Variational Neural Annealing

Mohamed Hibat-Allah[1,2]**, Estelle M. Inack**[1,3]**, Roeland Wiersema**[1,2]**, Roger G. Melko**[2,3]**, Juan Carrasquilla**[1,2]

[1] Vector Institute, MaRS Centre, Toronto, Ontario, M5G 1M1, Canada
[2] Department of Physics and Astronomy, University of Waterloo, Ontario, N2L 3G1, Canada
[3] Perimeter Institute for Theoretical Physics, Waterloo, ON N2L 2Y5, Canada
mohamed.hibat.allah@uwaterloo.ca

Abstract

Many combinatorial optimization problems relevant to computer science, computational biology and physics can be tackled with simulated annealing, which is a powerful framework for optimizing the properties of complex systems through the lens of statistical mechanics. However, simulated annealing (SA) and its quantum counterpart, simulated quantum annealing (SQA), are traditionally implemented via Markov chain Monte Carlo, often displaying slow convergence to optimal solutions for challenging optimization problems. Here we combine the variational principle in classical and quantum physics with recurrent neural networks, whose dynamics are naturally devoid of slow Markov chains to accurately emulate annealing in its classical and quantum formulations. We find that a variational implementation of classical annealing is not only superior to its quantum analog in terms of speed of convergence and accuracy of solutions but also better than traditional SA and SQA for the Edwards-Anderson model on system sizes up to 40×40 spins.

1   Introduction

Various heuristics have been used over the years to find approximate solutions to optimization problems. Among the most notable heuristic methods is SA [1, 2, 3], which refers to a family of in silico techniques for the optimization of complex systems. SA mirrors the analogous physical annealing process in materials science and metallurgy where a crystalline solid is heated and then slowly cooled down to its lowest energy and most structurally stable crystal arrangement. Through the addition of quantum fluctuations, the conceptual advances brought by thermal annealing can be transferred to the quantum realm. The resulting algorithm, known as quantum annealing (QA), has gained momentum since the advent of commercially available quantum annealing devices [4, 5, 6] which physically implement annealing in the presence of quantum fluctuations.

Machine learning, already explored as a condensed-matter and statistical physics research tool, provides a complementary paradigm to the above approaches to optimization of complex systems [7]. Artificial neural networks in particular have been used to identify phases of matter [8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18] and increase the performance of Monte Carlo simulations [19, 20, 21, 22, 23, 24, 25, 26]. A peculiar aspect of neural networks is their ability to accurately represent the underlying state of classical [27] and quantum systems [28, 29, 8, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39]. Notably, by combining them with variational principles, neural autoregressive models such as the recurrent neural network (RNN) can accurately describe the equilibrium properties of classical and quantum systems by taking advantage of autoregressive sampling which naturally avoids the use of Markov chains [38, 39]. These models learn compact representations of the correlations between the components in a physical system in a similar way to natural language.

In this paper, we take advantage of the expressive power of RNNs to devise a variational version of classical and quantum annealing to solve optimization problems. We show that our variational
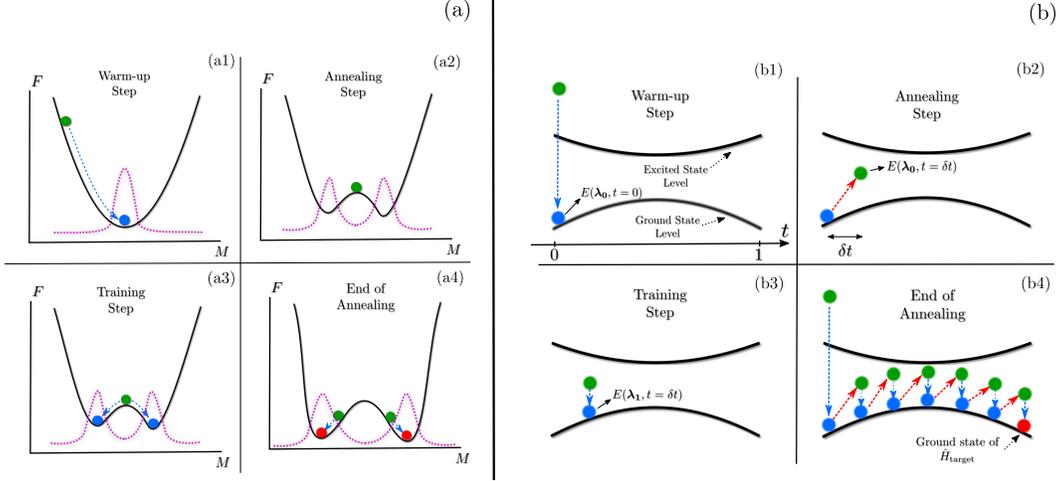
Figure 1: (a) An illustration of the variational classical annealing protocol. Here we illustrate the free energy against the order parameter of the system. The dashed curves denotes the probability distribution that minimizes the free energy (b) A simplified illustration of the variational quantum annealing protocol. Note that downward dashed arrows stands for gradient descent minimization.

emulation of classical annealing outperforms standard heuristics, namely SA and SQA [40, 41, 42] on the 2D Edwards-Anderson model.

2 Methods

2.1 Variational Classical Annealing

We first consider a variational implementation of simulated (or classical) annealing termed *variational classical annealing* (VCA). We take the variational approach of statistical mechanics [27], where a parameterized probability distribution $p_{\boldsymbol{\lambda}}(\boldsymbol{\sigma})$ approximates the Boltzmann distribution of the system. $\boldsymbol{\lambda}$ are the parameters of the model distribution which are optimized so that $p_{\boldsymbol{\lambda}}(\boldsymbol{\sigma})$ represented with RNNs, reproduces the equilibrium properties of the system at temperature T. In VCA, we search for the ground state of an optimization problem encoded in an Hamiltonian $\hat{H}_{\text{target}}$, by considering the variational free energy

$$F_{\boldsymbol{\lambda}}(t) = \langle \hat{H}_{\text{target}} \rangle_{\boldsymbol{\lambda}} - T(t) S_{\text{classical}}(p_{\boldsymbol{\lambda}}), \tag{1}$$

which provides an upper bound to the true instantaneous free energy and, can therefore be used at each annealing stage to update $\boldsymbol{\lambda}$ via gradient-descent techniques. The parameters that are optimized at temperature $T(t)$ are used as input at temperature $T(t + \delta t)$ to ensure that the model's distribution is always near its instantaneous equilibrium state. $\langle ... \rangle_{\boldsymbol{\lambda}}$ denotes ensemble averages taken over the probability $p_{\boldsymbol{\lambda}}(\boldsymbol{\sigma})$. The classical von Neumann entropy is given by $S_{\text{classical}}(p_{\boldsymbol{\lambda}}) = -\sum_{\boldsymbol{\sigma}} p_{\boldsymbol{\lambda}}(\boldsymbol{\sigma}) \log(p_{\boldsymbol{\lambda}}(\boldsymbol{\sigma}))$, where the sum runs over all the elements of the state space $\{\boldsymbol{\sigma}\}$. In our annealing setting, the temperature is decreased from an initial value T_0 to 0 using a linear schedule $T(t) = T_0(1 - t)$, where time $t \in [0, 1]$, which is analogous to traditional implementations of SA. We note that this annealing approach has been suggested in Ref. [27] as a possible solution to the mode collapse problem in the simulations of the equilibrium properties of classical systems. Here, we implement it to solve optimization problems with RNNs.

We summarize the VCA procedure in Fig. 1(a). We initially perform a warm-up step to bring a randomly initialized variational probability distribution to the equilibrium state of $F_{\boldsymbol{\lambda}}(t = 0)$. This is done by performing N_{warmup} gradient descent steps to minimize the free energy $F_{\boldsymbol{\lambda}}(t = 0)$ as illustrated in Fig. 1(a1). At each time step t, we reduce the temperature of the system from $T(t)$ to $T(t + \delta t)$. This effectively changes the shape of the free energy landscape of the system (see Fig. 1(a2)). We then perform a training step by taking N_{train} gradient descent steps in order to re-equilibrate the model to a new minimum of the free energy at $T(t + \delta t)$ as illustrated in Fig. 1(a3). Finally, by repeating over the last two steps $N_{\text{annealing}}$ times, we reach temperature $T(1) = 0$, which

is the end of the annealing protocol (see Fig. 1(a4)). In this case $N_{\text{annealing}} = (t_f - t_i)/\delta t = 1/\delta t$ where δt is a user-defined constant. In the absence of thermal fluctuations, $p_{\boldsymbol{\lambda}}(\boldsymbol{\sigma})$ is expected to converge to a distribution peaked around configurations with energies close to the ground states of the target Hamiltonian $\hat{H}_{\text{target}}$.

2.2 Variational Quantum Annealing

In quantum annealing [43, 44, 45, 46], the search for the ground state of an optimization problem encoded in the target Hamiltonian $\hat{H}_{\text{target}}$, is generally done by considering the following quantum Hamiltonian

$$\hat{H}(t) = \hat{H}_{\text{target}} + f(t)\hat{H}_D, \tag{2}$$

where quantum fluctuations are introduced via a driving term $\hat{H}_D$ that does not commute with $\hat{H}_{\text{target}}$. The function $f(t)$ is a user-defined time-dependent schedule function such that $f(0) = 1$ and $f(1) = 0$. Quantum annealing usually starts with a dominant driving term $\hat{H}_D \gg \hat{H}_{\text{target}}$ chosen so that its ground state is easy to prepare and simulate. The strength of the driving term is then subsequently reduced –typically adiabatically– using the schedule function f so that at the end of annealing, the system is in the lowest state of the target Hamiltonian. In our paper, we choose a linear schedule $f(t) = 1 - t$.

Here, we leverage the variational principle of quantum mechanics and propose a strategy to simulate quantum annealing dubbed *variational quantum annealing* (VQA). Our framework is based on the variational Monte Carlo (VMC) method, a quantum Monte Carlo that simulates equilibrium properties of quantum many-body systems at zero-temperature [47]. In VMC, the ground-state wave function of a quantum Hamiltonian $\hat{H}$ is modeled via an ansatz $|\Psi_{\boldsymbol{\lambda}}\rangle$ where $\boldsymbol{\lambda}$ are the variational parameters. The variational principle guarantees that the expectation value of the energy over the variational state $\langle\Psi_{\boldsymbol{\lambda}}|\hat{H}|\Psi_{\boldsymbol{\lambda}}\rangle$ is an upper bound to the ground state energy of $\hat{H}$. Thus, we use it as a cost function to optimize the parameters $\boldsymbol{\lambda}$. In a similar spirit, within our VQA setting, we define a time-dependent cost function $E(\boldsymbol{\lambda}, t) \equiv \langle\hat{H}(t)\rangle_{\boldsymbol{\lambda}} = \langle\Psi_{\boldsymbol{\lambda}}|\hat{H}(t)|\Psi_{\boldsymbol{\lambda}}\rangle$.

Our VQA setup is implemented via the protocol described in Fig. 1(b). We start by randomly initializing the parameters $\boldsymbol{\lambda}$ of the variational wave function. Then, we perform a warm-up step to prepare our ansatz close to the ground state of the Hamiltonian $\hat{H}(0)$, as illustrated in Fig. 1(b1). To do so, we apply N_{warmup} gradient descent steps to stochastically minimize the expectation value $E(\boldsymbol{\lambda}, t)$ at a fixed time $t = 0$. The obtained variational energy after this step is $E(\boldsymbol{\lambda}_0, t = 0)$. Next, we perform an annealing step by changing the time t from $t = 0$ to $t = \delta t$, while keeping the parameters $\boldsymbol{\lambda}_0$ of the variational wave function fixed. Hence, the obtained variational energy is $E(\boldsymbol{\lambda}_0, t = \delta t)$ as shown in Fig. 1(b2). Furthermore, we apply a training step by taking N_{train} gradient descent steps to bring the ansatz closer to the new instantaneous ground state. At the end of the training step, we obtain the energy $E(\boldsymbol{\lambda}_1, t = \delta t)$ as illustrated in Fig. 1(b3). Finally, we repeat the annealing step and the training step $N_{\text{annealing}}$ times until arriving at $t = 1$, which indicates the end of the annealing. Here, the system is expected to converge to the ground state of the optimization problem (see Fig. 1(b4)). Analogously to VCA, we choose RNN wave functions [38, 39] as ansätze to implement the protocol described above.

3 Results

We benchmark our variational neural annealing algorithms on a 2D problem Hamiltonian that not only exhibits disorder but also frustration, a property that makes it harder to solve as in constraint satisfaction problems. We focus on the Edwards-Anderson (EA) spin glass which has been studied experimentally [48], numerically [40, 41, 42] and theoretically [49]. This model is specified by the target Hamiltonian: $\hat{H}_{\text{EA}} = -\sum_{\langle i,j\rangle} J_{ij}\hat{\sigma}_i^z\hat{\sigma}_j^z$, where the sum runs over nearest neighbours and the coupling J_{ij} are drawn independently from a uniform distribution in the range $[-1, 1]$. In the absence of a longitudinal field for which solving the EA model is NP-hard, the ground state can be found in polynomial time [49]. For each random realization of the couplings J_{ij}, we use the spin-glass server [50] to obtain the exact ground state E_{G}. This feature makes EA model a good benchmark for our method particularly for large system sizes. To quantify the accuracy of our method, we use the residual energy ϵ_{res} [51, 40, 41, 52, 53], defined as $\epsilon_{\text{res}} = \left[\langle\hat{H}_{\text{EA}}\rangle_{\text{stat}} - E_{\text{G}}\right]_{\text{typ}}$. For VQA and VCA,
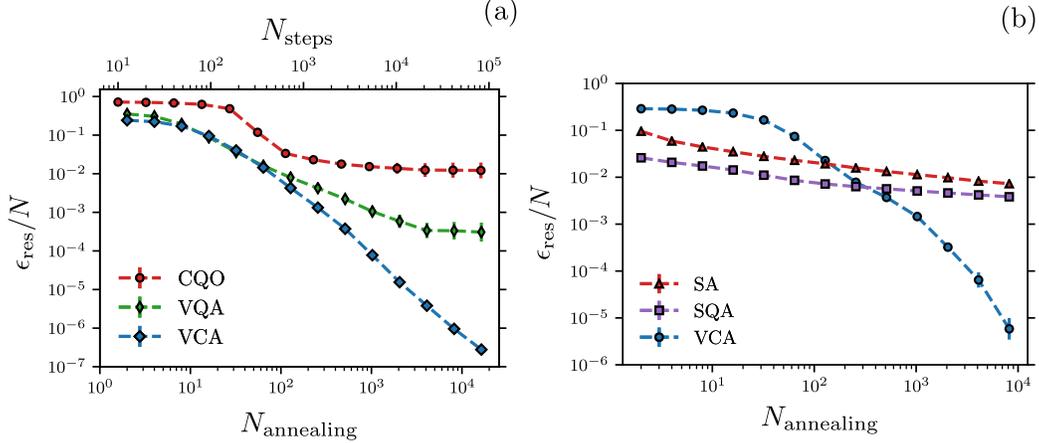
Figure 2: (a) Comparison of the residual energies per site between VCA, VQA and CQO for the Edwards-Anderson model with system size of 10×10 spins. (b) Similar comparison between VCA, SA and SQA for a system size of 40×40 spins.

$\langle ... \rangle_{\text{stat}}$ stands for arithmetic averages taken over 10^6 different independent configurations, that we autoregressively sample at the end of annealing from the RNN. Our choice for the RNN architecture is inspired by the success observed in Ref. [38], where a two-dimensional RNN was capable of representing the ground state of the two-dimensional transverse field Ising model with high accuracy. Here, we use 25 samples per each gradient update to train the RNNs. For SA and SQA, it denotes the arithmetic average over 25 independent Monte Carlo runs. $[...]_{\text{typ}}$ represents the typical (geometric) average over 25 different realizations of disorder.

As a first benchmark for VCA and VQA, we choose a system size $N = 10 \times 10$ spins. First, we run VQA with a one-body driving term $\hat{H}_D = -\Gamma_0 \sum_i \hat{\sigma}_i^x$ as shown in Eq. (2) and with an initial transverse magnetic field $\Gamma_0 = 1$. For VCA, we use the classical von Neumann entropy as a driving term in Eq. (1) and an initial temperature $T_0 = 1$. To demonstrate the advantage brought by the annealing protocols, we first perform VCA with zero thermal fluctuations, i.e., with an initial temperature $T_0 = 0$, known in the literature as classical-quantum optimization (CQO) [54, 55, 56]. In Fig. 2(a), we plot the residual energy per site against the number of annealing steps $N_{\text{annealing}}$ for VCA and VQA and against the number of training steps N_{steps} for CQO. We observe that finding the ground state of the EA model by performing directly a stochastic optimization of the variational energy of the problem Hamiltonian is less efficient compared to emulating an annealing dynamics in the optimization procedure. We equally observe that the VCA method is orders of magnitude more accurate than VQA and CQO for a large number of annealing steps.

Since VCA was found to be optimal in the previous benchmark, we use it to further demonstrate the ability to do variational annealing on a larger system size of 40×40 spins. Note that the VCA protocol remains the same for all system sizes provided that the hyperparameters of the model, namely the number of hidden states of the RNN cell, are sufficient to capture all the properties of the system at those sizes. For comparison, we use SA as well as SQA [40, 41] with $P = 20$ trotter slices. We show the results in Fig. 2(b), where we plot the residual energy per site against $N_{\text{annealing}}$ for SA, SQA and VCA. We first confirm the qualitative behavior of SA and SQA obtained in Refs. [40, 41]. We also observe that VCA can achieve a lower residual energy for large annealing steps, which is about three orders of magnitude better than SQA and SA.

Conclusion

We presented a variational emulation of classical and quantum annealing using recurrent neural networks on the 2D Edwards-Anderson spin glass. In contrast to known results in the literature [40, 42], we have found variational classical annealing superior to its quantum counterpart. Furthermore, compared to SA and SQA, VCA achieves significantly better results on the EA model. Additional preliminary results suggest that these advantages remain true for the Sherrington-Kirkpatrick model

and other fully connected glasses [57]. These results advocate for the use of VCA as a competitive algorithm to tackle real-world optimization problems.

Broader Impact

Optimization problems have a wide field of application in various areas such as science, industry, governance and medicine. For this reason, our framework could be used in those areas to tackle challenging problems and, potentially foster technological advancements, especially in the benchmarking of future quantum computers. We point out that the defense sector may also show interest in this research, as demonstrated by its current investment in quantum computing in general and quantum annealing technology in particular.

Acknowledgments and Disclosure of Funding

References

[1] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, 1983. ISSN 0036-8075. doi: 10.1126/science.220.4598.671. URL `https://science.sciencemag.org/content/220/4598/671`.

[2] V. Černý. Thermodynamical approach to the traveling salesman problem: An efficient simulation algorithm. *Journal of Optimization Theory and Applications*, 45(1):41–51, Jan 1985. ISSN 1573-2878. doi: 10.1007/BF00940812. URL `https://doi.org/10.1007/BF00940812`.

[3] Maliheh Aramon, Gili Rosenberg, Elisabetta Valiante, Toshiyuki Miyazawa, Hirotaka Tamura, and Helmut G. Katzgraber. Physics-inspired optimization for quadratic unconstrained problems using a digital annealer. *Frontiers in Physics*, 7:48, 2019. ISSN 2296-424X. doi: 10.3389/fphy.2019.00048. URL `https://www.frontiersin.org/article/10.3389/fphy.2019.00048`.

[4] MW Johnson, MHS Amin, S Gildert, T Lanting, F Hamze, N Dickson, R Harris, AJ Berkley, J Johansson, P Bunyk, et al. Quantum annealing with manufactured spins. *Nature*, 473(7346): 194–198, 2011. URL `https://www.nature.com/nature/journal/v473/n7346/full/nature10012.html`.

[5] Sergio Boixo, Tameem Albash, Federico M Spedalieri, Nicholas Chancellor, and Daniel A Lidar. Experimental signature of programmable quantum annealing. *Nat. Commun.*, 4:2067, 2013. URL `https://www.nature.com/articles/ncomms3067`.

[6] Sergio Boixo, Troels F Rønnow, Sergei V Isakov, Zhihui Wang, David Wecker, Daniel A Lidar, John M Martinis, and Matthias Troyer. Evidence for quantum annealing with more than one hundred qubits. *Nat. Phys.*, 10(3):218–224, 2014. URL `http://www.nature.com/nphys/journal/v10/n3/full/nphys2900.html`.

[7] Yoshua Bengio, Andrea Lodi, and Antoine Prouvost. Machine learning for combinatorial optimization: A methodological tour d'horizon. *European Journal of Operational Research*, 2020. ISSN 0377-2217. doi: https://doi.org/10.1016/j.ejor.2020.07.063. URL `http://www.sciencedirect.com/science/article/pii/S0377221720306895`.

[8] Juan Carrasquilla and Roger Melko. Machine learning phases of matter. *Nature Physics*, 14: 431–434, 2017. doi: https://doi.org/10.1038/nphys4035.

[9] E. van Nieuwenburg, Y. Liu, and S. Huber. Learning phase transitions by confusion. *Nature Physics*, 13:435—439, 2017. doi: https://doi.org/10.1038/nphys4037.

[10] Giacomo Torlai and Roger G. Melko. Learning thermodynamics with boltzmann machines. *Phys. Rev. B*, 94:165134, Oct 2016. doi: 10.1103/PhysRevB.94.165134. URL `https://link.aps.org/doi/10.1103/PhysRevB.94.165134`.

[11] Lei Wang. Discovering phase transitions with unsupervised learning. *Phys. Rev. B*, 94:195105, Nov 2016. doi: 10.1103/PhysRevB.94.195105. URL `https://link.aps.org/doi/10.1103/PhysRevB.94.195105`.

[12] Kelvin Ch'ng, Juan Carrasquilla, Roger G. Melko, and Ehsan Khatami. Machine learning phases of strongly correlated fermions. *Phys. Rev. X*, 7:031038, Aug 2017. doi: 10.1103/PhysRevX.7.031038. URL `https://link.aps.org/doi/10.1103/PhysRevX.7.031038`.

[13] Peter Broecker, Juan Carrasquilla, Roger G Melko, and Simon Trebst. Machine learning quantum phases of matter beyond the fermion sign problem. *Scientific Reports*, 7:8823, 2017. doi: https://doi.org/10.1038/s41598-017-09098-0.

[14] Yi Zhang and Eun-Ah Kim. Quantum loop topography for machine learning. *Phys. Rev. Lett.*, 118:216401, May 2017. doi: 10.1103/PhysRevLett.118.216401. URL `https://link.aps.org/doi/10.1103/PhysRevLett.118.216401`.

[15] Dong-Ling Deng, Xiaopeng Li, and S. Das Sarma. Machine learning topological states. *Phys. Rev. B*, 96:195145, Nov 2017. doi: 10.1103/PhysRevB.96.195145. URL `https://link.aps.org/doi/10.1103/PhysRevB.96.195145`.

[16] Frank Schindler, Nicolas Regnault, and Titus Neupert. Probing many-body localization with neural networks. *Phys. Rev. B*, 95:245134, Jun 2017. doi: 10.1103/PhysRevB.95.245134. URL `https://link.aps.org/doi/10.1103/PhysRevB.95.245134`.

[17] Yi-Ting Hsu, Xiao Li, Dong-Ling Deng, and S. Das Sarma. Machine learning many-body localization: Search for the elusive nonergodic metal. *Phys. Rev. Lett.*, 121:245701, Dec 2018. doi: 10.1103/PhysRevLett.121.245701. URL `https://link.aps.org/doi/10.1103/PhysRevLett.121.245701`.

[18] Patrick Huembeli, Alexandre Dauphin, and Peter Wittek. Identifying quantum phase transitions with adversarial neural networks. *Phys. Rev. B*, 97:134109, Apr 2018. doi: 10.1103/PhysRevB.97.134109. URL `https://link.aps.org/doi/10.1103/PhysRevB.97.134109`.

[19] E. M. Inack, G. E. Santoro, L. Dell'Anna, and S. Pilati. Projective quantum monte carlo simulations guided by unrestricted neural network states. *Phys. Rev. B*, 98:235145, Dec 2018. doi: 10.1103/PhysRevB.98.235145. URL `https://link.aps.org/doi/10.1103/PhysRevB.98.235145`.

[20] T. Parolini, E. M. Inack, G. Giudici, and S. Pilati. Tunneling in projective quantum monte carlo simulations with guiding wave functions. *Phys. Rev. B*, 100:214303, Dec 2019. doi: 10.1103/PhysRevB.100.214303. URL `https://link.aps.org/doi/10.1103/PhysRevB.100.214303`.

[21] S. Pilati, E. M. Inack, and P. Pieri. Self-learning projective quantum monte carlo simulations guided by restricted boltzmann machines. *Phys. Rev. E*, 100:043301, Oct 2019. doi: 10.1103/PhysRevE.100.043301. URL `https://link.aps.org/doi/10.1103/PhysRevE.100.043301`.

[22] B. McNaughton, M. V. Milošević, A. Perali, and S. Pilati. Boosting monte carlo simulations of spin glasses using autoregressive neural networks. *Phys. Rev. E*, 101:053312, May 2020. doi: 10.1103/PhysRevE.101.053312. URL `https://link.aps.org/doi/10.1103/PhysRevE.101.053312`.

[23] Junwei Liu, Yang Qi, Zi Yang Meng, and Liang Fu. Self-learning monte carlo method. *Phys. Rev. B*, 95:041101, Jan 2017. doi: 10.1103/PhysRevB.95.041101. URL `https://link.aps.org/doi/10.1103/PhysRevB.95.041101`.

[24] Xiao Yan Xu, Yang Qi, Junwei Liu, Liang Fu, and Zi Yang Meng. Self-learning quantum monte carlo method in interacting fermion systems. *Phys. Rev. B*, 96:041119, Jul 2017. doi:

10.1103/PhysRevB.96.041119. URL https://link.aps.org/doi/10.1103/PhysRevB.96.041119.

[25] Li Huang and Lei Wang. Accelerated monte carlo simulations with restricted boltzmann machines. *Phys. Rev. B*, 95:035105, Jan 2017. doi: 10.1103/PhysRevB.95.035105. URL https://link.aps.org/doi/10.1103/PhysRevB.95.035105.

[26] M. S. Albergo, G. Kanwar, and P. E. Shanahan. Flow-based generative models for markov chain monte carlo in lattice field theory. *Phys. Rev. D*, 100:034515, Aug 2019. doi: 10.1103/PhysRevD.100.034515. URL https://link.aps.org/doi/10.1103/PhysRevD.100.034515.

[27] Dian Wu, Lei Wang, and Pan Zhang. Solving statistical mechanics using variational autoregressive networks. *Physical Review Letters*, 122(8), Feb 2019. ISSN 1079-7114. doi: 10.1103/physrevlett.122.080602. URL http://dx.doi.org/10.1103/PhysRevLett.122.080602.

[28] J. Androsiuk, L. Kulak, and K. Sienicki. Neural network solution of the Schrödinger equation for a two-dimensional harmonic oscillator. *Chemical Physics*, 173(3):377–383, 1993. ISSN 0301-0104. doi: 10.1016/0301-0104(93)80153-Z.

[29] Artificial neural network methods in quantum mechanics. *Computer Physics Communications*, 104(1):1 – 14, 1997. ISSN 0010-4655. doi: https://doi.org/10.1016/S0010-4655(97)00054-4. URL http://www.sciencedirect.com/science/article/pii/S0010465597000544.

[30] Giuseppe Carleo and Matthias Troyer. Solving the quantum many-body problem with artificial neural networks. *Science*, 355(6325):602–606, Feb 2017. ISSN 1095-9203. doi: 10.1126/science.aag2302. URL http://dx.doi.org/10.1126/science.aag2302.

[31] Zi Cai and Jinguo Liu. Approximating quantum many-body wave functions using artificial neural networks. *Phys. Rev. B*, 97:035116, Jan 2018. doi: 10.1103/PhysRevB.97.035116. URL https://link.aps.org/doi/10.1103/PhysRevB.97.035116.

[32] Giacomo Torlai, Guglielmo Mazzola, Juan Carrasquilla, Matthias Troyer, Roger Melko, and Giuseppe Carleo. Neural-network quantum state tomography. *Nature Physics*, 14:447–450, 2018. doi: https://doi.org/10.1038/s41567-018-0048-5.

[33] Di Luo and Bryan K. Clark. Backflow transformations via neural networks for quantum many-body wave functions. *Phys. Rev. Lett.*, 122:226401, Jun 2019. doi: 10.1103/PhysRevLett.122.226401. URL https://link.aps.org/doi/10.1103/PhysRevLett.122.226401.

[34] David Pfau, James S. Spencer, Alexander G. de G. Matthews, and W. M. C. Foulkes. Ab-initio solution of the many-electron schrödinger equation with deep neural networks, 2019.

[35] Jan Hermann, Zeno Schätzle, and Frank Noé. Deep neural network solution of the electronic schrödinger equation, 2019.

[36] Kenny Choo, Antonio Mezzacapo, and Giuseppe Carleo. Fermionic neural-network states for ab-initio electronic structure. *Nature Communications*, 11:2368, 2020. doi: https://doi.org/10.1038/s41467-020-15724-9.

[37] Or Sharir, Yoav Levine, Noam Wies, Giuseppe Carleo, and Amnon Shashua. Deep autoregressive models for the efficient variational simulation of many-body quantum systems. *Physical Review Letters*, 124(2), Jan 2020. ISSN 1079-7114. doi: 10.1103/physrevlett.124.020503. URL http://dx.doi.org/10.1103/PhysRevLett.124.020503.

[38] Mohamed Hibat-Allah, Martin Ganahl, Lauren E. Hayward, Roger G. Melko, and Juan Carrasquilla. Recurrent neural network wave functions. *Physical Review Research*, 2 (2), Jun 2020. ISSN 2643-1564. doi: 10.1103/physrevresearch.2.023358. URL http://dx.doi.org/10.1103/PhysRevResearch.2.023358.

[39] Christopher Roth. Iterative retraining of quantum spin models using recurrent neural networks, 2020.

[40] G. E. Santoro. Theory of quantum annealing of an ising spin glass. *Science*, 295(5564): 2427–2430, Mar 2002. ISSN 1095-9203. doi: 10.1126/science.1068774. URL http://dx.doi.org/10.1126/science.1068774.

[41] Roman Martoňák, Giuseppe E. Santoro, and Erio Tosatti. Quantum annealing by the path-integral monte carlo method: The two-dimensional random ising model. *Phys. Rev. B*, 66: 094203, Sep 2002. doi: 10.1103/PhysRevB.66.094203. URL https://link.aps.org/doi/10.1103/PhysRevB.66.094203.

[42] B. Heim, T. F. Ronnow, S. V. Isakov, and M. Troyer. Quantum versus classical annealing of ising spin glasses. *Science*, 348(6231):215–217, Mar 2015. ISSN 1095-9203. doi: 10.1126/science.aaa4170. URL http://dx.doi.org/10.1126/science.aaa4170.

[43] Tadashi Kadowaki and Hidetoshi Nishimori. Quantum annealing in the transverse ising model. *Physical Review E*, 58(5):5355–5363, Nov 1998. ISSN 1095-3787. doi: 10.1103/physreve.58.5355. URL http://dx.doi.org/10.1103/PhysRevE.58.5355.

[44] Edward Farhi, Jeffrey Goldstone, Sam Gutmann, and Michael Sipser. Quantum computation by adiabatic evolution, 2000.

[45] E. Farhi, J. Goldstone, S. Gutmann, J. Lapan, A. Lundgren, and D. Preda. A quantum adiabatic evolution algorithm applied to random instances of an np-complete problem. *Science*, 292 (5516):472–475, Apr 2001. ISSN 1095-9203. doi: 10.1126/science.1057726. URL http://dx.doi.org/10.1126/science.1057726.

[46] Tameem Albash and Daniel A. Lidar. Adiabatic quantum computation. *Rev. Mod. Phys.*, 90: 015002, Jan 2018. doi: 10.1103/RevModPhys.90.015002. URL https://link.aps.org/doi/10.1103/RevModPhys.90.015002.

[47] F. Becca and S. Sorella. *Quantum Monte Carlo Approaches for Correlated Systems*. Cambridge University Press, 2017. doi: 10.1017/9781316417041.

[48] J. Brooke, D. Bitko, T. F. Rosenbaum, and G. Aeppli. Quantum annealing of a disordered magnet. *Science*, 284(5415):779–781, 1999. ISSN 0036-8075. doi: 10.1126/science.284.5415.779. URL https://science.sciencemag.org/content/284/5415/779.

[49] F Barahona. On the computational complexity of ising spin glass models. *Journal of Physics A: Mathematical and General*, 15(10):3241–3253, oct 1982. doi: 10.1088/0305-4470/15/10/028. URL https://doi.org/10.1088%2F0305-4470%2F15%2F10%2F028.

[50] Spin glass server. URL https://software.cs.uni-koeln.de/spinglass/.

[51] Satoshi Morita and Hidetoshi Nishimori. Mathematical foundation of quantum annealing. *Journal of Mathematical Physics*, 49(12):125210, Dec 2008. ISSN 1089-7658. doi: 10.1063/1.2995837. URL http://dx.doi.org/10.1063/1.2995837.

[52] Gary S. Grest, C. M. Soukoulis, and K. Levin. Cooling-rate dependence for the spin-glass ground-state energy: Implications for optimization by simulated annealing. *Phys. Rev. Lett.*, 56: 1148–1151, Mar 1986. doi: 10.1103/PhysRevLett.56.1148. URL https://link.aps.org/doi/10.1103/PhysRevLett.56.1148.

[53] Pam Ocampo-Alfaro and Hong Guo. Cooling-rate dependence of the ground-state energy using microcanonical simulated annealing. *Phys. Rev. E*, 53:1982–1985, Feb 1996. doi: 10.1103/PhysRevE.53.1982. URL https://link.aps.org/doi/10.1103/PhysRevE.53.1982.

[54] Joseph Gomes, Keri A. McKiernan, Peter Eastman, and Vijay S. Pande. Classical quantum optimization with neural network quantum states, 2019.

[55] Semyon Sinchenko and Dmitry Bazhanov. The deep learning and statistical physics applications to the problems of combinatorial optimization, 2019.

[56] Tianchen Zhao, Giuseppe Carleo, James Stokes, and Shravan Veerapaneni. Natural evolution strategies and quantum approximate optimization, 2020.

[57] Mohamed Hibat-Allah, Estelle Inack, Roeland Wiersema, Roger Melko, and Juan Carrasquilla. Variational neural annealing. In preparation.

[58] J. D. Hunter. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9 (3):90–95, 2007. doi: 10.1109/MCSE.2007.55.

[59] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. URL https://www.tensorflow.org/. Software available from tensorflow.org.

[60] Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. Array programming with numpy. *Nature*, 585(7825):357–362, Sep 2020. ISSN 1476-4687. doi: 10.1038/s41586-020-2649-2. URL https://doi.org/10.1038/s41586-020-2649-2.