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

Alev Orfi

Supervisor: Bill Coish

Collaborator: Felix Fehse

(Dated: April, 2021)

Many-body quantum systems are notoriously difficult to describe as the number of amplitudes required to describe a general state grows exponentially with the number of particles considered. A recent machine learning technique has been shown to efficiently reparameterize the many-body wavefunction of a transverse-field Ising model and Heisenberg model. This technique, restricted Boltzmann machine (RBM), has allowed more accurate simulations of large many-body systems as compared to variational methods [1][2].

We apply an RBM to a different many-body system, the Gaudin magnet. This model describes the interaction between one central-spin and many environmental spins. One limit of the model, with uniform coupling constants, has a closed-form analytical solution. However, the more general model, with varying coupling constants, has no known simple analytical solution. We compare the success of the RBM in both of these situations. The general model differs from the Hamiltonians on which the RBM was previously tested, which are known to be tractable [1].

An RBM gives an ansatz for a many-body wavefunction, and through minimizing the variational energy this state can represent the ground state of the system. Here following [3], we implement Monte Carlo energy estimation and stochastic reconfiguration minimization to ensure this minimization is efficient. The ground state of Gaudin magnets of size up to 13 spins was successfully found, illustrating the capability of the RBM. The success of this method was investigated as the system size scales for different Hamiltonian parameters, comparing the tractable case to one without an analytical solution. Additionally, the runtime scaling of the RBM method was shown to be favourable as compared to exact diagonalization.

I. INTRODUCTION

Many-particle quantum systems are notoriously computationally challenging to simulate, as the number of amplitudes required to describe a general quantum state grows exponentially with the number of particles. For a system of N spin-$\frac{1}{2}$ particles, an individual particle is described by a two-dimensional Hilbert space. Generally, the Hamiltonian is of size $2^N \times 2^N$. However, for certain systems this Hamiltonian can be reduced to block diagonal form leading to a tractable solution. In the case where this reduction is not possible, it is necessary to use numerical techniques, such as exact diagonalization. As the dimension of the Hilbert space grows exponentially with the number of spins, these numerical calculations of large systems become impractical. Specifically, numerical exact diagonalization involves finding the 2^N eigenstates of the system's Hamiltonian, a computationally challenging task.

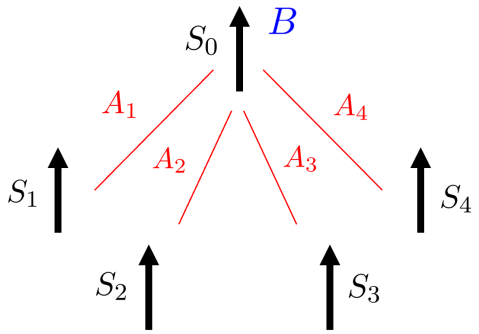

FIG. 1. Diagrammatic representation of the Gaudin magnet where S_i are the individual spin operators, A_i the coupling between the i^{th} spin and the central spin, and B the external field.

The Gaudin magnet [4] is a many-body system which describes the interactions between N spins, one central spin and $N-1$ bath spins. 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 bath spins is parameterized by A_k, a set of coupling constants. The bath spins act as an environment that influences the central spin. This model is represented with the following Hamiltonian,

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

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 has applications modelling the decoherence of an electron spin confined to a quantum dot. The above Hamiltonian can model the hyperfine interaction between an electron spin in a quantum dot and surrounding nuclear spins [5]. Therefore, by considering the Gaudin magnet's dynamics, the decoherence of an electron in a quantum dot due to neighbouring nuclear spins can be tracked. Additionally, a linear combination of commuting Gaudin magnets describes the Gaudin-Richardson model, which is often a starting point for a mean-field treatment of superconductivity[6]. In the case of arbitrary coupling constants A_i, the Gaudin magnet is not tractable. Numerical exact diagonalization quickly becomes too computationally challenging, thus the system requires new numerical techniques [7]. More sophisticated methods, such as the Bethe ansatz, lead to dynamics calculations of systems of around 30 spins [4]. Recently machine-learning methods have shown success in representing many-body wavefunctions [1]. Our goal is to implement these methods with the hope of reducing the computational complexity of this problem, thus allowing the possibility of 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.

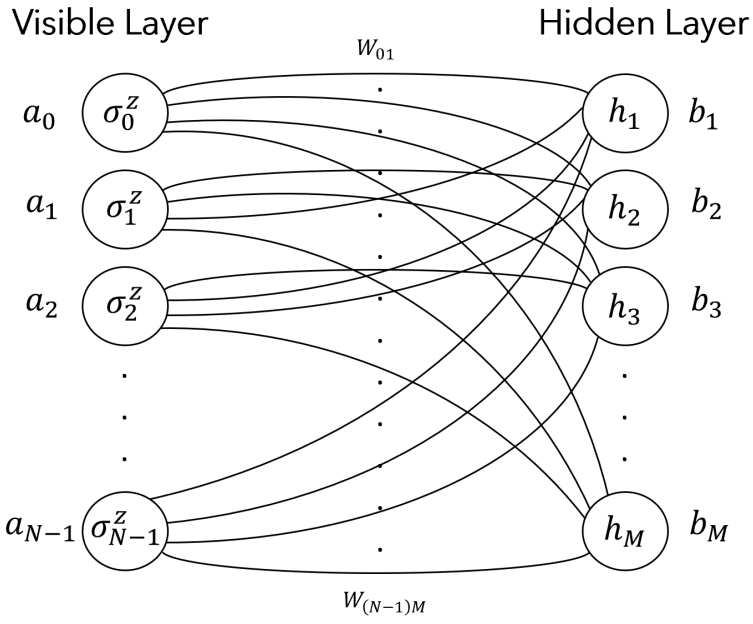

FIG. 2. A restricted Boltzmann machine architecture applied to the Gaudin magnet. The visible layer, shown on the left, is composed of visible spin variables $S = \{\sigma_0^z, \sigma_1^z, ..., \sigma_{N-1}^z\}$. The hidden layer, on the right, holds the hidden variables $h_1, h_2, ..., h_M$.

Machine-learning methods are designed to learn the important correlations of a system and thus disregard less impactful information. This is advantageous in quantum systems and as a network

can be trained to more efficiently represent a many-body wavefunction, allowing larger systems to be solved. Here we apply a specific network, a restricted Boltzmann machine (RBM), which is an effective many-body wavefunction reparameterization. Specifically, it 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 perform state tomography from experimental data [8].

A restricted Boltzmann machine is a two-layer network that gives a variational ansatz of the wavefunction in terms of the complex network parameters $\mathcal{W} = \{a_i, b_j, W_{ij}\}$. A pictorial representation of the network structure is shown in Fig. 2. The RBM parameterizes the coefficients of the wavefunction in the computational basis $\{\sigma\}$,

$$|\psi\rangle = \sum_{\{\sigma\}} C_\sigma |\sigma\rangle \simeq \sum_{\{\sigma\}} \Psi(\sigma; \mathcal{W}) |\sigma\rangle =: |\psi_{\mathrm{RBM}}(\mathcal{W}(t))\rangle . \tag{2}$$

The two layers of the RBM, seen in Fig. 2, are comprised of interacting stochastic variables, known as nodes. The visible layer nodes correspond to the system's physical parameters, here the eigenvalues of the Pauli-z operator, σ^z, acting at each spin site. Therefore there are N visible layer nodes, as seen in Fig. 2. 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 value of M can be modified to change the expressivity of the network. Each node is connected to every node in the opposite layer with a weight parameter W_{ij}. Additionally, the visible and hidden nodes have corresponding biases, $a_0, ..., a_{N-1}$ and $b_1, ..., b_M$, respectively. The specific form of this parameterization in terms of these network parameters $\mathcal{W} = \{a, b, W_{ij}\}$ is given by [1],

$$\Psi(S; \mathcal{W}(t)) = \sum_{\{h_i\}} e^{\sum_j a_j \sigma_j^z + \sum_i b_i h_i + \sum_{ij} W_{ij} h_i \sigma_j^z}. \tag{3}$$

Training of the RBM depends on the intended use of the network. Here we focus on ground-state determination, however, an RBM has been shown to also model system dynamics [1]. The network parameters are modified until an accurate approximation of the state is found. The cost function specified to found the ground state of the system is the variational energy,

$$E(\mathcal{W}) = \frac{\langle \psi_{\mathrm{RBM}}| H |\psi_{\mathrm{RBM}}\rangle}{\langle \psi_{\mathrm{RBM}}|\psi_{\mathrm{RBM}}\rangle}. \tag{4}$$

Minimizing this function thus gives the optimal network parameters, $\mathcal{W} = \{a, b, W_{ij}\}$, to represent the ground-state.

A. Previous Work

The work outlined here is a continuation of an undergraduate thesis project [9], as well as a Phys 396 project. During the thesis project a method for determining the success of the RBM training was developed and tested. An RBM was created and tested for small systems of both Heisenberg and Gaudin magnet systems. The network was later improved through the implementation of the NetKet package [3]. Netket offers a framework for using machine learning methods specifically designed for many-body quantum systems. Most notably, the use of Monte Carlo sampling and different minimization methods. Multiple training algorithms were investigated to minimize the variational energy, Eq. (4), the most successful being stochastic reconfiguration [3]. This term focused on updating previous work to NetKet version 3, then considering how the RBM performs with increasing system size.

II. ALGORITHM OVERVIEW

The real and imaginary components of RBM parameters are initialized from a gaussian distribution centered at zero with a standard deviation of 0.25. This value was found to lead to the largest number of successful runs for our Hamiltonian. This initialization gives enough variability for the minimization procedure to modify these parameters until the minimum variational energy, Eq. (4), is found.

The RBM gives the form of the variational wavefunction, but to represent the ground state the variational energy must then be minimized. To calculate the variational energy at each step of minimization, it is necessary to evaluate an expectation value with respect to the current estimate of the ground state, as shown in Eq. (4). Since this estimate generally contains an exponentially large number of amplitudes, directly evaluating the exact expectation value at each step of the minimization quickly becomes computationally intractable for large N. Instead, following [1], we approximate the expectation value at each step. Instead of using a basis of all spin configurations $\{\sigma\}$, only using a subset of these configurations $\{\tilde{\sigma}\}$,

$$\langle\psi_{\mathrm{RBM}}|\, H\,|\psi_{\mathrm{RBM}}\rangle = \frac{\sum_{\sigma,\sigma'}\psi^*_{\mathrm{RBM}}(\sigma)\,\langle\sigma|\, H\,|\sigma'\rangle\,\psi_{\mathrm{RBM}}(\sigma')}{\sum_{\sigma}|\psi_{\mathrm{RBM}}(\sigma)|^2} \tag{5}$$

$$= \sum_{\sigma}\Big(\sum_{\sigma'}\langle\sigma|\, H\,|\sigma'\rangle\,\frac{\psi_{\mathrm{RBM}}(\sigma')}{\psi_{\mathrm{RBM}}(\sigma)}\Big)\frac{|\psi_{\mathrm{RBM}}(\sigma)|^2}{\sum_{\sigma'}|\psi_{\mathrm{RBM}}(\sigma')|^2} \tag{6}$$

$$\approx \Big\langle\sum_{\sigma'}\langle\sigma|\, H\,|\sigma'\rangle\,\frac{\psi_{\mathrm{RBM}}(\sigma')}{\psi_{\mathrm{RBM}}(\sigma)}\Big\rangle_{\tilde{\sigma}} =: E_{\mathrm{local}}, \tag{7}$$

where in Eq. (7) the expectation value is taken over a sample of configurations $\{\tilde{\sigma}\}$ drawn from the following probability distribution,

$$\pi(\sigma) = \frac{|\psi_{\mathrm{RBM}}(\sigma)|^2}{\sum_{\sigma'}|\psi_{\mathrm{RBM}}(\sigma')|^2}. \tag{8}$$

The set $\{\tilde{\sigma}\}$ is generated using the Metropolis algorithm, which generates new configurations drawn from the distribution Eq. (8) [10].The accuracy of this expectation value estimate can be improved by increasing the number of the generated configurations.

Various minimization methods have been tested with the most successful found to be stochastic reconfiguration [3][11]. 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,

$$\mathcal{O}_k = \frac{1}{\psi_{\mathrm{RBM}}}\partial_{\mathcal{W}_k}\psi_{\mathrm{RBM}}. \tag{9}$$

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

$$\mathcal{W}(p+1) = \mathcal{W}(p) - \gamma S^{-1}(p)F(p). \tag{10}$$

Where here γ is a learning rate and S is the following covariance matrix,

$$S_{kk'}(p) = \langle\mathcal{O}^*_k\mathcal{O}_{k'}\rangle - \langle\mathcal{O}^*_k\rangle\langle\mathcal{O}_{k'}\rangle. \tag{11}$$

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

$$F_k(p) = \langle E_{\mathrm{local}}\mathcal{O}^*_k\rangle - \langle E_{\mathrm{local}}\rangle\langle\mathcal{O}^*_k\rangle, \tag{12}$$

with E_{local} defined in Eq. (7).

The expectation values in Eq. (11) and Eq. (12) are evaluated stochastically with the sampling method above to ensure subexponential scaling of this method. 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

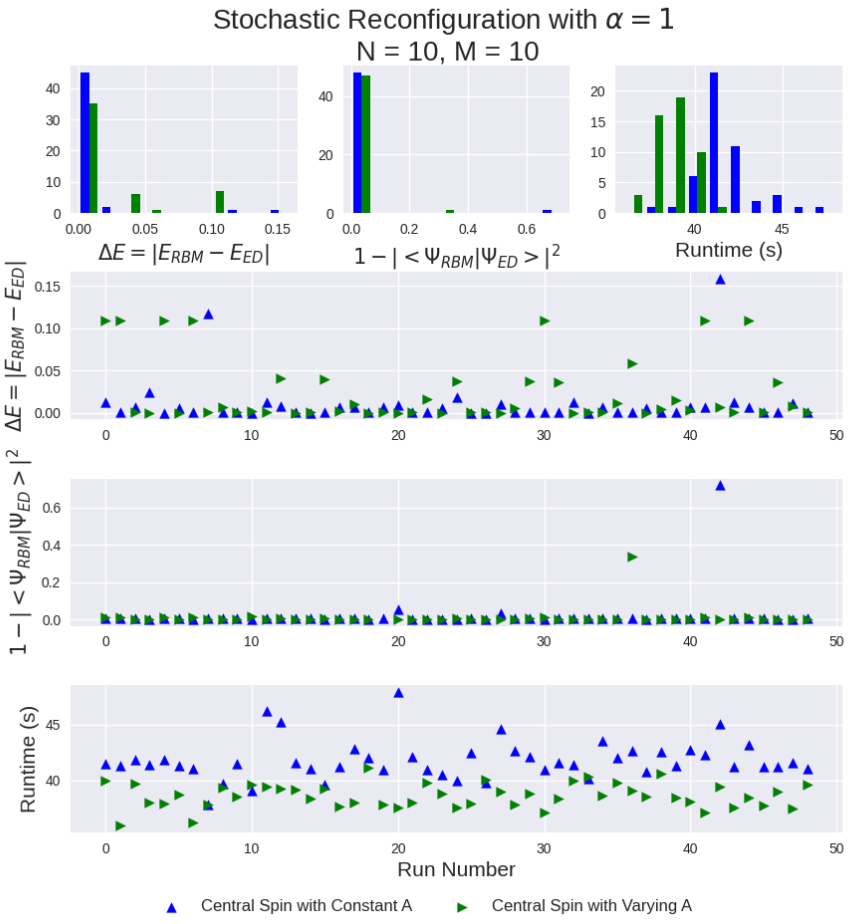

FIG. 3. Histogram of energy error, state error and runtime for both the constant coupling and exponential decay coupling case. The constant coupling Hamiltonian has has $B = 0.95$ and $A_k = 1$ for all k. The varying coupling constants Hamiltonian has $A_k = \frac{A}{N_0} e^{\frac{-k}{N_0}}$ with $B = 0.95\frac{N}{2}, A = N_0 = \frac{N}{2}$. Energy error, state error and runtime of each run are shown below.

To benchmark the restricted Boltzmann machine representation, an RBM was trained on systems of size $N = 2$ to $N = 13$ for two different choices of Hamiltonian parameters. The first has constant

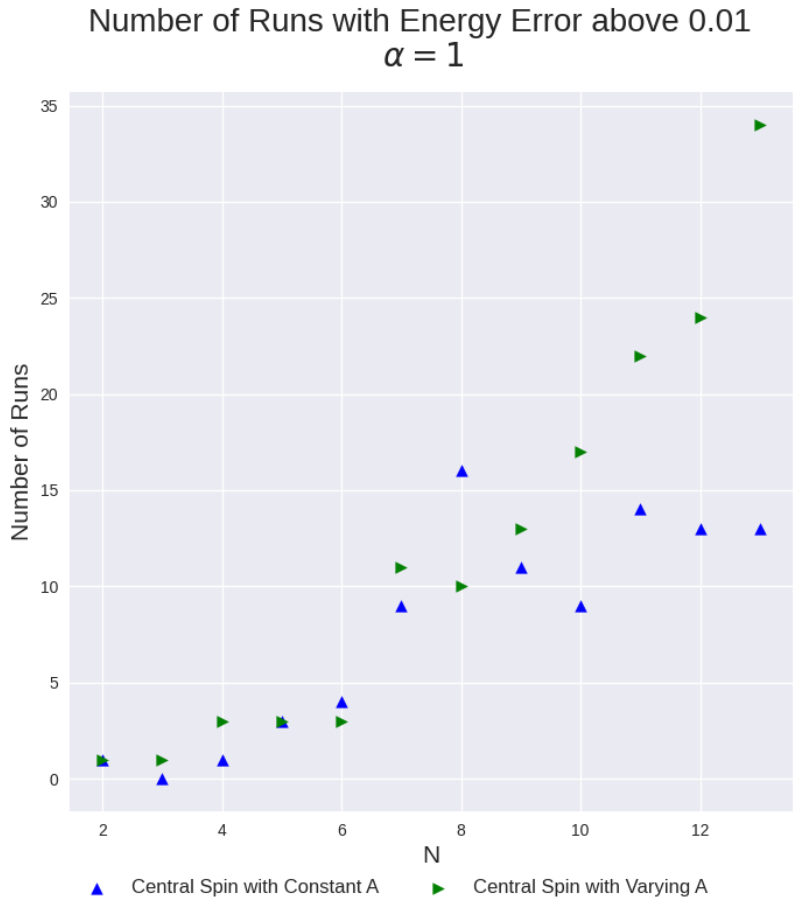

FIG. 4. Number of failed runs out of 50 with increasing system size N for constant coupling Hamiltonian and exponential decay coupling Hamiltonian. A failed run is defined as a final ground state energy error, $\Delta E = |E_{\mathrm{RBM}} - E_{\mathrm{ED}}|$, over 0.01. The constant coupling Hamiltonian has has $B = 0.95$ 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 $B = 0.95, A = N_0 = \frac{N}{2}$.

interaction between each environmental spin. Specifically, in reference to Eq. (1), the system has $B = 0.95$ and $A_k = 1$ for all k. The choice of $B = 0.95$ is the ensure there are no low level degeneracies which may cause the minimization to get stuck. 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 [12]. 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 $B = 0.95, A = N_0 = \frac{N}{2}$, and N the number of spins in the system. Both networks have the same number of hidden and visible nodes, $\alpha = \frac{N}{M} = 1$. All other hyper-parameters, such as the number of Monte Carlo samples, are

kept constant. Additionally all networks were initialized identically with RBM parameters drawn from a Gaussian distribution centered at zero with a standard deviation of 0.5.

Typically, some runs will terminate with significant energy error as the minimization gets stuck in a local minimum. Histograms of the state error, energy error, and runtime for $N = 10$ are shown in Fig. 3. In Fig. 3 the energy error, state error and runtime are displayed for each run.

To benchmark the likelihood of failure, a failed run is defined as having a final energy error over 0.01. In Fig. 4, the number of failed runs is shown as N increases for both choices of Hamiltonian parameters considered. Both systems see an increasing rate of failure with larger N. This result is expected as no hyperparameters, such as the number of minimization iterations, are changed with increasing N. For large N, there is a deviation between the constant coupling case and varying coupling case. The constant coupling case has a known analytic solution, whereas the varying case does not. This difference in RBM success may indicate this network is sensitive to these kinds of differences in the Hamiltonian structure.

IV. RUNTIME SCALING

An interesting metric for understanding an RBM method is runtime scaling. In Fig. 5 the RBM runtime is compared to exact diagonalization for a Hamiltonian with constant coupling. Specifically, the system has $B = 1$ and $A_k = 1$ for all k. For all runs $\alpha = \frac{N}{M} = 1$ and all other hyper-parameters are kept constant. Similarly to above, the RBM was initialized from a Gaussian distribution centered at zero with a standard deviation of 0.25.

There is a clear scaling advantage in the RBM runtime compared to exact diagonalization. Here you can see that even on systems small enough to run on a local computer, there is a crossover where the RBM has lower runtime. However, as the RBM hyperparameters are not varied with increasing N, we have an increasing ground state energy error. A better comparison would be between an RBM runtime with fixed error as bounding this error growth may have a significant computational cost.

V. CONCLUSION

Restricted Boltzmann machines have been shown to be a successful many-body ansatz for representing the transverse-field Ising and Heisenberg models [1]. The work presented here applies this

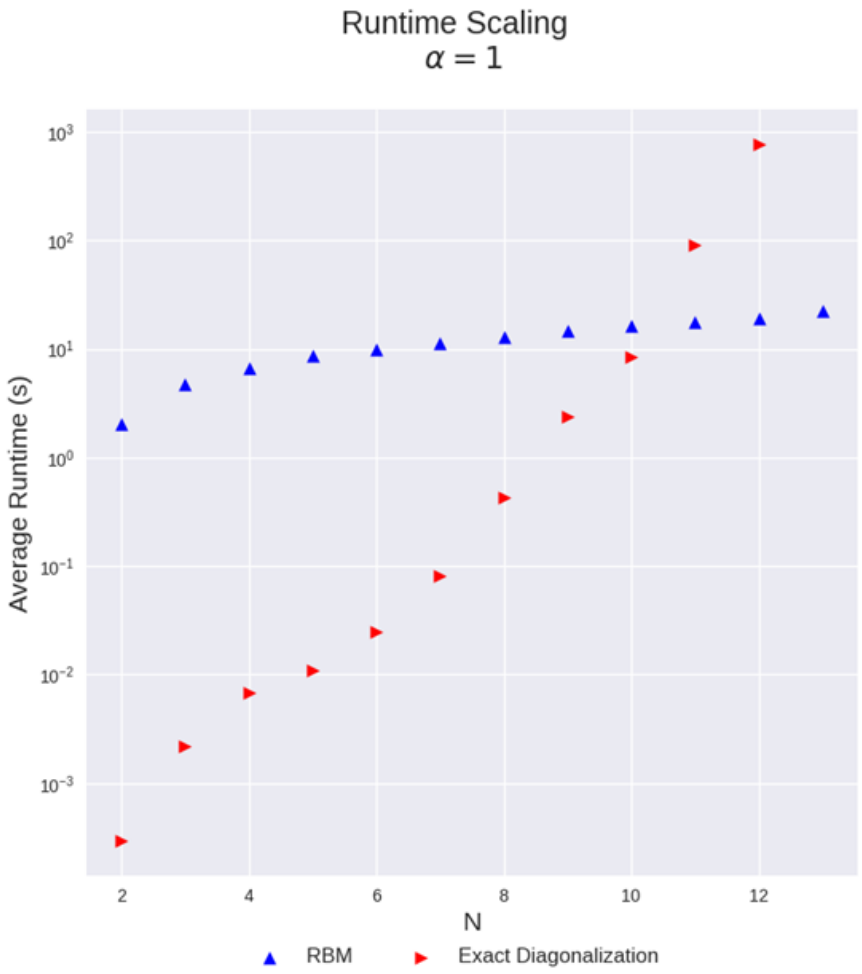

FIG. 5. Run time comparison between RBM and exact diagonalization. Here the Hamiltonian considered has constant coupling; $B = 1$ and $A_k = 1$ for all k.

RBM method to a different many-body system, the Gaudin magnet. Through the implementation of NetKet version 3 [3], which includes sampling estimates and stochastic reconfiguration minimization, an RBM can efficiently find the ground state of a Gaudin magnet. We considered this method's failure rate with increasing system size for two different sets of Hamiltonian parameters. We compared the runtime scaling of an RBM and exact diagonalization, showing a scaling advantage of this method. Machine learning techniques present an exciting tool for analyzing 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 physics 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] 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).

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

[5] 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).

[6] 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).

[7] A. Faribault and D. Schuricht, Spin decoherence due to a randomly fluctuating spin bath, Phys. Rev. B **88**, 085323 (2013).

[8] 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).

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

[10] 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).

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

[12] 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).