Fermionic neural-network states for ab-initio electronic structure

Kenny Choo,[1] Antonio Mezzacapo,[2] and Giuseppe Carleo[3, *]

[1]*Department of Physics, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland*[†]
[2]*IBM T.J. Watson Research Center, Yorktown Heights, NY, USA*[‡]
[3]*Center for Computational Quantum Physics, Flatiron Institute, 162 5th Avenue, New York, NY 10010, USA*

Neural-network quantum states have been successfully used to study a variety of lattice and continuous-space problems. Despite a great deal of general methodological developments, representing fermionic matter is however still early research activity. Here we present an extension of neural-network quantum states to model interacting fermionic problems. Borrowing techniques from quantum simulation, we directly map fermionic degrees of freedom to spin ones, and then use neural-network quantum states to perform electronic structure calculations. For several diatomic molecules in a minimal basis set, we benchmark our approach against widely used coupled cluster methods, as well as many-body variational states. On the test molecules, we recover almost the entirety of the correlation energy. We systematically improve upon coupled cluster methods and Jastrow wave functions, reaching levels of chemical accuracy or better. Finally, we discuss routes for future developments and improvements of the methods presented.

arXiv:1909.12852v1 [physics.comp-ph] 27 Sep 2019

Introduction.- Predicting the physical and chemical properties of matter from the fundamental principles of quantum mechanics is a central problem in modern electronic structure theory. In the context of ab-initio Quantum Chemistry (QC), a commonly adopted strategy to solve for the electronic wave-function is to discretize the problem on finite basis functions, expanding the full many-body state in a basis of anti-symmetric Slater determinants. Because of the factorial scaling of the determinant space, exact approaches systematically considering all electronic configurations, such as the full configuration interaction (FCI) method, are typically restricted to small molecules and basis sets. A solution routinely adopted in the field is to consider systematic corrections over mean-field states. For example, in the framework of the coupled cluster (CC) method [1, 2], higher level of accuracy can be obtained considering electronic excitations up to doublets, in CCSD, and triplets in CCSD(T). CC techniques are routinely adopted in QC electronic calculations, and they are often considered the "gold standard" in ab-initio electronic structure. Despite this success, the accuracy of CC is intrinsically limited in the presence of strong quantum correlations, in turn restricting the applicability of the method to regimes of relative weak correlations.

For strongly correlated molecules and materials, alternative, non-perturbative approaches have been introduced. Most notably, both stochastic and non-stochastic methods based on variational representations of many-body wave-functions have been developed and constantly improved in the past decades of research. Notable variational classes for QC are Jastrow-Slater wave-functions [3], correlated geminal wave-functions [4], and matrix product states [5–7]. Stochastic projection methods systematically improving upon variational starting points are for example the fixed-node Green's function Monte Carlo [8] and constrained-path auxiliary field Monte Carlo [9]. Main limitations of these methods stem, directly or indirectly, from the choice of the variational form. For example, matrix-product states are extremely efficient in quasi one-dimensional systems, but suffer from exponential scaling when applied to larger dimensions. On the other hand, variational forms considered so-far for higher dimensional systems typically rely on rigid variational classes and do not provide a systematic and computationally efficient way to increase their expressive power.

To help overcome some of the limitations of existing variational representations, ideas leveraging the power of artificial neural networks (ANN) have recently emerged in the more general context of interacting many-body quantum matter. These approaches are typically based on compact, variational parameterizations of the many-body wave-function in terms of ANN [10]. These approaches to fermionic problems are however comparatively less explored than for lattice spin systems. Two conceptually different implementations have been put forward. In the first, fermionic symmetry is encoded directly at the mean field level, and ANNs are used as a positive-definite correlator function [11]. Main limitation of this ansatz is that the nodal structure of the wave function is fixed, and the exact ground state cannot, in principle, be achieved, even in the limit of infinitely large ANN. The second method is to use ANNs to parametrize permutation symmetric many-body fermionic orbitals [12, 13], in the spirit of "backflow" many-body variational wave functions [14, 15], and only very recently applied to electronic structure [16, 17].

In this Article we provide an alternative representation of fermionic many-body quantum systems based on a direct encoding of electronic configurations. This task is achieved by mapping the fermionic problem onto an equivalent spin problem, and then solving the latter with spin-based neural-network quantum states. Using techniques from quantum information, we analyze different

---

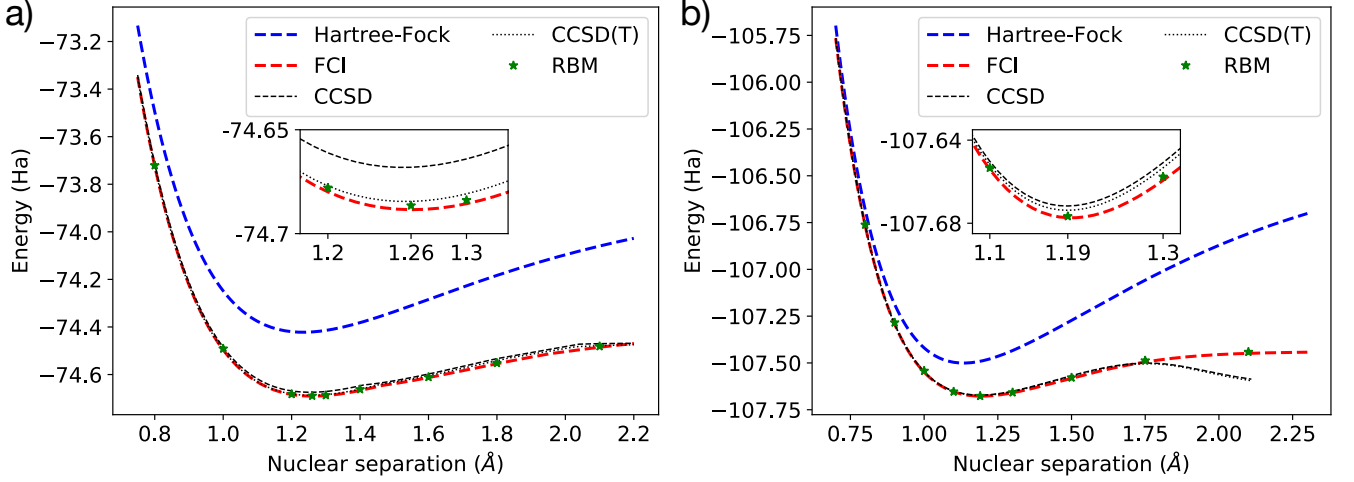* gcarleo@flatironinstitute.org
† kenny.choo@uzh.ch
‡ amezzac@us.ibm.com

FIG. 1. The accuracy of fermionic neural-network quantum states compared with other quantum chemistry approaches. Shown here are dissociation curves for a) C_2 and b) N_2, in the STO-3G basis with 20 spin-orbitals. The RBM used has 40 hidden units, and it is compared to both coupled-cluster approaches (CCSD, CCSD(T)) and exact FCI energies.

model agnostic fermion-to-spin mappings. We show results for several diatomic molecules in minimal Gaussian basis sets, where our approach reaches chemical accuracy (< 5kcal/mol) or better. The current challenges in extending the method to larger basis sets and molecules are also discussed.

Electronic structure on spin systems.- We consider many-body molecular fermionic Hamiltonians in second quantization formalism,

$$H = \sum_{i,j} t_{ij}\, c_i^\dagger c_j + \sum_{i,j,k,m} u_{ijkm}\, c_i^\dagger c_k^\dagger c_m c_j, \qquad (1)$$

where we have defined fermionic annihilation and creation operators with the anticommutation relation $\{c_i^\dagger, c_j\} = \delta_{i,j}$ on N fermionic modes, and one- and two-body integrals t_{ij} and u_{ijkm}. The Hamiltonian (1) can be mapped to interacting spin models via the Jordan-Wigner [18] mapping, or the more recent parity or Bravyi-Kitaev [19] encodings, which have been developed in the context of quantum simulations. These three encodings can all be expressed in the compact form

$$c_j \to \frac{1}{2} \prod_{i \in U(j)} \sigma_i^x \times \left(\sigma_j^x \prod_{i \in P(j)} \sigma_i^z - i\sigma_j^y \prod_{i \in R(j)} \sigma_i^z \right)$$

$$c_j^\dagger \to \frac{1}{2} \prod_{i \in U(j)} \sigma_i^x \times \left(\sigma_j^x \prod_{i \in P(j)} \sigma_i^z + i\sigma_j^y \prod_{i \in R(j)} \sigma_i^z \right), \qquad (2)$$

where we have defined an update $U(j)$, parity $P(j)$ and remainder $R(j)$ sets of spins, which depend on the particular mapping considered [20, 21], and $\sigma_i^{(x,y,z)}$ denote Pauli matrices acting on site i. In the familiar case of the Jordan-Wigner transformation, the update, parity and remainder sets become $U(j) = j$, $P(j) = \{0, 1, ...j-1\}$,

$R(j) = P(j)$, and the mapping takes the simple form

$$c_j \to \left(\prod_{i=0}^{j-1} \sigma_i^z \right) \sigma_j^-$$

$$c_j^\dagger \to \left(\prod_{i=0}^{j-1} \sigma_i^z \right) \sigma_j^+, \qquad (3)$$

where $\sigma_j^{+(-)} = (\sigma_j^x + (-)i\sigma_j^y)/2$. For all the spin encodings considered, the final outcome is a spin Hamiltonian with the general form

$$H_q = \sum_{j=1}^{r} h_j \boldsymbol{\sigma}_j, \qquad (4)$$

defined as a linear combination with real coefficients h_j of $\boldsymbol{\sigma}_j$, N-fold tensor products of single-qubit Pauli operators $I, \sigma^x, \sigma^y, \sigma^z$. Additionally, under such mappings, there is a one to one correspondence between spin configuration $\vec{\sigma}$ and the original particle occupations $\vec{n}_\sigma$. In the following, we will consider the interacting spin Hamiltonian (4) as a starting point for our variational treatment.

Neural-network quantum states.- Once the mapping is performed, we use neural-network quantum states (NQS) introduced in [10] to parametrize the ground state of the Hamiltonian (4). One conceptual interest of NQS is that, because of the flexibility of the underlying nonlinear parameterization, they can be adopted to study both equilibrium [22, 23] and out-of-equilibrium [24–29] properties of diverse many-body quantum systems. In this work we adopt a simple neural-network parameterization in terms of a complex-valued, shallow restricted Boltzmann machine (RBM) [10, 30]. For a system of N spins, the many-body amplitudes take the compact form

$$\Psi_M(\vec{\sigma}; \mathcal{W}) = e^{\sum_i a_i \sigma_i^z} \prod_{j=1}^{M} 2\cosh\theta_j(\vec{\sigma}), \text{ where} \quad (5)$$

$$\theta_j(\vec{\sigma}) = b_j + \sum_i^N W_{ij}\sigma_i^z. \quad (6)$$

Here, $\mathcal{W}$ are complex-valued network parameters $\mathcal{W} = \{a, b, W\}$, and the expressivity of the network is determined by the hidden unit density defined by $\alpha = M/N$ where M is number of hidden units. The simple RBM ansatz can efficiently support volume-law entanglement [31–34], and it has been recently used in several applications [35].

One can then train the ansatz Eq.(5) with a variational learning approach known as Variational Monte Carlo (VMC), by minimizing the energy expectation value

$$E(\mathcal{W}) = \frac{\langle \Psi_M | H_q | \Psi_M \rangle}{\langle \Psi_M | \Psi_M \rangle}. \quad (7)$$

This expectation value can be evaluated using Monte Carlo sampling using the fact that the energy (and, analogously, any other observable) can be written as

$$E(\mathcal{W}) = \frac{\sum_{\vec{\sigma}} E_{\text{loc}}(\vec{\sigma}) |\Psi_M(\vec{\sigma})|^2}{\sum_{\vec{\sigma}} |\Psi_M(\vec{\sigma})|^2}, \quad (8)$$

where we have defined the local energy

$$E_{\text{loc}}(\vec{\sigma}) = \sum_{\vec{\sigma}'} \frac{\Psi_M(\vec{\sigma}')}{\Psi_M^*(\vec{\sigma})} \langle \vec{\sigma}' | H_q | \vec{\sigma} \rangle. \quad (9)$$

Given samples $\mathcal{M}$ drawn from the distribution $\frac{|\Psi_M(\vec{\sigma})|^2}{\sum_{\vec{\sigma}} |\Psi_M(\vec{\sigma})|^2}$, the average over the samples $\hat{E}(\mathcal{W}) = \langle E_{\text{loc}}(\vec{\sigma}) \rangle_{\mathcal{M}}$ gives an unbiased estimator of the energy. Note that the computational cost of evaluating the local energy depends largely on the sparsity of the Hamiltonian H_q. In a generic QC problems, this cost scales in the worst case with $\mathcal{O}(N^4)$, as compared to the linear scaling in typical condensed matter systems with local interaction.

Sampling from $|\Psi_M(\vec{\sigma})|^2$ is performed using Markov chain Monte Carlo (MCMC), with a Markov chain $\vec{\sigma}_0 \to \vec{\sigma}_1 \to \vec{\sigma}_2 \to \ldots$ constructed using the Metropolis-Hastings algorithm [36]. Specifically, at each iteration, a configuration $\vec{\sigma}_{\text{prop}}$ is proposed and accepted with probability

$$P(\vec{\sigma}_{k+1} = \vec{\sigma}_{\text{prop}}) = \min\left(1, \left|\frac{\Psi_M(\vec{\sigma}_{\text{prop}})}{\Psi_M(\vec{\sigma}_k)}\right|^2\right). \quad (10)$$

The sample $\mathcal{M}$ then corresponds to the configurations of the Markov chain downsampled at an interval K, i.e. $\{\vec{\sigma}_0, \vec{\sigma}_K, \vec{\sigma}_{2K}, \ldots\}$. For the simulations done in this work, we typically use $K = 10N$ with sample size of approximately 100000.

Molecule	RBM	Jastrow	CCSD	CCSD(T)	FCI
H_2	-1.1373	-1.1373	-1.1373	-1.1373	-1.1373
LiH	-7.8826	-7.8814	-7.8828	-7.8828	-7.8828
NH_3	-55.5277	-55.4770	-55.5280	-55.5281	-55.5282
H_2O	-75.0232	-74.9784	-75.0231	-75.0232	-75.0233
C_2	-74.6892	-74.5001	-74.6745	-74.6876	-74.6908
N_2	-107.6767	-107.5924	-107.6717	-107.6738	-107.6774

TABLE I. Equilibrium energies (in Hartree) as obtained by different methods. The basis set considered here is STO-3G, and the corresponding geometries are reported in Appendix A. Energies are reported in Hartrees and statistical uncertainty on RBM and Jastrow states energies are on the last reported digits. The RBM used has a hidden unit density $\alpha = 1$ for all the molecules apart from C_2 and N_2 where we use $\alpha = 2$.

Since the Hamiltonians we are interested in have an underlying particle conservation law, it is helpful to perform this sampling in the particle basis $\vec{n}_\sigma$ rather than the corresponding spin basis $\vec{\sigma}$. The proposed configuration $\vec{\sigma}_{\text{prop}}$ at each iteration, then corresponds to a particle hopping between orbitals. Once a stochastic estimate of the expectation values is available, as well as its derivatives w.r.t. the parameters $\mathcal{W}$, the ansatz can be optimized using the stochastic reconfiguration method [37, 38], closely related to the natural-gradient method used in machine learning applications [10, 39].

Potential Energy surfaces.- We first consider small molecules in a minimal basis set (STO-3G). We show in Fig. 1 the dissociation curves for C_2 and N_2, compared to the CCSD and CCSD(T). It can be seen that on these small molecules in their minimal basis, the RBM is able to generate accurate representations of the ground
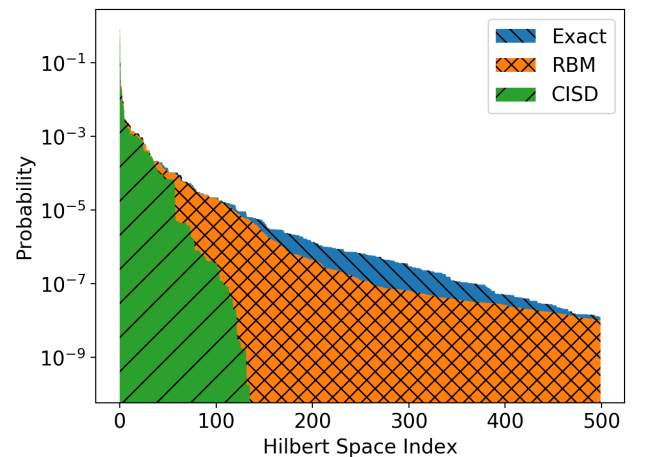

FIG. 2. Probabilities (in logarithmic scale) of the 500 most probable configurations in the exact (red), RBM (green) and CISD (blue) wavefunctions for the equilibrium nitrogen N_2 molecule in the STO-3G basis.

states, and remarkably achieve an accuracy better than standard QC methods. To further illustrate the expressiveness of the RBM, we show in Fig. 2 the probability distribution of the most relevant configurations in the wavefunction. We contrast between the RBM and configuration interaction limited to single and double excitations (CISD). In CISD, the Hilbert space is truncated to include only states which are up to two excitations away from the Hartree-Fock configuration. It is clear from the histogram that the RBM is able to capture correlations beyond double excitations.

Alternative encodings.- The above computations were done using the Jordan-Wigner mapping. To investigate the effect of the mapping choice on the performance of the RBM, we also performed select calculations using the parity and Bravyi-Kitaev mappings. All the aforementioned transformations require a number of spins equal to the number of fermionic modes in the model. However, the support of the Pauli operators $w_j = |\boldsymbol{\sigma}_j|$ in (4), i.e. the number of single-qubit Pauli operators in $\boldsymbol{\sigma}_j$ that are different from the identity I, depends on the specific mapping used. Jordan-Wigner and parity mappings have linear scalings $w_j = O(N)$, while the Bravyi-Kitaev encoding has a more favorable scaling $w_j = O(\log(N))$, due to the logarithmic spin support of the update, parity and remainder sets in (2). Note that one could in principle use generalized superfast mappings [40], which have a support scaling as good as $w_j = O(\log(d))$, where d is the maximum degree of the fermionic interaction graph defined by (1). However, such a mapping is not practical for the models considered here because the typical large degree of molecular interactions graphs makes the number of spins required for the simulation too large compared to the other model-agnostic mappings.

While these encodings are routinely used as tools to study fermionic problems on quantum hardware [41], their use in classical computing has not been systematically explored so far. Since they yield different structured many-body wave functions, it is then worth analyzing whether more local mappings can be beneficial for specific NQS representations. In Fig. 3, we analyze the effect of the different encodings on the accuracy of the variational ground-state energy for a few representative diatomic molecules. At fixed computational resources and network expressivity, we typically find that the RBM ansatz can achieve consistent levels of accuracy, independent of the nature of the mapping type. While the Jordan Wigner allows to achieve the lowest energies in those examples, the RBM is nonetheless able to efficiently learn the ground state also in other representations, and chemical accuracy is achieved in all cases reported in Fig. 3.

Sampling larger basis sets.- The spin-based simulations of the QC problems studied here show a distinctive MCMC sampling behavior that is not usually found in lattice model simulations of pure spin models. Specifically, the ground-state wave function of the diatomic
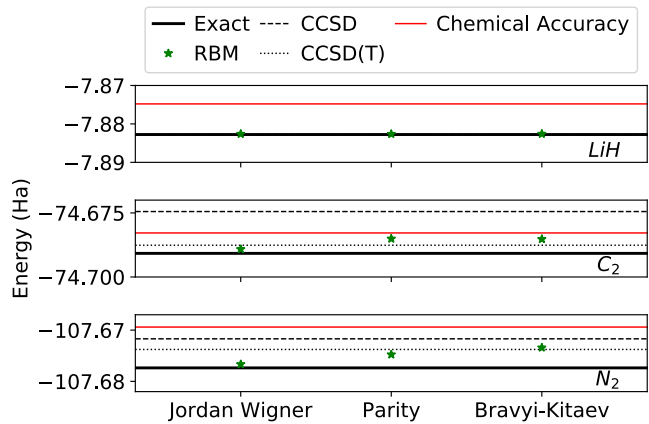

FIG. 3. Accuracy of the RBM (green star) representations for three different mapping types (Jordan-Wigner, Parity and Bravyi-Kitaev) and three different molecules (LiH, C_2 and N_2) in their equilibrium configuration in the STO-3G basis. The geometries used are reported in Appendix A.

molecules considered is typically sharply peaked around the Hartree-Fock state, and neighboring excited states. This behavior is prominently shown also in Fig. 2, where the largest peaks are several of order of magnitude larger than the distribution tail. As a result of this structure, any uniform sampling scheme drawing states $\vec{\sigma}$ from the VMC distribution $|\Psi_M(\vec{\sigma})|^2$, is bound to repeatedly draw the most dominant states, while only rarely sampling less likely configurations. To exemplify this peculiarity, we study the behavior of the ground state energy as a function of the number of MCMC samples used at each step of the VMC optimization. We concentrate on the water molecule in the larger 6-31g basis. In this case, the Metropolis sampling scheme exhibits acceptance rates as low as 0.1% or less, as a consequence of the presence of dominating states previously discussed.

In Fig. 4, we vary the sample size and also compare MCMC sampling with exact sampling. We can see that the accuracy of the simulation depends quite significantly on the sample size, and that chemical accuracy is reached only for a relatively large number of samples. The large number of samples needed in this case, together with a very low acceptance probability for the Metropolis Hasting algorithm, directly points to the inefficiency of uniform sampling from $|\Psi_M(\vec{\sigma})|^2$. At present, this represents the most significant bottleneck in the application of our approach to larger molecules and basis sets. This issue however is not a fundamental limitation, and alternatives to the standard VMC uniform sampling can be envisioned to efficiently sample less likely–yet important for chemical accuracy– states.

Outlook.- In this work we have shown that relatively simple shallow neural networks can be used to compactly encode, with high precision, the electronic wave function of model molecular problems in quantum chemistry. Our approach is based on the mapping between the fermionic
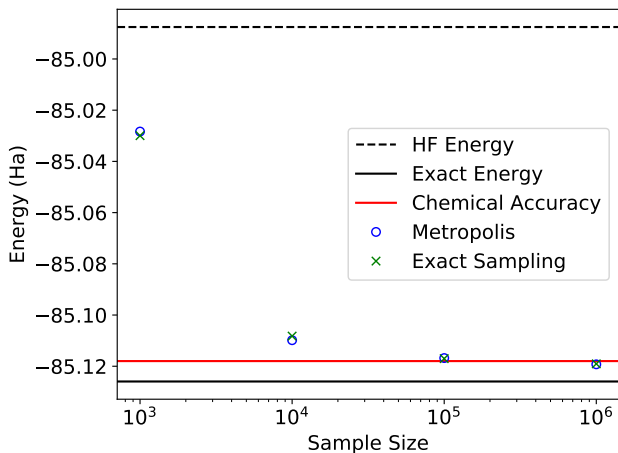
FIG. 4. Converged electronic energy of H_2O in the 6-31g basis (26 spin-orbitals) as the sample size used for each VMC iteration is varied. The converged energy for the sample obtained using the Metropolis algorithm (green cross) matches that obtain via exact sampling (blue circle), reaching chemical accuracy (red line) for the largest sample size.

quantum chemistry molecular Hamiltonian and corresponding spin Hamiltonians. In turn, the ground state of the spin models can be conveniently modeled with standard variational neural-network quantum states. On model diatomic molecules, we show that a RBM state is able to capture almost the entirety of the electronic excitations, improving on routinely used approaches as CCSD(T) and the Jastrow ansatz.

Several future directions can be envisioned. The distinctive peaked structure of the molecular wave function calls for development of alternatives to uniform sampling from the Born probability. These developments will allow to efficiently handle larger basis sets than the ones considered here. Second, our study has explored only a very limited subset of possible neural-network architectures. Most notably, the use of deeper networks might prove beneficial for complex molecular complexes. Another very interesting matter for future research is the comparison of different neural-network based approaches to quantum chemistry. Contemporary to this work, approaches based on antisymmetric wave-functions in continuous space have been presented [16, 17]. These have the advantage that they already feature a full basis set limit. However, the discrete basis approach has the advantage that boundary conditions and fermionic symmetry are much more easily enforced. As a consequence, simple-minded shallow networks can already achieve comparatively higher accuracy than the deeper and substantially more complex networks so-far adopted in the continuum case.

[1] F. Coester and H. Kümmel, Nuclear Physics **17**, 477 (1960).

[2] J. Čížek, *The Journal of Chemical Physics*, The Journal of Chemical Physics **45**, 4256 (1966).

[3] R. Jastrow, Physical Review **98**, 1479 (1955).

[4] M. Casula and S. Sorella, The Journal of Chemical Physics **119**, 6500 (2003).

[5] S. R. White, Physical Review Letters **69**, 2863 (1992).

[6] S. R. White and R. L. Martin, The Journal of Chemical Physics **110**, 4127 (1999).

[7] G. K.-L. Chan and S. Sharma, Annual Review of Physical Chemistry **62**, 465 (2011).

[8] J. B. Anderson, The Journal of Chemical Physics **63**, 1499 (1975).

[9] S. Zhang and H. Krakauer, Physical Review Letters **90**, 136401 (2003).

[10] G. Carleo and M. Troyer, Science **355**, 602 (2017).

[11] Y. Nomura, A. S. Darmawan, Y. Yamaji, and M. Imada, Physical Review B **96**, 205152 (2017).

[12] M. Ruggeri, S. Moroni, and M. Holzmann, Physical Review Letters **120**, 205302 (2018).

[13] D. Luo and B. K. Clark, Physical Review Letters **122**, 226401 (2019).

[14] R. P. Feynman and M. Cohen, Physical Review **102**, 1189 (1956).

[15] L. F. Tocchio, F. Becca, A. Parola, and S. Sorella, Physical Review B **78**, 041101 (2008).

[16] D. Pfau, J. S. Spencer, A. G. d. G. Matthews, and W. M. C. Foulkes, arXiv:1909.02487 [physics] (2019), arXiv: 1909.02487.

[17] J. Hermann, Z. Schätzle, and F. Noé, arXiv:1909.08423 [physics, stat] (2019), arXiv: 1909.08423.

[18] E. Wigner and P. Jordan, Z. Phys **47**, 631 (1928).

[19] S. Bravyi and A. Kitaev, Ann. of Phys. **298**, 210 (2002).

[20] J. Seeley, M. Richard, and P. Love, The Journal of Chemical Physics **137**, 224109 (2012).

[21] A. Tranter, S. Sofia, J. Seeley, M. Kaicher, J. McClean, R. Babbush, P. Coveney, F. Mintert, F. Wilhelm, and P. Love, International Journal of Quantum Chemistry **115**, 1431 (2015).

[22] K. Choo, G. Carleo, N. Regnault, and T. Neupert, Physical Review Letters **121**, 167204 (2018).

[23] F. Ferrari, F. Becca, and J. Carrasquilla, arXiv:1906.00463 [cond-mat] (2019), arXiv: 1906.00463.

[24] S. Czischek, M. Gärttner, and T. Gasenzer, Physical Review B **98**, 024311 (2018).

[25] G. Fabiani and J. Mentink, SciPost Physics **7**, 004 (2019).

[26] M. J. Hartmann and G. Carleo, Physical Review Letters **122**, 250502 (2019).

[27] A. Nagy and V. Savona, Physical Review Letters **122**, 250501 (2019).

[28] F. Vicentini, A. Biella, N. Regnault, and C. Ciuti, Physical Review Letters **122**, 250503 (2019).

[29] N. Yoshioka and R. Hamazaki, Physical Review B **99**, 214306 (2019).

[30] P. Smolensky (MIT Press, Cambridge, MA, USA, 1986) Chap. Information Processing in Dynamical Systems: Foundations of Harmony Theory, pp. 194–281.

[31] D.-L. Deng, X. Li, and S. Das Sarma, Physical Review X **7**, 021021 (2017).

[32] Y. Huang and J. E. Moore, arXiv:1701.06246 [cond-mat] (2017), arXiv: 1701.06246.

[33] J. Chen, S. Cheng, H. Xie, L. Wang, and T. Xiang, Physical Review B **97**, 085104 (2018).

[34] Y. Levine, O. Sharir, N. Cohen, and A. Shashua, Physical Review Letters **122**, 065301 (2019).

[35] R. G. Melko, G. Carleo, J. Carrasquilla, and J. I. Cirac, Nature Physics **15**, 887 (2019).

[36] W. K. Hastings, Biometrika **57**, 97 (1970).

[37] S. Sorella, Physical Review Letters **80**, 4558 (1998).

[38] S. Sorella, M. Casula, and D. Rocca, The Journal of Chemical Physics **127**, 014105 (2007).

[39] S.-i. Amari, Neural Computation **10**, 251 (1998).

[40] K. Setia, S. Bravyi, A. Mezzacapo, and J. D. Whitfield, arXiv preprint arXiv:1810.05274 (2018).

[41] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, Nature **549**, 242 (2017).

[42] G. Carleo, K. Choo, D. Hofmann, J. E. T. Smith, T. Westerhout, F. Alet, E. J. Davis, S. Efthymiou, I. Glasser, S.-H. Lin, M. Mauri, G. Mazzola, C. B. Mendl, E. van Nieuwenburg, O. O'Reilly, H. Théveniaut, G. Torlai, F. Vicentini, and A. Wietek, SoftwareX **10**, 100311 (2019).

[43] Q. Sun, T. C. Berkelbach, N. S. Blunt, G. H. Booth, S. Guo, Z. Li, J. Liu, J. D. McClain, E. R. Sayfutyarova, S. Sharma, S. Wouters, and G. K. Chan, "Pyscf: the python-based simulations of chemistry framework," (2017), https://onlinelibrary.wiley.com/doi/pdf/10.1002/wcms.1340.

[44] H. Abraham, I. Y. Akhalwaya, G. Aleksandrowicz, T. Alexander, G. Alexandrowics, E. Arbel, A. Asfaw, C. Azaustre, P. Barkoutsos, G. Barron, L. Bello, Y. Ben-Haim, L. S. Bishop, S. Bosch, D. Bucher, *et al.*, "Qiskit: An open-source framework for quantum computing," (2019).

[45] R. Johnson, NIST Standard Reference Database.

Appendix A: Geometries for diatomic molecules

The equilibrium geometries for the molecules presented in this work were obtained from the CCCBDB database [45]. For convenience, we present them in Table II.

Molecule	Basis	Geometry
H_2	STO-3G	H(0,0,0)
		H(0,0,0.734)
LiH	STO-3G	Li(0,0,0)
		H(0,0,1.548)
NH$_3$	STO-3G	N(0,0,0.149)
		H(0,0.947,-0.348)
		H(0.821,-0.474,-0.348)
		H(-0.821,-0.474,-0.348)
C$_2$	STO-3G	C(0,0,0)
		C(0,0,1.26)
N$_2$	STO-3G	N(0,0,0)
	.	N(0,0,1.19)
H$_2$O	STO-3G	H(0,0.769,-0.546)
		H(0,-0.769,-0.546)
		O(0,0,0.137)
H$_2$O	6-31G	H(0,0.795,-0.454)
		H(0,-0.795,-0.454)
		O(0,0,0.113)

TABLE II. Equilibrium configurations used for the ground-state calculations presented in the main text. The coordinates (x, y, z) are given in angstroms(Å)

Appendix B: Computing matrix elements

A crucial requirement for the efficient implementation of the stochastic variational Monte Carlo procedure to minimize the ground-state energy, is the ability to efficiently compute the matrix elements of the spin Hamiltonian $\langle \vec{\sigma}' | H_q | \vec{\sigma} \rangle$, appearing in the local energy, Eq. 9. Since H_q is a sum of products of Pauli operators, the goal is to efficiently compute matrix elements of the form

$$\mathcal{M}(\vec{\sigma}, \vec{\sigma}') = \langle \vec{\sigma}' | \sigma_1^{\nu_1} \sigma_2^{\nu_2} \ldots \sigma_N^{\nu_N} | \vec{\sigma} \rangle, \qquad \text{(B1)}$$

where $\sigma_i^{\nu_i}$ denotes a Pauli matrix with $\nu = I, x, y, z$ acting on site i. Because of the structure of the Pauli operators, these matrix elements are non-zero only for a specific $\vec{\sigma}'$ such that

$$\begin{cases} \sigma_i' = \sigma_i & \nu_i \in (I, Z) \\ \sigma_i' = -\sigma_i & \nu_i \in (X, Y) \end{cases} \qquad \text{(B2)}$$

and the matrix element is readily computed as

$$\mathcal{M}(\vec{\sigma}, \vec{\sigma}') = (i^{n_y}) \Pi_{k:v_k \in (y,z)} \sigma_k, \qquad \text{(B3)}$$

where n_y is the total number of σ^y operators in the string of Pauli matrices.