methods allowed the success of the RBM technique to be investigated as the system size scales, further probing the capacity of RBMs in describing many-body quantum systems.

I. INTRODUCTION

To model many-body quantum systems, the interactions between all the particles must be described. The number of degrees-of-freedom grows exponentially as the systems size grows; thus, analytic calculations become quickly impractical. Consider a system of N spin-$\frac{1}{2}$ particles. The state of each particle is described by a two-dimensional Hilbert space. To fully represent all of the spins' interactions, a Hamiltonian is of the size $2^N \times 2^N$. Exact diagonalization can be used to find the ground-state and the dynamics of this system. This process involves finding the 2^N eigenstates of the specific Hamiltonian used, then the system's state $|\psi\rangle$ can be represented as a linear combination of the eigenstates $|S\rangle$,

$$|\psi\rangle = \sum_{\{S\}} C_S \, |S\rangle \, . \tag{1}$$

Since exact diagonalization requires finding the eigenvectors of a $2^N \times 2^N$ matrix, a computationally challenging task, there is an inherent difficultly in solving large many-body systems.

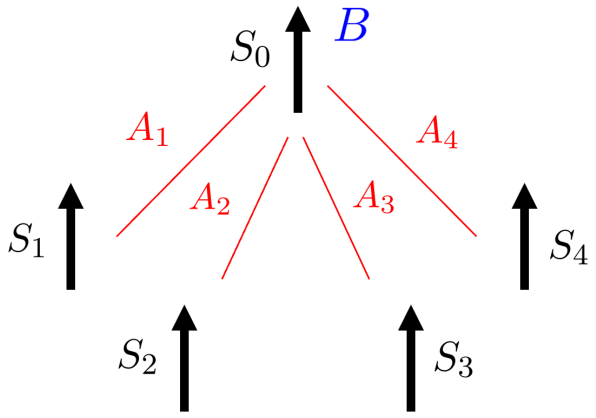

FIG. 1. Diagrammatic representation of the Gaudin magnet.

One such many-body spin system, which suffers from this non-ideal scaling, is known as the Gaudin magnet. This model describes the interactions between N spins, one central spin and $N-1$ bath spins. These bath spins act as an environment and influence the dynamics of the central spin. Thus, only the interactions between each bath spin and the central spin are modelled; the bath-bath interactions are not considered. A diagram of these interactions is shown in Fig. 1.

The central spin is subject to an external field B and its interaction with the environmental spins is parameterized by A_k, a set of coupling constants. This model is represented with the following Hamiltonian,

$$H = BS_0^z + \sum_{k=1}^{N-1} A_k \mathbf{S}_0 \cdot \mathbf{S}_k, \tag{2}$$

where $S_0^z = \frac{1}{2}\sigma_0^z$ is the spin operator for the z-component of the central spin. The vector $\mathbf{S}_k$ is defined as $\mathbf{S}_k = (S_k^x, S_k^y, S_k^z)$ which acts on the k^{th} spin.

The Gaudin magnet Hamiltonian can model the hyperfine interaction between an electron spin in a quantum dot and surrounding nuclear spins [3]. Thus, the decoherence of the quantum dot spin can be tracked by considering the dynamics of the Gaudin magnet. Additionally, a linear combination of commuting Gaudin magnets describe the Gaudin-Richardson model, which is often a starting point for a mean-field treatment of superconductivity[4]. As an N interacting spin system, the exact diagonalization of the Gaudin magnet is quickly too computationally challenging. Currently, more sophisticated methods such as the use of the Bethe ansatz allow the dynamics of larger Gaudin magnet systems to be calculated [5]. However, these methods are still limited to systems of around 30 spins [5]. Our goal is to implement machine learning techniques to reduce these calculations' complexity further, thus allowing larger systems to be modelled. If more efficient methods are found for the Gaudin magnet, they could be applied to both quantum dot decoherence and problems in non-equilibrium superconductivity.

Machine-learning methods are so successful as they learn which correlations are significant and therefore can disregard less impactful information. Applying these techniques to many-body systems leads to a lower-dimensional reparameterization of the wavefunction. It is more feasible to solve large-N systems if they are expressed with fewer parameters, thus decreasing the computational complexity. Here we consider one specific neural network, a restricted Boltzmann machine (RBM). This network choice has been shown to be an efficient method for finding a ground-state estimate and dynamics of the transverse field Ising model and the antiferromagnetic Heisenberg model [1]. Additionally, RBMs have been used to model open systems [2] and to preform state tomography from experimental data [6].

A RBM is a two-layer network defined by network parameters $\mathcal{W} = \{a, b, W_{ij}\}$. A pictorial representation of the network structure is shown in Fig. 2. To model a quantum system, an RBM

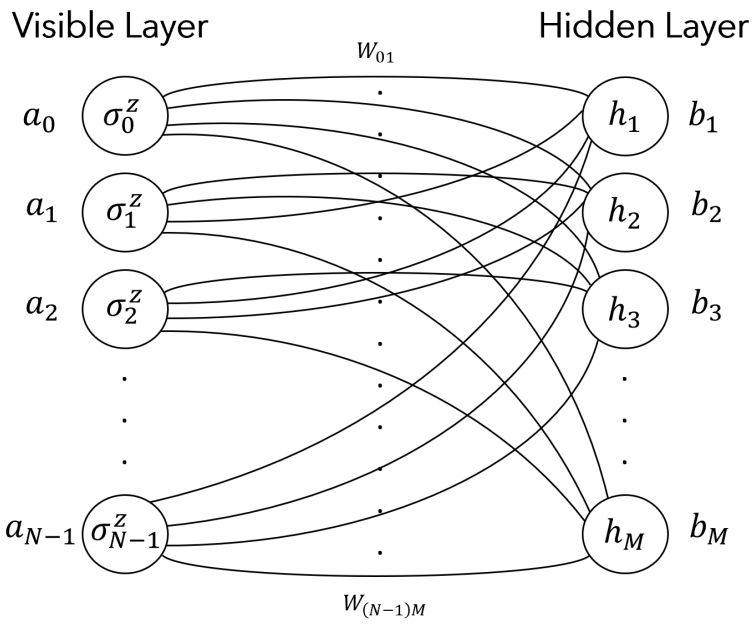

FIG. 2. A restricted Boltzmann machine architecture applied to the Gaudin magnet. The visible layer, shown on the left, is composed of visible spin variables $S = \{\sigma_0^z, \sigma_1^z, ..., \sigma_{N-1}^z\}$. The hidden layer, on the right, holds the hidden variables $h_1, h_2, ..., h_M$.

is used to approximate the coefficients C_S in Eq. (1),

$$C_S \simeq \Psi(S; \mathcal{W}). \tag{3}$$

Here, the network parameters $\mathcal{W} = \{a, b, W_{ij}\}$ are complex to describe the nature of the wavefunction. A RBM is comprised of layers of interacting stochastic variables, known as nodes. Here we consider two layers, a visible layer made from N nodes and a hidden layer of M nodes. Each node is connected to every node in the opposite layer with a weight parameter W_{ij}. The nodes of the visible layer are the physical parameters of the considered system. As seen in Fig. 2, in this context the visible nodes correspond to the Pauli-z operator, σ^z, acting at each spin site. This gives the following set of N visible nodes; $S = \{\sigma_0^z, \sigma_1^z, ..., \sigma_{N-1}^z\}$. The hidden layer consists of M auxiliary spin variables $h_1, h_2, ..., h_M$, which do not have a physical interpretation but allow the network to capture higher-order correlations between the spins. The number of hidden nodes M can be varied to change the expressiveness of the network. The visible and hidden nodes have corresponding biases, $a_0, ..., a_{N-1}$ and $b_1, ..., b_M$, respectively. Specifically, the set of the coefficients of Eq. (1) are reparameterized in terms of these network parameters $\mathcal{W} = \{a, b, W_{ij}\}$ as [1],

$$\Psi(S; \mathcal{W}(t)) = \sum_{\{h_i\}} e^{\sum_j a_j \sigma_j^z + \sum_i b_i h_i + \sum_{ij} W_{ij} h_i \sigma_j^z}. \tag{4}$$

Summing over the basis states gives the RBM representation of the system's wavefunction,

$$|\psi_{\mathrm{RBM}}(\mathcal{W}(t))\rangle = \sum_{\{S\}} \Psi(S;\mathcal{W}(t))\,|S\rangle\,. \tag{5}$$

The network is trained by modifying these network parameters until an accurate approximation of the state is found. This training process differs depending on the intended use of the network. We have been focusing on ground-state determination, however an RBM has been shown to also model system dynamics [1]. To find the ground-state of the system, a cost function can be defined as a variational energy,

$$E(\mathcal{W}) = \frac{\langle\psi_{\mathrm{RBM}}|\,H\,|\psi_{\mathrm{RBM}}\rangle}{\langle\psi_{\mathrm{RBM}}|\psi_{\mathrm{RBM}}\rangle}. \tag{6}$$

Training amounts to minimizing this function, thus finding the optimal network parameters to represent the ground-state.

A. Previous Work

As this work was a continuation of an undergraduate thesis project [7], the work completed this semester built upon previous results. Notably, the techniques use to probe the success of the network were previously created and tested. The RBM's ground-state results are compared to the exact diagonalization ground-state, thus allowing the accuracy of the RBM to be probed. It is essential that the numerical exact diagonalization results are accurate; therefore, they were additionally verified against analytic results. Particularly, in the case of constant coupling terms $A_k = A$, allows the Hamiltonian to be block diagonalized [7]. This formulation decomposed the Hilbert space in terms of 2×2 subspaces, which can be easily diagonalized, giving an analytic form of the ground-state. The numerical exact diagonalization and analytic methods yielded identical results. Thus, the numerical exact diagonalization is justified as a verification method for the RBM output.

Two errors are defined, one for comparing the ground-state energy and the other to compare the ground-state. The energy error ϵ_E,is defined to be the difference between the ground-state energy found using the RBM RBM, and that found with exact diagonalization E_{ED},

$$\epsilon_E = |E_{\mathrm{RBM}} - E_{\mathrm{ED}}|. \tag{7}$$

The state error ϵ_ψ, is given in terms of the overlap between the ground-state found through numerical

exact diagonalization, $|\psi_{\mathrm{ED}}\rangle$, and the ground-state found from the RBM, $|\psi_{\mathrm{RBM}}\rangle$,

$$\epsilon_\psi = 1 - |\langle\psi_{\mathrm{RBM}}|\psi_{\mathrm{ED}}\rangle|^2. \tag{8}$$

The quantities ϵ_{E} and ϵ_ψ provide measures for the quality of RBM training.

An RBM was created and tested for small Gaudin magnets with a constant interaction term in the thesis [7]. Multiple training algorithms were investigated to minimize the variational energy, Eq. (6). However, all suffered from large run times, making it not feasible to investigate systems of larger than around 5 spins. Additionally there were inaccuracies due to stagnation in local minima. The work presented here builds upon results presented in [7], specifically addressing these issues.

II. NETKET IMPLEMENTATION

An improvement made to the RBM training was the use of the NetKet package. NetKet offers an implementation of neural networks specifically designed to represent many-body quantum systems [8]. One such network is a restricted Boltzmann machine. NetKet also offers multiple minimization techniques and energy estimation to increase the algorithm efficiency. Two specific methods used in the NetKet package proved to be beneficial for modelling the Gaudin magnet: Monte Carlo estimations and stochastic reconfiguration.

1. Stochastic Sampling

Stochastic sampling is one crucial advantage the NetKet package provides[8]. Through this sampling, energy estimates can be made more efficiently, thus decreasing the algorithm's computational complexity. The training method requires the minimization of the variational energy, Eq. (6); therefore, the variational energy is calculated at each step of the minimization technique. As a result, an increase in the variational energy calculation efficiency has a large effect on the total runtime.

One can consider the basis of all spin configurations $\{\sigma\}$. A general wavefunction of the system may have finite weight for each spin configuration. Calculating the variation energy, an expectation value, would then requiring summing over an exponentially large number of terms. An estimate can be made by instead considering only a subset of these spin configurations. A Markov chain of these configurations is generated such that the spins are sampled with a predefined probability

distribution. Specifically, we have,

$$\langle\psi_{\mathrm{RBM}}|\, H\,|\psi_{\mathrm{RBM}}\rangle = \frac{\sum_{\sigma,\sigma'}\psi^{*}_{\mathrm{RBM}}(\sigma)\,\langle\sigma|\, H\,|\sigma'\rangle\,\psi_{\mathrm{RBM}}(\sigma')}{\sum_{\sigma}|\psi_{\mathrm{RBM}}(\sigma)|^{2}} \tag{9}$$

$$= \sum_{\sigma}\Big(\sum_{\sigma'}\langle\sigma|\, H\,|\sigma'\rangle\,\frac{\psi_{\mathrm{RBM}}(\sigma')}{\psi_{\mathrm{RBM}}(\sigma)}\Big)\frac{|\psi_{\mathrm{RBM}}(\sigma)|^{2}}{\sum_{\sigma'}|\psi_{\mathrm{RBM}}(\sigma')|^{2}} \tag{10}$$

$$\approx \Big\langle\sum_{\sigma'}\langle\sigma|\, H\,|\sigma'\rangle\,\frac{\psi_{\mathrm{RBM}}(\sigma')}{\psi_{\mathrm{RBM}}(\sigma)}\Big\rangle_{\tilde{\sigma}} =: E_{\mathrm{local}}, \tag{11}$$

where in Eq. (11) the expectation value is taken over a sample of configurations $\{\tilde{\sigma}\}$ drawn from the following probability distribution,

$$\pi(\sigma) = \frac{|\psi_{\mathrm{RBM}}(\sigma)|^{2}}{\sum_{\sigma'}|\psi_{\mathrm{RBM}}(\sigma')|^{2}}. \tag{12}$$

The set $\{\tilde{\sigma}\}$ is generated using the Metropolis algorithm, which creates a Markov chain of a predefined length P, $\sigma^{(1)} \longrightarrow \sigma^{(2)} \longrightarrow ... \longrightarrow \sigma^{(P)}$, where each new configuration is specifically chosen to ensure the set samples from Eq. (12) [9]. At each step of the Metropolis algorithm a new configuration is proposed by flipping one spin of the previous configuration. This new configuration is accepted according to the following probability,

$$A(\sigma^{k} \longrightarrow \sigma^{k+1}) = \min\Big(1, \Big|\frac{\psi_{\mathrm{RBM}}(\sigma^{k+1})}{\psi_{\mathrm{RBM}}(\sigma^{k})}\Big|^{2}\Big). \tag{13}$$

As each new configuration only varies from the last by a single spin flip there is only a slight change in the RBM wave function, ψ_{RBM}. Memory look-up tables are kept and updated at each iteration allowing for efficient calculation of the acceptance probabilities, Eq. (13). The estimate accuracy can be modified by increasing the number of the generated configurations. Overall, this stochastic sampling method gives an increase in algorithm efficiency, allowing larger systems to be modelled.

2. Stochastic Reconfiguration

In combination with the stochastic energy estimates, different minimization methods were investigated to minimize the variational energy efficiently. The most successful technique was found to be stochastic reconfiguration [8][10]. This gradient-based method updates the network parameters $\mathcal{W}$ based on their effect on the variational wavefunction. Specifically, the variational derivatives with respect to the k-th network parameter are defined as,

$$\mathcal{O}_{k} = \frac{1}{\psi_{\mathrm{RBM}}}\partial_{\mathcal{W}_{k}}\psi_{\mathrm{RBM}}. \tag{14}$$

This minimizer method updates the network parameter matrix $\mathcal{W}$ at the p$^{\text{th}}$-iteration as

$$\mathcal{W}(p+1) = \mathcal{W}(p) - \gamma S^{-1}(p) F(p). \tag{15}$$

Where here S is the following covariance matrix,

$$S_{kk'}(p) = \langle \mathcal{O}_k^* \mathcal{O}_{k'} \rangle - \langle \mathcal{O}_k^* \rangle \langle \mathcal{O}_{k'} \rangle. \tag{16}$$

Additionally, the forces F are vectors with one element defined as,

$$F_k(p) = \langle E_{\text{local}} \mathcal{O}_k^* \rangle - \langle E_{\text{local}} \rangle \langle \mathcal{O}_k^* \rangle, \tag{17}$$

with E_{local} defined in Eq. (11).

Thus, the partial derivatives of the wave function with respect to the network parameters are calculated at each step of the minimization. Due to the exponential form of the RBM ansatz, Eq. (4), these derivatives can obtained analytically. The expectation values in Eq. (16) and q. (17) are also evaluated statistically with the Metropolis-Hastings method outlined above. Overall the stochastic reconfiguration method gives an efficient minimization method, which is much less prone to local minima stagnation than a conventional gradient descent minimization.

III. GAUDIN MAGNET GROUND-STATE DETERMINATION

To test success of the restricted Boltzmann machines, ground-state determination was performed on various Gaudin magnets. The system size was varied from $N = 2$ to $N = 10$ for two different choices of Hamiltonian parameters. One weighs the interaction of each environmental spin equally. Specifically, in reference to Eq. (1), the system has $B = 1$ and $A_k = 1$ for all k. The second system has inhomogeneous coupling constants whose strength exponentially decays with the number of spins. This has physical significance in modelling the hyperfine interactions between a quantum dot and neighbouring nuclear spins. The interaction strength depends on the distance of the nuclei from the center of the quantum dot [11]. This exponential decay of the coupling constant models the increasing distance to surrounding atoms. Specifically we consider $A_k = \frac{A}{N_0} e^{\frac{-k}{N_0}}$ with $B = A = N_0 = \frac{N}{2}$, and N the number of spins in the system.

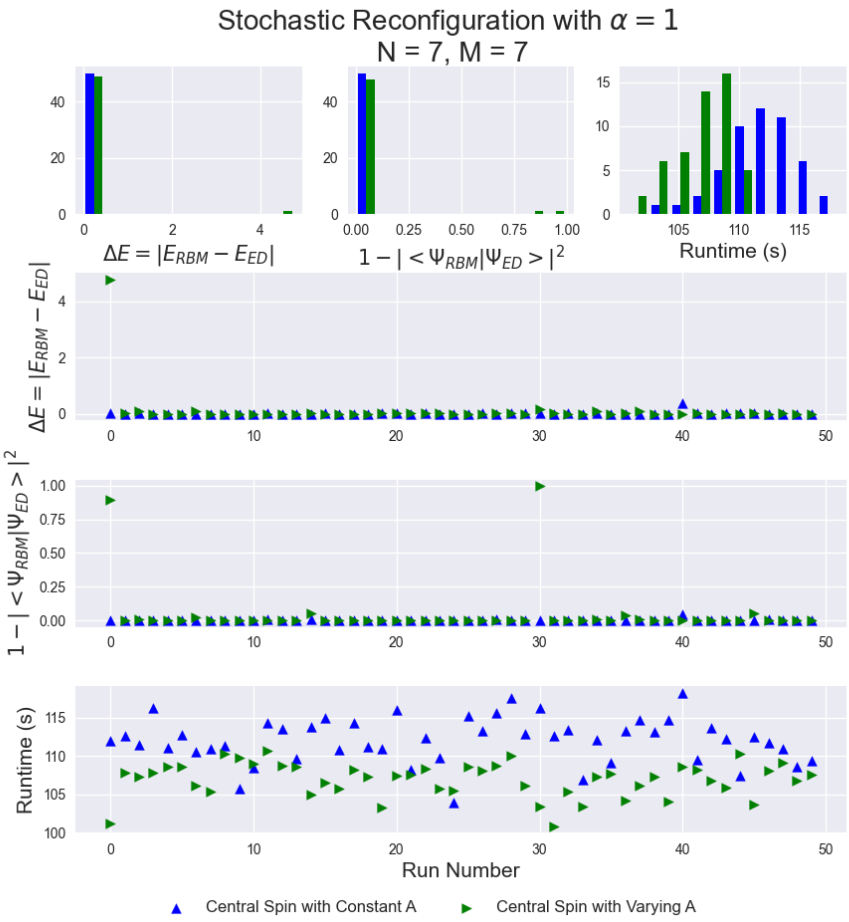

FIG. 3. Histogram of energy error, state error and runtime for both the constant coupling and exponential decay coupling case. The constant coupling Hamiltonian has has $B = 1$ and $A_k = 1$ for all k. The varying coupling constants Hamiltonian has $A_k = \frac{A}{N_0} e^{\frac{-k}{N_0}}$ with $B = A = N_0 = \frac{N}{2}$. Energy error, state error and runtime of each run are shown below.

All networks were initialized identically with RBM parameters drawn from a Gaussian distribution centered at zero with a standard deviation of 0.5. Additionally, the number of hidden nodes is kept equal to the number of visible nodes as the system size is scaled, $\alpha = \frac{N}{M} = 1$. All other hyper-parameters, such as the number of Monte Carlo samples, are kept constant for all runs. For each N, fifty runs were performed to benchmark the frequency of failure. Typically, a few runs will terminate with large energy error as the minimization gets stuck in a local minimum. A histogram of the state error, energy error, and runtime for $N = 7$ is shown in Fig. 3.

To benchmark the likelihood of failure, a failed run is defined as having a final energy error over

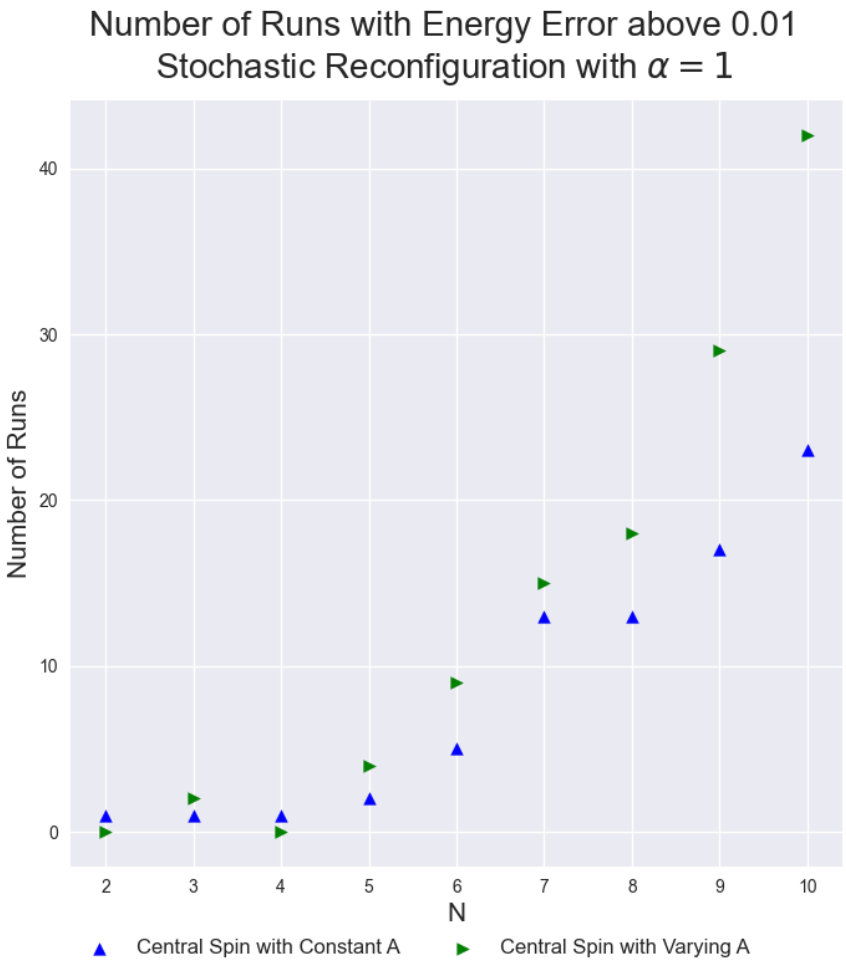

FIG. 4. Scaling of failed runs with growing system size. The constant coupling Hamiltonian has has $B = 1$ and $A_k = 1$ for all k. The varying coupling constants Hamiltonian has $A_k = \frac{A}{N_0} e^{\frac{-k}{N_0}}$ with $B = A = N_0 = \frac{N}{2}$. A failed run is defined as a final energy error over 0.01.

0.01. In Fig. 4, the number of failed runs is shown as N increases for both choices of Hamiltonian parameters considered. For large N, there is a large deviation between the constant coupling case and the case with exponentially decreasing coupling. The Hamiltonian for the constant coupling case has a symmetry allowing it to be block diagonalized [7], a known analytic process to simplify calculations. The difference in failed runs shows that the RBM is sensitive to these differences in the complexity of the Hamiltonian considered.

IV. CONCLUSION

Previously, RBMs have been shown to be successful in representing the transverse-field Ising and Heisenberg models [1]. Both of those systems describe the nearest-neighbour interactions of spins. Here, we instead considered a topology where one spin is connected to all others. Although a different connectivity, we showed an RBM is capable of describing the Gaudin magnet. By implementing sampling estimates and an efficient minimization method, an RBM was used to find the ground-state of two sets of Gaudin magnet Hamiltonian parameters. The success of the process was investigated with increasing system size. Results show an RBM is sensitive to the Hamiltonian complexity. Machine learning techniques present an exciting new tool for analyzing highly correlated many-body quantum systems. Not only do they allow the possibility to probe larger systems, but they have the potential to give further insights into the complexity of the system.

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

[2] M. J. Hartmann and G. Carleo, Neural-network approach to dissipative quantum many-body dynamics, Phys. Rev. Lett. **122**, 250502 (2019).

[3] J. Schliemann, A. Khaetskii, and D. Loss, Electron spin dynamics in quantum dots and related nanostructures due to hyperfine interaction with nuclei, Journal of Physics: Condensed Matter **15**, R1809 (2003).

[4] Yuzbashyan, Altshuler, Kuznetsov, and Enolskii, Solution for the dynamics of the BCS and central spin problems, Journal of Physics A: Mathematical and General **38**, 7831 (2005).

[5] M. Bortz and J. Stolze, Exact dynamics in the inhomogeneous central-spin model, Phys. Rev. B **76**, 014304 (2007).

[6] G. Torlai, B. Timar, E. P. L. van Nieuwenburg, H. Levine, A. Omran, A. Keesling, H. Bernien, M. Greiner, V. Vuletić, M. D. Lukin, R. G. Melko, and M. Endres, Integrating neural networks with a quantum simulator for state reconstruction, Phys. Rev. Lett. **123**, 230504 (2019).

[7] A. Orfi, F. Felix, and W. Coish, Application of machine learning algorithms to many-body quantum dynamics, PHYS 459 Final Report (2020).

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

[9] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, Equation of state calculations by fast computing machines, The Journal of Chemical Physics **21**, 1087 (1953).

[10] S. Sorella, Generalized lanczos algorithm for variational quantum monte carlo, Phys. Rev. B **64**, 024512 (2001).

[11] J. Schliemann, A. Khaetskii, and D. Loss, Electron spin dynamics in quantum dots and related nanostructures due to hyperfine interaction with nuclei, Journal of Physics: Condensed Matter **15**, R1809 (2003).