Data-Enhanced Variational Monte Carlo for Rydberg Atom Arrays

Stefanie Czischek,1,* M. Schuyler Moss,1 Matthew Radzihovsky,2,† Ejaaz Merali,1,3 and Roger G. Melko1,3
1Department of Physics and Astronomy, University of Waterloo, Ontario, N2L 3G1, Canada
2Department of Physics, Stanford University, Stanford, CA 94305, USA
3Perimeter Institute for Theoretical Physics, Waterloo, Ontario, N2L 2Y5, Canada
(Dated: March 11, 2022)

Rydberg atom arrays are programmable quantum simulators capable of preparing interacting qubit systems in a variety of quantum states. Due to long experimental preparation times, obtaining projective measurement data can be relatively slow for large arrays, which poses a challenge for state reconstruction methods such as tomography. Today, novel groundstate wavefunction ansätze like recurrent neural networks (RNNs) can be efficiently trained not only from projective measurement data, but also through Hamiltonian-guided variational Monte Carlo (VMC). In this paper, we demonstrate how pretraining modern RNNs on even small amounts of data significantly reduces the convergence time for a subsequent variational optimization of the wavefunction. This suggests that essentially any amount of measurements obtained from a state prepared in an experimental quantum simulator could provide significant value for neural-network-based VMC strategies.

Introduction. Rydberg atom arrays are powerful candidates for high-quality quantum simulation and computing platforms [1–3]. State-of-the-art experiments use optical tweezers to arrange and individually address atoms on arbitrary lattices [4–6], allowing them to strongly interact with a many-body Hamiltonian [7]. The combination of complex lattice structures together with the precise tuning of inter-atomic interactions has enabled the preparation of various novel phases and phase transitions [6, 8, 9], whose continuing experimental exploration is supported by a suite of rapidly advancing numerical simulation technologies [10–14].

The quantum state of an array is probed with fluorescent imaging techniques, which provide projective measurements in the Rydberg occupation basis [4, 6, 15]. Since each measurement is destructive, the repetition rate at which they can be performed is limited by a number of factors, in particular the preparation time of the target quantum state. The probabilistic loading of the array requires a non-trivial rearrangement of atoms, resulting in repetition rates on the order of a few measurements per second [4, 6], with exact time scales depending on the specific experimental setup. Thus, data acquisition is limited, especially when compared to competing quantum simulation platforms, such as ion traps which allow for hundreds of measurements per second [16], or superconducting circuits where orders of magnitude more measurements per second can be achieved [17].

Data acquisition rates have obvious consequences for state reconstruction and characterization, as well as the direct estimation of operator expectation values, which suffer variances that scale inversely proportional to the number of independent measurements. Recently, neural network wavefunctions have been explored as tools for leveraging limited measurement data, as they provide a powerful ansatz for representing quantum states with systematically tunable expressivity [18–24]. For example, standard generative models adopted from the field of machine learning have been used to tomographically reconstruct quantum states [25–31] and have demonstrated the ability to significantly reduce the amount of measurements required for the accurate reconstruction of operator expectation values [25]. In addition to their designed ability for data-driven learning, these ansätze have the ability to find ground state wavefunctions of a given Hamiltonian by variational energy minimization, via the

* sczischek@uwaterloo.ca
† Currently at Apple

FIG. 1. (a) Square lattice of 6 × 6 atoms in the striated phase, where atoms in the Rydberg state (blue) are separated by stripes of atoms in the ground state (white) due to the Rydberg blockade radius $R_b$. Projective measurements of individual atoms in the occupation number basis provide information about the quantum state. (b) A recurrent neural network (RNN), with green boxes representing a cell with $N_h$ hidden units and tunable parameters $W$. An input sequence $\sigma$ is iteratively provided to the network, generating an output depending on the current input $\sigma_i$ and the hidden unit state $h_i$. A single qubit state is used as input at iteration $i$ and the network output is the probability distribution underlying the state of qubit $i+1$, from which the input $\sigma_{i+1}$ is sampled [18].
same Hamiltonian-driven training methods common in variational Monte Carlo (VMC) [32–34]. Modern neural network strategies provide VMC ansatzes that can systematically be made powerful enough that their expressiveness is no longer the limiting bottleneck. Instead, the optimization process often requires long convergence times [35–39] and physics-inspired modifications of the network structure are sometimes needed to reach accuracies comparable to traditional VMC approaches [40, 41]. In addition, the power of wisely chosen initializations to improve convergence has already been demonstrated in traditional VMC methods [42, 43].

In this work, we leverage these unique features of neural network wavefunctions to explore the effect of combined data- and Hamiltonian-driven learning [44]. Beginning with a randomly initialized recurrent neural network (RNN) [18], we first optimize network parameters using a limited amount of simulated [12, 14, 45] Rydberg occupation data drawn from a two-dimensional array in the vicinity of a quantum phase transition. Then, we continue optimizing the network variationally, in the spirit of the recent work by Carrasquilla and Torlai [33]. We find a significant enhancement in variationally obtaining the ground state wavefunction by pretraining the RNN on a limited amount of simulated [12, 14, 45] Rydberg occupations, as illustrated in Fig. 1(b), and choose the gated recurrent unit (GRU) [50] as network cell, inspired by Refs. [18, 33]. The RNN generates an output based on a sequence of inputs \( \sigma \) and the state of \( N_h \) hidden neurons per network cell. During the training process, the network parameters \( W \) are tuned to generate a target output. The amount of tunable parameters is defined by \( N_h \), which can be increased to improve the network expressivity. More details on RNNs are given in [18, 33, 50].

RNNs are naturally designed to encode probability distributions [18]. The network output at iteration \( i \) can be interpreted as the conditional probability distribution \( p_{\text{RNN}}(\sigma_i|\sigma_{i-1}, \ldots, \sigma_1; W) \), providing the joint distribution \( p_{\text{RNN}}(\sigma; W) = \prod_i p_{\text{RNN}}(\sigma_i|\sigma_{i-1}, \ldots, \sigma_1; W) \). To represent a wavefunction, single qubit states are chosen as network input, which is iterated over the entire qubit system. The RNN is then trained to encode the probability distribution underlying projective measurements in the computational basis. In the case of positive-real valued wavefunctions, such as the ground state of Eq. (1), the RNN represents the full quantum state \( \Psi_W(\sigma) = \langle \sigma | \Psi \rangle = \sqrt{p_{\text{RNN}}(\sigma; W)} \). Further modifications can be used to reconstruct arbitrary complex wavefunctions or density matrices [18, 33].

Finally, samples from the generative step of the RNN-encoded distribution emulate projective measurement outcomes. The quantum state of the full system can be sampled by iteratively drawing single qubit states. We use these samples to evaluate the energy expectation value \( H_{\text{RNN}} = \langle \Psi_W | \hat{H} | \Psi_W \rangle \) via,

\[
H_{\text{RNN}} = \sum_{\{\sigma\}} |\Psi_W(\sigma)|^2 H_{\text{loc}}(\sigma) \approx \frac{1}{N_s} \sum_{\sigma \sim \text{prnn}(\sigma; W)} H_{\text{loc}}(\sigma),
\]

where we introduce the local energy,

\[
H_{\text{loc}}(\sigma) = \frac{\langle \sigma | \hat{H} | \Psi_W \rangle}{\langle \sigma | \Psi_W \rangle},
\]

which is calculated and averaged over \( N_s \) samples \( \sigma \) drawn from \( p_{\text{RNN}}(\sigma; W) \) and can efficiently be evaluated for local non-diagonal operators [18, 20, 21, 32, 33].

**RNN training procedures.** We first explore the reconstruction of the ground state of a Rydberg atom array based on a projective measurement data set \( \mathcal{D} \). In this data-driven setting, we optimize an RNN to approximate the probability distribution \( p_{\text{RNN}}(\sigma) \approx p_D(\sigma) \) underlying the data points \( \sigma \in \mathcal{D} \). We use the Kullback-Leibler divergence to define the loss function,

\[
\mathcal{L}_{\text{DKL}}(W) = \sum_{\{\sigma\}} p_D(\sigma) \log \frac{p_D(\sigma)}{p_{\text{RNN}}(\sigma; W)} \approx -\mathcal{S}_D - \frac{1}{|\mathcal{D}|} \sum_{\sigma \in \mathcal{D}} \log p_{\text{RNN}}(\sigma; W).
\]

In the last line we approximate \( p_D \) with the sum over the data in \( \mathcal{D} \) and introduce the entropy \( S_D = \frac{1}{|\mathcal{D}|} \sum_{\sigma \in \mathcal{D}} \log p_{\text{RNN}}(\sigma; W) \).
Data-enhanced VMC. In the above, we have shown that RNN wavefunctions can be made sufficiently expressive to represent the ground state of large Rydberg arrays, however naive variational optimization of the neural network parameters does not lead to accurate energies in a reasonable amount of simulation time. Further, due to the long state-preparation times in modern Rydberg experiments [4, 6], it is reasonable to hypothesize that accurate data-driven reconstruction will be challenging on a reasonable amount of simulation time. Further, due to the long state-preparation times in modern Rydberg experiments [4, 6], it is reasonable to hypothesize that accurate data-driven reconstruction will be challenging on a reasonable amount of simulation time.

Next, we train the RNN to represent the ground state of the same \( N = \times L \) system using the Hamiltonian-driven approach. In this procedure, the RNN parameters are optimized such that the energy expectation value \( H_{\text{RNN}} \) is minimized, corresponding to VMC. With this motivation we define the loss function,

\[
\mathcal{L}_H(W) = \frac{1}{N_s} \sum_{\sigma} H_{\text{loc}}(\sigma),
\]

and use again the Adam optimizer [51] to find optimal network parameters \( W \). Here we evaluate the local energy, Eq. (3), on \( N_s \) samples drawn from \( p_{\text{RNN}}(\sigma; W) \).

In Fig. 2(b) we consider the energy density difference as a function of \( N_s \) for different system sizes. The largest system size, \( N = 16 \times 16 \), is not shown due to exceeding computational runtimes. We choose \( N_s = 2L \) and average the measured energy densities over optimization iterations 9500 to 10,000. The measured differences decrease with increasing \( N_s \), before they saturate at \( \sim 10^{-2} \) for \( N_s / L \geq 2 \) in all cases. While higher reconstruction accuracies can be reached, increasing \( N_s \) comes at the price of higher computational costs. We thus focus on \( N_s = 2L \) as a practical compromise in the following.

In order to produce a dataset \( D \), we use the quantum Monte Carlo (QMC) algorithm introduced in [14], which is able to accurately emulate projective measurements of Rydberg atoms in the ground state of the Hamiltonian in Eq. (1). We consider systems with up to \( N = 16 \times 16 \) atoms, relevant for state-of-the-art experimental realizations [6], for which an unbiased estimate of the energy \( H_{\text{QMC}} \) is easily obtained. We generate \( |D| = 10^5 \) QMC samples for all considered system sizes, which gives estimates of the ground state energy density \( H_{\text{QMC}} / N \) with errors on the order of \( \sim 10^{-4} \) (caption of Fig. 2).

We implement our RNN based on the code provided in [33, 53] and choose similar network hyperparameters, fixing the learning rate to \( \eta = 0.001 \) unless otherwise stated. We first analyze the expressivity of the RNN in the data-driven approach to training. In Fig. 2(a) we train the network on the entire set of \( |D| = 10^5 \) QMC data points for different system sizes of \( N = L \times L \) atoms and plot the energy density difference \( \langle H_{\text{RNN}} - H_{\text{QMC}} \rangle / N \) as a function of the number of hidden neurons \( N_h \) per network cell. The energy density differences are averaged over optimization iterations 1350 to 1400, where the training is approximately converged. Each iteration corresponds to one training epoch where the full input dataset is given to the RNN in batches of 100 randomly chosen samples and the network parameters are updated. The energy expectation value \( H_{\text{RNN}} \) is calculated on 1000 samples drawn from \( p_{\text{RNN}}(\sigma; W) \). For \( N = 16 \times 16 \) we average the energy densities over iterations 950 to 1000 due to long computational runtimes, leading to larger variances as convergence is not yet reached. For all system sizes the energy density error shows a clear decreasing trend with increasing \( N_h \), which corresponds to higher network expressivity. We find that the observed differences reach values below \( 10^{-2} \) for \( N_h / L \geq 2 \) in all cases.

![Figure 2](image-url)
large arrays due to limited data. To test this, we generate a randomly chosen subset of $D$ containing 1000 projective Rydberg occupation measurements, representing a typical amount of experimental data. In Fig. 3(a) we show the energy density expectation value when training networks with different numbers of hidden neurons $N_h$ on the Rydberg ground state of $N = 16 \times 16$ atoms. Instead of converging towards the estimated value, the energies for $N_h \geq 24$ reach a minimum at $\sim 700$ iterations. This phenomenon is easily recognized as overfitting.

Clearly, such limited datasets are insufficient for accurate state reconstruction. However, we now demonstrate the ability of small datasets to enhance the performance of VMC in a hybrid training procedure, defined with a simple change of loss function. Namely, we begin training the RNN with the data-driven loss function Eq. (5), before switching to Hamiltonian-driven training via Eq. (6) after $t_{\text{trans}}$ iterations. Fig. 3(b) illustrates the effectiveness of this simple hybrid procedure. The green diamonds show the evolution of the energy using 1000 data points in data-driven training, before switching to Hamiltonian-driven training after $t_{\text{trans}}$ iterations. As a comparison, we also plot purely Hamiltonian-driven training results (i.e. $t_{\text{trans}} = 0$). We explore a number of different choices of $t_{\text{trans}}$, which show significantly better convergence than the pure Hamiltonian-driven variational method out to 4000 optimization steps. All hybrid-trained simulations reach similar energy densities at $\gtrsim 3500$ iterations, which approximate the estimated value within a margin of 0.015 (see Fig. 3(b) inset).

In order to quantify the performance improvement of our algorithm, we define a convergence time $t_{\text{conv}}$ as the iteration after which the energy density difference reaches $(H_{\text{RNN}} - H_{\text{QMC}})/N \leq 0.015$ for the first time. Results are illustrated in Fig. 3(c), where for better accuracy we consider the running average over 50 iterations. The convergence time is significantly reduced for all $t_{\text{trans}} > 0$, while Hamiltonian-driven training converges after $t_{\text{conv}} \sim 6600$ iterations. The shortest convergence time for our hybrid algorithm is observed for $t_{\text{trans}} = 800$, which is around the minimum of the data-driven training curve. Numerical studies on smaller system sizes have shown similar improvements in the hybrid training approach. While overfitting in the data-driven training starts at different points for different system sizes and amounts of hidden neurons, Fig. 3(c) proposes that $t_{\text{trans}}$ does not need to be optimized to ensure convergence within reasonable computational runtimes.

Conclusions. In this paper, we have explored the use of recurrent neural networks (RNNs) for studying ground state wavefunctions of interacting Rydberg atom arrays of sizes currently accessible to experiments. We consider RNNs both trained from projective measurement data that could be produced by a typical experiment, as well as trained variationally with knowledge of the target Hamiltonian. In the latter, we show that naive variational training of RNNs becomes challenging for arrays approaching current experimental sizes. However, we find that RNNs which undergo a preliminary training phase driven by a small amount of measurement data converge to accurate ground state energies with significantly less optimization steps than RNNs optimized without access to data. This result indicates the value of projective measurement data, obtained from Rydberg arrays and other quantum simulators, as a mechanism to significantly improve the convergence times of variational Monte Carlo simulations. This strategy should be immediately accessible to RNNs and other neural network wavefunction ansätze, which are amenable to both data-driven and Hamiltonian-driven training.

Our work also clearly indicates that the current gener-
ation of Rydberg atom arrays can produce data of high value to physicists, even at measurement rates which may be too low to be informationally complete for a full tomographic reconstruction of the quantum state. This strategy can be combined with the recent observation that Hamiltonian-driven optimization can also be used to mitigate errors in noisy experiments [44]. This suggests that the current generation of quantum hardware is on the cusp of bringing transformative improvement to the understanding of challenging quantum many-body systems, by providing data to enhance variational simulation strategies for any state that can be prepared by an experimental quantum simulator.

ACKNOWLEDGEMENTS

We thank J. Carrasquilla, E. Inack, I. De Vlugt, M. P. A. Fisher, D. Sels, R. Luo, and Y. H. Teoh for critically important discussions. Simulations were made possible by the facilities of the Shared Hierarchical Academic Research Computing Network (SHARCNET) and Compute Canada. This work was supported by NSERC, the Canada Research Chair program, the New frontiers in graduate education Fund, and the Perimeter Institute for Theoretical Physics. Research at Perimeter Institute is supported in part by the Government of Canada through the Department of Innovation, Science and Economic Development Canada and by the Province of Ontario through the Ministry of Economic Development, Job Creation and Trade.

[1] P. Scholl, M. Schuler, H. Williams, A. Eberharter, D. Barredo, K.-N. Schymik, V. Lienhard, L.-P. Henry, T. Lang, T. Lahaye, A. Läuchli, and A. Browaeys, Nature 595, 233 (2021).
[2] I. Cong, S.-T. Wang, H. Levine, A. Keesling, and M. D. Lukin, arXiv:2105.13501 [quant-ph] (2021).
[3] H. Levine, A. Keesling, G. Semeghini, A. Omran, T. T. Wang, S. Ebadi, H. Bernien, M. Greiner, V. Vuletić, H. Pichler, and M. D. Lukin, Phys. Rev. Lett. 123, 170503 (2019).
[4] M. Endres, H. Bernien, A. Keesling, H. Levine, E. R. Anschuetz, A. Krajenbrink, C. Senko, V. Vuletić, M. Greiner, and M. D. Lukin, Science 354, 1024 (2016).
[5] D. Barredo, V. Lienhard, S. de Léséleuc, T. Lahaye, and A. Browaeys, Nature 561, 79 (2018).
[6] S. Ebadi, T. T. Wang, H. Levine, A. Keesling, G. Semeghini, A. Omran, D. Bluvstein, R. Samajdar, H. Pichler, W. W. Ho, S. Choi, S. Sachdev, M. Greiner, V. Vuletić, and M. D. Lukin, Nature 595, 227 (2021).
[7] P. Fendley, K. Sengupta, and S. Sachdev, Phys. Rev. B 69, 075106 (2004).
[8] G. Semeghini, H. Levine, A. Keesling, S. Ebadi, T. T. Wang, D. Bluvstein, R. Verresen, H. Pichler, M. Kalinowski, R. Samajdar, A. Omran, S. Sachdev, A. Vishwanath, M. Greiner, V. Vuletić, and M. D. Lukin, Science 374, 1242 (2021).
[9] C. Miles, R. Samajdar, S. Ebadi, T. T. Wang, H. Pichler, S. Sachdev, M. D. Lukin, M. Greiner, K. Q. Weinberger, and E.-A. Kim, arXiv:2112.10789 [quant-ph] (2021).
[10] R. Samajdar, W. W. Ho, H. Pichler, M. D. Lukin, and S. Sachdev, Phys. Rev. Lett. 124, 103601 (2020).
[11] R. Samajdar, W. W. Ho, H. Pichler, M. D. Lukin, and S. Sachdev, Proceedings of the National Academy of Sciences 118, e2015758118 (2021).
[12] M. Kalinowski, R. Samajdar, R. G. Melko, M. D. Lukin, S. Sachdev, and S. Choi, arXiv:2112.10790 [quant-ph] (2021).
[13] R. Verresen, M. D. Lukin, and A. Vishwanath, Phys. Rev. X 11, 031005 (2021).
[14] E. Merali, I. J. S. De Vlugt, and R. G. Melko, arXiv:2107.00766 [cond-mat] (2021).
[15] W. Xu, A. V. Venkatramani, S. H. Cantú, T. Šumarać, V. Klüßner, M. D. Lukin, and V. Vuletić, Phys. Rev. Lett. 127, 050501 (2021).
[16] C. Monroe, W. C. Campbell, L.-M. Duan, Z.-X. Gong, A. V. Gorshkov, P. W. Hess, R. Islam, K. Kim, N. M. Linke, G. Pagano, P. Richeimer, C. Senko, and N. Y. Yao, Rev. Mod. Phys. 93, 025001 (2021).
[17] T. Walter, P. Kurpiers, S. Gasparinetti, P. Magnard, A. Potocnik, Y. Salathé, M. Pechal, M. Mondal, M. Oppliger, C. Eichler, and A. Wallraff, Phys. Rev. Applied 7, 054020 (2017).
[18] M. Hibat-Allah, M. Ganahl, L. E. Hayward, R. G. Melko, and J. Carrasquilla, Phys. Rev. Research 2, 023358 (2020).
[19] G. Torlai and R. G. Melko, Phys. Rev. B 94, 165134 (2016).
[20] G. Carleo and M. Troyer, Science 355, 602 (2017).
[21] G. Torlai, G. Mazzola, J. Carrasquilla, M. Troyer, R. Melko, and G. Carleo, Nature Physics 14, 447 (2018).
[22] G. Torlai and R. G. Melko, Phys. Rev. Lett. 120, 240503 (2018).
[23] J. Carrasquilla, G. Torlai, R. G. Melko, and L. Aolita, Nat. Mach. Intell. 1, 155 (2019).
[24] H. Ma, D. Dong, I.-R. Petersen, C.-J. Huang, and G.-Y. Xiang, arXiv:2111.09504 [quant-ph] (2021).
[25] G. Torlai, G. Mazzola, G. Carleo, and A. Mezzacapo, Phys. Rev. Research 2, 022060 (2020).
[26] 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, Phys. Rev. Lett. 123, 230504 (2019).
[27] G. Torlai and R. G. Melko, Annual Review of Condensed Matter Physics 11, 325 (2020).
[28] M. Neugebauer, L. Fischer, A. Jäger, S. Czischek, S. Jochim, M. Weidemüller, and M. Gärttner, Phys. Rev. A 102, 042604 (2020).
[29] K. Vogel and H. Risken, Phys. Rev. A 40, 2847 (1989).
[30] M. Cramer, M. B. Plenio, S. T. Flammia, R. Somma, D. Gross, S. D. Bartlett, O. Landon-Cardinal, D. Poulin, and Y.-K. Liu, Nat. Commun. 1, 149 (2010).
[31] B. P. Lanyon, C. Monz, M. Hennrich, T. P. Schibli, C. Monz, M. Hennrich, T. P. Schibli, C. Helbig, P. Zanardi, A. Wallraff, A. Brádler, and A. Wallraff, Nature Physics 5, 103601 (2020).
[32] R. G. Melko, G. Carleo, J. Carrasquilla, and J. I. Cirac, Nature Physics 15, 887 (2019).
