# Learnability Scaling of Quantum States: Restricted Boltzmann Machines

**Dan Sehayek**
Department of Physics and Astronomy, University of Waterloo, Ontario, N2L 3G1, Canada

**Anna Golubeva,* Michael S. Albergo, Roger G. Melko**
Department of Physics and Astronomy, University of Waterloo, Ontario, N2L 3G1, Canada
Perimeter Institute for Theoretical Physics, Waterloo, Ontario N2L 2Y5, Canada

**Bohdan Kulchytskyy**
Department of Physics and Astronomy, University of Waterloo, Ontario, N2L 3G1, Canada
Perimeter Institute for Theoretical Physics, Waterloo, Ontario N2L 2Y5, Canada
1QB Information Technologies (1QBit), Waterloo, ON, Canada

**Giacomo Torlai**
Department of Physics and Astronomy, University of Waterloo, Ontario, N2L 3G1, Canada
Perimeter Institute for Theoretical Physics, Waterloo, Ontario N2L 2Y5, Canada
Center for Computational Quantum Physics, Flatiron Institute, New York, NY 10010, USA

## Abstract

Generative modeling with machine learning has provided a new perspective on the data-driven task of reconstructing quantum states from a set of qubit measurements. As increasingly large experimental quantum devices are built in laboratories, the question of how these machine learning techniques scale with the number of qubits is becoming crucial. We empirically study the scaling of restricted Boltzmann machines (RBMs) applied to reconstruct ground-state wavefunctions of the one-dimensional transverse-field Ising model from projective measurement data. We define a learning criterion via a threshold on the relative error in the energy estimator of the machine. With this criterion, we observe that the number of RBM weight parameters required for accurate representation of the ground state in the worst case – near criticality – scales quadratically with the number of qubits. By pruning small parameters of the trained model, we find that the number of weights can be significantly reduced while still retaining an accurate reconstruction. This provides evidence that over-parametrization of the RBM is required to facilitate the learning process.

## 1 Introduction

Generative models are machine learning algorithms that seek to reconstruct an unknown probability distribution $p(\mathbf{x})$ from a set of data $\mathbf{x}$. While several generative modeling techniques are available for quantum state reconstruction, by far the most well-studied involves restricted Boltzmann machines (RBMs) [37, 36, 12, 13, 9]. We present a systematic study of the scaling of the computational resources required for accurate reconstruction of a quantum state with an RBM. In particular, we focus on the ground-state wavefunction of a one-dimensional transverse-field Ising model, which

*agolubeva@pitp.ca

Second Workshop on Machine Learning and the Physical Sciences (NeurIPS 2019), Vancouver, Canada.

has a positive-real representation. Our training data is a set of projective measurements sampled independently from a simulated tensor-network wavefunction. We define a learning criterion based on the accuracy of the energy estimator of the RBM. The state reconstruction is considered successful when the relative error of the energy estimator is smaller than a fixed threshold. We target in particular two contributions to the asymptotic scaling behavior in the many-qubit limit: the representational power of the RBM, i.e., the *expressiveness* of the parameterization of the state, and the amount of data required to train the model, also known as the *sample complexity*.

We find that deep within the ferromagnetic and paramagnetic phases, the number of RBM parameters required for accurate representation of the ground state is $\mathcal{O}(1)$. As the transverse field is varied to approach the quantum critical point between these two phases, the state becomes more challenging to reconstruct, as expected due to long-range quantum correlations that arise there. At the critical point, we observe that under standard RBM training procedures the number of parameters grows quadratically in the number of qubits, $\mathcal{O}(N^2)$. The minimum number of measurements required to train this number of parameters scales linearly with the number of qubits, $\mathcal{O}(N)$. Interestingly, we find that the number of parameters required for an accurate reconstruction can be significantly reduced post-training by pruning small weights and fine-tuning the RBM by a small number of additional training iterations. We argue that an RBM requires over-parameterization to facilitate the optimization procedure associated with learning.

## 2 Defining a scaling study

The goal is to find the asymptotic scaling of the computational resources required to reconstruct a quantum state using an RBM. The training set comprises projective measurement data produced from the ground-state wavefunction of the one-dimensional transverse-field Ising model (TFIM) defined by the Hamiltonian

$$H = -J \sum_{\langle ij \rangle} \sigma_i^z \sigma_j^z - h \sum_i \sigma_i^x, \tag{1}$$

where $\sigma^{x,y,z}$ are Pauli operators, defined over $N$ sites (or qubits), and $\langle ij \rangle$ denotes nearest-neighbor pairs on a one-dimensional lattice with open boundary conditions. This model is thoroughly studied in the condensed matter and quantum information literature, and serves as a standard benchmark for many numerical methods. We generate training data from a density matrix renormalization group (DMRG) simulation [15] for various values of $h/J$ using the ITensor library [1]. The measurements of the ground-state wavefunction are produced in the $\sigma^z$ basis. The Perron-Frobenius theorem guarantees that when the Hamiltonian Eq. (1) has negative off-diagonal matrix elements in the $\sigma^z$ (computational) basis, the ground-state wavefunction is positive-real. Thus, there is a direct mapping between the wavefunction and a probability distribution, $\psi(\boldsymbol{\sigma}) = \sqrt{p(\boldsymbol{\sigma})}$. This allows for a significant simplification in the RBM network structure, since complex phases or signs need not be parametrized. In addition, the computational basis is trivially informationally complete, enabling training from data produced only in the $\sigma^z$ basis [36].

### 2.1 Restricted Boltzmann machine

The RBM consists of two layers of binary variables $v_i, h_j \in \{0, 1\}$. The energy associated with each configuration is given by $E_{\boldsymbol{\lambda}}(\boldsymbol{v}, \boldsymbol{h}) = -\sum_{ij} W_{ij} v_i h_j - \sum_i^N b_i v_i - \sum_j^{N_h} c_j h_j$, where $N$ is the number of visible units, representing the qubits or spins, and $N_h$ is the number of hidden units parametrizing the interactions. The two layers are fully connected via the weight matrix $\boldsymbol{W}$ that, along with the bias terms $b_i$ and $c_j$, forms the set of learnable parameters $\boldsymbol{\lambda} = (\boldsymbol{W}, \boldsymbol{b}, \boldsymbol{c})$. The energy function defines the joint probability distribution $p_{\boldsymbol{\lambda}}(\boldsymbol{v}, \boldsymbol{h}) = \exp[-E_{\boldsymbol{\lambda}}(\boldsymbol{v}, \boldsymbol{h})]/Z_{\boldsymbol{\lambda}}$, where $Z_{\boldsymbol{\lambda}}$ is the partition function of the machine. The marginal distribution $p_{\boldsymbol{\lambda}}(\boldsymbol{v}) = \sum_{\boldsymbol{h}} p_{\boldsymbol{\lambda}}(\boldsymbol{v}, \boldsymbol{h}) = \sum_{\boldsymbol{h}} \exp[-E_{\boldsymbol{\lambda}}(\boldsymbol{v}, \boldsymbol{h})]/Z_{\boldsymbol{\lambda}}$ is obtained by tracing out the hidden units. It is this marginal distribution that forms the approximate representation of the ground state, $\psi_{\boldsymbol{\lambda}}(\boldsymbol{v}) = \sqrt{p_{\boldsymbol{\lambda}}(\boldsymbol{v})}$. Therefore, the training procedure is equivalent to conventional unsupervised learning of an RBM [3]. In particular, the objective of the training procedure is to minimize the Kullback-Leibler (KL) divergence, which defines the discrepancy between the distribution of projective measurements and the probability distribution parameterized by the RBM, through a method known as contrastive divergence [20]. We use the QuCumber software package to implement and train a positive-real RBM [7].

### 2.2 Learning criterion

In order to quantify the resources required for the data-driven reconstruction of the ground-state wavefunction for the TFIM, one must be able to assess when the learning is "complete". We define the *learning criterion* as follows: Take $\bar{U} = \langle H \rangle_{\mathrm{RBM}}$ to be the average of the energy estimator calculated from $n$ samples generated by the RBM. Since $n$ is finite, a statistical error exists in the estimator, quantified by the standard deviation $\sigma$. To account for this in a relative error measure, we compute the Gaussian confidence interval given by $\bar{U} \pm C\sigma/\sqrt{n}$. The value of $C = 2.576$ corresponding to $99\%$ confidence will be used throughout this paper. If $U = \langle H \rangle_{\mathrm{exact}}$ is the exact value of the energy estimator (calculated with DMRG), then we can upper-bound the ROE by the larger relative error value of the confidence interval: $\epsilon = \max \left| \left[ U - (\bar{U} \pm C\sigma/\sqrt{n}) \right] / U \right|$. Essentially, this means that we consider the learning to be "complete" when our desired upper bound on the ROE is satisfied 99% of the time on our sample size. We find empirically that $\epsilon = 0.002$ is a reasonable value that can be achieved by RBMs trained on TFIM data with conventional algorithms for $N \leq 100$ qubits.

## 3 Results

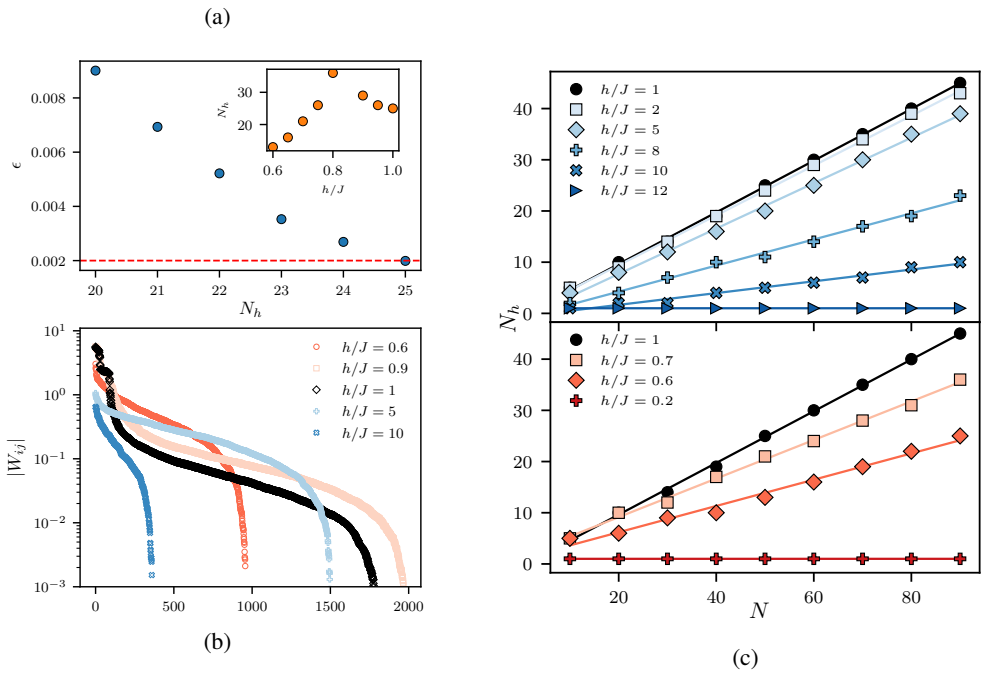


Figure 1: (a) The procedure used to determine the required RBM expressiveness for $N = 50$ qubits at $h/J = 1$. The number of hidden units $N_h$ is increased until the desired $\epsilon$ is achieved. Inset: $N_h$ required for convergence to $\epsilon \leq 0.002$ for different values of $h/J$. The position of the peak is discussed in the main text. (b) Weight magnitudes, sorted in descending order from left to right, for various transverse field values and $N = 60$. Converged RBM models from the parameter study shown in Fig. 1c are used here. (c) Minimum $N_h$ required for $\epsilon \leq 0.002$ for various values of $h/J$. Straight lines are fits to the data.

### 3.1 Scaling of the model parameters

To determine the minimal number of RBM parameters required to faithfully reproduce the ground-state energy of an $N$-qubit system, we study the scaling of the number of hidden units $N_h$. For each value of $N$, we produce projective measurements of $\sigma^z$ values using the DMRG simulation of the TFIM. Then, effectively assuming that the number of available training samples $M \to \infty$, we increase the number of hidden units $N_h$ until the learning criterion is uniquely satisfied for each value of $N$.

Table 1: The number of weights required to achieve $\epsilon \leq 0.002$ at the critical point $h/J = 1$.

| $N$ | 10 | 20 | 30 | 40 |
|---|---|---|---|---|
| original | 50 | 200 | 420 | 760 |
| pruned | 20 | 50 | 79 | 119 |

This procedure is illustrated in Fig. 1a for a fixed system size of $N = 50$. In the main plot, corresponding to $h/J = 1$ (critical point), we observe that the minimum number of hidden units required to accurately represent the ground-state wavefunction is $N_h = 25$. The inset illustrates the dependence of $N_h$ on field values near $h/J = 1$. One would expect that in the limit $N \to \infty$ the number of parameters required would be maximal at $h/J = 1$. Curiously, we find that this peak occurs around $h/J \approx 0.8$, slightly on the ferromagnetic side from the critical point. We hypothesize that this feature is tied to the magnetization of the underlying dataset used for training, which was produced by our DMRG simulations in ITensor. The result of repeating the above procedure for various numbers of qubits $N$ is illustrated in Fig. 1c. For values of $h/J$ far from the critical point, the required minimum number of hidden units scales as $N_h \sim \mathcal{O}(1)$ in the asymptotic limit of large $N$. This reflects the informational simplicity of the dataset in these regimes. Near $h/J \approx 1$ the scaling of $N_h$ is clearly linear, meaning that the leading asymptotic scaling of the number of parameters is $\mathcal{O}(N^2)$, as each additional hidden unit quadratically scales the number of parameters in the weight matrix $\boldsymbol{W}$. Finally, we note that for larger ROE thresholds $\epsilon > 0.002$ the prefactors and slopes are different, but the asymptotic scaling of the number of hidden units still remains linear near criticality.

### 3.2 Reducing the number of model parameters post-training

Implicit in the scaling result $N_h \approx \frac{1}{2}N$ from the previous section is the RBM optimization procedure: a stochastic gradient descent that minimizes the KL-divergence. This raises the question: Is it possible to find more efficient representations by modifying the learning protocol? Indeed, it has been found that the over-parameterization inherent to deep neural networks can ease and accelerate their optimization by (stochastic) gradient descent [29, 28, 5, 4, 35]. Figure 1b offers a clue that the RBM parameterization may not be optimal for the final trained wavefunction by demonstrating that the distribution of the weight magnitudes in a trained RBM is non-uniform. Recent machine learning literature has studied the relative importance of smaller weights with a procedure called *pruning*. Following the ideas of Refs. [19, 32], we define a pruning procedure for our scaling study in the following steps: (1) Start from the original, converged trained model. (2) Set a threshold $\delta$ for the weight magnitudes. If a given $|W_{ij}| < \delta$, set $W_{ij} = 0$, and freeze it for the following steps. (3) Fine-tune the pruned model by running several more training iterations until the desired accuracy is restored. (4) Repeat steps 2-4, pruning additional weights until the model fails to fulfill the learning criterion. We choose the pruning threshold such that 40% of the non-zero weights are pruned in the first iteration, and 5% of the non-zero weights in each following iteration.

We apply weight pruning to our trained RBM focusing on the critical point of the TFIM, and find that a significant reduction in the number of RBM parameters required to correctly capture the critical TFIM ground-state energy can be achieved for all system sizes. The results for several small numbers of qubits are presented in Table 1. We interpret this to mean that the standard training of an RBM benefits from employing more weights than is strictly required for accurate expression of the TFIM wavefunction in order to make the optimization more navigable. The success of the pruning procedure opens up the possibility of systematically searching for a change in scaling behavior. However, due to the significant increase in methodological complexity introduced by the pruning procedure, this analysis is out of scope for the current study and will be presented in another work.

## 4 Discussion

We have empirically studied the scaling of computational resources required for the accurate reconstruction of positive-real wavefunctions using generative modeling with a restricted Boltzmann machine (RBM), evaluated based on the relative error between the RBM estimator and the exact

energy value obtained from DMRG simulations.[2] For a standard optimization procedure with contrastive divergence, the number of weight parameters required for accurate reconstruction is at best constant (deep in the ferromagnetic/paramagnetic phases), and at worst quadratic (near the quantum critical point between the two phases). By employing a pruning technique, we present evidence that the number of parameters required to represent the ground state is drastically affected by the RBM learning procedure.It would then be interesting to compare the obtained results to theoretical expectations for the representational capacity of RBMs required for quantum ground-state wavefunctions [12, 17, 18]. It is natural to wonder what the scaling of computational resources is for reconstructing quantum states that are not real or positive. This question is especially pertinent for state-of-the-art experiments, such as fermionic quantum simulators [30], wavefunctions generated by quantum dynamics [26, 25], or quantum chemistry calculations with superconducting circuits [24]. In contrast to positive wavefunctions, the reconstruction (with a suitably modified RBM) demands training data from an extended set of measurement bases. The ability to theoretically identify the minimal set, and how the size of this set scales with the number of qubits, will ultimately determine the feasibility of integrating this type of machine learning technology into such near-term quantum devices.

## References


[1] *ITensor Library (version 2.0.11) http://itensor.org*.

[2] S. Aaronson. The learnability of quantum states. *Proc. R. Soc. A*, 463, 2007.

[3] D. H. Ackley, G. E. Hinton, and T. J. Sejnowski. A learning algorithm for boltzmann machines. *Cognitive Science*, 9(1):147 – 169, 1985.

[4] Z. Allen-Zhu, Y. Li, and Z. Song. A convergence theory for deep learning via over-parameterization. 2018.

[5] S. Arora, N. Cohen, and E. Hazan. On the optimization of deep networks: Implicit acceleration by overparameterization. *CoRR*, abs/1802.06509, 2018.

[6] M. J. S. Beach, R. G. Melko, T. Grover, and T. H. Hsieh. Making trotters sprint: A variational imaginary time ansatz for quantum many-body systems. 2019.

[7] M. J. S. Beach, I. D. Vlugt, A. Golubeva, P. Huembeli, B. Kulchytskyy, X. Luo, R. G. Melko, E. Merali, and G. Torlai. QuCumber: wavefunction reconstruction with neural networks. *SciPost Phys.*, 7:9, 2019.

[8] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborova. 2019.

[9] G. Carleo, Y. Nomura, and M. Imada. Constructing exact representations of quantum many-body systems with deep neural networks. *Nature Communications*, 9(1):5322, 2018.

[10] G. Carleo and M. Troyer. Solving the quantum many-body problem with artificial neural networks. *Science*, 355(6325):602–606, 2017.

[11] J. Carrasquilla, G. Torlai, R. G. Melko, and L. Aolita. Reconstructing quantum states with generative models. *Nature Machine Intelligence*, 1(3):155–161, 2019.

[12] J. Chen, S. Cheng, H. Xie, L. Wang, and T. Xiang. Equivalence of restricted boltzmann machines and tensor network states. *Phys. Rev. B*, 97:085104, Feb 2018.

[13] S. Cheng, J. Chen, and L. Wang. Information perspective to probabilistic modeling: Boltzmann machines versus born machines. *Entropy*, 20(8), 2018.

[14] S. Czischek, J. M. Pawlowski, T. Gasenzer, and M. Garttner. Sampling scheme for neuromorphic simulation of entangled quantum systems. 2019.


[2] We note that our scaling results are subject to the caveat that they could change if other criteria were to be considered, such as the convergence of fidelity or correlation functions.

[15] A. J. Ferris and G. Vidal. Perfect sampling with unitary tensor networks. *Phys. Rev. B*, 85:165146, Apr 2012.

[16] J. Frankle and M. Carbin. The lottery ticket hypothesis: Finding sparse, trainable neural networks. 2018.

[17] X. Gao and L.-M. Duan. Efficient representation of quantum many-body states with deep neural networks. *Nature Communications*, 8(1):662, 2017.

[18] I. Glasser, N. Pancotti, M. August, I. D. Rodriguez, and J. I. Cirac. Neural-network quantum states, string-bond states, and chiral topological states. *Phys. Rev. X*, 8:011006, Jan 2018.

[19] S. Han, J. Pool, J. Tran, and W. Dally. Learning both weights and connections for efficient neural network. In C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems 28*, pages 1135–1143. Curran Associates, Inc., 2015.

[20] G. E. Hinton. Training products of experts by minimizing contrastive divergence. *Neural Computation*, 14(8):1771–1800, 2002.

[21] W. W. Ho and T. H. Hsieh. Efficient variational simulation of non-trivial quantum states. *SciPost Phys.*, 6:29, 2019.

[22] W. W. Ho, C. Jonay, and T. H. Hsieh. Ultrafast variational simulation of nontrivial quantum states with long-range interactions. *Phys. Rev. A*, 99:052332, May 2019.

[23] E. M. Inack, G. Giudici, T. Parolini, G. Santoro, and S. Pilati. Understanding quantum tunneling using diffusion monte carlo simulations. *Phys. Rev. A*, 97:032307, Mar 2018.

[24] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta. Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. *Nature*, 549:242 EP –, 09 2017.

[25] A. Keesling, A. Omran, H. Levine, H. Bernien, H. Pichler, S. Choi, R. Samajdar, S. Schwartz, P. Silvi, S. Sachdev, P. Zoller, M. Endres, M. Greiner, V. Vuletic, and M. D. Lukin. Probing quantum critical dynamics on a programmable Rydberg simulator. sep 2018.

[26] B. P. Lanyon, C. Maier, M. Holzäpfel, T. Baumgratz, C. Hempel, P. Jurcevic, I. Dhand, A. S. Buyskikh, A. J. Daley, M. Cramer, M. B. Plenio, R. Blatt, and C. F. Roos. Efficient tomography of a quantum many-body system. *Nature Physics*, 13(12):1158–1162, sep 2017.

[27] N. Lee, T. Ajanthan, and P. H. S. Torr. Snip: Single-shot network pruning based on connection sensitivity. 2018.

[28] R. Livni, S. Shalev-Shwartz, and O. Shamir. On the computational efficiency of training neural networks. *CoRR*, abs/1410.1141, 2014.

[29] D. Lopez-Paz and L. Sagun. Easing non-convex optimization with neural networks, 2018.

[30] A. Mazurenko, C. S. Chiu, G. Ji, M. F. Parsons, M. Kanász-Nagy, R. Schmidt, F. Grusdt, E. Demler, D. Greif, and M. Greiner. A cold-atom fermi–hubbard antiferromagnet. *Nature*, 545:462 EP –, 05 2017.

[31] R. G. Melko, G. Carleo, J. Carrasquilla, and J. I. Cirac. Restricted boltzmann machines in quantum physics. *Nature Physics*, 2019.

[32] D. C. Mocanu, E. Mocanu, P. H. Nguyen, M. Gibescu, and A. Liotta. A topological insight into restricted boltzmann machines. *Machine Learning*, 104(2):243–270, Sep 2016.

[33] M. Paris and J. Rehacek, editors. *Quantum State Estimation*, volume 649 of *Lecture Notes in Physics*. Springer-Verlag Berlin Heidelberg, 2004.

[34] A. W. Sandvik. Stochastic series expansion method for quantum ising models with arbitrary interactions. *Phys. Rev. E*, 68:056701, Nov 2003.

[35] K. A. Sankararaman, S. De, Z. Xu, W. R. Huang, and T. Goldstein. The impact of neural network overparameterization on gradient confusion and stochastic gradient descent. 2019.

[36] G. Torlai, G. Mazzola, J. Carrasquilla, M. Troyer, R. Melko, and G. Carleo. Neural-network quantum state tomography. *Nature Physics*, 14(5):447–450, 2018.

[37] G. Torlai and R. G. Melko. Learning thermodynamics with Boltzmann machines. *Physical Review B*, 94(16):165134, Oct. 2016.

[38] G. Torlai and R. G. Melko. Machine learning quantum states in the nisq era. 2019.

[39] G. Torlai, B. Timar, E. P. van Nieuwenburg, H. Levine, A. Omran, A. Keesling, H. Bernien, M. Greiner, V. Vuletić, M. D. Lukin, R. G. Melko, and M. Endres. Integrating neural networks with a quantum simulator for state reconstruction. 2019.

[40] G. Vidal. Entanglement renormalization. *Phys. Rev. Lett.*, 99:220405, Nov 2007.

[41] H. Zhou, J. Lan, R. Liu, and J. Yosinski. Deconstructing lottery tickets: Zeros, signs, and the supermask. 2019.