Versão do arquivo anexado / Version of attached file:

Versão do Editor / Published Version

Mais informações no site da editora / Further information on publisher's website:

https://quantum-journal.org/papers/q-2023-12-18-1209

DOI: 10.22331/q-2023-12-18-1209

Direitos autorais / Publisher's copyright statement:

Synergy between deep neural networks and the variational Monte Carlo method for small $^4\mathrm{He}_N$ clusters

William Freitas and S. A. Vitiello

Instituto de Física Gleb Wataghin
University of Campinas - UNICAMP
13083-859 Campinas - SP, Brazil

We introduce a neural network-based approach for modeling wave functions that satisfy Bose-Einstein statistics. Applying this model to small $^4\mathrm{He}_N$ clusters (with N ranging from 2 to 14 atoms), we accurately predict ground state energies, pair density functions, and two-body contact parameters $C_2^{(N)}$ related to weak unitarity. The results obtained via the variational Monte Carlo method exhibit remarkable agreement with previous studies using the diffusion Monte Carlo method, which is considered exact within its statistical uncertainties. This indicates the effectiveness of our neural network approach for investigating many-body systems governed by Bose-Einstein statistics.

The field of artificial intelligence (AI) has seen significant advancements in recent years, making it a promising tool for addressing complex problems in quantum many-body physics [1–5]. Studies have demonstrated the effectiveness of AI in a wide range of applications [6], leading to increasing interest in exploring its potential for understanding nature as a whole. This paper builds upon this progress by proposing a neural network-based approach for modeling a wave function that satisfies Bose-Einstein statistics and applying it to the study of small clusters of $^4\mathrm{He}_N$ with $N = \{2, \dots 14\}$ atoms using the variational Monte Carlo (VMC) method.

Exact solutions of the Schrödinger equation are known for only a limited number of systems. In this context, quantum Monte Carlo methods have become a powerful tool for understanding quantum many-body systems. In recent years, machine learning, specifically neural networks, has gained traction as a method for studying fermionic systems, after the seminal work of Carleo and Troyer [7]. Although there have been some efforts [8, 9], the application of machine learning to the analysis of bosonic systems has generally received less attention. The present work

William Freitas: wfsilva@ifi.unicamp.br

aims to address this gap. Representing trial functions with neural networks and training them through an unsupervised VMC method yields results comparable to those obtained through the diffusion Monte Carlo (DMC) method, highlighting the potential of machine learning for studying Bose-Einstein systems. The diffusion Monte Carlo method, a projective method, requires configurations drawn from a trial function that is not orthogonal to the wavefunction of the system's ground state. Even when the available trial function is of low quality, this method can still yield very good results for the ground state energy of bosonic systems. However, in such cases, estimates of other non-commuting properties with the system's Hamiltonian may be biased due to the need for extrapolation in the estimation process. An alternative approach is to use a neural network-based representation of the ground state trial function, which can provide excellent energies and most likely better estimates of many properties in agreement with experimental results. This study of $^4\mathrm{He}_N$ clusters is not only a proof of concept, but also a topic that continues to be of recent interest [10–19].

Methods

The stationary states of small $^4\mathrm{He}_N$ clusters can be modeled by assuming that they are described by the Hamiltonian

$$\mathcal{H} = \frac{-\hbar^2}{2m} \sum_i^N \boldsymbol{\nabla}_i^2 + \sum_{i<j}^N V(r_{ij}), \qquad (1)$$

where the kinetic energy term depends on the atomic mass m, and V, the interatomic interaction, on the relative distance between atoms, $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$. The variational energies of the clusters were computed using the HFD-He interatomic potential, formulated by Aziz and collaborators [20]. This potential is widely employed in the study of $^4\mathrm{He}_N$ clusters [10, 21] due to its accurate modeling of the interactions between atoms in the Hamiltonian of Eq. 1. Exclusively to gain further insights into the optimization process, the Lennard-Jones [12-6] (LJ) potential was also utilized at this stage. The HFD-He potential offers a significantly higher level of accuracy for the system under
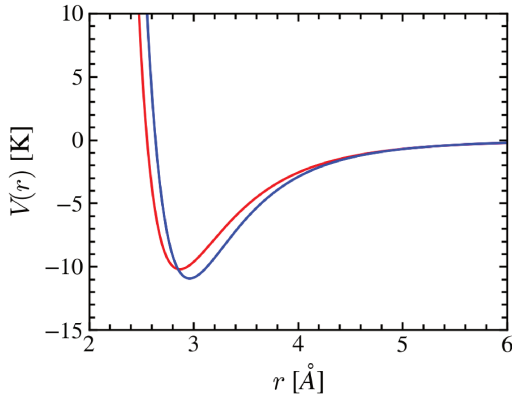
Figure 1: Potential energy V as a function of the pair separation r for the interatomic potentials Lennard-Jones (red) and HFD-He (blue).

investigation compared to the LJ potential. The minimum of the HFD-He potential is slightly shifted to large interatomic separations and is marginally deeper in comparison to the LJ interaction for helium atoms. A comparison of these potentials is presented in Fig.1. Among the very accurate potentials available to describe helium atom interactions, one of the motivations for choosing the HFD-He potential is the availability of existing literature results for the clusters we have analyzed.

The neural network architecture was built in search of efficiency. Thus, translational invariance and the symmetrical character of the wave function were encoded, since the Hamiltonian commutes with the translation operator and the wave function must obey Bose-Einstein statistics. The encoded neural network quantum state was introduced to model the ground state wave function

$$\Psi_{\text{DNN}}^{\text{B}}(R) = \exp\left[-\frac{1}{2}\sum_{i<j}^{N} r_{ij}^{-5}\right]\left(\sum_{\alpha=1}^{8}\omega_\alpha\chi_\alpha(R)\right),$$
(2)

where $R = \{\mathbf{r}_i \mid i = 1,...,N\}$ is a set of space coordinates associated with the N particles that form the many-body system. The first term of this trial function is a product of two-body terms of the Jastrow form with a pseudopotential of the McMillan form [22], which captures short-range correlations. It is multiplied by a sum of eight functions χ_α weighted by variational parameters ω_α. This is an accurate function that includes Feynman and Cohen backflow [23] correlations that cause an atom position to be dependent on the whole system configuration. In fact, it goes beyond the conventional manner [24] in which backflow correlations are included. It replaces a single function by a set of functions that includes backflow.

The simplicity of the trial function in Eq. 2, with only the variational parameters ω_α explicitly stated, may obscure a limitation of the variational method. This limitation arises from the presence of a substan-

tial number of parameters embedded within the functions χ_α, requiring optimization, which makes this process cumbersome. In the following sections, we will delve into strategies for overcoming these challenges and demonstrate how deep neural networks can be employed to effectively represent trial functions for the study of many-body bosonic systems.

Accurately handling atomic configurations is vital for an appropriate representation of the trial function. The information pertaining to each atom is presented as a vector

$$\mathbf{f}_i^0(\mathbf{r}_i, \{\mathbf{r}_{i/}\}) = \left(\mathbf{q}_i, q_i, \sum_j \frac{\mathbf{r}_{ij}}{N}, \sum_j \frac{r_{ij}}{N}\right),$$
(3)

where $\{\mathbf{r}_{i/}\}$ is the set of all space points other than the $\mathbf{r}_i$ one, $\mathbf{q}_i = \mathbf{r}_i - \mathbf{r}_{\text{cm}}$ is the particle coordinate relative to the center of mass $\mathbf{r}_{\text{cm}}$ and $q_i = |\mathbf{r}_i - \mathbf{r}_{\text{cm}}|$. The particle $\mathbf{r}_i$ is referenced as the main particle.

Any one of the four layers ℓ of the neural network has as input a single-atom stream $\mathbf{h}_i^\ell$ and a two-atoms stream $\mathbf{h}_{ij}^\ell$; for $\ell = 0$, the streams are $\mathbf{h}_i^0 = (\mathbf{q}_i, q_i)$ and $\mathbf{h}_{ij}^0 = (\mathbf{r}_{ij}, r_{ij})$. The average of both streams are employed to compute the intermediate single-atom stream vector $\mathbf{f}_i^\ell$ given by

$$\mathbf{f}_i^\ell = \left(\mathbf{h}_i^\ell - \sum_i \frac{\mathbf{h}_i^\ell}{N}, \sum_j \frac{\mathbf{h}_{ij}^\ell}{N}\right).$$
(4)

Each stream in layers with $\ell > 0$ are connected through a linear operation followed by a non-linear one. Information from the previous layer is also transmitted in the form of residual connections when both streams have the same shape. These steps can be summarized as

$$\mathbf{h}_i^{\ell+1} = \tanh\left(\mathbf{V}^\ell\mathbf{f}_i^\ell + \mathbf{b}^\ell\right) + \mathbf{h}_i^\ell,$$
(5)

$$\mathbf{h}_{ij}^{\ell+1} = \tanh\left(\mathbf{W}^\ell\mathbf{h}_{ij}^\ell + \mathbf{c}^\ell\right) + \mathbf{h}_{ij}^\ell.$$
(6)

The weights $\mathbf{V}^\ell$ and $\mathbf{W}^\ell$, as well as the biases $\mathbf{b}^\ell$ and $\mathbf{c}^\ell$ of the neural network, are variational parameters to be optimized. The final layer single-particle stream outputs $\mathbf{h}_i^L$ of each particle are reshaped into a matrix of elements $h_{\alpha\nu}$ used to construct the orbitals

$$\phi_\alpha(\mathbf{r}_i, \{\mathbf{r}_{i/}\}) = \sum_\nu h_{\alpha\nu}(\mathbf{r}_i, \{\mathbf{r}_{i/}\})\exp[-a_{\alpha\nu}q_i],$$
(7)

where each main particle is correlated to the centre of mass to bind the system and to give the correct behavior asymptotically when $r_i \to \infty$. The decaying rates $a_{\alpha\nu}$ are also variational parameters. The functions ϕ_α remain invariant under exchanges of pairs of atoms that do not include the i-th atom. This is due to the fact that the input features of the neural networks do not depend on the order of coordinates in $\{\mathbf{r}_{i/}\}$. Finally, symmetric functions χ_α are formed by the product

$$\chi_\alpha(R) = \prod_i \phi_\alpha(\mathbf{r}_i, \{\mathbf{r}_{i/}\}),$$
(8)

where particle exchanges leave χ_α invariant due to the commutativity of the product, which allows rearrangement of the terms in any order.

The foundational architecture of our approach was inspired by FermiNet [25], particularly in adopting a two-stream structure. However, there are several key distinctions between the two. Firstly, our input features $\mathbf{f}_i^0$ and intermediate stream vectors $\mathbf{f}_i^\ell$ are specialized for spin $\mathbf{s} = 0$ particles. Additionally, our single-particle input $\mathbf{h}_i^0$ bounds the i-th particle coordinates to the cluster's center of mass. This specialization arises from our treatment of the entire helium atom as a single particle, eliminating the Bohr-Oppenheimer approximation. This distinction is also evident in the intermediate stream vectors, where we employ subtraction, as outlined in Eq. 4, instead of concatenating the mean of single-particle vectors $\mathbf{h}_i^\ell$. Furthermore, our approach diverges significantly, starting from the $\mathbf{h}_i^L$ output single-particle stream onward, to address systems made from bosons. Nonetheless, it's worth noting that the linear combination of functions χ_α in Eq. 2 is equivalent to the multi-determinant expansion introduced in the fermionic function [2], establishing a final connection between the two Ansatz.

There is currently a growing focus on extending the neural network Ansatz to periodic systems [5, 26]. Notably, Pescia and colleagues [5] have investigated ^{4}He systems in one and two dimensions. Although both Eq. 3 and their trial function ensure that the input features remain invariant with respect to the symmetries of the problem, the approaches are distinct. In their work, periodic boundary conditions are implemented, while for the clusters, global translation invariance is considered. Additionally, the pooling operation, which confers permutation invariance, resembles the procedure outlined in Eq. 7. It's worth noting that their wave function construction is independent of system size, which proves advantageous once the network is trained. In contrast, our trial function is size-dependent, necessitating multiple training sessions that, nevertheless, also afford greater flexibility for the neural network to learn.

Unsupervised training of the neural network depends on the sampling of the probability distribution

$$p(R) = \frac{|\Psi_{\mathrm{DNN}}^{\mathrm{B}}(R)|^2}{\int dR \ |\Psi_{\mathrm{DNN}}^{\mathrm{B}}(R)|^2} \ . \tag{9}$$

The expectation value of the Hamiltonian E is estimated by approximating the integral $\int dR \, p(R) E_L(R)$ by the average values of the local energy $E_L(R) = \mathcal{H}\Psi_{\mathrm{DNN}}^{\mathrm{B}}(R)/\Psi_{\mathrm{DNN}}^{\mathrm{B}}(R)$ over the sampled configurations [27].

The set θ of all variational parameters is optimized using a second-order gradient-based method that accounts for correlations between pairs of variational parameters through derivatives of energy E and density probability p [2]

$$\boldsymbol{\nabla}_\theta E = 2 \int dR \ p(R) \times \tag{10}$$

$$\times \left[\frac{\mathcal{H}\psi(R)}{\psi(R)} - \int dR' \ p(R') \frac{\mathcal{H}\psi(R')}{\psi(R')} \right] \boldsymbol{\nabla}_\theta \log |\psi(R)| \ .$$

The optimization process updates the parameters repeatedly according to $\mathcal{F}^{-1}\boldsymbol{\nabla}_\theta E$, where $\mathcal{F}^{-1}$ is the inverse of the Fisher information matrix,

$$\mathcal{F}_{ij} = \int dR \ \frac{\partial \log p(R)}{\partial \theta_i} \frac{\partial \log p(R)}{\partial \theta_j}, \tag{11}$$

which correlates pairs of parameters.

The extensive parameter space inherent in neural networks can be handled in different ways. Particularly, the update of neural network parameters is a crucial step in the deep learning process. The iterative solver framework [28] is a powerful approach that can incorporate different methods to facilitate efficient parameter updates. In this work, we focus on employing one of these methods that may potentially be utilized within the iterative solver framework, namely the Kronecker-factored approximate curvature (K-FAC) method [29]. This method hinges on two critical principles. Firstly, it approximates the Fisher information matrix as block diagonal. This insight, rooted in the domain of neural networks, recognizes that for functions represented by neural networks, elements F_{ij} of the Fisher matrix effectively become zero if the network parameters Θ_i and Θ_j pertain to different layers. This approximation resonates with the hierarchical structure of neural networks, where parameters across distinct layers tend to exhibit a high degree of independence.

Secondly, to augment computational efficiency, the method employs Kronecker factorization to approximate the inverse of each of these blocks, including slight modifications to extend its applicability to unnormalized probability densities [2]. This technique offers a highly effective approach for managing the large-scale matrices inherent in neural network-based functions, thereby rendering the optimization process computationally manageable.

Collectively, these considerations underpin the methodology, facilitating the optimization of trial functions based on neural networks for intricate systems. In our specific case, this approach leads to a significant reduction in computational complexity, replacing the inversion of matrices with dimensions on the order of 10^5 with the inversion of ten matrices, each with dimensions of about 10^2.

In the optimization process, the selection of hyperparameters plays a crucial role in ensuring stability and efficiency. Many of these parameters can be extracted from the FermiNet approach [25]. The typical parameters for the simulations are provided in Table 1. The step size of proposed movements is not

included, since it is dynamically adjusted during the simulations. It is worth noting that the width of the layers can be reduced for smaller clusters, particularly for the single-atom stream width.

Table 1: Typical hyperparameter values employed in the simulations. An open-source version of the code used in this work to optimize the trial functions is available [30].

Batch size	8192
Single-atom channel width	128
Two-atom channel width	16
Symmetric functions	8
Number of layers	4
Learning rate	0.001
Learning rate decay	1.0
Learning rate delay	10000
Training iterations	2×10^5
Damping	0.001
Momentum	0.0
Covariance moving decay	0.95
Norm constrain	0.001

It is worthwhile to discuss why the results were obtained without a parameterization of the Jastrow factor in Eq. 2, as is typically done in variational calculations for systems composed of helium atoms. The main motivation for this choice was to enhance the efficiency of the optimization process. Preliminary tests, where this factor was parameterized, showed an unstable optimization process due to the intense character of the helium short-range interatomic interaction. While the HFD-He potential does not diverge at the origin, it does exhibit considerably high values for small pair separations. In finite-time runs, no sampled configuration will feature a pair at distances smaller than approximately 1 Å. However, eventually sampled configurations may involve a pair separation where the potential still has a substantial energy, and a local energy that significantly deviates from the average value. In principle, the Laplacian of the wave function should perfectly cancel out this value. However, the proposed trial wave function is only an approximation for the description of the ground state, and this perfect cancellation does not occur. Consequently, these imprecise estimations can significantly impede the optimization process by pushing variational parameters away from their optimal values and rendering the optimization process excessively slow.

The absence of the parameterization of the Jastrow factor in a trial function represented by a neural network for a system composed of helium atoms stands in contrast to approaches used in systems governed by Coulombic interactions [2, 31, 32]. The latter architectures prove especially effective in handling electron-electron correlations, scenarios where the Coulomb potential exhibits relatively predictable and manageable behavior.

These findings were validated through a simulation employing a Lennard-Jones potential with parameters $\varepsilon = 10.22$ K and $\sigma = 2.556$ Å for a 6-atom cluster. In this scenario, the Jastrow factor in Eq. 2, incorporating a McMillan pseudopotential with an r^{-5} dependency, ensures an asymptotically correct solution to the Schrödinger equation as $r \to 0$. The performance of the neural network with this Jastrow factor choice is depicted in the inset of Fig. 2, illustrating the total energies throughout the optimization process as a function of the performed iterations. The uncertainties associated with the energy are omitted due to significant variances in the estimated values during the initial iterations, making it difficult to observe the energy's progression as the optimization advances. Following the optimization, a standard VMC simulation with fixed variational parameters yielded a total binding energy of -1.8007(1) K, a value notably higher than that obtained with the HFD-He potential. This is the only result presented for the LJ potential, since its inclusion in this work was solely for the purpose of investigating the optimization stability.

The optimization of the parameters of the DNN Ansatz, Eq. 2, is performed to minimize the energy. Initially, the learning process was focused on determining parameters that yielded reasonable energy values. Subsequently, without an explicit directive, the neural network learned to reduce the variance, leading to more precise results as a consequence. This behavior can be seen in Fig. 2. More specifically, each optimization iteration consisted in obtaining a set of sampled configurations by applying the Metropolis algorithm followed by computing the energy and its gradients, the Fisher matrix, and updating the parameters to conclude the iteration. Typically, the number of configurations, *i.e.* the batch size, in each step was 2^{13}. After reaching a converged energy with a small variance, the optimization process was stopped. Then, a standard variational simulation was performed to get the final results and their associated variances.

At the beginning of the optimization process, large fluctuations in the total energy of the system often obscure the evolution of this quantity. To provide a clearer view of its progression during the optimization process, the uncertainties of the values shown in Fig. 2 are not displayed, as in the case of the LJ system.

Moreover, during the optimization, it was observed that this process evolved in a less than optimal manner, exhibiting slight instabilities after obtaining parameters that reasonably described the cluster energy. To address this challenge, we incorporated a hard sphere (HS) potential into the interatomic interaction given by the HFD-He potential during the course of the optimization. This addition addresses the issue that the r^{-5} dependence in the Jastrow factor of $\Psi_{\mathrm{DNN}}^{\mathrm{B}}$, Eq. 2, does not properly account for, in terms of behavior at short pair distances. The HS potential prevents atoms from getting too close
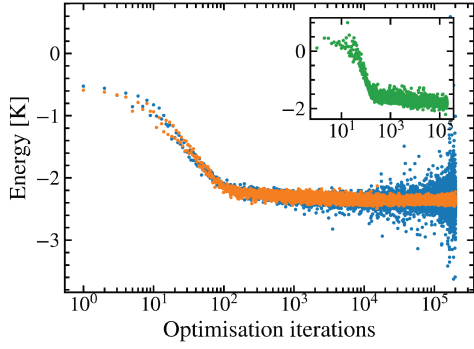
Figure 2: The total energy, without associated uncertainties, is presented to elucidate its evolution as a function of the optimization step for a cluster of $N = 6$ atoms employing the HFD-He potential (depicted as blue dots). The influence of incorporating a hard sphere (HS) potential is represented by the orange dots (refer to the text for details). The inset illustrates the optimization process for the Lennard-Jones system.

Table 2: Kinetic $\langle \mathcal{T} \rangle$ and total ground state $\langle E \rangle$ energies obtained with the DNN ansatz in units of Kelvin are presented for a ^{4}He cluster of size N. The fourth column shows DMC results from the literature [21].

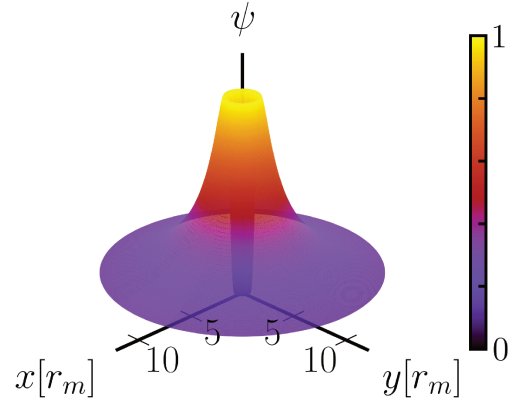| N	$\langle \mathcal{T} \rangle$	$\langle E \rangle$	DMC
2	0.1246(4)	-0.002142(6)	-
3	1.695(2)	-0.13323(9)	-0.135(2)
4	4.353(3)	-0.5775(1)	-0.573(2)
5	7.745(4)	-1.3341(2)	-1.334(2)
6	11.666(4)	-2.3710(3)	-2.367(3)
7	16.145(5)	-3.6510(4)	-3.646(4)
8	20.997(6)	-5.1448(4)	-5.144(5)
9	26.212(7)	-6.8207(9)	-6.827(6)
10	31.868(8)	-8.6766(7)	-8.673(6)
11	37.72(1)	-10.665(2)	-
12	43.50(1)	-12.801(2)	-
13	48.79(1)	-15.069(2)	-
14	55.92(1)	-17.444(2)	-

Figure 3: The wave function of the helium dimer is shown, with light surface colors emphasizing regions of large probability amplitude in an arbitrary scale. The space coordinates are given in terms of $r_m = 2.963$nm, where the potential is minimum.

to each other, overcoming the limitations of the Jastrow factor at small distances, which can lead to low-probability configurations.

The specific chosen value (1.8 Å) of the HS radius is not critical and did not produce any significant changes in the energy estimate. In fact, the weak unitarity satisfied by the helium clusters, makes the addition of an HS potential irrelevant, which will be discussed later. The optimization process with an infinite barrier at the HS radius is also shown in Fig. 2. The results indicate that the inadequate functional dependence of the pseudopotential in the Jastrow factor is a source of instability during the optimization process. The inclusion of the HS potential was able to lower the variances. Therefore, for all optimizations performed (and only in this stage), the HS potential was added to the HFD-He interatomic interaction.

Results

The reported final cluster energies were obtained in a standard VMC calculation with the interatomic interaction V given exclusively by the HFD-He potential while keeping the $\Psi_{\text{DNN}}^{\text{B}}$ parameters fixed. These results are presented in Table 2. Comparisons of the variational results with those from the DMC show that within statistical uncertainty they are in excellent agreement. This is remarkable because the DMC results are in principle exact in the Monte Carlo sense. For most of the clusters, the variational results are within one statistical standard deviation from the DMC results.

A three-dimensional plot of the two-body wave function is displayed in Fig. 3, where one of the atoms is at the origin and the probability amplitudes are shown for points (x, y) in the plane $z = 0$. The probability amplitude exhibits a peak at an interatomic separation of approximately 3.8 Å. The bond length, determined by the average value of the pair distance $\langle r_{ij} \rangle$, is 36.6 ± 0.6 Å. However, small pair separation is not unexpected, since the interatomic interaction used does not include relativistic and quantum electrodynamics contributions [33]. The ^{4}He$_2$ dimer is known to exist mostly in the quantum tunneling regime in a quantum halo state [34]. The plot demonstrates that even though radial symmetry was not explicitly imposed, the neural network was able to learn it.

The pair density function $\rho_N(r)$ for a cluster of N atoms is estimated by the operator

$$\hat{\rho}_N(r) = \sum_{i<j}^{N} \frac{\delta(r_{ij} - r)}{r^2}. \tag{12}$$

The integration of $\rho_N(r)$ in the whole space gives

$\int \rho_N(r) r^2 \mathrm{d}r = N(N-1)/2$. The pair density function is shown in Fig. 4.

Helium clusters are bound by the attractive van der Waals interaction, typically characterized by a $(-C_6/r^6)$ tail. This phenomenon arises due to the zero-point fluctuations of atomic dipole moments, leading to short-range correlations. In the study of these clusters, attention has recently been directed to the universal character of the short-range correlations [10, 15] by adapting the Tan relations [35–37] to clusters [38–40]. Within the framework of weak unitarity, the concept of contact that emerges from these relations serves to quantify the density of closely interacting pairs in strongly interacting systems, offering insights into their behavior at short distances.

While strong universality is not expected in helium clusters due to their characteristic distances (with an average pair distance of approximately 5 Å and a van der Waals length of about 5.4 Å [15]), it is well established [22] that the behavior of helium systems is primarily influenced by the closest pair of atoms. Nonetheless, provided there exists a strongly interacting model at short distances, the N-body wave function can still be factorized into the product of a universal two-body function and a state-dependent function [15]. This remains valid, because in close proximity, a correlated pair is minimally influenced by the surrounding particles. Consequently, its two-body wave function should exhibit consistency regardless of the system's size or state. This scenario is called weak universality [15].

The r-independent two-body contact $C_2^{(N)}$ can be estimated through the pair density function

$$\rho_N(r) \xrightarrow[\text{small } r]{} C_2^{(N)} \rho_2(r), \qquad (13)$$

by definition $C_2^{(2)} = 1$. The $C_2^{(N)}$ are treated as a fit parameter adjusted for values of r up to the maximum value of $\rho_N(r)$ in the minimization of the integral $\int (\rho_N(r) - C_2^{(N)} \rho_2(r))^2 \mathrm{d}r$.

The results of the pair density functions for $^4\mathrm{He}_N$ clusters, ranging from $N = 2$ to $N = 14$ atoms, normalized by contacts $C_2^{(N)}$ as a function of the radial distance, are displayed in Fig. 5. As shown, there is a clear collapse of the pair density functions for all cluster sizes. These results indicate that even the largest clusters, with up to 14 atoms, exhibit a weak universality. This is consistent with the coalescence of N-body atoms in a Bosonic system, as previously predicted [39]. The pair-atom contact $C_{2+1}^{(N)}$ [10] will be discussed in a separate publication.

Final comments

The use of neural network-based representation of the trial function for a bosonic system yielded results for core properties of $^4\mathrm{He}_N$ clusters that are in agreement
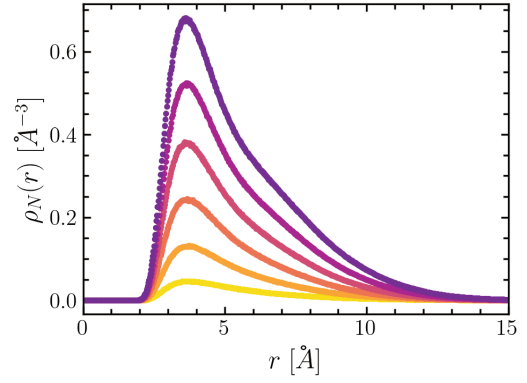

Figure 4: Pair density function for $^4\mathrm{He}_N$ clusters with $N = \{4, 6, 8, 12, 14\}$ atoms displayed by colors from light to dark.
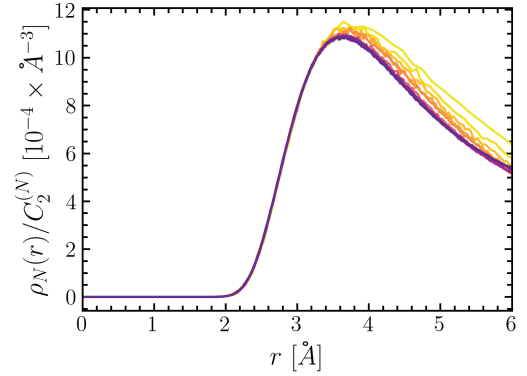

Figure 5: Pair density functions of $^4\mathrm{He}_N$ clusters with $N = 2$ to $N = 14$ atoms normalized with the appropriate contact $C_2^{(N)}$ shown by colors from light to dark.

with those obtained using the exact diffusion Monte Carlo method.

The implemented approach for the deep neural network does not rely on external information. The chosen representation of the trial function, by explicitly satisfying the Bose-Einstein statistics, avoids machine time spent considering non-physical solutions. Additionally, by satisfying global translation symmetry, the representation eliminates the need to consider infinite degenerate coordinates that represent the same physical situation. These features of the representation allow for an efficient optimization process, able to obtain results with modest computational resources.

Our results also demonstrated the importance, in the optimization process, of two-body correlation factors with the appropriate properties at the asymptotic limit $r \to 0$. The weak universality of $^4\mathrm{He}_N$ clusters enabled the inclusion of an HS potential in the atomic interaction, which stabilized the optimization process by eliminating configurations with very low probability, where two atoms are too close together. Previous literature [10, 41, 42] has presented various proposals for the two-body correlation factors of $^4\mathrm{He}_N$ clusters. The optimization process, due to its features at the limit $r \to 0$, could be valuable in determining the

correlation factor that most accurately describes the system. An improved functional form of the two-body correlation factor in $\Psi^{\mathrm{B}}_{\mathrm{DNN}}$ may remove the need for a cut-off in the interatomic interaction during optimization. An alternative and simpler approach that could also improve optimization is to use a soft sphere potential instead of an HS potential, which would allow the wave function to smoothly approach zero at short distance separations between atoms.

The weak universality of $^4\mathrm{He}_N$ clusters is shown through the collapse, up to the maxima, of the pair density functions normalized by their respective contacts $C_2^{(N)}$. These results were obtained by using a deep neural network to represent the trial function $\Psi^{\mathrm{B}}_{\mathrm{DNN}}$. This approach is validated both by previous findings [10, 15] and the specific analysis performed in this work.

One way to expand on this research is to study larger clusters and use the liquid drop model to extract the contact $C_2^{(2)}$ for bulk liquid $^4\mathrm{He}$. Another avenue for investigation could be to look into the effects of impurities in the clusters.

Acknowledgments

References

[1] Li Yang, Zhaoqi Leng, Guangyuan Yu, Ankit Patel, Wen-Jun Hu, and Han Pu. Deep learning-enhanced variational Monte Carlo method for quantum many-body physics. Physical Review Research, 2(1):012039, 2020-02. DOI: 10.1103/physrevresearch.2.012039.

[2] David Pfau, James S. Spencer, Alexander G. D. G. Matthews, and W. M. C. Foulkes. Ab initio solution of the many-electron schrödinger equation with deep neural networks. Physical Review Research, 2(3):033429, 2020-09. DOI: 10.1103/physrevresearch.2.033429.

[3] Jan Hermann, Zeno Schätzle, and Frank Noé. Deep-neural-network solution of the electronic schrödinger equation. Nature Chemistry, 12(10): 891–897, 2020-09. DOI: 10.1038/s41557-020-0544-y.

[4] Jan Kessler, Francesco Calcavecchia, and Thomas D. Kühne. Artificial neural networks as trial wave functions for quantum Monte Carlo. Advanced Theory and Simulations, 4(4):2000269, 2021-01. DOI: 10.1002/adts.202000269.

[5] Gabriel Pescia, Jiequn Han, Alessandro Lovato, Jianfeng Lu, and Giuseppe Carleo. Neural-network quantum states for periodic systems in continuous space. Physical Review Research, 4 (2):023138, 2022-05. DOI: 10.1103/physrevresearch.4.023138.

[6] Mario Krenn, Robert Pollice, Si Yue Guo, Matteo Aldeghi, Alba Cervera-Lierta, Pascal Friederich, Gabriel dos Passos Gomes, Florian Häse, Adrian Jinich, AkshatKumar Nigam, Zhenpeng Yao, and Alán Aspuru-Guzik. On scientific understanding with artificial intelligence. Nature Reviews Physics, 4(12):761–769, 2022-10. DOI: 10.1038/s42254-022-00518-3.

[7] Giuseppe Carleo and Matthias Troyer. Solving the quantum many-body problem with artificial neural networks. Science, 355(6325):602–606, feb 2017. DOI: 10.1126/science.aag2302.

[8] Michele Ruggeri, Saverio Moroni, and Markus Holzmann. Nonlinear network description for many-body quantum systems in continuous space. Physical Review Letters, 120(120):205302, May 2018. DOI: 10.1103/physrevlett.120.205302.

[9] Hiroki Saito and Masaya Kato. Machine learning technique to find quantum many-body ground states of bosons on a lattice. Journal of the Physical Society of Japan, 87(1):014001, 2018-01. DOI: 10.7566/jpsj.87.014001.

[10] A. J. Yates and D. Blume. Structural properties of $^4\mathrm{He}_N$ (N=2-10) clusters for different potential models at the physical point and at unitarity. Physical Review A, 105(2):022824, 2022-02. DOI: 10.1103/physreva.105.022824.

[11] J. Peter Toennies. Helium nanodroplets: Formation, physical properties and superfluidity. In Topics in Applied Physics, pages 1–40. Springer International Publishing, 2022. DOI: 10.1007/978-3-030-94896-2_1.

[12] P. Recchia, A. Kievsky, L. Girlanda, and M. Gattobigio. Subleading contributions to n-boson systems inside the universal window. Physical Review A, 106(2):022812, 2022-08. DOI: 10.1103/physreva.106.022812.

[13] Elena Spreafico, Giorgio Benedek, Oleg Kornilov, and Jan Peter Toennies. Magic numbers in boson $^4\mathrm{He}$ clusters: The auger evaporation mechanism. Molecules, 26(20):6244, 2021-10. DOI: 10.3390/molecules26206244.

[14] Daniel Odell, Arnoldas Deltuva, and Lucas Platter. van der waals interaction as the starting point for an effective field theory. Physical Review A, 104(2):023306, 2021-08. DOI: 10.1103/physreva.104.023306.

[15] B. Bazak, M. Valiente, and N. Barnea. Universal short-range correlations in bosonic helium clus-

ters. Physical Review A, 101(1):010501, 2020-01. DOI: 10.1103/physreva.101.010501.

[16] A. Kievsky, A. Polls, B. Juliá-Díaz, N. K. Timofeyuk, and M. Gattobigio. Few bosons to many bosons inside the unitary window: A transition between universal and nonuniversal behavior. Physical Review A, 102(6):063320, 2020-12. DOI: 10.1103/physreva.102.063320.

[17] B. Bazak, J. Kirscher, S. König, M. Pavón Valderrama, N. Barnea, and U. van Kolck. Four-body scale in universal few-boson systems. Physical Review Letters, 122(14), apr 2019. DOI: 10.1103/physrevlett.122.143001.

[18] A. Kievsky, M. Viviani, R. Álvarez-Rodríguez, M. Gattobigio, and A. Deltuva. Universal behavior of few-boson systems using potential models. Few-Body Systems, 58(2), 2017-01. DOI: 10.1007/s00601-017-1228-z.

[19] J. Carlson, S. Gandolfi, U. van Kolck, and S. A. Vitiello. Ground-state properties of unitary Bosons: From clusters to matter. Phys. Rev. Lett., 119:223002, Nov 2017. DOI: 10.1103/PhysRevLett.119.223002. URL https://link.aps.org/doi/10.1103/PhysRevLett.119.223002.

[20] Ronald A. Aziz, Frederick R.W. McCourt, and Clement C.K. Wong. A new determination of the ground state interatomic potential for He_2. Molecular Physics, 61(6):1487–1511, 1987-08. DOI: 10.1080/00268978700101941.

[21] Rafael Guardiola, Oleg Kornilov, Jesús Navarro, and J. Peter Toennies. Magic numbers, excitation levels, and other properties of small neutral he4 clusters (n≲50). The Journal of Chemical Physics, 124(8):084307, 2006-02. DOI: 10.1063/1.2140723.

[22] W. L. McMillan. Ground state of liquid ^{4}He. Phys. Rev., 138(2A):A442–A451, Apr 1965. DOI: 10.1103/PhysRev.138.A442.

[23] R. P. Feynman and Michael Cohen. Energy spectrum of the excitations in liquid helium. Phys. Rev., 102:1189–1204, Jun 1956. DOI: 10.1103/PhysRev.102.1189. URL http://link.aps.org/doi/10.1103/PhysRev.102.1189.

[24] K. E. Schmidt, Michael A. Lee, M. H. Kalos, and G. V. Chester. Structure of the ground state of a fermion fluid. Phys. Rev. Lett., 47:807–810, Sep 1981. DOI: 10.1103/PhysRevLett.47.807. URL http://link.aps.org/doi/10.1103/PhysRevLett.47.807.

[25] David Pfau James S. Spencer and FermiNet Contributors. FermiNet, 2020. URL http://github.com/deepmind/ferminet.

[26] Max Wilson, Saverio Moroni, Markus Holzmann, Nicholas Gao, Filip Wudarski, Tejs Vegge, and Arghya Bhowmik. Neural network ansatz for periodic wave functions and the homogeneous electron gas. Phys. Rev. B, 107:235139, Jun 2023. DOI: 10.1103/Phys-

RevB.107.235139. URL https://link.aps.org/doi/10.1103/PhysRevB.107.235139.

[27] D. M. Ceperley and M. H. Kalos. Quantum many-body problems. In K. Binder, editor, Monte Carlo Methods in Statistics Physics, volume 7 of Topics in Current Physics, chapter Quantum Many-Body Problems, pages 145–194. Springer-Verlag, Berlin, second edition, 1986. DOI: 10.1007/978-3-642-82803-4_4.

[28] Filippo Vicentini, Damian Hofmann, Attila Szabó, Dian Wu, Christopher Roth, Clemens Giuliani, Gabriel Pescia, Jannes Nys, Vladimir Vargas-Calderón, Nikita Astrakhantsev, and Giuseppe Carleo. NetKet 3: Machine learning toolbox for many-body quantum systems. SciPost Physics Codebases, 2022-08. DOI: 10.21468/scipostphyscodeb.7.

[29] James Martens and Roger B. Grosse. Optimizing neural networks with kronecker-factored approximate curvature. In ICML'15: Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37, 2015. DOI: 10.48550/arXiv.1503.05671. URL https://dl.acm.org/doi/10.5555/3045118.3045374.

[30] William Freitas. BoseNet Helium Clusters, 2023. URL https://github.com/freitas-esw/bosenet-helium-clusters.

[31] Nicholas Gao and Stephan Günnemann. Sampling-free inference for ab-initio potential energy surface networks. arXiv:2205.14962, 2022. DOI: 10.48550/arXiv.2205.14962.

[32] Ingrid von Glehn, James S. Spencer, and David Pfau. A self-attention ansatz for ab-initio quantum chemistry. axXiv:2211.13672, 2023. DOI: 10.48550/arXiv.2211.13672.

[33] M. Przybytek, W. Cencek, J. Komasa, G. Łach, B. Jeziorski, and K. Szalewicz. Relativistic and quantum electrodynamics effects in the helium pair potential. Physical Review Letters, 104(18):183003, 2010-05. DOI: 10.1103/physrevlett.104.183003.

[34] Stefan Zeller and et al. Imaging the He_2 quantum halo state using a free electron laser. Proceedings of the National Academy of Sciences, 113(51):14651–14655, 2016-12. DOI: 10.1073/pnas.1610688113.

[35] Shina Tan. Energetics of a strongly correlated Fermi gas. Ann. Phys., 323(12): 2952 − 2970, 2008. ISSN 0003-4916. DOI: http://dx.doi.org/10.1016/j.aop.2008.03.004. URL http://www.sciencedirect.com/science/article/pii/S0003491608000456.

[36] Shina Tan. Large momentum part of a strongly correlated Fermi gas. Ann. Phys., 323(12): 2971 − 2986, 2008. ISSN 0003-4916. DOI: http://dx.doi.org/10.1016/j.aop.2008.03.005.

URL   http://www.sciencedirect.com/science/article/pii/S0003491608000432.

[37] Shina Tan. Generalized virial theorem and pressure relation for a strongly correlated Fermi gas. Ann. Phys., 323(12): 2987 – 2990, 2008. ISSN 0003-4916. DOI: http://dx.doi.org/10.1016/j.aop.2008.03.003. URL   http://www.sciencedirect.com/science/article/pii/S0003491608000420.

[38] Gerald A. Miller. Non-universal and universal aspects of the large scattering length limit. Physics Letters B, 777:442–446, 2018-02. DOI: 10.1016/j.physletb.2017.12.063.

[39] Félix Werner and Yvan Castin. General relations for quantum gases in two and three dimensions. II. bosons and mixtures. Physical Review A, 86(5):053633, 2012-11. DOI: 10.1103/physreva.86.053633.

[40] Félix Werner and Yvan Castin. General relations for quantum gases in two and three dimensions: Two-component fermions. Physical Review A, 86(1):013626, 2012-07. DOI: 10.1103/physreva.86.013626.

[41] Yaroslav Lutsyshyn. Weakly parametrized jastrow ansatz for a strongly correlated bose system. J. Chem. Phys., 146(12):124102, Mar 2017. DOI: 10.1063/1.4978707.

[42] S. A. Vitiello and K. E. Schmidt. Optimization of ^{4}He wave functions for the liquid and solid phases. Phys. Rev. B, 46: 5442–5447, Sep 1992. DOI: 10.1103/PhysRevB.46.5442. URL http://link.aps.org/doi/10.1103/PhysRevB.46.5442.