



*mathematics*



Article

---

# A Real Neural Network State for Quantum Chemistry

---

Yangjun Wu, Xiansong Xu, Dario Poletti, Yi Fan, Chu Guo and Honghui Shang

Special Issue

Artificial Intelligence and Scientific Computing: Mathematical Techniques and Applications

Edited by

Dr. Jue Wang, Dr. Jidong Zhai and Dr. Weile Jia



<https://doi.org/10.3390/math11061417>

Article

# A Real Neural Network State for Quantum Chemistry

Yangjun Wu <sup>1</sup>, Xiansong Xu <sup>2,3</sup>, Dario Poletti <sup>2,4,5</sup> , Yi Fan <sup>6</sup>, Chu Guo <sup>7,8,\*</sup> and Honghui Shang <sup>1,\*</sup>

<sup>1</sup> Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China

<sup>2</sup> Science, Mathematics and Technology Cluster, Singapore University of Technology and Design, 8 Somapah Road, Singapore 487372, Singapore

<sup>3</sup> College of Physics and Electronic Engineering, and Center for Computational Sciences, Sichuan Normal University, Chengdu 610068, China

<sup>4</sup> EPD Pillar, Singapore University of Technology and Design, 8 Somapah Road, Singapore 487372, Singapore

<sup>5</sup> MajuLab, CNRS-UNS-NUS-NTU International Joint Research Unit, Singapore UMI 3654, Singapore

<sup>6</sup> Hefei National Research Center for Physical Sciences at the Microscale, University of Science and Technology of China, Hefei 230026, China

<sup>7</sup> Henan Key Laboratory of Quantum Information and Cryptography, Zhengzhou 450000, China

<sup>8</sup> Key Laboratory of Low-Dimensional Quantum Structures and Quantum Control of Ministry of Education, Department of Physics and Synergetic Innovation Center for Quantum Effects and Applications, Hunan Normal University, Changsha 410081, China

\* Correspondence: guochu604b@gmail.com (C.G.); shanghonghui@ict.ac.cn (H.S.)

**Abstract:** The restricted Boltzmann machine (RBM) has recently been demonstrated as a useful tool to solve the quantum many-body problems. In this work we propose tanh-FCN, which is a single-layer fully connected neural network adapted from RBM, to study ab initio quantum chemistry problems. Our contribution is two-fold: (1) our neural network only uses real numbers to represent the real electronic wave function, while we obtain comparable precision to RBM for various prototypical molecules; (2) we show that the knowledge of the Hartree-Fock reference state can be used to systematically accelerate the convergence of the variational Monte Carlo algorithm as well as to increase the precision of the final energy.

**Keywords:** neural network; variational Monte Carlo; quantum chemistry

**MSC:** 68T07



**Citation:** Wu, Y.; Xu, X.; Poletti, D.; Fan, Y.; Guo, C.; Shang, H. A Real Neural Network State for Quantum Chemistry. *Mathematics* **2023**, *11*, 1417. <https://doi.org/10.3390/math11061417>

Academic Editors: Jidong Zhai, Jue Wang and Weile Jia

Received: 10 January 2023

Revised: 27 February 2023

Accepted: 9 March 2023

Published: 15 March 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Ab initio electronic structure calculations based on quantum-chemical approaches (Hartree-Fock theory and post-Hartree-Fock methods) have been successfully applied in molecular systems [1]. For strongly correlated many-electron systems, the exponentially growing Hilbert space size limits the application scale of most numerical algorithms. For example, the full configuration interaction (FCI), which takes the whole Hilbert space into account, is currently limited within around 24 orbitals and 24 electrons [2]. The density matrix renormalization group (DMRG) algorithm [3,4] has been used to solve larger chemical systems of several tens of electrons [5,6]; however, it is essentially limited by the expressive power of its underlying variational ansatz: the matrix product state (MPS), which is a special instance of the one-dimensional tensor network state [7]. Therefore, DMRG could also be extremely difficult to use to approach even larger systems. The coupled cluster (CC) [8,9] method expresses the exact wave function in terms of an exponential form of a variational wave function ansatz, and a higher level of accuracy can be obtained by considering electronic excitations up to doublets in CCSD or triplets in CCSD(T). In practice, it is often accurate with a durable computational cost and is thus considered as the “gold standard” in electronic structure calculations. However, the accuracy of the CC method is only restricted in studying weakly correlated systems [10]. The multi-configuration self-consistent field (MCSCF) [11–13] method is crucial for describing molecular systems

containing nearly degenerate orbitals. It introduces a small number of (active) orbitals; then, the configuration interaction coefficients and the orbital coefficients are optimized to minimize the total energy of the MCSCF state. It has been applied to systems with around 50 active orbitals [14], but they are still limited by the exponential complexity that grows with the system size.

In recent years, the variational Monte Carlo (VMC) method in combination with a neural network ansatz for the underlying quantum state (wave function) [15], referred to as the neural network quantum states (NNQS), has been demonstrated to be a scalable and accurate tool for many-spin systems [16–18] and many-fermion systems [19]. NNQS allow very flexible choices of the neural network ansatz, and with an appropriate variational ansatz, it could often achieve comparable or higher accuracy compared to existing methods. NNQS has also been applied to solve ab-initio quantum chemistry systems in real space with up to 30 electrons [20–22], as well as in discrete basis after second quantization [23–25]. Up to now, various neural networks have been used, such as the restricted Boltzmann machine (RBM) [15], the convolutional neural network [16], recurrent neural networks [26] and the variational auto-encoder [25]. In all of those neural networks, the RBM is a very special instance in that: (1) it has a very simple structure that contains only a fully connected dense layer plus a nonlinear activation; (2) with such a simple structure, RBM can be more expressive than MPS [27]; in fact, it is equivalent to certain two-dimensional tensor network states [28] and can even represent certain quantum states with volume-law entanglement [29]. In practice, RBM achieves comparable accuracy to other more sophisticated neural networks for complicated applications such as frustrated many-spin systems [30,31].

For the ground state of molecular systems, the wave function is real. However, if one uses a real RBM as the variational ansatz for the wave function, then all of the amplitudes of the wave function will be positive, which means that it may be good for ferromagnetic states but will be completely wrong for anti-ferromagnetic states. Therefore, even for real wave functions one would have to use complex RBMs or two RBMs [32] in general. In this work, we propose a neural network with real numbers that is slightly modified from the RBM, such that its output can be both positive and negative, and use it as the neural network ansatz to solve quantum chemistry problems. To accelerate convergence of the VMC iterations, we explicitly use the Hartree–Fock reference state as the starting point for the Monte Carlo sampling after a number of VMC iterations such that the wave function ansatz has become sufficiently close to the ground state. We show that this technique can generally improve the convergence and the precision of the final result, even when using other neural networks. Our paper is organized as follows. In Section 2, we present our neural network ansatz. In Section 3, we present our numerical results demonstrating the effectiveness of our neural network ansatz and the technique of initializing the Monte Carlo sampling with the Hartree–Fock reference state. We conclude in Section 4.

## 2. Methods

### 2.1. Real Neural Network Ansatz

Before we introduce our model, we first briefly review the RBM used in NNQS. For a classical many-spin system, one could embed the system into a larger one consisting of visible spins (corresponding to the system) and hidden spins with the total (classical) Hamiltonian

$$\mathcal{H} = \sum_{j=1}^{N_v} a_j x_j + \sum_{i=1}^{N_h} b_i h_i + \sum_{i,j} W_{ij} h_i x_j, \quad (1)$$

where  $x_j$  represents the visible spin and  $h_i$  the hidden spin.  $N_v$  and  $N_h$  are the number of visible and hidden spins, respectively. The coefficients  $\theta = \{a, b, W\}$  are variational

parameters of the Hamiltonian. Since there is no coupling between the hidden spins, one could explicitly integrate them out and obtain the partition function of the system  $\mathcal{Z}$  as

$$\mathcal{Z} = \sum_{\mathbf{x}} p(\mathbf{x}), \tag{2}$$

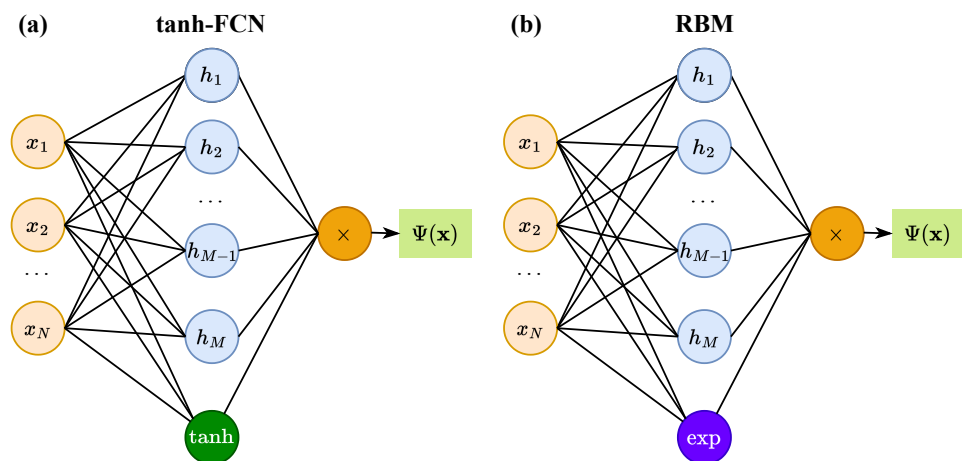
with  $\mathbf{x} = \{x_1, x_2, \dots, x_{N_v}\}$  a particular configuration and  $p(\mathbf{x})$  the unnormalized probability (in case of real coefficients) of  $\mathbf{x}$ , which can be explicitly written as

$$p(\mathbf{x}) = \sum_{\mathbf{h}} e^{\mathcal{H}} = e^{\sum_{j=1}^{N_v} a_j x_j} \times \prod_{i=1}^{N_h} 2 \cosh(b_i + \sum_{j=1}^{N_v} W_{ij} x_j). \tag{3}$$

When using RBM as a variational ansatz for the wave function of a quantum many-spin system,  $p(\mathbf{x})$  is interpreted as the amplitude (instead of the probability) of the configuration  $\mathbf{x}$ . Equation (3) can be seen as a single-layer fully connected neural work that accepts a configuration (a vector of integers) as input and outputs a scalar. For real coefficients, the output will always be positive by definition; therefore, one generally has to use complex coefficients even for real wave functions. In this work, we slightly change Equation (3) as follows so as to be able to output any real numbers with a real neural network:

$$p(\mathbf{x}) = \tanh\left(\sum_{j=1}^{N_v} a_j x_j\right) \times \prod_{i=1}^{N_h} 2 \cosh(b_i + \sum_{j=1}^{N_v} W_{ij} x_j). \tag{4}$$

In the following, we will write  $p(\mathbf{x})$  as  $\Psi_{\theta}(\mathbf{x})$  to stress its dependence on the variational parameters and that it is interpreted as a wave function instead of a probability distribution. We will also refer to our neural network in Equation (4) as tanh-FCN since it contains a fully connected layer followed by hyperbolic tangent as the activation function. The difference between RBM and tanh-FCN is demonstrated in Figure 1.



**Figure 1.** The architectures for (a) our tanh-FCN and (b) RBM. The major difference is that we use hyperbolic tangent as the activation function such that tanh-FCN could output both positive and negative numbers even if it only uses real numbers.

### 2.2. Variational Monte Carlo

The electronic Hamiltonian  $\hat{H}^e$  of a chemical system can be written in a second-quantized formulation:

$$\hat{H}^e = \sum_{p,q} h_q^p a_p^\dagger a_q + \frac{1}{2} \sum_{\substack{p,q \\ r,s}} g_{rs}^{pq} a_p^\dagger a_q^\dagger a_r a_s \tag{5}$$

where  $h_q^p$  and  $g_{rs}^{pq}$  are one- and two-electron integrals in molecular orbital basis, and  $a_p^\dagger$  and  $a_q$  in the Hamiltonian are the creation and annihilation operators. To treat the fermionic systems, we first use the Jordan–Wigner transformation to map the electronic Hamiltonian to a sum of Pauli operators, following Ref. [23], and then use our tanh-FCN in Equation (4) as the ansatz for the resulting many-spin system. The resulting spin Hamiltonian  $\hat{H}$  can generally be written in the following form:

$$\hat{H} = \sum_i c_i \prod_{j=1}^N \sigma_j^{v_{i,j}}, \tag{6}$$

where  $N = N_v$  is the number of spins,  $c_i$  is a real coefficient, and  $\sigma_j^{v_{i,j}}$  is a single spin Pauli operator acting on the  $j$ -th spin ( $v_{i,j} \in \{0, 1, 2, 3\}$  and  $\sigma^0 = I, \sigma^1 = \sigma^x, \sigma^2 = \sigma^y, \sigma^3 = \sigma^z$ ).

Given the wave function ansatz  $\Psi_\theta(\mathbf{x})$ , the corresponding energy can be computed as

$$E(\theta) = \frac{\langle \Psi_\theta | \hat{H} | \Psi_\theta \rangle}{\langle \Psi_\theta | \Psi_\theta \rangle} = \frac{\sum_{\mathbf{x}} E_{\text{loc}}(\mathbf{x}) |\Psi_\theta(\mathbf{x})|^2}{\sum_{\mathbf{y}} |\Psi_\theta(\mathbf{y})|^2}, \tag{7}$$

where the “local energy”  $E_{\text{loc}}(\mathbf{x})$  for a configuration  $\mathbf{x}$  is defined as

$$E_{\text{loc}}(\mathbf{x}) = \sum_{\mathbf{x}'} \frac{\Psi_\theta(\mathbf{x}')}{\Psi_\theta(\mathbf{x})} H_{\mathbf{x}'\mathbf{x}}, \tag{8}$$

with  $H_{\mathbf{x}'\mathbf{x}} = \langle \mathbf{x}' | \hat{H} | \mathbf{x} \rangle$ . The VMC algorithm evaluates Equation (7) approximately using Monte Carlo sampling, namely,

$$\tilde{E}(\theta) = \langle E_{\text{loc}} \rangle, \tag{9}$$

where the average is over a set of samples  $\{\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^{N_s}\}$  ( $N_s$  is the total number of samples), generated from the probability distribution  $|\Psi_\theta(\mathbf{x})|^2$ .  $\tilde{E}(\theta)$  will converge to  $E(\theta)$  if  $N_s$  is large enough. In this work, we use the Metropolis–Hastings sampling algorithm to generate samples [33]. A configuration  $\mathbf{x}$  is updated using the SWAP operation between nearest-neighbor pairs of spins to preserve the electron-number conservation. We also use the natural gradient of Equation (9) for the stochastic gradient descent algorithm in VMC, namely, the parameters are updated as

$$\theta^{k+1} = \theta^k - \alpha S^{-1} F, \tag{10}$$

where  $k$  is the number of iterations,  $\alpha$  is the learning rate ( $\alpha$  is dependent on  $k$  in general),  $S$  is the stochastic reconfiguration matrix [34,35], and  $F$  is the gradient of Equation (9). Concretely,  $S$  and  $F$  are computed by

$$S_{ij}(k) = \langle O_i^* O_j \rangle - \langle O_i^* \rangle \langle O_j \rangle, \tag{11}$$

and

$$F_i(k) = \langle E_{\text{loc}} O_i^* \rangle - \langle E_{\text{loc}} \rangle \langle O_i^* \rangle \tag{12}$$

respectively, with  $O_i(\mathbf{x})$  defined as

$$O_i(\mathbf{x}) = \frac{1}{\Psi_\theta(\mathbf{x})} \frac{\partial \Psi_\theta(\mathbf{x})}{\partial \theta_i}. \quad (13)$$

In general,  $S$  can be non-invertible, and a simple regularization is to add a small shift to the diagonals of  $S$ , namely, using  $S^{reg} = S + \epsilon I$  instead of  $S$  in Equation (10), with  $\epsilon$  a small number. The calculation of  $S$  can become the bottleneck in case the number of parameters is too large. This issue could be leveraged by representing  $S$  as a matrix function instead of building it explicitly [36], or by freezing a large portion of  $S$  during each iteration similar to DMRG [37]. Here, this is not a significant concern because we use at most about 1000 parameters to specify the network. To further enhance the stability of the algorithm, we add the contribution of an L2 regularization term when evaluating the gradient in Equation (10), that is, instead of directly choosing  $F$  as the gradient of  $\tilde{E}(\theta)$ ,  $F$  is chosen as the gradient of the function  $\tilde{E}(\theta) + \lambda \|\theta\|^2$  instead where  $\|\cdot\|^2$  means the square of the Euclidean norm. In this work, we choose  $\epsilon = 0.02$  and  $\lambda = 10^{-3}$  for our numerical simulations if not particularly specified.

### 3. Results

#### 3.1. Training Details

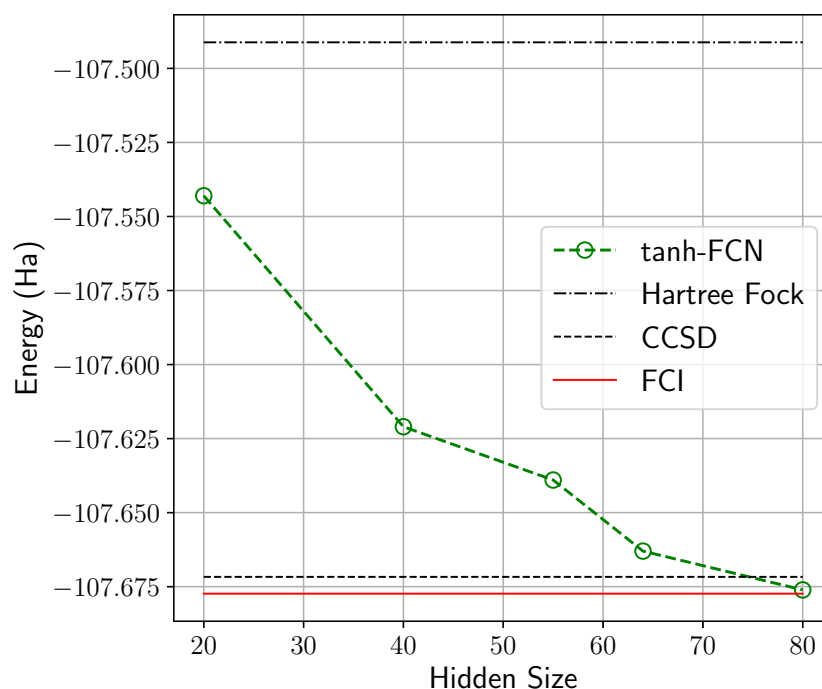
In this work, we use the Adam optimizer [38] for the VMC iterations, with an initial learning rate of  $\alpha = 0.001$ , and the decay rates for the first- and second-moment are  $\beta_1 = 0.9$ ,  $\beta_2 = 0.99$ , respectively. For the Metropolis–Hastings sampling, we will use a fixed  $N_s = 4 \times 10^4$  for our numerical simulations if not particularly specified (in principle, one should use a larger  $N_s$  for larger systems; however, in this work we focus on molecular systems with at most 30 qubits). We will also use a thermalization step of  $N_{th} = 2 \times 10^4$  (namely, throwing away  $N_{th}$  samples starting from the initial state). To avoid auto-correlation between successive samples we will only pick one out of every  $10N_v$  samples. In addition, for each simulation we run 8 Markov chains, and the energy is chosen to be the lowest of them. Since the energy will always contain some small fluctuations when  $N_s$  is not large enough, the final energy is evaluated by averaging over the energies of the last 20 VMC iterations.

#### 3.2. Effect of Hidden Size

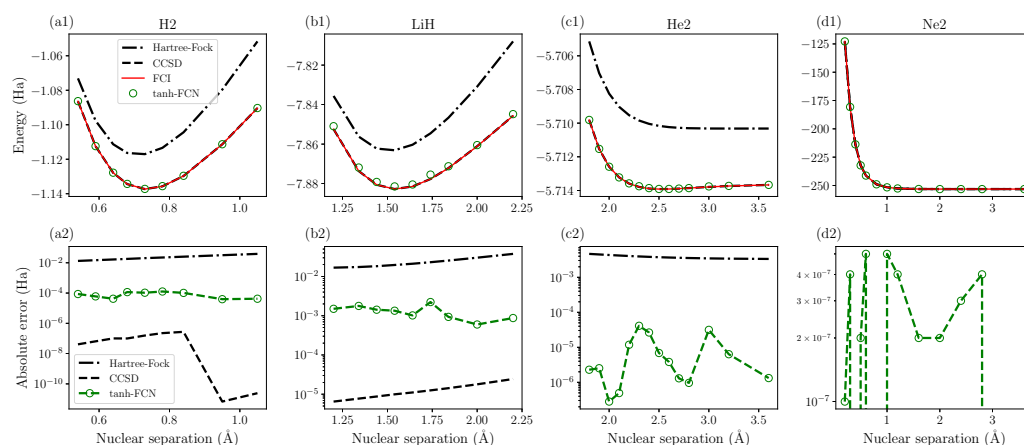
We first study the effect of  $N_h$ , which essentially determines the number of parameters and thus the expressivity of our tanh-FCN (analogously to RBM). The result is shown in Figure 2, where we have taken the N2 molecule as an example. We can see that by enlarging  $N_h$ , the precision of tanh-FCN can be systematically improved. With  $N_h = 4N_v = 80$ , we can already obtain a final energy that is lower than the CCSD results.

#### 3.3. Potential Energy Surfaces

Now we demonstrate the accuracy of our tanh-FCN by studying the potential energy surfaces of the two molecules  $H_2$  and  $LiH$  in the STO-3G basis, as shown in Figure 3(a1,b1). We can see that for both molecules under different bond lengths, our simulation can reach lower or very close to the chemical precision, namely error within  $1.6 \times 10^{-3}$  Hartree (Ha) or 1 kcal/mol (CCSD results are extremely accurate for these two molecules). To demonstrate of the effectiveness of our method for weakly correlated systems, we have also studied the potential energy surfaces of the two inert gas dimers  $He_2$  and  $Ne_2$  for completeness, which are shown in Figure 3(c1,d1). We can see that in the later cases our tanh-FCN can converge extremely well with the FCI results. We note that for  $Ne_2$  one may need to use a very large basis set to faithfully reproduce the actual potential energy surface, while here we have used the minimal STO-3G basis set due to the limitation of our current implementation.



**Figure 2.** Influence of the number of hidden spins in our tanh-FCN on the accuracy of the final energy. The N2 molecule in the STO-3G basis is used.



**Figure 3.** Potential energy surfaces of (a1) H<sub>2</sub>, (b1) LiH, (c1) He<sub>2</sub> and (d1) Ne<sub>2</sub>. We have used  $N_h/N_v = 2$  for H<sub>2</sub>, He<sub>2</sub>, Ne<sub>2</sub> and  $N_h/N_v = 4$  for LiH, which are sufficient for our tanh-FCN to reach chemical precision. We have also used  $N_s = 2 \times 10^4$  for both molecules during the training. (a2), (b2), (c2) and (d2) show the absolute error with respect to the FCI energy for H<sub>2</sub>, LiH, He<sub>2</sub> and Ne<sub>2</sub> respectively. We have used the STO-3G basis set for H<sub>2</sub>, LiH and Ne<sub>2</sub>, and the 6-31G basis set (using a (2e,4o) active space) for He<sub>2</sub>.

### 3.4. Final Energies for Several Molecular Systems

We further compare the precision of tanh-FCN with RBM and CCSD for several small-scale molecules, which are shown in Table 1. For these simulations we have used  $N_h/N_v = 2$ , while the RBM results are taken from Ref. [23]. As a proof of principle demonstration, we have mostly used the STO-3G basis set. However, we have also considered the LiH molecule in a larger basis set (6-31G) as well as in the localized molecular basis set (we have used the canonical molecular basis set for the rest ones) to show the effectiveness of our method in more general cases. Unlike DMRG which uses a one-dimensional matrix product state as the wave function ansatz, our neural network ansatz has an all to all structure which could represent certain volume-law quantum states [29], therefore it does

not significantly rely on localized orbitals and it seems that using a localized basis set could not improve the precision or significantly reduce the computational cost for us. From the runtime performance point of view, properly selected orbital localization scheme could reduce the number of Pauli terms in the Hamiltonian thus effectively accelerate the algorithm. However, this improvement is not universally achieved. For example, the number of Pauli terms for an equi-spaced  $H_{12}$  molecule with  $R(H-H) = 2.5$  Angstrom can be reduced from 14,905 to 4377 if natural atomic orbitals (NAO) [39] are used, while this number is increased to 23,109 if the bond length is changed to 0.7 Angstrom. An optimal choice of orbital localization methods is usually system-specific [40] and requires benchmark for the neural network ansatz.

These results show that even with a relatively small number of parameters and a real neural network, we can still obtain the ground state energies of a wide variety of molecules to very high precision (close to or lower than the CCSD energies). In the meantime, we note that the energies obtained using tanh-FCN is not as accurate as those obtained using RBM, however the computational cost of tanh-FCN is at least two times lower than RBM under with the same  $N_h$  and we could relatively easily study larger systems such as  $CO_2$  with 30 qubits. It should be noted that the total energy depends on the basis set size and the basis type, in principle, we should use a large basis set to obtain more reliable results.

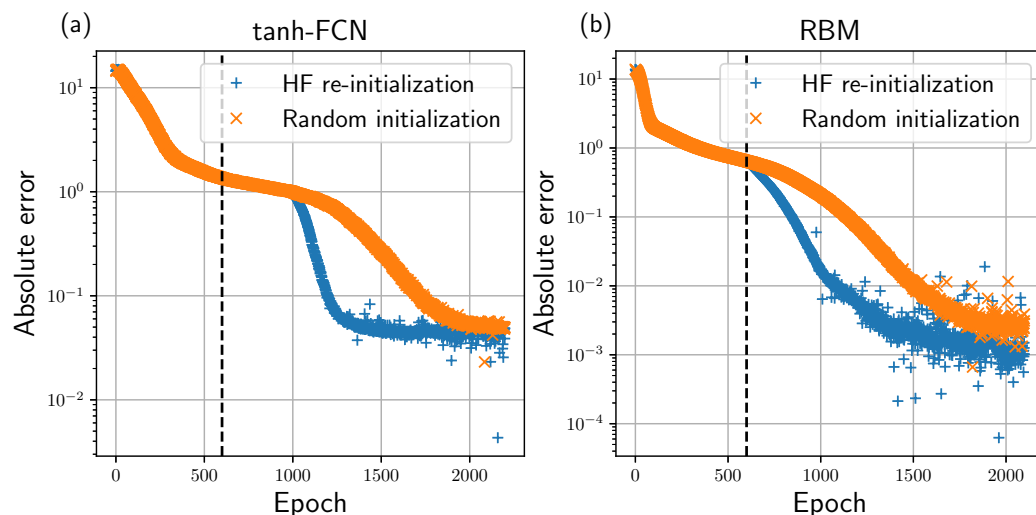
**Table 1.** List of molecules and the ground state energies computed using RBM, tanh-FCN, and CCSD. The FCI energy is also shown as a reference. The column  $N_v$  shows the number of qubits. We have used  $N_h/N_v = 2$  for all of the molecules studied.

Molecule	$N_v$	RBM [23]	tanh-FCN	CCSD	FCI
$H_2$	4	−1.1373	−1.1373	−1.1373	−1.1373
Be	10	-	−14.4033	−14.4036	−14.4036
C	10	-	−37.2184	−37.1412	−37.2187
$Li_2$	20	-	−14.6641	−14.6665	−14.6666
LiH	12	−7.8826	−7.8816	−7.8828	−7.8828
$NH_3$	16	−55.5277	−55.5101	−55.5279	−55.5282
$H_2O$	14	−75.0232	−75.0021	−75.0231	−75.0233
$C_2$	20	−74.6892	−74.6134	−74.6744	−74.6908
$N_2$	20	−107.6767	−107.622	−107.6716	−107.6774
$CO_2$	30	-	−185.1247	−184.8927	−185.2761

### 3.5. Effect of Hartree–Fock Re-Initialization

There are generally two ingredients which would affect the effectiveness of the NNQS algorithm: (1) the expressivity of the underlying neural network ansatz and (2) the ability to quickly approach the desired parameter regime during the VMC iterations. The former is dependent on an intelligent choice of the neural network ansatz. The effect of the latter is more significant for larger systems, and one generally needs to use a knowledgeable starting point such as transfer learning [41,42] for the VMC algorithm to guarantee success. For molecular systems, it is difficult to explore transfer learning since the knowledge for different molecules can hardly be shared. However, for molecular systems the Hartree–Fock reference state may have a large overlap with the exact ground state and is often used as a first approximation of the ground state. Here, we show that for quantum chemistry problems the ability to reach faster the ground state can be improved by using the knowledge of the Hartree–Fock reference state. Concretely, during the VMC iterations, after the energies have become sufficiently close to the ground state energy, we stop using random initialization for our Metropolis–Hastings sampling and use the Hartree–Fock reference state instead (Hartree–Fock re-initialization). The effect of the Hartree–Fock re-initialization is demonstrated in Figure 4, where we have taken the  $H_2O$  molecule as our example. To show the versatility of the Hartree–Fock re-initialization, we demonstrate its effect for RBM as well. We can see that for both tanh-FCN and RBM, using Hartree–Fock re-initialization after a number of VMC iterations can greatly accelerate the convergence and reach a lower

ground state energy than using random initialization throughout the VMC optimization. We can also see that for the H<sub>2</sub>O molecule, tanh-FCN is less accurate than RBM using the same  $N_h$ , which is probably due to the fact that under the same  $N_h$ , tanh-FCN has a different expressive power to RBM for H<sub>2</sub>O.



**Figure 4.** Effect of the Hartree–Fock (HF) re-initialization compared to random initialization for (a) tanh-FCN and (b) RBM. The H<sub>2</sub>O (STO-3G basis, 14 qubits) molecule is used here. The y-axis is the absolute error between the VMC energies and the FCI energy. For both methods, we start to use the HF re-initialization starting from 600-th VMC iteration marked by the vertical dashed lines. The other parameters used are  $N_s = 2 \times 10^4$ ,  $N_h/N_v = 1$  and  $\lambda = 10^{-4}$ .

#### 4. Conclusions

We propose a fully connected neural network inspired from the restricted Boltzmann machine to solve quantum chemistry problems. Compared to RBM, our tanh-FCN is able to output both positive and negative numbers even if the parameters of the network are purely real. As a result, we can directly study quantum chemistry problems using tanh-FCN with real numbers. In our numerical simulation, we demonstrate that tanh-FCN can be used to compute the ground states with high accuracy for a wide range of molecular systems with up to 30 qubits. In addition, we propose to explicitly use the Hartree–Fock reference state as the initial state for the Markov chain sampling used during the VMC algorithm and demonstrate that this technique can significantly accelerate the convergence and improve the accuracy of the final energy for both tanh-FCN and RBM. Our method could be used in combination with existing high performance computing devices that are well optimized for real numbers, such as to provide a scalable solution for large-scale quantum chemistry problems.

**Author Contributions:** Conceptualization, D.P. and H.S.; Methodology, C.G. and H.S.; Software, Y.W., X.X., Y.F. and C.G.; Validation, Y.W.; Visualization, Y.W. and C.G.; Writing—original draft, Y.W.; Writing—review & editing, D.P., C.G. and H.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** We thank Xiao Liang, Mingfan Li for helpful discussions of the algorithm. C.G. acknowledges support from National Natural Science Foundation of China under Grant No. 11805279. H.S. acknowledges support from the National Natural Science Foundation of China (22003073, T2222026). D.P. acknowledges support from the National Research Foundation, Singapore under its QEP2.0 programme (NRF2021-QEP2-02- P03).

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Helgaker, T.; Jørgensen, P.; Olsen, J. *Molecular Electronic-Structure Theory*; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 2000. [[CrossRef](#)]
2. Vogiatzis, K.D.; Ma, D.; Olsen, J.; Gagliardi, L.; de Jong, W.A. Pushing configuration-interaction to the limit: Towards massively parallel MCSCF calculations. *J. Chem. Phys.* **2017**, *147*, 184111.
3. White, S.R. Density matrix formulation for quantum renormalization groups. *Phys. Rev. Lett.* **1992**, *69*, 2863–2866. [[CrossRef](#)]
4. White, S.R. Density-matrix algorithms for quantum renormalization groups. *Phys. Rev. B* **1993**, *48*, 10345–10356. [[CrossRef](#)] [[PubMed](#)]
5. Brabec, J.; Brandejs, J.; Kowalski, K.; Xantheas, S.; Legeza, Ö.; Veis, L. Massively parallel quantum chemical density matrix renormalization group method. *J. Comput. Chem.* **2021**, *42*, 534–544.
6. Larsson, H.R.; Zhai, H.; Umrigar, C.J.; Chan, G.K.L. The chromium dimer: Closing a chapter of quantum chemistry. *J. Am. Chem. Soc.* **2022**, *144*, 15932–15937. [[CrossRef](#)] [[PubMed](#)]
7. Perez-Garcia, D.; Verstraete, F.; Wolf, M.M.; Cirac, J.I. Matrix Product State Representations. *Quantum Inf. Comput.* **2007**, *7*, 401–430. [[CrossRef](#)]
8. Purvis, G.D.; Bartlett, R.J. A full coupled-cluster singles and doubles model: The inclusion of disconnected triples. *J. Chem. Phys.* **1982**, *76*, 1910–1918. [[CrossRef](#)]
9. Čížek, J. On the Correlation Problem in Atomic and Molecular Systems. Calculation of Wavefunction Components in Ursell-Type Expansion Using Quantum-Field Theoretical Methods. *J. Chem. Phys.* **1966**, *45*, 4256–4266.
10. Coester, F.; Kümmel, H. Short-range correlations in nuclear wave functions. *Nucl. Phys.* **1960**, *17*, 477–485. [[CrossRef](#)]
11. Shepard, R. The Multiconfiguration Self-Consistent Field Method. In *Advances in Chemical Physics*; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 1987; pp. 63–200. [[CrossRef](#)]
12. Knowles, P.J.; Werner, H.J. An efficient second-order MC SCF method for long configuration expansions. *Chem. Phys. Lett.* **1985**, *115*, 259–267. [[CrossRef](#)]
13. Jensen, H.J.A. Electron Correlation in Molecules Using Direct Second Order MCSCF. In *Relativistic and Electron Correlation Effects in Molecules and Solids*; Malli, G.L., Ed.; Springer: Boston, MA, USA, 1994; pp. 179–206. [[CrossRef](#)]
14. Sun, Q.; Zhang, X.; Banerjee, S.; Bao, P.; Barbry, M.; Blunt, N.S.; Bogdanov, N.A.; Booth, G.H.; Chen, J.; Cui, Z.H.; et al. Recent developments in the PySCF program package. *J. Chem. Phys.* **2020**, *153*, 024109. [[CrossRef](#)] [[PubMed](#)]
15. Carleo, G.; Troyer, M. Solving the quantum many-body problem with artificial neural networks. *Science* **2017**, *355*, 602–606.
16. Choo, K.; Neupert, T.; Carleo, G. Two-dimensional frustrated  $J_1 - J_2$  model studied with neural network quantum states. *Phys. Rev. B* **2019**, *100*, 125124. [[CrossRef](#)]
17. Schmitt, M.; Heyl, M. Quantum Many-Body Dynamics in Two Dimensions with Artificial Neural Networks. *Phys. Rev. Lett.* **2020**, *125*, 100503. [[CrossRef](#)] [[PubMed](#)]
18. Yuan, D.; Wang, H.R.; Wang, Z.; Deng, D.L. Solving the Liouvillian Gap with Artificial Neural Networks. *Phys. Rev. Lett.* **2021**, *126*, 160401. [[CrossRef](#)]
19. Moreno, J.R.; Carleo, G.; Georges, A.; Stokes, J. Fermionic wave functions from neural-network constrained hidden states. *Proc. Natl. Acad. Sci. USA* **2022**, *119*, e2122059119.
20. Hermann, J.; Schätzle, Z.; Noé, F. Deep-neural-network solution of the electronic Schrödinger equation. *Nat. Chem.* **2020**, *12*, 891–897. [[CrossRef](#)]
21. Pfau, D.; Spencer, J.S.; Matthews, A.G.D.G.; Foulkes, W.M.C. Ab initio solution of the many-electron Schrödinger equation with deep neural networks. *Phys. Rev. Res.* **2020**, *2*, 033429. [[CrossRef](#)]
22. Humeniuk, S.; Wan, Y.; Wang, L. Autoregressive neural Slater-Jastrow ansatz for variational Monte Carlo simulation. *arXiv* **2022**, arXiv:2210.05871.
23. Choo, K.; Mezzacapo, A.; Carleo, G. Fermionic neural-network states for ab-initio electronic structure. *Nat. Commun.* **2020**, *11*, 2368. [[CrossRef](#)]
24. Barrett, T.D.; Malyshev, A.; Lvovsky, A. Autoregressive neural-network wavefunctions for ab initio quantum chemistry. *Nat. Mach. Intell.* **2022**, *4*, 351–358. [[CrossRef](#)]
25. Zhao, T.; Stokes, J.; Veerapaneni, S. Scalable neural quantum states architecture for quantum chemistry. *arXiv* **2022**, arXiv:2208.05637.
26. Wu, D.; Rossi, R.; Vicentini, F.; Carleo, G. From Tensor Network Quantum States to Tensorial Recurrent Neural Networks. *arXiv* **2022**, arXiv:2206.12363.
27. Sharir, O.; Shashua, A.; Carleo, G. Neural tensor contractions and the expressive power of deep neural quantum states. *Phys. Rev. B* **2022**, *106*, 205136. [[CrossRef](#)]
28. Glasser, I.; Pancotti, N.; August, M.; Rodriguez, I.D.; Cirac, J.I. Neural-Network Quantum States, String-Bond States, and Chiral Topological States. *Phys. Rev. X* **2018**, *8*, 011006. [[CrossRef](#)]
29. Deng, D.L.; Li, X.; Das Sarma, S. Quantum Entanglement in Neural Network States. *Phys. Rev. X* **2017**, *7*, 021021. [[CrossRef](#)]
30. Nomura, Y.; Imada, M. Dirac-Type Nodal Spin Liquid Revealed by Refined Quantum Many-Body Solver Using Neural-Network Wave Function, Correlation Ratio, and Level Spectroscopy. *Phys. Rev. X* **2021**, *11*, 031034. [[CrossRef](#)]
31. Liang, X.; Li, M.; Xiao, Q.; An, H.; He, L.; Zhao, X.; Chen, J.; Yang, C.; Wang, F.; Qian, H.; et al.  $2^{1296}$  Exponentially Complex Quantum Many-Body Simulation via Scalable Deep Learning Method. *arXiv* **2022**, arXiv:2204.07816.

32. Torlai, G.; Mazzola, G.; Carrasquilla, J.; Troyer, M.; Melko, R.; Carleo, G. Neural-network quantum state tomography. *Nat. Phys.* **2018**, *14*, 447–450. [[CrossRef](#)]
33. Hastings, W.K. Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika* **1970**, *57*, 97–109. [[CrossRef](#)]
34. Sorella, S.; Capriotti, L. Green function Monte Carlo with stochastic reconfiguration: An effective remedy for the sign problem. *Phys. Rev. B* **2000**, *61*, 2599–2612. [[CrossRef](#)]
35. Sorella, S.; Casula, M.; Rocca, D. Weak binding between two aromatic rings: Feeling the van der Waals attraction by quantum Monte Carlo methods. *J. Chem. Phys.* **2007**, *127*, 014105.
36. Vicentini, F.; Hofmann, D.; Szabó, A.; Wu, D.; Roth, C.; Giuliani, C.; Pescia, G.; Nys, J.; Vargas-Calderón, V.; Astrakhantsev, N.; et al. NetKet 3: Machine Learning Toolbox for Many-Body Quantum Systems. *SciPost Phys. Codebases* **2022**, *7*. [[CrossRef](#)]
37. Zhang, W.; Xu, X.; Wu, Z.; Balachandran, V.; Poletti, D. Ground state search by local and sequential updates of neural network quantum states. *arXiv* **2022**, arXiv:2207.10882.
38. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**, arXiv:1412.6980.
39. Reed, A.E.; Weinstock, R.B.; Weinhold, F. Natural population analysis. *J. Chem. Phys.* **1985**, *83*, 735–746. [[CrossRef](#)]
40. Ma, Y.; Ma, H. Assessment of various natural orbitals as the basis of large active space density-matrix renormalization group calculations. *J. Chem. Phys.* **2013**, *138*, 224105. [[CrossRef](#)] [[PubMed](#)]
41. Zen, R.; My, L.; Tan, R.; Hébert, F.; Gattobigio, M.; Miniatura, C.; Poletti, D.; Bressan, S. Transfer learning for scalability of neural-network quantum states. *Phys. Rev. E* **2020**, *101*, 053301. [[CrossRef](#)] [[PubMed](#)]
42. Hébert, F.; Zen, R.; My, L.; Tan, R.; Gattobigio, M.; Miniatura, C.; Poletti, D.; Bressan, S. Finding Quantum Critical Points with Neural-Network Quantum States. *arXiv* **2020**, arXiv:2002.02618

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.