

Machine learning for the prediction of converged energies from *ab initio* nuclear structure calculations

Marco Knöll^{a,*}, Tobias Wolfgruber^a, Marc L. Agel^a, Cedric Wenz^a, Robert Roth^{a,b}

^a Institut für Kernphysik, Fachbereich Physik, Technische Universität Darmstadt, Schlossgartenstr. 2, 64289 Darmstadt, Germany

^b Helmholtz Forschungsakademie Hessen für FAIR, GSI Helmholtzzentrum, 64289 Darmstadt, Germany

ARTICLE INFO

Article history:

Received 7 July 2022

Received in revised form 16 December 2022

Accepted 14 February 2023

Available online 17 February 2023

Editor: J.-P. Blaizot

ABSTRACT

The prediction of nuclear observables beyond the finite model spaces that are accessible through modern *ab initio* methods, such as the no-core shell model, poses a challenging task in nuclear structure theory. It requires reliable tools for the extrapolation of observables to infinite many-body Hilbert spaces along with reliable uncertainty estimates. In this work we present a universal machine learning tool capable of capturing observable-specific convergence patterns independent of nucleus and interaction. We show that, once trained on few-body systems, artificial neural networks can produce accurate predictions for a broad range of light nuclei. In particular, we discuss neural-network predictions of ground-state energies from no-core shell model calculations for ${}^6\text{Li}$, ${}^{12}\text{C}$ and ${}^{16}\text{O}$ based on training data for ${}^2\text{H}$, ${}^3\text{H}$ and ${}^4\text{He}$ and compare them to classical extrapolations.

© 2023 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>). Funded by SCOAP³.

1. Introduction

The major goal of nuclear structure theory is the accurate description of nuclear properties based on the underlying strong interaction. The low-energy properties of these complicated quantum many-body systems emerge from the interactions among nucleons as the relevant degrees of freedom. While the construction of these interactions poses a challenge in itself, also the solution of the nuclear many-body problem with controlled uncertainties is a formidable conceptual and computational task. Several successful *ab initio* methods are available nowadays, such as the no-core shell model (NCSM) [1–4], the coupled-cluster (CC) method [5], the self-consistent Green's function (SCGF) approach [6], the in-medium similarity renormalization group (IM-SRG) [7], or lattice and continuum quantum Monte Carlo (QMC) methods [8]. Besides the latter, all of these methods are built on the expansion of the many-body problem in a discrete many-body basis that is constructed from an underlying single-particle basis. To render the computational problem finite, various basis truncations are employed that define finite-dimensional model spaces. For *ab initio* methods the truncations are constructed in such a way that calculations for increasing model-space dimension eventually converge to the exact result for the full many-body Hilbert space.

In the past two decades *ab initio* methods have proven to provide accurate results for a range of observables and nuclei based on nucleon-nucleon (NN) and three-nucleon (3N) interactions from chiral effective field theory (EFT) [9–13]. Besides their success, all of these methods suffer from a rapid growth of the underlying basis and the resulting model spaces. Particularly for configuration interaction approaches, like the NCSM, the factorial growth of the model-space dimension with increasing particle number [14] sets severe limits for converged calculations. Even with access to high performance computing and methods to accelerate convergence, such as the similarity renormalization group (SRG) [15,16], one is inevitably confronted with incomplete convergence of the many-body calculation. Hence, there is a need for extrapolation procedures that provide robust predictions for the converged observables along with reliable uncertainty estimates.

Traditional extrapolation schemes typically rely on empirical exponential or polynomial parametrizations of the model-space dependence of observables [3,14,17]. Recent physics-motivated parametrizations, like the infrared extrapolation schemes derived from effective theories [18–22], have proven successful in specific cases, but impose additional constraints on the many-body calculations.

With increasing popularity of machine learning, artificial neural networks (ANNs) have entered the field of nuclear structure physics, e.g., through large-scale approaches based on experimental data [23,24], theoretical applications in the form of Bayesian machine-learning [25], neural-network quantum states [26–28], and many more. For a comprehensive overview and further read-

* Corresponding author.

E-mail addresses: mknoll@theorie.ikp.physik.tu-darmstadt.de (M. Knöll), robert.roth@physik.tu-darmstadt.de (R. Roth).

ing we refer to [29,30]. While ANNs have excelled in classification and interpolation tasks, precise extrapolations remain challenging [31]. However, first applications to NCSM and CC calculations have demonstrated the potential of machine learning as an extrapolation tool supplementing *ab initio* many-body methods [32,33]. So far, these applications extrapolate NCSM or CC ground-state observables by emulating their model-space dependence, more precise, their functional dependence on the model-space truncation parameter and the harmonic-oscillator (HO) frequency $\hbar\Omega$ of the underlying single-particle basis. The ANNs are trained to mimic this functional dependency for a specific interaction, nucleus, and eigenstate and are then used to extrapolate the observable to a sufficiently large model space. These approaches essentially replace the traditional exponential parametrization of the many-body data by an ANN, which contains many more parameters. The training process for the ANN requires a sufficient amount of data, i.e., a large number of many-body calculations for a specific target nucleus, interaction, and eigenstate with different model spaces. This can easily become a bottleneck for the applicability of this scheme. While the problem can partially be addressed through suitable interpolations of the training data as shown in [33], no data from large model spaces or actually converged results will ever inform the network. Moreover, the resulting ANN, by construction, is only useful for an extrapolation for the specific nucleus, interaction, and state it was trained for.

We present a new approach, build on a universal ANN for the prediction of the converged values of a specific observable for arbitrary nuclei, interactions, and eigenstates from sequences of NCSM calculations. The basic idea is akin to a pattern recognition task: Short sequences of NCSM results with increasing basis truncation parameter for different HO frequencies $\hbar\Omega$ form a pattern that allows an experienced many-body practitioner to guess what the converged value will be. For this, one does not even have to know which nucleus, interaction, or eigenstate the data was generated for—it is just the pattern of converging sequences that holds the relevant information. To extract quantitative predictions, we design an ANN that uses the raw many-body results for short N_{\max} sequences and several frequencies $\hbar\Omega$ as input and outputs the converged value of the observable. We train the ANN with calculations for few-body systems ($A \leq 4$), which can be fully converged in the NCSM, with a large number of different Hamiltonians. Once trained on a huge set of training data, the ANNs serve as universal tool for the prediction of the converged observable for expensive many-body runs, where only a very limited amount of data is available. In this work we discuss predictions of ground-state energies from ANNs that have been trained on converged NCSM calculations for the few-body systems ^2H , ^3H and ^4He . We subsequently apply the ANNs also to heavier p-shell nuclei, i.e., ^6Li , ^{12}C and ^{16}O .

2. No-core shell model

Our many-body method of choice is the NCSM, in which the stationary Schrödinger equation is cast into a matrix eigenvalue problem

$$\sum_j \langle \phi_i | H | \phi_j \rangle \langle \phi_j | \psi_n \rangle = E_n \langle \phi_i | \psi_n \rangle \quad \forall i, \quad (1)$$

with Hamiltonian H , its energy eigenvalues E_n , and the corresponding eigenstates $|\psi_n\rangle$. The eigenstates are expanded in a set of Slater determinants $\{|\phi_i\rangle\}$, which are constructed from HO single-particle states, introducing a dependence on the HO frequency $\hbar\Omega$. Furthermore, the model space is truncated by introducing an upper limit to the number of harmonic-oscillator excitation quanta N_{\max} . The convergence behavior is controlled by the N_{\max} truncation parameter and the HO frequency $\hbar\Omega$.

The Hamiltonian consists of a kinetic energy term and NN and 3N interactions obtained from chiral EFT, where the NN interaction comes with a short-range repulsion which induces strong short-range correlations. In a Slater determinant basis, the representation of these short-range correlations requires large model spaces, as they connect high and low-momentum states leading to slow many-body convergence. The convergence can be accelerated by using an SRG transformation of the Hamiltonian [34–36], which decouples high and low-momentum states and, thus, prediagonalizes the Hamiltonian. Therefore, with increasing flow parameter α of the SRG evolution, the model-space convergence is accelerated. In this work, we utilize the SRG transformation to generate a larger variety of convergence patterns for the training of the ANNs in order to make them robust.

For nuclei with $A \leq 4$ we employ the Jacobi-NCSM [37], a reformulation of the NCSM in a HO basis constructed in Jacobi coordinates. It enables the efficient computation of few-nucleon systems up to very large values of N_{\max} , reaching fully converged ground-state energies. An example for a set of Jacobi-NCSM calculations for ^3H is presented in Fig. 1a, which shows sequences of ground-state energies as a function of N_{\max} for multiple HO frequencies that converge towards the same limit. Due to the variational character of the NCSM approach, the ground-state energies always exhibit a monotonously decreasing convergence. A reasonable approximation for the N_{\max} -dependence of the energies for a single HO frequency is an exponential function. Therefore, predictions of the converged energy are traditionally extracted through exponential fits to manually selected frequencies [12,14,38], where the frequency with the minimum energy at the largest N_{\max} accessible yields the nominal extrapolated result. The extrapolations for neighboring frequencies serve as uncertainty estimates. Clearly, this is an empirical and potentially biased process that might strongly depend on the chosen frequencies. Given the simple and systematic convergence patterns, ANNs seem to provide a powerful alternative to these schemes.

3. Artificial neural networks

Generally speaking, ANNs are useful tools for pattern recognition. Being loosely modeled after the human brain, they consist of processing units called neurons arranged in layers [39]. In a dense feedforward network, as employed in this work, all neurons in a given layer receive inputs from all neurons of the preceding layer and cast their output to all neurons of the following layer. A schematic representation of the network topology is shown in Fig. 1b. The output x_i^{out} of the i -th neuron in a specific layer is calculated from the inputs x_j^{in} via

$$x_i^{\text{out}} = \sigma \left(\sum_j x_j^{\text{in}} w_{ji} + b_i \right) \quad (2)$$

with an activation function σ along with weights w_{ji} and biases b_i which correspond to the strengths of the connections between the neurons and their internal thresholds, respectively. These parameters are usually initialized randomly and are adjusted iteratively during the training process. The training or supervised learning can be understood as a high-dimensional fit to the training data based on a minimization of the loss function, which provides a measure for the deviation between the network outputs and the target values.

In each iteration a batch of training samples is passed through the network. This step is known as forward pass. Next, the loss is determined from the network's predictions for all samples in the batch. During the following backward pass, the weights and

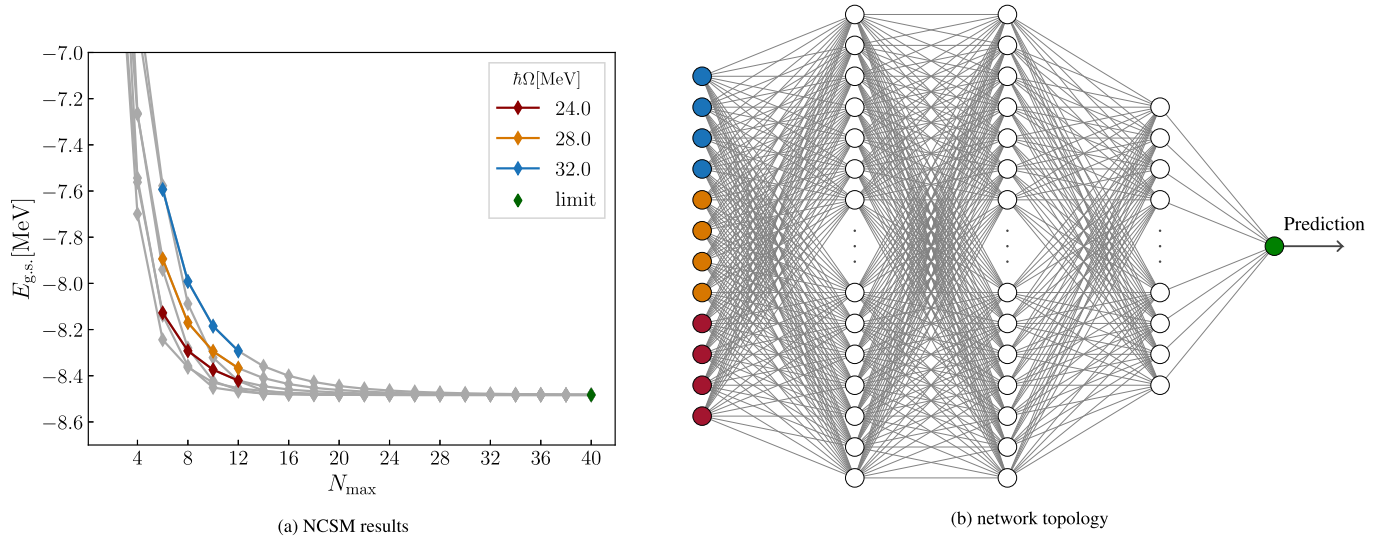


Fig. 1. (a) Exemplary results from NCSM calculations for the ground-state energy of ^3H . Shown in gray is the training data for one given interaction. The colored set of values for four consecutive N_{max} and three different HO frequencies together with the limit represents a single training sample for the ANNs. (b) Visualization of the topology of the fully connected feedforward ANN employed in this work. The colors in the first and last layer correspond to the colored sample in (a).

biases of the ANN are adjusted through a back-propagation algorithm [40]. This key process allows for a systematic optimization of every single parameter based on the gradient of the loss function, similar to steepest descent, in order to minimize the loss. Additionally, the adjustments are multiplied by a factor called learning rate, which controls the step size of the optimization process. Using a learning rate scheduler this factor will be reduced if the loss plateaus to achieve higher accuracy. Once trained, the network should have captured the convergence patterns and should provide an accurate description of the training data, leading to good predictions for unseen samples.

Note that we have introduced several so-called hyperparameters, i.e., the loss function, the back-propagation algorithm and the activation function, that come with a certain freedom of choice. In this work we employ a mean-square error (MSE) loss function, the AdamW algorithm [41], and choose a rectified linear unit (ReLU) [42] activation function. The latter has been found to yield a better performance compared to a sigmoid activation function.

4. Network design

When it comes to designing an ANN the first thing that needs to be addressed is the network topology and the structure of the data samples, as they define the size of the first (input) layer and the last (output) layer. While one neuron is sufficient for the output layer, which is supposed to provide a single real number that corresponds to the prediction for the converged observable, the ground-state energy in this case, one can think of different sets of input data that can be fed to the input layer of the ANN. In previous single-nucleus applications Negoita et al. [32] have established the two input values $\hbar\Omega$ and N_{max} and have trained their ANNs to predict or parameterize the results of NCSM calculations as a function of model-space size and HO frequency. Eventually, the trained ANNs are used to predict the result of an NCSM calculation at a large N_{max} , expected to coincide with the converged result. The functional dependence varies with nucleus and interaction and can, therefore, not be generalized.

In this work we aim for a universal ANN tool based on a different concept. We predict the actual converged value of an observable from a set of NCSM calculations in truncated model spaces. Therefore, we construct data samples from X sequences of NCSM calculations for different values of $\hbar\Omega$ with L data points for con-

secutive N_{max} , resulting in a total of $X \cdot L$ input neurons. For the prediction of the converged energy based on short sequences we choose a network topology with three hidden layers with $4X \cdot L$, $4X \cdot L$ and $8X$ neurons as depicted in Fig. 1b.

We discuss two different input modes, i.e., types of data formatting, to networks with slightly different topologies. First, we feed the raw energy eigenvalues into the ANN. The network resembles a mapping $M_{ABS} : S_{ABS} \rightarrow E^\infty$ where E^∞ is the converged energy and S_{ABS} is an input sample of the shape

$$S_{ABS}^{\mathcal{N}_{max}} = \left(E_{\hbar\Omega_1}^{\mathcal{N}_{max}-6}, E_{\hbar\Omega_1}^{\mathcal{N}_{max}-4}, E_{\hbar\Omega_1}^{\mathcal{N}_{max}-2}, E_{\hbar\Omega_1}^{\mathcal{N}_{max}}, \right. \\ \left. E_{\hbar\Omega_2}^{\mathcal{N}_{max}-6}, E_{\hbar\Omega_2}^{\mathcal{N}_{max}-4}, E_{\hbar\Omega_2}^{\mathcal{N}_{max}-2}, E_{\hbar\Omega_2}^{\mathcal{N}_{max}}, \right. \\ \left. E_{\hbar\Omega_3}^{\mathcal{N}_{max}-6}, E_{\hbar\Omega_3}^{\mathcal{N}_{max}-4}, E_{\hbar\Omega_3}^{\mathcal{N}_{max}-2}, E_{\hbar\Omega_3}^{\mathcal{N}_{max}} \right) \quad (3)$$

with \mathcal{N}_{max} being the highest N_{max} in the sample and $E_{\hbar\Omega}^{\mathcal{N}_{max}}$ being the NCSM energy eigenvalue for HO frequency $\hbar\Omega$ at a given N_{max} . This mode is referred to as “ABS”. We choose $X = 3$ and $L = 4$ and the hidden layers are scaled accordingly. The colored data points in Fig. 1a resemble one possible input sample along with the corresponding converged energy as a target value. Note that the parameter L limits the applicability of the ANN for heavier systems, as sufficiently large values of N_{max} need to be accessible through the NCSM. In our setup we need data at least up to $N_{max} = 8$ in order to be able to provide a prediction, since we choose to discard the $N_{max} = 0$ results.

A potential limitation of the ABS mode is the specific energy range of the input data. Heavier nuclei exhibit a broad range of ground-state energies that exceeds the -2 to -30 MeV range of the few-body systems that are employed in the training. One might consider to use the ABS mode in connection to the ground-state energies per nucleon to remedy this issue and we will come back to this idea later.

Further, we design a second mode, referred to as “DIFF”, for which the input is constructed from the energy differences between two consecutive N_{max} points. With this concept we aim to avoid any dependencies on the energy range in order to further improve the ANN predictions. This mode resembles a mapping $M_{DIFF} : S_{DIFF} \rightarrow \Delta^\infty$ where the input samples are given by

$$S_{\text{DIFF}}^{\mathcal{N}_{\text{max}}} = \left(\Delta_{\hbar\Omega_1}^{\mathcal{N}_{\text{max}}-4}, \Delta_{\hbar\Omega_1}^{\mathcal{N}_{\text{max}}-2}, \Delta_{\hbar\Omega_1}^{\mathcal{N}_{\text{max}}}, \right. \\ \left. \Delta_{\hbar\Omega_2}^{\mathcal{N}_{\text{max}}-4}, \Delta_{\hbar\Omega_2}^{\mathcal{N}_{\text{max}}-2}, \Delta_{\hbar\Omega_2}^{\mathcal{N}_{\text{max}}}, \right. \\ \left. \Delta_{\hbar\Omega_3}^{\mathcal{N}_{\text{max}}-4}, \Delta_{\hbar\Omega_3}^{\mathcal{N}_{\text{max}}-2}, \Delta_{\hbar\Omega_3}^{\mathcal{N}_{\text{max}}} \right) \quad (4)$$

with $\Delta_{\hbar\Omega}^{\mathcal{N}_{\text{max}}} = E_{\hbar\Omega}^{\mathcal{N}_{\text{max}}} - E_{\hbar\Omega}^{\mathcal{N}_{\text{max}}-2}$. The networks output will then be the difference from the lowest lying input energy to the predicted converged value, i.e.

$$\Delta^\infty = E^\infty - \min(E_{\hbar\Omega}^{\mathcal{N}_{\text{max}}}) \quad \text{for } \hbar\Omega \text{ in } S_{\text{DIFF}}^{\mathcal{N}_{\text{max}}}. \quad (5)$$

Note that the input layer for this mode has only $X \cdot (L - 1)$ input neurons to ensure that both ANN types employ the same number of raw NCSM results for each sample. Analogously, the hidden layers are also scaled with $(L - 1)$ instead of L .

5. Data preparation and training

As mentioned above, we aim to predict the ground-state energies of a broad range of p-shell nuclei based on data from few-body systems. Our training data consists of ${}^2\text{H}$, ${}^3\text{H}$, and ${}^4\text{He}$ calculations up to $N_{\text{max}} = 50, 40$, and 24 , respectively, for seven HO frequencies from $\hbar\Omega = 12$ to 32 MeV. The calculations were carried out for a family of non-local NN+3N interactions from chiral EFT at N^2LO , N^3LO and N^4LO and for three cutoffs $\Lambda = 450, 500$, and 550 MeV [11,43]. All interactions have been applied bare and SRG evolved with three different flow parameters $\alpha = 0.02 \text{ fm}^4$, 0.04 fm^4 , and 0.08 fm^4 . This accumulates to a total of 756 converging sequences, on which the ANNs can be trained. In order to study the impact of the selection of the training data, we also employ different physics-motivated filters on this data set, which are discussed in the result sections. For using the sequences in the training process, they need to be converted into samples that match the input layer of the networks. For each nucleus and interaction all subsets of X frequencies and their permutations are generated, before samples of L consecutive N_{max} are constructed. As mentioned earlier, we exclude $N_{\text{max}} = 0$ results from the training.

Since more data is beneficial for the quality of the trained ANN, additional samples are generated by randomly scaling and shifting the existing samples via

$$a \cdot S_{\text{ABS}}^{\mathcal{N}_{\text{max}}} + b \quad \text{and} \quad a \cdot S_{\text{DIFF}}^{\mathcal{N}_{\text{max}}} \quad (6)$$

for ABS and DIFF modes, respectively, where $a \in (0.25, 4)$ and $b \in (-20, 20)$ MeV for ABS and $a \in (0.5, 2)$ for DIFF are random numbers that are uniformly distributed over the given intervals. In this way an arbitrary number of training samples, which slightly differ in their convergence behavior and are spread over a broader energy range, can be generated. It also prevents the ANN from learning and reproducing only the three target values corresponding to the three training nuclei, which could be expected since all employed interactions produce very similar predictions for the ground-state energies of these nuclei. The distributions of the target values, shown in Fig. 2, indicate that the majority of training samples is still in the range of the few-body systems, but also samples with ground-state energies down to -150 MeV are represented in the training data. This is critical for the ABS mode as the ground-state energies of p-shell nuclei are significantly lower than energies of the training nuclei. Hence, we can avoid an inherent bias of the predictions. Similar effects apply to the DIFF mode, where the scaled training data covers corrections beyond -0.5 MeV instead of a few keV.

For the actual training three disjoint sets of samples are generated: (1) the training set consisting of 1 000 000 samples, (2) the development set consisting of 10 000 samples, and (3) the validation set consisting of 50 000 samples. The ANNs are trained on the

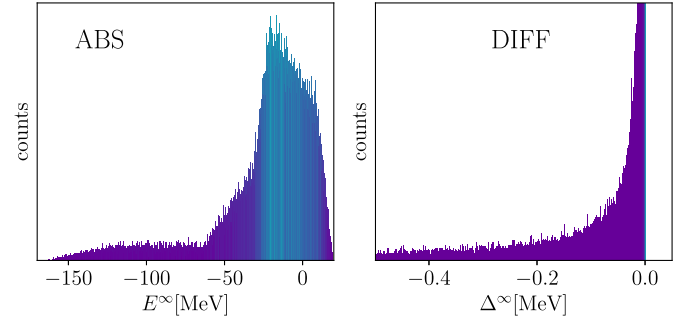


Fig. 2. Distributions of the target output values of all training samples for ABS and DIFF mode. A random scaling has been applied in order to spread the values over a larger energy range to avoid training the ANN on specific values. See text for details.

training set in batches of 512 samples at a time and we iterate through the training data a total of 20 times. During this process the average loss is monitored via an evaluation of the development set after every iteration in order to adjust any loss-dependent parameters like the learning rate, which is initialized at 0.01. Once the training is complete, we decide to keep or retrain the network based on its performance on the validation set. We found that the training is very robust and rarely any networks have to be discarded. The ANNs reach an average accuracy of about 60 keV (6 keV) for the ABS (DIFF) mode on the validation set. This indicates that the description of the training samples is more accurate when given as differences, due to the much smaller range of input values compared to the ABS method.

6. Statistical evaluation

The evaluation of the ANNs is done analogously to the training. We start from a small set of evaluation data from NCSM calculations for a specific nucleus, interaction, and eigenstate. Again, all possible subsets of $X = 3$ HO frequencies and their permutations are generated, from which input samples of $L = 4$ consecutive N_{max} are constructed. The samples are sorted by \mathcal{N}_{max} which describes the largest value of N_{max} present in the given sample. A network then evaluates all samples $S^{\mathcal{N}_{\text{max}}}$ for one given \mathcal{N}_{max} at a time. By increasing \mathcal{N}_{max} we construct a sequence of predictions using information from successively larger model spaces. In many relevant applications, the evaluation data will be limited to moderate values of N_{max} , typically $N_{\text{max}} \leq 12$ or even less. Therefore, we restrict our investigations to evaluation samples with $\mathcal{N}_{\text{max}} \leq 12$.

As one finds multiple samples for a given \mathcal{N}_{max} , we obtain multiple predictions from a single ANN, which we treat as equally plausible. In addition, we construct a total of 1000 networks with comparable performance on the validation set and run all the evaluation samples through all networks. The resulting histogram of all predictions obtained in this statistical evaluation shows a single peak structure reminiscent of a Gaussian distribution. Therefore, the final prediction for the ground-state energy can be characterized through a mean value and a standard deviation—this serves as nominal prediction for the converged ground-state energy along with a quantified statistical uncertainty. Fig. 3 shows examples for

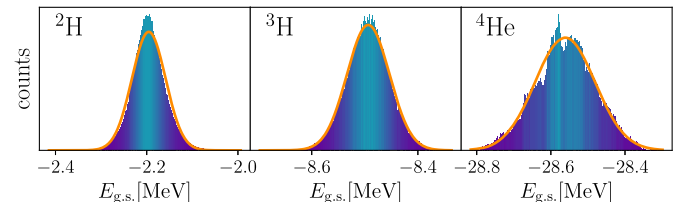


Fig. 3. Histograms of exemplary predictions from 1000 ANNs for ${}^2\text{H}$, ${}^3\text{H}$ and ${}^4\text{He}$ at $\mathcal{N}_{\text{max}} = 12$ with fitted Gaussian distributions.

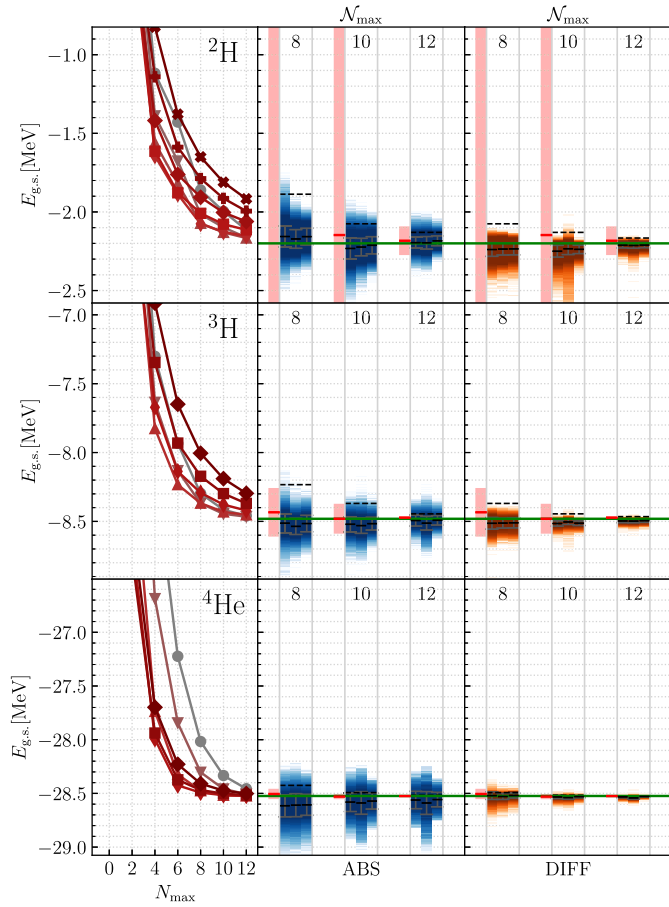


Fig. 4. Input data and predicted ground-state energies for ${}^2\text{H}$, ${}^3\text{H}$ and ${}^4\text{He}$. The left-hand column shows the input data for multiple HO frequencies around the variational minimum. In the middle and right-hand columns show the statistical predictions are shown as color-coded vertical histograms at a given N_{max} . The groups of three histograms correspond to the filters (a) to (c) on the training data (see text for details). The networks were trained and evaluated with either the ABS (middle) or DIFF (right) mode. The dashed black lines (---) are the variational boundaries for the respective model space and the horizontal green line (—) indicates the converged value. For comparison classical extrapolations are given as red based with uncertainty bands (-).

predictions for the training nuclei along with a Gaussian distribution obtained from the mean value and standard deviation, which represent the distribution well. In case of a multi-peak structure in the histogram one should look for an imbalance in either the training data or the evaluation data as discussed in [33].

7. Application to few-body systems

First, we will take a look at the networks' performance for the few-body nuclei used in the training. For investigations of these nuclei we need to ensure that we do not reproduce any of the training samples. We circumvent this by generating the evaluation data from another family of interactions that was not used for the training data. All NCSM results shown in the following are obtained from a realistic semi-local $\text{NN}+3\text{N}$ interaction from chiral EFT at N^2LO with cutoff $\Lambda = 450$ MeV [44,45] and SRG flow parameter $\alpha = 0.08 \text{ fm}^4$.

As mentioned earlier, we artificially truncate the evaluation data set at $N_{\text{max}} = 8, 10, 12$ to probe the robustness and consistency of the ANN prediction. We further employ different physics-motivated filters on the training and evaluation data in order to study the dependence of the predictions on the data selection. These are (a) all data available, (b) limitation of training data to

Table 1

ANN predictions for ABS and DIFF modes for all discussed nuclei. The ABS' results are obtained using the ABS mode with the ground-state energy per nucleon for training and evaluation. Variational bounds, classical extrapolations and, where available, converged results are given for comparison.

| | $N_{\text{max}} = 8$ | 10 | 12 |
|-------------------|----------------------|---------------|--------------|
| ${}^2\text{H}$ | | | |
| var. bound. | -1.887 | -2.075 | -2.130 |
| extrapol. | -2.720(*) | -2.147(*) | -2.183(91) |
| ABS | -2.156(69) | -2.223(64) | -2.196(37) |
| ABS' | -2.144(67) | -2.217(60) | -2.191(28) |
| DIFF | -2.239(43) | -2.250(38) | -2.212(17) |
| conv. result | -2.200 | -2.200 | -2.200 |
| ${}^3\text{H}$ | | | |
| var. bound. | -8.233 | -8.369 | -8.445 |
| extrapol. | -8.434(176) | -8.480(109) | -8.473(10) |
| ABS | -8.513(68) | -8.521(55) | -8.494(40) |
| ABS' | -8.503(76) | -8.516(47) | -8.495(29) |
| DIFF | -8.514(42) | -8.515(19) | -8.497(13) |
| conv. result | -8.481 | -8.481 | -8.481 |
| ${}^4\text{He}$ | | | |
| var. bound. | -28.425 | -28.493 | -28.517 |
| extrapol. | -28.505(48) | -28.530(24) | -28.525(9) |
| ABS | -28.615(103) | -28.581(87) | -28.562(81) |
| ABS' | -28.556(124) | -28.580(80) | -28.567(67) |
| DIFF | -28.521(36) | -28.533(15) | -28.532(8) |
| conv. result | -28.524 | -28.524 | -28.524 |
| ${}^6\text{Li}$ | | | |
| var. bound. | -31.014 | -31.535 | -31.813 |
| extrapol. | -31.718(278) | -31.906(189) | -32.098(317) |
| ABS | -31.653(324) | -31.770(227) | -31.964(185) |
| ABS' | -31.804(437) | -31.915(250) | -32.134(165) |
| DIFF | -31.699(186) | -31.919(106) | -32.126(147) |
| ${}^{12}\text{C}$ | | | |
| var. bound. | -97.00 | -97.42 | - |
| extrapol. | -97.55(150) | -97.43(44) | - |
| ABS | -97.59(43) | -97.25(35) | - |
| ABS' | -98.54(48) | -97.59(35) | - |
| DIFF | -97.88(18) | -97.64(7) | - |
| ${}^{16}\text{O}$ | | | |
| var. bound. | -145.85 | -146.84 | - |
| extrapol. | -148.35(134) | -147.28(1.86) | - |
| ABS | -147.31(64) | -146.96(47) | - |
| ABS' | -149.37(78) | -147.77(45) | - |
| DIFF | -147.96(52) | -147.46(12) | - |

Uncertainties denoted as (*) are unreasonably large due to a breakdown of the classical extrapolation method.

$N_{\text{max}} \leq 12$ as this is the relevant range for actual applications, and (c) omission of training data obtained with bare interactions, i.e., interactions that are not SRG evolved, as SRG transformed Hamiltonians are used in the applications.

Fig. 4 shows the results for the few-body nuclei and the corresponding numerical values are given in Table 1. The left-hand panels show the evaluation data that has been used as input for the ANNs, and the panels in the middle and on the right-hand side show the distributions of predictions from the ABS and DIFF modes, respectively. The solid black lines with error bars indicate the mean and standard deviation of the predictions. For comparison, classical extrapolations that are performed analogously to the procedure described in [12] are given as red bars with error bands. The green horizontal lines indicate the converged values, which have been obtained for very large N_{max} .

First of all, we observe that the ANN predictions are in good agreement with the known converged ground-state energies. Furthermore, two important features emerge: the variational boundaries are respected by the averaged predictions and the predictions become more accurate with increasing N_{max} , while the uncer-

tainty estimates are consistent with each other. Hence, we can obtain robust predictions from small model spaces already. On closer inspection, we find that the DIFF method yields significantly smaller uncertainties and, thus, more precise results compared to the ABS method, while the predictions are equally or even more accurate with respect to the known converged energies. One reason for the improved precision of the DIFF mode might be that the range of input values seen by the network is much smaller when using energy differences and does not change systematically from one nucleus to the other. In a broader sense this resembles a normalization of the data as it is often employed in machine learning applications [46].

As mentioned earlier, we can apply the ABS method to the ground-state energies per nucleon instead of the ground-state energy with the aim to reduce the range of possible target energies. We refer to this variant as ABS' and summarize results in Table 1. Contrary to expectations this scheme does not yield improvements compared to the ABS method, neither w.r.t. the mean predictions nor regarding the uncertainty estimates. It even produces a significant overbinding for some nuclei in small model spaces. Therefore, we refrain from discussing this scheme in more detail.

Compared to the classical extrapolations the ANN predictions, especially for the DIFF method, are more consistent while being in good agreement with the classical extrapolations. Note that the uncertainties for the classical extrapolations for ${}^2\text{H}$ up to $\mathcal{N}_{\text{max}} = 10$ are unreasonably large. This is due to the deuteron-specific convergence behavior that cannot be modeled well with exponential functions.

Regarding the filtered data sets (b) and (c) we find very little deviation from the unrestricted data set (a). Hence, the ANNs seem to be able to identify the required information from the training data making them robust against changes of the training set. In other words, the predictions are not significantly biased by the choice of training data. We note that there is no systematic improvement from a manual preselection of the best HO frequencies for the evaluation as it is done for the classical extrapolation. However, evaluating the ANNs with frequencies far from the variational energy minimum provides less accurate predictions and, therefore, the input data should use frequencies around this optimum.

8. Application to p-shell nuclei

We now proceed to the heavier p-shell nuclei ${}^6\text{Li}$, ${}^{12}\text{C}$ and ${}^{16}\text{O}$, which were not part of the training process. The results for ground-state energies are depicted in Fig. 5 and the numerical values are given in Table 1. Starting with ${}^6\text{Li}$ we still find consistent predictions that are in very good agreement with the classical extrapolations. Both, the ABS and DIFF modes produce very similar results. We note that some of the distributions exhibit tails towards lower energies, which indicates that a Gaussian is not the ideal choice for modeling the uncertainties. One could turn to a Bayesian treatment with more complicated probability density functions as suggested in [33].

The last nuclei we discuss are ${}^{12}\text{C}$ and ${}^{16}\text{O}$. Since they are much heavier, it is more challenging to converge the NCSM calculations and, therefore, these nuclei resemble perfect examples for the applications this machine learning tool is aiming for. We again find very consistent predictions for the ground-state energies that agree well with the classical extrapolations. However, some predictions for the ABS method at $\mathcal{N}_{\text{max}} = 10$ violate the variational bounds. The DIFF method on the other hand produces very accurate predictions for both nuclei that are compatible with independent results presented in [12].

For all p-shell nuclei, the predictions with the different filtered data sets do not differ much from each other. There might be a slight advantage in precision and consistency for filter (c), which

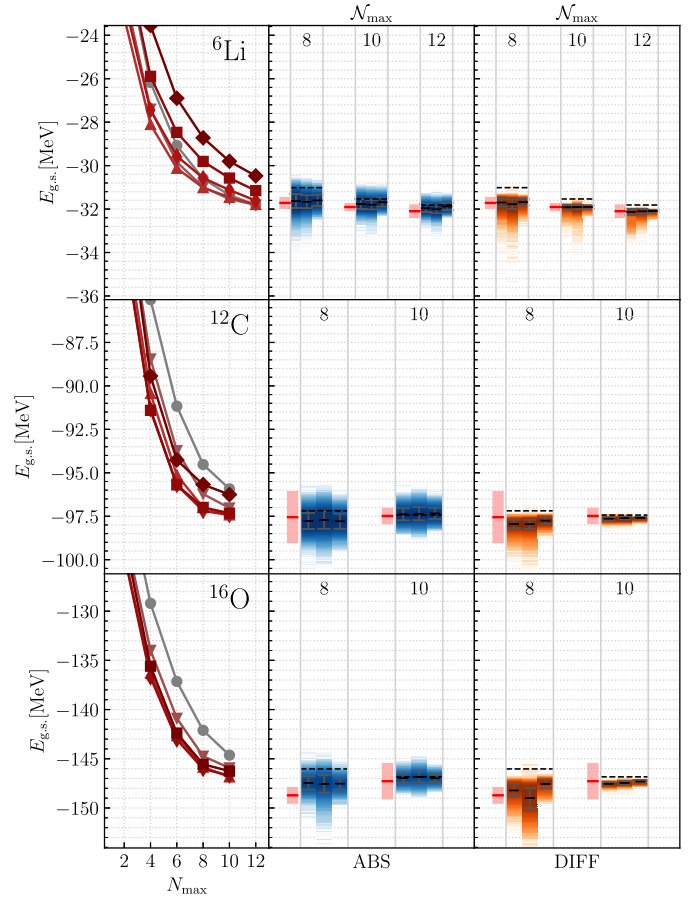


Fig. 5. Same as Fig. 4 but for ${}^6\text{Li}$, ${}^{12}\text{C}$ and ${}^{16}\text{O}$.

omits the bare interaction results from the training, but this might change for other evaluation sets. Generally, we recommend a full survey of all filters and all accessible \mathcal{N}_{max} for future applications, in order to check for consistency and robustness of the predictions.

9. Conclusions

We have developed a new machine learning tool for the prediction of converged ground-state energies based on input data from non-converged NCSM calculations. The key idea is to construct a universal ANN to identify the convergence pattern on NCSM calculations for arbitrary nuclei and interactions and to predict the converged ground-state energy directly. The training is based on a huge set of NCSM results for few-body systems with different chiral NN+3N interactions and SRG evolutions, where in each case the converged results are precisely known. The evaluation of the ANNs is embedded into a statistical framework that provides predictions with quantified uncertainties. We have shown that NCSM calculations for light systems up to $A = 4$ contain enough information on the convergence patterns to provide reliable predictions with competitive uncertainties for heavier nuclei. The quality of the predictions exceeds the capabilities of traditional extrapolations and is comparable with previous ANN extrapolations [32], which were trained specifically for a single nucleus and interaction. It is important to note that the ANNs presented here are only trained once, and are then ready for a broad range of applications in future calculations.

The scheme presented here can be easily transferred to other ab initio many-body approaches. It can also be generalized to other observables—work for the prediction of converged rms-radii is already in progress. In these cases the convergence behavior is not

constrained by the variational principle and, therefore, more complex. An important next development step to tackle other observables and to further improve the precision of energy predictions is, e.g., the normalization of the input sequences beyond the DIFF mode presented here. Furthermore, we can enrich the variety of training data by considering synthetic nuclei obtained with modified Hamiltonians, which cover a controlled range for converged observables. Work along these lines is in progress.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

This work is supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through the DFG Sonderforschungsbereich SFB 1245 (Project ID 279384907) and the BMBF through Verbundprojekt 05P2021 (ErUM-FSP T07, Contract No. 05P21RDFNB). Numerical calculations have been performed on the LICHTENBERG II cluster at the computing center of the TU Darmstadt.

References

- [1] B.R. Barrett, P. Navrátil, J.P. Vary, Ab initio no core shell model, *Prog. Part. Nucl. Phys.* 69 (2013) 131–181, <https://doi.org/10.1016/j.pnpnp.2012.10.003>.
- [2] P. Navrátil, S. Quaglioni, I. Stetcu, B. Barrett, Recent developments in no-core shell-model calculations, *J. Phys. G, Nucl. Part. Phys.* 36 (2009) 083101, <https://doi.org/10.1088/0954-3899/36/8/083101>.
- [3] R. Roth, Importance truncation for large-scale configuration interaction approaches, *Phys. Rev. C* 79 (2009) 064324, <https://doi.org/10.1103/PhysRevC.79.064324>.
- [4] D.C. Zheng, B.R. Barrett, L. Jaqua, J.P. Vary, R.J. McCarthy, Microscopic calculations of the spectra of light nuclei, *Phys. Rev. C* 48 (3) (1993) 1083, <https://doi.org/10.1103/PhysRevC.48.1083>.
- [5] K. Kowalski, D.J. Dean, M. Hjorth-Jensen, T. Papenbrock, P. Piecuch, Coupled cluster calculations of ground and excited states of nuclei, *Phys. Rev. Lett.* 92 (13) (2004) 132501, <https://doi.org/10.1103/PhysRevLett.92.132501>.
- [6] W. Dickhoff, C. Barbieri, Self-consistent Green's function method for nuclei and nuclear matter, *Prog. Part. Nucl. Phys.* 52 (2) (2004) 377–496, <https://doi.org/10.1016/j.pnpnp.2004.02.038>.
- [7] H. Hergert, S. Bogner, T. Morris, A. Schwenk, K. Tsukiyama, The in-medium similarity renormalization group: a novel ab initio method for nuclei, *Phys. Rep.* 621 (2016) 165, <https://doi.org/10.1016/j.physrep.2015.12.007>.
- [8] J. Carlson, S. Gandolfi, F. Pederiva, S.C. Pieper, R. Schiavilla, K.E. Schmidt, R.B. Wiringa, Quantum Monte Carlo methods for nuclear physics, *Rev. Mod. Phys.* 87 (2015) 1067, <https://doi.org/10.1103/RevModPhys.87.1067>.
- [9] R. Machleidt, D.R. Entem, Chiral effective field theory and nuclear forces, *Phys. Rep.* 503 (2011) 1, <https://doi.org/10.1016/j.physrep.2011.02.001>.
- [10] E. Epelbaum, H.-W. Hammer, U.-G. Meißner, Modern theory of nuclear forces, *Rev. Mod. Phys.* 81 (2009) 1773, <https://doi.org/10.1103/RevModPhys.81.1773>.
- [11] T. Hübner, K. Vobig, K. Hebeler, R. Machleidt, R. Roth, Family of chiral two-plus three-nucleon interactions for accurate nuclear structure studies, *Phys. Lett. B* 808 (2020) 135651, <https://doi.org/10.1016/j.physletb.2020.135651>.
- [12] P. Maris, E. Epelbaum, R. Furnstahl, J. Golak, K. Hebeler, T. Hübner, H. Kamada, H. Krebs, U.-G. Meißner, J. Melendez, et al., Light nuclei with semilocal momentum-space regularized chiral interactions up to third order, *Phys. Rev. C* 103 (5) (2021) 054001, <https://doi.org/10.1103/PhysRevC.103.054001>.
- [13] H. Hergert, S.K. Bogner, S. Binder, A. Calci, J. Langhammer, R. Roth, A. Schwenk, In-medium similarity renormalization group with chiral two- plus three-nucleon interactions, *Phys. Rev. C* 87 (2013) 034307, <https://doi.org/10.1103/PhysRevC.87.034307>.
- [14] P. Maris, J.P. Vary, A.M. Shirokov, Ab initio no-core full configuration calculations of light nuclei, *Phys. Rev. C* 79 (1) (2009) 014308, <https://doi.org/10.1103/PhysRevC.79.014308>.
- [15] S.K. Bogner, R.J. Furnstahl, R.J. Perry, Similarity renormalization group for nucleon-nucleon interactions, *Phys. Rev. C* 75 (2007) 061001(R), <https://doi.org/10.1103/PhysRevC.75.061001>.
- [16] S.K. Bogner, R.J. Furnstahl, A. Schwenk, From low-momentum interactions to nuclear structure, *Prog. Part. Nucl. Phys.* 65 (2010) 94–147, <https://doi.org/10.1016/j.pnpnp.2010.03.001>.
- [17] S.K. Bogner, R.J. Furnstahl, P. Maris, R.J. Perry, A. Schwenk, J.P. Vary, Convergence in the no-core shell model with low-momentum two-nucleon interactions, *Nucl. Phys. A* 801 (2008) 21, <https://doi.org/10.1016/j.nuclphysa.2007.12.008>.
- [18] R.J. Furnstahl, G. Hagen, T. Papenbrock, Corrections to nuclear energies and radii in finite oscillator spaces, *Phys. Rev. C* 86 (3) (2012) 031301, <https://doi.org/10.1103/PhysRevC.86.031301>.
- [19] S.A. Coon, M.I. Avetian, M.K.G. Kruse, U. van Kolck, P. Maris, J.P. Vary, Convergence properties of ab initio calculations of light nuclei in a harmonic oscillator basis, *Phys. Rev. C* 86 (5) (2012) 054002, <https://doi.org/10.1103/PhysRevC.86.054002>.
- [20] S.N. More, A. Ekström, R.J. Furnstahl, G. Hagen, T. Papenbrock, Universal properties of infrared oscillator basis extrapolations, *Phys. Rev. C* 87 (4) (2013) 044326, <https://doi.org/10.1103/PhysRevC.87.044326>.
- [21] R.J. Furnstahl, S.N. More, T. Papenbrock, Systematic expansion for infrared oscillator basis extrapolations, *Phys. Rev. C* 89 (2014) 044301, <https://doi.org/10.1103/PhysRevC.89.044301>.
- [22] K.A. Wendt, C. Forssén, T. Papenbrock, D. Sääf, Infrared length scale and extrapolations for the no-core shell model, *Phys. Rev. C* 91 (2015) 061301, <https://doi.org/10.1103/PhysRevC.91.061301>.
- [23] S. Athanassopoulos, E. Mavrommatis, K. Gernoth, J.W. Clark, Nuclear mass systematics using neural networks, *Nucl. Phys. A* 743 (4) (2004) 222–235, <https://doi.org/10.1016/j.nuclphysa.2004.08.006>.
- [24] S. Akkoyun, T. Bayram, S.O. Kara, A. Sinan, An artificial neural network application on nuclear charge radii, *J. Phys. G, Nucl. Part. Phys.* 40 (5) (2013) 055106, <https://doi.org/10.1088/0954-3899/40/5/055106>.
- [25] R. Utama, W.-C. Chen, J. Piekarewicz, Nuclear charge radii: density functional theory meets Bayesian neural networks, *J. Phys. G, Nucl. Part. Phys.* 43 (11) (2016) 114002, <https://doi.org/10.1088/0954-3899/43/11/114002>.
- [26] C. Adams, G. Carleo, A. Lovato, N. Rocco, Variational Monte Carlo calculations of $A < 4$ nuclei with an artificial neural-network correlator ansatz, *Phys. Rev. Lett.* 127 (2) (2021) 022502, <https://doi.org/10.1103/PhysRevLett.127.022502>.
- [27] A. Ekström, C. Forssén, C. Dimitrakakis, D. Dubhashi, H.T. Johansson, A.S. Muhammad, H. Salomonsson, A. Schliep, Bayesian optimization in ab initio nuclear physics, *J. Phys. G, Nucl. Part. Phys.* 46 (9) (2019) 095101, <https://doi.org/10.1088/1361-6471/ab2b14>.
- [28] K.A. Wendt, C. Forssén, T. Papenbrock, D. Sääf, Nonparametric Bayesian approach to extrapolation problems in configuration interaction method, *Phys. Rev. C* 102 (2) (2020) 024305, <https://doi.org/10.1103/PhysRevC.102.024305>.
- [29] J.W. Clark, T. Lindenau, M.L. Ristig, Scientific applications of neural nets, Springer, 1999, <https://doi.org/10.1007/BFb0104276>.
- [30] A. Boehnlein, M. Diefenthaler, N. Sato, M. Schram, V. Ziegler, C. Fanelli, M. Hjorth-Jensen, T. Horn, M.P. Kuchera, D. Lee, et al., Colloquium: Machine learning in nuclear physics, *Rev. Mod. Phys.* 94 (3) (2022) 031003, <https://doi.org/10.1103/RevModPhys.94.031003>.
- [31] K. Xu, M. Zhang, J. Li, S.S. Du, K.-i. Kawarabayashi, S. Jegelka, How neural networks extrapolate: from feedforward to graph neural networks, arXiv preprint, arXiv:2009.11848, <https://doi.org/10.48550/arXiv.2009.11848>.
- [32] G.A. Negoita, J.P. Vary, G.R. Luecke, P. Maris, A.M. Shirokov, I.J. Shin, Y. Kim, E.G. Ng, C. Yang, M. Lockner, G.M. Prabhu, Deep learning: extrapolation tool for ab initio nuclear theory, *Phys. Rev. C* 99 (2019) 054308, <https://doi.org/10.1103/PhysRevC.99.054308>.
- [33] W.G. Jiang, G. Hagen, T. Papenbrock, Extrapolation of nuclear structure observables with artificial neural networks, *Phys. Rev. C* 100 (2019) 054326, <https://doi.org/10.1103/PhysRevC.100.054326>.
- [34] R. Roth, T. Neff, H. Feldmeier, Nuclear structure in the framework of the unitary correlation operator method, *Prog. Part. Nucl. Phys.* 65 (2010) 50, <https://doi.org/10.1016/j.pnpnp.2010.02.003>.
- [35] R.J. Furnstahl, K. Hebeler, New applications of renormalization group methods in nuclear physics, *Rep. Prog. Phys.* 76 (12) (2013) 126301, <https://doi.org/10.1088/0034-4885/76/12/126301>.
- [36] R. Roth, A. Calci, J. Langhammer, S. Binder, Evolved chiral $nn + 3n$ Hamiltonians for ab initio nuclear structure calculations, *Phys. Rev. C* 90 (2014) 024325, <https://doi.org/10.1103/PhysRevC.90.024325>.
- [37] S. Liebig, U.-G. Meißner, A. Nogga, Jacobi no-core shell model for p-shell nuclei, *Eur. Phys. J. A* 52 (4) (2016) 1–18, <https://doi.org/10.1140/epja/i2016-16103-5>.
- [38] E.D. Jurgenson, P. Maris, R.J. Furnstahl, P. Navrátil, W.E. Ormand, J.P. Vary, Structure of p-shell nuclei using three-nucleon interactions evolved with the similarity renormalization group, *Phys. Rev. C* 87 (5) (2013) 054312, <https://doi.org/10.1103/PhysRevC.87.054312>.
- [39] I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning*, MIT Press, 2016.
- [40] R. Rojas, The backpropagation algorithm, in: *Neural Networks*, Springer, 1996, pp. 149–182, https://doi.org/10.1007/978-3-642-61068-4_7.
- [41] I. Loshchilov, F. Hutter, Decoupled weight decay regularization, arXiv preprint, arXiv:1711.05101, <https://doi.org/10.48550/arXiv.1711.05101>.

- [42] V. Nair, G.E. Hinton, Rectified linear units improve restricted Boltzmann machines, in: *Icml*, 2010.
- [43] D.R. Entem, R. Machleidt, Y. Nosyk, High-quality two-nucleon potentials up to fifth order of the chiral expansion, *Phys. Rev. C* 96 (2017) 024004, <https://doi.org/10.1103/PhysRevC.96.024004>.
- [44] P. Reinert, H. Krebs, E. Epelbaum, Semilocal momentum-space regularized chiral two-nucleon potentials up to fifth order, *Eur. Phys. J. A* 54 (5) (2018) 1–49, <https://doi.org/10.1140/epja/i2018-12516-4>.
- [45] P. Maris, E. Epelbaum, R. Furnstahl, J. Golak, K. Hebeler, T. Hüther, H. Kamada, H. Krebs, U.-G. Meißner, J. Melendez, et al., Light nuclei with semilocal momentum-space regularized chiral interactions up to third order, *Phys. Rev. C* 103 (5) (2021) 054001, <https://doi.org/10.1103/PhysRevC.103.054001>.
- [46] Y.A. LeCun, L. Bottou, G.B. Orr, K.-R. Müller, Efficient backprop, in: *Neural Networks: Tricks of the Trade*, Springer, 2012, pp. 9–48, https://doi.org/10.1007/3-540-49430-8_2.