Hidden-nucleons neural-network quantum states for the nuclear many-body problem

Alessandro Lovato [1,2,3] Corey Adams,[1,4] Giuseppe Carleo,[5] and Noemi Rocco[6]

[1]*Physics Division, Argonne National Laboratory, Argonne, Illinois 60439, USA*
[2]*Computational Science Division Division, Argonne National Laboratory, Argonne, Illinois 60439, USA*
[3]*INFN-TIFPA Trento Institute of Fundamental Physics and Applications, 38123 Trento, Italy*
[4]*Leadership Computing Facility, Argonne National Laboratory, Argonne, Illinois 60439, USA*
[5]*Institute of Physics, École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland*
[6]*Theoretical Physics Department, Fermi National Accelerator Laboratory, P.O. Box 500, Batavia, Illinois 60510, USA*

We generalize the hidden-fermion family of neural network quantum states to encompass both continuous and discrete degrees of freedom and solve the nuclear many-body Schrödinger equation in a systematically improvable fashion. We demonstrate that adding hidden nucleons to the original Hilbert space considerably augments the expressivity of the neural-network architecture compared to the Slater-Jastrow ansatz. The benefits of explicitly encoding in the wave function point symmetries such as parity and timereversal are also discussed. Leveraging on improved optimization methods and sampling techniques, the hidden-nucleon ansatz achieves an accuracy comparable to the numericallyexact hyperspherical harmonic method in light nuclei and to the auxiliary field diffusion Monte Carlo in ^{16}O. Thanks to its polynomial scaling with the number of nucleons, this method opens the way to highly-accurate quantum Monte Carlo studies of medium-mass nuclei.

I. INTRODUCTION

Since the early 2000s, the combination of nuclear effective field theories (EFTs) and sophisticated many-body methods has paved the way to a systematic description of atomic nuclei that is well rooted in the underlying theory of strong interactions [1]. In particular, numerical methods based on single-particle basis expansions, such as self-consistent Green's function [2], the in-medium similarity-renormalization group [3], and Coupled Cluster [4] have achieved a formidable success in treating binding energies and radii up to ^{208}Pb [5] and have provided a plausible solution to the quenching puzzle of betadecays [6]. Nevertheless, some prominent challenges remain open. As a chief example, most existing nuclear Hamiltonians cannot simultaneously describe light nuclear systems and the equation of state of infinite nucleonic matter [7,8]. Exhaustive tests of nuclear interactions require many-body methods that are suitable to retain the complexity of nuclear dynamics at short distances [9,10]. An accurate description of the latter is also critical to reproduce exclusive lepton-nucleus scattering cross sections [11–13], for the calculation of neutrinoless double-beta decay matrix elements [14], and for studying neutron-star matter in the high-density regime [8,9,15].

Continuum quantum Monte Carlo (QMC) methods [16] have no difficulties in treating short-range (or high-momentum) components of the nuclear wave function,

but are presently limited to light nuclei, with up to $A = 12$ nucleons. The primary reason preventing QMC calculations of medium-mass and large nuclei is the onset of the fermion-sign problem. Controlling it requires employing sophisticated wavefunctions, whose calculation either scales exponentially with the number of nucleons [17,18] or violates the factorization theorem [19–21]. Therefore, extending QMC calculations to medium-mass nuclei rests on our ability to develop wave functions that capture the vast majority of nuclear correlations while requiring computational time that scales polynomially with A. Developing such wavefunctions is also important to correctly evaluate observables that do not commute with the Hamiltonian—e.g., spatial and momentum distributions, electroweak transition, and responses—as their QMC estimates strongly depend upon the quality of the wave-function ansatz.

After the earliest application to prototypical interacting spins models [22], artificial neural networks (ANNs) have proven to compactly and accurately represent the wave function of a variety of strongly interacting systems [23–25]. The applicability of ANNs in solving the nuclear Schrödinger equation in momentum space has been first demonstrated in Ref. [26], where the groundstate of the deuteron is reproduced with remarkable accuracy by a single-layer ANN—see Ref. [27] for a detailed uncertainty-quantification analysis. Subsequently, an anti-symmetric coordinate-space ansatz defined through the product between a permutation-invariant ANNs Jastrow and a Slater determinant of single-particle orbitals has been utilized in a variational Monte Carlo (VMC) method to solve leading-order pionless-EFT Hamiltonians of $A \leqslant 6$ nuclei [28,29]. Detailed comparisons against Green's function Monte Carlo and the hyperspherical harmonics approaches have validated the expressivity of this ANN Slater-Jastrow (ANN-SJ) ansatz, but also highlighted its

shortcomings, essentially due to the incorrect nodal surface of the Slater determinant.

The authors of Ref. [30] have recently introduced a particularly expressive family of variational wave functions, consisting of augmented Slater determinants involving "hidden" additional fermionic degrees of freedom. This family of wave functions, which has proven to be universal for fermionic lattice systems, has been utilized to study the Hubbard model on the square lattice. In this paper, we generalize this approach to encompass both discrete and continuous degrees of freedom and solve the nuclear many-body problem in a systematically-improvable fashion. We couple the "hidden nucleons" wave function with the VMC method developed in Refs. [28,29] to compute the ground-state energies of ^{3}H, ^{3}He, ^{4}He, and ^{16}O nuclei starting from the pionless-EFT Hamiltonian of Ref. [31]. We note that ^{16}O becomes the largest nucleus studied with neural-network quantum states so far. Both parity and time-reversal symmetries are directly encoded in the variational state to appreciably improve the optimization of the ANN's parameters. The training is further accelerated capitalizing on a novel version of the stochastic reconfiguration [32] method, with a regularization based on RMSProp algorithm [33]. Our results for ^{3}H, ^{3}He, and ^{4}He nuclei are benchmarked against the numerically-exact hyperspherical-harmonics (HH) method [34]. For the larger ^{16}O nucleus, we compare the hidden-nucleon ground-state energies and point-nucleon density with the ones computed with the auxiliary-field diffusion Monte Carlo method [35].

This work is organized as follows. Section II provides an overview of the nuclear Hamiltonian, the hidden-nucleon wave function, and the optimization algorithm. In Sec. III we discuss our results, comparing them with existing hyperspherical-harmonics and QMC nuclear methods. Finally, in Sec. IV we state our conclusions and provide perspectives on future developments.

II. METHODS

To a remarkably large extent, the dynamics of atomic nuclei can be modeled by nonrelativistic Hamiltonians of the form

$$H = -\sum_i \frac{\nabla_i^2}{2m_N} + \sum_{i<j} v_{ij} + \sum_{i<j<k} V_{ijk}, \quad (1)$$

where v_{ij} and V_{ijk} denote the nucleon-nucleon (NN) and three-nucleon $3N$ potentials. In this work, following Refs. [28,29], we employ NN and $3N$ forces derived at leading order in a pionless-EFT expansion that consist of contact terms between nucleons [36]. More specifically, we use the leading-order NN potential "o" of Ref. [31] that is designed to reproduce the np scattering lengths and effective radii in $S/T = 0/1$ and $1/0$ channels. We assume the electromagnetic component of the NN potential to only include the Coulomb repulsion between finite-size protons. The authors of Ref. [31] explored different regulator values for the $3N$ force. Here we take $R_3 = 1.0$ fm because when used in conjunction with model "o", this value of R_3 provides binding energies that are in reasonably good agreement with experiments for various closed-shell nuclei across the nuclear chart.

To fix the notation, we introduce $R = \{\mathbf{r}_1 \dots \mathbf{r}_A\}$ and $S = \{\mathbf{s}_1 \dots \mathbf{s}_A\}$ to indicate the set of single-particle spatial three-dimensional coordinates and the zprojection of the spin-isospin degrees of freedom $\mathbf{s}_i = \{s_i^z, t_i^z\}$ of the A nucleons comprising a given nucleus. To automatically remove spurious center of mass contributions from all observables [37], in our calculations we employ intrinsic spatial coordinates, defined by $\mathbf{r}_i \rightarrow \mathbf{r}_i - \mathbf{R}_{\mathrm{CM}}$, with $\mathbf{R}_{\mathrm{CM}}$ being the center of mass coordinate.

A. Hidden-nucleon wave function

In Ref. [30] the "hidden-fermion" approach has been successfully applied to interacting fermionic hamiltonians with discrete degrees of freedom. To model the ground-state wave functions of atomic nuclei we introduce a "hidden-nucleon" (HN) ansatz that is specifically designed to encompass continuous and discrete coordinates. In addition to the visible coordinates R, S, the Hilbert space also contains fictitious degrees of freedom for the A_h hidden nucleons, which are in turn functions of the visible ones $(R_h, S_h) = f(R, S)$. The amplitudes of the hidden-nucleons wave function ansatz in the R, S basis are schematically given by

$$\Psi_{HN}(R, S) \equiv \det \begin{bmatrix} \phi_v(R, S) & \phi_v(R_h, S_h) \\ \chi_h(R, S) & \chi_h(R_h, S_h) \end{bmatrix}. \quad (2)$$

In the above equation, $\phi_v(R, S)$ denotes the $A \times A$ matrix representing visible single-particle orbitals computed on the visible coordinates—this would be the only component of the wave function in a Hartree-Fock description of the nucleus. Note that, in contrast with Ref. [30], the columns of our matrix denote different particles, while rows refer to different states. The $A_h \times A_h$ matrix $\chi_h(R_h, S_h)$ yields the amplitudes of hidden orbitals evaluated on hidden coordinates. Finally, $\chi_h(R, S)$ and $\phi_v(R_h, S_h)$ are $A_h \times A$ and $A \times A_h$ matrices that provide the amplitudes of hidden orbitals on visible coordinates and those of visible orbitals on hidden coordinates, respectively.

In the limit $A_h = 1$, $\chi_h(R, S) = 0$, and $\phi_v(R_h, S_h) = 0$ we recover the usual Slater-Jastrow formulation, which can be thereby interpreted as a limit of the hidden-nucleon ansatz. If the function f is permutation invariant, it is immediate to prove that $\Psi_{HN}(R, S)$ is anti-symmetric under the exchange of two-particles coordinates. The authors of Ref. [30] provided a mathematical proof of the expressivity of the hidden-fermion wave function for discrete degrees of freedom, such as discrete orbitals or spatial coordinates, provided that the functions χ_h and f are general. In order to bypass the combinatorial nature of the function f, we directly parametrize each column "i" of $\phi_v(R_h, S_h)$ and $\chi_h(R_h, S_h)$ in terms of independent, real-valued, permutation-invariant neural networks as

$$\phi_v^i(R_h, S_h) = e^{\mathcal{U}_\phi^i(R,S)} \tanh \left[\mathcal{V}_\phi^i(R, S) \right],$$

$$\chi_h^i(R_h, S_h) = e^{\mathcal{U}_\chi^i(R,S)} \tanh \left[\mathcal{V}_\chi^i(R, S) \right]. \quad (3)$$

As in Refs. [28,29,38], permutationinvariance is achieved by expressing the functions $\mathcal{U}_\phi^i$, $\mathcal{V}_\phi^i$, $\mathcal{U}_\chi^i$, and $\mathcal{V}_\chi^i$ in terms of the Deep-Sets architectures [39,40]. A number of artificial neural network architectures have been developed to efficiently represent permutation invariant functions, including 3D point cloud [41–43] and attention-based methods [44–46]. In

addition to satisfying the universal approximation theorem for permutation-invariant functions, we found Deep-Sets to be a practical choice for its simplicity and accurate enough for our purposes. Taking pair coordinates as input of the Deep-Sets instead of single-particle ones was found to accelerate the training when the sum pooling is used [29]. Here, we achieve similar performances with single-particle inputs employing the *logsumexp* pooling

$$\mathcal{F}(R, S) = \rho_{\mathcal{F}} \left[\log \left(\sum_i e^{\phi_{\mathcal{F}}(\mathbf{r}_i, \mathbf{s}_i)} \right) \right]. \quad (4)$$

Both $\phi_{\mathcal{F}}$ and $\rho_{\mathcal{F}}$ are dense feed-forward neural networks, comprised of two hidden layers with 16 nodes each. The latent space, i.e., the size of the output of $\phi_{\mathcal{F}}$ and the input of $\rho_{\mathcal{F}}$, is 16-dimensional. The output layers of $\rho_{\mathcal{F}}$ contain A nodes for $\mathcal{F} = \mathcal{U}_\phi^i$ and $\mathcal{F} = \mathcal{V}_\phi^i$, and A_h nodes for $\mathcal{F} = \mathcal{U}_\chi^i$ and $\mathcal{F} = \mathcal{V}_\chi^i$.

The single-particle orbitals defining $\phi_v(R, S)$ and $\chi_h(R, S)$ are represented by dense feed-forward neural networks that take as input the single particle coordinates of the nucleons. They are both comprised of two hidden layers with 10 nodes each, while their outputs are onedimensional with linear activation functions.

To gauge the optimal number of hyperparameters, we analyzed the convergence of the ^{4}He ground-state energy as obtained with a single hidden-nucleon ansatz. We found no significant improvements by using more than two hidden layers with 16 nodes each for ϕ_F and ρ_F and two hidden layers with 10 nodes each for ϕ_v and χ_v. On the other hand, using less than 12 nodes for ϕ_F and ρ_F and less than 8 nodes for ϕ_v and χ_v worsen the energy of ^{4}He by about 0.5 MeV.

The hyperbolic tangent is the activation function chosen for the hidden layers of all the dense feed-forward neural networks used in this work. We tried alternative differentiable functions, such as *Softplus* [47] and *GELU* [48], without finding appreciable differences. Note that our choices are restricted to differentiable activation functions because the calculation of the kinetic energy requires evaluating their second derivatives.

B. Symmetries and sampling

In this work, we enforce pointsymmetries, such as parity and time reversal, into the hidden nucleon ansatz. Since all the atomic nuclei considered in this work have ground-states with positive parity, we can construct such variational states by

$$\Psi_{HN}^P(R, S) \equiv \Psi_{HN}(R, S) + \Psi_{HN}(-R, S). \quad (5)$$

For even-even nuclei, such as ^{4}He and ^{16}O, we can additionally enforce time-reversal symmetry

$$\Psi_{HN}^{PT}(R, S) \equiv \Psi_{HN}^P(R, S) + \Psi_{HN}^P(R, \theta S), \quad (6)$$

where θS is obtained by applying the operator $-i\sigma_y$ to all single-particle spinors [49]. Note that no complex conjugate operation is required in the above definition since $\Psi_{HN}^P(R, S)$ is a real-valued function. As discussed in Sec. III, enforcing pointsymmetries considerably accelerates the training and augments the expressivity of the ANN for a fixed number of hyperparameters. In principle, additional pointsymmetries,

such as "signature", and "simplex" [50], can be imposed in a similar fashion as Eqs. (5) and (6). Their impact on ANN quantum states will be investigated in future works.

The set of variational parameters **p** entering the ANN variational state are optimized minimizing the expectation value of the Hamiltonian $E = \langle\Psi|H|\Psi\rangle/\langle\Psi|\Psi\rangle$. The $3A$ spatial dimensional integrals and the spin-isospin summations are evaluated in a stochastic fashion, using the Metropolis-Hastings sampling algorithm discussed in the supplemental material of Ref. [28]. It has to be noted that the hidden-nucleon ansatz is completely general and does not prevent the Metropolis-Hastings algorithm to sample nonphysical states. To better elucidate this issue, consider a nucleus with a given total isospin projection T_z. Even though nuclear interactions are charge conserving, the hidden-nucleon ansatz yields nonzero amplitudes for states S such that $\sum_i t_i^z \neq T_z$. The latter are avoided by constraining the Metropolis walk onto states with $\sum_i t_z^i = T_z$, which is equivalent to multiplying $\Psi_{HN}(R, S)$ by the Kronecker Delta $\delta(\sum_i t_z^i - T_z)$. Similarly, the LO pionless-EFT Hamiltonian of Eq. (1) does not contain a tensor force and preserves the total spin projection on the $z$axis S_z. As a consequence, to prevent sampling nonphysical states, we restrict the Metropolis walk to S with $\sum_i s_i^z = S_z$. In practice, this is achieved by starting the walk with a state that satisfies the above requirements and sampling the new ones by exchanging the spin and the isospin projections of two, not necessarily different, pairs of nucleons.

C. Optimization

Given the neural-network architecture, and the sampling procedure as described in the previous discussions, we further need to specify a procedure to minimize the variational energy and find the optimal parameters of the model wave function. The stochastic-reconfiguration (SR) algorithm [32], closely related to the natural gradient descent method [51] in unsupervised learning, has proven to efficiently optimize neural network quantum states for a variety of applications, ranging from spin models [22,30,52] to periodic bosonic systems [38] and atomic nuclei [28,29]. The variational parameters are updated as $\mathbf{p}_t = \mathbf{p}_{t-1} - \eta S_{t-1}^{-1}\mathbf{g}_{t-1}$, where η is the learning rate, S is the quantum Fisher-information matrix, and $\mathbf{g} = \partial E/\partial\mathbf{p}$ is the gradient of the energy. The inversion of matrix S is typically stabilized by adding a small positive diagonal matrix $S \to S + \epsilon I$, implying that all diagonal elements are shifted by the same amount, thereby neglecting potential order of magnitudes differences in the parameters' changes [53]. To remedy this shortcoming, inspired by the RMSProp method [33], we accumulate the exponentially-decaying averages of the squared gradients

$$\mathbf{v}_t = \beta\mathbf{v}_{t-1} + (1 - \beta)\mathbf{g}_t^2, \quad (7)$$

and regularize the the Fisher-information matrix by $S \to S + \epsilon \, \text{diag}(\sqrt{\mathbf{v}_t} + 10^{-8})$. In the original formulation of the SR algorithm, taking larger values of ϵ reduces the magnitude of the parameters' update and rotates it towards the stochastic gradient descent direction. In our RMSProp version, the regularization term rotates the update towards the RMSProp direction, which typically yields faster training than the simple stochastic gradient descent. Since the dimension of the Fisher
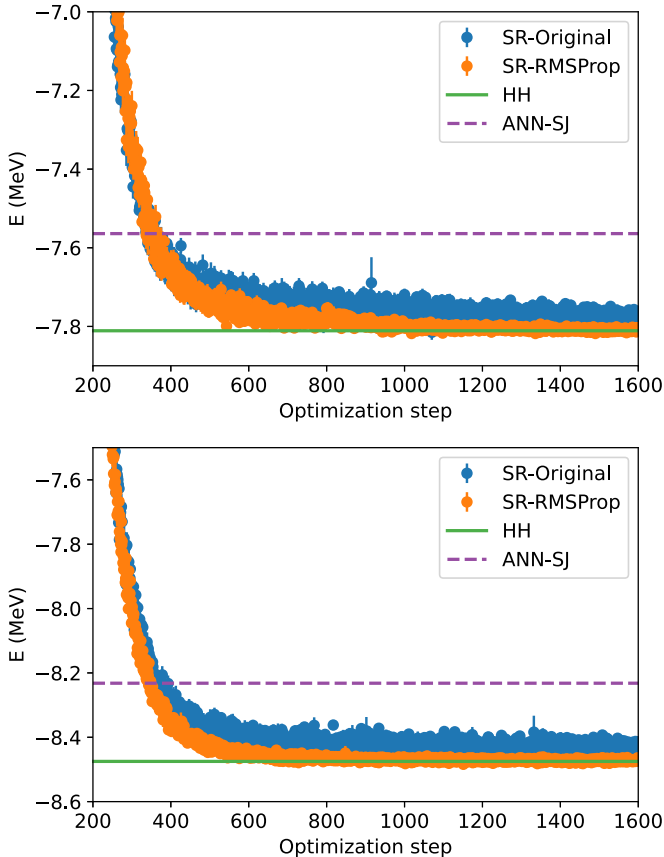
FIG. 1. Convergence of the SR algorithm for ^{3}He (upper panel) and ^{3}H (lower panel) with the original (blue solid circles) and RMSProp-like (orange solid circles) diagonal shifts. The ANN-SJ and the HH energies of Ref. [29] are displayed by the purple dashed and solid green lines, respectively.

FIG. 2. Upper panel: ^{4}He ground-state energy convergence with the ansatz that conserves parity (blue solid circles) and parity plus time reversal (orange solid circles). The ANN-SJ and the HH ground-state energies of Ref. [29] are displayed by the purple dashed and solid green lines, respectively. Lower panel: Convergence of the parity-conserving ansatz utilizing a wider ANN than in the upper panel.

matrix scales quadratically with the number of variational parameters, storing it in memory becomes unfeasible for a nucleus as large as ^{16}O. To overcome this limitation, as in Ref. [54], we use the iterative conjugate-gradient method to solve the linear system associated with the SR parameters' update. Finally, the *Adaptive Epsilon* heuristic scheduler introduced in Ref. [28] has been adopted to determine the best value of the regularization parameter ϵ at each optimization step.

III. RESULTS

We begin our analysis comparing the performances of the original SR method with the version that includes the RMSProp-inspired diagonal shift in the Fisher-information matrix. In Fig. 1, we show the convergence of the ground-state energy of ^{3}He (upper panel) and ^{3}H (lower panel) as obtained with $A_h = 3$ hidden nucleons and the positive-parity ansatz of Eq. (5). The orange solid circles, corresponding to the energies obtained using the RMSProp-like regularization, are noticeably closer to the numerically-exact HH result of Ref. [28] than those obtained with the original version of the SR algorithm, displayed by solid blue circles. The fact that the SR-RMSProp estimates are less scattered that the
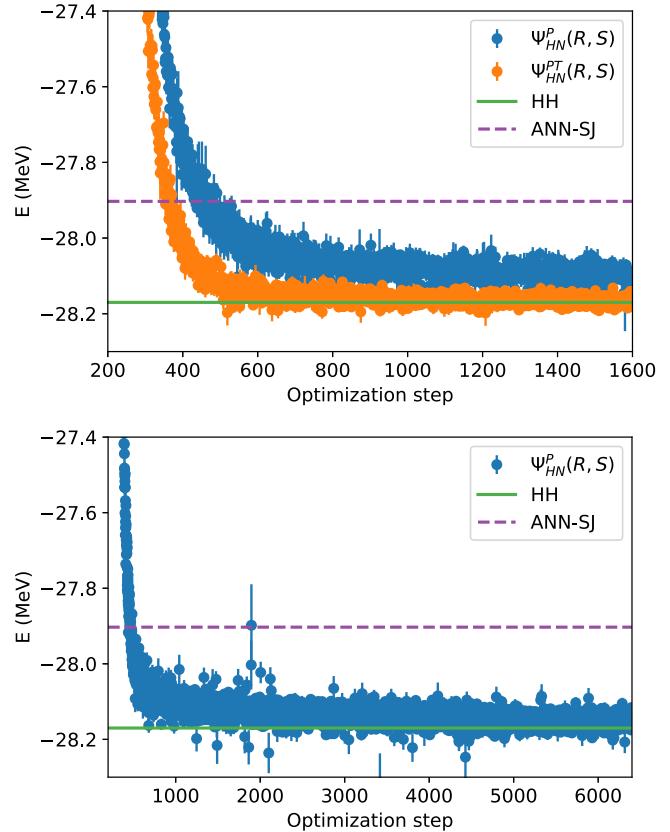
SR ones is another indication of the better minima found by the new version of the algorithm. Most notably, independent of the particular regularization choice, both the SR and SR-RMSProp energies are appreciably lower than the ANN-SJ value reported in Ref. [29].

Because of its superior training performances with respect to the original version of the algorithm, in the remainder of the paper we will only show results obtained with the SR-RMSProp minimizer. The convergence of the ^{4}He ground-state energy computed with $A_h = 4$ hidden nucleons is displayed in Fig. 2. The parity-conserving wave function $\Psi^P_{HN}(R, S)$ is outperformed by $\Psi^{PT}_{HN}(R, S)$, which additionally preserves the time-reversal symmetry. Both of them provide significantly better energies than the original ANN-SJ model, as they can improve the nodal surface of the single-particle Slater determinant. More importantly, $\Psi^{PT}_{HN}(R, S)$ provides a variational energy that is consistent with the numerically-exact HH estimate of Ref. [29]. It has to be noted that $\Psi^P_{HN}(R, S)$ should in principle converge to the exact energy, even with $A_h = 4$ hidden nucleons, but it requires wider (or deeper) ANN architectures. To prove this point, in the lower panel of Fig. 2 we show the training of $\Psi^P_{HN}(R, S)$ with $A_h = 4$ in which the number of nodes in the hidden layers in $\phi_{\mathcal{F}}$

and $\rho_{\mathcal{F}}$ has been increased from 16 to 24. After about 4800 optimization steps, the parity-conserving ansatz yields energies that are consistent with the HH method. Nevertheless, our results indicate that enforcing time-reversal symmetry is effective in reducing the training time and augments the expressivity of the hidden-nucleon ANN architecture.

Neural-network quantum states applications to nuclear systems have so far been limited to light nuclei, with up to $A = 6$ nucleons [26,28,29]. Here, we significantly extend the reach of this method by computing the ground-state of ^{16}O utilizing the hidden-nucleon ansatz. In Ref. [31], the AFDMC method has been employed to study this nucleus using as input the LO pionless-EFT Hamiltonian of Eq. (1). The AFDMC trial wave function takes the factorized form $\Psi_T(R, S) \equiv \langle RS | \mathcal{F} | \Phi \rangle$. The Slater determinant of single-particle orbitals $\Phi(R, S)$ determines the long-range behavior of the wave function. The correlation operator is expressed as

$$\mathcal{F} = \left(\prod_{i<j<k} F_{ijk}^c \right) \left(\prod_{i<j} F_{ij}^c \right) \left(1 + \sum_{i<j} F_{ij}^{op} \right). \qquad (8)$$

The spin-isospin independent three-body correlations F_{ijk}^c act on all triplets of nucleons. Similarly, the central two-body Jastrow F_{ij}^c is applied to all nucleon pairs, while the spin-isospin dependent term, F_{ij}^{op}, appears in a linearized form [55]. This approximation reduces the computational cost of evaluating $\Psi_T(R, S)$ from exponential to polynomial in A but makes the trial wave function nonextensive: if the system is split in two (or more) subsets of particles that are separated from each other, the $\mathcal{F}$ does not factorize into a product of two factors in such a way that only particles belonging to the same subset are correlated. As a consequence, the correlation operator of Eq. (8) becomes less effective for nuclei larger than ^{16}O, preventing the applicability of the AFDMC method to medium-mass nuclei.

The AFDMC projects out the groundstate of the system from the starting trial wave function performing an evolution in imaginary time τ

$$|\Psi_0\rangle \propto \lim_{\tau \to \infty} |\Psi(\tau)\rangle = e^{-H\tau} |\Psi_T\rangle. \qquad (9)$$

The fermion-sign problem is mitigated by means of the constrained-path approximation, which essentially limits the imaginary-time propagation to regions where the propagated and trial wave functions have a positive overlap [16]. Contrary to the fixed-node approximation, the constrained-path approximation does provide an upper bound to the true ground-state energy of the system [56]. The accuracy of the trial wave function is critical to reduce this bias, as the constrained-path approximation becomes exact when the trial wave function coincides with the groundstate one.

In Fig. 3, we display the ground-state energy of ^{16}O as a function of the number of hidden nucleons A_h for the parity and timereversal conserving ansatz of Eq. (6). For comparison, the VMC energy of ^{16}O obtained with the correlation operator of Eq. (8) is represented in Fig. 3 by the dashed green line, while the shaded area is the Monte Carlo statistical uncertainty. The solid horizontal line and the shaded area indicate the constrained-path AFDMC energy and its statistical uncertainty as listed in Ref. [31]. Already for $A_h = 2$,
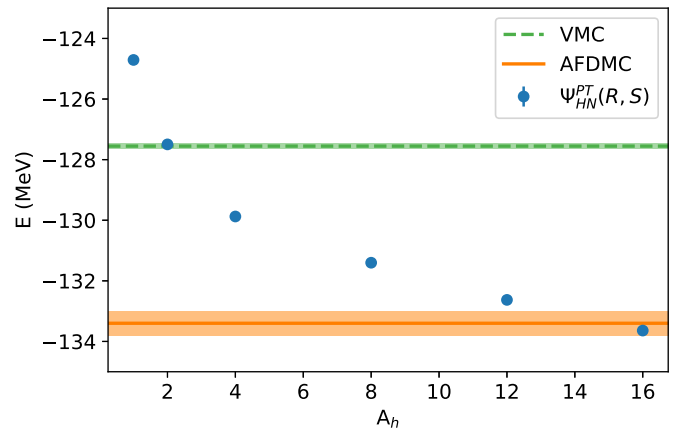

FIG. 3. Ground-state energy of ^{16}O as a function of the number of hidden nucleons A_h (solid blue points). The VMC and AFDMC energies—the latter taken from Ref. [31]—are shown by the green dashed and orange solid lines. The shaded areas represent the Monte Carlo statistical uncertainties.

the hidden-nucleon wave function matches the VMC value. By further increasing A_h, the variational energy lowers until it becomes consistent with the AFDMC value, within error bars, demonstrating the accuracy of the hidden-nucleon ansatz even in the p-shell region.

It has to be noted that the Authors of Ref. [57] showed that the hidden-fermion wave function is a universal approximator for fermionic systems with discrete degrees of freedom even when a single hidden fermion is used ($A_h = 1$ in our case). However, their proof relies on constructing lookup tables and cannot directly be applied to continuous coordinates. On the other hand, it can be shown that an augmented Slater determinant containing A_h hidden states yields A_h-particle and A_h-holes excitations of the original Hartree-Fock state. Hence, we conjecture that when $A_h \geqslant A$, the hidden-nucleon ansatz can also represent any fermionic wave function on the continuum.

Unless a forward-walk propagation is used [58,59], within diffusion Monte Carlo methods, expectation values of operators that do not commute with the Hamiltonian are typically estimated at first order in perturbation theory as

$$\frac{\langle \Psi(\tau) | O | \Psi(\tau) \rangle}{\langle \Psi(\tau) | \Psi(\tau) \rangle} \simeq 2 \frac{\langle \Psi_T | O | \Psi(\tau) \rangle}{\langle \Psi_T | \Psi(\tau) \rangle} - \frac{\langle \Psi_T | O | \Psi_T \rangle}{\langle \Psi_T | \Psi_T \rangle}. \qquad (10)$$

Hence, in addition to controlling the fermion-sign problem, the accuracy of $\Psi_T(R, S)$ is critical to evaluate observables that do not commute with the Hamiltonian, such as density distributions in coordinate and momentum space. In any case, the extrapolated estimator is always biased in a quantity difficult to assess. On the other hand, neural-network quantum states provide pure estimators and no extrapolations are required for computing expectations values of operator that do not commute with the Hamiltonian.

One such operator is the point-nucleon density, which is defined as

$$\rho_N(r) = \frac{1}{4\pi r^2} \langle \Psi_V | \sum_i \delta(r - |\mathbf{r}_i|) | \Psi_V \rangle, \qquad (11)$$
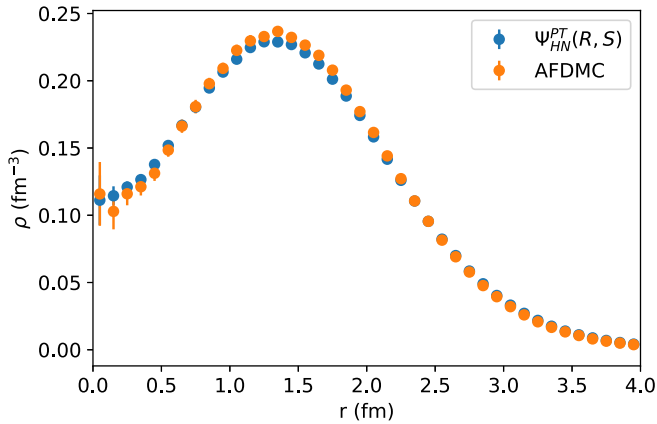
FIG. 4. Point nucleon density of ^{16}O as obtained with the hidden nucleon ansatz (solid blue circles) compared with the perturbatively-corrected AFDMC estimates of Eq. (10).

where again, $\mathbf{r}_i$ is the intrinsic spatial coordinate of the ith nucleon.

In Fig. 4, we display the point-nucleon density of ^{16}O as obtained with $A_h = 16$ hidden nucleons, compared with the AFDMC results, which are obtained in perturbation theory as in Eq. (10). Remarkably, despite the lack of mean-field information encoded in the hidden nucleon anstaz—the single-particle orbitals are randomly-initialized feed-forward neural networks—the training procedure yields a point-nucleon density of ^{16}O that is very close to the AFDMC one. This excellent agreement confirms the accuracy of the hidden-nucleon ansatz in modeling the wave functions of atomic nuclei at both short and long distances.

IV. CONCLUSIONS

We have developed a h idden-nucleon, neural-network ansatz suitable to solve the nuclear many-body Schrödinger equation of a leading-order pionless-EFT Hamiltonian in a systematically improvable fashion. To this aim, we extend the hidden-fermion family of variational states to encompass both continuous and discrete degrees of freedom, corresponding to the spatial and spin-isospin coordinates of the nucleons. Point symmetries such as parity and time reversal are built in the neural-network wave function to augment its expressivity. We have concurrently improved the stochastic-reconfiguration algorithm by introducing a nonconstant diagonal regularization inspired by the RMSProp optimization method.

We first gauge the hidden-nucleon wave function in ^{3}H, ^{3}He, and ^{4}He nuclei, whose ground-state energies turn out in excellent agreement with numerically exact hyperspherical-hamonics results reported in Ref. [29]. In $A = 3$ nuclei, we observe that the RMSProp-inspired diagonal shift considerably accelerates the convergence of the training compared to the default version of the SR method and adopt it for all the subsequent calculations. We observe that enforcing time-reversal symmetry noticeably accelerates the optimization of the ^{4}He ground-state wave function and improves the expressivity of the neural-network ansatz for a fixed number of variational parameters.

We then applied the hidden-nucleon architecture to study the groundstate of ^{16}O, the largest nuclear system yet computed with neural-network quantum states. Even using only two hidden nucleons, we are able to match the VMC energies obtained with conventional spin-isospin dependent variational wavefunctions. By further increasing A_h, the hidden-nucleon architecture gains the capability to reproduce the AFDMC constrained-path energy, thereby proving the expressivity of this ansatz for continuous degrees of freedom.

In addition to the ground-state energy, the neural-network architecture is capable of learning the single-particle density of ^{16}O, at both short and long distances. This is particularly remarkable, as, in contrast with other QMC methods, the hidden-nucleon ansatz is agnostic to the mean-field properties of the specific nucleus of interest. Besides the Hamiltonian of choice, the only information entering the calculation are the number of protons and neutrons, the total spin, and its point symmetries; the short- and long-range components of the ground-state wave functions are learned by minimizing the energy expectation value.

The hidden-nucleon ansatz provides a compact representation of nuclear wave function in terms of a relatively-small number of variational parameters. For instance, the ground-state wave function of ^{16}O with A_h hidden nucleons is completely determined by about 77,000 neural-network parameters. This is an order of magnitude smaller than the number of coefficients required to represent the wave functions by state-of-the-art nuclear many-body methods based on a single-particle basis expansion, such as the no-core shell model [60]. The wave functions computed in this work, as well as the Google-JAX [61] program used to generate them, are publicly available [62]. They can be readily used to compute quantities of experimental interests, without the need of training the ANN.

In principle, generalized backflow transformations similar to those employed to model the groundstate of electronic systems [24,25] can be used to augment the expressivity of the hidden-nucleon architecture. We anticipate using them when dealing with more sophisticated nuclear forces, characterized by strong tensor components. Since the latter do not conserve the total spin projection on the z axis S_z, it may well be beneficial to employ an over-complete spin basis, similar to the one employed by the AFDMC method.

Finally, it has to be noted that the hidden-nucleon approach exhibits a favorable polynomial scaling in computational time with the number of nucleons. The calculation of the wave function requires A^3 operations, while evaluating the expectation value of the spin-isospin dependent component of the NN potential costs A^5 operations. On the other hand, including spin-isospin dependent 3N forces without resorting to density-dependent approximations would make the method to scale as A^6. We envision treating medium-mass nuclei as large as ^{40}Ca by exploiting leadership class hybrid GPU/CPU computers.

[1] H. Hergert, A guided tour of *ab initio* nuclear many-body theory, Front. Phys. **8**, 379 (2020).

[2] V. Somà, Self-consistent Green's function theory for atomic nuclei, Front. Phys. **8**, 340 (2020).

[3] H. Hergert, S. K. Bogner, T. D. Morris, A. Schwenk, and K. Tsukiyama, The in-medium similarity renormalization group: A novel *ab initio* method for nuclei, Phys. Rep. **621**, 165 (2016).

[4] G. Hagen, T. Papenbrock, M. Hjorth-Jensen, and D. J. Dean, Coupled-cluster computations of atomic nuclei, Rep. Prog. Phys. **77**, 096302 (2014).

[5] B. Hu, W. Jiang, T. Miyagi, Z. Sun, A. Ekström, C. Forssén, G. Hagen, J. D. Holt, T. Papenbrock, S. R. Stroberg, and I. Vernon, *Ab initio* predictions link the neutron skin of ^{208}Pb to nuclear forces, arXiv:2112.01125.

[6] P. Gysbers *et al.*, Discrepancy between experimental and theoretical β-decay rates resolved from first principles, Nat. Phys. **15**, 428 (2019).

[7] Y. Nosyk, D. R. Entem, and R. Machleidt, Nucleon-nucleon potentials from Δ-full chiral effective-field-theory and implications, Phys. Rev. C **104**, 054001 (2021).

[8] A. Lovato, I. Bombaci, D. Logoteta, M. Piarulli, and R. B. Wiringa, Benchmark calculations of infinite neutron matter with realistic two- and three-nucleon potentials, Phys. Rev. C **105**, 055808 (2022).

[9] M. Piarulli, I. Bombaci, D. Logoteta, A. Lovato, and R. B. Wiringa, Benchmark calculations of pure neutron matter with realistic nucleon-nucleon interactions, Phys. Rev. C **101**, 045801 (2020).

[10] O. Benhar, Scale dependence of the nucleon–nucleon potential, Int. J. Mod. Phys. E **30**, 2130009 (2021).

[11] M. Duer *et al.* (CLAS Collaboration), Direct Observation of Proton-Neutron Short-Range Correlation Dominance in Heavy Nuclei, Phys. Rev. Lett. **122**, 172502 (2019).

[12] J. R. Pybus, I. Korover, R. Weiss, A. Schmidt, N. Barnea, D. W. Higinbotham, E. Piasetzky, M. Strikman, L. B. Weinstein, and O. Hen, Generalized contact formalism analysis of the 4He(e,e'pN) reaction, Phys. Lett. B **805**, 135429 (2020).

[13] A. Schmidt *et al.* (CLAS Collaboration), Probing the core of the strong nuclear interaction, Nature **578**, 540 (2020).

[14] R. Weiss, P. Soriano, A. Lovato, J. Menendez, and R. B. Wiringa, Neutrinoless double-beta decay: Combining quantum Monte Carlo and the nuclear shell model with the generalized contact formalism, arXiv:2112.08146.

[15] O. Benhar, A. Lovato, and G. Camelio, Modeling neutron star matter in the age of multimessenger astrophysics, arXiv:2205.11183.

[16] J. Carlson, S. Gandolfi, F. Pederiva, S. C. Pieper, R. Schiavilla, K. E. Schmidt, and R. B. Wiringa, Quantum Monte Carlo methods for nuclear physics, Rev. Mod. Phys. **87**, 1067 (2015).

[17] B. S. Pudliner, V. R. Pandharipande, J. Carlson, and R. B. Wiringa, Quantum Monte Carlo Calculations of A $\leqslant$ 6 Nuclei, Phys. Rev. Lett. **74**, 4396 (1995).

[18] B. S. Pudliner, V. R. Pandharipande, J. Carlson, S. C. Pieper, and R. B. Wiringa, Quantum Monte Carlo calculations of nuclei with A $\leqslant$ 7, Phys. Rev. C **56**, 1720 (1997).

[19] D. Lonardoni, J. Carlson, S. Gandolfi, J. E. Lynn, K. E. Schmidt, A. Schwenk, and X. B. Wang, Properties of Nuclei Up to $A = 16$ using Local Chiral Interactions, Phys. Rev. Lett. **120**, 122502 (2018).

[20] D. Lonardoni, S. Gandolfi, J. E. Lynn, C. Petrie, J. Carlson, K. E. Schmidt, and A. Schwenk, Auxiliary field diffusion Monte Carlo calculations of light and medium-mass nuclei with local chiral interactions, Phys. Rev. C **97**, 044318 (2018).

[21] S. J. Novario, D. Lonardoni, S. Gandolfi, and G. Hagen, Trends of neutron skins and radii of mirror nuclei from first principles, arXiv:2111.12775.

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

[23] K. Choo, A. Mezzacapo, and G. Carleo, Fermionic neural-network states for *ab-initio* electronic structure, Nat. Commun. **11**, 2368 (2020).

[24] J. Hermann, Z. Schätzle, and F. Noé, Deep-neural-network solution of the electronic Schrödinger equation, Nat. Chem. **12**, 891 (2020).

[25] D. Pfau, J. S. Spencer, A. G. D. G. Matthews, and W. M. C. Foulkes, *Ab initio* solution of the many-electron Schrödinger equation with deep neural networks, Phys. Rev. Res. **2**, 033429 (2020).

[26] J. W. T. Keeble and A. Rios, Machine learning the deuteron, Phys. Lett. B **809**, 135743 (2020).

[27] J. R. Sarmiento, J. W. T. Keeble, and A. Rios, Machine learning the deuteron: new architectures and uncertainty quantification, arXiv:2205.12795.

[28] C. Adams, G. Carleo, A. Lovato, and N. Rocco, Variational Monte Carlo Calculations of A$\leqslant$4 Nuclei with an Artificial Neural-Network Correlator Ansatz, Phys. Rev. Lett. **127**, 022502 (2021).

[29] A. Gnech, C. Adams, N. Brawand, G. Carleo, A. Lovato, and N. Rocco, Nuclei with up to $A = 6$ nucleons with artificial neural network wave functions, Few-Body Syst. **63**, 7 (2022).

[30] J. Robledo Moreno, G. Carleo, A. Georges, and J. Stokes, Fermionic wave functions from neural-network constrained hidden states, arXiv:2111.10420.

[31] R. Schiavilla, L. Girlanda, A. Gnech, A. Kievsky, A. Lovato, L. E. Marcucci, M. Piarulli, and M. Viviani, Two- and three-nucleon contact interactions and ground-state energies of light- and medium-mass nuclei, Phys. Rev. C **103**, 054003 (2021).

[32] S. Sorella, Wave function optimization in the variational Monte Carlo method, Phys. Rev. B **71**, 241103(R) (2005).

[33] G. H. Tijmen Tieleman, Lecture 6.5-RMSProp, Coursera Neural Networks for Machine Learning Lecture 6 (2018).

[34] A. Kievsky, S. Rosati, M. Viviani, L. E. Marcucci, and L. Girlanda, A High-precision variational approach to three- and four-nucleon bound and zero-energy scattering states, J. Phys. G: Nucl. Part. Phys. **35**, 063101 (2008).

[35] K. E. Schmidt and S. Fantoni, A quantum Monte Carlo method for nucleon systems, Phys. Lett. B **446**, 99 (1999).

[36] P. F. Bedaque and U. van Kolck, Effective field theory for few nucleon systems, Annu. Rev. Nucl. Part. Sci. **52**, 339 (2002).

[37] P. Massella, F. Barranco, D. Lonardoni, A. Lovato, F. Pederiva, and E. Vigezzi, Exact restoration of Galilei invariance in density functional calculations with quantum Monte Carlo, J. Phys. G: Nucl. Part. Phys. **47**, 035105 (2020).

[38] G. Pescia, J. Han, A. Lovato, J. Lu, and G. Carleo, Neural-Network Quantum States for Periodic Systems in Continuous Space, Phys. Rev. Res. **4**, 023138 (2022).

[39] M. Zaheer, S. Kottur, S. Ravanbakhsh, B. Poczos, R. Salakhutdinov, and A. Smola, Deep sets, arXiv:1703.06114.

[40] E. Wagstaff, F. B. Fuchs, M. Engelcke, I. Posner, and M. Osborne, On the limitations of representing functions on sets, arXiv:1901.09006.

[41] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, PointNet: Deep learning on point sets for 3D classification and segmentation, arXiv:1612.00593.

[42] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, PointNet++: Deep hierarchical feature learning on point sets in a metric space, arXiv:1706.02413.

[43] C. R. Qi, W. Liu, C. Wu, H. Su, and L. J. Guibas, Frustum PointNets for 3D object detection from RGB-D data, arXiv:1711.08488.

[44] O. Vinyals, S. Bengio, and M. Kudlur, Order matters: Sequence to sequence for sets, arXiv:1511.06391.

[45] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, Attention is all you need, arXiv:1706.03762.

[46] J. Lee, Y. Lee, J. Kim, A. R. Kosiorek, S. Choi, and Y. Whye Teh, Set transformer: A framework for attention-based permutation-invariant neural networks, arXiv:1810.00825.

[47] C. Dugas, Y. Bengio, F. Bélisle, C. Nadeau, and R. Garcia, Incorporating second-order functional knowledge for better option pricing, in *Advances in Neural Information Processing Systems 13*, edited by T. K. Leen, T. G. Dietterich, and V. Tresp (MIT Press, Cambridge, 2001), pp. 472–478.

[48] D. Hendrycks and K. Gimpel, Gaussian error linear units (GELUs), arXiv:1606.08415.

[49] A. Bohr and B. R. Mottelson, *Nuclear Structure*, Nuclear Structure Vol. 1 (World Scientific, Singapore, 1998).

[50] J. Dobaczewski, J. Dudek, S. G. Rohozinski, and T. R. Werner, Point symmetries in the Hartree-Fock approach: Densities, shapes and currents, Phys. Rev. C **62**, 014310 (2000).

[51] Shun-ichi Amari, Natural gradient works efficiently in learning, Neural Comput. **10**, 251 (1998).

[52] M. Bukov, M. Schmitt, and M. Dupont, Learning the ground state of a non-stoquastic quantum Hamiltonian in a rugged neural network landscape, SciPost Phys. **10**, 147 (2021).

[53] S. Sorella, M. Casula, and D. Rocca, Weak binding between two aromatic rings: Feeling the van der waals attraction by quantum monte carlo methods, J. Chem. Phys. **127**, 014105 (2007).

[54] E. Neuscamman, C. J. Umrigar, and G. K.-L. Chan, Optimizing large parameter sets in variational quantum Monte Carlo, Phys. Rev. B **85**, 045103 (2012).

[55] S. Gandolfi, A. Lovato, J. Carlson, and K. E. Schmidt, From the lightest nuclei to the equation of state of asymmetric nuclear matter with realistic nuclear interactions, Phys. Rev. C **90**, 061306 (2014).

[56] R. B. Wiringa, S. C. Pieper, J. Carlson, and V. R. Pandharipande, Quantum Monte Carlo calculations of A = 8 nuclei, Phys. Rev. C **62**, 014001 (2000).

[57] J. Robledo Moreno, G. Carleo, A. Georges, and J. Stokes, Fermionic wave functions from neural-network constrained hidden states, Proc. Natl. Acad. Sci. USA **119**, e2122059119 (2022).

[58] S. Zhang and M. H. Kalos, Bilinear quantum Monte Carlo: Expectations and energy differences, J. Stat. Phys. **70**, 515 (1993).

[59] J. Casulleras and J. Boronat, Unbiased estimators in quantum Monte Carlo methods: Application to liquid ^{4}He, Phys. Rev. B **52**, 3654 (1995).

[60] B. R. Barrett, P. Navratil, and J. P. Vary, *Ab initio* no core shell model, Prog. Part. Nucl. Phys. **69**, 131 (2013).

[61] J. Bradbury, R. Frostig, P. Hawkins, M. J. Johnson, C. Leary, D. Maclaurin, and S. Wanderman-Milne, JAX: Composable transformations of Python+NumPy programs (2018).

[62] Nuclear Physics with Machine Learning, https://github.com/Nuclear-Physics-with-Machine-Learning/JAX_QMC.