# Study into the Optimal Variational Ansatz for Frustrated J1-J2 Heisenberg Models

Yuhuang Ou, *King's College London*

*Supervisor: Joe Bhaseen, Examiner: George Booth*

***Abstract*—Despite the exceptional developments in the utility of machine learning in studying many-body systems, their empirical successes are not well understood. We present a study into popular machine learning approaches and their respective variational ansatz, namely Jastrow wavefunctions and Restricted Boltzmann Machines. Then, by simulating Frustrated J1-J2 Heisenberg models we draw correlations between their capabilities with quantitative performance in accuracy and computational efficiency. Based on our results, we make strategic recommendations for future lines of research in using machine learning to study quantum many-body systems.**

***Index Terms*—Jastrow wavefunctions, machine learning, Monte Carlo Markov, Restricted Boltzmann Machines, quantum many-body systems**

## I. Introduction

In the last five to ten years, we have seen an explosion of high impact development in machine learning and its application towards resolving the outstanding challenges that lie at the heart of quantum physics. Ever-improving computational resources, datasets, and libraries have enabled more efficient and effective numerical tools in solving computational tasks previously considered intractable.

In the realm of condensed matter physics, a multitude of machine learning applications have been developed. Recent advances have included the exploration of tensor-network representations of many-body quantum states [1], and the development of Quantum Monte Carlo approaches [2]. Consequently, these developments have enabled a speeding up of many-body simulations [3], the classifying of many-body quantum phases [4], and the performance of quantum state tomography [5]. Despite these advances, the origin of the empirical successes obtained so far among these methods are not well understood. As such, this report seeks to study these methods and draw empirical correlations in their capabilities.

The two classical approaches in resolving the non-trivial challenges of simulating many-body quantum states can be ordered in two categories. The first is through the representation of quantum states in stochastic Variational Monte Carlo (VMC) calculations, Jastrow wavefunctions being the prime example. This approach carries high entanglement but also limited variational freedom. The second, more recent approach, utilises tensor-networks for non-stochastic variational optimisation, which operate more specifically on entanglement-limited variational wavefunctions. In response to the limitations of these approaches, recent developments have been made towards the utility of Artificial Neural Networks (ANNs) in tackling these issues.

Arguably the most well-known development in this area is that devised by Carleo and Troyer, who introduce a new variational quantum states representation based on the Restricted Boltzmann Machine (RBM) [6]. A RBM stands out as having a bipartite arrangement of two layers of neurons, a visible and hidden layer. Its lack of intra-layer connections is what makes it a "Restricted" Boltzmann Machine, and which enables more efficient training algorithms such as Stochastic Gradient Descent (SGD). This Neural-network Quantum States (NQS) approach couples the parametrisation of wavefunction amplitudes as Feed-Forward Neural Networks (FFNNs) and the utility of a reinforcement-learning scheme to achieve high accuracy in the determination of equilibrium and dynamical properties of interacting spin models in both one and two dimensions.

The best architecture for describing a many-body quantum system may differ from one case to another. As such, in this report we consider the case of the spin ½ antiferromagnetic $J_1$-$J_2$ chain, for which exact solutions may be calculated using the Lanczos algorithm as a method of brute-force exact diagonalisation. This particular model has been chosen as an extensive number of computational methods have already been applied to study this problem, including the application of NQS and tensor-network matrix product states [7]. Furthermore, the $J_1$-$J_2$ Heisenberg model presents outstanding theoretical challenges pertaining to its phase transitions, ones which our study seeks to explore as we evaluate the results of our approaches and their respective ansatz. This makes the model a useful and non-trivial benchmark for ANN techniques.

With the development of NQS came further innovations in the tools for use in their application. *Carleo et al* introduced Netket, a comprehensive open-source framework for the study of quantum many-body systems [8]. This facilitated a good starting point for exploring the implementation of NQS, SGDs, VMC, and supervised and unsupervised learning environments. As such, in our work, we make full use of its Python library to explore the implementation of the RBM and Jastrow wavefunction. Finally, we will discuss the results gathered in solving the many-body system and draw numerical correlations for their limitations and strengths to provide strategies for future developments.

## II. Hamiltonian

We began by implementing the spin ½ antiferromagnetic $J_1 - J_2$ Heisenberg model which is defined by the Hamiltonian:

$$H = \sum_{i=1}^{L} J_1 \vec{\sigma}_i \cdot \vec{\sigma}_{i+1} + J_2 \vec{\sigma}_i \cdot \vec{\sigma}_{i+2} \tag{1}$$

We considered a lattice system in which the wavefunction could be expanded to $\Psi(\vec{\sigma}) = \sum_{\vec{\sigma}} C[\vec{\sigma}]\vec{\sigma}$, where $\vec{\sigma}$ was considered a vector corresponding to the spin configuration of the system. Such that $\vec{\sigma} = \vec{\sigma}_1 \dots \vec{\sigma}_n$ are $S_i^z$ eigenstates spanning the Hilbert space. Consequently, the challenge was to approximate the function $C[\sigma]$ with fewer parameters than the given Hilbert space dimensionality. Given the scope of our study, we constrained our formulation of the Hilbert space to the zero-magnetisation subsector, such that $S_z = 0$. Furthermore, we focused primarily on the instances where the nearest and next-nearest neighbour interactions were antiferromagnetic, where $J_1, J_2 \geq 0$, such that the magnetic interactions were frustrated.

In focusing on the range of antiferromagnetic frustration, we noted three regions in the classical limit of this model's long-range order. Firstly, $J_2/J_1 < 0.5$, where the Néel two sub-lattice antiferromagnet is that of the system's ground-state. Secondly, $J_2/J_1 > 0.5$, where a col-linear antiferromagnet with four sub-lattices can be found. Finally, $J_2/J_1 \approx 0.5$, where the Bethe-Hulthen solution has shown that quantum fluctuations at this level of frustration can destroy the antiferromagnetic long-range order for 1 dimensional chains and result in a high degree of degeneracy across different states [9]. Competing theoretical approaches towards evaluating these regions have arrived at different results. Exact diagonalisation studies, such as that through the Lanczos algorithm, have predicted the destruction of the classical Néel order [10]. While spin-wave theories within a Hartree-Fock scheme, such as that through the Jastrow wavefunction, do not find this intermediate phase [11]. Our report will explore these phenomena and relate them to the performance of our chosen ansatz. For simplicity we set $J_1 = 1$ throughout this report, with $J_2 = 0.4$ being the default value unless otherwise stated.

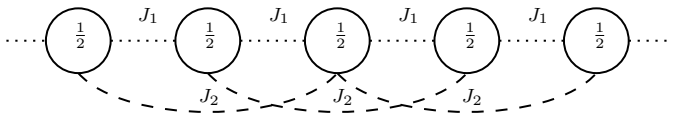


Fig. 1: The frustrated spin ½ antiferromagnetic $J_1$-$J_2$ Heisenberg chain is one of the simplest frustrated spin models.

The model is implemented as a graph with two types of edges representing $J_1$ and $J_2$ respectively, as depicted in figure 1. NetKet readily provides access to pre-defined Hamiltonians for the Heisenberg and transverse field Ising models. However, for our study we constructed a custom Hamiltonian as defined in equation 1, by using the Local Operator class. This enabled the construction of any common lattice model from a set of $k$-local operators which acted on each site. Therefore, given the Hilbert space and list of operators acting on the $n$ sites, we encoded the spatial structure of the $J_1$-$J_2$ Hamiltonian. We proceeded to implement two popular variational ansatz: the Jastrow wavefunction and the NQS.

## III. Jastrow Wavefunction

The Jastrow-Slater wavefunction was mentioned earlier as being a prominent example of VMC calculations [12]. This approach is summarised as the product of the Slater determinants and the totally symmetric non-negative Jastrow correlation factor. The former is usually obtained from accurate Linear Discriminant Analysis or Hartree-Fock calculations, while the latter is an explicit function of electron-electron separations [13]. Principally, the wavefunction is defined as:

$$\Psi(s_1, \dots, s_n)_{JTR} = e^{\sum_{ij} s_i W_{ij} s_j} \tag{2}$$

where $W_{ij}$ is the network weight. In order to implement the trial wavefunction, we must also consider the two other primary components: the optimiser and the sampler. The optimiser seeks to select the best results of the function, the most popular example of which is the above-noted Stochastic Gradient Descent. A Gradient Descent optimising method, in general, is an iterative method that follows the downhill direction of the surface gradient created by the objective function until a minimum is reached. An SGD, in particular, introduces a randomised element by performing a parameter update for each training example; thereby increasing the chances of achieving the global minima. Meanwhile, the sampling method defines the approach in which elements out of an entire set are chosen. One of the most popular approaches is through the Metropolis-Hastings algorithm, an MCMC method that obtains a sequence of random samples from a probability distribution [14]. In the context of our study, these two components form part of our iterative scheme that seeks to minimise the expectation value of the energy, $E(W_k) = \langle \Psi_m | \mathcal{H} | \Psi_m \rangle \,/\, \langle \Psi_m \mid \Psi_m \rangle$. A Monte Carlo sampling is then performed for a set of variational parameters in $W_k$ for each iteration $k$. Meanwhile stochastic estimates of the energy gradient were then calculated in order to establish the next set of network weights, $W_{k+1}$. Futhermore, our method uses parallel tempering to improve the properties of our many-body Monte Carlo simulation. In essence, multiple copies of the system are randomly initialised to enable maximal approximation of the global optimum. For simplicity, we kept constant all parameters for our optimisation and sampling algorithms, opting for a sample size of 300 and setting the learning rate, $\eta$, to 0.02. Meanwhile the number of Markov chains executed in parallel to operate the Metropolis exchange was kept at 16. Further details pertaining to our implementation of these optimisation algorithms are addressed in appendix B.

NetKet's Jastrow wavefunction is known to be a long-range description. As such, it is stipulated to be widely sufficient for the calculation of accurate ground-state energies for weak and strong frustrations [15]. Consequently, through the quantitative analysis of our implementation we can measure its effectiveness with a correspondingly competitive model, the RBM, and its respective NQS ansatz.

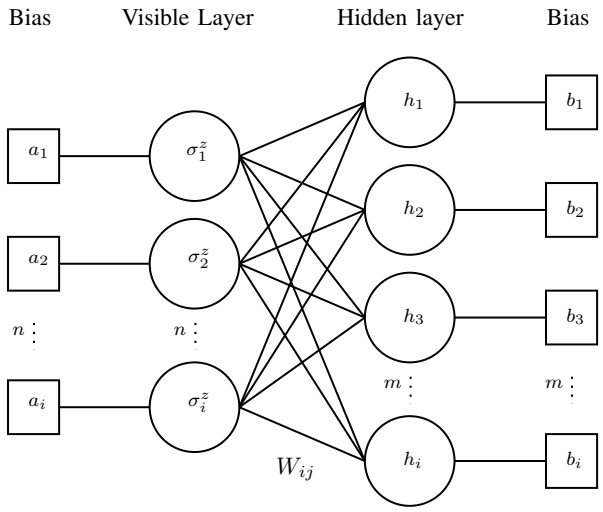


Fig. 2: Diagram showing the bipartite relationship between the visible and hidden layer and their respective biases. The spin values of the system are mapped to a visible layer of $n$ neurons in the RBM. While the number of hidden nodes, $m$, is determined by the $\alpha$ value. Both the visible and hidden layers take binary values of $\sigma_i$ and $h_i$, each connected to their respective biases. The left hand side represents the input while the right represents the output.

## IV. Restricted Boltzmann Machine

Moving on to the RBM architecture, we can implement the aforementioned widely known NQS. The RBM's ability to update with high acceptance rates and low autocorrelations in identifying features such as collective modes and correlations, is the source of its power in Monte Carlo simulations [17]. As such, the RBM can build an adaptive model for the physical probability distribution. When coupled with its effectiveness in estimating physical observables, the RBM can accelerate Monte Carlo simulations [18]. This underlying property has enabled the proliferation of the RBM in Monte Carlo approaches to the many-body problem. In our work we use the RBM spin machine provided by the NetKet library - which in practice is equivalent to a 2-layer FFNN with a nonlinear activation function in between. Our implementation of this RBM is encapsulated in figure 2, while the relationship between the visible and hidden nodes is fully described by the following ansatz:

$$\Psi_{\mathrm{RBM}}(\sigma_1^z, \ldots, \sigma_n^z) = e^{\sum_{i=1}^{n} a_i \sigma_i^z} \prod_{i=1}^{m} \cosh(b_i + \sum_j W_{ij} \sigma_j^z \tag{3}$$

where $a_i$ and $b_i$ are the visible and hidden biases respectively. Together with the weights $W_{ij}$, they are variational parameters that will minimise the energy calculation. The number of hidden nodes, $m$, is determined by the hidden unit density, $\alpha$, such that its total is defined by $m = \alpha n$. The network weights were produced as complex-valued which provided a complete description of both the wavefunction's phase and its amplitude.

In general, both the weights and biases were determined through a stochastic optimisation process which involved local Metropolis moves, the same aforementioned method for the Jastrow wavefunction. However, in contrast, the minimisation of energy was then achieved through the Stochastic Reconfiguration (SR) algorithm and the AdaMax optimiser from the same library. In the context of our study, this formulation of the RBM architecture as a wavefunction ansatz represented one of a number which are collectively known as NQS. Such approaches are currently reported to be the cutting edge in machine learning applications towards solving many-body quantum states.

## V. Variational Results

In the Jastrow wavefunction and the RBM cases, we can obtain analytical comparisons in accuracy and efficiency by comparing the results of the target function – the ground-state wavefunctions – and varying their parameters for operation. In figure 3 we draw one such comparison between the NQS and Jastrow ansatz. In this instance, the latter method outperforms the former by converging towards the exact ground-state energy within fewer iterations. It is stipulated that the RBM architecture can be significantly improved by formulating the NQS representation through lattice translation symmetry as the use of symmetry reduces the number of variational parameters in the ansatz. For the left panel shown in the figure, the number of parameters was indeed reduced from 528 to 24. At initial review the symmetric RBM approach succeeds in resolving the challenge of minimising the variational parameters in approximating the function $C[\psi]$. However, on closer inspection this change only accounted for a 6% reduction in computational time. This is in contrast to the Jastrow approach, which resulted in 231 parameters and required 29% more time relative to the pure RBM approach. The inconsistency between the number of parameters and computational resource requirements are not well known, and further studies can be performed in this area. Meanwhile, the right panel shows the limiting behaviour of the three methods as the number of sites, $n$, increases towards infinity, each adhering to a time complexity of $O(n^2)$. Both the RBM and symmetric RBM behave similarly and perform the best in large systems, with the latter being marginally better. As such, for the remainder of this report we will refer to the symmetric RBM in place of the pure RBM to simplify the study.

In figure 4, we show the accuracy of the RBM architecture quantified by the relative error on the ground-state energy $\epsilon_{rel} = (E_{RBM}(J_2) - E_{exact})/E_{exact}$, where $E_{exact}$ is computed from the Lanczos algorithm. An exponential correlation between the $J_2$ parameter and computational requirements is shown. After 200 iterations, in the instance of a chain of $n = 22$, the symmetric RBM did not show advantages in accuracy at higher values of $J_2$. Consequently, this has shown that exponentially more iterations are required to attain accurate results with larger values of $J_2$ and therefore frustration. It has also been shown that the Jastrow wavefunction is capable of tackling strong frustration, $J_2 \approx 4$, competitively with the RBM. However, the theoretical correspondence between $J_2 > 0.5$ and decreasing accuracy is not well understood; further research can be conducted in this area.

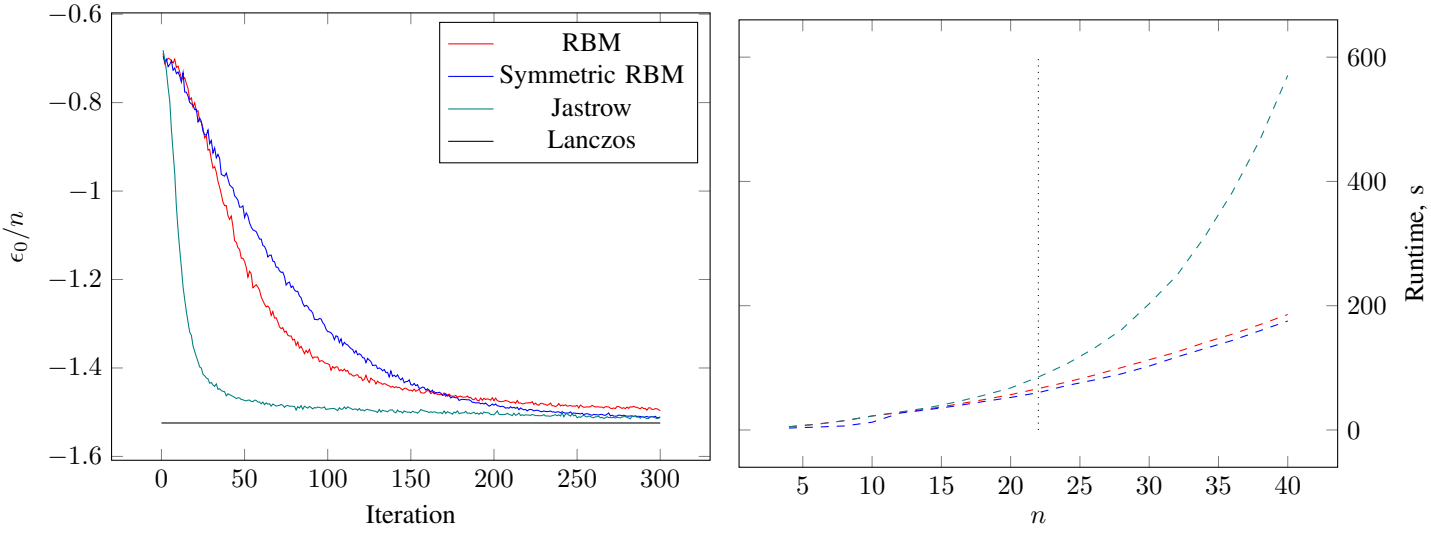


Fig. 3: The left panel shows the performance of three methods in finding the ground-state energy per site of a 1D AFH $J_1$-$J_2$ chain for a $J_2$ parameter of 0.4 with 22 sites. Results were attained from an average across 20 batches, and sample size of 300. Most notably, the rate of convergence for Jastrow was highest, while the symmetric RBM attained a more accurate result overall. The right panel compares the time taken for computation for each method against the number of sites. The dotted line marks the instance represented in the left panel.

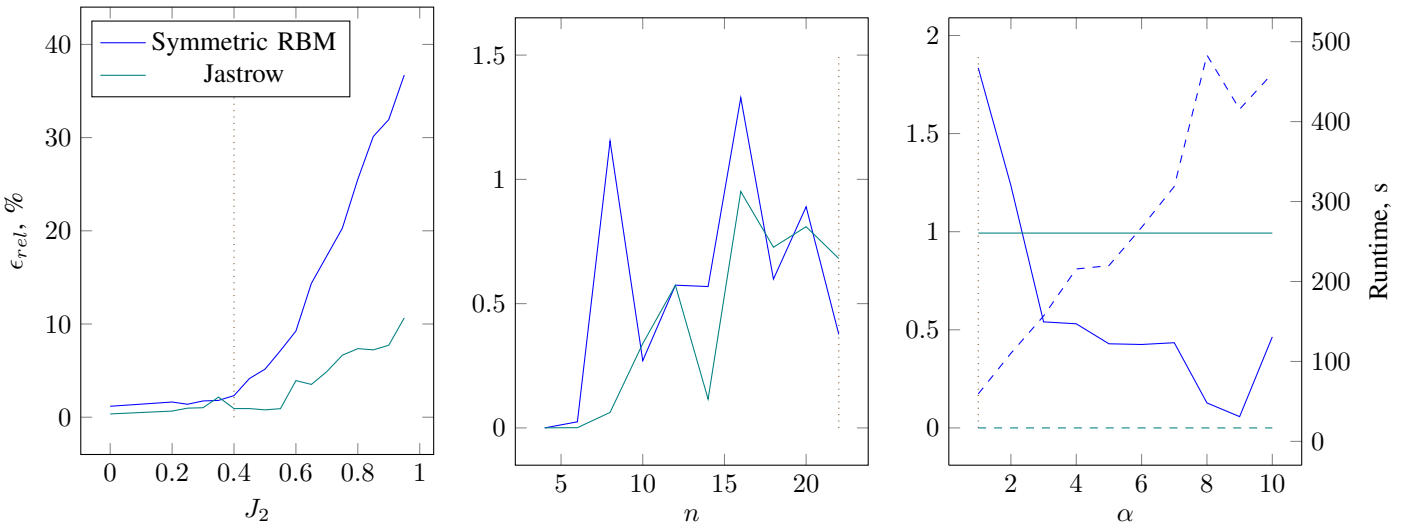


Fig. 4: Comparing the accuracy of the Symmetric RBM and Jastrow methods. In the left panel, the graph shows the relative error between the computed ground-state energies and the exact values for a given $J_2$. In the middle panel, the graph shows the relative error against the number of spin states. The right panel shows the inverse correlation between accuracy and computational time with increasing hidden node density, $\alpha$. In all three cases the values were obtained with a fixed number of 200 iterations, a constant $J_1$ value of 1, and a sample size of 300. The dotted lines represent the default parameters which produced the results in seen in the previous figure. Note that the results of the pure RBM approach align very closely with the symmetric RBM, and therefore has been omitted for greater clarity.

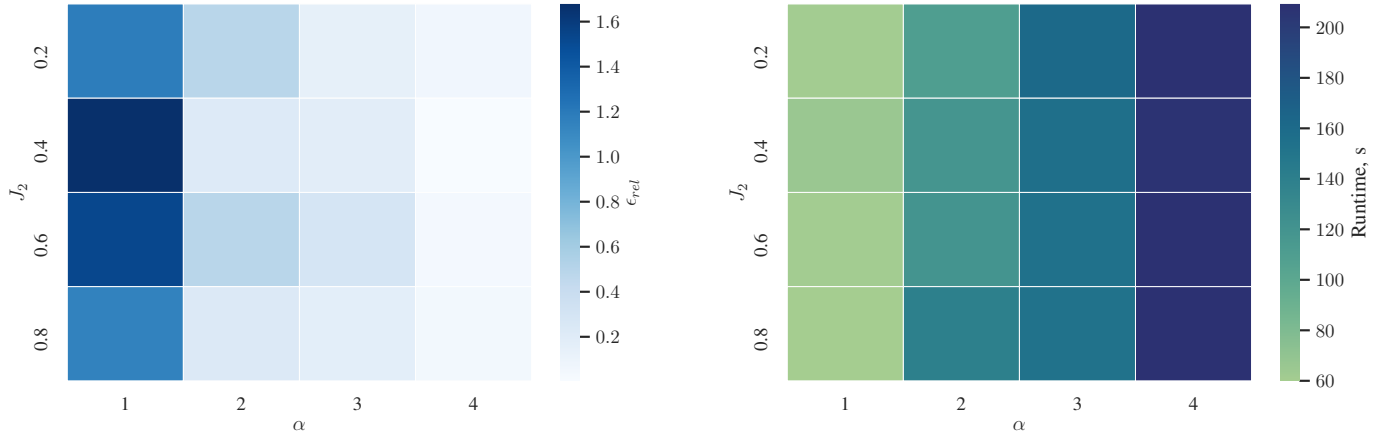


Fig. 5: Visual comparison of accuracy and efficiency across a range of $J_2$ and $\alpha$ parameters. The left panel most notably shows acute deviation from the Lanczos result in convergence between 0.4 and 0.6. This is consistent with the proposal that $J_2 \approx 0.5$, representing the intermediate regime, is highly frustrated [16]. The right panel shows consistency in the correlation between $\alpha$ and the runtime, and the contrary with $J_2$. Both results were obtained for an AFH chain of length 22, but all correlations should remain relatively consistent for all values of $n$ as shown before.

Previously, in formulating the $J_1$-$J_2$ Hamiltonian, we noted that competing theoretical approaches towards evaluating the regions of frustration have computed different results. Considering this, the relative error calculation is not just a parametrisation of the accuracy of the calculations, but also a parametrisation of the alignment of the two differing approaches. The small and constant $\epsilon_{rel}$ between $J_2$ values of $0$ and $0.4$ reflects consistency for the two distinct theoretical methods. Meanwhile, the increasing $\epsilon_{rel}$ at $J_2 \approx 0.5$ reflects the conflicting results of the two methods. This is consistent with the analysis by *Ceccatto et al* which indicates good alignment in results for both methods at small values of frustration, and the contrary for strong frustrations [19]. It is interesting to note that the symmetric RBM approach deviates much earlier at $J_2 \approx 0.4$, further studies could also be made to determine whether this was a result of limited computational resources or a result of fundamental differences in theoretical methodology.

In the case of increasing numbers of sites, $n$, both the Jastrow wavefunction and the symmetric RBM are shown to be equally as effective in accurately determining ground-state energies, with fluctuations between $0\%$ to $1.5\%$ in relative error. Our study was not able to determine the effects in which $n > 22$ due to computational limitations which prevented the accurate determination of an exact ground-state energy, and therefore prevented further calculations of relative errors.

A purported characteristic of the NQS is that the greater the value of $\alpha$, the more accurate the results become. This is partially evinced in the right panel, which depicts significantly increased accuracy between $\alpha$ values 1 and 3; however further increases do not significantly improve performance. One possible reason is that for an ANN with greater numbers of hidden nodes, the parameter space also becomes more complex; therefore the computational expressibility in finding the optimal representation is increased. Furthermore, we also noted in our study that larger hidden unit densities incurred greater costs in computational time. We show that a linear time complexity correlation exists between the RBM methods with $\alpha$.

The correlations depicted in figure 5 can be used to visually comprehend the trends between the relationship of the input parameters of the RBM and the resultant accuracy and runtime of its computation. For our default values, $J_2 = 0.4$ and $n = 22$, we achieved an approximate $80\%$ improvement in accuracy when $\alpha$ was increased from 1 to 3, however at a cost of an approximate $130\%$ increase in computational time. The computational costs remains consistent throughout values of $J_2$, providing evidence of little correlation between the level of frustration and computational requirements for the RBM. Meanwhile, the trend between $J_2$ and $\alpha$ with $\epsilon_{rel}$ is less consistent. A particularly anomalous result was the greater than expected error for $J_2 = 0.2$. Further work could be completed in uncovering the corresponding computational or theoretical reason behind this.

## VI. Conclusion

In this report, we demonstrated the application of two simple approaches to approximating the ground-state energy of a $J_1 - J_2$ AFH spin ½ chain. We compared the use of the Jastrow wavefunction to the RBM in the context of ANNs and found optimal conditions for operation in both. Consequently, we can put forward recommendations for their use and further avenues for study.

Firstly, for large systems containing more than 20 sites the RBM should be used in place of the Jastrow wavefunction due to its shallower time complexity. Secondly, larger hidden unit densities correlate with drastically higher accuracy at the cost of slower computation. Given a target value of relative error, an

optimisation function can be formulated to the optimal number of hidden units. Finally, our analysis showed that the Jastrow wavefunction is capable of producing more accurate results at stronger levels of frustration in line with expectations.

We showed that the Hamiltonian symmetries on the ANN reduce the number of variational parameters but does not result in a significant increase in efficiency. Furthermore, we did not ascertain a strong correlation between the number of spin states and increasing demand for computational resource. This reflects the stipulation that these methods have been well adapted for simulating higher numbers of spin states and exploiting the symmetries which are present. Further studies could be made with greater number of spin states.

In our simulations we found that when $\alpha = 3$, where there are three hidden nodes for every visible node, a drastic improvement in accuracy is measured when compared to a one-to-one ratio between the two node types. However, we also showed that further increasing the number of nodes does not significantly improve performance. As such we would not recommend greater numbers of hidden nodes.

The direct correlation between the effectiveness of these methods towards applications in identifying and classifying phase transitions are yet to be studied. As such, further work could be carried out to uncover the numerical correlations in adopting ANNs in this end. Furthermore, the incorporation of deep learning in this context is still an open question and would require further studies.

Due to the nature of ANNs being heuristic algorithms whose effectiveness largely coincides with the designer's capabilities, further work in developing our systematic understanding of the validity and limitations of these approaches in condensed matter physics is still required. In doing so we hope to tighten the intersection between the computational and physical sciences so as to develop further innovations in the numerical tools that will help investigate theoretical models and eventually resolve the outstanding challenges at hand.

*Outlook* — Machine learning provides the framework which enables computers to solve a given challenge without being provided the explicit instructions to do so. Recent technological developments in computational power have spurred greater research into the application of machine learning in solving scientific challenges that were previously considered impossible. One such challenge sits at the core of quantum mechanics, a branch of physics which helps decipher how the universe operates at a sub-atomic scale. In such systems, there is an exponentially increasing data requirement for accurately simulating interactions between objects. Due to the limitations of our classical numerical tools, an infinite amount of time and energy would also be required as we increase the scale of these systems. However, recent developments in applying machine learning to such quantum mechanical challenges have resulted in substantive breakthroughs. Our report unravels these developments and draws comparisons between recent approaches to highlight their strengths and weaknesses, and to propose further areas of research. In particular, we focused on comparing the Jastrow wavefunction with Neural Quantum States, representing the classical and cutting edge approaches respectively. Our study found no perfect solution, but determined the effective range of applications for the two machine learning methods. As such it is clear that further developments in this area are required - ones which could inevitably help solve the challenges which lie at the heart of condensed matter, chemical, and quantum physics.

## REFERENCES


[1] F. Verstraete, V. Murg, and J. Cirac, Advances in Physics **57**, 143–224 (2008).
[2] O. Sharir, Y. Levine, N. Wies, G. Carleo, and A. Shashua, Physical Review Letters **124** (2020).
[3] D. Wu, L. Wang, and P. Zhang, Physical Review Letters **122** (2019).
[4] J. Carrasquilla and R. G. Melko, Nature Physics **13**, 431–434 (2017).
[5] J. Carrasquilla, G. Torlai, R. G. Melko, and L. Aolita, Nature Machine Intelligence **1**, 155–161 (2019).
[6] G. Carleo and M. Troyer, Science **355**, 602–606 (2017).
[7] T. Westerhout, N. Astrakhantsev, K. S. Tikhonov, M. I. Katsnelson, and A. A. Bagrov, Nature Communications **11**, 1593 (2020).
[8] G. Carleo et al., SoftwareX **10**, 100311 (2019).
[9] W. J. Caspers, M. Labuz, and A. Wal, Journal of Physics Conference Series (Online) **30**, 73 (Feb 2006).
[10] J. Richter, Phys. Rev. B **47**, 5794 (1993).
[11] A. F. Barabanov and A. V. Mikheyenkov, Zeitschrift für Physik B Condensed Matter **93**, 349 (1994).
[12] R. Jastrow, Phys. Rev. **98**, 1479 (1955).
[13] W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal, Rev. Mod. Phys. **73**, 33 (2001).
[14] W. K. Hastings, Biometrika **57**, 97 (1970).
[15] K. Retzlaff, J. Richter, and N. B. Ivanov, Zeitschrift für Physik B Condensed Matter **93**, 21 (1993).
[16] K. Choo, T. Neupert, and G. Carleo, Physical Review B **100** (2019).
[17] K. Abhinav, B. Chandrasekhar, V. M. Vyas, and P. K. Panigrahi, Physics Letters A **381**, 457 (2017).
[18] L. Huang and L. Wang, Physical Review B **95** (2017).
[19] H. A. Ceccatto, C. J. Gazza, and A. E. Trumper, Phys. Rev. B **45**, 7832 (1992).
[20] M. D. Zeiler, Adadelta: An adaptive learning rate method, 2012.
[21] S. Sorella, M. Casula, and D. Rocca, The Journal of Chemical Physics **127**, 014105 (2007).

## APPENDIX A
## NETKET DEPENDENCIES

Our chosen approach for executing our study is through the use of the NetKet Python library. This library provides all the necessary tools in implementing and studying the Lanczos algorithm and the VMC methods. In our study, we implemented the $J_1 - J_2$ hamiltonian via the Local Operator. This function enables the summing of an arbitrary number of operators which act locally on a limited set of quantum numbers, which in this instance is the Hilbert space obtained from the tensor product of local spin states determined by the Hilbert spin function. The spin function extracts the number of sites from a specified graph. In this instance, we define the 1D chain using a hypercube lattice of side length $n$ and $n_{dim} = 1$ dimension.

In implementing the VMC method we made use of the two aforementioned components: the sampler and the optimiser. Our sampler consisted of a Metropolis Exchange which acted on either the RBM, symmetric RBM, or the Jastrow ansatz. These three approaches and their respective parameters represented the independent variables in our study. Optimisers were then applied to our VMC, in particular Stochastic Reconfiguration (SR) and SGD algorithms were applied irrespective of the ansatz used. In contrast, the Adadelta optimiser was applied in addition to the other optimisers for the RBM and symmetric RBM. A visual representation of this description can be seen in figure 6.

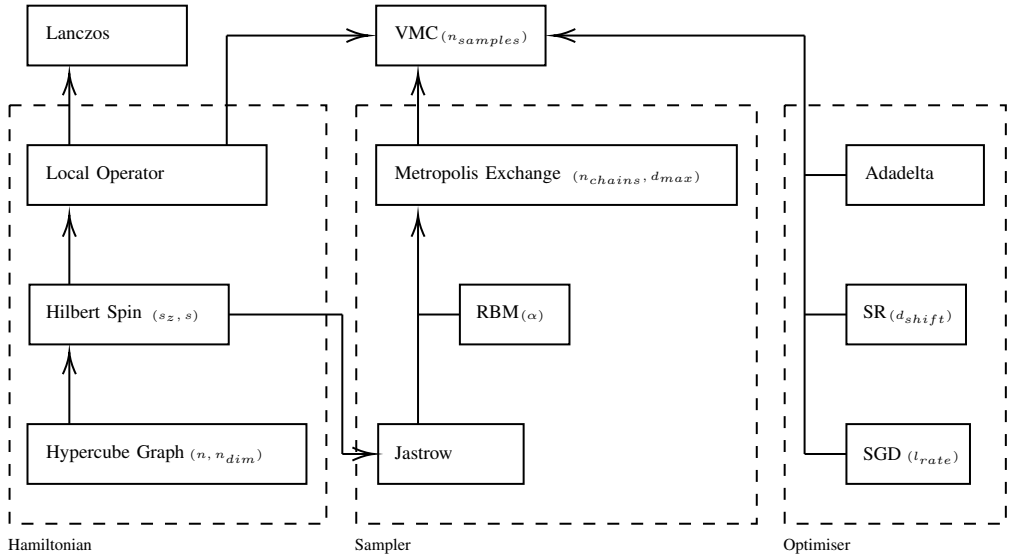


Fig. 6: An illustration of the submodule and parameter dependencies for our implementation of the Lanczos calculations and Variational Monte Carlo approach in computing the ground-state energies. Note that the RBM in this diagram represents both the pure RBM and the symmetric RBM.

## APPENDIX B
## OPTIMISATION ALGORITHMS

The NetKet implementation of the SGD was achieved through a stochastic estimate of a cost function, $G(\mathbf{p})$, performing a parameter update, $p'_k = p_k - \eta G_k(\mathbf{p})$, where $\eta$ is the learning rate. In our study, we employed a fixed learning rate in the order of $10^{-3}$ and a constant small sample size of 300 samples for estimating the gradient.

Meanwhile, the AdaGrad optimiser is an adaptive learning rate algorithm which scales $\eta$ with a sum across past gradients. Similarly to the SGD, given a stochastic estimate of the gradient of a cost function, $G(\mathbf{p})$, updates are made for parameters $p_k$ and additionally $g_k$:

$$p'_k = p_k - \frac{\eta}{\sqrt{g_k + \epsilon}} G_k(\mathbf{p}) \tag{4}$$

$$g'_k = g_k + G_k(\mathbf{p})^2 \tag{5}$$

A particular strength of the AdaGrad algorithm is its ability to perform well when gradients are space. However, as the squares of past gradients are cumulative, $\eta$ risks becoming too small with increasing numbers of iterations.

However, the monotonic decay of $\eta$ is addressed in the AdaDelta algorithm, which entirely removes the need for the global learning rate [20]. This particular formulation equates $E[g^2]$ and $s$, such that:

$$E[g^2]'_k = \rho E[g^2] + (1 - \rho) G_k(\mathbf{p})^2 \tag{6}$$

$$\Delta p_k = -\frac{\sqrt{E[\Delta x^2] + \epsilon}}{\sqrt{E[g^2] + \epsilon}} G_k(\mathbf{p}) \tag{7}$$

$$E[\Delta x^2]'_k = \rho E[\Delta x^2] + (1-\rho)\Delta p_k^2 \tag{8}$$

$$p'_k = p_k + \Delta p_k \tag{9}$$

where $\rho$ represents the exponential decay rate, $\epsilon$ represents the small cutoff, default values of which in the NetKet implementation are $0.95$ and $10^{-7}$ respectively.

Earlier in our report we stated our use of the Stochastic Reconfiguration algorithm in minimising the energy to find the optimal solution. More specifically, we used the method proposed by *Sorella et al*, which can be visualised as an imaginary-time evolution in the variational subspace [21]. Within the SR minimisation, the variational parameters are changed at each iteration. The algorithm is defined by the SR matrix, where $S$ is a positive definite covariance matrix defined as:

$$s_{k,k'} = \langle O_k O_{k'} \rangle - \langle O_k \rangle \langle O_{k'} \rangle \tag{10}$$

where $O_k$ are operators diagonal in the Hilbert space spanned by the their respective configurations such that,

$$O_k(\mathcal{S}) = \frac{1}{\Psi_m(S)} \partial_{w_k} \Psi_m(S). \tag{11}$$

The SR matrix relates to the network weights with respect to the x-th update iteration, $W$, through the Moore-Penrose pseudo-inverse function, $S^{-1}$, $W(x+1) = W(x) - \gamma S^{-1}(x) F(x)$. Initially $W$ is set to an arbitrarily small value and then procedurally optimised using the method above.

## Appendix C
## Phase Diagram

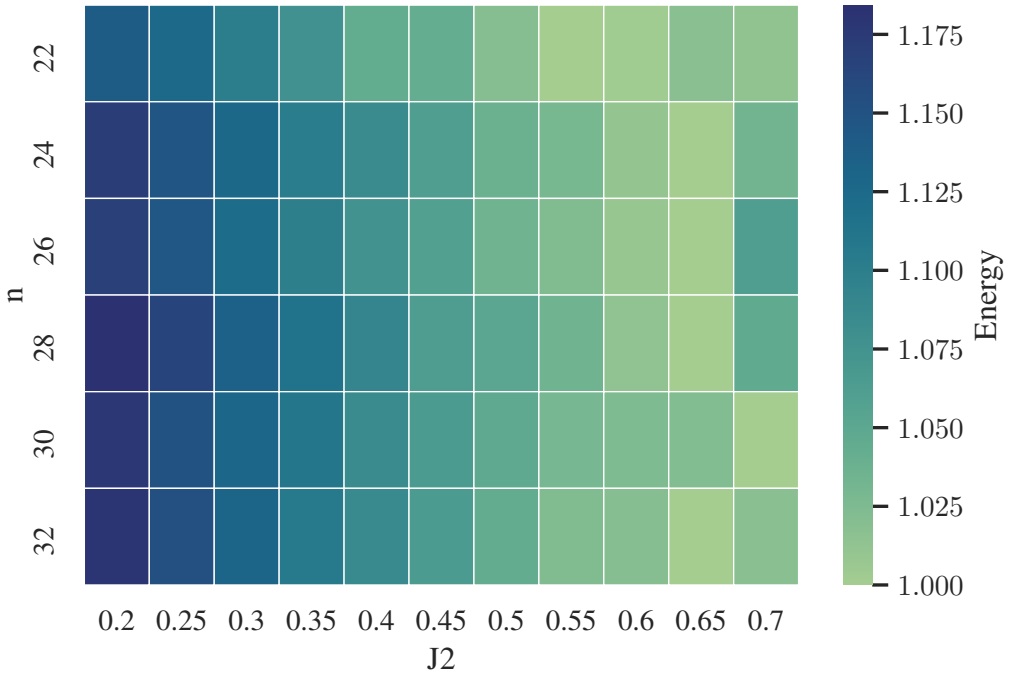


Fig. 7: Heatmap showing phase changes for values of J2 across $J_1$-$J_2$ spin ½ antiferromagnetic Heisenberg chains from 22 particles to 32. The Energy values represent max normalisations of ground-state energies across values of parameter $J_2$, between 0.2 and 0.7.

## Appendix D
## Code Availability

All simulations were conducted using the NetKet python package. Some of the corresponding code to carry out the study may be publicly accessed at: https://richardou.com/ml-qm