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

Alessandro Lovato,1, 2 Corey Adams,1, 3 Giuseppe Carleo,4 and Noemi Rocco5
1Physics Division, Argonne National Laboratory, Argonne, IL 60439
2INFN-TIFPA Trento Institute of Fundamental Physics and Applications, 38123 Trento, Italy
3Leadership Computing Facility, Argonne National Laboratory, Argonne, IL 60439
4Institute of Physics, École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland
5Theoretical Physics Department, Fermi National Accelerator Laboratory, P.O. Box 500, Batavia, Illinois 60510, USA
(Dated: June 22, 2022)

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 time-reversal are also discussed. Leveraging on improved optimization methods and sampling techniques, the hidden-nucleon ansatz achieves an accuracy comparable to the numerically-exact hyperspherical harmonic method in light nuclei and to the auxiliary field diffusion Monte Carlo in 160O. 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 208Pb [5] and have provided a plausible solution to the quenching puzzle of beta-decays [6]. Nevertheless, some prominent challenges remain open. As a chief example, most of 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 the product between a permutation-invariant ANNs Jastrow ansatz, for the calculation 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 wave-functions, 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 wave-functions 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 ground-state 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 ≤ 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 an extremely 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 neutral-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, $^4$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 AFDMC.

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

\[ H = - \sum_i \frac{\nabla_i^2}{2m} + \sum_{i<j} v_{ij} + \sum_{i<j<k} V_{ijk}, \]

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 35. 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 $R = \{r_1, \ldots, r_A\}$ and $S = \{s_1, \ldots, s_A\}$ to indicate the set of single-particle spatial three-dimensional coordinates and the z-projection of the spin-isospin degrees of freedom $s_i = \{s_i^z, t_i\}$ of the $A$ nucleons comprising a given nucleus.

A. Hidden-nucleon wave function

Generalizing the “hidden-fermion” approach of Ref. 29 to encompass continuous and discrete coordinates, we introduce a “hidden-nucleon” (HN) ansatz to model the ground-state wave functions of atomic nuclei. 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

\[ \Psi_{HN}(R, S) \equiv \det \begin{bmatrix} \phi_v(R, S) & \phi_v(R_h, S_h) \\ \chi_h(R, S) & \chi_h(R_h, S_h) \end{bmatrix} \]  

In the above equation, $\phi_v(R, S)$ denotes the $A \times A$ matrix representing visible single-particle orbitals evaluated 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 $\chi_h(f(R, S)) = 0$ 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 have demonstrated that the hidden-fermion ansatz can represent any anti-symmetric functions on the lattice, provided that the functions $\chi$ 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

\[ \phi_v^i(R_h, S_h) = e^{U_v^i(R_h, S_h)} \tanh[V_v^i(R, S)] \]

\[ \chi_h^i(R_h, S_h) = e^{U_h^i(R_h, S_h)} \tanh[V_h^i(R, S)] . \]

As in Refs. 28, 29, 36, permutation-invariance is achieved by expressing the functions $U_v^i, V_v^i, U_h^i, V_h^i$ in terms of the Deep-SetS architectures 27, 38. Taking pair coordinates as input 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 $\log \sum \exp$ pooling

\[ F(R, S) = \rho_F \left[ \log \left( \sum_i e^{\phi_v(r_i, s_i)} \right) \right] \]
Both $\phi_F$ and $\rho_F$ 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 $\phi_F$ and the input of $\rho_F$, is 16-dimensional. The output layers of $\rho_F$ contain $A$ nodes for $F = U_0^i$ and $F = V_0^i$ and $A_h$ nodes for $F = U_h^i$ and $F = V_h^i$.

The single-particle orbitals defining $\phi_c(R, S)$ 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 one-dimensional with linear activation functions.

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 [39] and GELU [40], 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 point-symmetries, 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

$$
\Psi_{HN}^p(R, S) = \Psi_{HN}(R, S) + \Psi_{HN}(-R, S).
$$

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

$$
\Psi_{HN}^{PT}(R, S) = \Psi_{HN}^p(R, S) + \Psi_{HN}^p(R, \theta S),
$$

where $\theta S$ is obtained by applying the operator $-i\sigma_y$ to all single-particle spinors [11]. 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. [11], enforcing point-symmetries considerably accelerates the training and augments the expressivity of the ANN for a fixed number of hyperparameters. In principle, additional point-symmetries, such as “signature”, and “simplex” [12], 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 non-zero 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 $\sum_i t_i^z = T_z$, which is equivalent to multiplying $\Psi_{HN}(R, S)$ by the Kronecker Delta $\delta(\sum_i t_i^z - T_z)$. 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 [43] in unsupervised learning, has proven to efficiently optimize neural network quantum states for a variety of applications, ranging from spin models [22, 30, 44] to periodic bosonic systems [36] and atomic nuclei [28, 29]. The variational parameters are updated as $p_i = p_{i-1} - \eta S_i^{-1} g_{i-1}$, where $\eta$ is the learning rate, $S$ is the quantum Fisher-information matrix, and $g = \partial E / \partial p$ is the gradient of the energy. The inversion of matrix $S$ is typically stabilized by adding a small positive diagonal matrix $S \rightarrow S + \epsilon I$, implying that all diagonal elements are shifted by the same amount, thereby neglecting potential order of magnitudes differences in the parameters’ changes [45]. To remedy this shortcoming, inspired by the RMSProp method [33], we accumulate the exponentially-decaying averages of the squared gradients

$$
v_t = \beta v_{t-1} + (1 - \beta) g_t^2
$$

and regularize the the Fisher-information matrix by $S \rightarrow S + \epsilon \text{diag}(\sqrt{v_t} + 10^{-8})$. In the original formulation of the SR algorithm, taking larger values of $\epsilon$ 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 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 Refs. [46], 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 $\epsilon$ 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 blue solid circles, corresponding to the energies obtained using the RMSProp-like regularization, are noticeably closer to the numerically-exact HH result of Ref. [29] than those obtained with the original version of the SR algorithm. The fact that the SR-RMSProp estimates are less scattered that the SR ones is another indication of the better minima found by 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 in displayed in Fig. 2. The parity-conserving wave function $\Psi_{HN}^P(R,S)$ is outperformed by $\Psi_{HN}^{PT}(R,S)$, 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, $\Psi_{HN}^{PT}(R,S)$ provides a variational energy that is consistent with the numerically-exact HH estimate of Ref. [29]. It has to be noted that $\Psi_{HN}^P(R,S)$ should in principle converge to the exact energy, even with $A_h = 4$ hidden nucleons.

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.
but it requires wider (or deeper) ANN architectures. To prove this point, in the lower panel of Fig. 4 we show the training of $\Psi_{HN}^\tau(R,S)$ with $A_h = 4$ in which the number of nodes in the hidden layers in $\phi_F$ and $\rho_F$ have 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 augment 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 \cite{26,28,29}. Here, we significantly extend the reach of this methods by computing the ground-state of $^{16}$O utilizing the hidden-nucleon ansatz. In Ref. \cite{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 $\Psi_T(R,S) = \langle RS| F|\Phi\rangle$. 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

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

(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 \cite{17}. This approximation reduces the computational cost of evaluating $\Psi_T(R,S)$ from exponential to polynomial in $A$ but makes the trial wave function non-extensive: if the system is split in two (or more) subsets of particles that are separated from each other, the $F$ 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 ground-state of the system from the starting trial wave function performing an evolution in imaginary time $\tau$

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

(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 \cite{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 \cite{16}. 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 is coincides with the ground-state one.

In Fig. 3, we display the point-nucleon density 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. \cite{31} — are shown by the green-dashed and orange solid lines. The shaded areas represent the Monte Carlo statistical uncertainties.

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

(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.

In Fig. 4, we display the point-nucleon density of $^{16}$O as obtained with $A_h = 16$ hidden nucleons, compared
We observe that enforcing time-reversal symmetry no-
method and adopt it for all the subsequent calculations.

The training compared to the default version of the SR
agonal shift considerably accelerates the convergence of
A hyperspherical-hamonics results reported in Ref. [29]. In
turn out in excellent agreement with numerically exact
3mization method.

We have concurrently improved the stochastic-
reconfiguration algorithm by introducing a non-constant
sivity. We anticipate
sional states to encompass both continuous and dis-
imals such as parity and time reversal are built in
the hidden nu-
3 nucleons without resorting to density-dependent

tions. On the other hand, including spin-isospin depen-
tion 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 ex-
pection value.

The hidden-nucleon ansatz provides a compact repre-
sentation of nuclear wave function in terms of a relatively-
small number of variational parameters. For instance,
the ground-state wave function of 16O with Ah hidden
ucleons is completely determined by about 77,000 neural-
network parameters. This is an order of magnitudes
smaller than the number of coefficients required to repre-
sent 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 [51]. The wave functions
computed in this work, as well as the Google-JAX pro-
gram used to generate them are publicly available [52].
They can be readily used to compute quantities of exper-
imental interests, without the need of training the ANN.

In principle, generalized backflow transformations sim-
lar to those employed to model the ground-state of elec-
tronic systems [24, 25] can be used to augment the expres-
sivity of the hidden-nucleon architecture. We anticipate
using them when dealing with more sophisticated nu-
clear 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 employed by the
AFDMC method.

Finally, it has to be noted that the hidden-nucleon ap-
proach exhibit a favorable polynomial scaling in computa-
tional time with the number of nucleons. The calcula-
tion of the wave function requires $A^3$ operations, while
evaluating the expectation value of the spin-isospin de-
pendent component of the NN potential costs $A^2$ op-
ations. On the other hand, including spin-isospin depend-
ent 3N forces without resorting to density-dependent

VI.

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].

We then applied the hidden-nucleon architecture to
study the ground-state 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 wave-functions. By further in-
creasing $A_h$, 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 inter-
est. Besides the Hamiltonian of choice, the only informa-
tion 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 ex-
pection value.

The hidden-nucleon ansatz provides a compact repre-
sentation 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
ucleons is completely determined by about 77,000 neural-
network parameters. This is an order of magnitudes
smaller than the number of coefficients required to repre-
sent 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 [51]. The wave functions
computed in this work, as well as the Google-JAX pro-
gram used to generate them are publicly available [52].
They can be readily used to compute quantities of exper-
imental interests, without the need of training the ANN.

In principle, generalized backflow transformations sim-
lar to those employed to model the ground-state of elec-
tronic systems [24, 25] can be used to augment the expres-
sivity of the hidden-nucleon architecture. We anticipate
using them when dealing with more sophisticated nu-
clear 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 employed by the
AFDMC method.

Finally, it has to be noted that the hidden-nucleon ap-
proach exhibit a favorable polynomial scaling in computa-
tional time with the number of nucleons. The calcula-
tion of the wave function requires $A^3$ operations, while
evaluating the expectation value of the spin-isospin de-
pendent component of the NN potential costs $A^2$ op-
ations. On the other hand, including spin-isospin depend-
ent 3N forces without resorting to density-dependent

IV. CONCLUSIONS

We have developed a novel, hidden-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 vari-
ational states to encompass both continuous and dis-
crete degrees of freedom, corresponding to the spatial
and spin-isospin coordinates of the nucleons. Point sym-
metries such as parity and time reversal are built in
the neural-network wave function to augment its expres-
sivity. We have concurrently improved the stochastic-
reconfiguration algorithm by introducing a non-constant
diagonal regularization inspired by the RMSProp optimi-
mation 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 di-
agonal 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 no-
ticeably 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 ground-state 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 wave-functions. By further in-
creasing $A_h$, 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 inter-
est. Besides the Hamiltonian of choice, the only informa-
tion 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 ex-
pection value.

The hidden-nucleon ansatz provides a compact repre-
sentation 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
ucleons is completely determined by about 77,000 neural-
network parameters. This is an order of magnitudes
smaller than the number of coefficients required to repre-
sent 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 [51]. The wave functions
computed in this work, as well as the Google-JAX pro-
gram used to generate them are publicly available [52].
They can be readily used to compute quantities of exper-
imental interests, without the need of training the ANN.

In principle, generalized backflow transformations sim-
ilar to those employed to model the ground-state of elec-
tronic systems [24, 25] can be used to augment the expres-
sivity of the hidden-nucleon architecture. We anticipate
using them when dealing with more sophisticated nu-
clear 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 employed by the
AFDMC method.

Finally, it has to be noted that the hidden-nucleon ap-
proach exhibit a favorable polynomial scaling in computa-
tional time with the number of nucleons. The calcula-
tion of the wave function requires $A^3$ operations, while
evaluating the expectation value of the spin-isospin de-
pendent component of the NN potential costs $A^2$ op-
ations. On the other hand, including spin-isospin depend-
ent 3N forces without resorting to density-dependent

VI.
approximations would make the method to scale as \( A^4 \). We envision treating medium-mass nuclei as large as \(^{40}\text{Ca}\) by exploiting leadership class hybrid GPU/CPU computers.

V. ACKNOWLEDGMENTS

Enlightening discussions with K. E. Schmidt and F. Vicentini are gratefully acknowledged. The present research is supported by the U.S. Department of Energy, Office of Science, Office of Nuclear Physics, under contracts DE-AC02-06CH11357, by the NUCLEI SciDAC program (A.L.), the DOE Early Career Research Program, and Argonne LDRD awards, and by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the U.S. Department of Energy, Office of Science, Office of High Energy Physics (N.R.). This research used resources of the Argonne Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract DE-AC02-06CH11357 and the Laboratory Computing Resource Center of Argonne National Laboratory. Part of the numerical calculations presented in this work were carried out through a CINECA-INFN agreement that provides access to resources on MARCONI100 at CINECA.

[1] H. Hergert, “A Guided Tour of \textit{ab initio} Nuclear Many-Body Theory,” Front. in Phys. 8, 379 (2020), arXiv:2008.05061 [nucl-th].
[2] Vittorio Somà, “Self-consistent Green’s function theory for atomic nuclei,” Front. in Phys. 8, 340 (2020), arXiv:2003.11321 [nucl-th].
[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. Rept. 621, 165–222 (2016), arXiv:1512.06956 [nucl-th].
[4] G. Hagen, T. Papenbrock, M. Hjorth-Jensen, and D.J. Dean, “Coupled-cluster computations of atomic nuclei,” Rept. Prog. Phys. 77, 096302 (2014), arXiv:1312.7872 [nucl-th].
[5] Baishan Hu et al., “\textit{Ab initio} predictions link the neutron skin of \(^{206}\text{Pb}\) to nuclear forces,” (2021), arXiv:2112.01125 [nucl-th].
[6] P. Gysbers et al., “Discrepancy between experimental and theoretical \(\beta\)-decay rates resolved from first principles,” Nature Phys. 15, 428–431 (2019), arXiv:1903.00047 [nucl-th].
[7] Y. Nosyk, D. R. Entem, and R. Machleidt, “Nucleon-nucleon potentials from \(\Delta\)-full chiral effective-field-theory and implications,” Phys. Rev. C 104, 054001 (2021), arXiv:2107.06452 [nucl-th].
[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,” (2022), arXiv:2202.10293 [nucl-th].
[9] M. Piarulli, I. Bombaci, D. Logoteta, A. Lovato, and R.B. Wiringa, “Benchmark calculations of pure neutron matter with realistic nuclear interactions,” Phys. Rev. C 101, 045801 (2020), arXiv:1908.04426 [nucl-th].
[10] Omar Benhar, “Scale dependence of the nucleon–nucleon potential,” Int. J. Mod. Phys. E 30, 2130009 (2021).
[11] M. Duer et al. (CLAS), “Direct Observation of Proton-Neutron Short-Range Correlation Dominance in Heavy Nuclei,” Phys. Rev. Lett. 122, 172502 (2019), arXiv:1810.05343 [nucl-ex].
[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 \(4\text{He}(e,e'pN)\) reaction,” Phys. Lett. B 805, 135429 (2020), arXiv:2003.02318 [nucl-th].
[13] A. Schmidt et al. (CLAS), “Probing the core of the strong nuclear interaction,” Nature 578, 540–544 (2020), arXiv:2004.11221 [nucl-ex].
[14] Ronen Weiss, Pablo Soriano, Alessandro Lovato, Javiw Menendez, and R. B. Wiringa, “Neutrinoless double-beta decay: combining quantum Monte Carlo and the nuclear shell model with the generalized contact formalism,” (2021), arXiv:2112.08146 [nucl-th].
[15] Omar Benhar, Alessandro Lovato, and Giovanni Camellio, “Modeling Neutron Star Matter in the Age of Multimesenger Astrophysics,” (2022), arXiv:2205.11183 [nucl-th].
[16] J. Carlson, S. Gandolfi, F. Pederiva, Steven C. Pieper, R. Schiavilla, K.E. Schmidt, and R.B. Wiringa, “Quantum Monte Carlo methods for nuclear physics,” Rev. Mod. Phys. 87, 1067 (2015), arXiv:1412.3081 [nucl-th].
[17] B. S. Pudliner, V. R. Pandharipande, J. Carlson, and Robert B. Wiringa, “Quantum Monte Carlo calculations of \(A <= 6\) nuclei,” Phys. Rev. Lett. 74, 4396–4399 (1995), arXiv:nucl-th/9502031.
[18] B. S. Pudliner, V. R. Pandharipande, J. Carlson, Steven C. Pieper, and Robert B. Wiringa, “Quantum Monte Carlo calculations of nuclei with \(A <= 7\),” Phys. Rev. C 56, 1720–1750 (1997), arXiv:nucl-th/9705009.
[19] D. Lonardoni, J. Carlson, S. Gandolfi, J. E. Lynn, K. E. Schmidt, A. Schwenk, and X. Wang, “Properties of nuclei up to \(A = 16\) using local chiral interactions,” Phys. Rev. Lett. 120, 122502 (2018), arXiv:1709.09143 [nucl-th].
[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), arXiv:1802.08932 [nucl-th].
[21] S. J. Novario, D. Lonardoni, S. Gandolfi, and G. Hagen, “Trends of neutron skins and radii of mirror nuclei from first principles,” (2021), arXiv:2111.12775 [nucl-th].
[22] Giuseppe Carleo and Matthias Troyer, “Solving the quantum many-body problem with artificial neural networks,” Science 355, 602–606 (2017).
[23] Kenny Choo, Antonio Mezzacapo, and Giuseppe Carleo, “Fermionic neural-network states for ab-initio electronic structure,” Nature Communications 11, 2368 (2020).
[24] Jan Hermann, Zeno Schätzle, and Frank Nor, “Deep-neural-network solution of the electronic Schrödinger
