Many-body Quantum State Representation with Neural Networks

Andrew Gleeson*[1] and Ben Safvati[2]

Department of Physics, University of California at Berkeley
366 LeConte Hall, Berkeley, CA

**Department of Electrical Engineering and Computer Science, University of California at Berkeley*
253 Cory Hall, Berkeley, CA

[1]gleeson@berkeley.edu
[2]bsafvati@berkeley.edu

***Abstract*— Describing the quantum state of a system of many interacting particles is a central research area in Quantum Mechanics and Computational Physics. Simulation of many-body quantum states proves difficult using standard descriptions of these states as elements of a Hilbert Space because the dimensionality of this space grows exponentially large with respect to the number of particles. Modern efforts have aimed to find a more efficient model to describe the nonlocal quantum correlations between particles without suffering computationally due to the dimensionality of the state space. We demonstrate the performance of the Restricted Boltzmann Machine neural network in representing the quantum state, training on a smaller amount of variational parameters with computational costs growing polynomially with respect to number of variables trained. Using two physical models for spin interactions, we show how to train the RBM to learn the ground state of the system by Monte Carlo sampling, stochastically updating the parameters in the direction of steepest descent to the state of lowest energy. This method develops the Restricted Boltzmann Machine as a "black box" for the quantum state of the particles, carrying the information underlying the the quantum correlations between spins and effectively computing the ground state energy and spin correlations for two spin models with the trained model parameters. Our computations show that this architecture's representational power is comparable to the state of the art in computational many-body methods.**

I. THE CURSE OF DIMENSIONALITY

The Axioms of Quantum Mechanics [1] as developed in the 1930s describe the quantum state of a physical system as a ray in a Hilbert Space. This state vector contains all relevant information about the physical properties of the system under study, with each coefficient of the state providing a probability distribution over the set of possible observation outcomes for any measurement of the system. While this physical model is generally accepted as the most accurate description of reality so far, when dealing with systems of many particles the methods of Quantum Mechanics become intractable; specifically, for a system of N particles each with d degrees of freedom[1], the Hilbert Space of the state vector is d^N-dimensional, exponentially large in the number of particles and thus far too large for computations involving systems on practical scales. Computational physicists have worked extensively on methods for approximating properties of many-body systems, using either stochastic approaches that compute observable quantities as expectation values over samples drawn from the target distribution [3], or by finding compressed representations [4, 5] of the quantum state within the Hilbert Space that efficiently carry the state's information with sub-exponential scaling of parameters with system size.

Because the Many-Body Problem ultimately involves approximating a function of exponentially many parameters, Machine Learning provides ample tools for finding efficient models that approximate the coefficients of the state vector. In this paper we adopt the Neural Quantum State Ansatz of Carleo and Troyer [6] for analysis of the representational power of Neural Networks in describing a quantum state, where the network is trained on data sampled from the probability distribution of the system under study and iteratively optimized in a feedback scheme until convergence to the desired state. After describing the project architecture and the training procedure, we continue to demonstrate the effectiveness of this method in capturing particle-particle correlations for different physical environments. We will use our calculated correlations after training to develop a geometric intuition of quantum states when acted upon by archetypal many-body environments, where the range of the correlations acts as a quantitative measure of the complexity of the quantum state.

II. ILLUSION OF HILBERT SPACE

In order to efficiently characterize a quantum system, we must have an idea of the underlying structure of the interactions that govern which quantum states are preferred and which are less common. We begin by formalizing notation and restricting our attention to systems of spin 1/2 particles on a lattice, a typical platform for exploring particle interactions. Each particle σ_i at lattice site i is a qubit—a two-level system with states "spin-up" $\sigma_i = 1$ or "spin-

[1]We are purposefully vague here, and will specialize to the specific case of spin systems that we study later. More generally, degrees of freedom can be any parameter that when varied changes the overall state, and they are generally specified by quantum numbers [2].

down" $\sigma_i = -1$ corresponding to energetically different preferred orientations of the particle. For a system of n particles on such a lattice, the state of the entire system is

$$|\Psi\rangle = \sum_{S \in \{-1,1\}^n} \psi(S) \, |S\rangle$$

This is known in Quantum Physics as a superposition of states where the squared norm of the state amplitudes $|\psi(S)|^2$ provide the probability of measuring that spin configuration upon observation. Here $|S\rangle$ is standard quantum notation for a column vector (traditionally called a ket) that represents one of the basis states of the Hilbert Space formed by n spins. $S \in \{-1, 1\}^n$ represents a possible configuration of the n spins, and $\psi(S)$ is a complex-valued coefficient associated with the basis vector $|S\rangle$. Because there are 2^n possible arrangements of the n spins, we see that representation of a general quantum state quickly becomes impractical as we involve more spins.

In order to find a state representation model that approximates quantum states accurately for practical purposes, we must investigate the manifold of physically relevant states within Hilbert Space to see if this exponentially large ambient dimension is ever fully explored. The evolution of a quantum state is governed by the Hamiltonian operator H, a Hermitian operator whose eigenvalues are the energy levels of the system and whose eigenstates correspond to the characteristic energy states induced by this Hamiltonian that the spin configuration can occupy. These Hamiltonians are completely general and can model any type of quantum evolution, but if we utilize our knowledge of the range of interactions induced by this Hamiltonian it is possible to characterize the space of relevant states as those where distant particles have limited correlations.

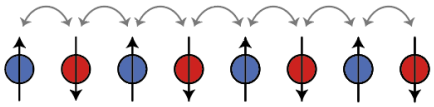

Fig. 1. An example of a 1D antiferromagnetic spin chain where nearest neighbors are anti-correlated. Atomic spin orderings such as this help to illuminate the basis of macroscopic magnetic phenomena in solid materials. Hamiltonian models that have these orderings as ground states of minimum energy are heavily studied and simulated to predict properties of similar many-body systems.

Characterizing the complexity of different Hamiltonians [8] is an emerging field with similarities to classical complexity theory, and in the context of quantum simulation our interest is in those Hamiltonians with local structure. These Hamiltonians, often used as toy models for interactions in solid-state materials, have exponentially decaying correlations [9] as the distance between particles increases, and so the local Hamiltonian will only have non-zero entries corresponding to nearest-neighbor particle interactions. This characterization allows us to interpret the Hamiltonian as a sparse matrix, drastically reducing the necessary computations involved when solving for the energy eigenstates of the system. This class of Hamiltonian operators with local interactions is also limited in that the evolution of a state under this Hamiltonian will be closed within a sub-exponential manifold of the exponentially large Hilbert Space [4], and so we can deduce that the states we seek to represent under the evolution of these Hamiltonians will be those with limited correlations between non-neighboring particles.

III. Learning Quantum States

Because the Hamiltonian's dimensions grow exponentially with the number of particles, typical matrix diagonalization to find the energy eigenvalues is again made intractable due to the burden of high dimensionality, so we must leverage the sparsity assumption of H to develop a more efficient means of finding the eigenstates of these operators.

The Hamiltonian is the central component of the Schrodinger Equation, which governs quantum state evolution. For the nth energy eigenstate $|\Psi_n\rangle$ this equation tells us that

$$H \, |\Psi_n\rangle = E_n \, |\Psi_n\rangle$$

Where E_n is the energy of the nth eigenstate. From this equation we can cast this eigenvalue problem as a variational optimization over the set of all states in the Hilbert Space. The expectation value of the energy for an arbitrary state $|\Psi\rangle$ can be found as

$$\langle E \rangle = \langle \Psi | \, H \, | \Psi \rangle$$

where $\langle \Psi|$ is a row vector, the conjugate transpose of $|\Psi\rangle$.Then utilizing the quadratic structure of the expectation we know that this expectation is bounded below by the lowest energy eigenvalue of the Hamiltonian, also known as the ground state energy. Thus we can formulate this equation as our objective function for finding the ground state of the Hamiltonian with the optimization problem

$$|\Psi_0\rangle = \min_{|\Psi\rangle} \langle \Psi | \, H \, | \Psi \rangle$$

Although this quadratic form still suffers from exponentially large dimensions relative to system size, a landmark result [10] by McMillan showed how this expectation can be calculated numerically by averaging over the probability distribution of the target quantum state. We can show this by algebraic manipulation of objective function, noting that we can rewrite the quadratic form as a sum over pairs of configurations (S, S') where the element $H_{S,S'} = \langle S | \, H \, | S' \rangle$ is explicitly operated on in the calculation below[2]

$$\frac{\langle \Psi | \, H \, | \Psi \rangle}{\langle \Psi | \, | \Psi \rangle} = \frac{\sum_{S,S'} \psi(S)^* \, \langle S | \, H \, | S' \rangle \, \psi(S')}{\sum_S |\psi(S)|^2}$$

[2]We divide our objective by a normalizing factor $\langle \Psi | \, | \Psi \rangle$ by convention and to prevent the amplitudes from growing too large.

$$\frac{\langle\Psi|\, H\, |\Psi\rangle}{\langle\Psi|\, |\Psi\rangle} = \frac{\sum_S (|\psi(S)|^2 \sum_{S'} H_{S,S'} \frac{\psi(S')}{\psi(S)})}{\sum_S |\psi(S)|^2}$$

In this form the expectation value of the energy can be interpreted as a statistical average over the probability distribution determined by the quantum state. the quantity $E_{loc}(S) = \sum_{S'} H_{S,S'} \frac{\psi(S')}{\psi(S)}$, known as the local energy of configuration S, is efficiently computed for local Hamiltonians where the sparsity of non-zero elements is guaranteed. Finding these non-zero elements depends on the kind of Hamiltonian, and we will outline this method for two Hamiltonian models that we implemented in the Data Analysis section.

This variational approach is the backbone of Quantum Monte Carlo methods [11], where high-dimensional matrix products (or in the case of continuous variables high-dimensional integrals) are framed as expectations over a probability distribution. From this energy objective function we can find the direction of steepest descent by taking the gradient with respect to each parameter in the quantum state model, allowing us to step toward the state of minimum energy iteratively. If we shape the model parameters into a vector $\theta \in \mathbb{C}^{N_{var}}$ where N_{var} is the number of model parameters, then we can update our parameters using the equation

$$\theta^{t+1} = \theta^t - \gamma F^t$$

where F is the gradient of the energy function, called in literature as the generalized forces of the system by analogy with concepts in physics. When we define the variational derivative of the ith parameter as

$$O_i(S) = \frac{1}{\psi(S)} \frac{\partial}{\partial \theta_i} \psi(S)$$

then the generalized force for the ith parameter can also be framed as an expectation over the quantum state probability distribution (see [12] for the derivation)

$$F_i = \langle E_{loc} O_i^* \rangle - \langle O_i^* \rangle \langle E_{loc} \rangle$$

This gradient expression iteratively updates the parameters similar to the stochastic gradient descent methods that we have studied. This formulation, however, fails to capture the notion of distance within the Hilbert Space that the quantum state lives in. In particular, this descent method is implicitly defined on a Euclidean metric for the parameter space [13], whereas in describing the quantum state the update should be measured on the Fubini-Study metric [6] of the Hilbert Space. We must thus transform our gradient update to account for this new metric, where small steps in the parameter space can correspond to large updates to the complex state amplitudes that the model computes as a function of those parameters. This transformation is achieved by the Stochastic Reconfiguration algorithm as formulated by Sorella et al. [14] with the new parameter update

$$\theta^{t+1} = \theta^t - \gamma S^{-1} F^t$$

with the matrix S defined as

$$S_{i,j} = \langle O_i^* O_j \rangle - \langle O_i^* \rangle \langle O_j \rangle$$

This matrix performs the transformation of the updates to the appropriate lengths induced by the Hilbert Space that quantum states occupy. This data augmentation is similar to other work in finding the "natural gradient" [15] for a set of parameters based on the geometry of the data space. We find that this correction to the parameter updates leads to more consistent convergence of the model to the correct ground state energies for all models learned.

IV. MONTE CARLO STATE SAMPLING

The developed training model's accuracy depends on the set of sample configurations $\{S_i\}$ used to compute the expectations involved in the stochastic parameter updates. Because our model approximates the ground state $|\Psi_0\rangle$, we want the expectation to be over the probability distribution determined by the amplitudes of $|\Psi_0\rangle$. This can be done by taking random walks in the space of spin configurations and transitioning to that state with probability determined by the ratio of quantum state amplitudes. This process has a Markov chain interpretation [6] where a new spin configuration is added to the chain with probability

$$P(S^{k+1}) = \min\left(1, \left|\frac{\psi(S^{k+1})}{\psi(S^k)}\right|^2\right)$$

After several thousand iterations we have a large enough training set to compute parameter updates using the Stochastic Reconfiguration Algorithm. As our model parameters improve the estimates for the state amplitudes will also improve, which then can be used to generate a new Markov Chain of spin configurations that more closely resembles the ground state. This feedback process is readily implemented with the Restricted Boltzmann Machine neural network as explained below, allowing us to iteratively approach the exact ground state in times polynomial with respect to the number of variational parameters, avoiding the exploding dimensionality of the system as we increase the number of particles.

V. MODEL ARCHITECTURE

While quantum states can be represented by many neural network topologies depending on the specific Hamiltonian, Restricted Boltzmann Machines have proven to be excellent models for quantum states because they have a physically motivated interpretation in terms of a energy function that allows us to use the ideas above for sampling and training of the parameters.

A Restricted Boltzmann Machine consists of a layer of n visible neurons $\sigma_1, ..., \sigma_n$ which correspond to the spins, attached to a layer of m hidden units $h_1, ..., h_m$. We can then

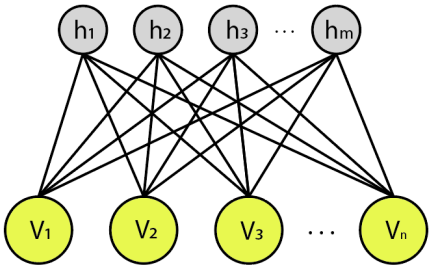

Fig. 2. A Restricted Boltzmann Machine (RBM) with n visible units and m hidden units.

express a quantum state

$$\psi(\sigma) = \sum_{\{h_i\}} \exp\left(\sum_{j=1}^{n} a_j\sigma_j + \sum_{i=1}^{m} b_i h_i + \sum_{i=1,\, j=1}^{m,\, n} W_{ij} h_i \sigma_j\right)$$

where a_j, b_i, and W_{ij} are the parameters of neural network. We take these parameters to be complex, which allows us to compute both a magnitude and a phase when calculating the quantum state. The Restricted Boltzmann Machine is of special interest because there are no interactions between neurons in the same layer, which allows us to explicitly calculate out the values of h_i, finding an explicit represtation of the quantum state ϕ in terms of just the network parameters and visible spins:

$$\psi(\sigma) = \exp\left(\sum_{i=1}^{n} a_i\sigma_i\right) \prod_{i=1}^{m} \left(2\cosh\left[b_i + \sum_{j=1}^{n} W_{ij}\sigma_j\right]\right)$$

We define a hyperparameter $\alpha \equiv m/n$, the ratio of hidden units to variable units. In principle, by increasing α we can continually improve the representative ability of the RBM, because adding hidden units adds complexity to the model. For simple Hamiltonians, a small α can work, but as the complexity of the Hamiltonian grows having enough hidden units becomes essential.

VI. DATA ANALYSIS

We trained Restricted Boltzmann Machines on two spin models, the 1-D Ising model and the 1-D Heisenberg model. These models are characteristic in the field of condensed matter, as they consider both how an external magnetic field affects particles with spin on a lattice and how the spins interact with each other.

For both of these models, we use periodic boundary conditions, which means that the particles at each end of the lattice are considered to be neighbors (consider a ring of atoms).

The Ising model is parameterized by the magnetic field strength h. When h is large, it dominates inter-spin interactions and dampens long-range spin interactions. In a weaker field, the spins are able to interact more freely.

$$H_{\text{Ising}} = -\sum_{i=1}^{n} \sigma_i\sigma_{i+1} - h\sum_{i=1}^{n} \sigma_i$$

The Heisenberg model is more complex in that there is an extra term in the Hamiltonian proportional to a parameter J that is added to the local energy calculation for lattice sites where neighboring spins are anticorrelated.

$$H_{\text{Heisenberg}} = -J\sum_{i=1}^{n} \sigma_i\sigma_{i+1} - \sum_{i=1}^{n} \sigma_i$$

When J is positive, differing adjacent spins lower the energy, so the ground states should exhibit sets of anticorrelated spins relative to their neighbors. When J is negative, anticorrelated neighboring spins increase the energy of the total system, so in the ground state all spins are aligned in one direction. The magnitude of J determines the strength of this effect. Because these models consider only local interactions, their Hamiltonians are sparse and computations will be efficient for reasons discussed previously.

We typically trained for 1000 epochs at 3000 iterations per epoch, and were able to achieve error rates on the ground state energies for both models between 10^{-5} and 10^{-7}, which is consistent with other recent results in the literature [6]. Additionally, we can improve the error without bound by increasing the number of epochs and iterations. The algorithm is extremely sensitive to its hyperparameters, and requires careful tuning to find suitable configurations for convergence to the ground state.

A. HYPERPARAMETER TUNING

Restricted Boltzmann Machine
- $\alpha = 4$

Stochastic Reconfiguration Algorithm
- Learning Rate = 0.8
- Epochs = 1000

MCMC Sampling
- Iterations = 1000
- Thermalization
  - Thermalization Factor = 0.01
  - Sweep Factor = 1

Fig. 3. This is an example of a hyperparameter configuration which converges for the 1-D Ising Hamiltonian.

In order to achieve optimal performance, we need to choose values for each of the hyperparameters that determine a particular instance of the algorithm. For some, like the thermalization factor and the sweep factor in Stochastic Reconfiguration, we are able to use common values found in the literature because the overall algorithm is not very sensitive to them. Other parameters, like the number of epochs to run the Stochastic Reconfiguration algorithm and the number of iterations for each sampling procedure, face a trade-off between accuracy and time, as increasing the number of steps can only improve the accuracy of the predictions. Therefore, we can strategically pick values for these hyperparameters

depending on the desired level of accuracy and computation time.

1) LEARNING RATE AND REGULARIZATION: The Stochastic Reconfiguration algorithm is very sensitive to the learning rate at which we update the parameters of our model. Additionally, we found that because S is a sparse matrix, inverting it can cause numerical divergence problems, so we added an L2 regularization term λ to the matrix when solving for the parameter updates. Both of these model parameters must be taken into account in order to make the algorithm converge. If the learning rate and regularization are both large, then the Stochastic Reconfiguration algorithm is taking large steps across the parameter space with updates that have high bias but low variance. We begin with large values because we are initializing our parameters randomly—we do not expect the parameters to be initially near the optimal configuration, so large changes are useful for roughly approaching the true solution. However, we want finer resolution as the parameters approach the optimum so that there is no risk of taking steps around the true solution and missing it. However, if the learning rate gets small and λ remains large, then the updates are still high bias, which will draw the solution away from the true optimum. But if λ gets too small, $S + \lambda I$ approaches a singular matrix and can cause divergence. Thus we must carefully lower both parameters together in order to achieve convergence. We found good performance with a decaying learning rate and ridge parameter

$$\text{Learning Rate(Iteration } i) = \frac{c}{\sqrt{i}}$$

$$\lambda(\text{Iteration } i) = 100 \cdot (0.85)^i$$

When we refer to learning rate values later in the text, we will refer to the constant c. Additionally, to prevent singularities, we set a lower bound of 10^{-4} on λ for instances with a high number of update iterations.

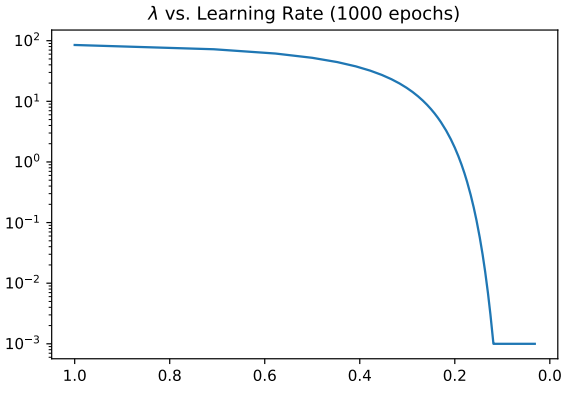

Fig. 4. An example of how the learning rate and regularization parameter change together after 1000 iterations ($c = 1$).

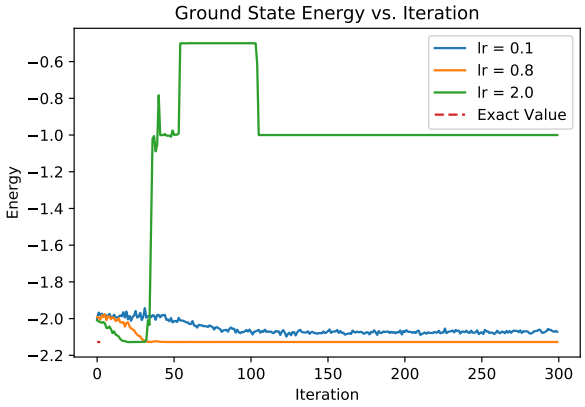

Fig. 5. With a learning rate of 0.1, the algorithm does not converge within our desired window when training with the 1-D Ising Hamiltonian. With a learning rate of 0.8, it converges. With a learning rate of 2, it diverges because the step size is too large for the fine optimization landscape of the parameter space.

2) α: The hyperparameter α, which determines the ratio of hidden units to visible units in the RBM, can be increased without bound for increased accuracy. Large values, however, will slow down the computation because the number of hidden units in a Restricted Boltzmann Machine will grow with increased α to the point where this method will provide no improvement within a reasonable computational time. We found that $\alpha = 4$ provides acceptable performance without slowing down the algorithm too much for spin chains of size 10-100 spins, though as the number of spins increases having a large α improves the robust convergence of the algorithm.

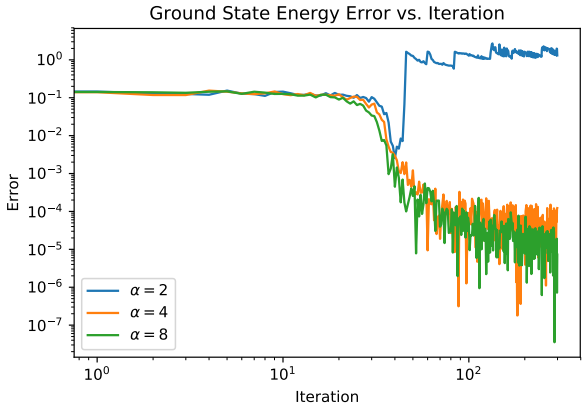

Fig. 6. With 8 spins and a learning rate of 0.5, we can use validation against the exact ground state energy to see that the algorithm fails with $\alpha = 2$, but succeeds with $\alpha = 4$ and $\alpha = 8$.

B. CORRELATIONS

One way we can verify that Restricted Boltzmann Machines accurately represent quantum states is by computing the expectation values of correlations between each spin and validating them against what we expect through physical intuition.

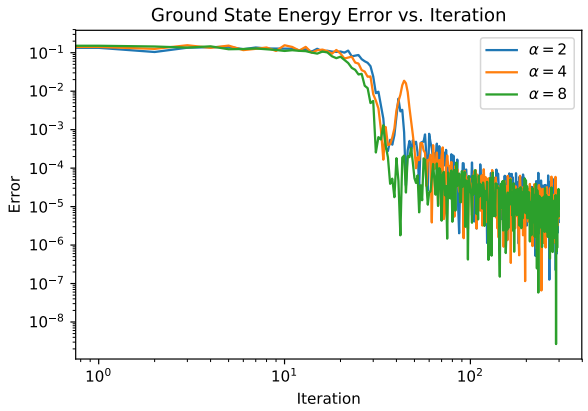

Fig. 7. While the $\alpha = 2$ case is fixable by increasing the learning rate to 0.8, this shows that having a larger α helps the algorithm converge when dealing with suboptimal hyperparameter settings.

1) ISING MODEL: For the 1-D Ising Hamiltonian, we expect that spins adjacent to each other are more correlated, while spins farther apart are less or not correlated because the strength of the interaction decays with distance. With a strong magnetic field, these interactions should be limited to close neighbors and drop off sharply.

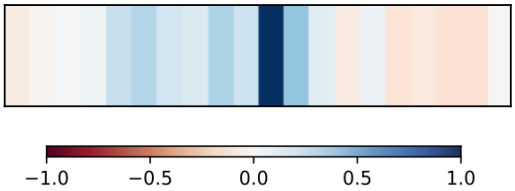

Fig. 8. When plotting the correlations of each spin in the 1-D Ising model with field strength $h = 2$ with the first spin (center, to account for periodic boundary conditions) after training, we see that the spins close to the first spin are correlated with it, while those at the edges are not.

With a weaker field, we should see the neighboring spins interact more with each other, and see correlations decay more slowly with distance.

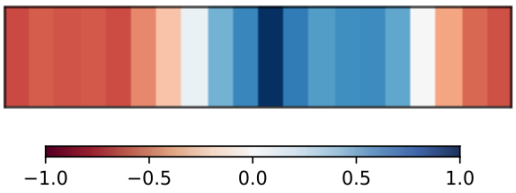

Fig. 9. In the 1-D Ising model with a weak field ($h = 0.2$) we see that correlations of each spin with the first spin (center, to account for periodic boundary conditions) are large, being strongly correlated near the first spin and strongly anticorrelated far away from it.

2) HEISENBERG MODEL: For the 1-D Heisenberg model with positive J, we expect that in the ground state spins adjacent to each other are anticorrelated, as is the case with antiferromagnetic materials that obey these kinds of interactions at high orders. Because the field strength is weak compared to the spin interaction, this anticorrelation is limited by distance and dies off quickly.

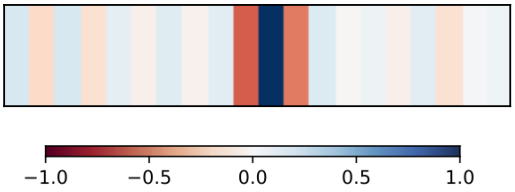

Fig. 10. In the 1-D Heisenberg model with $J = 1$, we see that adjacent spins are anticorrelated, but the strength dies off quickly with distance.

If we increase the strength of the spin neighbor interaction, the spins will lock into a completely anti-correlated phase that resembles complete antiferromagnetic spin order. If J is negative, then the spins in ground state should all be aligned in the same direction because now the Hamiltonian gains energy with anti-correlated adjacent spins and so the lower energy states should have spins positively correlated.

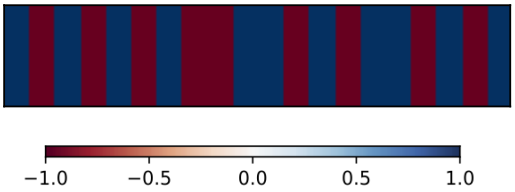

Fig. 11. In the 1-D Heisenberg model with $J = 5$, we see that adjacent spins are anticorrelated strongly everywhere.

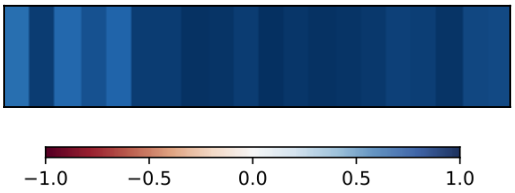

Fig. 12. For the 1-D Heisenberg Hamiltonian with $J = -0.5$, we see that all spins are positively correlated, which confirms our physical intuition.

VII. CONCLUSIONS

Our Restricted Boltzmann Machine quantum model has proven to be capable of competing with other modern representations in describing the energy and expected correlations induced on a many-body particle system for two common spin models. This model converges to the ground state energy with accuracy only dependent on the number of hidden units used. This neural network,

trained on spin configurations artificially generated using Monte Carlo sampling methods, is capable of learning the ordered spin phases of ground states for Hamiltonians with different kinds of interactions between particles. The success of neural networks for this problem can be traced to the implicit non-locality of the hidden units in storing information about the particle correlations. A surprising feature of quantum mechanics when first discovered was that the correlation between particles, called the quantum entanglement of the system by physicists, allows for two particles to be completely correlated in their measurement outcomes while being arbitrarily far from each other. This non-locality in quantum mechanics is a unique feature of the Hilbert Space geometry, and quantitative descriptions of the amount of quantum entanglement in a system have shown [16] that for the local Hamiltonians studied in this paper the entanglement between particles must be concentrated at the boundary between spins. This drastically simplifies the complexity of our representation, and in practice we can represent increasingly more complex models by increasing the hidden unit density. Using these ideas in future work, it would be interesting to find how the non-locality of the Hamiltonian model affects the convergence in training a neural network quantum state. Additionally, developing higher dimensional lattice models would be useful to see how the neural network model complexity must scale with the amount of nearest neighbors interacting with each spin.

Although the RBM architecture has proven capable for these locally interacting many-body systems, recent interest in strongly correlated spin systems [16, 17] has shown the need to simulate hamiltonians with extended correlations. This naturally suggests the development of a Deep Boltzmann Machine [18] model to accomplish the task of representing quantum states with greater entanglement, with one or more extra hidden layers acting to store more information about the greater number of nonlocal correlations determining the energetically minimal ground states. These results further suggest the practicality of machine learning techniques to conquer the intractable dimensionality of the quantum realm, using efficient representations of the quantum state that learn the unseen quantum interactions between particles without explicit construction of the Hamiltonian. By analyzing the effectiveness of the Restricted Boltzmann Machine with varied number of hidden units, we can gain knowledge about the storage of quantum information with spin chains of different dimensions, shedding light on the entanglement geometry of a wide range of solid state lattice structures.

REFERENCES

[1] Dirac, P. A. M. The Principles of Quantum Mechanics. Clarendon Press, 1958.

[2] Griffiths, David J. (2004). Introduction to Quantum Mechanics (2nd ed.). Prentice Hall. ISBN 0-13-805326-X.

[3] Ceperley, D. and Alder, B. Quantum Monte Carlo. Science 231, 555560 (1986)

[4] Orus, R. A practical introduction to tensor networks: Matrix product states and projected entangled pair states. Ann. Phys. 349, (2014).

[5] Schollwck, U. The density-matrix renormalization group in the age of matrix product states. Ann. Phys. 326, 96192 (2011).

[6] G. Carleo and M. Troyer. Solving the quantum many-body problem with artificial neural networks. arXiv:1606.02318, 2016.

[7] White, S. R., 1992, Phys. Rev. Lett. 69, 2863.

[8] S. Gharibian, Y. Huang, Z. Landau, and S. W. Shin, Quantum Hamiltonian complexity, arXiv:1401.3916.

[9] Hastings, M. B., and T. Koma, 2006, Commun. Math. Phys. 265, 781.

[10] McMillan, W. L., 1965, Phys. Rev. 138, A442.

[11] Foulkes W.M.C., Mitas L., Needs R.J., Rajagopal G. (2001). Rev. Mod. Phys. 73, 33-83.

[12] Carleo, G. (2017) *Neural Network Quantum States: A Lecture for the Machine Learning and Many-Body Physics Workshop*

[13] Casula, M., C. Attaccalite, and S. Sorella, 2004, The Journal of Chemical Physics 121(15), 7110.

[14] S. Sorella, M. Casula, and D. Rocca, J. Chem. Phys. 127, 014105 2007.

[15] Amari, S. (1998). Natural gradient works efficiently in learning. Neural Computation, 10(2) 251276

[16] A. Garg, M. Randeria, and N. Trivedi, Nature Physics 4, 762 (2008).

[17] Georges, A., 2002, in Strongly Correlated Fermions and Bosons in Low-Dimensional Disordered Systems, edited by I. V. Lerner, B. L. Althsuler, V. I. Falko, and T. Giamarchi (Kluwer Academic Publishers, Dordrecht, Netherlands), NATO Science Series II.

[18] Salakhutdinov, R. Hinton, G. Deep Boltzmann machines. In Proc. International Conference on Artificial Intelligence and Statistics 448455 (2009).