Solving the nuclear pairing model with neural network quantum states

Mauro Rigo [●],[1] Benjamin Hall,[2,3] Morten Hjorth-Jensen [●],[2,4] Alessandro Lovato [●],[3,5,6] and Francesco Pederiva [●][1,6]

[1]*Physics Department, University of Trento, via Sommarive 14, I-38123 Trento, Italy*

[2]*Department of Physics and Astronomy and Facility for Rare Isotope Beams, Michigan State University, East Lansing, Michigan 48824, USA*

[3]*Physics Division, Argonne National Laboratory, Argonne, Illinois 60439, USA*

[4]*Department of Physics and Center for Computing in Science Education, University of Oslo, N-0316 Oslo, Norway*

[5]*Computational Science Division, Argonne National Laboratory, Argonne, Illinois 60439, USA*

[6]*INFN-TIFPA Trento Institute for Fundamental Physics and Applications, Via Sommarive, 14, 38123 Trento, Italy*

We present a variational Monte Carlo method that solves the nuclear many-body problem in the occupation number formalism exploiting an artificial neural network representation of the ground-state wave function. A memory-efficient version of the stochastic reconfiguration algorithm is developed to train the network by minimizing the expectation value of the Hamiltonian. We benchmark this approach against widely used nuclear many-body methods by solving a model used to describe pairing in nuclei for different types of interaction and different values of the interaction strength. Despite its polynomial computational cost, our method outperforms coupled-cluster and provides energies that are in excellent agreement with the numerically exact full configuration-interaction values.

I. INTRODUCTION

The nuclear many-body problem entails nontrivial challenges, primarily due to the nonperturbative nature and the strong spin-isospin dependence of realistic nuclear forces. Solving it has been the springboard for the development of sophisticated numerical methods that are designed to capitalize on each generation of high-performance computing resources [1–4]. Among them, continuum quantum Monte Carlo approaches have proven extremely accurate in modeling nuclear dynamics at short and long range, but are either limited to relatively small systems or to somewhat simplified nuclear potentials [5]. On the other hand, although methods based on single-particle basis expansions can treat nuclei up to ^{208}Pb—including their binding energies, radii, and electroweak transitions—in terms of the individual interactions among their constituents [6–8], they are not ideally suited to model the high-momentum components of the nuclear wave function. Therefore, a comprehensive description of short- and long-range nuclear dynamics for medium-mass and heavy nuclei has not been achieved using available nuclear many-body methods.

Since their pioneering application to interacting spin models [9], neural network quantum states (NQSs) have seen widespread and successful applications to quantum chemistry [10–12] and condensed-matter [13,14] problems. When solving many-particle systems of fermions, NQSs must respect the Pauli exclusion principle; in the first quantization formalism, this constraint is imposed by using representations that are antisymmetric by construction, such as the Slater-Jastrow [15], the generalized backflow [11,12], and the hidden-fermion [16] *Ansätze*. On the other hand, within the second quantization formalism, the Pauli principle is automatically taken care of,

thereby allowing one to use simpler NQS architectures, including restricted Boltzmann machines [10,17].

Within low-energy nuclear physics, from their initial applications to solving the deuteron in momentum space [18,19], NQSs have been subsequently combined with variational Monte Carlo (VMC) techniques to approximately solve $A \leqslant 4$ [20] and $A \leqslant 6$ nuclei [21]. Most recently, the hidden-nucleon architecture has been proposed to overcome the limitations of the Slater-Jastrow *Ansatz* and successfully applied to nuclei up to ^{16}O [22]. Nuclear physics applications of NQSs have so far been limited to the first-quantization formalism. However, the possibility of using nondiverging effective nuclear forces at short distances makes NQSs in the occupation-number formalism ideally suited to tackle the nuclear many-body problem.

In this work, we put forward a novel VMC method suitable for solving the nuclear many-body problem using NQS to compactly represent the ground state of the system in the occupation-number formalism. As a first application, we consider different versions of the nuclear pairing model, which is one of the most important ingredients of the effective nuclear interaction, as recognized in the early work of Bohr, Mottelson, and Pines [23]. Pairing plays an important role in the calculations of a number of properties that depend on the low-energy microscopic structure of the nucleus, including their ground-state energies, low-lying excitations, level densities, odd-even staggering effects, single-particle occupancies, and electromagnetic transition rates. Pairing is also relevant in our understanding of nuclear reactions and particle decays [24,25].

Solving the pairing Hamiltonian, even in the simple case of constant pairing strength, involves nontrivial difficulties. Exact diagonalization techniques [26–29] and particle-number

conserved-seniority approaches [30,31] suffer from growing exponential complexity and are therefore very hard to use for large systems. However, different families of integrable pairing models were recognized after the pioneering work of Richardson [32,33] and are now known as Richardson-Gaudin models [34]. As an example of their physical relevance, the authors of Ref. [35] found that a separable pairing interaction and nondegenerate single-particle energies with two free parameters provides an excellent approximation to the Gogny pairing. It has to be noted that the energies of the Richardson-Gaudin models are found by solving a set of coupled nonlinear equations, which by itself is a complicated problem [36] and it is still an area of active research. In this regard, the number-projected BCS approximation has proven to provide accurate results for superconducting properties of ultrasmall Al grains [37,38].

In this work, we compare the VMC-NQS method with many-body perturbation theory, coupled-cluster calculations—both widely used in *ab initio* studies of nuclear systems—exact-diagonalization techniques, and iterative solutions of the Richardson-Gaudin equations. We consider different values of the interaction strength, and both constant and separable couplings.

This work is organized as follows: In Sec. II we discuss the different versions of the pairing model employed in this work and their exact solutions. In Sec. III we introduce the VMC method based on NQSs. Section IV is dedicated to coupled-cluster theory, with particular emphasis to pair coupled cluster doubles theory. Section V describes a many-body perturbation theory approach. In Sec. VI we present our results and concluding remarks are given in Sec. VII.

II. NUCLEAR PAIRING MODEL

A. The model

In 1963, Richardson proposed a model consisting of fermions occupying nondegenerate energy levels which interact solely through the pairing force [32,33]. This pairing model consists of P nondegenerate energy levels, occupied by M pairs of fermions of opposite spin (see Fig. 1). Its Hamiltonian is given by

$$H = \sum_{p\sigma} d_p a_{p\sigma}^\dagger a_{p\sigma} - \sum_{pq} g_{pq} a_{p+}^\dagger a_{p-}^\dagger a_{q-} a_{q+}, \qquad (1)$$

where the indices p and q sum over the set $\{1, \ldots, P\}$, representing the different energy levels, while σ runs over the set $\{+, -\}$, corresponding to the spin of each fermion. The coefficients d_p denote the single-particle energies, which for our our analysis we set to grow linearly as $d_p = p$. The coefficients g_{pq} are the so-called pairing strengths, representing the energy associated with moving a pair of fermions from the qth to the pth energy level.

The pairing Hamiltonian can be rewritten in terms of so-called pairing operators as

$$H = \sum_{p=1}^P d_p N_p - \sum_{p,q=1}^P g_{pq} A_p^\dagger A_q. \qquad (2)$$

Here, N_p is the pair number operator, which counts the number of fermions occupying the pth energy level. Furthermore, $A_p^\dagger$
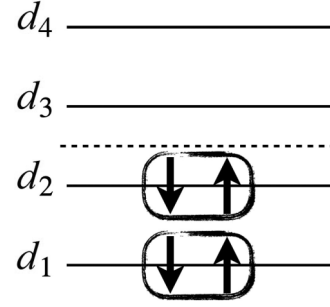

FIG. 1. Example of model space with $P = 4$ and $M = 2$. Shown are four energy levels with single-particle energies d_1, d_2, d_3, d_4, of which the bottom two are initially filled by pairs. The dashed line represents the Fermi level which divides the energy levels with energies d_1 and d_2 (the hole states) from those with energies d_3 and d_4 (the particle states).

and A_p are the pair fermionic creation and annihilation operators respectively, which create and annihilate pairs of fermions on the pth energy level. These operators are defined in terms of fermionic creation and annihilation operators as follows:

$$\begin{aligned} N_p &= \sum_\sigma a_{p\sigma}^\dagger a_{p\sigma}, \\ A_p^\dagger &= a_{p+}^\dagger a_{p-}^\dagger, \\ A_p &= a_{p-} a_{p+}, \end{aligned} \qquad (3)$$

where σ, again, sums over the set $\{+, -\}$. The purpose of this rewriting becomes clear when one notices that these operators satisfy the SU(2) algebra described by the following commutation relations:

$$\begin{aligned} [A_p, A_q^\dagger] &= \delta_{pq}(1 - N_p), \\ [N_p, A_q^\dagger] &= 2\delta_{pq} A_p^\dagger, \\ [N_p, A_q] &= -2\delta_{pq} A_p. \end{aligned} \qquad (4)$$

B. Pairings and exact solutions

When Richardson originally introduced the pairing model, he accompanied it with an exact solution for the case of constant pairing strengths, $g_{pq} = g$ for all p and q,

$$H = \sum_{p=1}^P d_p N_p - g \sum_{p,q=1}^P A_p^\dagger A_q. \qquad (5)$$

The *Ansatz* that solves the model with M pairs is

$$|\Psi\rangle = \prod_{\alpha=1}^M B_\alpha^\dagger |0\rangle, \qquad (6)$$

where

$$B_\alpha^\dagger = \sum_{\kappa=1}^P \frac{1}{2d_\kappa - E_\alpha} A_\kappa^\dagger. \qquad (7)$$

When plugged into the Schrödinger equation, this solution leads to the so-called Richardson equations [32,33]

$$1 - \sum_{\kappa=1}^{P} \frac{g}{2d_\kappa - E_\alpha} - \sum_{\beta=1, \beta \neq \alpha}^{M} \frac{2g}{E_\alpha - E_\beta} = 0. \qquad (8)$$

This is a set of coupled, nonlinear equations which one solves for the terms E_α and the ground-state energy of the Hamiltonian is found as $E = \sum_{\alpha=1}^{M} E_\alpha$. Unfortunately, the algebraic solution of the Richardson equations becomes numerically unstable at certain critical values of the interaction strength, and dedicated stable algorithms have been developed to this aim [36,39].

In addition to the constant-pairing Hamiltonian above, in this work we also consider the exactly solvable Hamiltonian with separable pairing interaction and nondegenerate single-particle energies

$$H = \sum_{p=1}^{P} d_p N_p - 2g \sum_{p,q=1}^{P} \sqrt{(\alpha - d_p)(\alpha - d_q)} A_p^\dagger A_q, \qquad (9)$$

where g is the pairing strength and α the interaction cutoff. The above Hamiltonian belongs to the hyperbolic family of Richardson-Gaudin models, whose exact solution can be obtained by again solving a set of coupled nonlinear equations similar to Eq. (8). Despite its simplicity, this model is capable of reproducing properties of heavy nuclei as described by a more realistic Gogny interaction [35].

III. VARIATIONAL MONTE CARLO WITH NEURAL QUANTUM STATES

Variational Monte Carlo (VMC) algorithms solve the many-body Schrödinger equation of a given Hamiltonian H by approximating the true ground state of the system with a variational state $|\psi_V\rangle$, defined in terms of a set of variational parameters $\boldsymbol{\theta}$. Their optimal values are found by minimizing the variational energy

$$E_V(\boldsymbol{\theta}) \equiv \frac{\langle \psi_V | H | \psi_V \rangle}{\langle \psi_V | \psi_V \rangle}, \qquad (10)$$

since the Rayleigh-Ritz variational principle ensures that the true ground-state energy $E_0 \leqslant E_V(\boldsymbol{\theta})$. Therefore, a minimal $E_V(\boldsymbol{\theta})$ corresponds to the best possible approximation to the true ground-state wave function given the *Ansatz* $|\psi_V\rangle$.

In this work, we expand the variational wave function in the occupation-number basis, where each state $|N\rangle = |n_1, n_2, \ldots, n_P\rangle$ is characterized by a set of occupation numbers $n_p = 0, 1$ representing the number of pairs occupying each energy level $p = 1, \ldots, P$. By inserting a completeness relation over the occupation states, $1 = \sum_N |N\rangle\langle N|$, in Eq. (10), one obtains

$$E_V(\boldsymbol{\theta}) = \sum_N \frac{|\langle \psi_V | N \rangle|^2}{\langle \psi_V | \psi_V \rangle} E_L(N), \qquad (11)$$

where we define the local energy as

$$E_L(N) = \frac{\langle N | H | \psi_V \rangle}{\langle N | \psi_V \rangle}. \qquad (12)$$

The sum entering Eq. (11) quickly becomes intractable by means of exact-quadrature methods due to the high dimensionality of the Hilbert space, and it is typically evaluated by means of Monte Carlo methods. Indeed, Eq. (11) has already the form of an expectation value of the local energy over the normalized probability distribution

$$P(N) = \frac{|\langle \psi_V | N \rangle|^2}{\langle \psi_V | \psi_V \rangle}. \qquad (13)$$

Therefore, according to the central limit theorem, $E_V(\boldsymbol{\theta})$ can be estimated by the average of the local energy $E_L(N)$ calculated for a set of configurations N sampled according to the probability $P(N)$. This latter task can be accomplished by employing the Metropolis-Hastings algorithm [40,41]. In order for the exclusion principle to be satisfied and the total number of pairs $\sum_p n_p$ to be conserved, in the random walk, the proposed state $|N'\rangle$ is obtained by swapping two random occupation numbers in the array of zeros and ones defining a given $|N\rangle$. Note that this procedure is not limited to the Hamiltonian operator, but it can be used to estimate the ground-state expectation value of any quantum-mechanical operator.

When evaluating $E_L(N)$, the one-body term of the Hamiltonian and the diagonal part of the pairing interaction can be computed as

$$\langle N| \sum_p (d_p N_p - g_{pp} A_p^\dagger A_p)|\psi_V\rangle$$
$$= \sum_p (2d_p - g_{pp}) n_p \langle N|\psi_V\rangle. \qquad (14)$$

To compute the nondiagonal terms, we insert a completeness relation:

$$\langle N| \sum_{p \neq q} g_{pq} A_p^\dagger A_q |\psi_V\rangle$$
$$= \sum_{N'} \langle N| \sum_{p \neq q} g_{pq} A_p^\dagger A_q |N'\rangle \langle N'|\psi_V\rangle. \qquad (15)$$

In this expression, for a given a state $|N\rangle$, only the indices p, q such that $n_p = 1$ and $n_q = 0$ contribute to the inner sum. Furthermore, for fixed p, q, only the $|N'\rangle$ having $n'_p = 0$, $n'_q = 1$ and $n'_r = n_r$ for $r \neq p$, q yields a nonzero matrix element. Thus, these terms can be calculated using two nested loops, one with index p running over the nonzero occupation numbers of $|N\rangle$ and the other with index q over the null ones, and multiplying g_{pq} by the variational wave function evaluated for a state $|N'\rangle$ with the same occupation numbers as $|N\rangle$ but $n'_p = 0, n'_q = 1$.

A. Neural network quantum states

Inspired by the recently introduced NQS *Ansätze* [21,22], we parametrize the variational wave function as

$$\psi_V(N) \equiv \langle N|\psi_V\rangle = e^{\mathcal{U}(N)} \tanh[\mathcal{V}(N)], \qquad (16)$$

where $\mathcal{U}(N)$ and $\mathcal{V}(N)$ are fully connected neural networks (FCNNs). In contrast with previous applications in the field based on complex variational states, our *Ansatz* is real-valued; the function $\mathcal{U}(N)$ determines the magnitude of $\psi_V(N)$, while $\tanh[\mathcal{V}(N)]$ is a smooth version of its sign.

A chief advantage of using the occupation number basis is that the states $|N\rangle$ inherently encode the exchange symmetry required for a wave function describing a system of fermions, thus the FCNNs can simply take as inputs the string of occupation numbers $n_p = 0, 1$ defining each state. We typically use between two to three hidden layers and tanh activation functions, since these choices proved to be stable and yield the most accurate results. As we typically use hidden layers of $\approx P$ neurons, evaluating the wave function exhibits a polynomial scaling in the number of energy levels [$O(P^2)$], independent of the number of pairs. With a complexity for the calculation of the local energy at worst quadratic in the number of energy levels, this gives the whole algorithm a complexity in the number of energy levels of $O(P^4)$.

B. Training algorithm

Generally, in the context of artificial neural networks, a *loss function* $\mathcal{L}$ measures the performance of the model, and training a network, i.e., optimizing its parameters $\boldsymbol{\theta}$, corresponds to numerically minimizing $\mathcal{L}(\boldsymbol{\theta})$. This procedure is analogous to what is typically done in VMC, in this case the loss function being the variational energy defined in Eq. (10). Therefore, training the neural network quantum state *Ansatz* of Eq. (16) consists in minimizing the variational energy with respect to the parameters entering the *Ansatz*. A significant improvement in the execution time and in the stability of the optimization can be achieved by using the so-called stochastic reconfiguration (SR) algorithm [42], which is closely related to the natural gradient descent method [43,44]. Within the SR algorithm, the variational parameters $\boldsymbol{\theta}$ are updated as

$$\boldsymbol{\theta} \rightarrow \boldsymbol{\theta} + \delta\boldsymbol{\theta}, \quad \boldsymbol{S}\delta\boldsymbol{\theta} = -\delta\tau\boldsymbol{G}, \tag{17}$$

where $\delta\tau$ is the learning rate, $\boldsymbol{S}$ is the overlap matrix, and $\boldsymbol{G}$ is the gradient of the variational energy with respect to the variational parameters.

The matrix elements of $\boldsymbol{S}$ and $\boldsymbol{G}$ are calculated as the following covariances:

$$S_{mn} = \langle \mathcal{O}_m^\dagger \mathcal{O}_n \rangle - \langle \mathcal{O}_m^\dagger \rangle \langle \mathcal{O}_n \rangle, \tag{18}$$

$$G_n = 2\langle \mathcal{O}_n^\dagger H \rangle - 2\langle \mathcal{O}_n^\dagger \rangle \langle H \rangle, \tag{19}$$

where $\mathcal{O}_n|\psi_V\rangle = \partial|\psi_V\rangle/\partial\theta_n$ is the derivative operator with respect to the variational parameter θ_n. An optimization using this update is essentially equivalent to performing an imaginary-time evolution of the variational wave function with time step $\delta\tau$ [42,45] in the manifold spanned by the wave function and its derivatives. Since the overlap matrix is positive definite, the linear system of Eq. (17) could be solved using, e.g., the Cholesky decomposition. However, this requires storing the entire matrix $\boldsymbol{S}$ and, since the number of variational parameters is of order P^2, its size scales as P^4, which quickly becomes prohibitively large as P increases. For this reason, we solve the linear system using the conjugate gradient algorithm, which only requires the definition of a function that calculates the matrix-vector product $\boldsymbol{Sv}$ for an arbitrary vector $\boldsymbol{v}$. The expectation values in Eqs. (18) and (19) can once again be expressed as averages over configurations sampled according to the probability $P(N)$ of Eq. (13); the
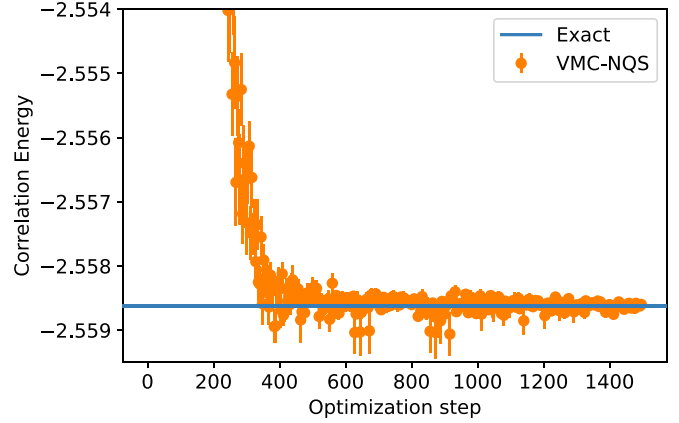

FIG. 2. Convergence of the SR-RMSProp training algorithm for a system of $P = 10$ energy levels occupied by $M = 5$ pairs of fermions. The correlation energy obtained from exact-diagonalization is displayed by the solid blue line.

product $\boldsymbol{Sv}$ given an arbitrary $\boldsymbol{v}$ is then rewritten in order to make its evaluation time and memory efficient [46].

Following Ref. [22], to further improve the stability of the convergence, we regularize the SR equations by adding a small RMSProp-inspired diagonal shift as

$$\boldsymbol{v} \rightarrow \beta\boldsymbol{v} + (1-\beta)\boldsymbol{G}^2,$$

$$\left[\boldsymbol{S} + \varepsilon\, \mathrm{diag}\left(\sqrt{\frac{\boldsymbol{v}}{(1-\beta^k)}} + 10^{-8}\right)\right]\delta\boldsymbol{\theta} = -\delta\tau\boldsymbol{G}. \tag{20}$$

In the above equation, ε is the strength of the diagonal shift, whose typical values range from $\varepsilon = 0.0001$ to $\varepsilon = 0.001$ for different Hamiltonians. As with the exponential decay rates for the second moment estimates of the RMSprop algorithm [47], we take $\beta = 0.99$. Finally, the factor $(1 - \beta^k)$ corrects the bias in the second-moment estimates as in Adam [48], with k being the current optimization step number.

To illustrate the performance of the SR-RMSProp training algorithm, in Fig. 2 we plot the behavior of the correlation energy as a function of the optimization step for the constant-pairing Hamiltonian of Eq. (5) in the case of $P = 10$ energy levels occupied by $M = 5$ pairs of fermions. After about one thousand steps, the VMC-NQS correlation energy agrees with the exact-diagonalization one. For this specific case, we achieve a final 10^{-7} precision level.

IV. PAIR COUPLED CLUSTER DOUBLES THEORY

Coupled cluster theory was first developed to solve the nuclear Schrödinger equation by Coester and Kümmel in the late 1950s and early 1960s [49,50]. In this work, we focus on the pair coupled cluster doubles theory (pCCD), which is ideally suited to solve the pairing problem. The pCCD *Ansatz* is given by

$$|\Psi\rangle = e^T|\Phi_0\rangle, \tag{21}$$

where $|\Phi_0\rangle$ is the reference state and T is the cluster operator restricted to moving pairs of fermions:

$$T \equiv \sum_{ia} t_{ai} A_a^\dagger A_i. \tag{22}$$

Here, $A_a^\dagger$ and A_i are the pair fermionic creation and annihilation operators defined in Eq. (3). Inspired by the notation commonly adopted in quantum chemistry, we reserve the labels $i, j, k, \ldots$ for the hole states and $a, b, c, \ldots$ for the particle states.

Starting from the time-independent Schrödinger equation

$$H|\Psi\rangle = E|\Psi\rangle, \tag{23}$$

with the coupled cluster *Ansatz* (21), left-multiplying both sides by e^{-T} and left-multiplying by either the reference state or an excited state yields

$$E = \langle\Phi_0|\overline{H}|\Phi_0\rangle, \tag{24}$$

$$0 = \langle\Phi_i^a|\overline{H}|\Phi_0\rangle, \tag{25}$$

where the excited state $\langle\Phi_i^a|$ is obtained as

$$\langle\Phi_i^a| = \langle\Phi_0|A_i^\dagger A_a, \tag{26}$$

and $\overline{H}$ is the similarity-transformed pairing Hamiltonian,

$$\overline{H} = e^{-T}He^T. \tag{27}$$

Using the Baker-Campbell-Hausdorff (BCH) identity, $\overline{H}$ can be expanded as

$$\overline{H} = H + [H, T] + \tfrac{1}{2}[[H, T], T] + \cdots. \tag{28}$$

When the latter expression is plugged into the pCCD equations (24) and (25), one can truncate the resulting expressions by noting that only certain terms in the infinite sum (those which can be fully contracted) are nonzero. Separating the pairing model Hamiltonian into a single-body and a two-body term $H = F + V$, where

$$F = \sum_p d_p N_p,$$

$$V = -\sum_{pq} g_{pq} A_p^\dagger A_q, \tag{29}$$

the truncated pCCD equations read

$$E = \langle\Phi_0|F + V + VT|\Phi_0\rangle, \tag{30}$$

$$0 = \langle\Phi_i^a|V + FT + VT + \tfrac{1}{2}VT^2 - TVT|\Phi_0\rangle. \tag{31}$$

Recognizing that the first two terms of the truncated pCCD energy in Eq. (30) equal the reference energy, the correlation energy equation can readily be identified with

$$\Delta E = \langle\Phi_0|VT|\Phi_0\rangle = -\sum_{ia} t_{ai}g_{ia}, \tag{32}$$

where in the last equality we used Wick's theorem [51]. Similarly, Eqs. (31) become

$$0 = g_{ai} + 2\left(d_i - d_a - g_{ii} + g_{ia}t_{ai} - \sum_b g_{ib}t_{bi}\right.$$
$$\left. - \sum_j g_{ja}t_{aj}\right)t_{ai} + \sum_b g_{ab}t_{bi} + \sum_j g_{ji}t_{aj}$$
$$+ \sum_{bj} g_{jb}t_{aj}t_{bi}. \tag{33}$$

In practice, the amplitudes t are found by solving the non-linear and coupled Eqs. (33), which is accomplished using an iterative root-finding algorithm such as Newton's method, and plugging them into Eq. (32). To facilitate the convergence of Newton's method, a good initial guess for each t is necessary, which is informed here by many-body perturbation theory.

V. MANY-BODY PERTURBATION THEORY

Many-body perturbation theory (MBPT) [52–54] adds correlations between particles as perturbations to the Hartree-Fock wave function, the ground state of a collection of noninteracting particles subject to a mean-field potential that approximates their interaction. Analogously to single-body perturbation theory, it is possible to approximate the energy of a system by writing the Hamiltonian as the sum of two terms, one of which is readily diagonalizable. In this theory, we assume that the Hamiltonian H can be written as the sum of an unperturbed Hamiltonian H_0 and an interacting Hamiltonian H_I, namely, $H = H_0 + H_I$, where $H_0|\Phi_n\rangle = E_n|\Phi_n\rangle$ is easily solvable. Expanding the exact wave-function $|\Psi_0\rangle$ in terms of the unperturbed wave function and its excited states and the defining the projection operators $P = |\Phi_0\rangle\langle\Phi_0|$ and $Q = \sum_{n=1}^\infty |\Phi_n\rangle\langle\Phi_n|$, we can expand the correlation energy ΔE perturbatively in terms of the interaction H_I. A standard approach for this perturbative expansion is given by Rayleigh-Schrödinger (RS) perturbation theory, see, for example, Refs. [52–54]. With RS many-body perturbation theory the correlation energy reads

$$\Delta E = \sum_{n=1}^\infty \Delta E^{(n)}, \tag{34}$$

where, up to second order in the interaction, we have

$$\Delta E^{(1)} = \langle\Phi_0|H_I|\Phi_0\rangle, \tag{35}$$

$$\Delta E^{(2)} = \langle\Phi_0|H_I \frac{Q}{E_0 - H_0} H_I|\Phi_0\rangle. \tag{36}$$

For the pairing model Hamiltonian (1) with the partitioning

$$H_0 = \sum_{p\sigma} \epsilon_p a_{p\sigma}^\dagger a_{p\sigma}, \tag{37}$$

where

$$\epsilon_p = d_p - \sum_i \delta_{pi}g_{pp}, \tag{38}$$

we can rewrite the second-order correlation energy contribution (36) as

$$\Delta E^{(2)} = \frac{1}{2} \sum_{ia} \frac{g_{ai}g_{ia}}{d_i - d_a - g_{ii}}. \tag{39}$$

We can use this result to inform a good initial guess for the cluster parameters t by comparing the pCCD correlation energy (32) with the second-order contribution to the correlation energy from MBPT, which allows us to identify

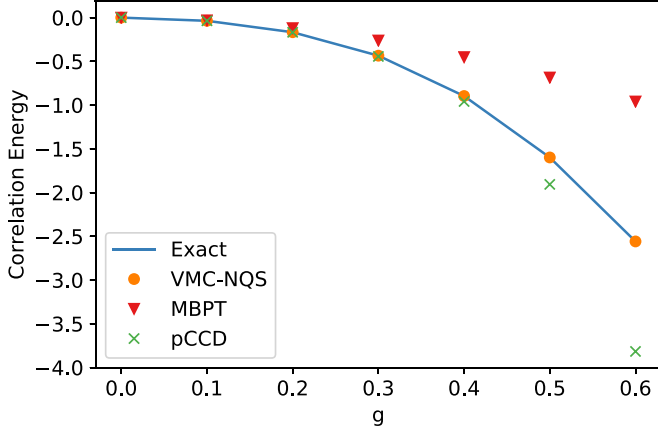$$t_{ai}^{(0)} = -\frac{1}{2} \frac{g_{ai}}{d_i - d_a - g_{ii}}. \tag{40}$$

FIG. 3. Correlation energy for the constant-coupling model for $P = 10$ energy levels and $M = 5$ pairs as a function of the interaction strength. VMC-NQS (solid orange circles), MBPT (solid red triangles), and pCCD (green crosses) calculations are compared with the exact answer (solid line).

VI. RESULTS

The VMC-NQS values reported in this section were obtained by averaging over the energy estimates of 100 optimization steps at convergence. We initially present results for the Hamiltonian with constant pairing strength g of Eq. (5) considering $P = 10$ energy levels occupied by $M = 5$ pairs of fermions. The correlation energies as functions of the pairing strength computed within MBPT, pCCD, and VMC-NQS are displayed in Fig. 3. At small values of g, all many-body methods yield similar correlation energies. Already at $g \lesssim -0.2$, MBPT results start deviating from the exact-diagonalization ones, indicating the shortcomings of perturbation theory in addressing the large negative coupling constant region. A similar trend is observed by the pCCD calculations, albeit they remain closer to the exact ones down to $g \lesssim -0.4$. At deeper values of g, the pCCD energies become significantly lower than the exact ones, showing a clear violation of the variational principle. On the other hand, the correlation energies computed using VMC-NQS are in excellent agreement with the exact ones over the whole range of g values considered and beyond, with discrepancies remaining always smaller than 10^{-4}. Crucially, in contrast with MBPT and pCCD methods, VMC-NQS fulfils by construction the variational principle.

We then extend our comparative analysis considering a larger model space, with up to $P = 40$ energy levels and $M = 20$ pairs of fermions. In Table I, we compare the correlation energies obtained within VMC-NQS, pCCD, and the highly accurate iterative approach for solving the Richardson equations developed in Ref. [36]. To make contact with the results of the latter reference, the single-particle energies are rescaled as $d_p = p/10$ and the pairing strength is taken to be $g = 0.05$. Similarly to the smaller model space, the correlation energies computed within VMC-NQS are almost always in perfect agreement with those obtained from the iterative method. The only exception is the case with $P = 40$ and $M = 20$, for which the VMC-NQS yields -1.785, while the iterative method provides -2.111. In this regard, we successfully executed the *Mathematica* program attached to Ref. [36]

TABLE I. Correlation energies obtained with the VMC-NQS method compared with pCCD and the iterative approach of Ref. [36] for the constant-coupling Hamiltonian of Eq. (5) with $g = 0.05$ and $d_p = p/10$.

	P	$M = 5$	$M = 10$	$M = 15$	$M = 20$
VMC-NQS	10	-0.160	0.000		
pCCD	10	-0.193	0.000		
Iterative	10	-0.160	0.000		
VMC-NQS	20	-0.394	-0.515	-0.394	0.000
pCCD	20	-0.987	-1.743	-1.110	0.000
Iterative	20	-0.394	-0.515	-0.394	0.000
VMC-NQS	30	-0.606	-0.946	-1.057	-0.946
pCCD	30	-1.750	3.500	-6.467	-5.697
Iterative	30	-0.606	-0.946	-1.057	-0.946
VMC-NQS	40	-0.809	-1.356	-1.678	-1.785
pCCD	40	-4.231	124.923	7.598	-44.346
Iterative	40	-0.809	-1.356	-1.678	-2.110

for all combinations of P and M, except for $P = 40$ and $M = 20$, where we could not obtain converged results. As expected and numerically verified in the $P = 10$ case using exact diagonalization, the correlation energy as a function of M for fixed P is convex and has a minimum for $M = P/2$. This behavior is correctly reproduced by the VMC-NQS algorithm; on the other hand, the behavior of the iterative method is less regular, with a change from positive to negative concavity at $M = 20$. To further support the VMC-NQS result, we also verified its consistency within statistical error (of order 10^{-6}) when doubling the width of the inner neural networks layers from 40 to 80 neurons.

Contrarily to VMC-NQS, pCCD struggles already for $P = 10$ and $M = 5$, where it yields $\approx 20\%$ overbinding. Increasing the number of single-particle states and pairs, pCCD becomes more and more unreliable, with large departures from both the exact and VMC-NQS values. This behavior is due to the pairing interaction terms growing faster with the number of states than the mean-field one, making the exact ground-state very different from the Hartree-Fock solution.

To further gauge the accuracy of the VMC-NQS method, we consider a parametrization of the constant-pairing Hamiltonian of Eq. (5) suitable to describe ultrasmall superconducting grains [37,38]. Specifically, to make contact with density-matrix renormalization group and exact results of Refs. [37,38], we take $P = 100$ energy levels with half filling ($M = 50$) and $g = 0.4$. We find the VMC-NQS correlation energy, $-40.500\,741(20)$, to be statistically compatible with the exact value of $-40.500\,756$.

We now turn to the separable-pairing Hamiltonian of Eq. (9). In Fig. 4, we display the correlation energy for this Hamiltonian for $P = 10$ energy levels and $M = 5$ pairs as a function of the interaction strength. Consistent with Ref. [35], we use the cutoff $\alpha = 10$. The exact-diagonalization and VMC-NQS results overlap nicely over the entire range of pairing strength that we consider, in some cases matching up to eight significant digits, corroborating once more the accuracy of NQS in representing complicated many-body wave functions. Similarly to the results displayed in Fig. 3 for
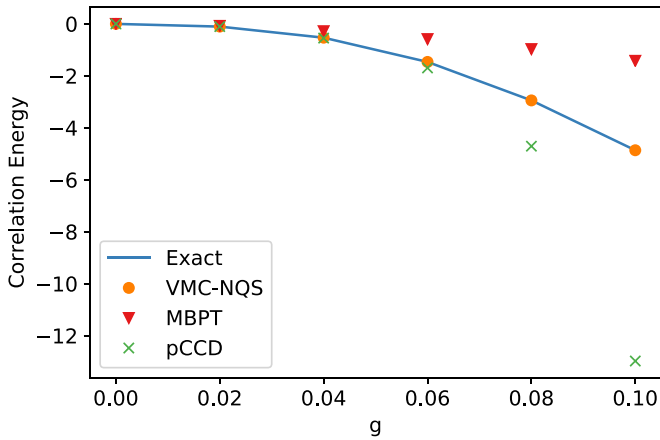
FIG. 4. Correlation energy for the separable-coupling model for $P = 10$ energy levels and $M = 5$ pairs as a function of the interaction strength. VMC-NQS (solid orange circles), MBPT (solid red triangles), and pCCD (green crosses) calculations are compared with the exact answer (solid line).

the constant-pairing case, MBPT yields accurate correlation energies only for very small values g, while for larger values of the pairing strength large deviations are observed. The energies computed within pCCD start deviating from the exact ones for $g \gtrsim 0.06$, exhibiting abrupt violations of the variational principle. Consistent with the constant-pairing energies listed in Table I, pCCD becomes less accurate with increasing system size, while VMC-NQS does not seem to suffer similar limitations. In addition, we verified that the nearly perfect agreement between VMC-NQS and exact-diagonalization energies persist at least up until $g = 0.6$, confirming once again the robustness of the method in the nonperturbative regime.

VII. CONCLUSIONS

We have introduced a variational Monte Carlo method based on neural-network quantum states that solves the nuclear many-body problem in the occupation-number formalism. A tailored version of the stochastic-reconfiguration algorithm, with a regularization term inspired by RMSprop, is utilized to train the neural networks and minimize the Hamiltonian expectation value.

As a specific application of this method, which exhibits a polynomial scaling with the number of single-particle levels, we considered two classes of exactly solvable pairing models both with constant and separable pairing interaction strength. In addition to exact-diagonalization techniques, which are limited to relatively small model spaces, we benchmark the VMC-NQS approach against virtually exact methods that solve the Gaudin-Richardson equations in an iterative fashion. We also gauge the accuracy of many-body perturbation theory and pair coupled cluster doubles theory, both routinely used to solve the nuclear many-body problem.

We find the VMC-NQS results to be in excellent agreement with the exact solution, independent of the magnitude of the pairing strength. As expected, the correlation energies obtained from MBPT start deviating from the exact values when the pairing strength becomes large. A similar trend is observed by pCCD calculations, although they remain closer to the exact solutions for a broader range of pairing strengths values. Most notably, VMC-NQS is guaranteed to respect the variational principle, which is in general violated by both MBPT and pCCD.

It has to be mentioned that VMC-NQS is not limited to constant or separable pairing interactions. As an immediate application of this method, we will consider the state dependent, nonseparable, Gogny interaction in the Hartree-Fock basis and compute the ground-state energies of nuclei going beyond the mean-field approximation. In this regard, it would be interesting to validate the conclusions of Ref. [35], which were obtained by fitting an exactly solvable hyperbolic model to the gaps and pairing tensors of Gogny Hartree-Fock-Bogoliubov calculations.

The applicability of VMC-NQS extends to pairing Hamiltonians relevant to model interacting systems in condensed matter. An example is the characterization of number-conserving topological superconductors, or superfluids, and the fate of Majorana zero-modes beyond mean-field [55]. While the Richardson-Gaudin-Kitaev model is integrable for periodic and antiperiodic boundary conditions, no exact solutions are known for open boundaries. Since the interactions are long-range, DMRG approaches struggle to converge [56], while deep neural quantum states should be able to provide more accurate solutions.

Although interesting in its own right, solving the pairing model is only the first application of NQS to the nuclear many-body problem in the occupation number formalism. As a next step, we will consider more realistic shell-model Hamiltonians [57]. We plan on carrying out benchmark calculations with exact-diagonalization methods and coupled cluster theory in pf-shell nuclei. The long-range goal is to provide accurate solutions to Hamiltonians that are systematically derived within chiral effective-field theory, extending the work of Ref. [58] that was limited to a matrix product state *Ansatz*.

[1] G. Hagen, T. Papenbrock, M. Hjorth-Jensen, and D. J. Dean, Rep. Prog. Phys. **77**, 096302 (2014).

[2] H. Hergert, S. K. Bogner, T. D. Morris, A. Schwenk, and K. Tsukiyama, Phys. Rep. **621**, 165 (2016).

[3] V. Somà, Front. Phys. **8**, 340 (2020).

[4] D. Lee, Front. Phys. **8**, 174 (2020).

[5] J. Carlson, S. Gandolfi, F. Pederiva, S. C. Pieper, R. Schiavilla, K. E. Schmidt, and R. B. Wiringa, Rev. Mod. Phys. **87**, 1067 (2015).

[6] P. Gysbers *et al.*, Nat. Phys. **15**, 428 (2019).

[7] S. Malbrunot-Ettenauer, S. Kaufmann, S. Bacca, C. Barbieri, J. Billowes, M. L. Bissell, K. Blaum, B. Cheal, T. Duguet, R. F. Ruiz *et al.*, Phys. Rev. Lett. **128**, 022502 (2022).

[8] B. Hu *et al.*, Nat. Phys. **18**, 1196 (2022).

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

[10] K. Choo, A. Mezzacapo, and G. Carleo, Nat. Commun. **11**, 2368 (2020).

[11] J. Hermann, Z. Schätzle, and F. Noé, Nat. Chem. **12**, 891 (2020).

[12] D. Pfau, J. S. Spencer, A. G. D. G. Matthews, and W. M. C. Foulkes, Phys. Rev. Res. **2**, 033429 (2020).

[13] G. Pescia, J. Han, A. Lovato, J. Lu, and G. Carleo, Phys. Rev. Res. **4**, 023138 (2022).

[14] M. Wilson, S. Moroni, M. Holzmann, N. Gao, F. Wudarski, T. Vegge, and A. Bhowmik, arXiv:2202.04622 [physics.chem-ph] (2022).

[15] J. Stokes, J. R. Moreno, E. A. Pnevmatikakis, and G. Carleo, Phys. Rev. B **102**, 205122 (2020).

[16] J. R. Moreno, G. Carleo, A. Georges, and J. Stokes, Proc. Natl. Acad. Sci. USA **119**, e2122059119 (2022).

[17] T. Zhao, J. Stokes, and S. Veerapaneni, arXiv:2208.05637.

[18] J. W. T. Keeble and A. Rios, Phys. Lett. B **809**, 135743 (2020).

[19] J. R. Sarmiento, J. W. T. Keeble, and A. Rios, arXiv:2205.12795.

[20] C. Adams, G. Carleo, A. Lovato, and N. Rocco, Phys. Rev. Lett. **127**, 022502 (2021).

[21] A. Gnech, C. Adams, N. Brawand, G. Carleo, A. Lovato, and N. Rocco, Few-Body Syst. **63**, 7 (2022).

[22] A. Lovato, C. Adams, G. Carleo, and N. Rocco, Phys. Rev. Res. **4**, 043178 (2022).

[23] A. Bohr, B. R. Mottelson, and D. Pines, Phys. Rev. **110**, 936 (1958).

[24] D. J. Dean and M. Hjorth-Jensen, Rev. Mod. Phys. **75**, 607 (2003).

[25] P. Moller and J. R. Nix, Nucl. Phys. A **536**, 20 (1992).

[26] H. Molique and J. Dudek, Phys. Rev. C **56**, 1795 (1997).

[27] A. Volya, B. A. Brown, and V. Zelevinsky, Phys. Lett. B **509**, 37 (2001).

[28] V. Zelevinsky and A. Volya, Phys. At. Nucl. **66**, 1781 (2003).

[29] X. Liu and C. Qi, Comput. Phys. Commun. **259**, 107349 (2021).

[30] N. Sandulescu, J. Blomqvist, T. Engeland, M. Hjorth-Jensen, A. Holt, R. J. Liotta, and E. Osnes, Phys. Rev. C **55**, 2708 (1997).

[31] L. Y. Jia, Phys. Rev. C **88**, 044303 (2013).

[32] R. W. Richardson, Phys. Lett. **3**, 277 (1963).

[33] R. W. Richardson, Phys. Lett. **5**, 82 (1963).

[34] M. Gaudin, J. Phys. (Paris) **37**, 1087 (1976).

[35] J. Dukelsky, S. Lerma H., L. M. Robledo, R. Rodriguez-Guzman, and S. M. A. Rombouts, Phys. Rev. C **84**, 061301(R) (2011).

[36] X. Guan and C. Qi, Comput. Phys. Commun. **275**, 108310 (2022).

[37] J. Dukelsky and G. Sierra, Phys. Rev. Lett. **83**, 172 (1999).

[38] J. Dukelsky and G. Sierra, Phys. Rev. B **61**, 12302 (2000).

[39] S. Rombouts, D. Van Neck, and J. Dukelsky, Phys. Rev. C **69**, 061303(R) (2004).

[40] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).

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

[42] S. Sorella, Phys. Rev. B **71**, 241103(R) (2005).

[43] S.-i. Amari, Neural Comput. **10**, 251 (1998).

[44] J. Stokes, J. Izaac, N. Killoran, and G. Carleo, Quantum **4**, 269 (2020).

[45] S. Sorella, Phys. Rev. Lett. **80**, 4558 (1998).

[46] E. Neuscamman, C. J. Umrigar, and G. K.-L. Chan, Phys. Rev. B **85**, 045103 (2012).

[47] G. H. Tijmen Tieleman, Lecture 6.5—RMSProp, Course Neural Networks for Machine Learning Lecture 6 (2018).

[48] D. P. Kingma and J. Ba, arXiv:1412.6980.

[49] F. Coester, Nucl. Phys. **7**, 421 (1958).

[50] F. Coester and H. Kümmel, Nucl. Phys. **17**, 477 (1960).

[51] G. C. Wick, Phys. Rev. **80**, 268 (1950).

[52] J. Paldus, *Springer Handbook of Atomic, Molecular, and Optical Physics*, edited by G. Drake (Springer, New York, 2006), pp. 101–114.

[53] I. Shavitt and R. J. Bartlett, *Many-Body Methods in Chemistry and Physics: MBPT and Coupled-Cluster Theory*, Cambridge Molecular Science (Cambridge University Press, 2009).

[54] J. Lietz, G. Jansen, G. Hagen, and M. Hjorth-Jensen, Computational nuclear physics and post-Hartree-Fock methods, in *An Advanced Course in Computational Nuclear Physics Bridging the Scales from Quarks to Neutron Stars* (Springer International Publishing, 2017), pp. 327–333.

[55] G. Ortiz, J. Dukelsky, E. Cobanera, C. Esebbag, and C. Beenakker, Phys. Rev. Lett. **113**, 267002 (2014).

[56] F. Iemini, L. Mazza, D. Rossini, R. Fazio, and S. Diehl, Phys. Rev. Lett. **115**, 156402 (2015).

[57] M. Hjorth-Jensen, T. T. Kuo, and E. Osnes, Phys. Rep. **261**, 125 (1995).

[58] A. Tichai, S. Knecht, A. T. Kruppa, O. Legeza, C. P. Moca, A. Schwenk, M. A. Werner, and G. Zarand, arXiv:2207.01438.