TECHNISCHE
UNIVERSITÄT
DARMSTADT

Precision No-Core Shell Model Calculations with Artificial Neural Networks

Vom Fachbereich Physik
der Technischen Universität Darmstadt

zur Erlangung des Grades
Doctor rerum naturalium
(Dr. rer. nat.)

genehmigte Dissertation von
Tobias Wolfgruber, M.Sc.

Referent: Prof. Dr. Robert Roth
Korreferent: Priv.-Doz. Dr. Kai Hebeler

Darmstadt 2026
D17

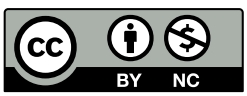
Precision No-Core Shell Model Calculations with Artificial Neural Networks

Präzise No-Core Schalenmodellrechnungen mit Künstlichen Neuronalen Netzen

Genehmigte Dissertation von Tobias Wolfgruber, M.Sc.
Darmstadt, Technische Universität Darmstadt

Referent: Prof. Dr. Robert Roth
Korreferent: Priv.-Doz. Dr. Kai Hebeler

Wolfgruber, Tobias: Precision No-Core Shell Model Calculations with Artificial Neural Networks
Darmstadt, Technische Universität Darmstadt
Jahr der Veröffentlichung der Dissertation auf TUprints: 2026
DOI: https://doi.org/10.26083/tuda-7951
URN: https://nbn-resolving.org/urn:nbn:de:tuda-tuda-153683
Tag der mündlichen Prüfung: 13.04.2026



Abstract

Achieving quantitatively reliable *ab initio* predictions of nuclear structure observables remains a central challenge in theoretical nuclear physics, particularly for quantities that exhibit slow convergence in many-body calculations, such as nuclear radii. The aim of this thesis is to enable high-precision *ab initio* radius calculations in light nuclei using state-of-the-art interactions derived from chiral effective field theory and the no-core shell model (NCSM), while overcoming the practical limitations imposed by incomplete model-space convergence.

To this end, we develop a machine-learning–based framework that employs artificial neural networks to infer infinite model-space limits from partially converged NCSM calculations. Using large data sets of *ab initio* results, the networks are trained to predict converged ground-state energies and radii from truncated model-space inputs. A statistically robust evaluation scheme is introduced that yields probability distributions for the predicted observables, enabling a robust extraction of converged values together with associated many-body and extrapolation uncertainties. The framework is further enhanced by a dedicated post-processing strategy that exploits correlations between related observables, allowing for high-precision predictions of radius differences both within individual nuclei and between isotopes. In addition, we present a coherent approach for combining the network-derived many-body uncertainties with order-by-order truncation uncertainties of the chiral effective field theory interactions.

We show a benchmark study that demonstrates that our proposed method significantly outperforms established extrapolation techniques, including both classical approaches and previous neural-network–based methods, particularly in small model spaces where conventional extrapolations are least reliable.

As a physics application, we apply the framework to a comprehensive *ab initio* investigation of ground-state energies and radii of boron isotopes using two families of chiral nucleon-nucleon and three-nucleon interactions. In particular, we study ^{8}B as a candidate for a proton halo and provide high-precision theoretical predictions for squared proton radius differences between boron isotopes, facilitating direct comparison with current and future isotope-shift experiments.

Zusammenfassung

Präzise No-Core Schalenmodellrechnungen mit Künstlichen Neuronalen Netzen

Die Erzielung quantitativ zuverlässiger *ab initio*-Vorhersagen für Kernstrukturobservablen stellt nach wie vor eine zentrale Herausforderung der theoretischen Kernphysik dar, insbesondere für Größen, die in Vielteilchenrechnungen nur langsam konvergieren, wie etwa Kernradien. Ziel dieser Arbeit ist es, hochpräzise *ab initio*-Berechnungen von Kernradien in leichten Kernen unter Verwendung moderner Wechselwirkungen aus der chiralen effektiven Feldtheorie und des No-Core Schalenmodell (NCSM) zu ermöglichen und dabei die praktischen Einschränkungen durch unvollständige Modellraumkonvergenz zu überwinden.

Zu diesem Zweck entwickeln wir einen auf maschinellem Lernen basierenden Ansatz, der auf künstliche neuronale Netzen aufbaut, um Grenzwerte im unendlichen Modellraum aus unvollständig konvergierten NCSM-Berechnungen zu bestimmen. Auf Grundlage umfangreicher Datensätze von *ab initio*-Ergebnissen werden die Netze darauf trainiert, konvergierte Grundzustandsenergien sowie Kernradien aus Daten aus trunkierten Modellräumen vorherzusagen. Es wird ein statistisch robustes Auswertungsschema eingeführt, das Wahrscheinlichkeitsverteilungen für die vorhergesagten Observablen liefert und damit eine robuste Bestimmung der konvergierten Werte zusammen mit den zugehörigen Vielteilchen- und Extrapolationsunsicherheiten erlaubt. Der Ansatz wird durch eine gezielte Postprocessing-Prozedur erweitert, welche Korrelationen zwischen verwandten Observablen ausnutzt und so hochpräzise Vorhersagen für Radiusdifferenzen sowohl innerhalb einzelner Kerne als auch zwischen Isotopen ermöglicht. Darüber hinaus präsentieren wir ein konsistentes Verfahren zur Kombination der netzwerkbasierten Viele-Teilchen-Unsicherheiten mit den ordnungsweisen Trunkationsunsicherheiten der chiralen effektiven Feldtheorie-Wechselwirkungen.

In einem Vergleich mit anderen Methoden zeigen wir, dass der vorgeschlagene Ansatz etablierte Extrapolationsmethoden, einschließlich sowohl klassischer Verfahren als auch früherer neuronaler Netzwerkansätze, deutlich übertrifft, insbesondere in kleinen Modellräumen, in denen konventionelle Extrapolationen am unzuverlässigsten sind.

Als physikalische Anwendung wird die entwickelte Methode auf eine umfassende *ab initio*-Untersuchung der Grundzustandsenergien und Radien von Bor-Isotopen unter Verwendung zweier moderner Familien chiraler Nukleon-Nukleon- und Drei-Nukleon-Wechselwirkungen angewandt. Insbesondere wird ^{8}B als Kandidat für eine Protonen-Halo-Struktur untersucht, und es werden hochpräzise theoretische Vorhersagen für quadratische Protonenradiusdifferenzen zwischen Bor-Isotopen bereitgestellt, die einen direkten Vergleich mit aktuellen und zukünftigen Isotopen-Shift-Experimenten ermöglichen.

Contents

List of Abbreviations

3N	three nucleon
ANN	artificial neural network
CC	coupled-cluster
CFP	coefficient of fractional parentage
chiral EFT	chiral effective field theory
CI	configuration interaction
CNN	convolutional neural network
CODATA	Committee on Data for Science and Technology
EMN	Entem–Machleidt–Nosyk
FFNN	feed-forward neural network
Gen1	first generation
Gen2	second generation
Gen2+	second generation with sequence selection
HO	harmonic oscillator
IM-NCSM	in-medium no-core shell model
IM-SRG	in-medium similarity renormalization group
IR	infrared
ISU	Iowa State University
IT	importance truncation
IT-NCSM	importance-truncated no-core shell model
Jacobi-NCSM	Jacobi no-core shell model
lattice EFT	lattice effective field theory
LEC	low-energy constant
LENPIC	Low-Energy Nuclear Physics International Collaboration
LO	leading order

MSE	mean-square error
N^{2}LO	next-to-next-to-leading order
N^{3}LO	next-to-next-to-next-to-leading order
N^{4}LO	next-to-next-to-next-to-next-to-leading order
N^{5}LO	next-to-next-to-next-to-next-to-next-to-leading order
NCSM	no-core shell model
NLO	next-to-leading order
NN	nucleon nucleon
NN+3N	nucleon-nucleon and three-nucleon
OTN	observable transcoder network
PDG	Particle Data Group
QCD	quantum chromo dynamics
QMC	quantum Monte Carlo
ReLU	rectified linear unit
rms	root-mean-square
RNN	recurrent neural network
SCGF	self-consistent Green's function
SCS	semi-local coordinate space
SMS	semi-local momentum space
SRG	similarity renormalization group
UV	ultraviolet

Introduction

Understanding the structure of atomic nuclei from the underlying interactions between nucleons is a long-standing and central objective of nuclear physics. Over the past two decades, substantial progress toward this goal has been achieved through the development of *ab initio* methods, which aim to describe nuclear systems directly from realistic nuclear forces without phenomenological adjustments. These approaches have enabled increasingly quantitative predictions for a broad range of observables in light and medium-mass nuclei.

A key theoretical foundation of modern *ab initio* nuclear structure calculations is chiral effective field theory (chiral EFT) [1–6], which provides a systematically improvable description of nuclear interactions based on the underlying theory of quantum chromo dynamics (QCD) [7]. Chiral interactions naturally generate a hierarchy of two and many-body forces and offer a well-defined framework for estimating uncertainties associated with the truncation of the chiral expansion. In particular, Bayesian order-by-order analyses developed within the BUQEYE collaboration enable a quantitative and systematic assessment of these truncation uncertainties [8–10].

Despite these conceptual and methodological advances, a persistent challenge in *ab initio* nuclear structure theory is the slow convergence of many observables with respect to the size of the many-body model space. This problem is exacerbated for chiral interactions, whose strong short-range repulsion and associated high-momentum components couple low- and high-energy states in the Hamiltonian, thereby shifting important interaction contributions to large model spaces and further slowing convergence [11, 12]. To mitigate this issue, the nuclear Hamiltonian is commonly pre-processed using the similarity renormalization group (SRG), a continuous unitary transformation that softens the interaction and approximately pre-diagonalizes the Hamiltonian [13]. Owing to its effectiveness, the SRG has become a standard tool in modern many-body calculations, particularly in conjunction with chiral EFT interactions.

In parallel, the landscape of *ab initio* many-body methods has expanded significantly and is now more diverse than ever [14]. Established configuration

interaction (CI) approaches such as the no-core shell model (NCSM) [12, 15–17] and its importance-truncated extension (IT-NCSM) [17–19] coexist with coupled-cluster (CC) [20–23], quantum Monte Carlo (QMC) [24–26], self-consistent Green's function (SCGF) [27–30], lattice effective field theory (lattice EFT) [31, 32], and in-medium methods such as the in-medium similarity renormalization group (IM-SRG) [33–35], which have also inspired hybrid approaches like the in-medium no-core shell model (IM-NCSM) [36–38]. Together, these methods have enabled *ab initio* calculations of unprecedented precision across a wide range of nuclear masses.

Among these approaches, the NCSM plays a unique role due to its variational character [39], its exact treatment of translational invariance [40, 41], and its ability to deliver systematically improvable results in light nuclei [12, 16]. These features make the NCSM particularly well suited for controlled convergence studies, benchmarking, and the development of new extrapolation strategies.

As with most CI methods, however, the principal limitation of the NCSM is the rapid growth of the model-space dimension with increasing basis size. Even for light nuclei, achieving well-converged results requires extremely large model spaces, effectively restricting fully converged calculations to very light systems with mass number $A \lesssim 4$. For heavier nuclei, theoretical predictions necessarily rely on extrapolation techniques to infer the infinite-model-space limit from calculations performed in truncated spaces. While ground-state energies often converge comparatively rapidly, observables that probe the spatial extent of the nuclear wave function, most notably nuclear radii and electromagnetic moments, exhibit significantly slower convergence. At the same time, precisely because of their strong sensitivity to the detailed structure of the nuclear state, these observables provide particularly powerful and discriminating benchmarks for comparison with experimental data.

Traditional extrapolation methods are often based on empirically motivated functional forms. Although successful in specific regimes, these approaches generally lack statistically robust uncertainty estimates and tend to be less reliable and accurate in small model spaces [42].

Against this backdrop, machine-learning techniques have emerged as powerful tools for regression and pattern recognition [43] across many areas of science [44, 45], including nuclear physics [46–49]. These developments naturally motivate the question of whether machine learning can be exploited to improve the extrapolation of *ab initio* many-body calculations and to provide more robust uncertainty estimates for slowly converging observables.

The central objective of this thesis is to develop and validate a machine-learning–based framework for model-space extrapolation in *ab initio* nuclear structure calculations using artificial neural networks (ANNs). Furthermore, by integrating the network-derived many-body uncertainties with order-by-order truncation uncertainties from chiral effective field theory, the framework enables a coherent uncertainty quantification that is essential for meaningful comparisons with high-precision experimental data.

This thesis is structured as follows. We begin by introducing the theoretical and computational tools required for modern nuclear many-body calculations, including the construction of chiral EFT-based nuclear interactions and their use in NCSM calculations. We then review the fundamentals of ANNs, with a focus on their application to model-space extrapolation. Subsequently, we present the design, training, and application of our first generation (Gen1) networks for extrapolating ground-state energies in the NCSM. This is followed by the introduction of the improved second generation (Gen2) networks, including a variant capable of predicting converged radii, as well as the development of an enhanced evaluation and uncertainty quantification scheme. We then benchmark our neural-network–based approach against widely used extrapolation methods and a competing ANN-based strategy, considering both ground-state energies and proton radii. Next, we introduce two post-processing techniques: one that enables high-precision predictions of difference observables, such as excitation energies and proton–neutron radius differences, and another that combines the network-derived many-body uncertainties with chiral interaction uncertainties. Finally, the complete framework is applied to the boron isotopes, yielding high-precision *ab initio* predictions for energies and radii, investigating the possible emergence of a proton halo in ^{8}B, and providing precise theoretical estimates for the squared proton radius difference between ^{10}B and ^{11}B for comparison with current and future isotope-shift experiments.

1

Nuclear Structure Calculations

Theoretical calculations of nuclear properties can be carried out using a variety of different methods. In this work, we focus on the computation of nuclear observables using the NCSM, as the main goal of this work is to achieve high fidelity and precision *ab initio* results for ground-state radii of light nuclei as well as other observables. For the purposes of this work, *ab initio* mean that any truncation and approximation within the calculation is controllable and systematically improvable, and that the resulting uncertainties are equally systematically quantifiable. To facilitate this *ab initio* approach throughout the whole calculation, we generally start with a nuclear interaction derived from chiral EFT. To improve the convergence of the subsequent many-body calculation, the interaction first undergoes a unitary transformation using the SRG formalism. Then, the many-body calculation in the NCSM can be carried out to obtain observables such as the ground-state energy and the root-mean-square (rms) radius.

1.1 The No-Core Shell Model

The NCSM is the bread and butter many-body method used throughout this work. We will use it in its standard form to calculate most results for light nuclei, while using extensions or combinations with other methods for all other results. Few-body systems can be calculated very efficiently with the Jacobi no-core shell model (Jacobi-NCSM) [50–52], and the reach of the regular NCSM [12, 15–17] can be extended further with the importance truncation (IT) creating the importance-truncated no-core shell model (IT-NCSM) [17–19]. Another

approach enabling calculations of heavier nuclei is the combination of the IM-SRG with the NCSM, called the IM-NCSM [36–38]. This extension to the NCSM can extend the reach of the NCSM into the medium-mass regime, which is well beyond the scope of this work. Therefore, we will concentrate mostly on the NCSM as well as the Jacobi-NCSM.

1.1.1 Standard NCSM

The NCSM, at its core, is a configuration interaction many-body method designed to solve the stationary Schrödinger equation

$$\hat{H}\,|\psi_n\rangle = E_n\,|\psi_n\rangle\,, \tag{1.1}$$

with the Hamiltonian $\hat{H}$ and the energy E_n associated with the nth eigenstate $|\psi_n\rangle$. Contrary to the shell model, from which it derives its name, it aims to solve the Schrödinger equation without conceptual approximations and treats all nucleons as active degrees of freedom. While this approach is computationally costly, it offers the ability to calculate all observables consistently and precisely with controllable and systematically improvable truncation effects. As a basis expansion method, it translates the operator equation Eq. (1.1) into a matrix eigenvalue problem by inserting a complete, discrete many-body basis set $|\phi_i\rangle$. The standard choice for such a basis set is the many-body harmonic oscillator (HO) basis, constructed from Slater determinants of single-particle HO basis states. This choice of basis has some specific advantages we will highlight a bit later. Expanding Eq. (1.1) in such a basis, we obtain the matrix-eigenvector form of the time-independent Schrödinger equation

$$\sum_j \langle\phi_i\,|\hat{H}\,|\,\phi_j\rangle\langle\phi_j\,|\,\psi_n\rangle = E_n\langle\phi_i\,|\,\psi_n\rangle\,. \tag{1.2}$$

If we can diagonalize the matrix $\langle\phi_i\,|\hat{H}\,|\,\phi_j\rangle$ and find the eigenvalues E_n and corresponding eigenvectors $\langle\phi_j\,|\,\psi_n\rangle = C_j^{(n)}$, we can calculate the full solution of the eigenstates $|\psi_n\rangle = \sum_j C_j^{(n)}\,|\phi_j\rangle$ from Eq. (1.1). However, before we can numerically diagonalize the matrix, we first need to apply a suitable truncation scheme to limit the otherwise infinite-dimensional many-body matrix. For that, the standard choice is the so-called $N_{\max}$ truncation, which limits the total excitation energy above the lowest-energy many-body HO Slater determinant to $N_{\max}\hbar\Omega$, where Ω is the oscillator frequency of the single-particle HO basis and $N_{\max}$ is

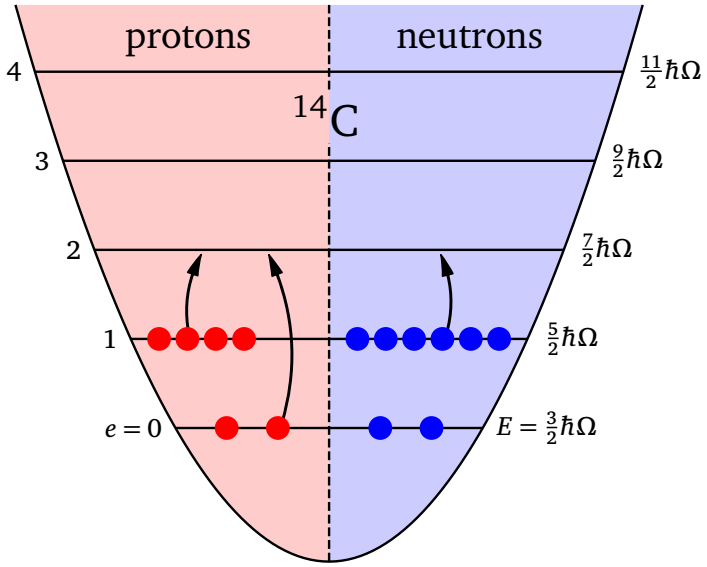


Figure 1.1: Illustration of the lowest-energy Slater determinant of ^{14}C in the harmonic oscillator potential. The black arrows indicate a possible excitation of two protons and one neutron with a total excitation energy of $4\hbar\Omega$, which is included in $N_{\max} \geq 4$ model spaces.

an integer ≥ 0. This is obvious if we recall the energy of the three-dimensional harmonic oscillator in relation to the principal n and the orbital momentum quantum number l is

$$E_{\mathrm{HO}} = \hbar\Omega\left(2n + l + \frac{3}{2}\right) = \hbar\Omega\left(e + \frac{3}{2}\right), \tag{1.3}$$

with e being the single-particle energy. Thus, single-particle excitations always result in an energy difference that is an integer multiple of $\hbar\Omega$. Figure 1.1 illustrates the $N_{\max}$ truncation for ^{14}C, showing the lowest energy Slater determinant for this nucleus. The excited state containing proton and neutron excitations, indicated by arrows, is contained in all model spaces with $N_{\max} \geq 4$, since the sum of all energy deltas of the excitations is $4\hbar\Omega$. It is important to note that $N_{\max}$ only increases in steps of two, since a change in energy of an odd multiple of $\hbar\Omega$ in the harmonic oscillator is only possible by increasing the orbital momentum quantum number l by an odd amount, thereby flipping the parity of the whole state. States of unnatural parity, however, cannot be connected to states of natural parity by the Hamiltonian. We can therefore exclude these states when calculating eigenstates for one particular parity, reducing the model space size significantly. A common alternative parametrization of the HO basis is via the

oscillator length

$$a_{\mathrm{HO}} = \sqrt{\frac{\hbar}{m_N \Omega}}, \tag{1.4}$$

where m_N is the nucleon mass. This is effectively just a rescaling of the $\hbar\Omega$ basis parameter, and is often useful when targeting observables sensitive to the spacial extent of the wave function, for example the radius, as it simplifies the usage of a spatially equidistant parameter grid. Another important aspect to highlight is that since the many-body HO states form a complete basis set for $N_{\mathrm{max}} \to \infty$, the exact solution is independent of $\hbar\Omega$. This is, however, not true for finite N_{max}.

After truncation, the now finite dimensional matrix can be diagonalized with standard methods like the Lanczos method [53]. The Hamiltonian only depends on the relative coordinates of the nucleons. However, since we are using an absolute coordinate HO basis which is therefore not translationally invariant, spurious center-of-mass contaminated states emerge during the diagonalization. These states are degenerate in their eigenenergies, but differ in their center-of-mass component. One very important advantage of choosing both the HO basis and the N_{max} truncation is that this specific combination allows the intrinsic and the center of mass components to factorize exactly [40]

$$|\Psi_n\rangle = |\Psi_n\rangle_{\mathrm{cm}} \otimes |\Psi_n\rangle_{\mathrm{int}} \,. \tag{1.5}$$

As a consequence, one can easily eliminate the center-of-mass contamination by adding the so-called Lawson term [41] to the intrinsic Hamiltonian

$$\hat{H}_{\mathrm{cm}} = \hat{T}_{\mathrm{cm}} + \frac{mA\Omega^2}{2}\hat{\vec{R}}^2 - \frac{3}{2}\hbar\Omega, \tag{1.6}$$

with the center-of-mass coordinate $\hat{\vec{R}}$, and solve the nuclear eigenvalue problem for

$$\hat{H} = \hat{H}_{\mathrm{int}} + \beta \hat{H}_{\mathrm{cm}} \,. \tag{1.7}$$

The purpose of this added term is to set the ground-state center-of-mass energy to zero while excitations are shifted outside the relevant part of the spectrum to higher energies for $\beta > 0$.

The main reason why we chose to construct our many-body basis in the lab frame in the first place is that the antisymmetrization of the basis states is much

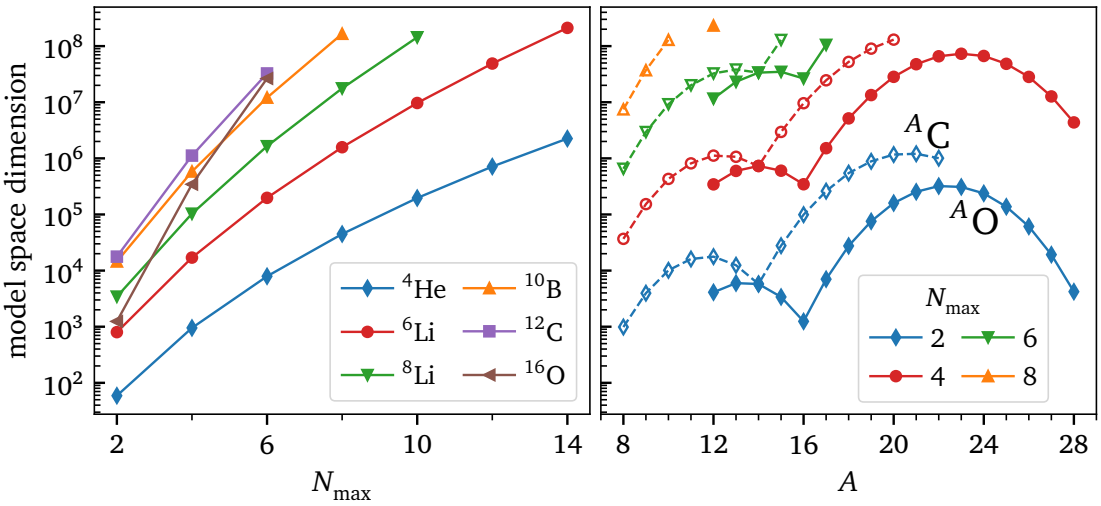


Figure 1.2: Illustration of the rapid model-space dimension growth with increasing N_{max} for a selection of nuclei (left panel), and the behavior with growing nucleon number A (right panel) for the carbon (open symbols) and oxygen isotopic chain (filled symbols). While the increase as a function of N_{max} is almost exponential, the behavior, while generally increasing, is more complicated for the nucleon number due to shell closures.

simpler in the absolute HO basis. However, this choice does also incur several disadvantages which make it less efficient. The main one is that the total angular momentum J is no longer a good quantum number. While its projection M can be used to reduce the number of basis states one targets, this still leads to an inflated number of basis states. The model space dimensions for a selection of nuclei as a function of N_{max} is shown in the left-hand panel of Fig. 1.2. Model space dimensions rapidly grow at an almost exponential pace. The same is generally also true for increasing nucleon number A, though this dependence is more complicated in the vicinity of shell closures as is demonstrated in the right-hand panel of Fig. 1.2. The doubly magic ^{16}O has a significantly smaller model-space dimension than some of its lighter isotopes, especially at smaller N_{max}, as there are less unique basis states available at the respective truncation levels. This effect lessens for larger N_{max}, where the increased number of possible basis states starts do dominate the model-space dimension through sheer combinatorics due to more available nucleons. For this reason, even very light nuclei can usually not reach full convergence, as the required model space size grows too large with increasing N_{max}. Therefore, we often employ the NCSM with a relative basis to reach full convergence in very light systems, called the Jacobi-NCSM.

1.1.2 Jacobi-NCSM

The Jacobi-NCSM is a variation of the standard NCSM that uses a relative HO basis, also known as a Jacobi basis as its many-body basis [50–52]. In coordinate space, this basis is constructed in the following way. The center-of-mass coordinate is

$$\vec{\xi}_0 = \sqrt{\frac{1}{A}} \sum_{i=1}^{A} \vec{r}_i \tag{1.8}$$

and the coordinate of the $(n+1)$-th nucleon is defined as

$$\vec{\xi}_n = \sqrt{\frac{n}{n+1}} \left(\frac{1}{n} \sum_{i=1}^{n} \vec{r}_i - \vec{r}_{n+1} \right), \tag{1.9}$$

where $\vec{r}_{A+1} = 0$. Here, the Jacobi coordinate of the $(n+1)$-th nucleon is defined relative to the coordinates of the first n nucleons. Analogously, a Jacobi momentum can be defined as

$$\vec{\pi}_0 = \sqrt{\frac{1}{A}} \sum_{i=1}^{A} \vec{p}_i \tag{1.10}$$

$$\vec{\pi}_n = \sqrt{\frac{n}{n+1}} \left(\frac{1}{n} \sum_{i=1}^{n} \vec{p}_i - \vec{p}_{n+1} \right), \tag{1.11}$$

with $\vec{p}_{A+1} = 0$.

In these coordinates, the defining equation for the harmonic oscillator, in particular the Hamiltonian, reads

$$\hat{H}_{\mathrm{HO}} = \sum_{i=0}^{A-1} \left(\frac{\hat{\vec{\pi}}_i^{\,2}}{2m} + \frac{1}{2} m \Omega \hat{\vec{\xi}}_i^{\,2} \right), \tag{1.12}$$

i.e. the equation has the exact same form as with the standard coordinates $\hat{\vec{r}}_i$ and $\hat{\vec{p}}_i$. As such, one obtains the same eigenfunctions and therefore also the same generating functions for the Jacobi-based harmonic oscillator basis. This allows both for a clean separation of the center-of-mass portion of the basis, and the targeting of specific J states, as it remains a good quantum number. Both lead to significant savings in model space dimension, which enable larger N_{max} compared to standard NCSM calculations, as illustrated in Fig. 1.3 for ^{4}He. Here, the model space dimension grows significantly slower for increasing N_{max} in the Jacobi-

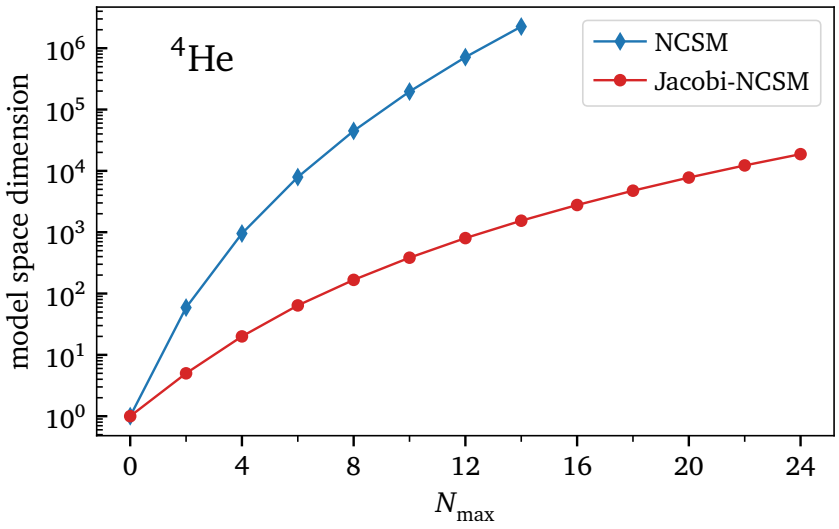


Figure 1.3: Illustration of the model-space dimensions for ^{4}He as a function of $N_{\max}$ for the standard NCSM and the Jacobi-NCSM. This demonstrated the massively increased efficiency of the Jacobi-NCSM for very light nuclei.

NCSM compared with the standard NCSM, with model space dimensions of 714 288 at $N_{\max} = 12$ in the latter, versus 801 at $N_{\max} = 12$ and still only 18 724 at $N_{\max} = 24$ for the Jacobi-NCSM. This underlines the significant efficiency advantage of the Jacobi-NCSM, which enables us to compute fully-converged NCSM results for few-body systems.

One wrinkle, however, is that the antisymmetrization of the many-body states, as is required since nucleons are fermions, is significantly more involved. In the single-particle formulation, the antisymmetrization is simply taken care of by using Slater determinants of single-particle HO states, which is no longer possible in the Jacobi basis. This is generally a limiting factor for the Jacobi-NCSM, with some codes able to achieve antisymmetrization up to $A \leq 8$ [50–52, 54]. Our Jacobi-NCSM code, however, is limited to $A \leq 4$, as we use pre-computed coefficient of fractional parentages (CFPs), which are the main ingredient of the antisymmetrization. Higher particle numbers quickly run into size constraints and make pre-computing infeasible.

Another disadvantage is that the basis does not track the isospin of individual nucleons. This makes a direct computation of observables that are sensitive to the isospin, e.g. the point-proton rms radius, impossible. For the purposes of this work, this limitation is fortunately largely irrelevant, since all computations of these observables can and will be carried out using the standard NCSM, as none of them need the specific advantages the Jacobi-NCSM offers.

1.2 The Many-Body Hamiltonian

The basis of all nuclear many-body calculations is the Hamiltonian, which encodes the interactions between protons and neutrons. Nuclear forces therefore constitute the essential input to theoretical descriptions of nuclei, incorporating both nucleon nucleon (NN) and three nucleon (3N) interactions in a form suitable for modern many-body methods. Forces beyond the two-body level arise because protons and neutron are baryons, i.e. composite particle consisting of three valence quarks. Unfortunately, a quantitative description of the strong interactions among nucleons at low energies cannot be obtained directly from the underlying quantum field theory of QCD, which governs the behavior of these particles. For a long time, this limitation restricted realistic nuclear structure calculations to phenomenological NN interactions [55–58], whose parameters were adjusted directly to fit experimental data. Modern nuclear structure theory has largely overcome these limitations through the framework of chiral EFT, which exploits the symmetries of QCD, most notably its approximate chiral symmetry, to construct nuclear interactions in a systematic manner [2, 5, 6].

1.2.1 Chiral Interactions

Constructing realistic nuclear interactions remains a central challenge in nuclear structure physics due to the nonperturbative nature of QCD at low energies. Chiral effective field theory addresses this challenge by employing nucleons and pions as explicit degrees of freedom and exploiting the approximate chiral symmetry of QCD [1–6].

Using the power-counting scheme developed by Weinberg [7, 59, 60], all interaction terms consistent with the symmetries of QCD can be systematically generated and ordered according to an expansion in powers of Q^{ν}. Here, the expansion parameter Q represents the typical momentum or the pion mass [7] over the breakdown scale Λ_{χ} of chiral EFT [61]. This procedure leads to the hierarchy of interaction diagrams shown in Fig. 1.4, where individual Feynman diagrams are organized by both their chiral order ν and their particle rank. The leading order (LO) contributions correspond to $\nu = 0$, while terms at $\nu = 1$ are forbidden by parity and time-reversal symmetry [62]. As a result, the next-to-leading order (NLO) contributions appear at $\nu = 2$. Three-nucleon interactions first arise at

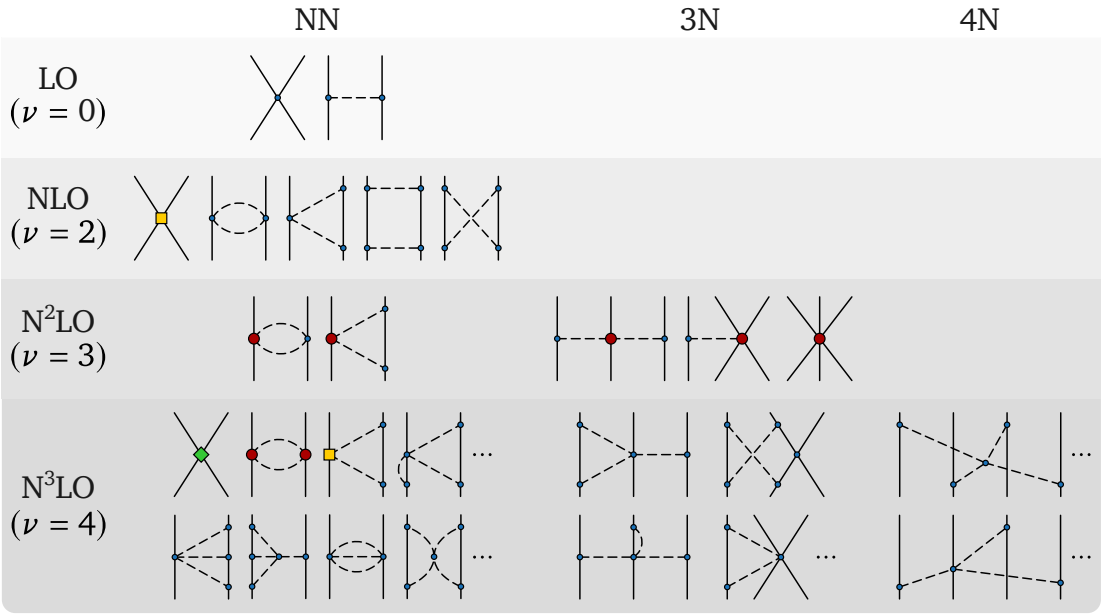


Figure 1.4: Illustration of the hierarchical structure of nuclear interaction terms in chiral effective field theory (chiral EFT), represented by Feynman diagrams. Solid lines denote nucleons, while dashed lines indicate pion exchanges. Vertex dimensions 0, 1, 2, and 4 are encoded by blue dots, red circles, yellow squares, and green diamonds, respectively.

next-to-next-to-leading order (N^2LO) ($\nu = 3$), with higher orders introducing additional interaction topologies and higher-body forces. Although four-nucleon forces formally appear at next-to-next-to-next-to-leading order (N^3LO), they are generally neglected in practical applications, as their inclusion is computationally extremely expensive and their impact on nuclear observables is found to be small [63].

The terms in the chiral expansion are accompanied by a set of low-energy constants (LECs), which determine the strength of different parts of the interaction. At present, these constants are largely determined by fits to experimental data. With continued progress in lattice QCD calculations, future nuclear interactions may be constrained directly from first-principles QCD calculations, reducing or eliminating the reliance on experimental input.

Even within chiral EFT, naïve calculations of nuclear forces lead to ultraviolet divergences. These are regulated by introducing cutoff functions that suppress high-momentum components above a regulator scale Λ. The cutoff must be chosen large compared to the typical momentum scale of the system, but below the breakdown scale Λ_χ of the chiral expansion. In principle, physical observables should be independent of the specific choice of regulator and cutoff, as regulator

Interaction family	Cutoff	Abbreviation
Entem–Machleidt–Nosyk	450 MeV	EMN[450]
	500 MeV	EMN[500]
	550 MeV	EMN[550]
Semi-local momentum-space regularized	450 MeV	SMS[450]
	500 MeV	SMS[500]

Table 1.1: List of interaction shorthands used throughout this thesis.

dependence can be absorbed into a redefinition of the LECs. In practice, however, residual regulator artifacts do affect many-body calculations [64], and certain regulator choices offer numerical or conceptual advantages.

A central strength of chiral EFT is that its systematic expansion provides a natural framework for estimating theoretical uncertainties. One important class of uncertainties arises from the choice of regulator and cutoff, which can be explored by comparing results obtained with interactions employing different regularization schemes and cutoff values. Arguably the dominant theoretical uncertainty, however, is typically associated with truncation of the chiral expansion, i.e. the impact of the neglected higher-order terms. In this work, we estimate these truncation uncertainties using the BUQEYE point-wise Bayesian model [8], which estimates uncertainties from the order-by-order convergence of each observable. While alternative approaches to chiral uncertainty quantification exist [61], the BUQEYE framework provides a consistent and statistically well-defined means of combining order-by-order information into quantitative uncertainty estimates.

In the following sections, we introduce the two families of chiral EFT interactions employed in this work and discuss their specific regularization schemes and cutoff choices.

1.2.1.1 Entem–Machleidt–Nosyk Interactions. The Entem–Machleidt–Nosyk (EMN) interaction family [65] represents a modern extension of the original Entem–Machleidt NN interaction [66, 67]. These interactions are available at chiral orders ranging from NLO to next-to-next-to-next-to-next-to-leading order (N^4LO) and employ a non-local regulator with cutoffs $\Lambda = 450$, 500, and 550 MeV. They are complemented by consistent three-nucleon interactions available up to N^3LO, as described in Refs. [68, 69]. In this work, we exclusively use the many-body optimized version of the EMN interaction, in which the three-nucleon low-energy constant c_D is fitted to the ground-state energy of ^{16}O. Up to N^3LO,

the two- and three-nucleon forces are constructed consistently at the same chiral order. At the highest order, however, the N^{4}LO NN interaction is combined with the N^{3}LO 3N force, as no N^{4}LO three-nucleon interaction is currently available. As this leads to shifts in some of the LECs [69], we denote this combined two and three-body interaction as N^{4}LO′.

1.2.1.2 Semi-Local Momentum-Space Regularized Interactions. The semi-local momentum space (SMS) regularized interactions are developed by the Low-Energy Nuclear Physics International Collaboration (LENPIC). While there is also a complementary semi-local coordinate space (SCS) regularized interaction [61, 70], only the SMS interactions [71] are used in this work.

Unlike fully local or fully non-local schemes, the SMS interactions employ a mixed regularization strategy. The long-range pion-exchange contributions are regularized using local regulators in order to minimize regulator artifacts [61, 64], while the short-range part is regularized non-locally. This approach is applied consistently to both NN and 3N interactions. Constructing a fully consistent three-body force within this framework is technically challenging, and as a result, the NN interaction is available up to N^{4}LO and beyond, whereas the 3N force is currently only available at N^{2}LO [72].

As part of the LENPIC effort, these interactions have been applied to nuclei extending into the medium-mass region [73, 74]. We contributed to these efforts by calculating and providing SRG-transformed two and thee-body matrix elements and performing many-body calculations.

1.2.2 Free-Space Similarity Renormalization Group

While chiral interactions provide a systematic and theoretically well-founded description of nuclear forces, they often exhibit relatively hard momentum-space behavior. As a consequence, they couple low and high-momentum states strongly, leading to large off-diagonal matrix elements in the Hamiltonian [11, 12]. In the context of the NCSM, this significantly slows model-space convergence, as important correlations only emerge at large values of $N_{\max}$.

To mitigate these issues, chiral interactions are commonly preprocessed using the free-space SRG [75–78]. The SRG applies a continuous unitary transformation to the Hamiltonian

$$\hat{H}(s) = \hat{U}^\dagger(s)\,\hat{H}\,\hat{U}(s)\,, \tag{1.13}$$

where $\hat{H} = \hat{H}(0)$ denotes the initial Hamiltonian and s is the flow parameter controlling the evolution. The goal of the transformation is to pre-diagonalize the Hamiltonian, i.e. shift large off-diagonal contributions towards a band-diagonal structure [11, 12]. Differentiating Eq. (1.13) with respect to s yields the SRG flow equation

$$\frac{\mathrm{d}\hat{H}(s)}{\mathrm{d}s} = \left[\hat{\eta}(s), \hat{H}(s)\right], \tag{1.14}$$

which defines the anti-Hermitian generator $\hat{\eta}(s)$ of the transformation. A common choice for the generator is based on the intrinsic kinetic energy operator $\hat{T}_{\mathrm{int}} = \hat{T} - \hat{T}_{\mathrm{cm}}$ and reads

$$\hat{\eta}(s) = (2\mu)^2 \left[\hat{T}_{\mathrm{int}}, \hat{H}(s)\right], \tag{1.15}$$

with the reduced nucleon mass $\mu = m_N/2$, which fixes the units of the flow parameter to fm^4.

In principle, the SRG transformation is exactly unitary and therefore preserves the eigenvalues of the Hamiltonian. In practice, however, the commutator structure of Eq. (1.14) induces many-body operators of increasing particle rank during the evolution. Since the explicit treatment of induced four- and higher-body terms is computationally prohibitive, the evolution is typically truncated at the three-body level. As a result, the transformation is no longer strictly unitary, and a residual dependence on the flow parameter remains [79]. This effect is usually small for moderate values of s and can be quantified by repeating calculations for several flow parameters. Typical values for practical calculations range from about $s = 0.02$ to $0.16\,\mathrm{fm}^4$, though larger flow parameters are possible at the expense of larger effects of the neglected terms.

The three-body SRG evolution is computationally demanding, and several techniques are employed to reduce the associated cost. Three-body matrix elements are stored in a relative harmonic-oscillator basis and grouped according to conserved quantum numbers such as isospin T_{12}, total angular momentum J_{12}, and parity Π_{12}. This block structure allows the SRG evolution to be performed independently in each channel. Moreover, the SRG evolution is carried out at a fixed oscillator frequency, typically $\hbar\Omega_{\mathrm{SRG}} = 36\,\mathrm{MeV}$. Matrix elements at other oscillator frequencies are then obtained via frequency conversion [11]. See Refs. [11, 69, 80] for more information and technical details regarding the three-body SRG evolution.

1.3 Observable Calculation

The calculation of observables within the NCSM framework is rather straight forward. Since one obtains the full eigenstates $|\psi_n\rangle$ from the diagonalization, any observable O can simply be calculated via the expectation value

$$O_{n,m} = \langle \psi_m | \hat{O} | \psi_n \rangle , \tag{1.16}$$

with $m \neq n$ for transition observables. This requires the operator $\hat{O}$ to be represented in the same HO basis as the many-body calculation. If the Hamiltonian has been transformed using the free-space SRG, consistency demands that all operators be evolved accordingly. This is achieved by calculation the unitary transformation matrix U of the free-space SRG, and simply apply that transformation to the observable

$$\hat{O}(s) = \hat{U}^\dagger(s)\, \hat{O}\, \hat{U}(s) . \tag{1.17}$$

The effects of this transformation are usually very small. Still, we apply this transformation for all high-precision radius calculations, though only up to the two-body level, to improve consistency.

Among nuclear observables, radii play a particularly important role as they probe the spatial structure of the nuclear wave function, especially its asymptotic behavior. However, their calculation within the HO basis of the NCSM poses a well-known challenge. While the physical wave function exhibits an exponential decay $\propto \exp(-r)$ at large distances, HO basis functions fall off as $\propto \exp(-r^2)$. For finite $N_{\max}$, this mismatch leads to systematic distortions: small oscillator lengths a_{HO} truncate the long-range tail, resulting in underestimated radii, whereas large a_{HO} values artificially extend the wave function, leading to overestimated radii. Consequently, no finite model space can fully capture the correct asymptotic behavior, and observables sensitive to the tail, such as radii and electromagnetic moments, converge significantly slower than, for example, ground-state energies.

In this work, we focus on mass, point-proton, and point-neutron rms radii. Comparisons with experimental data are typically performed using the charge radius, which incorporates additional corrections discussed below.

1.3.1 Point-Proton, Neutron, and Mass Radii

Typically, many body calculations mainly look at the point-proton and the mass rms radius. The mass rms radius as a two-body operator defined as

$$R = \sqrt{\left\langle \psi_n \left| \frac{1}{A} \sum_{i=1}^{A} (\hat{\vec{r}}_i - \hat{\vec{R}})^2 \right| \psi_n \right\rangle}, \tag{1.18}$$

where $|\psi_n\rangle$ is the state for which we want to compute the radius, in our case usually the eigenstate obtained from an NCSM calculation, and

$$\hat{\vec{R}} = \frac{1}{A} \sum_{i=1}^{A} \hat{\vec{r}}_i \tag{1.19}$$

is the center-of-mass operator. To calculate the point-proton rms radius, we first need to introduce the proton projection operator

$$\hat{\Pi}_{p,i} = \frac{1}{2} + \hat{t}_{3,i}\,, \tag{1.20}$$

that evaluates to 1 if the ith nucleon is a proton and 0 otherwise, making use of the isospin operator. Together with Eq. (1.18), we can construct the point-proton rms radius as

$$R_p = \sqrt{\left\langle \psi_n \left| \frac{1}{Z} \sum_{i=1}^{A} \hat{\Pi}_{p,i} (\hat{\vec{r}}_i - \hat{\vec{R}})^2 \right| \psi_n \right\rangle}. \tag{1.21}$$

For some calculations we will also compute the point-neutron rms radius R_n which can be defined analogously to the point-proton rms radius in Eq. (1.21) by swapping the proton projection operator with the neutron projection operator $\hat{\Pi}_{n,i} = 1 - \hat{\Pi}_{p,i}$ and adjusting the normalization accordingly:

$$R_n = \sqrt{\left\langle \psi_n \left| \frac{1}{N} \sum_{i=1}^{A} \hat{\Pi}_{n,i} (\hat{\vec{r}}_i - \hat{\vec{R}})^2 \right| \psi_n \right\rangle}. \tag{1.22}$$

All radius operators used throughout this thesis assume exact isospin symmetry, i.e. identical masses for protons and neutrons.

1.3.2 Charge Radii

The charge radius of a given nucleus is, opposed to the point-proton rms radius, directly observable and therefore targeted by many experiments. Its calculation is therefore very important to compare results with experimental data. We can calculate the charge radius from the point-proton rms radius by including corrections for the fact that protons and neutrons are not point-like particles but rather composite particles with a substructure and internal charge distribution. The charge radius and the point-proton rms radius are connected via

$$R_{\mathrm{ch}}^2 = r_p^2 + R_p^2 + \frac{3}{4m_p^2} + \frac{N}{Z} r_n^2 + \mathcal{O}(\mathrm{N^3LO})\,, \tag{1.23}$$

where m_p and r_p are the proton mass and charge radius, and N and Z are the number of neutrons and protons in the nucleus, respectively. While the neutron has no net charge, it does have a charge distribution which leads to a nonzero squared neutron charge radius r_n^2. We neglect higher-order contributions, e.g. relativistic spin-orbit corrections or charge-density exchange terms. However, even without these terms, a conversion from R_{ch} to R_p is, at least at the current time, not straight forward, as there is still some tension regarding the experimental value of the proton charge radius r_p [81–88]. Where necessary, the above formula is used to convert R_{ch} to R_p or vice versa using the Committee on Data for Science and Technology (CODATA) 2022 recommended value for the proton charge radius $r_p = 0.840\,75(64)\,\mathrm{fm}$ [89], combined with the current Particle Data Group (PDG) value for the neutron charge radius squared $r_n^2 = -0.1155(17)\,\mathrm{fm}^2$ [90] and the Darwin-Foldy correction $3/(4m_p^2) = 0.033\,17\,\mathrm{fm}^2$ [91, 92].

In some cases we can circumvent the conversion nearly completely. When comparing charge radii within an isotopic chain, we can simply calculate and compare the squared charge radius difference. In this case, the proton charge radius terms, along with the proton mass terms, cancel out, and the neutron charge radius squared is suppressed by $\Delta N/Z$. This is especially convenient since this difference itself is also accessible experimentally in isotope-shift experiments and can be measured to very high precision [93–96].

1.4 Convergence

Dealing with the convergence behavior of NCSM calculations is the main focus of this thesis. It is therefore paramount to obtain a thorough understanding of the convergence behaviors exibited by different observables, their differences and similarities, to be able to spot patterns we can use to estimate fully-converged results from calculations with incomplete convergence. The two main drivers limiting convergence are the nucleon number A and the truncation parameter $N_{\max}$. Increasing either leads to a lot more possible basis states through combinatorics, increasing the model-space dimension significantly. This directly correlates with a larger memory footprint, increased compute time to construct the basis set, and significantly more computational effort needed to diagonalize the resulting, though sparse, Hamilton matrix. As shown in Section 1.1.1, model spaces grow a lot larger for nuclei in the *sd*-shell beyond the reach of especially efficient methods like the Jacobi-NCSM. Simply due to both memory and computational limitations, we are therefore often limited to a few steps in $N_{\max}$, which is generally not enough to reach full convergence.

There is another parameter in both NCSM formulations that influences the convergence: the harmonic oscillator frequency Ω, usually written in terms of an energy as $\hbar\Omega$, or, equivalently, the harmonic oscillator length a_{HO}. This parameter determines the energy and spacial extend of the HO wave functions in the many-body basis, and has a profound impact on the convergence. Figure 1.5 illustrates this for the ground-state energy of ^{4}He. We will discuss the specifics of the convergence of energies in the next section, however, universal for all observable is the spread of the convergence pattern in relation to the harmonic oscillator parameter. We can observe that the difference between different $\hbar\Omega$ values decreases for increasing $N_{\max}$. Furthermore, as discussed previously, all curves have to converge to the same value for $N_{\max} \to \infty$, irrespective of the HO basis parameter. There is usually a HO parameter for which convergence is fastest, as is the case for $\hbar\Omega = 32\,\mathrm{MeV}$ in Fig. 1.5. One of the strategies to optimize convergence is to try to find and target such an oscillator parameter and a few values around it.

As already hinted at, another important aspect to understand is how different observables have vastly different convergence patterns and constraints. We will therefore take a look at the convergence behavior of the two most important observable in this work, energies and radii.

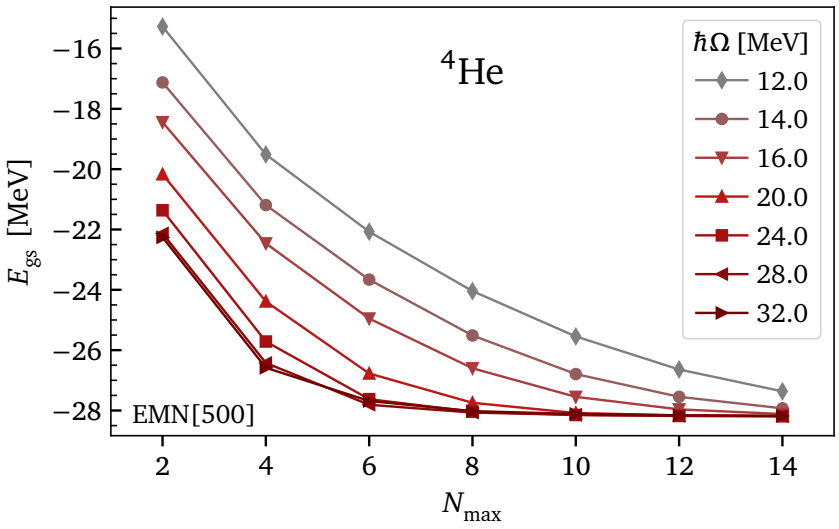


Figure 1.5: Typical ground-state energy convergence in the NCSM. The HO basis parameter dependence decreases with increasing model-space size determined by the truncation parameter N_{max}.

1.4.1 Energy Convergence

Energies display a very regular convergence pattern, as the energy is constrained via the Hylleraas–Undheim theorem [39], since the NCSM is a variational many-body method. This means that energies always decrease monotonically as the model space grows, i.e. for larger N_{max}. As a consequence, the lowest calculated energy for a specific N_{max} is always an upper bound for the fully converges value. Obviously, this means that the lowest calculated energy is also an upper bound for the ground-state energy.

Figure 1.6 shows such converging sequences for the ground-state energy of ^{4}He as a function of the model-space truncation parameter N_{max} in the left-hand panel. As discussed previously, all sequences converge monotonically towards the ground-state energy, and the spread between the different sequences corresponding to different HO basis parameters gets smaller and eventually almost vanishes for large N_{max}. The right panel of Fig. 1.6 shows the 1^+ ground and the 3^+ and 2^+ excited-state energies of ^{6}Li and demonstrates that, as expected, the monotonicity of the convergence also holds for exited-state energies, which equally follow a very structured and constrained convergence pattern just like the ground-state. This exceptionally regular and stable convergence behavior makes it one of the easiest non-trivial observables to extrapolate. It is important to highlight that this pattern is largely independent of the interaction, the nucleus,

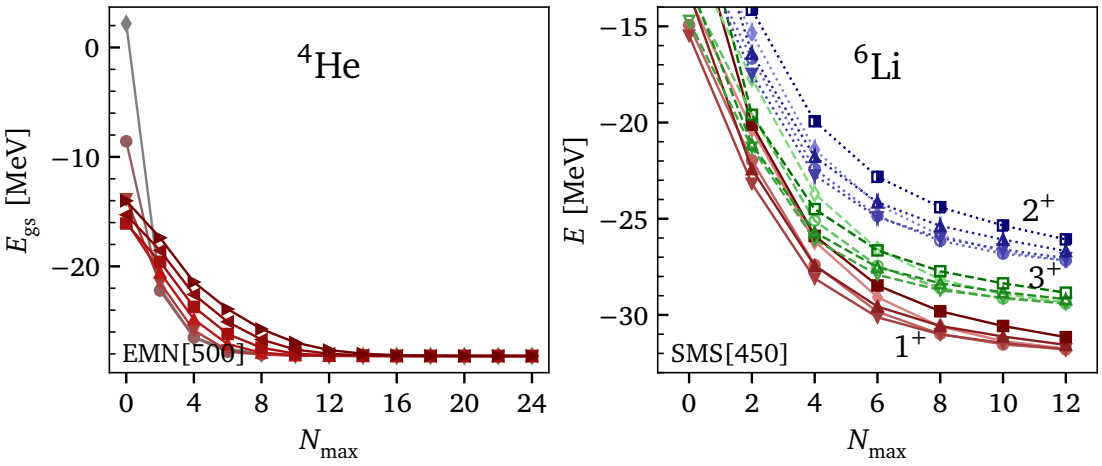


Figure 1.6: Illustration of energy convergence in the NCSM. The left panel shows the convergence of the ground-state energy of ^{4}He with an equidistant a_{HO} grid in the Jacobi-NCSM, while the right panel shows the convergence behavior of the 1^+ ground-state and the 3^+ and 2^+ state energies of ^{6}Li in the NCSM for an $\hbar\Omega$ grid.

or whether one chooses an $\hbar\Omega$ or a_{HO} grid. All these parameters are different between the two panels in Fig. 1.6, i.e. the left panel is calculated with an a_{HO} grid and the EMN[450] interaction SRG evolved to $s = 0.04\,\mathrm{fm}^4$, while the right panel uses an $\hbar\Omega$ grid and the SMS[500] interaction evolved to $s = 0.08\,\mathrm{fm}^4$, yet the resulting convergence patterns appear extremely similar.

Of course, quantitatively these parameters as well as the nucleus and the window of the HO parameter grid have an enormous impact on the convergence speed and absolute numbers and are therefore important aspects to consider for any extrapolation scheme. One thing we can notice is that the monotonously falling energy curves look roughly like an exponential decay, a fact that is often used to estimate and extrapolate the fully converged value as we will in a later section.

1.4.2 Radius Convergence

Contrary to the very well-constrained convergence behavior of energies, radii are a lot more varied. There is no equivalent to the variational principle for radii or a similarly strong constraint, thus, radii do not converge monotonously.

However, radii behave a lot more predictably in dependence of the HO basis parameter than energies, at least in very small model spaces. In fact, for $N_{\mathrm{max}} = 0$,

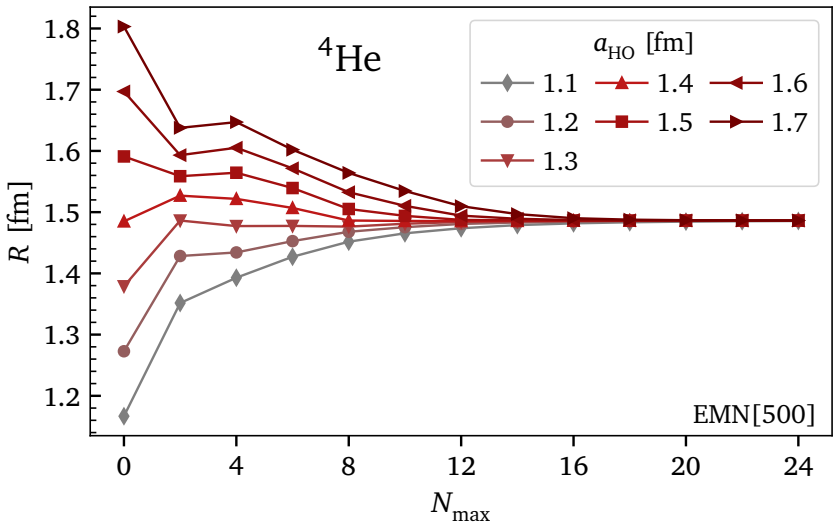


Figure 1.7: Convergence behavior of the radius as a function of N_{max} in ^{4}He for a few different a_{HO} values. Contrary to energies, sequences can converge from above as well as from below, depending on the HO basis parameter, and are generally not monotonous.

the calculated radius is directly proportional to the HO length scale a_{HO}

$$R \propto a_{\mathrm{HO}}\,, \tag{1.24}$$

as the basis simply cannot produce a wave function of a different length scale, which directly affects the radius. This equidistant spread at $N_{\mathrm{max}} = 0$ when using an a_{HO} grid is demonstrated in Fig. 1.7, which shows the rms radius converge for ^{4}He. One simple but very important consequence of this fact is that we can control whether the convergence has a generally increasing or decreasing trend, depending on the chosen HO basis parameter. This will become important for choosing suitable a_{HO} parameter windows for our calculations to improve the performance of radius extrapolations later on.

There is also often a rather flat converging sequence for an optimal a_{HO}, though it is not always guaranteed to stay flat until convergence is reached. In some nuclei and interaction combinations, there is even a slightly oscillatory behavior visible before convergence is reached, especially for sequences that converge from above. We will explore this further in the next section, as this type of behavior is especially pronounced in the deuteron.

However varied the details of converging sequences for different nuclei and interactions may be, the general shape remains extremely similar, just not quite as constrained as for energies. Especially the latter makes constructing extrapolation schemes much harder.

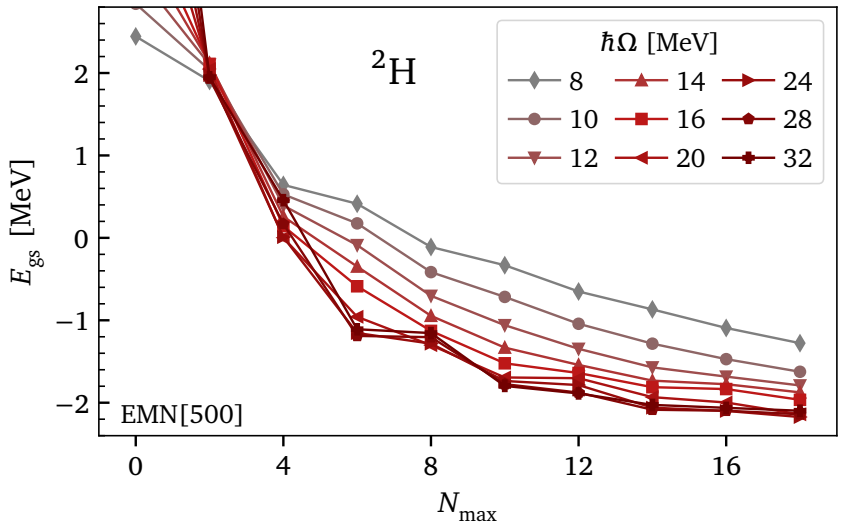


Figure 1.8: Ground-state energy convergence as a function of N_{max} for the deuteron. Note the very pronounced stairstep behavior of many sequences.

1.4.3 Deuteron Convergence Behavior

The convergence behavior of the deuteron warrants its own section, as it is the most extreme case for some pathologies that in some cases are rather unique to its exceptionally weak binding energy and restrictive angular-momentum coupling, which together with the limited combinatorics of just two particles leads to very slow model-space growth with increasing N_{max} compared to larger nuclei. However, it is important to note here that, though to a lesser extent, some artifacts or quirks exhibited by the deuteron also show up in the convergence of other nuclei.

Before diving any deeper into the specifics of the deuteron, we can take a look at what exactly differentiates the convergence behavior of deuteron from that of larger nuclei. For that, let us look at the ground-state energy for the EMN[450] interaction first, as is show in Fig. 1.8. The ground-state energy converges at a value of −2.22 MeV, with an experimental value of −2.224 566 MeV [97], making it an extremely weakly bound but still stable nucleus, with a binding energy per nucleus of just over 1.1 MeV. This exceptionally weak binding energy leads to comparatively extended nucleus and therefore a very slow convergence. Furthermore, instead of smoothly converging sequences, we observe a "stair-step" behavior for larger ħΩ values, where clearly not every N_{max} step leads to a significant—if any—improvement of the variational energy.

One way to think about this effect is to consider the combination of $\hbar\Omega$ and $N_{\max}$ as cutoffs when analyzing the infrared (IR) behavior of the NCSM. For large $\hbar\Omega$ or, equivalently small a_{HO}, although every even $N_{\max}$ adds additional s-wave radial basis states, these are highly localized and do not significantly extend the effective IR length scale of the model space. This is due to the fact that the IR length scale only increases $\propto \sqrt{N_{\max}} a_{\mathrm{HO}}$, i.e. very slowly for large $\hbar\Omega$. Additionally, these rather small changes are balanced by the kinetic energy and the potential energy changes associated with pushing the wave function further out. This leads to scenarios in which the change in IR scale must be large enough to reorganize the wave function composition based on the available basis states to better describe the asymptotic tail and without negatively affecting the description closer to the center, where the repulsive short-range behavior of the interaction is dominant. Therefore, not all new basis functions have a net positive effect on the binding energy, leading to the stair-step behavior.

The radius convergence shows a similar effect, though here, due to not being constrained by the variational principle, the sequences exhibit very pronounced zigzag pattern. In total, the deuteron exhibits the most extreme and least structured convergence patterns out of all nuclei accessible by the NCSM, while still conforming to the basic convergence behavior previously discussed for both observables.

1.5 Established Extrapolation Techniques

As discussed in Section 1.1.1, even for fairly light nuclei like ^{6}Li, achieving fully converged results is out of reach except for very soft NN-only interactions. When increasing the nucleon number a bit further, full convergence irrespective of the underlying interaction is computationally infeasible. For any quantifiable results, we are therefore often reliant on extrapolation techniques when using the NCSM. To facilitate the extrapolation and judge the convergence, one typically calculates all desired observables for all $N_{\max}$ up to the largest one compute constrains allow, for at least a few different HO basis parameters. Based on this data, extrapolation techniques can be employed. Such extrapolations are generally required to not only give a robust estimate of the fully-converged value of the observable, we also need these techniques to give an uncertainty estimate of the extrapolation itself. Furthermore, we require these extrapolation techniques to handle a wide variety of different cases, and ideally be constructed in such a way that the

quoted uncertainties can easily be combined with other sources of uncertainty, e.g. many-body uncertainties and chiral order-by-order interaction uncertainties. Of all non-trivial observables commonly calculated in the NCSM, the energy is the only one with established extrapolation schemes due to its highly constrained nature. There do exist some schemes for other observables like radii, however, these techniques are generally a lot less precise and robust compared to the ones for the energy and cannot be considered as established or widely used. To give an overview of different extrapolation techniques, we will take a closer look at the—at least to our knowledge—most commonly used techniques.

1.5.1 Exponential Extrapolations

Extrapolations based on the roughly exponential decay-shaped convergence of energies as a function of N_{max} are rather common and comparatively easy to do. In the most basic form, one simply fits an exponential decay function

$$E(N_{\text{max}}) = E^{\infty} + a\mathrm{e}^{-bN_{\text{max}}} \tag{1.25}$$

to a single $\hbar\Omega$ sequence to determine an estimate for the converged energy E^{∞}. However, since the convergence is typically not perfectly exponential, one usually only considers the last four N_{max} points for the extrapolation [98, 99]. Usually, one chooses the sequence closest to the variational minimum, i.e. the sequence with the lowest energy at the highest available N_{max}, for the best possible extrapolation result. Often, repeating this process for the neighboring sequences or different N_{max} windows can give a reasonable estimate of the robustness of the extrapolation [99].

A formalized method based upon this basic approach is outlined in Ref. [100] and is used in this thesis whenever we refer to exponential extrapolation. This recipe is very robust and gives reasonable, though rather conservative, uncertainty estimates.

While this technique is generally very robust and easy to implement, it does have some drawbacks. For one, the energy convergence is not exponential, as it generally converges slower than a best-fit exponential function. This can sometimes lead to a slight underestimation of the binding energy. While this can be partially remedied by raising the N_{max} in the exponential to the power of another fit parameter c, however the resulting function is still not a perfect fit for the convergence behavior, and the introduction of another fit parameter is often

not worth the increased effort and number of data points for a stable fit. There are other variants based upon the basic exponential extrapolation technique, for example fitting exponentials as shown in Eq. (1.25) simultaneously to several $\hbar\Omega$ sequences with the constraint that E^∞ is the same for all of them [101]. However, these techniques do not seem to see widespread use.

1.5.2 UV/IR Extrapolations

Another ansatz for an extrapolation technique for energies emerges when investigating the IR behavior of the NCSM [17, 102–104]. Casting the two main convergence parameters $\hbar\Omega$ and $N_{\max}$ into ultraviolet (UV) and IR cutoffs or length scales exposes the underlying structure of the HO basis within the NCSM. The idea is that if the NCSM is sufficiently UV converged, the remaining dependence on the effective IR length scale L_{eff} is exponential [105], and can simply, and analogously to the exponential extrapolation, be modeled as

$$E(L_{\text{eff}}) = E^\infty + a\mathrm{e}^{-2bL_{\text{eff}}}, \tag{1.26}$$

where E^∞ is the converged energy. This is due to the fact that for comparatively long wavelengths, the hard Gaussian cutoff of the HO basis functions effectively like a spherical potential well, where the exponential IR dependence can be shown analytically [102].

Reaching UV convergence in practice requires calculations at rather large $\hbar\Omega$, since the UV cutoff scales with $\sqrt{N_{\max}\Omega}$ and increasing $N_{\max}$ enough to compensate for smaller HO frequencies closer to the variational minimum is computationally infeasible for most nuclei and interactions. Therefore, calculations up to $\hbar\Omega = 50\,\text{MeV}$, well beyond the typical parameter range, are necessary. Furthermore, for this method to work well we need quite a few calculations for different $\hbar\Omega$ and $N_{\max}$ combinations, which are rather expensive to compute.

Analogous to our publication in Ref. [42], we use two variants of these IR extrapolations in this work. The first, simply referred to as “IR extrapolation” closely follows the approach in Ref. [104] and excludes data with $\Lambda_{\text{eff}} < 800\,\text{MeV}$ from the fit. A slightly modified alternative version we refer to as “manually optimized IR extrapolation” involves an additional manual selection of the data to ensure that “all points fall on a single narrow envelope” [104], since purely algorithmic or algebraic constraints sometimes fail to achieve this requirement. This results in a limited selection of 7–9 data points with the highest L_{eff} after the

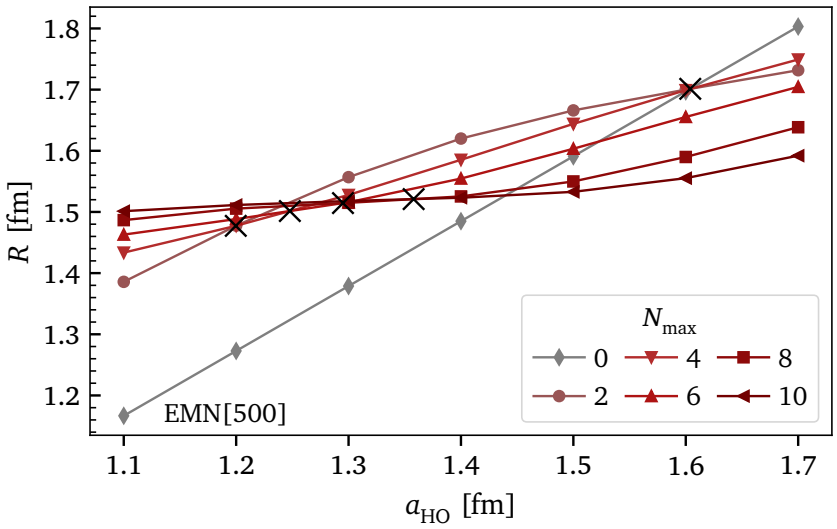


Figure 1.9: Illustration of the crossing point method to approximately determine the converged value of the radius by looking at the crossing point (black crosses) of successive N_{max} lines, which show the radius as a function of a_{HO}.

previously mentioned Λ_{eff} cut [42]. The uncertainties of this method are purely derived from the uncertainty of the exponential fit, limiting their meaningfulness and reliability as they lack the robustness to account for systematic errors [103]. We should also note that while corrections to radii using this method exist and were reported in Ref. [102], these are almost never applied, and no error analysis or uncertainty estimation is provided for them. Therefore, IR extrapolations will be used exclusively for ground-state energies in this work.

1.5.3 Crossing Point Method for Radii

As mentioned before, there do not exist widely established extrapolation techniques for radii. One reasonable approach to get at least a rough estimate of the converged radius is what we will refer to as the crossing point method. The idea of this technique is easiest understood when plotting the radius over the HO basis parameter, as is shown in Fig. 1.9. When viewing the data this way, we can clearly make out spots where many of the $R(a_{HO})$ curves for one fixed N_{max} cross the sequence of the neighboring $N_{max} - 2$. Ideally, these crossing points, shown in the figure as black crosses, are all at approximately the same radius, which can be used as a rough estimate for the converged radius. However, as the example shows, this does not always work quite as neatly, and the extrapolated

value generally exhibits a $N_{\max}$ dependence that only vanished for quite large and therefore expensive to compute $N_{\max}$ values.

We consider this approach a generalization of using the flattest convergence sequence as an estimate for the radius. That is because a completely flat behavior of the radius as function of $N_{\max}$ would be considered the optimal choice of oscillator length. In this scenario, the crossing point with any other sequence would still be on this line, hence the crossing point method is equivalent to finding the flattest sequence in this case. However, due to not relying on a single converging sequence, the crossing point method is not as dependent on the a_{HO} grid employed in the NCSM calculation, and also does not rely on a somewhat ambiguous definition of an *optimal* oscillator length. Furthermore, it is more robust to fluctuations in the radius convergence, where the flattest curve in the lower $N_{\max}$ range does not always correlate perfectly with the converged value. We will refrain from trying to estimate an uncertainty for this method, but will still give the estimate at different $N_{\max}$ to have a comparison point for benchmarking.

2

Artificial Neural Networks

Machine learning and ANNs have been an interesting and fruitful subject of research for about 70 years [106–108]. While research progressed steadily, larger-scale practical applications remained limited by the compute resources available at the time. With the increase of available computational power the possible usages of deep neural networks grew rapidly. This includes applications in across science, where the increased interest, together with easily usable libraries around machine learning and artificial neural networks [109–111] has spurred rapid development in applying these new techniques to various problems [44, 45]. Nuclear physics is no exception in this regard. There has been a wealth of published applications of machine learning methods and neural networks in particular, covering nearly all subdomains. A fairly current and comprehensive overview can be found in Refs. [46–49].

Before we look at the details of our technique, which applies ANNs to NCSM output data to predict the fully converged values, we will first introduce the basic building blocks of ANNs, give a general overview of the training process and the various methods involved in it, and explain the most important so-called hyperparameters that determine the structure as well as the behavior and training steps.

2.1 Fully-Connected Neural Networks

ANNs are inspired by biological neurons and their connections. The neurons in this sense are abstractions of their biological counterparts, simplifying the

behavior to a simple function that takes n inputs and computes a single output. We can simply write this process as

$$\vec{x} \longrightarrow \sigma(\vec{w} \cdot \vec{x} + b)\,, \tag{2.1}$$

where $\vec{x}$ is the vector of inputs, $\vec{w}$ are the weights, which determine the relative importance of each input of the input vector, b is the bias and σ is the so-called activation function. A graphical representation of a single neuron is shown on the left side of Fig. 2.1. Both $\vec{w}$ and b are free parameters that are optimized during the training of the network. The activation function introduces non-linearity, which is important to increase the effectiveness of the network.

ANNs consist of many of these neurons, and their arrangement and connections to each other determine the exact type of network, ranging from feed-forward neural network (FFNN) [112–114] and convolutional neural networks (CNNs) [115–118] to recurrent neural networks (RNNs) [119–123]. One of the simplest forms, which we will also be using for our application, are FFNN. Here, neurons are organized in layers, and the outputs of one layer are fed into the input of the next layer, thereby propagating through the whole network from the input layer, which receives the data, to the output layer, which return the result. The layers in between are called hidden layers and generally contain the bulk of the training parameters and are key to providing the necessary complexity.

A special type of these networks are fully-connected FFNNs. In these, the output of every neuron is piped into the input of every neuron of the next layer. Such a network is illustrated on the right side of Fig. 2.1. Every circle represents a neuron and the lines show each connection between the output of the left and the input of the right neuron.

With the help of Eq. (2.1) we can now write the effect of the ith layer as

$$\vec{L}_{i+1}(\vec{x}_i) = \vec{\sigma}(\mathbf{w}_i \cdot \vec{x}_i + \vec{b}_i)\,, \tag{2.2}$$

where $\vec{\sigma}$ is simply the scalar activation function applied element-wise to its argument, $\mathbf{w}_i$ is a matrix of weights, and $\vec{b}$ is a vector of biases associated with the ith layer. In fully-connected networks the weight matrix $\mathbf{w}_i$ of every layer is dense, since no connections are *a priori* disabled and therefore set to 0, which is why these are also called dense neural networks.

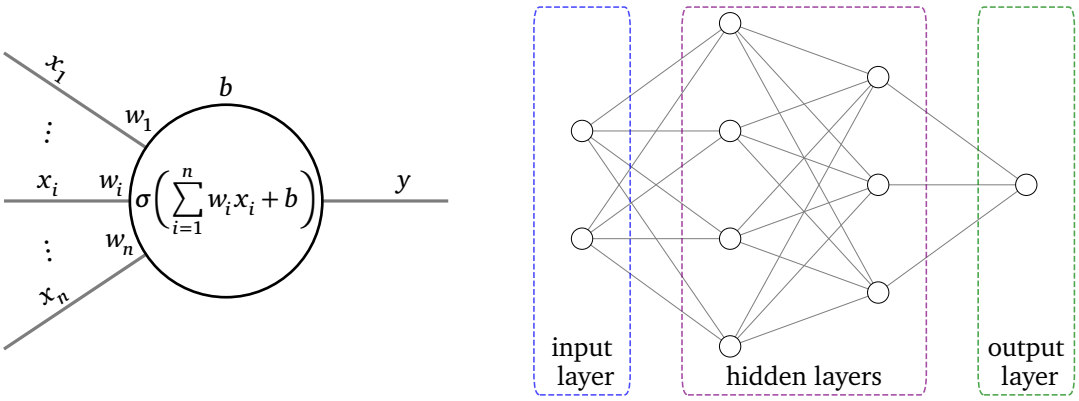


Figure 2.1: On the left: Schematic illustration of how a single neuron processes its inputs x_i weighted by w_i together with the bias b and the activation function σ to produce the output y. The right-hand side shows an abstraction of a fully-connected feed-forward neural network with input, hidden, and output layers. The circles correspond to neurons as depicted on the left, each with its own weights and bias.

With this definition, we can now easily construct a feed-forward network by simple composition of the different layers

$$\vec{y} = N_{L_1,\dots,L_n}(\vec{x}) = (\vec{L}_n \circ \vec{L}_{n-1} \circ \dots \circ \vec{L}_1)(\vec{x}), \tag{2.3}$$

where $\vec{x}$ is the input and $\vec{y}$ is the corresponding output vector produced by the network $N_{L_1,\dots,L_n}$.

2.2 Training of Artificial Neural Network

Having constructed a simple feed-forward neural network in the previous section, we now need to find a way to constrain all free parameters within the network, i.e. $\mathbf{w}_i$ and $\vec{b}_i$, in such a way that the network most accurately does its task. This is a high-dimensional, non-trivial optimization problem, where performance is key to enable larger network topologies. Before we can optimize the parameters, we first have to define a target and a way to measure the network's performance. To achieve this, we define a loss function

$$C\left(\vec{y}_{\text{target}}, N_{L_1,\dots,L_n}(\vec{x})\right) = C(\vec{y}_{\text{target}}, \vec{y}) = \|\vec{y}_{\text{target}} - \vec{y}\|, \tag{2.4}$$

which measures the distance between the network output based on the input $\vec{x}$ and the expected or target output $\vec{x}_{\text{target}}$. Both of these are part of the training data, which contains sets of inputs together with the corresponding expected output. Essentially, the networks are trained by example, and the loss function provides feedback how well the expected output is reproduced. This process is called supervised learning.

In practice, one typically does not calculate just a single loss for one specific input, rather one aggregates many different inputs and expected outputs, for example using the mean square error (MSE) defined as

$$C_{\text{MSE}} = \frac{1}{2\dim(L_n)} \sum_{i=1}^{m} (\vec{y}_{i,\text{target}} - \vec{y}_i)^2 , \tag{2.5}$$

summing over m-many network inputs resulting in network outputs $\vec{y}_i$ and targets $\vec{y}_{i,\text{target}}$, with $\dim(L_n)$ being the size of the output layer as a normalization. The number m is also known as the batch size.

The parameters of the network, i.e. the weights and bias, which are typically initialized randomly at the start of the training. With every batch, they are successively updated with every step t to minimize the loss via gradient descent

$$w_{t+1} = w_t - \eta \nabla_w C \tag{2.6}$$

$$b_{t+1} = b_t - \eta \nabla_b C , \tag{2.7}$$

where the gradient with respect to the weights or bias is scaled with the learning rate η, another important hyperparameter. The learning rate determines the step size at which the parameters are adjusted, and must be carefully chosen such that one avoids premature convergence into a suboptimal solution or needlessly waste computation time if the learning rate is too small. On the flip side, a learning rate that is chosen too large can lead to erratic behavior where the parameters do not properly settle down and converge. Typically, one wants to start with a comparatively large learning rate to quickly approach a minimum, and then systematically decrease the step size to converge to a minimum. For this purpose, learning rate schedulers are often used that adaptively decrease the step size.

The key insight that allowed the optimization of these vast parameter spaces of large neural networks in the first place is the so-called backpropagation algorithm, which enables an extraordinarily efficient computation of the gradient of the loss function with respect to the network parameters $\vec{\nabla}_w C$ and $\vec{\nabla}_b C$. For example, one obtains the following expression for the gradient with respect to the weights

w

$$\nabla_{\mathbf{w}_i} C = \vec{\sigma}_i' \circ \mathbf{w}_{i+1} \vec{\sigma}_{i+1}' \circ \ldots \circ \mathbf{w}_{n-1} \vec{\sigma}_{n-1}' \circ \mathbf{w}_n \vec{\sigma}_n' \circ \vec{\nabla}_{\vec{x}_n} C \cdot \vec{x}_{n-1}^{\mathrm{T}} \,. \tag{2.8}$$

by applying the chain rule. This expression can be unraveled and computed very efficiently by starting from the last layer and computing backwards throughout the network, calculating the gradients for the parameters for *i*th layer step by step. The individual terms can be expressed as matrix and vector operations, making them very fast to implement and run on modern computer hardware.

While we've focused on the most basic scenario, it's essential to acknowledge that modern backpropagation algorithms have evolved to incorporate more sophisticated modifications to these fundamental equations. Moreover, in practice, training is typically carried out using batches rather than individual samples, which significantly enhances the learning process and can be easily adapted for batch-based optimization. A full cycle through the entire dataset constitutes an epoch, and successful training usually requires multiple epochs. At first glance, it may seem straightforward to train neural networks. However, there are significant pitfalls and challenges that must be addressed. The primary concern is dealing with numerous free parameters, which necessitates access to substantial amounts of data. We will address this issue in more detail in the next section. Another critical issue related to parameter selection is under- or overfitting. Ideally, an ANN should be designed with sufficient parameters to capture correlations in the data without overly complicating it. If the network size is too small, it will fail to describe the underlying relationships. Conversely, a network that is excessively large may accurately fit training data but struggle with making reliable predictions on unseen information. Furthermore, the topology, i.e. the layout and dimension of each layer can have a large influence on the results. The numerous adjustments involved in training an ANN come with their own set of hyperparameters, which must be carefully tuned for optimal results.

We can monitor the training progress by looking at the loss over the course of the training process. For our purposes, we divide the training data into several independent subsets, including the training set which is used for the back propagation, and the development set, also referred to as test set, which is used at the end of every epoch to independently evaluate training progress and dynamically adjust the learning rate. Figure 2.2 shows the behavior of the loss on the development set over the course of many epochs for two different networks. The blue line shows a comparatively large network with 3 hidden layers that is fairly well suited for learning the patterns within the training data, while the

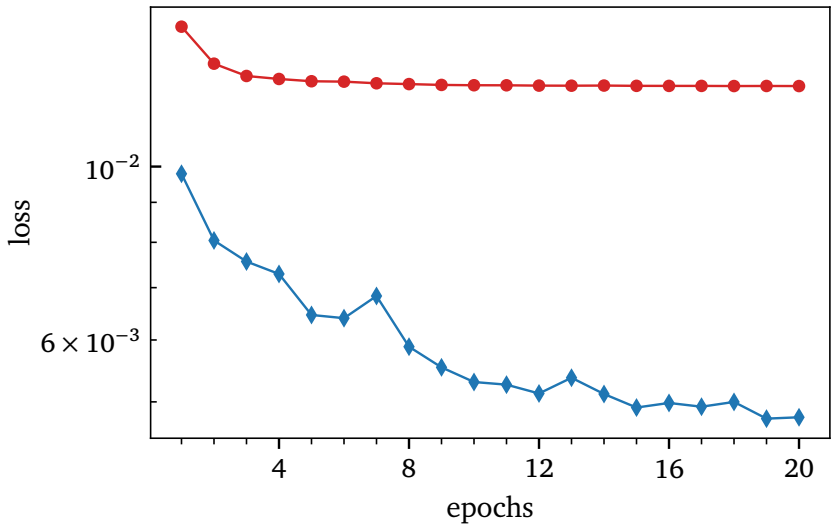


Figure 2.2: Illustration of the loss for two different network topologies, which work very well (blue line) or rather poorly (red line) in learning the pattern in the training data.

network which generated the loss shown by the red line only has one fairly small hidden layer that is not enough for the task, as it performs significantly worse and the loss changes comparatively very little over the course of the training. Compared with that the network which generated the blue line is clearly very sensitive to changes to the parameters during training, and shown significant training progress and improvement over the 20 epochs shown in the plot. Another aspect clearly visible here is the impact of the learning-rate scheduler, which leads to the slight stair-step behavior.

The last step before a network is accepted after training is called validation. Here, it is common to test the performance of the network on data not used during training to gauge their performance on unseen data. This is typically done by evaluating the network on a set of validation data and checking if the resulting loss exceeds a certain validation threshold. Additionally, other constraints on the network behavior can be checked in this step as well, for example to check for a certain behavior of the output in specific parameter regions, like monotonicity in one parameter. These additional constraints might not be fully covered by the training data and are often difficult to enforced explicitly during training. If a network does not pass the validation checks, it is discarded. This ensures that networks whose training does not generalize beyond the data it was trained on or shows other undesirable properties can be picked out automatically without intervention.

2.3 Training Data

While the basic architecture and other hyperparameters are very important for the successful deployment of a ANN-based method, at least as important are the data used to train the networks. As briefly discussed previously, we focus our efforts on supervised learning, where the training data is labeled, i.e. for every network input in the training data is labeled with its corresponding expected output. Since the neural networks can only learn correlations that are present in the training data, it is crucial that all relevant patterns are present in this data. Furthermore, care must be taken when assembling the training data set such that no biases are introduced, as these can manifest themselves in the networks trained with this data as well.

Another important aspect is that the training covers the range of input values expected in the application of the networks. This is important to reduce the chance of running into scenarios where for certain extreme input parameter combinations far away from the ones seen during training, the network parameters are not well constrained, which can result in nonsensical outputs. Put another way, we can think about this in terms of interpolation versus extrapolation. In this context, interpolation refers to the process of predicting or estimating a value within the range of training data. When an input falls within the domain covered by the training set, the network can make accurate predictions based on its learned patterns and correlations between inputs and outputs. This is typically performed when the input space is well-represented in the training dataset. In contrast, extrapolation occurs when the ANN is asked to predict or estimate a value outside the range covered by the training data. This can be problematic because the network lacks sufficient information to make accurate predictions. Extrapolation often requires additional assumptions about the underlying relationships between inputs and outputs, that are either built into the network architecture or used during training as additional constraints, or after training to validate the network and possibly discard it and train a new one if it does not meet the requirements. All these mitigations come with drawbacks and significant additional effort during the training process.

In practice, the aggregation of suitable training data is an important step in the development of neural networks, and involves not only the calculation of many different training sets, but also the filtering of the data to exclude undesirable behaviors and remove bias. Further treatment might be necessary on the level of

individual samples that are fed into the network to reduce for example the risk of overfitting. Often, a normalization step or similar schemes are used in this case.

3

Artificial Neural Networks for NCSM Energy Post-Processing

In previous chapters, we have discussed the NCSM as a method for solving the nuclear many-body problem and the associated difficulties with reaching converged results, as well as the basics of artificial neural networks. The goal for this chapter is to outline the development of ANNs capable of predicting the converged value of the ground-state energy given not fully converged NCSM results. This includes a robust uncertainty estimate for the prediction obtained from this method.

A framework to tackle the model-space extrapolation problem was first developed by Negoita et al. in Refs. [124, 125]. In this approach, the networks are trained to emulate a specific convergence pattern of the nucleus, interaction, and observable of interest. However, due to this specificity, new networks must be trained for each distinct evaluation. This potentially raises the requirements on the minimum number of data points required to successfully deploy this method, limiting its universality and increasing computational cost. Since networks are not transferable to other evaluations and even the addition of new data requires training a new set of networks, this can also slow down and increase the effort of analyzing NCSM results.

We want to produce a general set of neural networks that are only trained once and can then be applied to arbitrary NCSM calculation results to estimate the converged value. Thus, the ANNs discussed in this chapter follow a completely different *ansatz*.

3.1 Basic Concept

The goal is to train networks to recognize the NCSM convergence pattern of a specific observable and be able to predict the fully converged value. In this approach, ANNs are trained only once per observable, and applying them to a new nucleus or interaction only requires a limited set of evaluation data.

The basic idea is that the input for our neural networks are several converging sequences, from which the networks are then supposed to estimate the fully converged value. We can imagine this similar to if a human was tasked to estimate the converged value from a selection of different sequences that must all converge to the same value. When having seen enough NCSM data, one can at least roughly gauge the region where the final value is likely located. Using ANNs, we can turn this approach into a quantitative extrapolation or rather prediction method. By training the networks with a lot of converging sequences for which we know the fully converged value, the networks can pick up on the patterns within the data and be able to estimate the fully converged value. The trained networks can then be applied to data for which we lack the ability to calculate large enough model spaces to obtain converged results.

Since neural networks have a set input layer size, we need to fix the number and length of the sequences and the format of the data to feed into the network. To keep the minimal number of NCSM calculations required for a network evaluation low, as these can get very computationally expensive especially for larger nuclei, we choose to use three $\hbar\Omega$ sequences with a length of four consecutive $N_{\max}$ steps, as indicated on the left side of Fig. 3.1. From the NCSM data shown in gray, the selected data which is fed into the network in this example is highlighted in blue, yellow, and red. We call this selection of 12 data points which serves as the input data for the network a sample, where the input neurons are color-coded accordingly. One network evaluation results in a single number, which is the prediction for the fully converged value for $N_{\max} \to \infty$, shown in green. The network itself is a fully-connected feed forward neural network. We will explore different input dimensions and network topologies a bit later in this chapter.

It is important to note, that with enough, varied training data, the network is essentially interpolating between already seen convergence patterns to produce the output prediction. This is why we try not to refer to this method as an extrapolation method as we reformulated the problem into an interpolation problem instead, for which ANNs are very well suited.

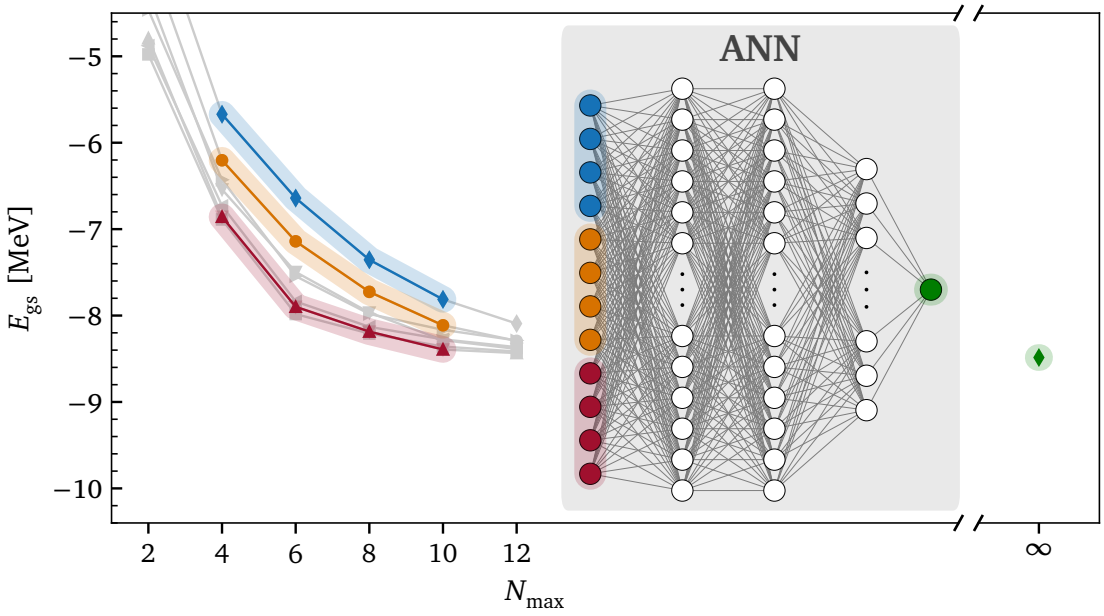


Figure 3.1: Illustration of the basic principle and in and outputs of our ANN design. Data for 4 successive N_{max} from 3 different $\hbar\Omega$ sequences (blue, yellow, red) are fed into the network as indicated. Based on this input, the network calculates its output prediction for the fully converged value (green). The topology of our ANN consists of an input layer of size 12, three hidden layers of depth 48, 48, and 24, as well as a single output neuron.

3.2 Ground-State Energy Training Data

As discussed in Section 2.3, the amount and quality of training data is paramount to the quality of the resulting networks. Since our network architecture has a prediction for the fully converged NCSM observable result as its output, the training data needs to include this value as well. For NCSM results, this essentially limits us to utilizing the Jacobi-NCSM and therefore $A \leq 4$, as these are the only cases for which we can reliably reach sufficient convergence using realistic interactions. In particular, we chose the nuclei ^{2}H, ^{3}H, and ^{4}He, which also cover a wide range from very weakly to very strongly bound. Since we want the training data to resemble typical cases, for which we want to evaluate the networks later, we chose to use data calculated with the realistic, chiral EFT-based non-local EMN NN and 3N interaction at N^{2}LO, N^{3}LO, and N^{4}LO$'$ for 3 different cutoffs $\Lambda = 450,\ 500$, and 550 MeV. In addition to using the bare interaction without SRG, we also utilize flow parameters of $s = 0.02,\ 0.04$, and 0.08 fm^4. We intentionally do not utilize the SMS family of interactions for training to be able to

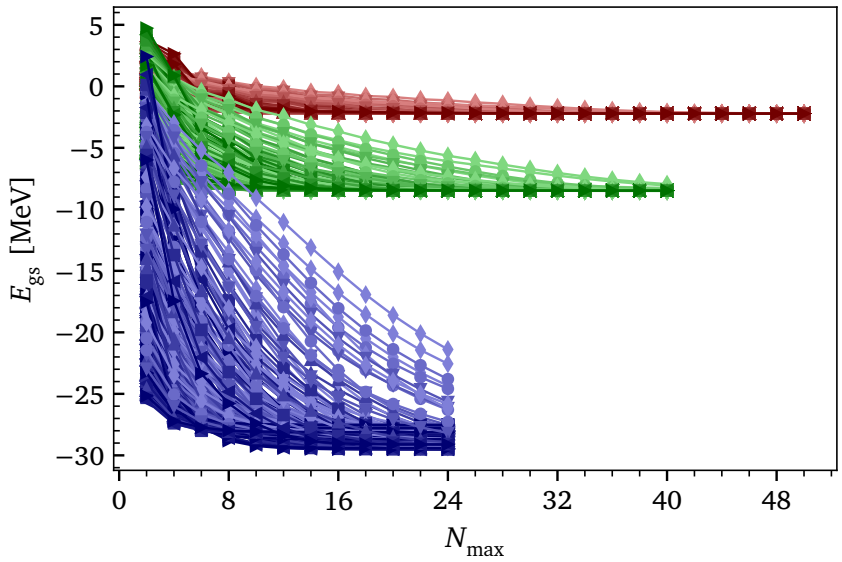


Figure 3.2: Plot of the complete training data for the ground-state energy networks comprised of NCSM calculations of ^{2}H (red), ^{3}H (green), and ^{4}He (blue) for 36 different versions of the non-local interaction.

use it to evaluate the networks for testing purposes with a completely independent interaction, even if the nucleus and observable was part of the training. Since the main focus is initially energy extrapolation, we used an HO grid of 7 different $\hbar\Omega$ parameters, namely $\hbar\Omega = 12,\ 14,\ 16,\ 20,\ 24,\ 28,$ and $32\,\mathrm{MeV}$.

A plot showing all training data for the ground-state energy obtained with the above-mentioned parameters is available in Fig. 3.2. The plot clearly highlights the different convergence speeds of the different interactions, which are mainly driven by the flow parameter of the SRG. Bare interactions, i.e. interactions with no SRG applied, converge much slower compared to ones with a large flow parameter, e.g. $s = 0.08\,\mathrm{fm}^4$. The figure also highlights the different model space size, for which data is available. For ^{4}He we are limited to $N_{\max} = 24$, and for ^{3}H we are limited to $N_{\max} = 40$, as we run out of CFP for larger model spaces. We artificially limit the data for ^{2}H to $N_{\max} = 50$ to strike a balance between including as much training data as possible, the quality of the data, and the goal to prevent ^{2}H data being overrepresented too much in the total training data set. In total, the training data consists of the results of 14364 individual Jacobi-NCSM calculations, 6300 for ^{2}H, 5040 for ^{3}H, and 3024 for ^{4}He.

However, from Fig. 3.2 it is also clear that, due to being limited to very light systems, the training data only covers a miniscule interval of the expected energy ranges of nuclei accessible to the NCSM. More worryingly, the fully converged results the networks are trained on are essentially only take on three different values, namely the ground-state energy of the three nuclei with only compar-

atively minor variations due to interaction and SRG differences, which could easily lead to overfitting or other deficiencies. We developed several strategies to alleviate the scale-dependence otherwise intrinsic to the training data, which will be discussed in Sections 3.4.3 and 4.2. Additionally, even irrespective of scale, an obvious question is whether our training set is really representative of NCSM data, especially for heavier nuclei than the very light systems in the training data. While the convergence behavior is clearly quantitatively similar to that of heavier nuclei, it is not clear whether it is also quantitatively close enough to be a useful foundation for training. We will address this question a bit later in Section 4.3.

Furthermore, not all generated data are equally valuable or reliable. To ensure the quality of our dataset, we implemented a filtering process that eliminated samples exhibiting erratic behavior, e.g. from numerical instability, or excessively slow convergence rates.

Finally, the fully converged value for each training sequence was extracted through an exponential fit applied to the data at the four highest $N_{\max}$ values, with the procedure, including the frequency selection, discussed in Section 1.5.1. While it may seem circular to train networks on values extracted via a classical extrapolation scheme, due to the extremely large $N_{\max}$ values and therefore almost completely converged nature of the data, the extrapolation length is minimal such that in this extreme case, the exponential extrapolation gives near perfect results in all cases.

3.3 Statistical Evaluation

Evaluating only a single neural network is not sufficient to meet the demand of also providing a robust uncertainty estimate for the prediction. Integrating an uncertainty estimator within the network architecture is a very hard problem, so similarly to Refs. [124, 125], we chose a different approach of training and evaluating multiple networks to obtain a distribution of predictions to estimate the reliability and uncertainty. The idea is that the evaluation of our ANNs can be significantly enhanced by employing statistical methods to mitigate the effects of outliers and provide a more robust estimate of the predicted observable. So instead of relying on a single network, we train and evaluate $n_{\text{nets}} = 1000$ networks, which only differ in the initialization of the weights and biases.

In cases where we have more than the minimum of 12 NCSM calculations, i.e. more than three $\hbar\Omega$ sequences for at least 4 overlapping consecutive $N_{\max}$, at

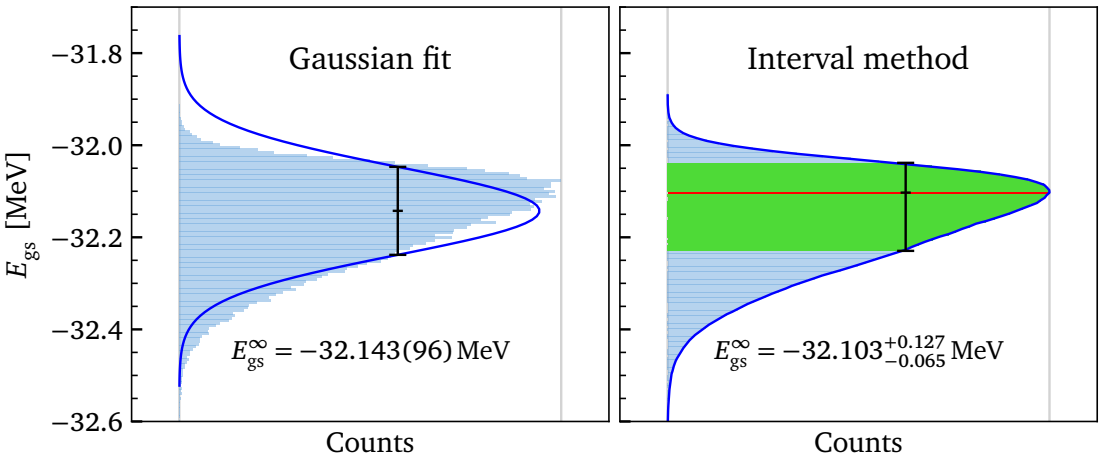


Figure 3.3: Example histogram analysis showing the Gaussian fit method in the left panel and the interval method in the right panel. See text for details.

our disposal, we can further increase the amount of output data. We generate n_{samples} many samples of the evaluation data by chosing all possible combinations and permutations of three $\hbar\Omega$ sequences. Evaluating all available sample with all networks results in an even larger set of predictions that can be used for statistical analysis. The first step is to construct a histogram from the output data, i.e. the $n_{\text{total}} = n_{\text{nets}} \cdot n_{\text{samples}}$ network predictions of the converged observable. The bin size and therefore the number of bins of the histogram is determined heuristically via the Freedman-Diaconis rule [126], which is based on the interquartile range. This makes it rather robust and well suited for a range of distributions the networks produce, especially slightly skewed distributions as well as the wide range of available evaluation data, since n_{samples} can vary significantly and is usually between just a (few) thousand and several hundred thousand and more depending on the amount of evaluation data available.

Over time, we refined how we extract the final prediction of the whole evaluation procedure from the histogram. The method described and used in Refs. [127, 128] is based on fitting a Gaussian to the histogram to obtain an estimate for the mean and standard deviation as an indicator of the converged value and 1σ uncertainty, respectively. An example for such an analysis is shown in the left panel of Fig. 3.3, where the histogram is shown in light blue, the fitted Gaussian in dark blue, and the error bar indicates the mean and uncertainty based on the standard deviation. Fitting a Gaussian works very well for output distributions that are symmetrical and generally fairly close to a normal distribution. The advantage of fitting a Gaussian to the histogram compared to just calculating the mean and average of the distribution is that the former is slightly more robust to outliers

and smaller multimodal peaks that are sometimes present in the evaluation. However, while a majority of distributions obtained from the networks are very close to a normal distribution, this approach is less well suited if the distribution deviates significantly from a simple Gaussian, for example if it is asymmetrical, as is the case for the histogram in the right panel of Fig. 3.3. Especially skewed distributions are not well described, as the obtained center value of the Gaussian is shifted away from the peak of the distribution. Additionally, by construction this model can only produce symmetric uncertainties around the mean value, which does not reflect the shape of a skewed distribution.

To improve the results in such cases, a new approach is needed to provide more robust and accurate estimates that better reflect the underlying distribution. We start by smoothing the histogram via a moving mean with a width of 10 histogram bins. This is important to remove high-frequency noise from the histogram, which would otherwise negatively affect the next steps. The next step is selecting the peak of the distribution, i.e. the most likely histogram bin, as the best estimate of the converged value. This is both sensible when considering the results obtained in Refs. [127, 128], and is independent of the skewness of the distribution. To determine the uncertainty, we add the next-largest adjacent bin next to the previously selected largest histogram bin. Repeating this process iteratively, always adding up the next-largest bin on either side of all previously added bins, until 68 % of the total distribution is reached. This procedure is shown in the right panel of Fig. 3.3, where the largest peak and therefore the final prediction is shown in red, and the gathered bins corresponding to 68 % of the histogram are shown in green. As before for the Gaussian fit model, the final prediction and uncertainty is indicated by the black error bar. This asymmetric uncertainty is better at capturing the underlying skewed distribution while at the same time being equivalent to the previous approach for Gaussian distributions and even more resistant to outliers.

3.4 Artificial Neural Networks Development

Before showing results for our ANN-method, we first want to dive deeper into some of the design decisions and hyperparameter choices and give a general overview over all important aspects of the networks. This section focuses on the networks as presented in Ref. [127], though the design is identical except for the input formatting to later networks. The Gen1 networks actually consist of two

distict but very similar network architectures, where the main difference is the form in which the data of 12 NCSM calculations is fed into the networks, which also has cascading effects on the topology and some other hyperparameters as well. We will highlight the differences between these two network types, called Abs and Diff and show all relevant hyperparameters used for training. Additionally, we will highlight the influence of altering the topology and input layer size.

Differences compared to the Gen2 networks that are described and used in our newer publications such as Refs. [42, 128–130] will be highlighted in a later chapter.

3.4.1 Main Hyperparameters and Network Training

The behavior of neural networks can be tuned and significantly affected by the so-called hyperparameters. These include properties of the network itself, like its topology, size, and activation function, i.e. model hyperparameters, as well as training hyperparameters that include for example the batch size, optimizer, or learning rate scheduler.

Often, a structured, algorithmic approach for the hyperparameter optimization is employed. However, we decided to manually search and optimize the hyperparameters for our networks for a number of reasons. Most importantly, since we use labeled data, i.e. Jacobi-NCSM results for ^{2}H, ^{3}H, and ^{4}He for which we can calculate the fully converged value, an algorithmic optimization scheme would only optimize for the best performance in this regime. However, this is not necessarily the best for larger nuclei, where only very coarse rejection-based criteria, like the violation of the variational principle for energy predictions, could be implemented. Additionally, we found that the networks are not overly sensitive to relatively small variations in, for example, the dimensions of the hidden layers, so a manual optimization is still feasible. Therefore, the hyperparameters of the Gen1 networks were determined via a manual, semi-random search of the hyperparameter space. While there might still be room for improvement of these parameters, we do not expect these to significantly affect the results after taking our sampling and statistical evaluation procedure into account.

Table 3.1 shows the resulting hyperparameters for the Gen1 networks. The input format will be discussed in detail in a later section, but it is important to note that it does determine the size of the input layer, so either 12 in the case of Abs, or 9 for the Diff input mode, as well as a few other parameters.

Hyperparameter	Gen1 Abs	Gen1 Diff
input mode	Abs	Diff
input layer dimension	12	9
hidden layer dimensions	48, 48, 24	36, 36, 24
output layer dimension	1	
activation function	ReLU	
loss function	MSE	
optimizer	AdamW	
batch size	512	
learning rate scheduler	ReduceOnPlateau	
initial learning rate	0.01	
epochs	20	
threshold	50 keV	10 keV
weight/bias initialization	random from $\left[-\frac{1}{\sqrt{\#\text{inputs}}}, \frac{1}{\sqrt{\#\text{inputs}}}\right]$	

Table 3.1: Summary of the main hyperparameters used for the Gen1 networks.

Since we scale the hidden layer dimensions based on the input dimensions, these differ slightly as well. All weights and biases are initialized uniformly random from an interval that scales inversely proportional with the root of the number of inputs for the corresponding neuron. For the activation function we chose a rectified linear unit (ReLU) [131] for all except the output layer, as it performed significantly better than other available options such as a sigmoid function. As previously discussed, we calculate the loss via a mean-square error (MSE) loss function and combine it with the AdamW optimizer algorithm [132], using a batch size of 512. The initial learning rate is set to 0.01 and dynamically reduced by half if the loss plateaus for more than two epochs in a row. The threshold for changes to still be detected as a plateau is set to 0.1 MeV. It is important to note that the loss used for the learning rate scheduler is not calculated from the training, but rather from the entirely separate development or test set, and is therefore completely independent from the data used for the weights and bias optimization. Each network is trained for exactly 20 epochs, and is evaluated on the entirely separate validation set. If the loss on this set is larger than the threshold value, the ANN is rejected and an entirely new network is trained instead. For the parameters shown above, hardly any networks are rejected at this last step.

We will discuss the effects of a small selection of the most important of these hyperparameters in more detail in the next section. For brevity, these will only be discussed for the Abs mode, though the main insights apply equally to both

input modes and also the improved Gen2 networks we will cover in the next chapter.

3.4.2 Network Topology Optimization

The input layer dimensions of our ANN plays a crucial role in determining its performance and computational requirements. Specifically, we require sequences of consecutive N_{max} values to capture the underlying physics of nuclear systems. A minimum of one such sequence is necessary to establish a baseline, but longer and more numerous sequences are expected to yield better results due to the increased ability to capture complex correlations. However, there are several drawbacks associated with increasing the number of input layer dimensions in this manner. Firstly, as we require more NCSM data points for each sequence, the total amount of training samples that can be constructed from the data decreases rapidly with an increase in their length. While this is not a big problem for the training nuclei, which are fairly cheap to calculate, an increase in the length or the number of sequences is more significant for larger nuclei. This is due to the fact that all subsequent evaluations using these networks require at least NCSM calculations for n_{seq}-many different $\hbar\Omega$ sequences with at least l_{seq} consecutive N_{max} steps. While the calculation of more $\hbar\Omega$ sequences increases the cost only linearly, calculating larger N_{max} can be excessively expensive to calculate for larger nuclei, hence limiting the applicability of the networks in practice.

To strike a balance between these competing demands, we default to using three different $\hbar\Omega$ sequences with four consecutive N_{max} values each in our input layer, resulting in an overall dimensionality of 12. This compromise allows us to calculate sufficient data points even for slightly larger nuclei. Another important feature is the possibility to estimate the convergence and stability of the network predictions for cases in which there are data for more than 4 consecutive N_{max} steps available. By simply grouping the evaluation samples by their maximum N_{max}, which we refer to as $\mathcal{N}_{\text{max}}$, and performing the network evaluation for each $\mathcal{N}_{\text{max}}$ in the available data, we can compare the predictions and judge the stability of the results.

A selection of different input layer dimensions is shown in Fig. 3.4 for several numbers of sequences n_{seq} in the upper panels and several lengths l_{seq} in the lower panels. The upper left plot clearly demonstrates that a single sequence ($n_{\text{seq}} = 1$) is not sufficient for a precise and accurate prediction of the converged energy. Adding a second sequence per sample leads to a significant improvement, though

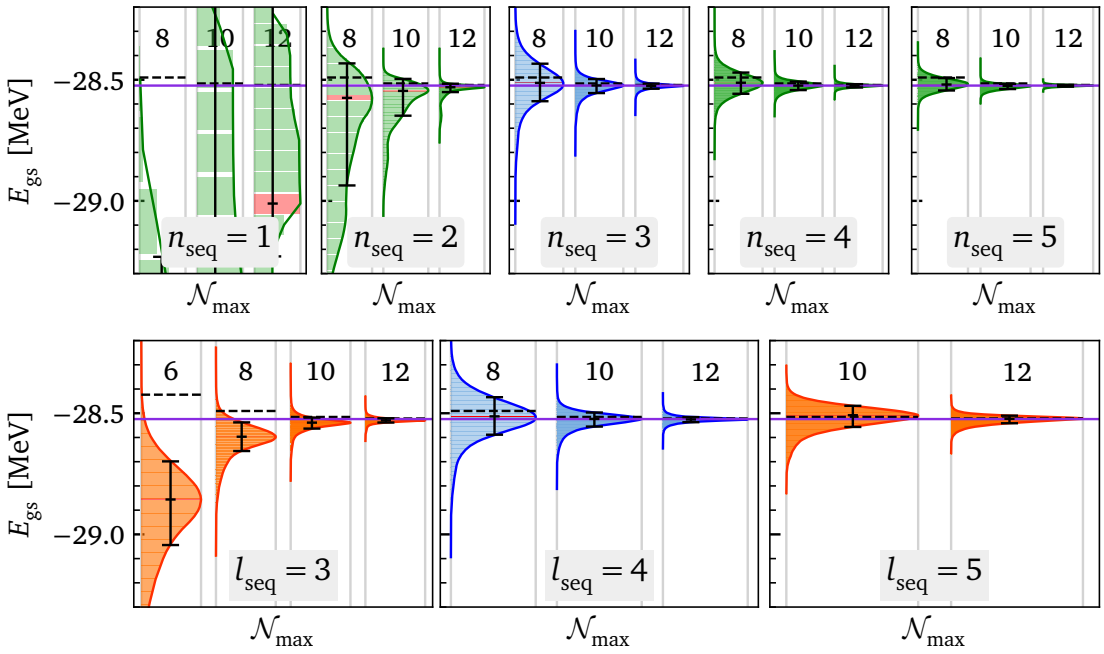


Figure 3.4: Network evaluation of the ground-state energy of ^{4}He, illustrating the effects of networks trained and evaluated with different numbers of sequences, i.e. n_{seq}-many differerent $\hbar\Omega$ per sample (top panels) with $l_{seq} = 4$, and different sequence lengths l_{seq} for $n_{seq} = 3$ fixed. Larger l_{seq} lead to fewer available $\mathcal{N}_{max}$ in the evaluation. The standard parameters for the Abs networks are shown in blue, and the fully-converged value is indicated by the violet line.

n_{seq} outperforms it significantly in precision, accuracy and stability in $\mathcal{N}_{max}$. More sequences give diminishing returns, improving mainly on the precision of the predictions while the accuracy remains about the same. Using three distinct sequences is therefore a good compromise between accuracy, stability, and computational requirements.

Looking at the length of the sequences l_{seq} next, using data from just 3 N_{max} steps as shown in the lower left panel of Fig. 3.4 results in predictions that are not stable with respect to $\mathcal{N}_{max}$. While it does allow evaluations with one less, and potentially very expensive to calculate, N_{max} step, the resulting accuracy is clearly insufficient for practical applications. Adding data for one more N_{max} step to the sequences dramatically improves the accuracy and gives extremely stable results. Going to a sequence length of 5 does not amount to any improvements in the predictions, and takes away one more $\mathcal{N}_{max}$ step from the evaluation. Given these results and considering the very stark increase in computational cost associated with every N_{max} step in NCSM calculations, a sequence length of $l_{seq} = 4$ is the overall best choice.

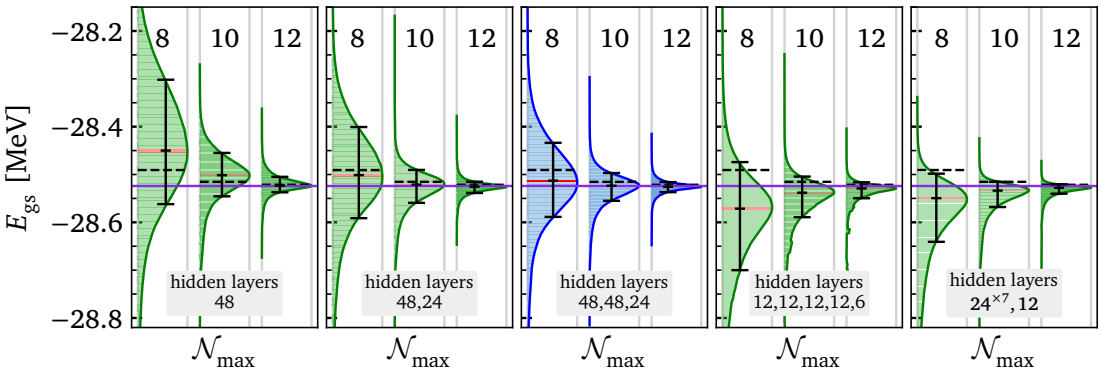


Figure 3.5: Evaluation of the ground-state energy of ^{4}He for networks with 12 input nodes, one output nodes and varying hidden layer dimensions indicated in each panel. The notation in the right-most panel means 7 hidden layers of size 24, and one of size 12. From left to right, the total number of free parameters for the network topologies shown are 673, 1825, 4177, 709, and 4225.

We should note that the network topology was scaled according to the input layer dimension, with hidden layer sizes of $4n_{\text{seq}}l_{\text{seq}}$, $4n_{\text{seq}}l_{\text{seq}}$, and $8l_{\text{seq}}$. Furthermore, the validation threshold had to be increased significantly for some of the worse-performing input dimensions to produce any networks. While the example shown in Fig. 3.4 is only a single case of many, it is representative of the results obtained for other interactions and nuclei. Therefore, the default input sample consists of data for 4 consecutive N_{max} steps from 3 different $\hbar\Omega$ sequences.

Another important hyperparameter that significantly affects the performance of the ANNs is the number and size of hidden layers. These are the main drivers of the network's ability to learn and recognize patterns. As discussed previously, a network with a single hidden layer of arbitrary size combined with a non-linear activation function is sufficient to fully describe an arbitrary scalar function of a single variable [112]. However, to learn more complex patterns arising from the correlations of multiple input parameters, more hidden layers are needed, as is demonstrated in Fig. 3.5. Here, it is obvious that a network with a single hidden layer shown in the left panel is not quite enough to process the complex correlations within the data, as its predictions are above the variational minimum and therefore physically nonsensical, and is therefore not able to provide reasonable predictions for the converged ground-state energy. Adding another hidden layer with 24 neurons before the output layer improves the predictions noticeably. The decrease in the layer size of the last hidden layer is to facilitate the condensation of the information towards the single-node output layer. The topology with 48, 48, and 24 neuron in its respective hidden layer, shown in the center panel, shows the best performance in both accuracy and

stability across $\mathcal{N}_{\text{max}}$. More, smaller hidden layers as shown in the last two panels of Fig. 3.5 perform overall worse, though the results of the network depicted in the right most panel is significantly improved compared to the one to its left. Together with the number of free parameters listed in the caption for all these topologies, it is likely that the worse performing topologies simply do not have enough free parameters to properly learn and model the convergence patterns and correlations within the data. However, the arrangement of hidden layers also plays a role, since the network in the last panel performs notably worse in accuracy and stability compared to our standard topology, even though the former has more free parameters at its disposal.

In conclusion, this supports our choice of default network topology using an input layer of size 12, with data from 3 different $\hbar\Omega$ sequences for 4 consecutive N_{max} steps each, three hidden layers of depth 48, 48, and 24 and a single output neuron for the Abs network architecture. This results in networks with around 4200 free parameters, which is significantly less than the amount of labeled training samples we are able to produce, as we will explore in the next section. It is important to highlight that while the results shown in Figs. 3.4 and 3.5 are exemplary for the general behavior we observe in the different network topologies, our general assessment is based on a lot more network evaluations. This includes other interactions and few-body systems for which we can calculate the converged value, and also heavier nuclei for which we can check the stability and assess the general behavior. This comprehensive review of the different architectures resulted in the above-mentioned topologies for the Abs and Diff networks, as well as the other hyperparameters listed in Table 3.1.

3.4.3 Input Format and Sample Generation

Another very important aspect we need to cover is how exactly the input to the networks is generated from the training or evaluation data. This has a considerable influence on the training process, performance of individual networks as well as the statistical evaluation.

As discussed previously, we found that taking information from three distinct HO parameters and four consecutive N_{max} to be the best compromise between network accuracy and data availability. This is achieved by arranging corresponding

observable data calculated in the NCSM together into an $\mathbb{R}^{12}$ vector

$$\begin{aligned} S_{\text{Abs}}^{\mathcal{N}_{\max}} = (&O_{\hbar\Omega_1}^{\mathcal{N}_{\max}-6}, O_{\hbar\Omega_1}^{\mathcal{N}_{\max}-4}, O_{\hbar\Omega_1}^{\mathcal{N}_{\max}-2}, O_{\hbar\Omega_1}^{\mathcal{N}_{\max}}, \\ &O_{\hbar\Omega_2}^{\mathcal{N}_{\max}-6}, O_{\hbar\Omega_2}^{\mathcal{N}_{\max}-4}, O_{\hbar\Omega_2}^{\mathcal{N}_{\max}-2}, O_{\hbar\Omega_2}^{\mathcal{N}_{\max}}, \\ &O_{\hbar\Omega_3}^{\mathcal{N}_{\max}-6}, O_{\hbar\Omega_3}^{\mathcal{N}_{\max}-4}, O_{\hbar\Omega_3}^{\mathcal{N}_{\max}-2}, O_{\hbar\Omega_3}^{\mathcal{N}_{\max}}), \end{aligned} \tag{3.1}$$

where $O_{\hbar\Omega}^{N_{\max}}$ is the value of the observable for the HO basis parameter $\hbar\Omega$ and $N_{\max}$, and $\mathcal{N}_{\max}$ signifies the largest $N_{\max}$ of the $N_{\max}$ window from which the data in the sample is taken. For this input mode, the networks are trained on the calculated fully-converged value of the observable O^{∞}. When evaluating the networks, the output is therefore directly the prediction for O^{∞}. For the Abs input mode, Eq. (3.1) is already the full input sample. In this mode, the data is simply fed into the networks as-is.

It is important to note that, irrespective of the input mode, we always generate all possible samples from the available input data. This means we generate data for all possible combinations of three distinct HO basis parameters, as well as all permutations of these within the sample. So for data with $n_{\hbar\Omega}$-many available HO frequencies in a certain $N_{\max}$ window, a total of

$$n_{\text{samples}} = \binom{n_{\hbar\Omega}}{3} \cdot 3! = n_{\hbar\Omega}^3 - 3n_{\hbar\Omega}^2 + 2n_{\hbar\Omega} \tag{3.2}$$

samples are generated per $\mathcal{N}_{\max}$. Together with the 1000 networks that are evaluated at the same time, this means there are always at least 6000 predictions available for the statistical analysis discussed previously in Section 3.3.

Going back to the input modes, the Diff mode aims to compress the input data compared to the Abs mode and reduce it to the essential information by taking the difference between subsequent $N_{\max}$. Since we want to keep the raw information used by the networks for their predictions of the converged value the same, a Diff sample therefore contains only 9 values

$$\begin{aligned} S_{\text{Diff}}^{\mathcal{N}_{\max}} = (&\Delta_{\hbar\Omega_1}^{\mathcal{N}_{\max}-4}, \Delta_{\hbar\Omega_1}^{\mathcal{N}_{\max}-2}, \Delta_{\hbar\Omega_1}^{\mathcal{N}_{\max}}, \\ &\Delta_{\hbar\Omega_2}^{\mathcal{N}_{\max}-4}, \Delta_{\hbar\Omega_2}^{\mathcal{N}_{\max}-2}, \Delta_{\hbar\Omega_2}^{\mathcal{N}_{\max}}, \\ &\Delta_{\hbar\Omega_3}^{\mathcal{N}_{\max}-4}, \Delta_{\hbar\Omega_3}^{\mathcal{N}_{\max}-2}, \Delta_{\hbar\Omega_3}^{\mathcal{N}_{\max}}), \end{aligned} \tag{3.3}$$

with

$$\Delta_{\hbar\Omega}^{N_{\max}} = O_{\hbar\Omega_1}^{N_{\max}} - O_{\hbar\Omega_1}^{N_{\max}-2}. \tag{3.4}$$

For the DIFF input mode, the networks are trained on the smallest difference between the observable values at the largest $\mathcal{N}_{\max}$ and the converged value

$$\Delta^{\infty} = O^{\infty} - \min\left(\left\{ O_{\hbar\Omega}^{\mathcal{N}_{\max}} \text{ for all } \hbar\Omega \text{ in } S_{\text{DIFF}}^{\mathcal{N}_{\max}} \right\}\right). \tag{3.5}$$

The main advantage of the DIFF method is that it vastly reduces the range of possible input values and especially output values. Ground-state energies for nuclei accessible by the NCSM range from around $-2\,\text{MeV}$ for deuteron to around $-130\,\text{MeV}$ for ^{16}O and beyond. This vast difference in scales, which is of course also present in the input data, is alleviated by the DIFF method. We can visualize this difference by looking at the target values the networks are trained to predict, i.e. O^{∞} for ABS and Δ^{∞} for DIFF. A stacked histogram of the distribution of these values for all samples used to train both network types, separated by nucleus and using the ground-state energy as the observable, is shown in the upper two panels of Fig. 3.6. The left plot shows the distribution of the converged ground-state energies in the training data for the ABS method. There are three clear peaks, corresponding to the ground-state energies of ^{2}H, ^{3}H, and ^{4}He. We also notice there are slightly more ^{2}H samples than ^{3}H samples, and significantly more than there are for ^{4}He. This is simply due to the larger number of $N_{\max}$ steps available for ^{2}H than for ^{3}H and especially ^{4}He. The slight variations of the converged ground-state energy, especially visible for ^{4}He, is due to the slightly different results one obtains for different orders, cutoffs, and SRG evolutions of the 36 versions of the EMN interaction used. As is, this data is virtually unusable for training ANNs, as it would very likely lead to networks which simply reproduce the three distinct energy values corresponding to the three peaks in the histogram without learning to predict the converged energy from an arbitrary sample, especially one not present in the training data with a very different converged ground-state energy.

The upper left panel of Fig. 3.6 analogously depicts the target values $\Delta_{E_{\text{gs}}}^{\infty}$ for the DIFF method. While the distribution strongly peaks for values at or slightly below 0 MeV, it does not exhibit the strong grouping of data for each nucleus. Though it does not necessarily cover the whole range of values expected for larger nuclei, where worse convergence and therefore larger absolute differences between the minimum energy in the sample and the converged energy are expected.

To combat these issues of the distribution of the converged values the networks are trained on, we introduce randomness into the samples. For the ABS input

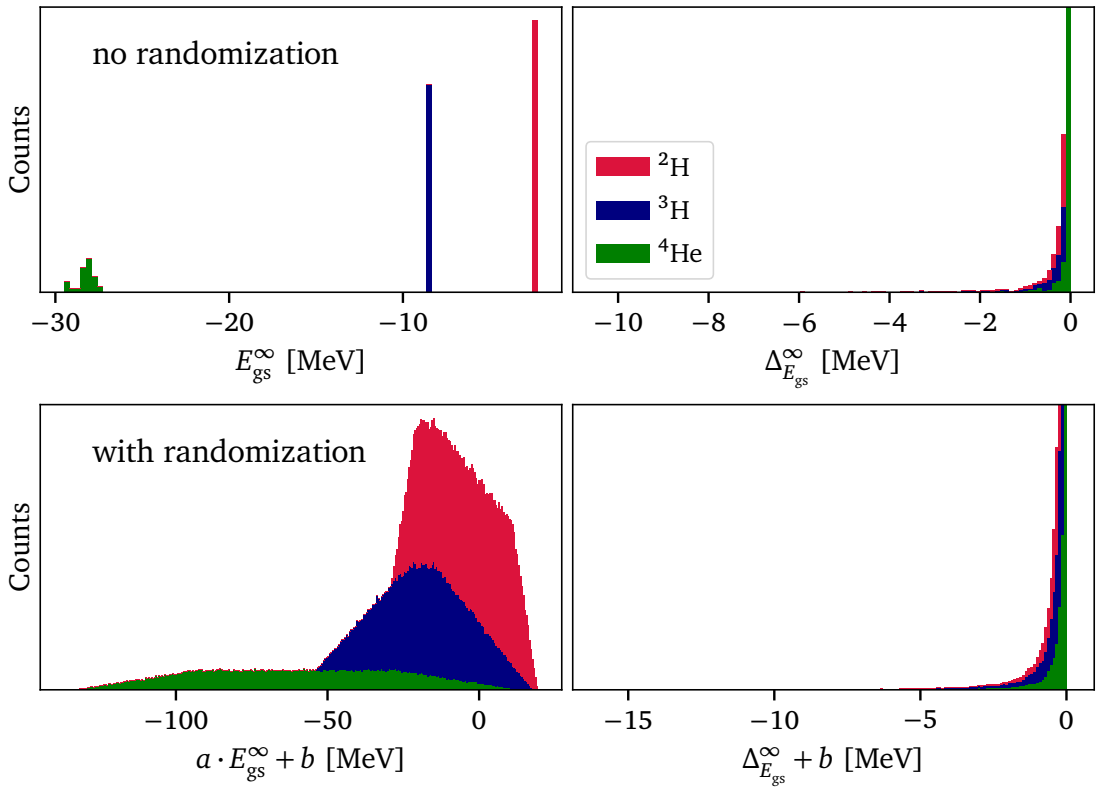


Figure 3.6: Stacked histograms showing the distribution of target values in the ground-state energy training data without randomization (upper panels) and with randomization (lower panels). The left-hand histograms are for the Abs input mode and the right-hand histograms show the distribution for the Diff mode. The latter shows a zoomed-in view of the distribution, and the peak is cut off at approximately 20 % of its total height.

method, the training samples are randomly scaled and shifted via

$$S_{\text{Abs,rand}}^{\mathcal{N}_{\max}} = a \cdot S_{\text{Abs}}^{\mathcal{N}_{\max}} + b\,, \tag{3.6}$$

with two independent, uniform random variables a and b, where $a \in (0.25, 4)$ and $b \in (-20, 20)$ MeV. Since shifting the Diff method samples after calculating the differences interferes with the general convergence structure of these samples and shifting them before is pointless due to Eqs. (3.4) and (3.5), we only scale these according to

$$S_{\text{Diff,rand}}^{\mathcal{N}_{\max}} = a \cdot S_{\text{Diff}}^{\mathcal{N}_{\max}}\,, \tag{3.7}$$

with $a \in (0.5, 2)$. For both methods, the labels of each training sample, i.e. O^{∞} for Abs and Δ^{∞}, is transformed via the same values as the sample it belongs to.

The lower panels of Fig. 3.6 show the distribution of the target value of each sample with the discussed randomization, specifically for the ABS mode via Eq. (3.6) in the left panel, for DIFF via Eq. (3.7) in the right panel. For the DIFF input mode, the effects of the random scaling are mainly a widening of the distribution towards smaller values. This extends the possible target values towards ones more representative of larger, less converged nuclei. However, the extremely high peak at or close to zero, which is cut off in both the upper and lower panel and therefore even more dominant than shown here, is still present.

The effect of the randomization is quite different for the ABS mode, where the random scaling and shifting is able to smear out the raw target values into a fairly wide distribution, reaching down to around $-130\,\mathrm{MeV}$, i.e. regions of the ground-state energy values found in much heavier nuclei such as ^{16}O. This breaks up the clustering observed before, and the distribution is not nearly as concentrated on a few values as before, though it is still far away from uniform. There is therefore still a significant bias towards higher ground-state energies, peaked at around $-20\,\mathrm{MeV}$. Some ground-state energy training targets are even shifted to positive values, though of course the corresponding sequences would also be shifted up accordingly. While this is nonsensical physically, we have to remind ourselves that this data only serves to train ANNs to learn the general, universal convergence pattern hidden within the data. As long as it does not impede the training process, this transformation is perfectly acceptable for our purposes. Furthermore, we see that the data stretching below $-50\,\mathrm{MeV}$, i.e. the regime for heavier nuclei accessible by the NCSM, consists exclusively of shifted and scaled ^{4}He data. However, we consider this a significant improvement compared to the data without randomization, as it alleviates the main concern with that data, and enables us to start training networks with the few-body Jacobi-NCSM data we calculated.

Beyond spreading out the training data and alleviating problematic structures within it, the randomization also allows us to inflate the number of training samples beyond their initial number, since we can in principle generate an unlimited amount of different random samples with this method. There are, however, some limitations to this, since at some point, more training data does not improve the resulting networks by any significant margin. In fact since these new data points are heavily correlated with each other via Eqs. (3.6) and (3.7), many of these do not contain an appreciable amount of new, independent information, and are therefore of limited effectiveness for the training.

In practice, this means that from the 362 880 raw samples we can generate from the 14 364 Jacobi-NCSM calculations of ^{2}H, ^{3}H, and ^{4}He, we generate a

total of 1 150 000 randomized samples, divided into three independent sets: The training set which consists of 1 000 000 samples and is used for the main training loop of the networks, the development set, also referred to as test set, consisting of 100 000 samples and is used for the dynamic adjustment of the learning rate via the learning rate scheduler, and the validation set with 50 000 sample which acts as a discriminator for fully trained networks to reject ones that perform too poorly on this set of fresh, unseen data.

3.5 Results

After training a thousand networks according to the procedure described in the previous sections and the parameters outlined in Table 3.1 for both the Abs and the Diff input mode, we can finally look at the results we can achieve with both sets of ANNs. First, however, we should note that while the training data only contains NCSM data obtained with the EMN interaction, all evaluations shown in this section are based on a completely separate and independent family of interactions. To be more precise, we show results calculated with the SMS family of interactions, specifically the N^{2}LO nucleon-nucleon and three-nucleon (NN+3N) interaction with a cutoff of $\Lambda = 450\,\text{MeV}$ and SRG evolved with a flow parameter of $s = 0.08\,\text{fm}^4$. The evaluation data was calculated in the NCSM and Jacobi-NCSM for the same selection of 7 HO frequencies as in the training data, i.e. $\hbar\Omega = 12,\ 14,\ 16,\ 20,\ 24,\ 28$, and $32\,\text{MeV}$, though we artificially limit the data for the network evaluations to $N_{\text{max}} \leq 12$ to properly test the ANNs, even if data for larger model spaces is available.

Since both networks were designed for the extrapolation of the ground-state energy, we will restrict ourselves to this observable in this section. An expansion towards radii will lead directly to the development of the Gen2 networks discussed in the following section. The results shown in this section mirror the results presented in our 2023 publication [127], though with an updated visualization and improved statistical evaluation as detailed previously in Section 3.3.

As a first step, we can examine the networks' performance by looking at the results obtained on nuclei in the training set, i.e. few-body systems for which we can calculate the fully converged result as a reference, though now for the SMS[450] interaction. The results of such a calculation are shown in Fig. 3.7. As a reference to gauge the convergence, the evaluation data is shown in the

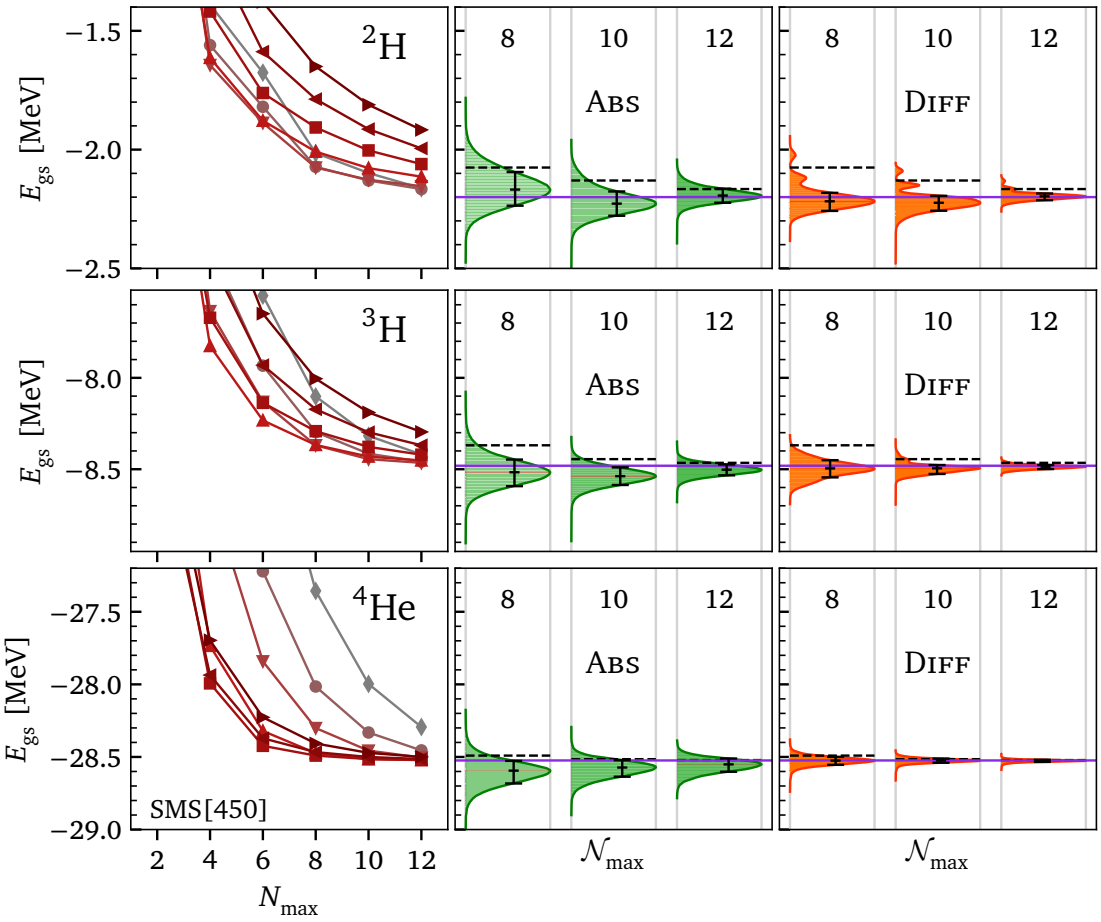


Figure 3.7: Gen1 network results for the Abs and Diff input modes for ^{2}H, ^{3}H, and ^{4}He. The left panels show the evaluation data for $\hbar\Omega$ = 12, 14, 16, 20, 24, 28, and 32 MeV, while the center and right panels show the corresponding statistical evaluation of the networks for Abs and Diff, respectively. The dashed lines indicate the variational minimum at each N_{max}, while the red line shows the converged value E_{gs}^{∞}.

left panels, and the results for the Abs and Diff network evaluations for several $\mathcal{N}_{max}$ steps are shown in the center and right panels, respectively.

Generally, we notice that the predictions get more precise and also more accurate with increasing $\mathcal{N}_{max}$, as one might expect. The distributions get smaller and the peak gets closer to the red line, which indicates the fully converged value for each nucleus for the SMS[450] interaction. The distributions and the extracted predictions and uncertainties are also very well compatible with each other, specifically they are very consistent with increasing $\mathcal{N}_{max}$. There is a lot of overlap between distributions for different model space cutoffs, wich generally points to very stable and robust predictions. Additionally, only very few individual network predictions and none of the extracted final predications or

their uncertainty bands violate the variational principle discussed in Section 1.4.1. That means that no statistical evaluation result is at or above the minimal energy at the highest $N_{\max}$ in the samples used for the evaluation, indicated by the dashed black line.

Looking at the differences between Abs and Diff, we notice that the histograms of the Abs input mode look generally more Gaussian, while the Diff distributions are both more sharply peaked than a normal distribution.There are also some artifacts visible in ^{2}H, likely a result of the rather untypical convergence behavior compared to other nuclei (see Section 1.4.3). In general the Diff input method based networks seem to give significantly more precise results, while also being slightly more accurate. Specifically for ^{3}H and ^{4}He, the Abs networks tend to underpredict the ground-state energy, while the Diff networks are very accurate already at $\mathcal{N}_{\max} = 8$. Though in some cases, e.g. at $\mathcal{N}_{\max} = 10$ for both ^{2}H and ^{3}H, the Diff networks slightly underpredict both the ground-state energy and their own uncertainty, though in general we should point out that this is still very much compatible with the 68 % uncertainty interval we chose for the error bar threshold.

The results for larger nuclei, which are depicted in Fig. 3.8, show a similar picture as the smaller nuclei. These results use smaller and, in the case of ^{12}C, a slightly different $\hbar\Omega$ grid than the training data as well as fewer $N_{\max}$ due to availability. However, this serves as a perfect test case for real world calculations, specifically when performed without an ANN evaluation in mind and one is limited to a certain compute budgets. The Abs networks generally produces less precise final results than the Diff method, though this precision becomes a slight detriment for the latter, since there is a more pronounced shift visible for increasing $\mathcal{N}_{\max}$, which is especially pronounced in ^{6}Li and ^{16}O. Here, the Diff method seems to slightly underestimate its uncertainty, since the results for larger $\mathcal{N}_{\max}$ do not overlap with the results for previous $\mathcal{N}_{\max}$. For the Abs method on the other hand, these shifts with increasing $\mathcal{N}_{\max}$ are less pronounced, and the overlap with distributions from smaller $\mathcal{N}_{\max}$ is additionally increased by the larger, and in these cases more realistic and appropriate, uncertainty estimates. However, we also notice that the prediction for ^{12}C at $\mathcal{N}_{\max} = 10$ violates the variational principle. This seems to be part of a general trend that sees the results of the Abs method to be slightly higher than those from the Diff method.

In general, while the Gen1 networks provide good predictions for ground-state energy and compare favorably to the exponential extrapolation method as we demonstrated in Ref. [127], the discussed shortcomings of both Abs and Diff ANNs together with additional challenges arising when trying to extend this

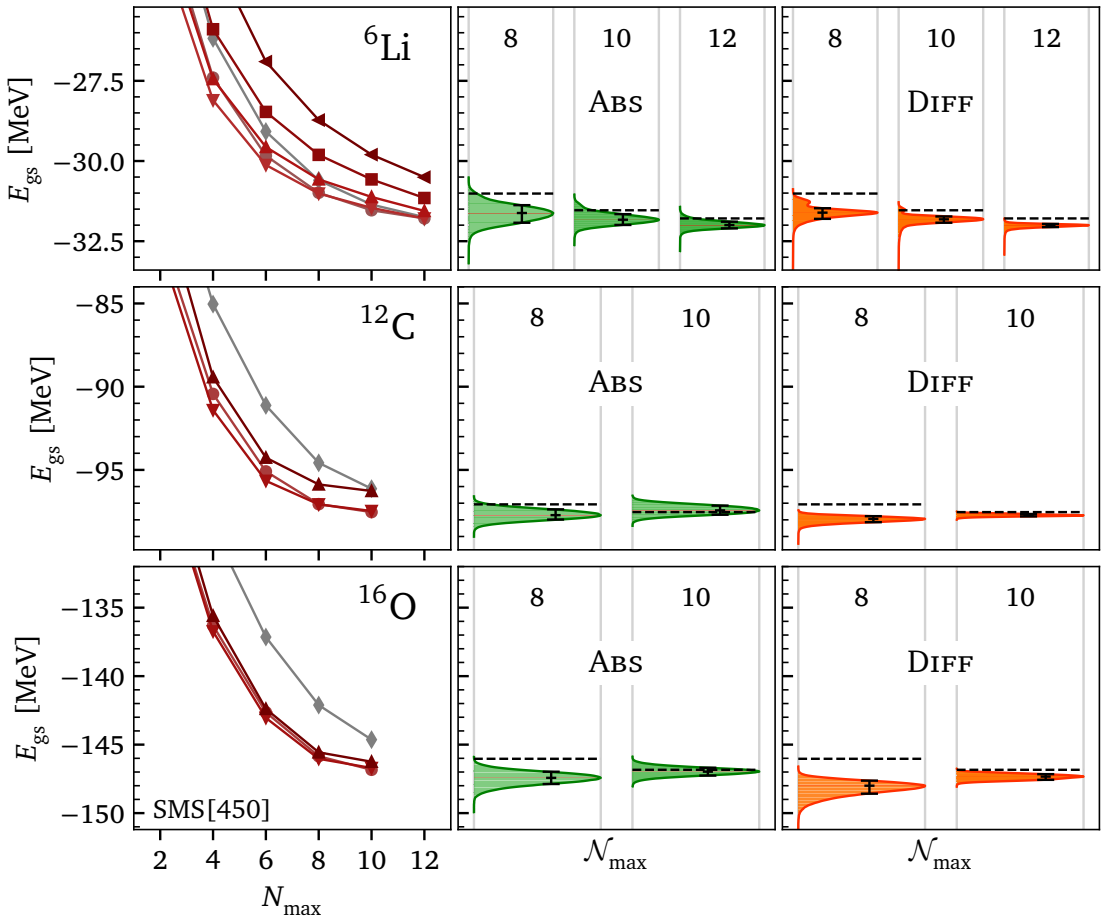


Figure 3.8: Similar to Fig. 3.7, but for ^{6}Li, ^{12}C, and ^{16}O. The available HO frequencies are $\hbar\Omega = 14,\ 16,\ 20,\ 24,\ 28,\ 32\,\mathrm{MeV}$ for ^{6}Li, $\hbar\Omega = 16,\ 20,\ 24,\ 28\,\mathrm{MeV}$ for ^{12}C, and $\hbar\Omega = 16,\ 20,\ 22,\ 24\,\mathrm{MeV}$ for ^{16}O.

approach to radii directly led to the development of the Gen2 networks, which we will discuss in the next chapter.

4

Improved Artificial Neural Networks for Energies and Radii

After developing the Gen1 networks for ground-state energy predictions based on NCSM data, a natural next step is to extend this ANN-based approach to radii. Unfortunately, the Gen1 networks do not perform well with this new observable, as we will discover in this chapter. While the range of possible values is much more constrained than, e.g. the ground-state energy for different nuclei, the same is not true for the convergence pattern itself. The lack of the monotonous convergence constraint also raises questions on how to choose the reference value the Diff networks are trained on, since the lowest value, as chosen for energies, is in general not the best converged result in an arbitrary sample. Furthermore, we want to improve on the various shortcomings of the Gen1 networks highlighted in the last chapter.

To address these issues, we developed a second generation of our artificial neural networks. The basic concept remains the same and almost all of the Abs network architecture and topology will be used, though with a new input mode which normalizes the samples before feeding them to the network and slight adjustments to some hyperparameters. Input normalization is a fairly common technique in neural network applications, however, it was not our first choice initially, since it removes the ability of our random scaling and shifting method to increase the amount of training data almost arbitrarily. However, the clear advantage is that the network only needs to learn pattern recognition in a constrained, limited scale, while the range of its inputs can vary wildly. This led us directly to a new input mode called Minmax, which significantly improves

the networks' performance compared to the ABS/DIFF networks, especially for radius predictions.

4.1 Radius Training Data

The training data for the Gen2 networks for the ground-state energy remains the same data that was used for the Gen1 networks, which was discussed in Section 3.2. Since our main motivation to develop a second generation of neural networks was the prediction of nuclear radii, we obviously need training data for this observable. Analogous to the energy training data, we need NCSM radius calculations for which we can also compute the fully-converged radius. Therefore, we are limited to few-body systems up to ^{4}He calculated in the Jacobi-NCSM. Furthermore, we again only include data for one family of chiral interactions in the training data, so that we can utilize the other one to evaluate the performance as independently as possible, even for nuclei that are part of the training set.

So, completely analogous to the energy training data, we use Jacobi-NCSM calculations for ^{2}H, ^{3}H, and ^{4}He, performed with the chiral NN+3N non-local interaction at N^{2}LO, N^{3}LO, and N^{4}LO$'$ for cutoffs $\Lambda = 450,\ 500$, and 550 MeV, which are used bare or SRG transformed with 3 different flow parameters $s = 0.02,\ 0.04$, and 0.08 fm^4. The calculations were performed for 7 HO basis parameters $\hbar\Omega = 12,\ 14,\ 16,\ 20,\ 24,\ 28$, and 32 MeV. Future networks might benefit from using an HO osczillator length a_{HO} grid instead, as it produces more equally spaced convergence sequences, though we did not observe any obvious deficiencies based on this choice to stick to the $\hbar\Omega$ grid for the radius training data.

Since the training data is calculated in the Jacobi-NCSM, where the isospin of individual nucleons is not tracked, point-proton and point-neutron rms radii cannot be calculated exactly. Fortunately, the convergence behavior of these compared to the rms radii we can calculate for training is so similar, that networks trained on one can easily be applied to data from other radius types. Therefore, calculating only mass rms radii and training one set of networks on them is enough to also cover point-proton rms and point-neutron rms radii. Similarly, the effect of the SRG correction to the radius operator is so small, that it too does not significantly alter the values or convergence pattern, so networks trained on data calculated without SRG correction can simply be applied to data which includes this correction. This massively simplifies the training requirements, and

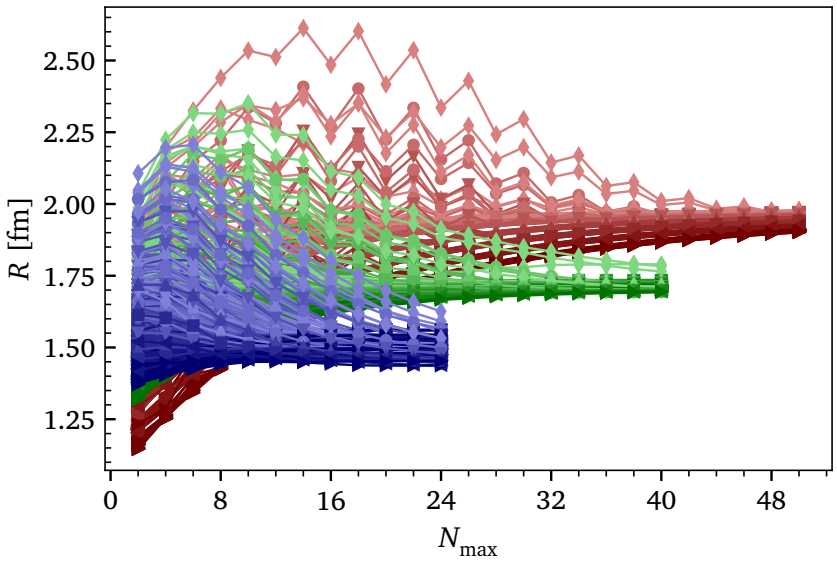


Figure 4.1: Plot of the complete training data for the radius networks comprised of NCSM calculations of ^{2}H (red), ^{3}H (green), and ^{4}He (blue) for 36 different versions of the non-local interaction.

allows us to train just one set of universal radius prediction networks to cover all radius observable data from the NCSM.

All Jacobi-NCSM results included in the training data are depicted in Fig. 4.1. The three different nuclei, separated by color, exhibit very similar convergence patterns at this scale. Though it has to be pointed out that only the ^{2}H data, especially for small $\hbar\Omega$, exhibits the typical zigzag behavior discussed in Section 1.4.3. Also visible in the plot is that the ^{2}H and ^{3}H data predominantly converges from below, while the ^{4}He data shows a more mixed convergence direction. Analogous to the energy training data, the data available for training is limited to $N_{\max} = 24$ for ^{4}He, $N_{\max} = 40$ for ^{3}H, and $N_{\max} = 50$ for ^{2}H. This results in a total of 14 364 NCSM data points usable for training the ANNs, split between 6300 for ^{2}H, 5040 for ^{3}H, and 3024 for ^{4}He, exactly the same as for the energy training data.

4.2 MINMAX Input Mode

Normalizing the input values, also referred to as feature scaling, is a very common technique in neural networks applications [43, 133]. There are several different standard methods to achieve an optimal normalization over a wide range of inputs. We chose to employ the min-max normalization (MINMAX), which scales and shifts every sample so all its values are in the interval $[a, b]$. Specifically, the

minimum value in the sample is transformed to be a and the maximum value to be b. For our use-case we set $a = 0$, $b = 1$ and therefore an interval of $[0, 1]$. While it seems redundant at first since multi-layer ANNs can in principle incorporate this linear transformation into the first layer, it aids in keeping the weights of natural size [43]. This makes a significant difference, as the initialization of the network parameters before training is much closer to the optimal values, greatly improving training results. We should note that we also tried other common normalization techniques, since the optimal normalization method is often domain and problem specific. For instance, we also tried a method called standardization or Z-score normalization, which transforms the data such that the mean is zero and it has unit variance or standard deviation [43]. However, this and other normalization techniques performed consistently worse than MINMAX which is why we focused our efforts on the latter.

For our networks, we retain the core concept of feeding in sequences for $n_{\text{seq}} = 3$ different HO basis parameter, which each contain $l_{\text{seq}} = 4$ value of the observable for subsequent N_{max} values. A MINMAX-normalized sample can therefore simply be described as an ABS sample (see Eq. (3.1)) that is shifted and scaled to the interval $[0, 1]$ as follows:

$$S_{\text{MINMAX}}^{\mathcal{N}_{\text{max}}} = \frac{S_{\text{ABS}}^{\mathcal{N}_{\text{max}}} - \min\left(S_{\text{ABS}}^{\mathcal{N}_{\text{max}}}\right)}{\max\left(S_{\text{ABS}}^{\mathcal{N}_{\text{max}}}\right) - \min\left(S_{\text{ABS}}^{\mathcal{N}_{\text{max}}}\right)}. \tag{4.1}$$

The converged value O^{∞} for the training samples are converted alongside their respective sample via the same transformation to serve as labels for the network training. To recover the actual value of the observable from a network prediction, we simply apply the inverse transformation

$$O^{\infty} = O_{\text{MINMAX}}^{\infty} \cdot \left(\max\left(S_{\text{ABS}}^{\mathcal{N}_{\text{max}}}\right) - \min\left(S_{\text{ABS}}^{\mathcal{N}_{\text{max}}}\right)\right) + \min\left(S_{\text{ABS}}^{\mathcal{N}_{\text{max}}}\right). \tag{4.2}$$

This transformation is independent of the label and therefore works exactly the same for evaluation data. Furthermore, the normalization works independent of the observable, so all NCSM data, whether it is for energies or radii, can be transformed and compressed to the unit interval using this technique. In practice, we simply store the shift value $\min\left(S_{\text{ABS}}^{\mathcal{N}_{\text{max}}}\right)$ and scaling value $\left[\max\left(S_{\text{ABS}}^{\mathcal{N}_{\text{max}}}\right) - \min\left(S_{\text{ABS}}^{\mathcal{N}_{\text{max}}}\right)\right]^{-1}$ for each sample to revert the transformation of each network prediction after the evaluation in bulk.

Given the transformation in Eq. (4.1), it is easy to see that the previously applied random scaling and shifting of each sample to increase the amount of training data would simply be reverted by the MINMAX normalization. Therefore,

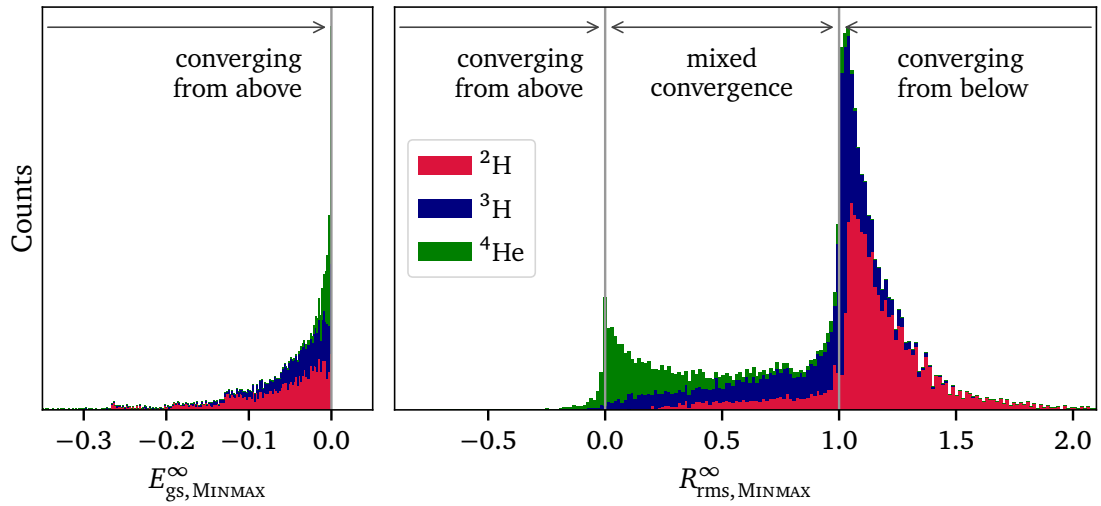


Figure 4.2: Stacked histogram of MINMAX-normalized target values for the ground-state energy (left panel) and the mass radius (right panel) of all samples in the training data, separated by nucleus.

in contrast to ABS and DIFF, we are limited to the roughly 360 000 native samples generated from the training data. However, due to their scale invariance, the MINMAX samples are more effective for the training of the neural networks, and the training data seems to be sufficient to arrive at well-behaved networks.

Similar to the ABS and DIFF samples, we can visualize the distribution of the ground-state energy target values of the normalized MINMAX samples, which are shown in the left panel of Fig. 4.2. The distribution looks overall very similar to the DIFF distribution (compare Fig. 3.6), though not as sharply peaked at zero. Due to the monotonicity constraint of the energy convergence as a function of N_{max}, all target values are always strictly less than zero. Also, similar to the DIFF target values, there is no clustering of the data corresponding to the three source nuclei, which demonstrates the universality and scale invariance achieved with the normalization in the MINMAX input method.

With energies out of the way, we can shift our attention to radii. Here, as discussed earlier in Section 1.4.2, the convergence behavior is a lot less constrained and therefore more challenging. Since convergence is no longer monotonous, samples can now also converge from below, or change their direction during convergence, some sequences even follow a zigzag pattern before eventually converging. For both the ABS and the DIFF method, these different convergence patterns can pose challenges, which are not present in the MINMAX input mode. The normalization works exactly the same as for the energies, as shown in Eq. (4.1). However, while the sample continues to be strictly within the interval $[0, 1]$, due to the varied convergence behaviors and directions, the target

value can now be any value, though in practice it remains of natural size. This is demonstrated in the right panel of Fig. 4.2, which we can divide into three regions. If the target value is below the lowest value in the sample, it falls into the region less than zero and the convergence is therefore from above, similar to the energy convergence. However, for radii the target value can also be larger than the largest value in the sample, indicating that the sample converges from below with a target value greater than one. As a third option, the target value can be between the minimum and maximum value in the sample, which we call mixed convergence, leading to a target value between zero and one. According to Fig. 4.2, a majority of the samples generated from the training data converges from below. This is caused by the HO frequency selection used for the training data Jacobi-NCSM calculations. In fact, only ^{4}He provides some samples which converge completely or mainly from above, due to its rather compact radius compared to ^{3}H and especially ^{2}H. The imbalance of training data towards samples that contain sequences which converge from below compared to the other two convergence patterns is not necessarily intrinsically worse. While networks trained on this data tend to produce more accurate and precise predictions, as we will see and utilize later in this chapter, it is not clear whether that is due to the bias in the training data or because of the much more structured convergence behavior of sequences converging from below.

4.3 Sample Universality

Before moving on to training new networks, we are now, with the help of the MINMAX normalization, in a position to answer a fundamental question about the generality of the training data. The assumption that underpins our ANN method is whether the samples generated from the training data from ^{2}H, ^{3}H, and ^{4}He are representative of the evaluation data obtained in significantly heavier nuclei. So far, we have not demonstrated this to be accurate. Another way to phrase this question is whether the network prediction is truly, as we have claimed so far, an interpolation between convergence patterns already known from the training data, or if the evaluation samples are too different from the known training data and there is a significant amount of extrapolation necessary on behalf of the networks. Because the MINMAX normalization compresses all samples into the $[0, 1]$ interval, we can now look at an arbitrary sample from the evaluation data of a much heavier nucleus than the training data, and see if we can find samples

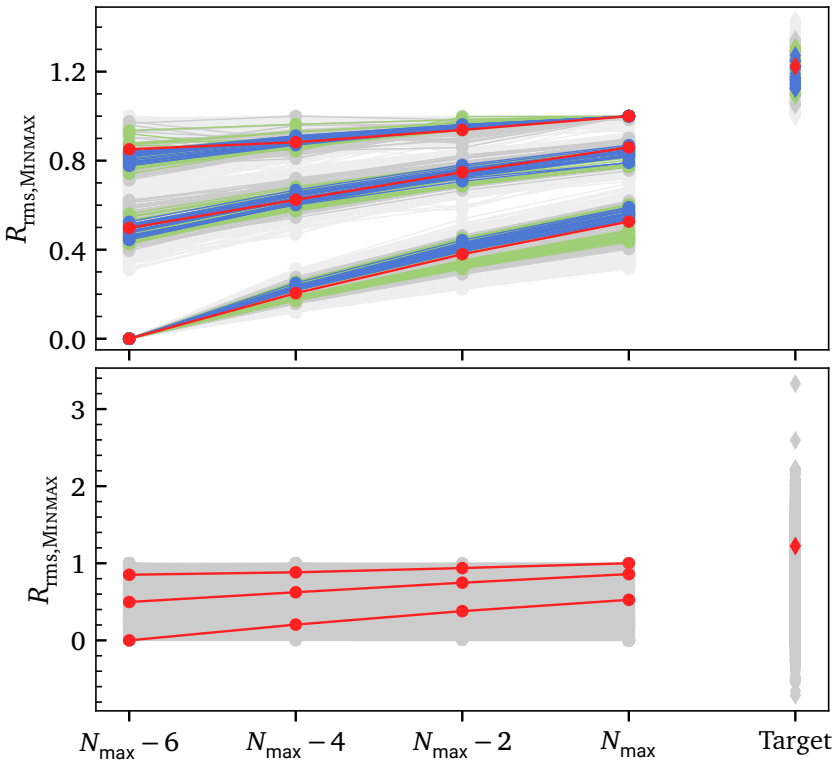


Figure 4.3: Illustration of MinMax, i.e. normalized and therefore dimensionless, radius samples. In red is a ^{9}Be evaluation sample together with the networks' prediction. The upper panel shows training samples that are the 10 (blue), 50 (green), 100 (gray), and 500 (light gray) most similar samples from the training set, including their respective target value the networks are trained on. The lower panel shows 10 000 random training samples and their targets in gray, demonstrating the wide range of different sample and target combinations in the training set.

that are similar in the training data. For simplicity, we define similarity with the help of the standard Euclidean norm of the difference between the evaluation from a heavier nucleus and each training sample. The upper panel of Fig. 4.3 illustrates this for an arbitrary evaluation sample for the radius of ^{9}Be shown in red, where the target value is the prediction of the networks. The blue lines show the 10 closest samples from the training data, with the other colors casting a wider and wider net around the ^{9}Be sample. Many training samples are indeed very close to the evaluation sample, and the target value of the training samples is also very similar to the network prediction based on the evaluation sample. This clearly demonstrates that there is sufficiently similar training data to justify the claim that the networks mainly interpolate between already seen convergence patterns and are not forced to extrapolate towards a parameter space range not seen during training.

The lower panel of Fig. 4.3 once again shows the ^{9}Be evaluation sample in red, though it now also depicts 10 000 randomly selected training samples in gray. These samples fill the full range from zero to one in the plot, with the point being that any reasonable convergence is almost guaranteed to be fairly closely represented in the training data. Indeed, we can find similar samples as demonstrated in the upper panel for other evaluation samples, not just the one shown and irrespective of the nucleus.This indicates a certain generality of the training data, even though it just contains data from few-body systems, which seem to encode much of the convergence behavior in the NCSM generalizable to larger nuclei. However, this is not to say that our current training data set is optimal. Further research could almost certainly yield improvements by optimizing the training set. Though it is encouraging that even with the naïve selection, we obtained a rather universal and productive selection of training data.

4.4 Second Generation Networks

The new MinMax input mode does, of course, require a review of the hyperparameters of the networks. After optimization, similar to the steps described in Section 3.4.1, we concluded that the hyperparameters used in the Abs networks also work very well for the MinMax input mode. Specifically the topology investigation outlined in that section, both regarding the input layer dimension and shape as well as the hidden layer count and dimensions transfer nearly one-to-one from the Abs network optimization. Due to this similarity, we omit a more detailed discussion of the network topology and other hyperparameters and instead refer to the relevant sections for the Gen1 networks. For a quick overview, Table 4.1 lists all relevant hyperparameters used in the Gen2 networks for both the energy and the radius variant of the networks.

Having fixed the network topology and other hyperparameters, we can train a new set of networks for energy and radius using the training data outlined in Sections 3.2 and 4.1. We retain the statistical evaluation approach from the Gen1 networks, meaning we train 1000 MinMax networks for both observables. To evaluate the resulting networks, we can compare them to both Abs and Diff networks, for which we also trained 1000 networks for the radius.

Starting with ground-state energies, Fig. 4.4 shows the results of the new MinMax networks in the right column, while the results from the Gen1 Abs and Diff

Hyperparameter	Gen2 for energies	Gen2 for radii
input mode	MINMAX	
input layer dimension	12	
hidden layer dimensions	48, 48, 24	
output layer dimension	1	
activation function	ReLU	
loss function	MSE	
optimizer	AdamW	
batch size	512	
learning rate scheduler	ReduceOnPlateau	
initial learning rate	0.01	
epochs	20	
threshold	0.05	0.1
weight/bias initialization	random from $\left[-\frac{1}{\sqrt{\#\text{inputs}}}, \frac{1}{\sqrt{\#\text{inputs}}}\right]$	

Table 4.1: Summary of the main hyperparameters used for the Gen2 networks.

networks is shown in the left and center panel, respectively. The histograms for ABS and DIFF are the same as the ones shown in Fig. 3.7, though we omit the evaluation data for brevity. We observe that MINMAX significantly outperforms ABS in both accuracy and precision for all three nuclei. Even for the smallest $\mathcal{N}_{\text{max}} = 8$, the MINMAX predictions are already extremely close to the fully-converged values. Compared to DIFF, the MINMAX networks perform in general very similarly, with almost identical predictions for the converged values, though with generally slightly larger uncertainties. However, since the DIFF uncertainties tend to be a bit to optimistic as concluded previously in Section 3.5, this is not necessarily a disadvantage. The comparison also reveals similar, though noticeably less pronounced artifacts, e.g. in ^{2}H at $\mathcal{N}_{\text{max}} = 10$. Overall, the MINMAX networks perform extremely well, giving accurate predictions for the converged value even from calculations in rather limited model spaces. Additionally, the predictions converge for increasing $\mathcal{N}_{\text{max}}$ as expected, and predictions for different $\mathcal{N}_{\text{max}}$ are very compatible with each other.

Exploring the performance of all three networks on radii is the next step. To obtain ABS and DIFF networks, we performed the training as outlined in Section 3.4.1 using the same hyperparameters as listed in Table 3.1, though with a slightly different validation threshold of 0.01 fm (ABS) and 0.001 fm as well as a modified range of $[-1, 1]$ for the random shifts of the ABS samples, which is more appropriate for the range of expected radius values. This results in the predictions for the rms radius depicted in Fig. 4.5, where the left panels once

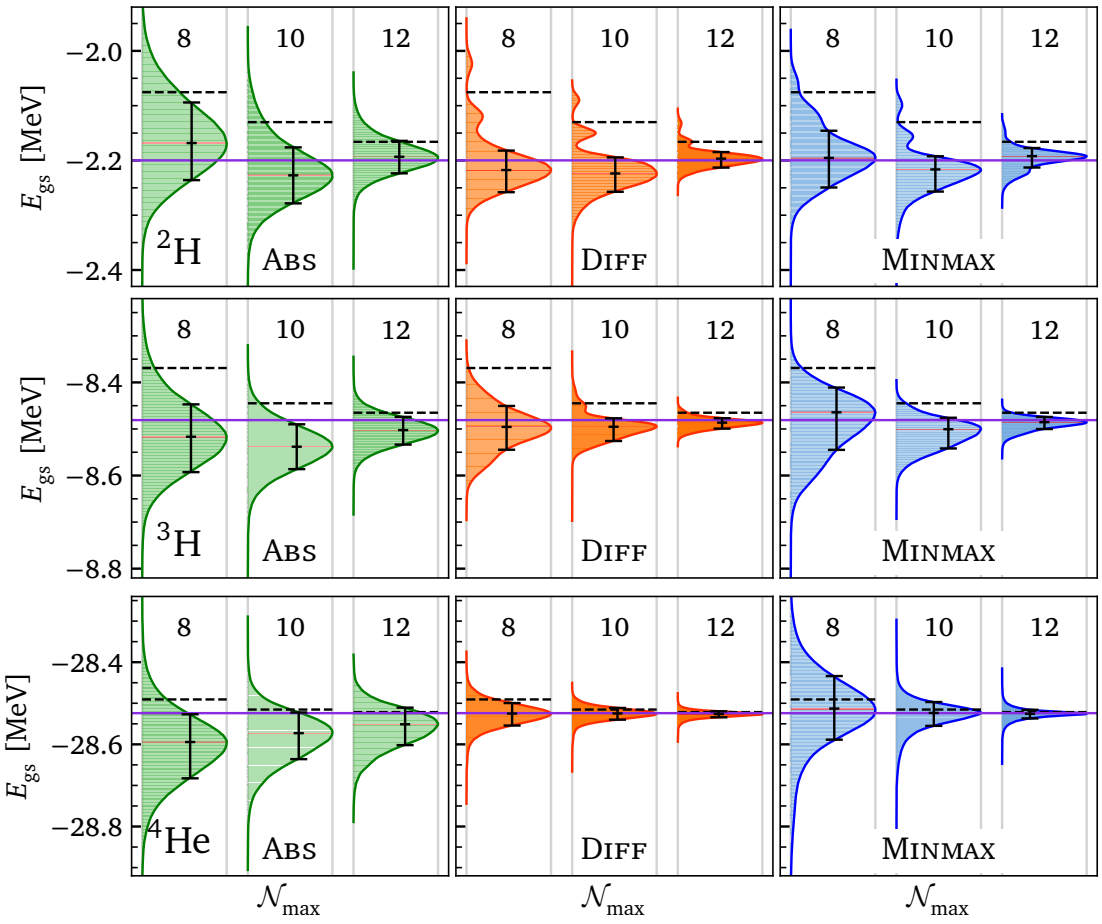


Figure 4.4: Comparison of the Gen1 networks using the Abs and Diff input modes and the Gen2 networks using the MinMax input normalization for the ground-state energy of ^{2}H, ^{3}H, and ^{4}He. The evaluation data is the same as in Fig. 3.7.

again show the histograms at various $\mathcal{N}_{max}$ for Abs, with the Diff results in the center. Even though these nuclei, albeit with a different interaction, are part of the training data, both the Abs and the Diff networks struggle to accurately predict the fully converged value, indicated by the purple horizontal line. While Abs works reasonably well for predicting the converged value for ^{4}He, both ^{2}H and ^{3}H are not described well at all, with five out of six predictions not containing the fully converged value within their uncertainties. The situation is worse for Diff, where the predictions are even farther off of the target value. The Gen2 MinMax networks on the other hand, shown in the right column, do a much better job at predicting the converged radius, only struggling a bit with ^{3}H, where a slight systematic downward trend is visible for increasing $\mathcal{N}_{max}$, converging onto the target value. ^{2}H is predicted very consistently, and ^{4}He is described nearly perfectly.

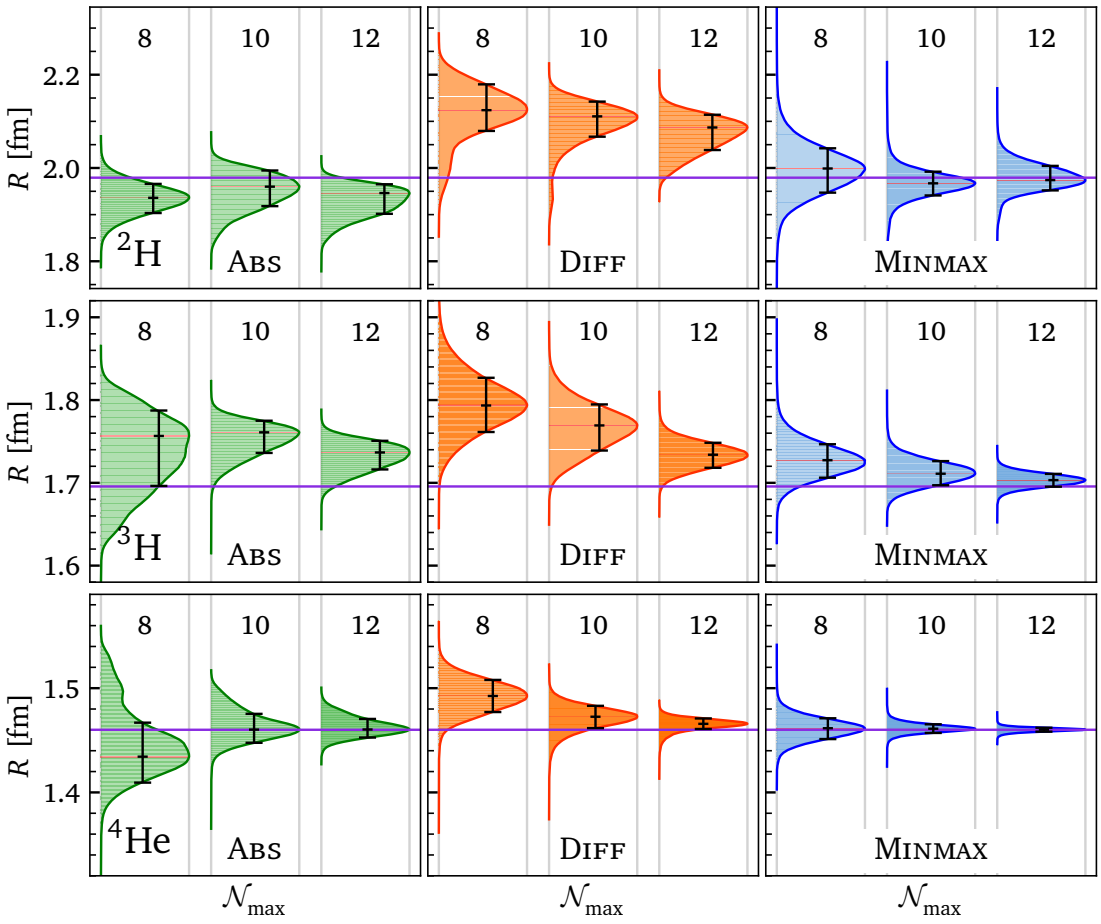


Figure 4.5: Comparison of the Gen1 networks using the Abs and Diff input modes and the Gen2 networks using the MinMax input normalization for the rms radius of ^{2}H, ^{3}H, and ^{4}He.

Altogether, these results show the power of the MinMax input mode and associated networks. While energies are still predicted very accurately by these networks, their true advantage over the previous Gen1 Abs and Diff networks is clearly the far superior description of the rms radius and, due to very similar convergence, radii in general. Going forward, all predictions for both energies and radii will be based on these Gen2 networks.

4.5 Evaluation Data Selection

Another important and interesting aspect to explore is where the uncertainties of the final network predictions actually originate. For the statistical evaluation, there are two sources that determine the width of the resulting distribution and

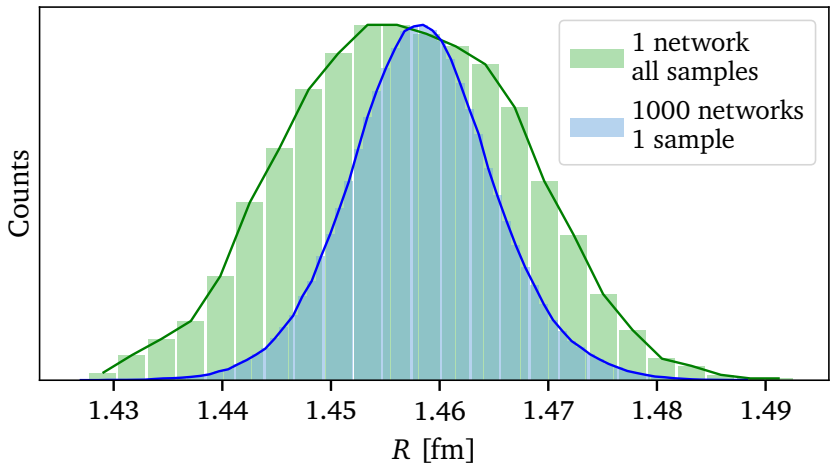


Figure 4.6: Histograms resulting from evaluating a single network with all samples (green), and evaluating all 1000 networks with a single sample (blue) for ^{4}He.

therefore the size of the uncertainties. One is the evaluation of 1000 different networks, each trained with different weight and bias initializations, resulting in slightly different predictions. The other is the difference between input samples the networks are evaluated on. These samples contain sequences from different HO parameters, heavily influencing the convergence and therefore the network input values. Naturally, one might expect these two effects to be of roughly comparable size, but this is not the case. Figure 4.6 demonstrates this for ^{4}He. Here, evaluating a single network for all samples ($\hbar\Omega = 12,\ 14,\ 16,\ 20,\ 24,\ 28,\ 32\,\mathrm{MeV}$) shown in green results in a significantly wider histogram than evaluating all 1000 networks on just a single sample (in this example $\hbar\Omega = 16,\ 20,\ 24\,\mathrm{MeV}$) shown in blue. This points to the fact that not all samples are equally useful for the networks in predicting the converged value, and precision and possibly also accuracy could be improved with a clever selection of the best-suited evaluation samples. We can test this hypothesis by artificially down-selecting the evaluation data. We find that we can indeed improve the prediction accuracy and precision by selecting an optimal subset of the evaluation data according to the following recipe. To distinguish predictions with all available evaluation data and this new optimized sequence selection, we will refer to the latter as second generation with sequence selection (Gen2+).

For energies, we select five sequences with the lowest energy at the largest N_{max} in the evaluation range. This typically means we select the variational minimum and two HO basis parameters to either side, though a slightly asymmetric selection around the variational minimum is also not uncommon. This choice should result in the best-converging sequences for the networks to base their pre-

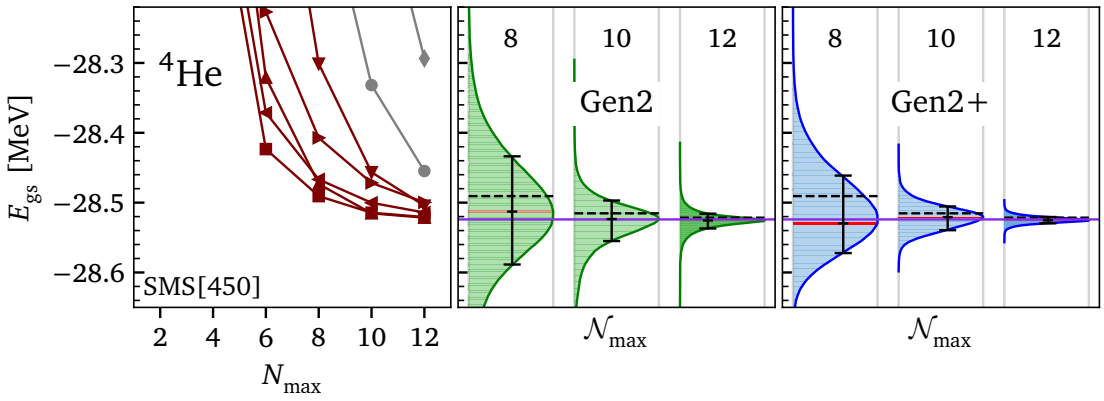


Figure 4.7: Comparison of network evaluations of the ground-state energy of ^{4}He using our Gen2 networks with all evaluation data (center panel), and with an optimized selection of sequences (right panel). The left panel shows all evaluation sequences, with the ones in the optimized selection highlighted red.

dictions on. A comparison between predictions without and with the optimized sequence selection is shown in Fig. 4.7 for ^{4}He. Here, performing the network evaluation with only a subset of the available evaluation data mainly leads to smaller uncertainties, while keeping the predicted values virtually unchanged. However, we should note that fewer network predictions violate the variational minimum at each $\mathcal{N}_{max}$, indicated by the dashed black lines. In our experience this evaluation data selection noticeably improves the final predictions obtained from the networks, especially when the number of sequences increases. We will show more examples comparing Gen2 and Gen2+ results in the next chapter.

For radii, we select the five sequences resulting in the largest radius that are still converging monotonously from below. We can justify this choice by referring back to Fig. 4.2 of Section 4.2, which clearly shows that sequences converging from below are significantly overrepresented compared to mixed convergence and especially convergence from above, thereby optimizing the selection towards data more similar to the training data. However, there is also a physical motivation for this choice. As we explored in Section 1.4.2, sequences converging from below generally exhibit a more consistent and predictable convergence behavior that other sequences. It is therefore reasonable to exploit this fact not only by training the networks predominantly with such data, but also select the evaluation data accordingly. In practice this usually results in an a_{HO} selection that is a bit lower (higher for $\hbar\Omega$) than one might typically calculate, though typically not too extreme, i.e. too far away from the variational minimum of the ground-state energy. The effect of this downselection on radii predictions is shown exemplary in Fig. 4.8 for the point-proton rmsradius of ^{8}Li. Similarly

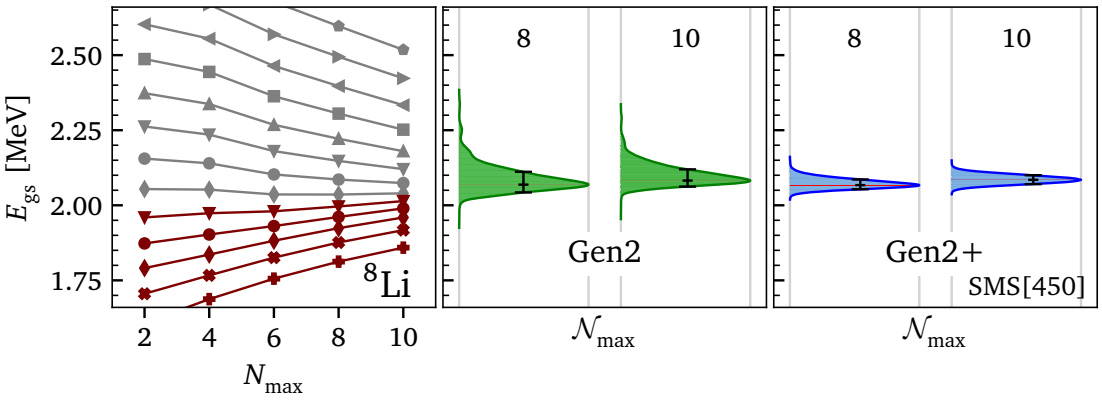


Figure 4.8: Analogous to Fig. 4.7 but for the point-proton rms radius of ^{8}Li.

to the ground-state energies, this generally leads to sharper distributions and therefore smaller uncertainties. It also eliminated many of the artifacts some prediction distributions exhibit and leads to distributions that are more Gaussian. This is clearly visible when comparing the distributions in the center panel with the ones in the right panel. While the Gen2 predictions are almost triangle shaped with some minor artifacts at larger radii, the predictions for Gen2+ are more Gaussian and do not exhibit any artifacts.

In conclusion, we are able to significantly improve the predictions of the MINMAX networks with the described selection of evaluation data. We will refer to this combination of the improved Gen2 networks with this evaluation recipe as Gen2+ and use it as the standard method throughout later chapters. The decision of using Gen2+ as the standard evaluation scheme is explored further in the next chapter, which offers additional comparisons and benchmarks between Gen2 and Gen2+ as well as other model-space extrapolation methods.

5

Comparison with Other Extrapolation Methods

In the preceding chapters, we detailed the development of our ANN-based extrapolation framework, culminating in the Gen2+ architecture and its associated evaluation scheme. We also presented representative applications of this method, demonstrating its ability to reliably infer fully-converged observables from partially converged NCSM calculations, both for few-body systems included in the training data and for heavier nuclei beyond it. A critical next step is a systematic comparison of our approach with established model-space extrapolation techniques, as well as with a competing neural-network–based method.

For this comparison, we look at the ground-state energy as well as the point-proton charge radius of ^{6}Li, ^{7}Li, and ^{8}Li. This allows for a detailed assessment of the performance, stability, and uncertainty quantification of a range of extrapolation methods, while simultaneously serving as a validation of our ANN-based approach. The analysis presented here closely follows Ref. [42], which was carried out in collaboration with the Iowa State University (ISU) group.

5.1 Prerequisites

The NCSM calculations for this comparison use the Daejeon16 NN potential [134], a completely different interaction than the chiral NN+3N interactions used for the training of our ANNs. Utilizing this interaction offers several advantages. First, since it is only a two-body interaction, it is significantly cheaper to calculate,

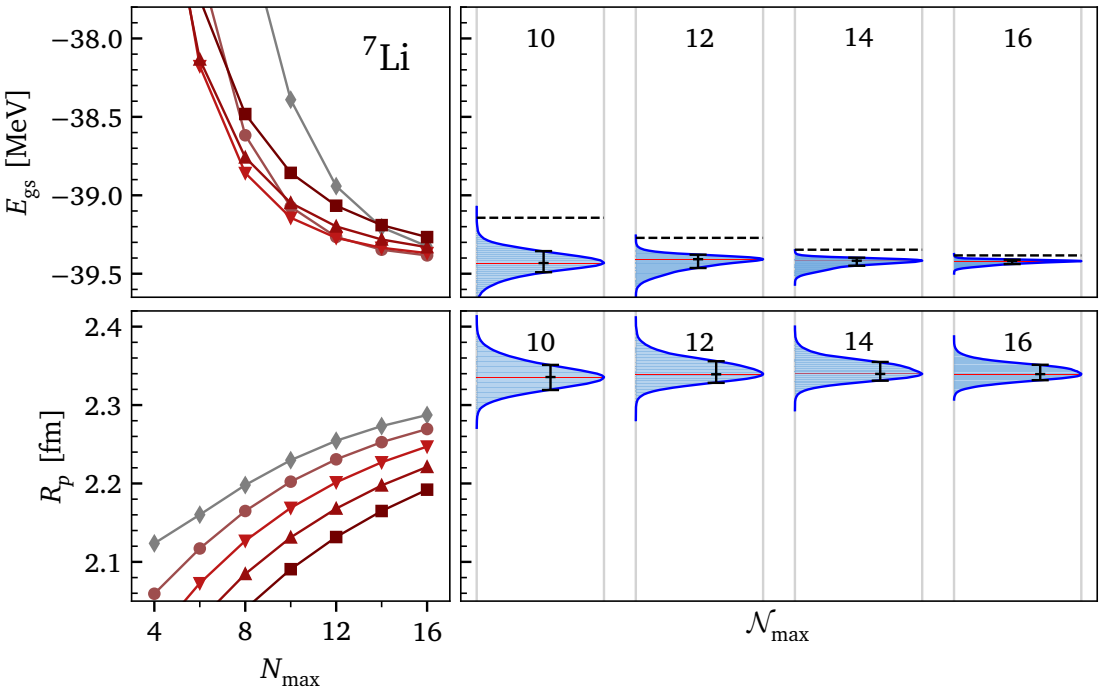


Figure 5.1: Evaluation using our Gen2+ networks for the ground-state energy and point-proton rms radius of ^{7}Li.

increasing the achievable range of N_{max}. The rather soft nature of the interaction further aids in reducing computational cost. Additionally, this interaction is completely separate from the training and even evaluation data we used so far to evaluate our own ANNs, so these calculations offer a completely new window into their behavior and should provide a good test of our claim of universality of our network approach.

The Daejeon16 interaction is essentially a modified and heavily SRG evolved version of the N^{3}LO NN Idaho (Entem-Machleidt) interaction [66, 67], with a flow parameter of $\lambda = 1.5\,\mathrm{fm}^{-1}$. Hence, this interaction is very soft and significantly accelerates the convergence of many-body methods such as the NCSM, which allows for larger model-space calculations. This resulted in NCSM calculations spanning a grid of $\hbar\Omega = 10$ to 30 MeV with a step size of 2.5 MeV for N_{max} up to 18 (^{6}Li), 16 (^{7}Li), and 14 (^{8}Li) for both the ground-state energy and the point-proton rms radius. Initially, a larger $\hbar\Omega$ set with more extreme values was calculated, however, due to computational limitations these are not available for all cases. So, to keep the evaluation data consistent across all nuclei and N_{max}, we reduced the grid to the above-mentioned $\hbar\Omega$ values, which also reflect a typical range of basis parameters for NCSM calculations. The evaluation based on this data for ^{7}Li using our Gen2+ networks, which selects only the optimal

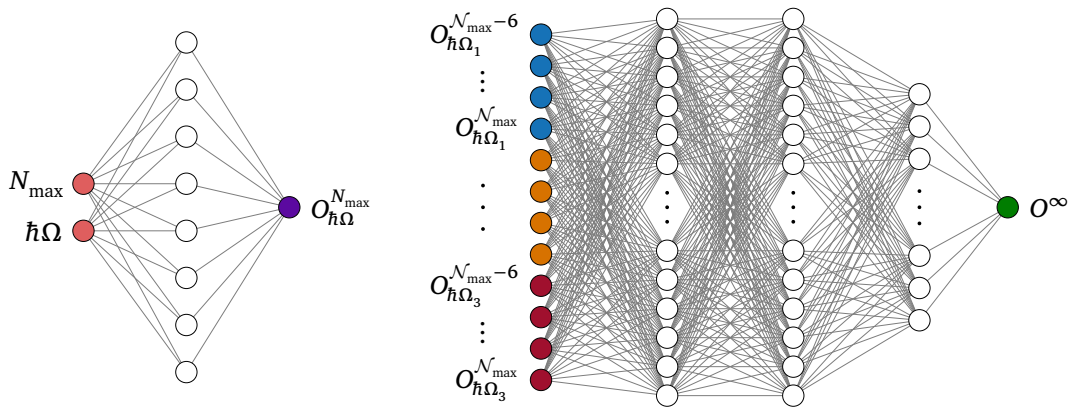
$\hbar\Omega$ sequences for evaluation, is shown in Fig. 5.1 for both the ground-state energy and the point-proton rms radius. Both the energy and radius results are remarkably stable across $\mathcal{N}_{\text{max}}$ and nicely converge and increase in precision, and the distributions show no significant artifacts. Similar results are obtained for the other nuclei in the evaluation set. This also underpins our claim that networks trained on mass rms radius data can simply be applied as-is to other radius observables, such as the point-proton rms radius due to the extremely similar convergence behaviors.

For the benchmark, we provide results for both the Gen2, which uses all available evaluation data, and the improved Gen2+ variant, which downselects the data. The methods we compare our ANNs with are classical, more or less established extrapolation methods. For the ground-state energy, we compare with the exponential extrapolation technique, and UV/IR-based extrapolations. Since radii convergence is less constrained, there is no well established method, which is why we compare with the rather simple and heuristic crossing point method. Each of these methods is discussed briefly in Section 1.5. Additionally, we also benchmark our networks against a completely different neural network based approach, which will be shortly described in the next section, before looking at the results of the benchmarks.

5.2 The ISU Neural Network Approach

This neural network based approach that is conceptually very different from our networks was developed at the ISU and first described in Refs. [124, 125]. For this comparison, a slightly updated version outlined in Ref. [135] is used. To easily tell our ANN approach and this one apart, we simply refer to it as the ISU ANNs.

This technique trains fully-connected neural networks with a single hidden layer with 8 neurons to match the inputs $(\hbar\Omega, N_{\text{max}})$ with the output $O_{\hbar\Omega}^{N_{\text{max}}}$, i.e. the value of the observable calculated in the NCSM with the parameters $\hbar\Omega$ and N_{max}. The input and output as well as the topology of the ISU ANNs are schematically depicted in Fig. 5.2, and our Gen2 networks are shown on the right for comparison. It is important to note the vastly different ansatz and therefore working principle of the ISU ANN compared to our method. The goal of the ISU ANNs is to learn the convergence pattern for one specific nucleus, interaction, state, and observable as a function of N_{max} and $\hbar\Omega$. What we referred to as evaluation data are technically



ISU ANN topology | Our Gen2 ANN topology

Figure 5.2: Schematic illustration of the topology of the ISU ANNs and our Gen2 ANNs. It is important to note that the input and output values and therefore the underlying working principle of both approaches are vastly different.

the training data for the ISU networks. Essentially, these networks produce a fit of a two-dimensional surface describing the value of the observable in the $N_{\max}$ and $\hbar\Omega$ space from the few data points calculated in the NCSM. The trained networks are then evaluated at significantly larger $N_{\max}$ values than the ones in the training data, which forces the ANNs to extrapolate in the $N_{\max}$ dimension to estimate the converged value of the observable. While these networks can be applied to new observables fairly easily, it is important to note that these networks are specific to each nucleus, interaction, state, and observable. Hence, new networks must be trained for each evaluation, increasing computational cost while also potentially raising the requirements on the minimum number of NCSM data available for the adequate training of the networks. Similarly to our approach, multiple networks with different initializations are trained, in this case 200 000, from which the 50 best performing networks are selected for the evaluation. The evaluation works by simply evaluating the networks at a large $N_{\max}$ value ($N_{\max} = 70$ for the ground-state energy and $N_{\max} = 90$ for the radius) for all HO basis parameters in the evaluation data. Even for very light nuclei, these $N_{\max}$ values are several multiples of the largest $N_{\max}$ achievable in NCSM calculations and therefore potentially incur a significant extrapolation error. Therefore, one criterion to select the best performing networks is to require that the output remains approximately constant for all HO frequencies at the large evaluation $N_{\max}$ value. For more details on the training and selection criteria, see Refs. [125, 135].

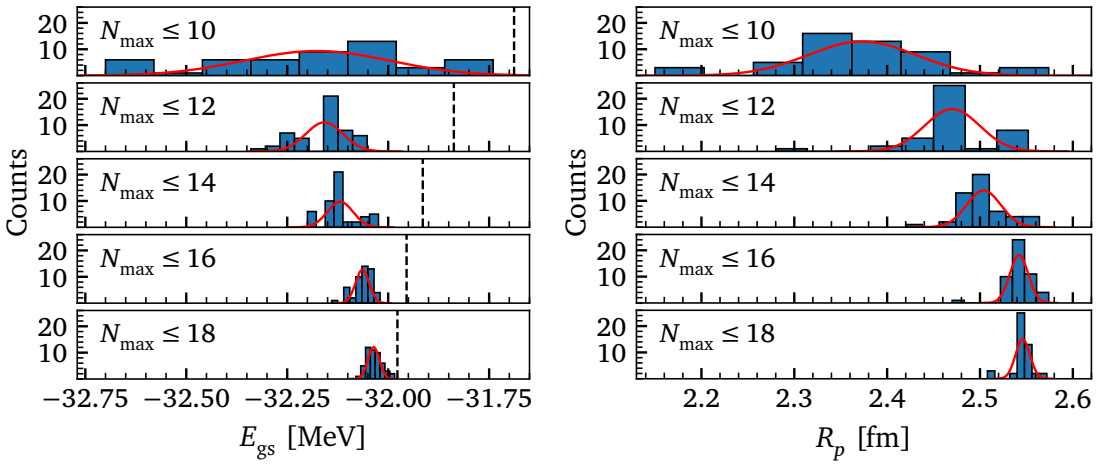


Figure 5.3: Evaluation using the ISU network approach to extrapolate the ground-state energy (left panels) and the point-proton rms radius (right panels) of ^{6}Li at different N_{max} cutoffs. The different panels in each row show histograms of networks trained using more and more data (top to bottom) by increasing the maximum N_{max} in the training data.

The resulting predictions of the remaining 50 networks are then collected into a histogram and a Gaussian is fitted to the distribution to obtain the final prediction and its uncertainty as the center and standard deviation of the Gaussian. An evaluation for ^{6}Li using the ISU ANNs based on the evaluation data described previously is illustrated in Fig. 5.3 for the ground-state energy and the point-proton rms radius. Here, the N_{max} values available for training the networks increases from top to bottom, with for example the top row showing the histograms obtained from networks only trained with data truncated at $N_{max} = 10$. It is immediately obvious that due to the relatively small number of networks after the strict selection criteria, the statistics and therefore the histogram is rather coarse, especially for smaller N_{max} where the extrapolated values are more spread out. This leads to comparatively large uncertainties, which get systematically smaller with increasing model-space size. For more details on the training and evaluation of these networks, we refer to Refs. [42, 135].

5.3 Benchmark Results

We are now in a position to directly compare the various extrapolation methods with our ANN-based approach. In doing so, we deliberately omit experimen-

tal values as a reference point. Because the fully converged theoretical result depends on the chosen interaction, agreement with experiment is not a reliable indicator of extrapolation quality. A method may accurately recover the converged many-body value yet deviate from experiment, while a less robust approach could coincidentally appear more accurate. For this reason, comparisons to experimental data are deferred to a later chapter focused on physical interpretation.

The results obtained from the various methods for the ground-state energy and the point-proton rms radius of ^{6}Li, ^{7}Li, and ^{8}Li grouped by the maximum $N_{\max}$ used in the respective evaluations $\mathcal{N}_{\max}$, is shown in Figs. 5.4 to 5.6, respectively. In these figures, "ISU extrap. B" refers to the exponential extrapolation technique for energies discussed in Section 1.5.1, and "IR extrap." ("IR extrap. man. opt.") refers to IR extrapolations with (without) manual optimization, as described in Section 1.5.2. All results are also given in Table 5.1.

Starting with the ground-state energy extrpolations shown in the upper panels of the figures, we can first of all observe that only the naïve IR extrapolation violates the variational limit given by the dashed black line. Specifically for ^{6}Li and ^{8}Li this deviation is multiple standard deviations large, indicating that the published algorithmic criteria for the data selection breaks down and that the purely fit-based uncertainties are also not reliable in these cases. Both IR methods see a systematic shift towards more binding energy for increasing model-space sizes of the underlying data, though this behavior is significantly suppressed in the manually optimized IR extrapolation, where all subsequent extrapolations are compatible within uncertainties. A similar behavior can be observed for the exponential extrapolation. The ISU ANNs also exhibit a seemingly systematic shift across mode-space sizes, though here the results seem to overestimate the binding energy at smaller $\mathcal{N}_{\max}$ and subsequently converge upwards when information for larger model spaces is added. Our Gen2 networks on the other hand show remarkable consistency and stability across $\mathcal{N}_{\max}$, as the prediction for the converged ground-state energy for all three nuclei barely changes and uncertainties get smaller from the smallest to the largest model spaces. We do see, however, a significant increase in precision for the Gen2+ results, as the selection of more optimal sequences for evaluation tightens the obtained distributions while remaining just as stable as the Gen2 results. See Table 5.1 for the selected $\hbar\Omega$ windows of each evaluation.

Except for the naïve IR method, all methods produce predictions that are compatible with each other within uncertainties at the largest available $\mathcal{N}_{\max}$ for ^{6}Li and ^{7}Li. For ^{8}Li, however, this is not true for the ISU ANN prediction,

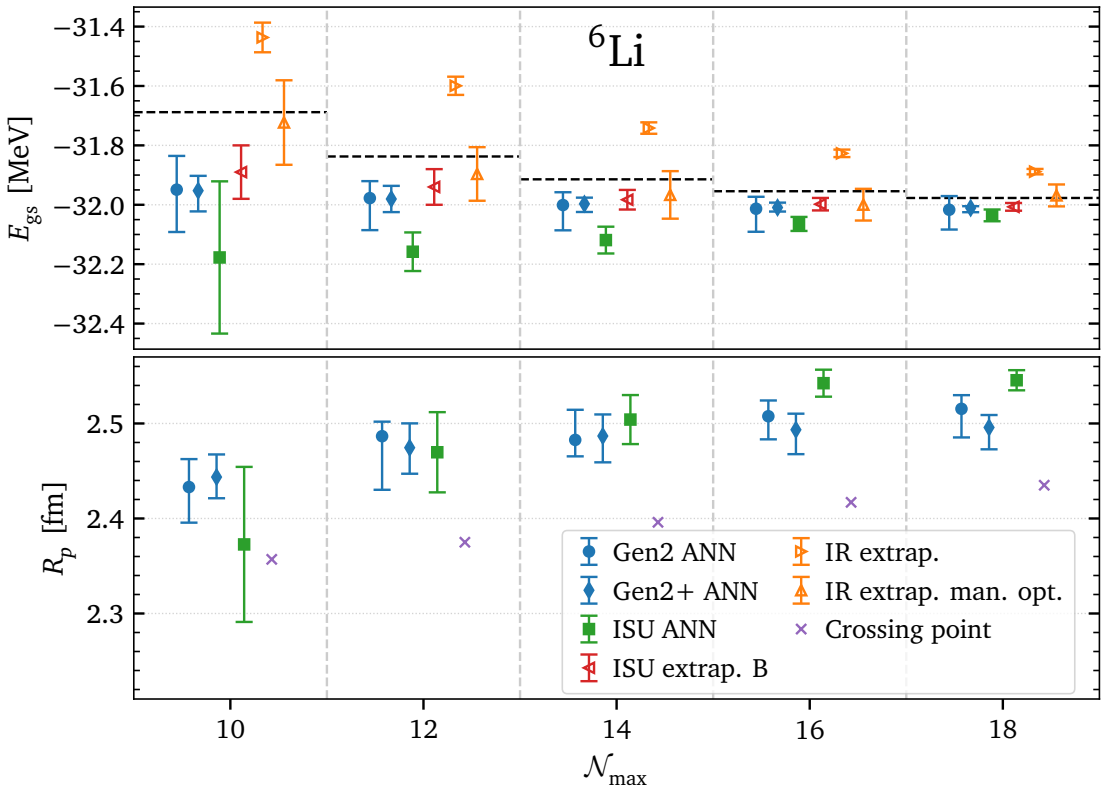


Figure 5.4: Results of ANN based (filled symbols) and traditional (open symbols) extrapolation methods for the ground-state energy and point-proton rms radius of ^{6}Li based on NCSM data with increasing model-space sizes. The variational minimum of the energies is shown with dashed black lines. Note that we do not provide uncertainties for the crossing point method, and that the fit-based uncertainties of the IR methods are not directly comparable to the uncertainties of the other methods.

which overestimates the binding energy compared to all other methods, and only slightly overlaps with the Gen2 prediction within uncertainties. Furthermore, while the extrapolation for $\mathcal{N}_{max} = 10, 12$ are almost identical and seem stable and converged, the following prediction abruptly jump upwards, decreasing confidence that the convergence and reliability of extrapolations can be estimated adequately from a more limited number of $\mathcal{N}_{max}$ evaluations.

Moving on to the results for the point-proton rms radius for all three nuclei, shown in the lower panels of the corresponding plots and also detailed in Table 5.1, we first of all notice that this observable seems to be significantly harder to extrapolate. Especially ^{6}Li shows a systematic shift towards larger radii for both ANN-based methods and the crossing point estimate. We should note that we do not quote an uncertainty for the crossing point method, as it is essentially just a poor man's estimate for the converged value, without a clear or reliable

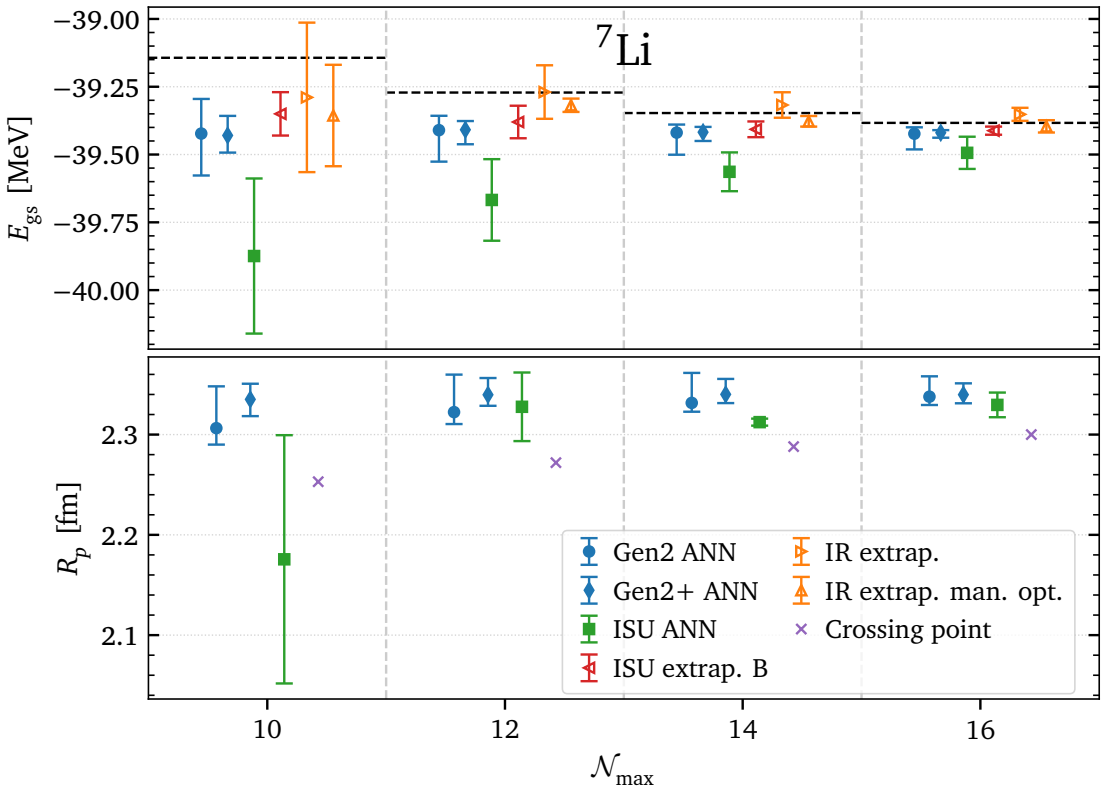


Figure 5.5: Analogous to Fig. 5.4, but for ^{7}Li.

way to add an uncertainty estimate. Despite the upwards trend, which is least severe for our ANN results and stabilizes after one or two $\mathcal{N}_{max}$ steps, the Gen2+ predictions are all compatible with each other within uncertainties, and the same holds for the Gen2 predictions if we exclude the prediction for the smallest $\mathcal{N}_{max}$.

The ISU ANNs show a significantly more pronounced upwards shift, as the prediction for $\mathcal{N}_{max} = 10$ is substantially below our network predictions and the prediction for the largest $\mathcal{N}_{max}$ is a good bit above them. Additionally, none of the smaller model-space extrapolations is compatible with the $\mathcal{N}_{max} = 16, 18$ predictions within the calculated uncertainties.

For ^{7}Li and ^{8}Li, our ANN approach shows remarkably stable predictions with respect to $\mathcal{N}_{max}$. While the Gen2 evaluation for ^{7}Li shows a slight upwards trend, it is fully contained within the quoted uncertainties and fully remedied my the limited frequency selection of Gen2+. Results from both methods are also always compatible with each other.

The ISU networks show a few interesting artifacts in these two nuclei as well. For ^{7}Li, the first extrapolation for $\mathcal{N}_{max} = 10$ is surprisingly small before the predictions jump upwards to be compatible with the predictions of our ANNs. Furthermore, the uncertainty at $\mathcal{N}_{max} = 14$ is suddenly uncharacteristically small,

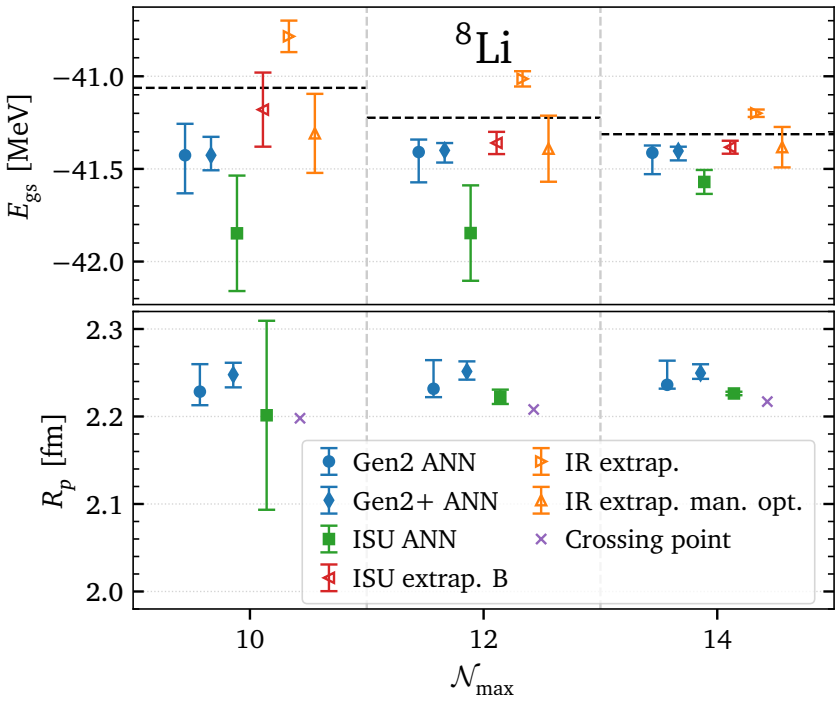


Figure 5.6: Analogous to Fig. 5.4, but for ^{8}Li.

decreasing by an order of magnitude before increasing again by a factor of three at $\mathcal{N}_{max} = 16$. This sudden and extreme decrease in the uncertainty also happens in ^{8}Li from $\mathcal{N}_{max} = 10$ to 12, though it continues to decrease also rather fast again in the next $\mathcal{N}_{max}$ step. The volatile nature of the uncertainty estimate of the ISU ANNs makes an assessment rather difficult. We suspect this is related to the strict selection criteria together with the relatively small number of networks, retaining only the 50 best performing ANNs out of 200 000. Enlarging and diversifying the pool of remaining networks by slightly weakening the selection criteria might improve uncertainty estimates. And while the result for ^{7}Li for the largest model space is fully compatible with our network predictions, the ^{8}Li extrapolation is only compatible with the Gen2 predictions within uncertainties.

The crossing point method estimate seems to systematically underpredict the radius compared to the other methods, and shows a strong $\mathcal{N}_{max}$ dependence with a very slow convergence. It is virtually unusable for precision extrapolations, further compounded by the lack of an uncertainty estimate, that it can only serve as a very rough estimate of the converged result if multiple $\mathcal{N}_{max}$ steps are available.

Method	$\mathcal{N}_{\max} = 10$	12	14	16	18
	^{6}Li ground-state energy [MeV]				
Variational min.	-31.688	-31.837	-31.914	-31.954	-31.977
Gen2 ANN	$-31.949^{+0.114}_{-0.143}$	$-31.978^{+0.057}_{-0.108}$	$-32.001^{+0.043}_{-0.085}$	$-32.013^{+0.040}_{-0.078}$	$-32.017^{+0.046}_{-0.067}$
Gen2+ ANN*	$-31.953^{+0.050}_{-0.070}$	$-31.980^{+0.044}_{-0.044}$	$-31.997^{+0.021}_{-0.027}$	$-32.009^{+0.017}_{-0.013}$	$-32.011^{+0.006}_{-0.014}$
ISU ANN	$-32.18(26)$	$-32.158(65)$	$-32.119(45)$	$-32.065(24)$	$-32.036(20)$
ISU extrap. B	$-31.89(09)$	$-31.94(06)$	$-31.983(33)$	$-31.998(21)$	$-32.007(13)$
IR extrap.	$-31.437(50)$	$-31.599(31)$	$-31.742(19)$	$-31.827(13)$	$-31.888(9)$
IR ex. man. opt.	$-31.72(14)$	$-31.896(90)$	$-31.967(80)$	$-32.000(53)$	$-31.969(37)$
	^{6}Li point-proton rms radius [fm]				
Gen2 ANN	$2.433^{+0.029}_{-0.037}$	$2.487^{+0.015}_{-0.056}$	$2.483^{+0.032}_{-0.017}$	$2.508^{+0.017}_{-0.024}$	$2.515^{+0.014}_{-0.030}$
Gen2+ ANN†	$2.444^{+0.024}_{-0.022}$	$2.474^{+0.026}_{-0.027}$	$2.487^{+0.023}_{-0.028}$	$2.493^{+0.017}_{-0.026}$	$2.496^{+0.013}_{-0.023}$
ISU ANN	2.373(82)	2.470(42)	2.504(26)	2.542(14)	2.546(11)
Crossing point	2.37	2.41	2.43	2.44	
	^{7}Li ground-state energy [MeV]				
Variational min.	-39.143	-39.271	-39.347	-39.383	
Gen2 ANN	$-39.423^{+0.128}_{-0.155}$	$-39.410^{+0.053}_{-0.116}$	$-39.419^{+0.030}_{-0.082}$	$-39.423^{+0.024}_{-0.058}$	
Gen2+ ANN*	$-39.430^{+0.072}_{-0.063}$	$-39.409^{+0.032}_{-0.053}$	$-39.418^{+0.019}_{-0.032}$	$-39.419^{+0.009}_{-0.018}$	
ISU ANN	$-39.87(29)$	$-39.67(15)$	$-39.564(71)$	$-39.494(59)$	
ISU extrap. B	$-39.35(8)$	$-39.38(6)$	$-39.407(29)$	$-39.412(15)$	
IR extrap.	$-39.29(28)$	$-39.270(99)$	$-39.317(47)$	$-39.352(24)$	
IR ex. man. opt.	$-39.36(19)$	$-39.318(25)$	$-39.377(20)$	$-39.396(23)$	
	^{7}Li point-proton rms radius [fm]				
Gen2 ANN	$2.306^{+0.042}_{-0.016}$	$2.322^{+0.037}_{-0.012}$	$2.332^{+0.030}_{-0.009}$	$2.338^{+0.020}_{-0.008}$	
Gen2+ ANN†	$2.335^{+0.016}_{-0.017}$	$2.340^{+0.017}_{-0.011}$	$2.340^{+0.015}_{-0.009}$	$2.340^{+0.011}_{-0.009}$	
ISU ANN	2.18(12)	2.328(34)	2.312(4)	2.330(12)	
Crossing point	2.28	2.30	2.31	2.32	
	^{8}Li ground-state energy [MeV]				
Variational min.	-41.062	-41.224	-41.313		
Gen2 ANN	$-41.426^{+0.170}_{-0.205}$	$-41.409^{+0.067}_{-0.164}$	$-41.413^{+0.039}_{-0.115}$		
Gen2+ ANN*	$-41.426^{+0.099}_{-0.081}$	$-41.400^{+0.040}_{-0.066}$	$-41.403^{+0.023}_{-0.051}$		
ISU ANN	$-41.85(31)$	$-41.85(26)$	$-41.570(65)$		
ISU extrap. B	$-41.18(20)$	$-41.36(6)$	$-41.383(35)$		
IR extrap.	$-40.785(85)$	$-41.014(41)$	$-41.199(20)$		
IR ex. man. opt.	$-41.31(21)$	$-41.39(18)$	$-41.38(11)$		
	^{8}Li point-proton rms radius [fm]				
Gen2 ANN	$2.228^{+0.031}_{-0.015}$	$2.232^{+0.033}_{-0.010}$	$2.236^{+0.028}_{-0.004}$		
Gen2+ ANN†	$2.248^{+0.014}_{-0.015}$	$2.252^{+0.011}_{-0.009}$	$2.250^{+0.010}_{-0.007}$		
ISU ANN	2.20(11)	2.223(8)	2.226(2)		
Crossing point	2.21	2.22	2.23		

Table 5.1: Extrapolation results for a variety of ANN-based and classical methods for the ground-state energy and point-proton rms radius of ^{6}Li, ^{7}Li, and ^{8}Li, grouped by $\mathcal{N}_{\max}$, which limits the NCSM data used for the evaluation to $N_{\max} \leq \mathcal{N}_{\max}$. Gen2+ ANN rows marked with (*) use a HO selection of $\hbar\Omega = 10$ to 20 MeV, ones with (†) use $\hbar\Omega = 15$ to 25 MeV with a step size of 2.5 MeV.

5.4 Conclusion

In conclusion, the Gen2+ networks and evaluation procedure offer the most stable and, given the agreement of almost all methods at the largest model-space, probably also the most accurate predictions with in comparison very high precision. The fact that the predictions for both Gen2 and Gen2+ remain almost unchanged across $\mathcal{N}_{\text{max}}$ makes these methods very well suited for calculations in heavier nuclei or more with more demanding nuclear interactions, where computational cost limits the maximum achievable N_{max} significantly. For energies, only the exponential extrapolation method or the IR extrapolation with manual optimization are reliable enough to serve as good sanity checks for the network predictions, though both methods still suffer a slight downward shift of the extrapolated values for increasing model-space sizes. The ISU ANNs on the other hand show severe deficiencies in this benchmark, often severely underestimating the ground-state energy, especially for smaller $\mathcal{N}_{\text{max}}$. This severely limits the applicability of this method beyond the lower half of the p-shell once 3N forces are included in the calculation. Furthermore, the displayed artifacts, including the sudden jumps at certain $\mathcal{N}_{\text{max}}$ values and uncertainty estimation volatility for radii suggest there are systematic effects not accounted for within this method, lowering the confidence in the overall extrapolation results, especially for high-precision applications. However, we should highlight that the ISU approach is much more flexible and very easy to adapt to a wide variety of observables compared to our ANN method, as it does not rely on training data with known fully-converged observable results for training. For example, Ref. [135] successfully adapted this method to electric quadrupole moments.

6

Post-Processing of Network Predictions

In previous chapters, we constructed ANNs capable of predicting the converged energy and radius from NCSM data. For the development of these methods, our main focus was on the ground state and only a single observable or nucleus. However, we can extend the capability of our network framework to difference-based observables, like excitation energies, point-proton rms and point-neutron rms radius differences, or even differences between nuclei, by post-processing our network results.

Another important step towards *ab initio* calculations is the ability to robustly combine order-by-order uncertainties of the underlying chiral interactions with the many-body or extrapolation uncertainties our networks provide. This can also be achieved via post-processing of the network predictions, as we will show in this chapter.

6.1 Difference-Based Observables

While the neural networks we have developed and evaluated in previous chapters work exceptionally well for individual observables like the ground-state energy or radii, the relation between some of these observables is often also of great interest. One example is the excitation energy, which is simply the difference between the ground-state energy and an excited-state energy. Since our ANNs can make predictions for the two observables individually, as we can simply

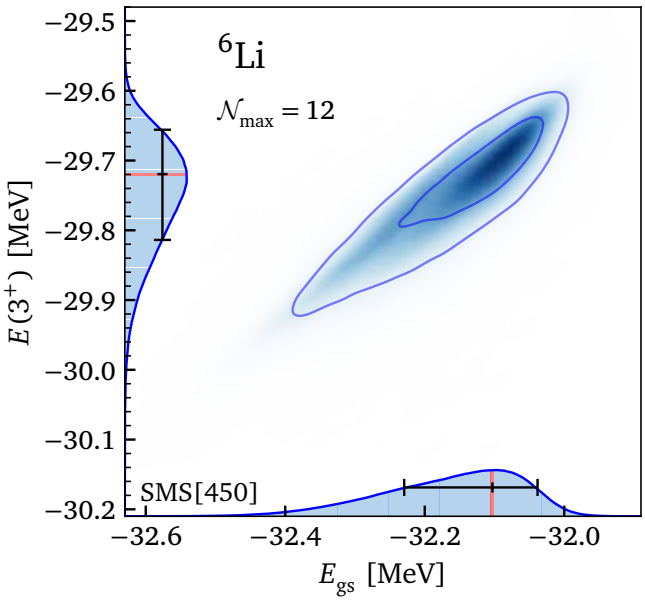


Figure 6.1: Depiction of the correlation between ground state and the 3^+ excited-state energy of ^{6}Li for the same networks and corresponding samples, using the SMS[450] interaction. The density is constructed via a Gaussian kernel-density estimator of all corresponding pairs of networks predictions $(E_{gs}, E(3^+))$. The inner and outer contours of the density contain about 39 % and 86 % of the density, which correspond to 1σ and 2σ intervals of the one dimensional projections.

reuse the ground-state energy trained networks for the excited-state predictions due to the extremely similar convergence behavior ob both observables, we can calculate an estimate for the excitation energy by subtracting the network predictions from each other. With a naïve approach, given two predictions a and b with associated uncertainties $\delta a_\pm$ and $\delta b_\pm$, the difference using simple error propagations computes to

$$a^{+\delta a_+}_{-\delta a_-} - b^{+\delta b_+}_{-\delta b_-} = \left(a-b\right)^{+\sqrt{a_+^2+b_-^2}}_{-\sqrt{a_-^2+b_+^2}}. \tag{6.1}$$

While this approach is definitely viable, it typically leads to rather large uncertainties for the difference observable we are interested in, since it explicitly assumes there are no correlations between the two results. However, this is clearly not the case, as demonstrated in Fig. 6.1, which depicts the correlation between the ground and the first 3^+ state energy of ^{6}Li for corresponding network predictions. To create this plot, we turn the point cloud from the large number of corresponding $(E_{gs}, E(3^+))$ tuples for each network prediction into a density via a Gaussian kernel-density estimator. Clearly, predictions for samples and networks

that result in smaller than average ground-state energies also tend to result in smaller predictions for the 3^+ state, and vice versa. This results in the elongated shape of the distribution along the diagonal visible in the figure, opposed to a circular shape which would indicate there is no correlation.

This correlation is due to two factors. First, the observables themselves are strongly correlated and so are their convergence patterns. Therefore, samples created from similarly behaving evaluation sequences will also lead to, after normalization, similar samples and therefore correlated predictions. SEcond, predictions by the same network with the same hidden parameters will obviously be correlated with each other.

We can use these correlations to improve the predictions and tighten the uncertainties associated with these difference-based observables. Instead of just looking at the final prediction and its uncertainty, we can take a step back and use the correlations between individual predictions, which in turn make up the final histogram from which we extract the value and uncertainty of the observable. So by subtracting the two observables sample-wise, i.e. calculating the difference between the predictions for like samples from the same ANN, we can construct a new histogram based on these calculated differences. Like samples can in the simplest term mean samples constructed from the same $\hbar\Omega$ HO sequences, which is typically the case if the values of the resulting observable are relatively close to each other, e.g. for comparatively small excitation energies. For other observables, where the value can vary more widely, as is the case for point-proton rms and point-neutron rms radius differences in nuclei with a relatively large proton–neutron asymmetry, it is often more helpful to match similar HO sequences. Using the Gen2+ with its convergence-related selection of sequences, this is easily achieved. In this case, predictions from samples with closely matched $\hbar\Omega$ from the same ANN are subtracted from each other. This also guarantees that there are exactly the same number of sample for both observables, which is the only rigid requirement for this method to work.

Figure 6.2 demonstrated the technique for the case of the excitation energy of the 3^+ state of ^{6}Li, where we first compute the network predictions for the ground state and the 3^+ state energies separately. These are then subtracted from each other sample-wise and collected into histograms, shown as blue histograms in the lower right panel. This results in remarkably stable and precise predictions. The results for the naïve approach discussed in Eq. (6.1), shown as red error bars in the plot, are both consistent with each other and the sample-wise predictions. However, the sample-wise predictions are much more precise, while the final predictions are almost identical, which supports our claim that the sample-wise

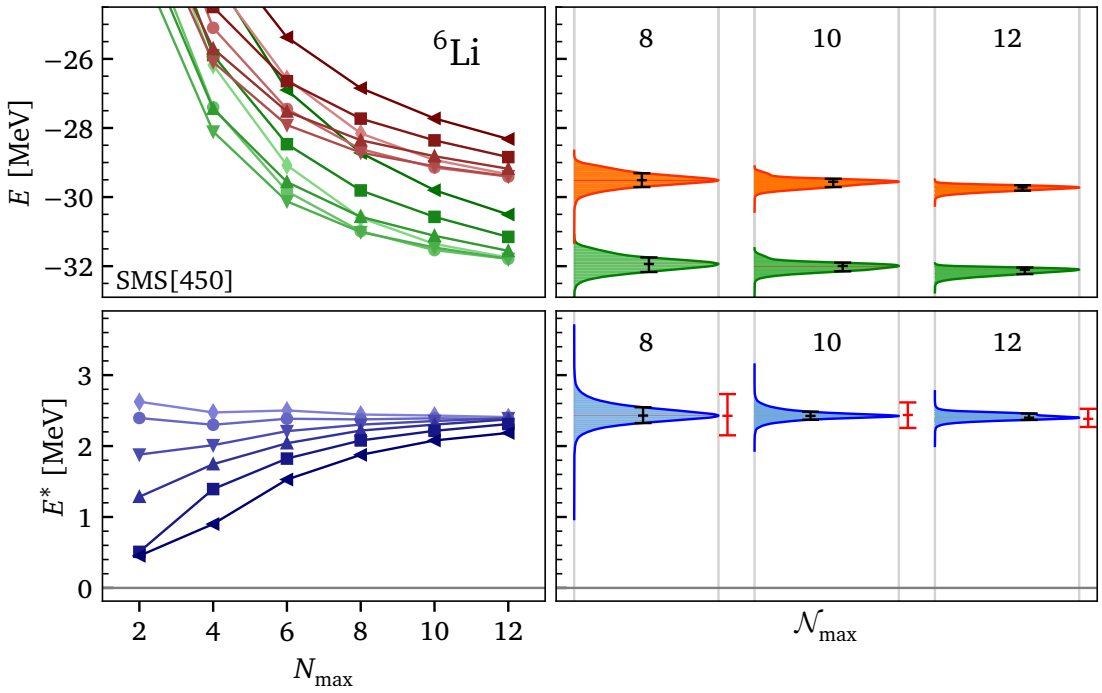


Figure 6.2: Illustration of the post-processing results for the excitation energy of ^{6}Li for the SMS[450] interaction. The ground-state energy (green) and the energy of the 3^+ state (red) are calculated and predicted separately (upper panels), and the histogram of the sample-wise differences is depicted in the lower right panel. The sequences shown in the lower left panel are simply the difference of the sequences depicted above, and are only shown for illustrative purposes and are not used for the network predictions. Results for the naïve approach of simply subtracting the final predictions results are shown as red error bars.

approach's main advantage is using hidden correlations in the predictions of related observables to tighten otherwise inflated uncertainties.

The following two sections show some more examples and practical applications of this method for a variety of observables and nuclei.

6.1.1 Excitation Energies

The most obvious application for this technique are the already mentioned excitation energies, which we can calculate from predictions for the ground-state energy and the individual excited-state energies. Figure 6.3 shows selected excitation energy predictions for ^{6}Li and ^{8}Li. For both nuclei, the predictions for the excitation energies are very stable and robust, though the highest states display a slight downward shift for increasing $\mathcal{N}_{\max}$. We will not show more

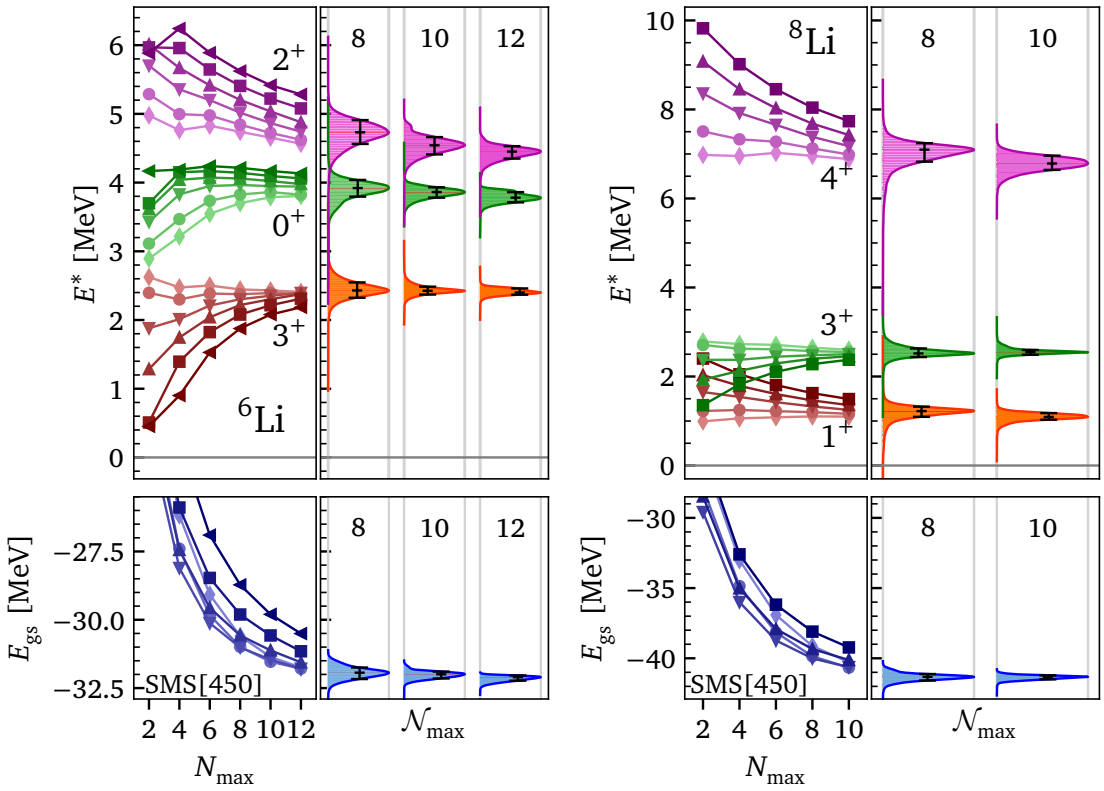


Figure 6.3: Excitation energy predictions for ^{6}Li and ^{8}Li using the SMS[450] interaction. The lower panels show the ground-state energy evaluation data and network predictions, while the upper panels show the excitation energy predictions based on ground-state energy results. The sequences shown in the upper left panels is only shown for illustrative purposes and is not used for the network predictions, which follow the recipe described in the text.

results for excitation energies in this work, focussing our attention more on the radius differences discussed in the next sections. However, more examples for this type of evaluation have been published in Ref. [128]. Overall, we note that the difference sampling works very well for excitation energies and is a simple and effective post-processing step to obtain robust predictions for this observable without the need to modify our core network framework or train new networks.

6.1.2 Radius Differences

Another exciting application of the sample-wise difference method are differences between radii, which will be used in the high-precision boron analysis in the next chapter. Here, being able to precisely determine both the difference between the proton and neutron radius in the same nucleus and the difference between radii

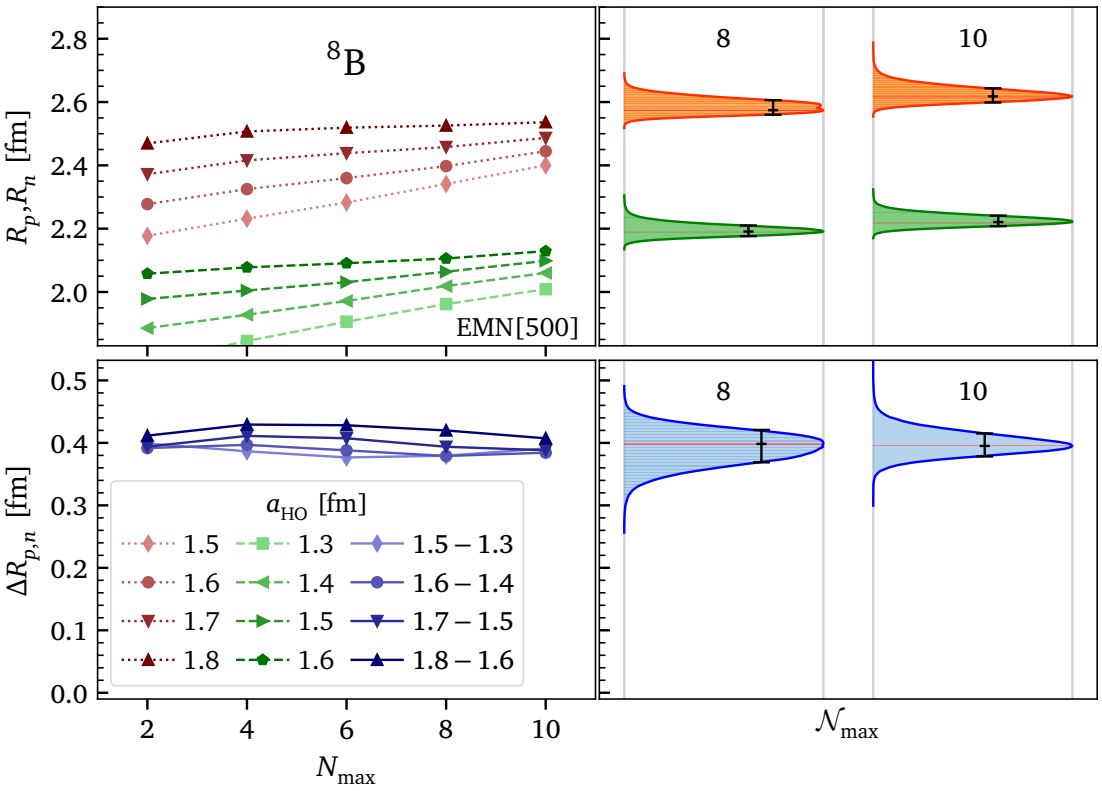


Figure 6.4: Example of a sample-wise precision evaluation for the difference between the point-proton rms (red) and the point-neutron rms radius (green) of ^{8}B using the EMN[450] interaction. Since the values of the proton and neutron radius differ significantly, different a_{HO} ranges were used for each observable according to the Gen2+ recipe. The blue lines in the lower left panel show the difference of matching a_{HO} sequences, though these are just shown to illustrate the method and are not used by the networks to generate the histograms in the right panel.

of different nuclei, especially within an isotope chain, is of particular interest for high-precision comparisons with experimental data.

6.1.2.1 Proton–Neutron Radius Difference. While the difference between the proton and neutron radius in heavy nuclei is mostly used to measure the neutron skin thickness, in the smaller systems accessible with the NCSM this difference is very helpful in identifying possible proton or neutron halo nuclei. However, especially in these types of nuclei, the difference between the proton and neutron radius is almost always large enough to require different a_{HO} or $\hbar\Omega$ ranges for the optimal evaluation in the Gen2+ framework. In these cases, we match these sequences, as is shown in Fig. 6.4, where we look at the point-proton rms and point-neutron rms radius difference $\Delta R_{p,n} = R_p - R_n$ in ^{8}B. The values of

both radii require different optimal a_{HO} ranges as shown in the legend, and the blue lines in the lower left panel illustrate what the resulting difference of the matching sequences looks like. Of course, as discussed before, these sequences are not used to obtain the network predictions in the lower right panel, as these are calculated via the sample-wise difference method. However, one important aspect when looking at the network results in the right-hand side panels of Fig. 6.4 is that while the predictions for the proton and neutron radius show a slight shift towards larger radii for increasing $\mathcal{N}_{\mathrm{max}}$, this trend is eliminated entirely in the difference prediction below. The elimination of these systematic trends is a key feature of this method, improving and stabilizing the resulting predictions.

6.1.2.2 Inter-Isotopic Radius Difference. Analogously, this technique can also be applied to radius differences between isotopes, which similarly results in very precise predictions for the radius difference. Of course, as is the case in nuclei with significant proton and neutron radius differences, we generally expect to have to use different optimal a_{HO} ranges for both isotopes. While we could use this technique to extract just, for example, the plain point-proton rms radius difference between two isotopes, with respect to a potential comparison to isotope shift experiments it is more useful to compute the squared radius difference. The final prediction in computed by evaluating our radius networks on the point-proton rms data calculated in the NCSM to obtain predictions for the proton radius of both nuclei. Then, each individual prediction is squared and subtracted sample-wise from its partner from the other isotope. With this process, we once again obtain a distribution of the squared point-proton rms radius difference which we can the statistically analyze. This process is illustrated in Fig. 6.5, which shows the squared point-proton rms radius difference $\Delta R_p^2(^{11}\mathrm{B}, ^{10}\mathrm{B}) = R_p^2(^{11}\mathrm{B}) - R_p^2(^{10}\mathrm{B})$ between ^{10}B and ^{11}B, subtracting the former from the latter. Due to computational constraints, only data up to $N_{\mathrm{max}} = 8$ and therefore just a single $\mathcal{N}_{\mathrm{max}}$ could be computed in this case. As expected, slightly different a_{HO} windows were selected by the Gen2+ criteria, indicating that the converged values show a significant difference. Unfortunately, the a_{HO} grid chosen beforehand did not allow for the selection of 5 a_{HO} for both cases which fulfill the selection criteria, reducing the number of sequences for both nuclei to 4. As discussed before, this number must be the same for both observables in this approach to be able to properly match all evaluations with a partner from the other evaluation. The computation of the squared radius difference can then be carried out as described above. Despite squaring the individual predictions, the resulting distribution appears

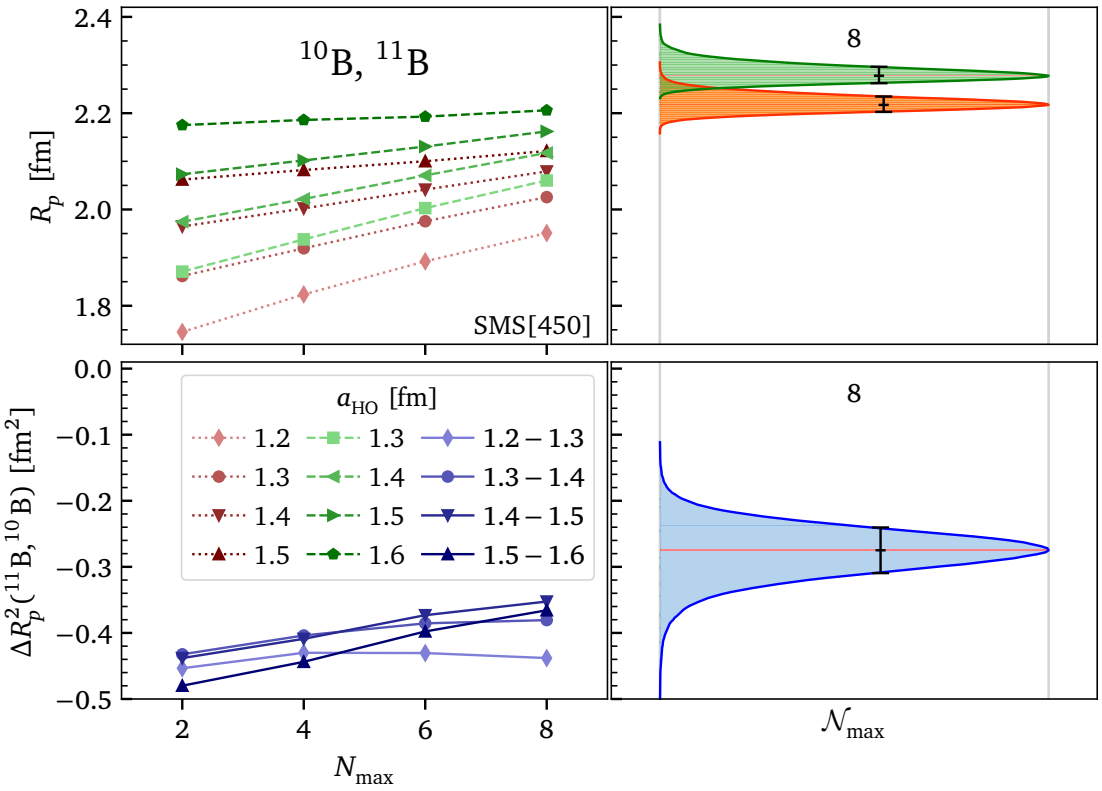


Figure 6.5: Example of a sample-wise precision evaluation for the squared point-proton rms radius difference (blue) between ^{10}B (green) and ^{11}B (red) using the SMS[450] interaction. We use slightly different a_{HO} windows for both nuclei according to the Gen2+ selection criteria, limited to just 4 HO parameters due to availability. The sequences in the lower left panel are only shown for illustrative purposes and are not used to generate the final network prediction for the difference.

to be very symmetrical shows no signs of artifacts, similar to previous examples. This combined with other evaluations and previously shown results for other observables increases our confidence in the merits of this general approach and this application in particular, qualifying it for high-precision calculations of nuclear radii.

6.2 Including Chiral Order-By-Order Uncertainties

Incorporating uncertainties associated with chiral EFT interactions is an essential ingredient of any high-precision *ab initio* nuclear structure calculation. As discussed in Section 1.2.1, the dominant source of interaction uncertainty arises

from the truncation of the chiral expansion. In this work, we estimate these truncation effects using the BUQEYE point-wise Bayesian model [8], which assumes that omitted higher-order contributions are of natural size and suppressed according to $(Q/\Lambda_\chi)^\nu$.

The BUQEYE point-wise model is formulated for scalar observable values and therefore does not directly apply to the probability distributions obtained from our ANN-based statistical evaluation, which yields histograms of predictions rather than single numbers. To retain the conceptual structure of the BUQEYE approach while consistently incorporating it into our framework, we adapt the model to operate on these prediction histograms. Specifically, we follow the BUQEYE procedure up to the construction of the Student's t distribution that represents the truncation uncertainty of the highest available chiral order. The inputs to the point-wise model are the most probable values of the observable at successive chiral orders, obtained from our ANN evaluation and statistical analysis described in Section 3.3. From these, the BUQEYE model constructs a Student's t distribution $t_\nu(0,\tau^2)$ with mean zero, scale τ, and ν degrees of freedom, where both τ and ν are inferred from the order-by-order pattern of the results [8].

Rather than shifting this distribution to the highest-order prediction and extracting a confidence interval, as is done in the standard scalar implementation, we retain the full Student's t distribution and incorporate it directly into our prediction histogram via convolution. Concretely, if the ANN evaluation yields a normalized histogram $h(s)$ over possible observable values s, and $t_\nu(0,\tau^2)$ denotes the Student's t distribution from the BUQEYE model, we define the updated histogram as the convolution

$$\tilde{h}(s) = \int_{-\infty}^{\infty} h(r)\, t_\nu(s-r,\tau^2)\,\mathrm{d}r\,, \tag{6.2}$$

or, for the discrete case,

$$\tilde{h}[s] = \sum_{r\in S} h[r]\, t_\nu(s-r,\tau^2) \quad \text{with } r\in S\,, \tag{6.3}$$

where S denotes the set of positions of the histogram bins, extended to positive and negative infinity. The updated histogram $\tilde{h}(s)$ simultaneously encodes the many-body and extrapolation uncertainties from the ANN evaluation and the truncation uncertainty from the chiral interaction. The same iterative procedure described in Section 3.3 is then applied to extract the final prediction and its associated uncertainty. Because the convolution with the Student's t distribution

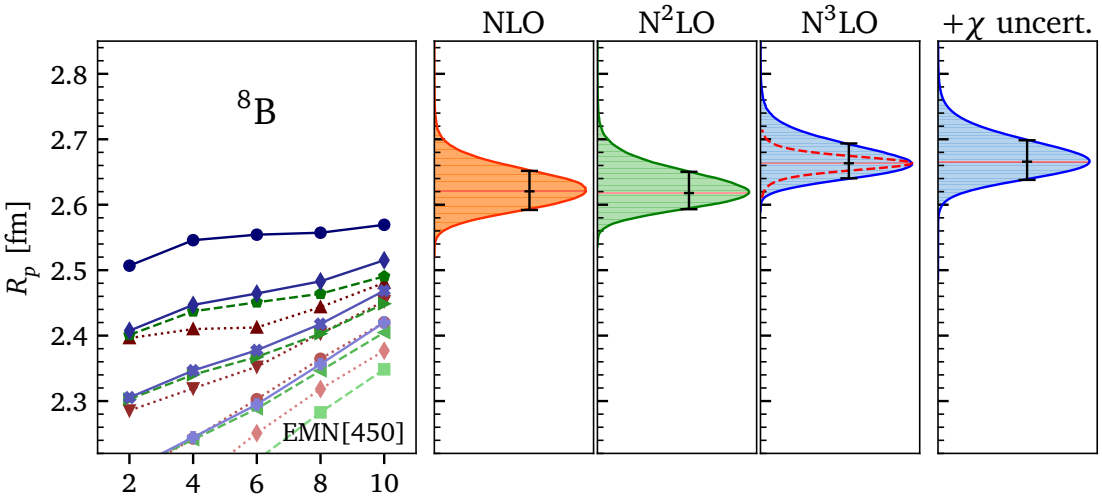


Figure 6.6: Example for a full network evaluation for several orders of the EMN[450] interaction to calculate the point-proton rms radius of ^{8}B with chiral order-by-order interaction uncertainties. The network predictions in the middle panel use the evaluation data depicted in the left panel corresponding to the same color. A Student's t distribution is calculated from the results of these evaluations and shown as a dashed red line in the N^{3}LO column. The right-most panel shows the convolution of the highest-order network predictions with the Student's t distribution.

produces a sufficiently smooth histogram, the additional moving-average smoothing step used in the standard evaluation can be omitted. This is expected, since the convolution is mathematically equivalent to smoothing with a kernel that is typically much broader than the rectangular function used for the high-frequency noise suppression.

The full procedure is illustrated in Fig. 6.6. For each chiral order, here NLO, N^{2}LO, and N^{3}LO of the EMN[450] interaction, we first select the optimal a_{HO} range and evaluate the ANNs to obtain a prediction histogram. These histograms are smoothed to determine the most probable value at each order, indicated by the red lines and error bars in the figure. The resulting order-by-order values serve as inputs to the BUQEYE point-wise model, which yields the parameters ν and τ of the Student's t distribution. This distribution, shown as a dashed red curve in the N^{3}LO panel, is then convolved with the original (unsmoothed) histogram of the highest-order ANN prediction. The final histogram, shown in the rightmost panel, includes both ANN and chiral truncation uncertainties and is used to extract the final result.

The same strategy applies straightforwardly to derived observables, such as the proton–neutron radius difference or isotope-dependent squared point-proton rms radius differences. In these cases, the post-processed ANN predictions for

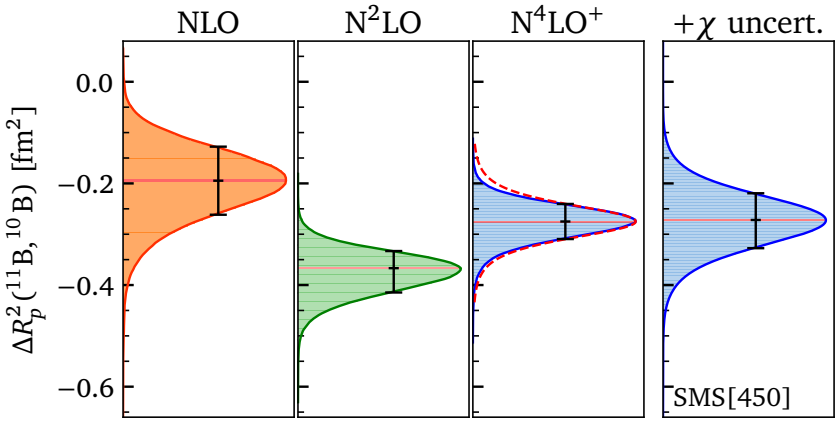


Figure 6.7: Analogous to Fig. 6.6, but for the squared point-proton rms radius difference between ^{10}B and ^{11}B calculated with the SMS[450] interaction. Here, the chiral uncertainty is about the same size as the many-body and extrapolation uncertainty.

the derived observable, obtained as described in Section 6.1, are used as inputs to the BUQEYE point-wise model. The resulting Student's t distribution is then convolved with the prediction histogram of the highest available chiral order analogous to the procedure outlined above. An explicit example is shown in Fig. 6.7 for the squared point-proton rms radius difference between ^{10}B and ^{11}B. Because one fewer consistent NN+3N order is available for the SMS interaction, the expected suppression of higher-order contributions is reduced. As a consequence, the truncation uncertainty inferred from the BUQEYE model is comparable in size to the ANN-derived many-body uncertainty at N^{4}LO$^+$, leading to a noticeably broader final prediction distribution, as expected.

The ability to coherently and robustly combine many-body, extrapolation, and chiral truncation uncertainties is a central ingredient of the analysis presented in the following chapter. It enables us to perform *ab initio* calculations for the boron isotopic chain, with uncertainty estimates that are both transparent and quantitatively reliable.

7

High-Precision Boron Radius Analysis

Having introduced all essential ingredients and techniques, ranging from chiral EFT-based interactions to the SRG and the NCSM, and finally our Gen2+ network architecture, evaluation procedure, and post-processing steps, we are now in a position to combine these elements and apply them to a series of boron isotopes for high-precision radius calculations. We begin by discussing the underlying NCSM data generated for this study and by briefly looking at the influence of the SRG and the model-space convergence of our network predictions. Next, we analyze the order-by-order behavior of the two different interaction families in a few select cases. We then turn to ground-state energy results in order to assess potential systematic deviations from experimental values, which could also affect the extracted radii. Next, we present results for the point-proton and point-neutron rms radii and compare them with available experimental data, as well as compute the proton–neutron radius difference. The latter quantity is of particular interest for ^{8}B, which is a potential candidate for exhibiting a proton halo. We subsequently provide high-precision estimates for the squared point-proton rms radius differences between various boron isotopes and compare the ^{10}B and ^{11}B difference with available isotope-shift measurement data of comparable precision, which are specifically designed to extract these differences. In addition, we offer theoretical predictions to complement and compare with future experimental efforts. The results presented in this chapter largely coincide with those reported in our publication on this subject [129]. However, we extend the discussion by including two additional nuclei as well as the squared point-proton rms radius difference between ^{10}B and ^{11}B.

7.1 Evaluation Data

The data underpinning this chapter comprise an extensive set of NCSM calculations for ^{8}B, ^{10}B, ^{11}B, and ^{12}B. To further support the analysis, several additional nuclei were computed and included in the evaluation. These include ^{6}He, serving as a benchmark for a well-established (neutron) halo nucleus, as well as the $N = 3$ isotonic chain from ^{6}Li and ^{7}Be to ^{8}B. The latter enables the identification of systematic trends toward the potential proton-halo candidate ^{8}B by tracking the evolution of the proton radius as one proton is added at each step. In addition, ^{8}Li was included as the mirror nucleus of ^{8}B to investigate the theoretical differences between proton-rich and neutron-rich systems in very light nuclei.

In line with the *ab initio* philosophy of this high-precision study, each nucleus and observable was calculated using two distinct families of chiral chiral EFT interactions at several chiral orders, allowing for a comprehensive order-by-order assessment of interaction uncertainties. One interaction family is the EMN NN+3N interaction at two cutoff values $\Lambda = 450$ and 500 MeV (EMN[450] and EMN[500]), evaluated at NLO, N^{2}LO, and N^{3}LO. The second family consists of the SMS NN+3N interactions developed within the LENPIC group, computed at NLO, N^{2}LO, and N^{4}LO$^+$ for the same cutoff values, $\Lambda = 450$ and 500 MeV (SMS[450] and SMS[500]). All interactions and operators were evolved using the SRG with a flow parameter of $s = 0.08\,\mathrm{fm}^4$ to accelerate NCSM model-space convergence. In addition, a limited number of calculations were performed at N^{2}LO with a smaller flow parameter, $s = 0.04\,\mathrm{fm}^4$, to assess the sensitivity of the observables to the free-space SRG evolution.

All NCSM calculations presented in this chapter were carried out by Pieter Maris [136] using the MFDn code [137, 138], with the required interaction matrix elements and SRG corrections to the radius operator computed and provided by us. The calculations employed an HO basis with oscillator lengths spanning 1.2 to 2.4 fm in steps of 0.1 fm. The maximum model-space truncation $N_{\max}$ depends on the nucleus and ranges from $N_{\max} = 12$ for ^{6}He, to $N_{\max} = 10$ for $7 \leq A \leq 8$, and $N_{\max} = 8$ for the heavier boron isotopes. For all cases, ground-state energies, point-proton radii, and point-neutron rms radii were computed.

The network predictions shown in this chapter exclusively employ the Gen2+ architecture and evaluation scheme. At the time these calculations were scheduled, the optimized HO sequence selection had not yet been finalized, as the radii data set described above does not provide five radius sequences converging from

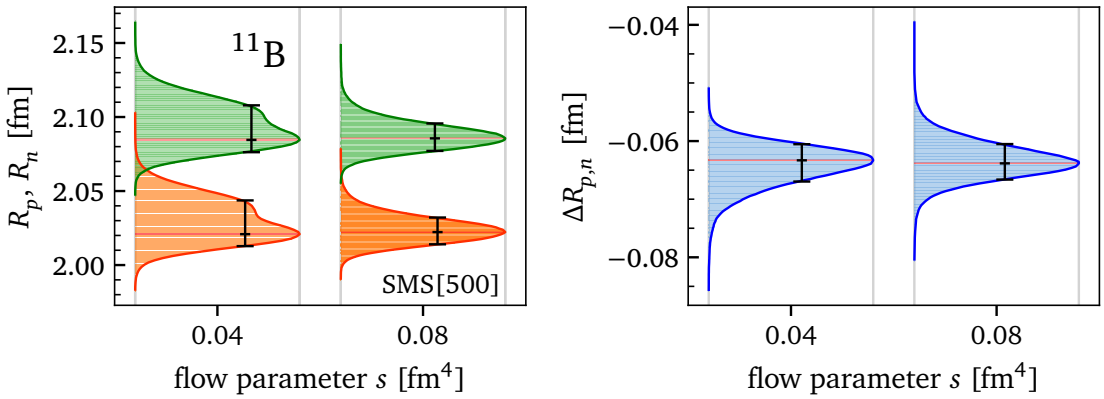


Figure 7.1: Proton and neutron radius as well as the proton–neutron radius difference of ^{11}B for two different SRG flow parameters of the SMS[500] interaction at N^{2}LO to test the SRG sensitivity. The evaluation was performed for $\mathcal{N}_{\text{max}} = 8$.

below for all nuclei and interaction combinations. Under these circumstances, a modified a_{HO} grid, either extending to one or two values below 1.2 fm or employing a different grid spacing, would have been preferable. However, rather than generating additional NCSM data, we opted to reduce the number of a_{HO} sequences used in the optimized Gen2+ selection to four. While this reduction decreases the number of samples underlying the resulting network-prediction distributions, its impact on the final results was found to be minor in cases where a direct comparison with the standard five-sequence selection was possible.

7.2 Network Predictions

Building on the extensive data set described in the previous section, we now apply our Gen2+ ANNs to all nuclei and interaction combinations to obtain predictions for ground-state energies as well as proton and neutron radii. Before interpreting the resulting physical trends, it is essential to assess the robustness of these predictions. In particular, we examine their stability with respect to the maximum model-space truncation $\mathcal{N}_{\text{max}}$, and quantify the influence of the free-space SRG evolution on the predicted observables.

We begin by considering the sensitivity to the SRG flow parameter and focus on one of the cases for which calculations are available at both $s = 0.04\,\text{fm}^4$ and $s = 0.08\,\text{fm}^4$. The corresponding results for ^{11}B are shown in Fig. 7.1, where we display the prediction distributions for the proton and neutron radii, as well as their difference, computed using the post-processing technique introduced in

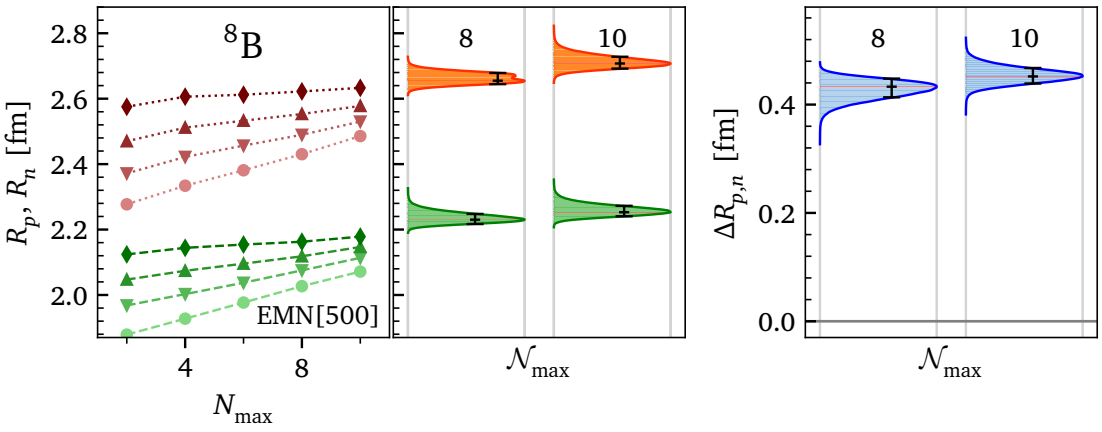


Figure 7.2: Model-space stability of the proton (red) and neutron radius (green) as well as the proton–neutron radius difference Gen2+ network predictions of ^{8}B using the EMN[500] interaction at N^{3}LO.

Section 6.1. For the less strongly evolved interaction, the prediction distributions for the individual proton and neutron radii exhibit noticeable artifacts and a larger spread compared to the results obtained at $s = 0.08\,\mathrm{fm}^4$. This behavior can be attributed to the slower convergence of the underlying NCSM calculations, which poses a more challenging extrapolation task for the networks. Despite these differences in the shape of the prediction distributions, the final network predictions, indicated by the error bars in the figures, remain essentially unchanged. In particular, the artifacts observed in the individual radii are absent in the proton–neutron radius difference, for which the predictions at both flow parameters are nearly identical. For all three observables, the maximum deviation between the two SRG flow parameters is below 2 %. Similar behavior is observed in the few additional cases where results at both flow parameters are available. Although this comparison does not provide a direct benchmark against the unevolved (bare) interaction, it suggests that residual SRG-induced effects are likely at the level of a few percent at most, and significantly smaller for difference observables.

We now turn to the stability of the network predictions with respect to the maximum $\mathcal{N}_{max}$ used in the input data. As an illustrative example, Fig. 7.2 shows the point-proton and point-neutron rms radius predictions for ^{8}B, together with their difference, obtained for different values of $\mathcal{N}_{max}$. While the proton-radius histogram at $\mathcal{N}_{max} = 8$ exhibits mild artifacts that lead to a slightly smaller final prediction, the predictions are overall stable and mutually consistent within uncertainties. Furthermore, the proton–neutron radius difference shows no such artifacts and is more stable than the individual results, as expected. Notably, ^{8}B

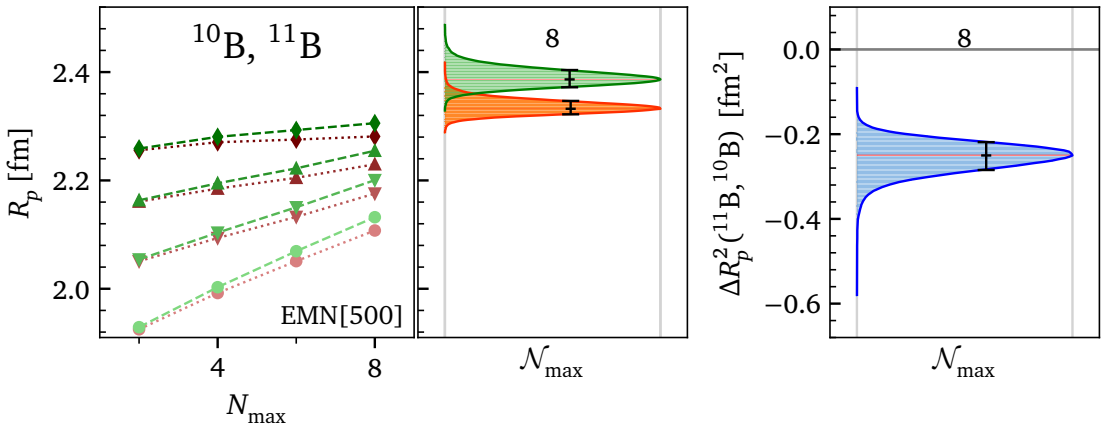


Figure 7.3: Example inter-isotope point-proton rms difference analysis for ^{10}B (green) and ^{11}B (red) using the EMN[500] interaction at N^{3}LO.

represents one of the more challenging cases in the present data set, for nearly all other nuclei and interaction combinations, the corresponding predictions are more stable.

Finally, an example of the inter-isotope radius difference analysis is shown in Fig. 7.3, where we present the individual point-proton rms radius predictions for ^{10}B and ^{11}B, along with their squared radius difference. Neither the individual predictions nor the derived difference exhibit signs of artifacts, and the resulting values are show a relatively high level of precision. This behavior is representative of the inter-isotope analyses performed for other nucleus pairs in the data set and underlines the reliability of the network-based approach for extracting precise radius differences.

7.3 Order-By-Order Analysis

A key feature of employing chiral EFT-based interactions is the existence of a systematic hierarchy of improving orders, which enables controlled estimates of truncation uncertainties. Using the BUQEYE point-wise model, we exploit this hierarchy to quantify the uncertainty associated with neglected higher-order terms in the chiral expansion and combine it with the many-body uncertainties inferred from the ANN predictions, as outlined in Section 6.2. To this end, we use predictions at three successive chiral orders for both the EMN and SMS interaction families.

For the EMN interactions, the highest available order consists of consistent NN+3N forces at N^3LO. In contrast, the highest-order SMS interaction combines a NN force at N^4LO^+, i.e. N^4LO with a few select next-to-next-to-next-to-next-to-next-to-leading order (N^5LO) terms, with a 3N interaction at N^2LO, as three-nucleon forces beyond this order are not yet available for this interaction. Owing to this inconsistency in chiral order between the two and three-body force, we treat the highest-order SMS interaction effectively as a third-order result within the BUQEYE uncertainty analysis. This choice limits the expected suppression of higher-order contributions at the highest order and consequently leads to generally larger estimated interaction uncertainties for the SMS family.

Representative examples of the resulting order-by-order calculations and uncertainty estimates are shown in Fig. 7.4. Specifically, we display results for the point-proton rms radius and the proton–neutron radius difference of ^{8}B, as well as the squared proton-radius difference between ^{10}B and ^{11}B, obtained with both the EMN and SMS interactions at a cutoff of $\Lambda = 500\,\mathrm{MeV}$.

Considering first the point-proton radius R_p shown in the top row, both interaction families exhibit a reduction of the predicted radius at N^2LO, which is partially or fully compensated at the subsequent order. For the EMN[500] interaction, the NLO and N^3LO predictions are nearly identical, indicating good convergence of the chiral expansion. In contrast, the SMS[500] interaction displays a more pronounced dip at N^2LO, followed by the highest-order N^4LO^+ result that lies between the NLO and N^2LO values. The combination of larger order-to-order variations and the reduced suppression of higher-order terms at the highest effective order leads to a significantly larger interaction uncertainty for the SMS interaction.

The proton–neutron radius difference, $\Delta R_{p,n}$, exhibits broadly similar trends for both interaction families, although the N^2LO dip observed for the SMS interaction is less pronounced, resulting in a correspondingly smaller interaction uncertainty. Moreover, the values of $\Delta R_{p,n}$ show substantially reduced order-to-order variation compared to the individual absolute radii. This behavior is consistent with the expectation that taking differences removes correlated systematic effects that similarly affect proton and neutron radii.

Turning to the squared proton-radius difference between ^{10}B and ^{11}B, shown in the bottom row of Fig. 7.4, we observe a remarkable degree of stability for the EMN interaction across all chiral orders. This robustness reflects both the stability of the underlying individual radius predictions and the additional cancellation of systematic effects inherent in isotope-shift observables. The SMS interaction yields comparably stable results, with only a mild reduction at N^2LO. As a

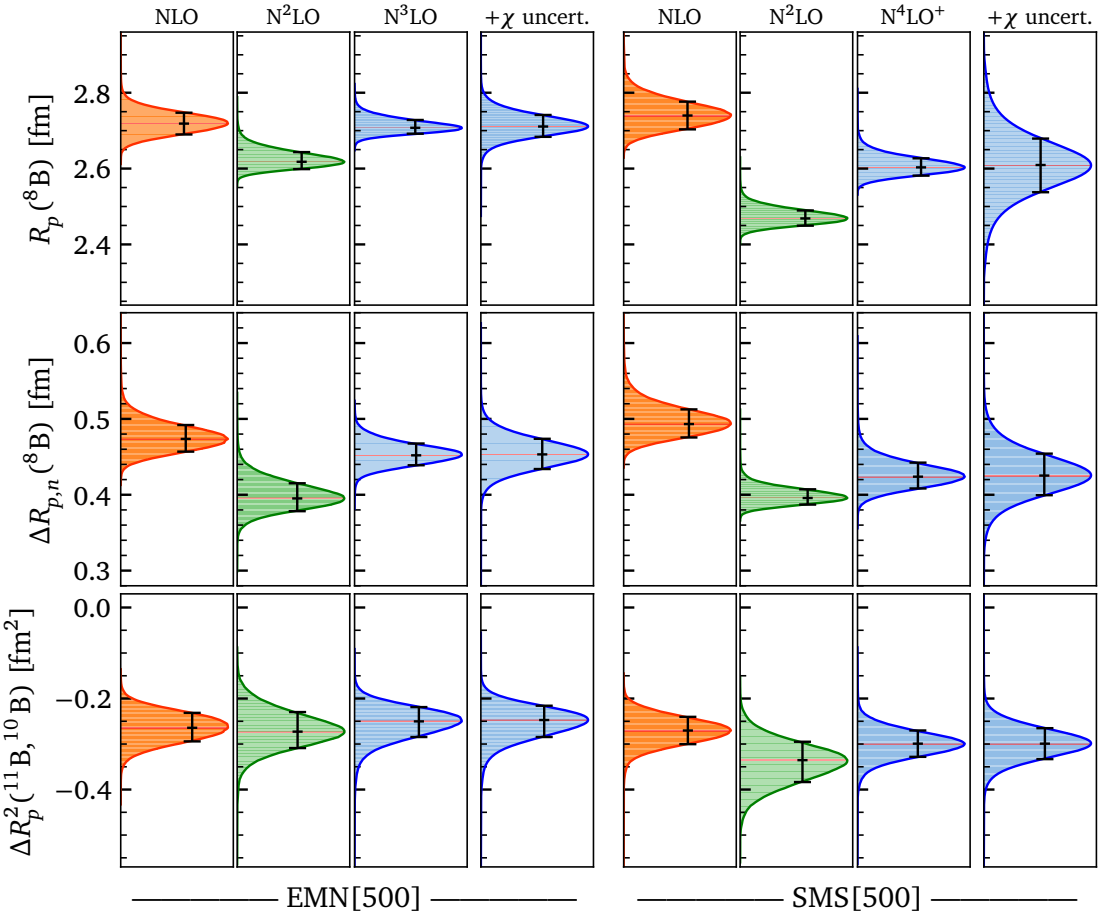


Figure 7.4: Order-by-order interaction uncertainty analysis for the point-proton rms radius (top row) and the proton–neutron radius difference (middle row) of ^{8}B as well as the squared proton radius difference between ^{10}B and ^{11}B (bottom row) for the EMN[500] and SMS[500] interactions. From left to right, the columns show the results for increasing orders of the interaction, and the right-most column for both interactions show the network predictions at the highest order combined with the order-by-order uncertainty from the BUQEYE model.

consequence, the total uncertainties for this observable are similar for both interaction families and are dominated by the many-body uncertainties rather than by the chiral truncation error.

All individual predictions at each chiral order, as well as the combined results including chiral uncertainty estimates for all four interaction variants, are summarized in Chapter A.

7.4 Results and Discussion

In the following, we present the final predictions obtained with our ANN-based method developed in this work for all observables, nuclei, and interaction variants considered. For each quantity, we quote a single consolidated result together with its total uncertainty, obtained by combining the many-body uncertainties inferred from the Gen2+ ANNs with the chiral order-by-order uncertainties discussed in the previous section.

7.4.1 Ground-State Energies

Before turning to the central results of this study, namely the radii of the boron isotopes, we first examine the predicted ground-state energies. Ground-state energies are of intrinsic interest and, moreover, provide an important consistency check for the quality of the underlying nuclear wave functions. Significant deviations from experimental binding energies would indicate deficiencies in the description of the eigenstates, which would in turn undermine the reliability of derived observables such as nuclear radii. Even more modest discrepancies can be informative, as systematic overbinding or underbinding is often correlated with corresponding trends toward smaller or larger radii, respectively, and guide the interpretation of the results.

The predicted ground-state energies for all nuclei considered, obtained with both interaction families and the two cutoff values $\Lambda = 450$ and $500\,\mathrm{MeV}$, are shown in Fig. 7.5. The corresponding numerical values are listed in Table A.1 of Chapter A. Overall, the calculated energies are in very good agreement with experimental data, with a slight but systematic tendency towards underbinding. As expected, the associated uncertainties increase with mass number, reflecting the growing model-space extrapolation uncertainties of the underlying NCSM calculations for heavier nuclei.

Among the interaction variants, the EMN[450] and SMS[500] interactions exhibit a modest accuracy advantage, particularly for the heavier nuclei starting from ^{10}B. Nevertheless, all interactions are mutually consistent within uncertainties, and no pronounced outliers are observed. Taken together, these results demonstrate that the ground-state energies of the nuclei under consideration are reproduced with sufficient accuracy, lending confidence that any visible feature or

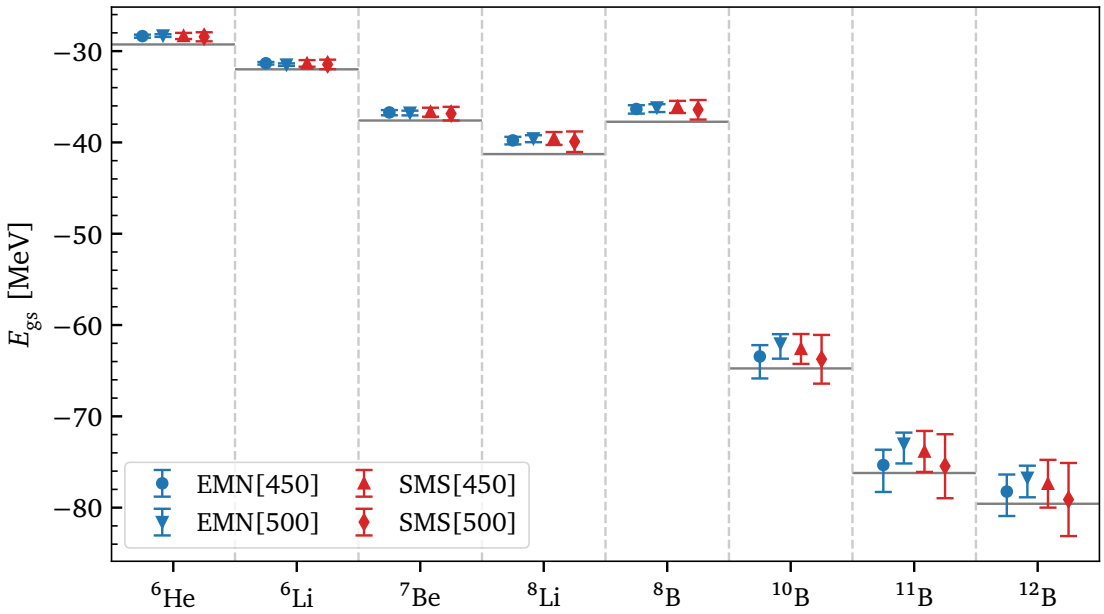


Figure 7.5: Overview of the ANN predictions for the ground-state energy of various nuclei including order-by-order uncertainties for two chiral interaction families with two cutoffs. The gray lines indicate experimental results taken from Ref. [97], the associated uncertainties are too small to be visible at this scale.

deviations in the radius predictions discussed below are not a mere consequence of an inadequate description of the ground-state energy.

7.4.2 Proton and Neutron Radii

We now turn to the proton and neutron radius predictions. An overview of the results for all nuclei and all four interaction variants is shown in Fig. 7.6. Just like for the ground-state energies, the corresponding numerical values are listed in Chapter A in Tables A.2 and A.3. The top and middle panels display the point-proton and point-neutron rms radii, respectively, with uncertainties indicated by error bars. Where available, experimental values from Ref. [139] are shown in gray and converted to point-proton radii using Eq. (1.23).

In contrast to the ground-state energies, the radii exhibit a stronger dependence on the choice of interaction. In particular, the SMS interactions consistently predict smaller proton and neutron radii than the EMN interactions. Interestingly, the cutoff dependence differs qualitatively between the two interaction families. For the EMN interactions, increasing the cutoff from $\Lambda = 450\,\text{MeV}$ to $500\,\text{MeV}$ leads to slightly larger radii, whereas the opposite trend is observed for the SMS interactions. In most cases, the $\Lambda = 450\,\text{MeV}$ variants of both interaction families

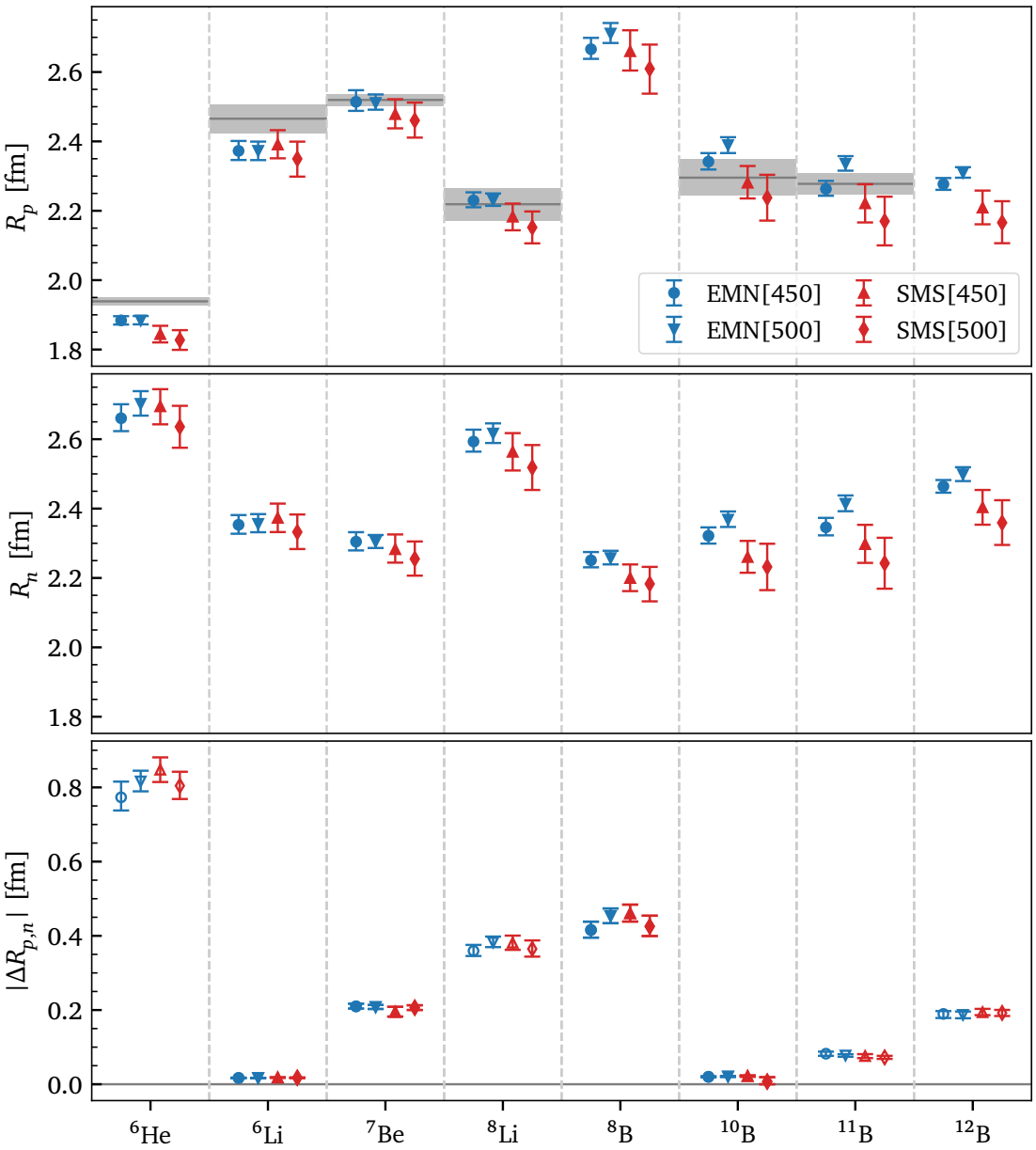


Figure 7.6: Overview of the ANN predictions for the point-proton rms radius (top panel), the point-neutron rms radius (center panel), and the absolute value of their difference (bottom panel) of various nuclei including order-by-order uncertainties for two chiral interaction families with two cutoffs. Open symbols in the bottom panel indicate that the proton radius is smaller than the neutron radius. Where available, the gray lines and shaded intervals indicate experimental results and uncertainties taken from Ref. [139].

agree within uncertainties, while the 500 MeV results show more pronounced differences. Comparison with experimental proton radii reveals generally good agreement within uncertainties, with the exception of a slight underprediction for ^{6}He and ^{6}Li. Overall, the EMN[450] and SMS[450] interactions provide the best

reproduction of the available experimental data. No experimental constraints are currently available for the neutron radii.

Additional insight into the structure of these nuclei can be gained by considering the proton–neutron radius difference, $\Delta R_{p,n}$, shown in the bottom panel of Fig. 7.6. Notably, the agreement between different interactions is substantially improved for this observable compared to the individual radii. At the same time, the uncertainties are significantly reduced, reflecting the strong correlations between proton and neutron radii that are efficiently exploited by our evaluation scheme. Tracing the $N = 3$ isotonic chain from ^{6}Li to ^{7}Be and finally to ^{8}B, where one proton is added at each step, we observe an approximately linear increase of $\Delta R_{p,n}$ from near zero to about 0.45 fm in ^{8}B. This trend arises from a simultaneous increase of the proton radius and a slight decrease of the neutron radius as the proton excess grows. In contrast, the heavier boron isotopes do not display such pronounced differences. The nearly symmetric nucleus ^{10}B exhibits an almost vanishing $\Delta R_{p,n}$, while the modest neutron excess in ^{11}B and ^{12}B is reflected in slightly larger neutron radii.

A particularly striking feature is the sharp reduction of $\Delta R_{p,n}$ between ^{8}B and ^{10}B, which is driven primarily by a substantial decrease in the proton radius. The absolute value of $\Delta R_{p,n}$ is slightly larger in ^{8}B than in its mirror nucleus ^{8}Li with an analogous neutron excess. This asymmetry can plausibly be attributed to additional Coulomb repulsion in the proton-rich system. Although the proton–neutron radius difference in ^{8}B is smaller than that of the well-established halo nucleus ^{6}He, the absolute proton radius of ^{8}B is comparable to the neutron radius of ^{6}He. Taken together, these findings provide indications for, and are fully consistent with, the presence of a proton halo in ^{8}B. A definitive theoretical confirmation, however, would require a more targeted analysis, for instance based on spatial density distributions.

7.4.3 Radius Differences Between Isotopes

Finally, we present high-precision theoretical predictions for isotope-dependent radius differences, focusing on squared point-proton radius differences ΔR_p^2. These observables are particularly well suited for comparison with experiment, as the corresponding squared charge-radius differences can be measured with very high precision. In particular, isotope-shift measurements provide access to these differences with significantly smaller uncertainties than those achievable for absolute radii [95, 96].

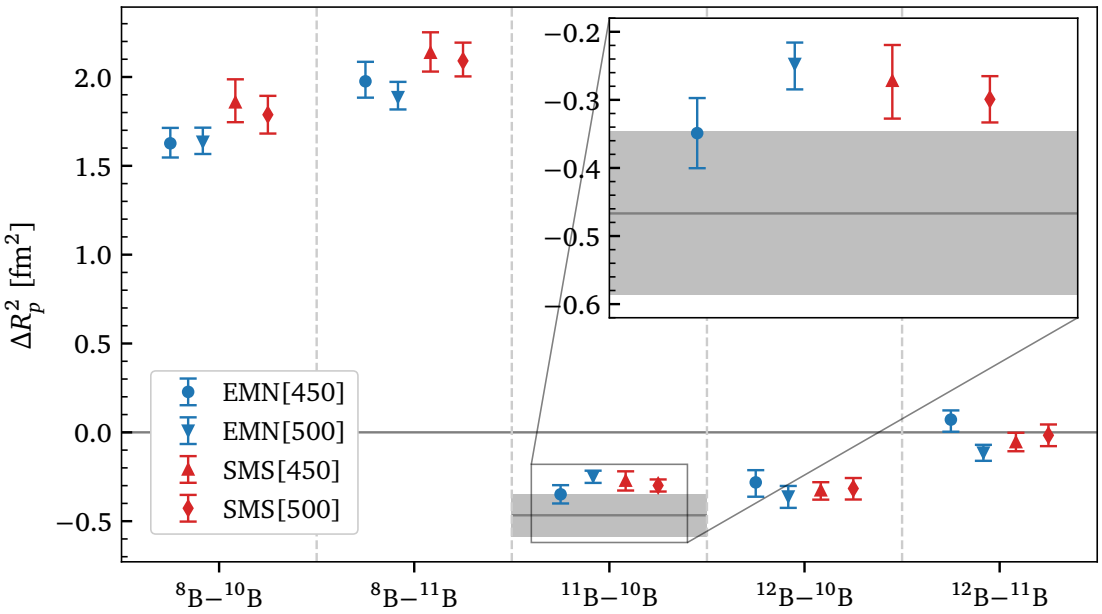


Figure 7.7: Squared point-proton rms differences between boron isotopes for different chiral interactions. The inset provides a magnified view of the $^{11}\mathrm{B}-{}^{10}\mathrm{B}$ difference, which is currently the only case for which an experimental value [95] is available for comparison.

The calculated values of ΔR_p^2 for various pairs of boron isotopes are shown in Fig. 7.7. Following current and planned experimental strategies [94, 95], the stable isotopes ^{10}B and ^{11}B are used as reference points for the other isotopes. Overall, the results show a high degree of consistency across all interaction variants. The proton radius of ^{8}B is found to be significantly larger than those of the heavier boron isotopes, again supporting the interpretation of a halo-like structure. The remaining boron isotopes cluster more closely, although ^{11}B and ^{12}B appear noticeably more compact than ^{10}B.

For the squared proton-radius difference between ^{10}B and ^{11}B, we additionally compare our predictions with the only currently available experimental result, taken from Ref. [95] and converted to a squared point-proton radius difference using the current PDG value for the neutron charge radius squared [90]. Our theoretical predictions are close to the experimental value, with the EMN[450] interaction being fully compatible within uncertainties. While the remaining interaction variants show somewhat larger deviations, these are not statistically significant enough to conclusively rule them out. Owing to the strong correlations exploited in our analysis, the resulting theoretical uncertainties are highly competitive with current experimental precision and substantially smaller than previously reported theoretical estimates [95].

Conclusion and Outlook

Conclusion

In this thesis, we developed a machine-learning–based framework to robustly extract fully-converged *ab initio* results from NCSM calculations with incomplete convergence. Using a large set of Jacobi-NCSM calculations for few-body systems, specifically ^{2}H, ^{3}H, and ^{4}He, where fully-converged results for both ground-state energies and radii can be obtained explicitly, we trained ANNs to predict the infinite model-space limits from partially converged input data. Building on these networks, we introduced a statistical evaluation scheme that yields a distribution of predictions, from which both the converged value and a well-defined many-body and extrapolation uncertainty can be extracted.

A key improvement for the networks' performance is the normalization of the network input samples, which substantially improves the quality of the predictions and enables the treatment of nuclear radii as a new class of observables alongside energies. We demonstrated that the resulting Gen2 networks, particularly when combined with a selective evaluation scheme (Gen2+), consistently outperform commonly used extrapolation techniques. This includes both traditional extrapolation methods and an existing ANN-based approach, with the largest performance gains observed in smaller model spaces where convergence is most limited. In addition, a special post-processing procedure allowed us to compute differences between correlated observables, such as proton and neutron radii within a nucleus or proton radii between isotopes, with markedly enhanced precision by exploiting the strong correlations between closely related observables. The many-body uncertainties extracted from the network predictions were further combined in a robust manner with chiral order-by-order interaction uncertainties, treating both sources of uncertainties on equal footing.

Applying this framework to a comprehensive set of NCSM calculations for boron isotopes with two state-of-the-art families of chiral interactions, the EMN

and SMS interactions, we produced high-precision *ab initio* predictions for ground-state energies and nuclear radii. Both interaction families yield stable and largely consistent results across all nuclei considered, with only minor deviations from experimental data in selected cases. By analyzing proton and neutron radii across isotopic and isotonic chains, as well as by comparison with the mirror nucleus, we identified clear signatures consistent with the presence of a proton halo in ^{8}B. While these signatures are less pronounced than those observed for the well-established neutron halo nucleus ^{6}He, they strongly suggest that ^{8}B is a compelling candidate for a proton halo and merits further theoretical and experimental investigation.

Finally, we provided high-precision predictions for squared proton radius differences between boron isotopes, including the experimentally measured ^{10,11}B pair. Our results are competitive with, and in some cases comparable to, the precision of existing isotope-shift measurements, while substantially improving upon previously available theoretical estimates. Predictions for additional isotope pairs were also presented in anticipation of forthcoming high-precision experimental data. Taken together, the methodological developments and physics results presented in this thesis demonstrate the potential of machine-learning–assisted *ab initio* approaches to significantly extend the reach and precision of NCSM-based calculations in nuclear structure theory.

Outlook

While this work establishes a robust and versatile framework for model-space extrapolation based on artificial neural networks, several promising directions for further development remain. One natural avenue for improvement lies in tailoring the training data more closely to the eventual evaluation strategy. Preliminary investigations indicate that jointly optimizing the training procedure and the evaluation scheme could yield further gains in predictive accuracy and precision. In addition, there are opportunities for refining the network architecture itself, both through further hyperparameter optimization and by incorporating advanced techniques from machine-learning and neural network research. However, even without further improvements, there are still more applications for our networks in combination with further post-processing techniques as demonstrated in Refs. [140, 141], where our network architecture was used to predict the hyperon separation energy.

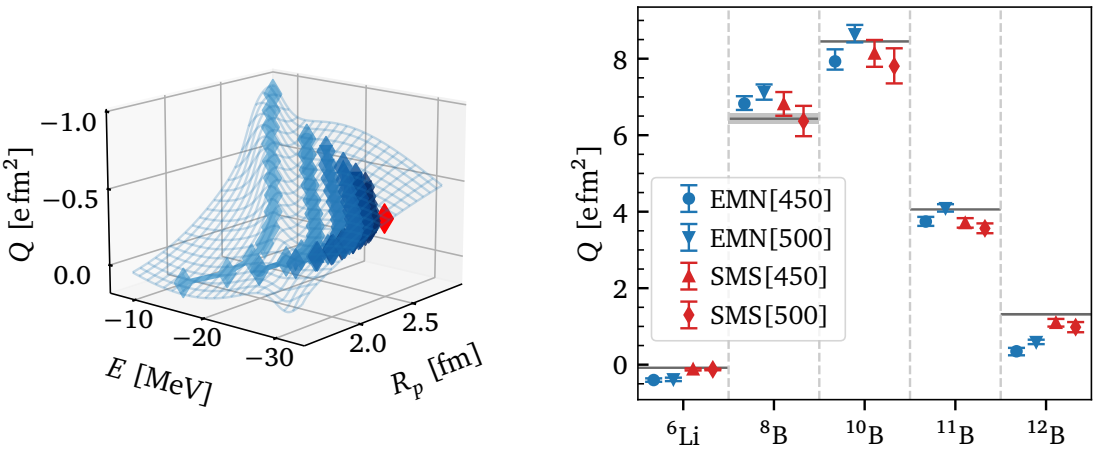


Figure 7.8: The wireframe in the left plot illustrates the OTN outputs for the electric quadrupole moment for each (E_{gs},R_p) input, showing the learnt hypersurface. The lines connect NCSM data points with the same N_{max} and varying a_{HO} are the training data. The red diamond is the network predicion for the converged value. A small selection of the final results including an order-by-order uncertainty analysis presented in Ref. [130] is shown in the right panel.

Another promising direction for future development is the generation of artificial training data. Such data could provide a more diverse and systematically controlled training set for our current ANN architecture, while simultaneously addressing the limited availability of suitable training data for certain observables in few-body systems.

A particularly promising extension of the ANN-based extrapolation paradigm is the concept of observable transcoder networks (OTNs). In this approach, networks are trained to model the correlations between multiple observables, specifically the ground-state energy, the point-proton radius, and another observable, for example electromagnetic moments. Once the network has learned to interpolate the hypersurface spanned by these quantities, as illustrated in the left panel of Fig. 7.8, the converged value of the third observable can be obtained by evaluating the network at the converged energy and radius predicted by the networks developed in this work. Within this framework, the procedure corresponds to an interpolation in the radius and a small, well-controlled extrapolation in the ground-state energy.

We have demonstrated the effectiveness of this architecture in Ref. [130], where OTNs were applied to predict electric quadrupole moments Q for a range of nuclei, exploiting the strong correlation between R_p and Q [142, 143] and using E_{gs} as a proxy for the convergence. A representative subset of those results, including a full order-by-order uncertainty analysis, is shown in the

right panel of Fig. 7.8. In that study, the OTN approach not only outperformed a direct extrapolation strategy based on networks similar to those employed in the present work, but also proved more flexible. In particular, it does not rely on the availability of fully converged training data in few-body systems, which can be a significant limitation for some observables. This flexibility is especially advantageous for electromagnetic observables, many of which vanish in very light nuclei such as ^{2}H, ^{3}H, or ^{4}He, or lack suitable transition states. Extending the OTN framework to a broader class of electromagnetic observables, including transition strengths, therefore represents a natural next step in advancing the neural-network–enhanced *ab initio* methodologies introduced in this thesis.

A

Numerical Boron ANN Results

The following tables list the network predictions for the ground-state energy, the point-proton and the point-neutron rms radius and its difference for a variety of nuclei as well as the proton radius difference between boron isotopes. The tables are split by interaction family, and each interaction is calculated with two different cutoffs $\Lambda = 450$ and 500 MeV.

	non-local (EMN) $\Lambda = 450\,\mathrm{MeV}$				non-local (EMN) $\Lambda = 500\,\mathrm{MeV}$			
Observable	NLO	N^{2}LO	N^{3}LO	$+\chi$ uncert.	NLO	N^{2}LO	N^{3}LO	$+\chi$ uncert.
E_{gs} (^{6}He)	$-28.894^{+0.033}_{-0.041}$	$-28.874^{+0.057}_{-0.061}$	$-28.343^{+0.065}_{-0.080}$	$-28.357^{+0.162}_{-0.171}$	$-27.743^{+0.039}_{-0.055}$	$-28.540^{+0.058}_{-0.062}$	$-28.285^{+0.063}_{-0.077}$	$-28.294^{+0.142}_{-0.151}$
E_{gs} (^{6}Li)	$-31.819^{+0.032}_{-0.039}$	$-31.532^{+0.051}_{-0.071}$	$-31.317^{+0.058}_{-0.081}$	$-31.331^{+0.141}_{-0.146}$	$-30.686^{+0.042}_{-0.044}$	$-31.376^{+0.057}_{-0.070}$	$-31.471^{+0.064}_{-0.068}$	$-31.480^{+0.139}_{-0.143}$
E_{gs} (^{7}Be)	$-37.295^{+0.065}_{-0.080}$	$-37.236^{+0.158}_{-0.193}$	$-36.715^{+0.187}_{-0.229}$	$-36.715^{+0.257}_{-0.312}$	$-35.792^{+0.081}_{-0.093}$	$-36.872^{+0.150}_{-0.198}$	$-36.736^{+0.163}_{-0.215}$	$-36.749^{+0.228}_{-0.293}$
E_{gs} (^{8}Li)	$-40.563^{+0.068}_{-0.095}$	$-40.981^{+0.180}_{-0.220}$	$-39.745^{+0.208}_{-0.291}$	$-39.778^{+0.391}_{-0.440}$	$-38.618^{+0.084}_{-0.124}$	$-40.443^{+0.182}_{-0.222}$	$-39.529^{+0.187}_{-0.268}$	$-39.562^{+0.349}_{-0.414}$
E_{gs} (^{8}B)	$-37.050^{+0.084}_{-0.103}$	$-37.515^{+0.191}_{-0.253}$	$-36.296^{+0.247}_{-0.354}$	$-36.339^{+0.419}_{-0.505}$	$-35.194^{+0.098}_{-0.119}$	$-36.961^{+0.188}_{-0.286}$	$-36.104^{+0.221}_{-0.408}$	$-36.174^{+0.361}_{-0.501}$
E_{gs} (^{10}B)	$-64.040^{+0.159}_{-0.181}$	$-65.696^{+1.130}_{-1.776}$	$-63.439^{+1.237}_{-2.149}$	$-63.439^{+1.237}_{-2.409}$	$-60.291^{+0.223}_{-0.209}$	$-64.178^{+0.813}_{-1.497}$	$-62.030^{+0.851}_{-1.389}$	$-62.030^{+1.030}_{-1.657}$
E_{gs} (^{11}B)	$-75.966^{+0.164}_{-0.164}$	$-78.472^{+1.433}_{-2.115}$	$-75.331^{+1.357}_{-2.793}$	$-75.331^{+1.676}_{-2.953}$	$-71.362^{+0.223}_{-0.223}$	$-76.084^{+1.077}_{-1.693}$	$-72.996^{+1.003}_{-1.637}$	$-72.996^{+1.214}_{-2.165}$
E_{gs} (^{12}B)	$-81.751^{+0.187}_{-0.165}$	$-81.239^{+1.489}_{-2.007}$	$-78.239^{+1.542}_{-2.679}$	$-78.239^{+1.867}_{-2.679}$	$-76.191^{+0.184}_{-0.225}$	$-79.556^{+1.392}_{-2.054}$	$-76.731^{+1.096}_{-1.788}$	$-76.731^{+1.327}_{-2.134}$
	semi-local (SMS) $\Lambda = 450\,\mathrm{MeV}$				semi-local (SMS) $\Lambda = 500\,\mathrm{MeV}$			
Observable	NLO	N^{2}LO	N^{4}LO+	$+\chi$ uncert.	NLO	N^{2}LO	N^{4}LO+	$+\chi$ uncert.
E_{gs} (^{6}He)	$-28.836^{+0.044}_{-0.054}$	$-29.044^{+0.045}_{-0.055}$	$-28.329^{+0.049}_{-0.056}$	$-28.339^{+0.335}_{-0.338}$	$-27.371^{+0.044}_{-0.054}$	$-29.216^{+0.046}_{-0.056}$	$-28.417^{+0.048}_{-0.059}$	$-28.432^{+0.495}_{-0.495}$
E_{gs} (^{6}Li)	$-31.917^{+0.043}_{-0.060}$	$-32.040^{+0.051}_{-0.054}$	$-31.333^{+0.053}_{-0.057}$	$-31.344^{+0.361}_{-0.364}$	$-30.444^{+0.049}_{-0.060}$	$-32.295^{+0.041}_{-0.057}$	$-31.439^{+0.051}_{-0.063}$	$-31.454^{+0.518}_{-0.522}$
E_{gs} (^{7}Be)	$-37.124^{+0.113}_{-0.121}$	$-37.724^{+0.112}_{-0.119}$	$-36.636^{+0.119}_{-0.145}$	$-36.671^{+0.471}_{-0.515}$	$-35.254^{+0.110}_{-0.135}$	$-38.085^{+0.102}_{-0.116}$	$-36.794^{+0.116}_{-0.141}$	$-36.846^{+0.740}_{-0.749}$
E_{gs} (^{8}Li)	$-39.490^{+0.122}_{-0.149}$	$-41.394^{+0.117}_{-0.133}$	$-39.508^{+0.133}_{-0.162}$	$-39.557^{+0.693}_{-0.722}$	$-37.283^{+0.131}_{-0.160}$	$-42.008^{+0.109}_{-0.133}$	$-39.867^{+0.131}_{-0.170}$	$-39.916^{+1.114}_{-1.134}$
E_{gs} (^{8}B)	$-35.982^{+0.134}_{-0.188}$	$-37.736^{+0.129}_{-0.167}$	$-36.038^{+0.183}_{-0.196}$	$-36.088^{+0.639}_{-0.689}$	$-33.850^{+0.161}_{-0.225}$	$-38.322^{+0.130}_{-0.148}$	$-36.340^{+0.154}_{-0.177}$	$-36.408^{+1.055}_{-1.089}$
E_{gs} (^{10}B)	$-61.322^{+0.269}_{-0.269}$	$-67.044^{+0.241}_{-0.295}$	$-62.502^{+0.291}_{-0.355}$	$-62.609^{+1.626}_{-1.647}$	$-57.254^{+0.265}_{-0.302}$	$-69.003^{+0.272}_{-0.255}$	$-63.667^{+0.345}_{-0.393}$	$-63.739^{+2.655}_{-2.679}$
E_{gs} (^{11}B)	$-72.387^{+0.239}_{-0.292}$	$-80.422^{+0.246}_{-0.300}$	$-73.787^{+0.336}_{-0.382}$	$-73.857^{+2.256}_{-2.233}$	$-67.478^{+0.270}_{-0.270}$	$-82.863^{+0.246}_{-0.263}$	$-75.373^{+0.394}_{-0.368}$	$-75.449^{+3.491}_{-3.516}$
E_{gs} (^{12}B)	$-76.152^{+0.241}_{-0.295}$	$-85.348^{+0.348}_{-0.348}$	$-77.243^{+0.374}_{-0.457}$	$-77.381^{+2.616}_{-2.616}$	$-70.542^{+0.278}_{-0.278}$	$-87.978^{+0.284}_{-0.348}$	$-78.992^{+0.461}_{-0.461}$	$-79.111^{+4.003}_{-4.003}$

Table A.1: Numerical values of the predicted ground-state energy using our Gen2+ ANNs for the EMN and SMS interactions. Results are given in units of MeV for each order of the respective interaction and for the highest order combined with interactions uncertainties.

Observable	non-local (EMN) $\Lambda = 450\,\mathrm{MeV}$ NLO	$\mathrm{N^2LO}$	$\mathrm{N^3LO}$	$+\chi$ uncert.	non-local (EMN) $\Lambda = 500\,\mathrm{MeV}$ NLO	$\mathrm{N^2LO}$	$\mathrm{N^3LO}$	$+\chi$ uncert.
R_p (^{6}He)	$1.843^{+0.008}_{-0.008}$	$1.876^{+0.010}_{-0.008}$	$1.883^{+0.010}_{-0.010}$	$1.884^{+0.012}_{-0.012}$	$1.885^{+0.008}_{-0.009}$	$1.864^{+0.009}_{-0.009}$	$1.883^{+0.009}_{-0.009}$	$1.884^{+0.013}_{-0.011}$
R_p (^{6}Li)	$2.389^{+0.017}_{-0.017}$	$2.371^{+0.021}_{-0.023}$	$2.370^{+0.026}_{-0.024}$	$2.373^{+0.029}_{-0.026}$	$2.441^{+0.018}_{-0.018}$	$2.350^{+0.027}_{-0.020}$	$2.371^{+0.025}_{-0.022}$	$2.373^{+0.027}_{-0.027}$
R_p (^{7}Be)	$2.462^{+0.026}_{-0.024}$	$2.508^{+0.027}_{-0.025}$	$2.514^{+0.029}_{-0.026}$	$2.514^{+0.033}_{-0.026}$	$2.527^{+0.022}_{-0.024}$	$2.490^{+0.023}_{-0.014}$	$2.511^{+0.019}_{-0.018}$	$2.511^{+0.025}_{-0.019}$
R_p (^{8}Li)	$2.161^{+0.011}_{-0.014}$	$2.211^{+0.021}_{-0.013}$	$2.228^{+0.021}_{-0.016}$	$2.230^{+0.023}_{-0.020}$	$2.208^{+0.012}_{-0.015}$	$2.202^{+0.018}_{-0.014}$	$2.229^{+0.017}_{-0.013}$	$2.231^{+0.019}_{-0.017}$
R_p (^{8}B)	$2.620^{+0.031}_{-0.028}$	$2.618^{+0.032}_{-0.025}$	$2.663^{+0.030}_{-0.023}$	$2.666^{+0.033}_{-0.028}$	$2.719^{+0.029}_{-0.029}$	$2.618^{+0.026}_{-0.019}$	$2.708^{+0.021}_{-0.016}$	$2.711^{+0.030}_{-0.027}$
R_p (^{10}B)	$2.257^{+0.019}_{-0.016}$	$2.310^{+0.020}_{-0.015}$	$2.340^{+0.021}_{-0.019}$	$2.342^{+0.025}_{-0.023}$	$2.344^{+0.017}_{-0.015}$	$2.319^{+0.015}_{-0.010}$	$2.386^{+0.017}_{-0.015}$	$2.389^{+0.023}_{-0.023}$
R_p (^{11}B)	$2.208^{+0.014}_{-0.011}$	$2.237^{+0.015}_{-0.020}$	$2.261^{+0.021}_{-0.016}$	$2.263^{+0.023}_{-0.020}$	$2.287^{+0.015}_{-0.013}$	$2.262^{+0.016}_{-0.011}$	$2.332^{+0.015}_{-0.010}$	$2.335^{+0.022}_{-0.020}$
R_p (^{12}B)	$2.142^{+0.013}_{-0.010}$	$2.271^{+0.009}_{-0.012}$	$2.276^{+0.012}_{-0.011}$	$2.277^{+0.018}_{-0.016}$	$2.219^{+0.013}_{-0.011}$	$2.270^{+0.012}_{-0.010}$	$2.309^{+0.010}_{-0.008}$	$2.310^{+0.016}_{-0.015}$
$\Delta R_{p,n}$ (^{6}He)	$-0.863^{+0.021}_{-0.021}$	$-0.737^{+0.031}_{-0.040}$	$-0.773^{+0.032}_{-0.042}$	$-0.773^{+0.036}_{-0.042}$	$-0.866^{+0.023}_{-0.023}$	$-0.763^{+0.020}_{-0.027}$	$-0.816^{+0.022}_{-0.027}$	$-0.816^{+0.027}_{-0.029}$
$\Delta R_{p,n}$ (^{6}Li)	$0.019^{+0.001}_{-0.001}$	$0.017^{+0.001}_{-0.001}$	$0.017^{+0.001}_{-0.001}$	$0.017^{+0.001}_{-0.001}$	$0.018^{+0.001}_{-0.001}$	$0.017^{+0.001}_{-0.001}$	$0.017^{+0.001}_{-0.001}$	$0.017^{+0.001}_{-0.001}$
$\Delta R_{p,n}$ (^{7}Be)	$0.192^{+0.019}_{-0.019}$	$0.205^{+0.007}_{-0.005}$	$0.209^{+0.007}_{-0.006}$	$0.210^{+0.007}_{-0.006}$	$0.193^{+0.016}_{-0.019}$	$0.203^{+0.009}_{-0.015}$	$0.208^{+0.006}_{-0.005}$	$0.208^{+0.006}_{-0.005}$
$\Delta R_{p,n}$ (^{8}Li)	$-0.370^{+0.017}_{-0.014}$	$-0.325^{+0.009}_{-0.014}$	$-0.358^{+0.011}_{-0.014}$	$-0.359^{+0.014}_{-0.016}$	$-0.383^{+0.014}_{-0.018}$	$-0.336^{+0.008}_{-0.013}$	$-0.381^{+0.008}_{-0.011}$	$-0.382^{+0.012}_{-0.016}$
$\Delta R_{p,n}$ (^{8}B)	$0.448^{+0.020}_{-0.019}$	$0.389^{+0.015}_{-0.012}$	$0.416^{+0.021}_{-0.021}$	$0.416^{+0.022}_{-0.021}$	$0.474^{+0.018}_{-0.017}$	$0.395^{+0.020}_{-0.017}$	$0.452^{+0.016}_{-0.013}$	$0.453^{+0.021}_{-0.019}$
$\Delta R_{p,n}$ (^{10}B)	$0.021^{+0.001}_{-0.001}$	$0.019^{+0.001}_{-0.001}$	$0.020^{+0.001}_{-0.001}$	$0.020^{+0.001}_{-0.001}$	$0.022^{+0.001}_{-0.001}$	$0.019^{+0.001}_{-0.001}$	$0.021^{+0.001}_{-0.001}$	$0.021^{+0.001}_{-0.001}$
$\Delta R_{p,n}$ (^{11}B)	$-0.052^{+0.012}_{-0.014}$	$-0.072^{+0.005}_{-0.006}$	$-0.082^{+0.005}_{-0.006}$	$-0.082^{+0.006}_{-0.006}$	$-0.068^{+0.002}_{-0.001}$	$-0.074^{+0.005}_{-0.005}$	$-0.077^{+0.003}_{-0.004}$	$-0.077^{+0.003}_{-0.004}$
$\Delta R_{p,n}$ (^{12}B)	$-0.146^{+0.010}_{-0.013}$	$-0.186^{+0.009}_{-0.007}$	$-0.190^{+0.010}_{-0.007}$	$-0.189^{+0.011}_{-0.008}$	$-0.167^{+0.007}_{-0.008}$	$-0.183^{+0.007}_{-0.007}$	$-0.188^{+0.009}_{-0.009}$	$-0.187^{+0.009}_{-0.009}$
ΔR_p^2 (^{8,10}B)	$1.786^{+0.120}_{-0.109}$	$1.523^{+0.105}_{-0.115}$	$1.627^{+0.073}_{-0.080}$	$1.620^{+0.073}_{-0.094}$	$1.893^{+0.111}_{-0.111}$	$1.474^{+0.074}_{-0.081}$	$1.630^{+0.054}_{-0.064}$	$1.630^{+0.069}_{-0.079}$
ΔR_p^2 (^{8,11}B)	$1.980^{+0.127}_{-0.167}$	$1.864^{+0.120}_{-0.157}$	$1.976^{+0.092}_{-0.101}$	$1.967^{+0.083}_{-0.118}$	$2.159^{+0.128}_{-0.128}$	$1.741^{+0.077}_{-0.084}$	$1.881^{+0.053}_{-0.069}$	$1.881^{+0.069}_{-0.086}$
ΔR_p^2 (^{10,11}B)	$-0.202^{+0.049}_{-0.070}$	$-0.351^{+0.050}_{-0.050}$	$-0.349^{+0.052}_{-0.052}$	$-0.349^{+0.052}_{-0.052}$	$-0.264^{+0.033}_{-0.030}$	$-0.273^{+0.043}_{-0.036}$	$-0.250^{+0.031}_{-0.034}$	$-0.247^{+0.031}_{-0.037}$
ΔR_p^2 (^{12,10}B)	$-0.483^{+0.048}_{-0.064}$	$-0.185^{+0.056}_{-0.061}$	$-0.282^{+0.057}_{-0.075}$	$-0.282^{+0.069}_{-0.081}$	$-0.567^{+0.048}_{-0.053}$	$-0.231^{+0.045}_{-0.054}$	$-0.361^{+0.046}_{-0.051}$	$-0.361^{+0.059}_{-0.064}$
ΔR_p^2 (^{12,11}B)	$-0.282^{+0.023}_{-0.023}$	$0.178^{+0.033}_{-0.060}$	$0.080^{+0.031}_{-0.056}$	$0.072^{+0.052}_{-0.068}$	$-0.306^{+0.024}_{-0.029}$	$0.040^{+0.023}_{-0.030}$	$-0.112^{+0.021}_{-0.023}$	$-0.114^{+0.044}_{-0.046}$

Table A.2: Same as Table A.1 but for radius results of the EMN interactions. All results are in units of fm except for ΔR_p^2 which are in fm^2.

	semi-local (SMS) $\Lambda = 450\,\mathrm{MeV}$				semi-local (SMS) $\Lambda = 500\,\mathrm{MeV}$			
Observable	NLO	$\mathrm{N^2LO}$	$\mathrm{N^4LO+}$	$+\chi$ uncert.	NLO	$\mathrm{N^2LO}$	$\mathrm{N^4LO+}$	$+\chi$ uncert.
R_p (^{6}He)	$1.818^{+0.010}_{-0.011}$	$1.793^{+0.007}_{-0.007}$	$1.844^{+0.009}_{-0.008}$	$1.844^{+0.024}_{-0.024}$	$1.857^{+0.012}_{-0.009}$	$1.776^{+0.007}_{-0.007}$	$1.827^{+0.008}_{-0.008}$	$1.827^{+0.028}_{-0.028}$
R_p (^{6}Li)	$2.388^{+0.021}_{-0.021}$	$2.301^{+0.016}_{-0.020}$	$2.393^{+0.017}_{-0.020}$	$2.391^{+0.041}_{-0.040}$	$2.430^{+0.021}_{-0.021}$	$2.256^{+0.018}_{-0.016}$	$2.350^{+0.017}_{-0.020}$	$2.350^{+0.050}_{-0.051}$
R_p (^{7}Be)	$2.458^{+0.029}_{-0.027}$	$2.395^{+0.022}_{-0.020}$	$2.474^{+0.025}_{-0.023}$	$2.479^{+0.043}_{-0.041}$	$2.507^{+0.033}_{-0.031}$	$2.358^{+0.021}_{-0.016}$	$2.454^{+0.025}_{-0.019}$	$2.460^{+0.051}_{-0.049}$
R_p (^{8}Li)	$2.155^{+0.015}_{-0.017}$	$2.077^{+0.013}_{-0.012}$	$2.183^{+0.013}_{-0.017}$	$2.183^{+0.038}_{-0.039}$	$2.205^{+0.019}_{-0.018}$	$2.048^{+0.013}_{-0.011}$	$2.150^{+0.014}_{-0.013}$	$2.152^{+0.046}_{-0.046}$
R_p (^{8}B)	$2.686^{+0.029}_{-0.034}$	$2.525^{+0.024}_{-0.022}$	$2.657^{+0.026}_{-0.028}$	$2.660^{+0.060}_{-0.056}$	$2.740^{+0.037}_{-0.037}$	$2.469^{+0.021}_{-0.019}$	$2.603^{+0.024}_{-0.022}$	$2.609^{+0.070}_{-0.072}$
R_p (^{10}B)	$2.271^{+0.026}_{-0.019}$	$2.148^{+0.016}_{-0.014}$	$2.278^{+0.019}_{-0.016}$	$2.281^{+0.048}_{-0.045}$	$2.374^{+0.025}_{-0.023}$	$2.105^{+0.012}_{-0.012}$	$2.233^{+0.015}_{-0.012}$	$2.238^{+0.066}_{-0.066}$
R_p (^{11}B)	$2.231^{+0.020}_{-0.016}$	$2.059^{+0.012}_{-0.008}$	$2.217^{+0.017}_{-0.015}$	$2.221^{+0.055}_{-0.055}$	$2.315^{+0.022}_{-0.020}$	$2.022^{+0.010}_{-0.008}$	$2.166^{+0.015}_{-0.013}$	$2.170^{+0.071}_{-0.070}$
R_p (^{12}B)	$2.160^{+0.016}_{-0.015}$	$2.043^{+0.010}_{-0.008}$	$2.206^{+0.014}_{-0.014}$	$2.209^{+0.049}_{-0.048}$	$2.240^{+0.020}_{-0.015}$	$2.013^{+0.009}_{-0.007}$	$2.164^{+0.013}_{-0.013}$	$2.166^{+0.062}_{-0.059}$
$\Delta R_{p,n}$ (^{6}He)	$-0.880^{+0.027}_{-0.032}$	$-0.795^{+0.019}_{-0.024}$	$-0.848^{+0.025}_{-0.025}$	$-0.848^{+0.033}_{-0.033}$	$-0.867^{+0.025}_{-0.030}$	$-0.751^{+0.017}_{-0.022}$	$-0.804^{+0.021}_{-0.023}$	$-0.804^{+0.036}_{-0.038}$
$\Delta R_{p,n}$ (^{6}Li)	$0.018^{+0.001}_{-0.001}$	$0.018^{+0.001}_{-0.001}$	$0.018^{+0.001}_{-0.001}$	$0.018^{+0.001}_{-0.001}$	$0.017^{+0.001}_{-0.001}$	$0.017^{+0.001}_{-0.001}$	$0.017^{+0.001}_{-0.001}$	$0.017^{+0.001}_{-0.001}$
$\Delta R_{p,n}$ (^{7}Be)	$0.206^{+0.005}_{-0.005}$	$0.205^{+0.005}_{-0.005}$	$0.194^{+0.014}_{-0.011}$	$0.195^{+0.014}_{-0.013}$	$0.188^{+0.021}_{-0.017}$	$0.202^{+0.005}_{-0.005}$	$0.207^{+0.004}_{-0.005}$	$0.206^{+0.007}_{-0.006}$
$\Delta R_{p,n}$ (^{8}Li)	$-0.405^{+0.018}_{-0.019}$	$-0.349^{+0.010}_{-0.010}$	$-0.380^{+0.012}_{-0.013}$	$-0.381^{+0.019}_{-0.019}$	$-0.436^{+0.018}_{-0.018}$	$-0.348^{+0.012}_{-0.014}$	$-0.364^{+0.010}_{-0.011}$	$-0.365^{+0.021}_{-0.023}$
$\Delta R_{p,n}$ (^{8}B)	$0.492^{+0.020}_{-0.020}$	$0.420^{+0.013}_{-0.011}$	$0.458^{+0.014}_{-0.014}$	$0.461^{+0.023}_{-0.022}$	$0.493^{+0.019}_{-0.018}$	$0.396^{+0.011}_{-0.009}$	$0.424^{+0.019}_{-0.016}$	$0.425^{+0.029}_{-0.026}$
$\Delta R_{p,n}$ (^{10}B)	$0.023^{+0.002}_{-0.002}$	$0.020^{+0.001}_{-0.001}$	$0.022^{+0.002}_{-0.001}$	$0.022^{+0.002}_{-0.002}$	$0.025^{+0.002}_{-0.002}$	$0.019^{+0.001}_{-0.001}$	$0.007^{+0.011}_{-0.013}$	$0.007^{+0.012}_{-0.013}$
$\Delta R_{p,n}$ (^{11}B)	$-0.046^{+0.014}_{-0.017}$	$-0.065^{+0.003}_{-0.003}$	$-0.076^{+0.003}_{-0.003}$	$-0.076^{+0.005}_{-0.005}$	$-0.073^{+0.003}_{-0.002}$	$-0.064^{+0.003}_{-0.003}$	$-0.073^{+0.003}_{-0.003}$	$-0.073^{+0.004}_{-0.004}$
$\Delta R_{p,n}$ (^{12}B)	$-0.167^{+0.010}_{-0.012}$	$-0.187^{+0.009}_{-0.009}$	$-0.195^{+0.007}_{-0.007}$	$-0.194^{+0.008}_{-0.009}$	$-0.196^{+0.011}_{-0.009}$	$-0.180^{+0.009}_{-0.010}$	$-0.193^{+0.007}_{-0.006}$	$-0.192^{+0.008}_{-0.008}$
ΔR_p^2 (^{8,10}B)	$2.021^{+0.140}_{-0.140}$	$1.755^{+0.081}_{-0.096}$	$1.857^{+0.103}_{-0.094}$	$1.848^{+0.112}_{-0.130}$	$1.863^{+0.118}_{-0.118}$	$1.661^{+0.064}_{-0.091}$	$1.780^{+0.083}_{-0.098}$	$1.780^{+0.098}_{-0.114}$
ΔR_p^2 (^{8,11}B)	$2.218^{+0.130}_{-0.130}$	$2.144^{+0.118}_{-0.099}$	$2.127^{+0.097}_{-0.115}$	$2.127^{+0.106}_{-0.115}$	$2.148^{+0.148}_{-0.125}$	$2.003^{+0.091}_{-0.100}$	$2.076^{+0.072}_{-0.103}$	$2.076^{+0.087}_{-0.103}$
ΔR_p^2 (^{10,11}B)	$-0.195^{+0.067}_{-0.067}$	$-0.367^{+0.034}_{-0.048}$	$-0.275^{+0.035}_{-0.035}$	$-0.272^{+0.053}_{-0.056}$	$-0.270^{+0.030}_{-0.030}$	$-0.335^{+0.040}_{-0.048}$	$-0.299^{+0.029}_{-0.029}$	$-0.299^{+0.034}_{-0.034}$
ΔR_p^2 (^{12,10}B)	$-0.498^{+0.050}_{-0.070}$	$-0.439^{+0.040}_{-0.044}$	$-0.326^{+0.039}_{-0.039}$	$-0.326^{+0.046}_{-0.052}$	$-0.602^{+0.056}_{-0.067}$	$-0.373^{+0.037}_{-0.045}$	$-0.316^{+0.033}_{-0.033}$	$-0.316^{+0.059}_{-0.062}$
ΔR_p^2 (^{12,11}B)	$-0.314^{+0.030}_{-0.030}$	$-0.068^{+0.009}_{-0.010}$	$-0.054^{+0.015}_{-0.015}$	$-0.055^{+0.053}_{-0.051}$	$-0.334^{+0.036}_{-0.036}$	$-0.037^{+0.007}_{-0.010}$	$-0.015^{+0.011}_{-0.013}$	$-0.017^{+0.061}_{-0.061}$

Table A.3: Same as Table A.2 but for radius results of the SMS interactions.

Bibliography

[1] E. Epelbaum, H. Krebs, and P. Reinert, **High-Precision Nuclear Forces From Chiral EFT: State-of-the-Art, Challenges, and Outlook**, *Frontiers in Physics* **8**, 98 (2020), DOI: 10.3389/fphy.2020.00098.

[2] R. Machleidt, **Historical perspective and future prospects for nuclear interactions**, *International Journal of Modern Physics E* **26**, 1730005 (2017), DOI: 10.1142/S0218301317300053.

[3] M. Piarulli and I. Tews, **Local Nucleon-Nucleon and Three-Nucleon Interactions Within Chiral Effective Field Theory**, *Frontiers in Physics* **7**, 245 (2020), DOI: 10.3389/fphy.2019.00245.

[4] R. Machleidt and F. Sammarruca, **Chiral EFT based nuclear forces: achievements and challenges**, *Physica Scripta* **91**, 083007 (2016), DOI: 10.1088/0031-8949/91/8/083007.

[5] R. Machleidt and D. Entem, **Chiral effective field theory and nuclear forces**, *Physics Reports* **503**, 1–75 (2011), DOI: 10.1016/j.physrep.2011.02.001.

[6] E. Epelbaum, H.-W. Hammer, and U.-G. Meißner, **Modern theory of nuclear forces**, *Reviews of Modern Physics* **81**, 1773–1825 (2009), DOI: 10.1103/RevModPhys.81.1773.

[7] S. Weinberg, **Nuclear forces from chiral lagrangians**, *Physics Letters B* **251**, 288–292 (1990), DOI: 10.1016/0370-2693(90)90938-3.

[8] J. A. Melendez, R. J. Furnstahl, D. R. Phillips, M. T. Pratola, and S. Wesolowski, **Quantifying correlated truncation errors in effective field theory**, *Physical Review C* **100**, 044001 (2019), DOI: 10.1103/PhysRevC.100.044001.

[9] J. A. Melendez, S. Wesolowski, and R. J. Furnstahl, **Bayesian truncation errors in chiral effective field theory: Nucleon-nucleon observables**, *Physical Review C* **96**, 024003 (2017), DOI: 10.1103/PhysRevC.96.024003.

[10] R. J. Furnstahl, N. Klco, D. R. Phillips, and S. Wesolowski, **Quantifying truncation errors in effective field theory**, *Physical Review C* **92**, 024005 (2015), DOI: 10.1103/PhysRevC.92.024005.

[11] R. Roth, A. Calci, J. Langhammer, and S. Binder, **Evolved chiral N N + 3 N Hamiltonians for *ab initio* nuclear structure calculations**, *Physical Review C* **90**, 024325 (2014), DOI: 10.1103/PhysRevC.90.024325.

[12] B. R. Barrett, P. Navrátil, and J. P. Vary, **Ab initio no core shell model**, *Progress in Particle and Nuclear Physics* **69**, 131–181 (2013), DOI: 10.1016/j.ppnp.2012.10.003.

[13] S. Bogner, R. Furnstahl, and A. Schwenk, **From low-momentum interactions to nuclear structure**, *Progress in Particle and Nuclear Physics* **65**, 94–147 (2010), DOI: 10.1016/j.ppnp.2010.03.001.

[14] H. Hergert, **A Guided Tour of ab initio Nuclear Many-Body Theory**, *Frontiers in Physics* **8**, 379 (2020), DOI: 10.3389/fphy.2020.00379.

[15] P. Navrátil, S. Quaglioni, I. Stetcu, and B. R. Barrett, **Recent developments in no-core shell-model calculations**, *Journal of Physics G: Nuclear and Particle Physics* **36**, 083101 (2009), DOI: 10.1088/0954-3899/36/8/083101.

[16] P. Maris, **Ab Initio Nuclear Structure Calculations of Light Nuclei**, *Journal of Physics: Conference Series* **402**, 012031 (2012), DOI: 10.1088/1742-6596/402/1/012031.

[17] M. K. G. Kruse, **Extensions to the No-Core Shell Model: Importance-Truncation, Regulators and Reactions** (Springer Theses). Cham: Springer International Publishing (2013), ISBN: 978-3-319-01392-3 978-3-319-01393-0 DOI: 10.1007/978-3-319-01393-0.

[18] R. Roth and P. Navrátil, ***Ab Initio* Study of Ca 40 with an Importance-Truncated No-Core Shell Model**, *Physical Review Letters* **99**, 092501 (2007), DOI: 10.1103/PhysRevLett.99.092501.

[19] R. Roth, **Importance truncation for large-scale configuration interaction approaches**, *Physical Review C* **79**, 064324 (2009), DOI: 10.1103/PhysRevC.79.064324.

[20] K. Kowalski, D. J. Dean, M. Hjorth-Jensen, T. Papenbrock, and P. Piecuch, **Coupled Cluster Calculations of Ground and Excited States of Nuclei**, *Physical Review Letters* **92**, 132501 (2004), DOI: 10.1103/PhysRevLett.92.132501.

[21] G. Hagen, T. Papenbrock, D. J. Dean, A. Schwenk, A. Nogga, M. Włoch, and P. Piecuch, **Coupled-cluster theory for three-body Hamiltonians**, *Physical Review C* **76**, 034302 (2007), DOI: 10.1103/PhysRevC.76.034302.

[22] G. Hagen, D. J. Dean, M. Hjorth-Jensen, T. Papenbrock, and A. Schwenk, **Benchmark calculations for H 3 , He 4 , O 16 , and Ca 40 with *ab initio* coupled-cluster theory**, *Physical Review C* **76**, 044305 (2007), DOI: 10.1103/PhysRevC.76.044305.

[23] G. Hagen, T. Papenbrock, M. Hjorth-Jensen, and D. J. Dean, **Coupled-cluster computations of atomic nuclei**, *Reports on Progress in Physics* **77**, 096302 (2014), DOI: 10.1088/0034-4885/77/9/096302.

[24] J. Carlson, S. Gandolfi, F. Pederiva, S. C. Pieper, R. Schiavilla, K. E. Schmidt, and R. B. Wiringa, **Quantum Monte Carlo methods for nuclear physics**, *Reviews of Modern Physics* **87**, 1067–1118 (2015), DOI: 10.1103/RevModPhys.87.1067.

[25] J. Lynn, I. Tews, S. Gandolfi, and A. Lovato, **Quantum Monte Carlo Methods in Nuclear Physics: Recent Advances**, *Annual Review of Nuclear and Particle Science* **69**, 279–305 (2019), DOI: 10.1146/annurev-nucl-101918-023600.

[26] S. Gandolfi, D. Lonardoni, A. Lovato, and M. Piarulli, **Atomic Nuclei From Quantum Monte Carlo Calculations With Chiral EFT Interactions**, *Frontiers in Physics* **8**, 117 (2020), DOI: 10.3389/fphy.2020.00117.

[27] W. Dickhoff and C. Barbieri, **Self-consistent Green's function method for nuclei and nuclear matter**, *Progress in Particle and Nuclear Physics* **52**, 377–496 (2004), DOI: 10.1016/j.ppnp.2004.02.038.

[28] V. Somà, C. Barbieri, and T. Duguet, ***Ab initio* Gorkov-Green's function calculations of open-shell nuclei**, *Physical Review C* **87**, 011303 (2013), DOI: 10.1103/PhysRevC.87.011303.

[29] V. Somà, A. Cipollone, C. Barbieri, P. Navrátil, and T. Duguet, **Chiral two- and three-nucleon forces along medium-mass isotope chains**, *Physical Review C* **89**, 061301 (2014), DOI: 10.1103/PhysRevC.89.061301.

[30] V. Somà, **Self-Consistent Green's Function Theory for Atomic Nuclei**, *Frontiers in Physics* **8**, 340 (2020), DOI: 10.3389/fphy.2020.00340.

[31] A. Nicholson, **Lattice Methods and Effective Field Theory**, in *An Advanced Course in Computational Nuclear Physics*, M. Hjorth-Jensen, M. P. Lombardo, and U. Van Kolck, Eds., 936, Cham: Springer International Publishing (2017), pp. 155–235, ISBN: 978-3-319-53335-3 978-3-319-53336-0 DOI: 10.1007/978-3-319-53336-0_5.

[32] D. Lee, **Recent Progress in Nuclear Lattice Simulations**, *Frontiers in Physics* **8**, 174 (2020), DOI: 10.3389/fphy.2020.00174.

[33] H. Hergert, S. Bogner, T. Morris, A. Schwenk, and K. Tsukiyama, **The In-Medium Similarity Renormalization Group: A novel ab initio method for nuclei**, *Physics Reports* **621**, 165–222 (2016), DOI: 10.1016/j.physrep.2015.12.007.

[34] H. Hergert, S. K. Bogner, S. Binder, A. Calci, J. Langhammer, R. Roth, and A. Schwenk, **In-medium similarity renormalization group with chiral two- plus three-nucleon interactions**, *Physical Review C* **87**, 034307 (2013), DOI: 10.1103/PhysRevC.87.034307.

[35] H. Hergert, S. K. Bogner, T. D. Morris, S. Binder, A. Calci, J. Langhammer, and R. Roth, ***Ab initio* multireference in-medium similarity renormalization group calculations of even calcium and nickel isotopes**, *Physical Review C* **90**, 041302 (2014), DOI: 10.1103/PhysRevC.90.041302.

[36] E. Gebrerufael, K. Vobig, H. Hergert, and R. Roth, ***Ab Initio* Description of Open-Shell Nuclei: Merging No-Core Shell Model and In-Medium Similarity Renormalization Group**, *Physical Review Letters* **118**, 152503 (2017), DOI: 10.1103/PhysRevLett.118.152503.

[37] K. Vobig, **Electromagnetic Observables and Open-Shell Nuclei from the In-Medium No-Core Shell Model**, (2020), DOI: 10.25534/TUPRINTS-00011375.

[38] M. Frosini, T. Duguet, J.-P. Ebran, B. Bally, T. Mongelli, T. R. Rodríguez, R. Roth, and V. Somà, **Multi-reference many-body perturbation theory for nuclei: II. Ab initio study of neon isotopes via PGCM and IM-NCSM calculations**, *The European Physical Journal A* **58**, 63 (2022), DOI: 10.1140/epja/s10050-022-00693-y.

[39] E. A. Hylleraas and B. Undheim, **Numerische Berechnung der 2S-Terme von Ortho- und Par-Helium**, *Zeitschrift für Physik* **65**, 759–772 (1930), DOI: 10.1007/BF01397263.

[40] B. Verhaar, **A method for the elimination of spurious states in the nuclear harmonic oscillator shell model**, *Nuclear Physics* **21**, 508–525 (1960), DOI: 10.1016/0029-5582(60)90073-0.

[41] D. Gloeckner and R. Lawson, **Spurious center-of-mass motion**, *Physics Letters B* **53**, 313–318 (1974), DOI: 10.1016/0370-2693(74)90390-6.

[42] M. Knöll, M. Lockner, P. Maris, R. J. McCarty, R. Roth, J. P. Vary, and T. Wolfgruber, **Benchmarking artificial neural network extrapolations of the ground-state energies and radii of Li isotopes**, *Physical Review C* **111**, (2025), DOI: 10.1103/physrevc.111.064304.

[43] C. M. Bishop, **Neural Networks for Pattern Recognition**. Oxford University PressOxford (1995), ISBN: 978-0-19-853849-3 978-1-383-02638-2 DOI: 10.1093/oso/9780198538493.001.0001.

[44] J. W. Clark, T. Lindenau, and M. L. Ristig, Eds., **Scientific Applications of Neural Nets** (Lecture Notes in Physics). Springer Berlin Heidelberg (1999), 522, ISBN: 978-3-540-65737-8 DOI: 10.1007/BFb0104276.

[45] F. A. Rodrigues, **Machine learning in physics: A short guide**, *Europhysics Letters* **144**, 22001 (2023), DOI: 10.1209/0295-5075/ad0575.

[46] A. Boehnlein, M. Diefenthaler, N. Sato, M. Schram, V. Ziegler, C. Fanelli, M. Hjorth-Jensen, T. Horn, M. P. Kuchera, D. Lee, W. Nazarewicz, P. Ostroumov, K. Orginos, A. Poon, X.-N. Wang, A. Scheinker, M. S. Smith, and L.-G. Pang, ***Colloquium* : Machine learning in nuclear physics**, *Reviews of Modern Physics* **94**, 031003 (2022), DOI: 10.1103/RevModPhys.94.031003.

[47] P. Bedaque, A. Boehnlein, M. Cromaz, M. Diefenthaler, L. Elouadrhiri, T. Horn, M. Kuchera, D. Lawrence, D. Lee, S. Lidia, R. McKeown, W. Melnitchouk, W. Nazarewicz, K. Orginos, Y. Roblin, M. Scott Smith, M. Schram, and X.-N. Wang, **A.I. for nuclear physics**, *The European Physical Journal A* **57**, 100 (2021), DOI: 10.1140/epja/s10050-020-00290-x.

[48] W. He, Q. Li, Y. Ma, Z. Niu, J. Pei, and Y. Zhang, **Machine learning in nuclear physics at low and intermediate energies**, *Science China Physics, Mechanics & Astronomy* **66**, 282001 (2023), DOI: 10.1007/s11433-023-2116-0.

[49] W.-B. He, Y.-G. Ma, L.-G. Pang, H.-C. Song, and K. Zhou, **High-energy nuclear physics meets machine learning**, *Nuclear Science and Techniques* **34**, 88 (2023), DOI: 10.1007/s41365-023-01233-z.

[50] P. Navrátil, G. P. Kamuntavičius, and B. R. Barrett, **Few-nucleon systems in a translationally invariant harmonic oscillator basis**, *Physical Review C* **61**, 044001 (2000), DOI: 10.1103/PhysRevC.61.044001.

[51] S. Liebig, U.-G. Meißner, and A. Nogga, **Jacobi no-core shell model for p-shell nuclei**, *The European Physical Journal A* **52**, 103 (2016), DOI: 10.1140/epja/i2016-16103-5.

[52] X.-X. Sun, H. Le, U.-G. Meißner, and A. Nogga, **Radii of light nuclei from the Jacobi no-core shell model**, *Physical Review C* **112**, 024317 (2025), DOI: 10.1103/j4ky-tn1j.

[53] C. Lanczos, **An iteration method for the solution of the eigenvalue problem of linear differential and integral operators**, *Journal of Research of the National Bureau of Standards* **45**, 255 (1950), DOI: 10.6028/jres.045.026.

[54] H. Le, J. Haidenbauer, U.-G. Meißner, and A. Nogga, **Jacobi no-core shell model for p-shell hypernuclei**, *The European Physical Journal A* **56**, 301 (2020), DOI: 10.1140/epja/s10050-020-00314-6.

[55] R. B. Wiringa, V. G. J. Stoks, and R. Schiavilla, **Accurate nucleon-nucleon potential with charge-independence breaking**, *Physical Review C* **51**, 38–51 (1995), DOI: 10.1103/PhysRevC.51.38.

[56] S. Veerasamy and W. N. Polyzou, **Momentum-space Argonne V18 interaction**, *Physical Review C* **84**, 034003 (2011), DOI: 10.1103/PhysRevC.84.034003.

[57] R. Machleidt, **High-precision, charge-dependent Bonn nucleon-nucleon potential**, *Physical Review C* **63**, (2001), DOI: 10.1103/physrevc.63.024001.

[58] R. Machleidt, K. Holinde, and Ch. Elster, **The bonn meson-exchange model for the nucleon—nucleon interaction**, *Physics Reports* **149**, 1–89 (1987), DOI: 10.1016/S0370-1573(87)80002-9.

[59] S. Weinberg, **Effective chiral lagrangians for nucleon-pion interactions and nuclear forces**, *Nuclear Physics B* **363**, 3–18 (1991), DOI: 10.1016/0550-3213(91)90231-L.

[60] S. Weinberg, **Three-body interactions among nucleons and pions**, *Physics Letters B* **295**, 114–121 (1992), DOI: 10.1016/0370-2693(92)90099-P.

[61] E. Epelbaum, H. Krebs, and U.-G. Meißner, **Precision Nucleon-Nucleon Potential at Fifth Order in the Chiral Expansion**, *Physical Review Letters* **115**, 122301 (2015), DOI: 10.1103/PhysRevLett.115.122301.

[62] R. Machleidt and F. Sammarruca, **Can chiral EFT give us satisfaction?**, *The European Physical Journal A* **56**, 95 3 (2020), DOI: 10.1140/epja/s10050-020-00101-3.

[63] S. Schulz, **Four-Nucleon Forces in Ab Initio Nuclear Structure**, Ph.D. dissertation, Technische Universität Darmstadt (2018).: https://tuprints.ulb.tu-darmstadt.de/7384/.

[64] A. Dyhdalo, R. J. Furnstahl, K. Hebeler, and I. Tews, **Regulator artifacts in uniform matter for chiral interactions**, *Physical Review C* **94**, 034001 (2016), DOI: 10.1103/PhysRevC.94.034001.

[65] D. R. Entem, R. Machleidt, and Y. Nosyk, **High-quality two-nucleon potentials up to fifth order of the chiral expansion**, *Physical Review C* **96**, 024004 (2017), DOI: 10.1103/PhysRevC.96.024004.

[66] D. Entem and R. Machleidt, **Accurate nucleon–nucleon potential based upon chiral perturbation theory**, *Physics Letters B* **524**, 93–98 (2002), DOI: 10.1016/S0370-2693(01)01363-6.

[67] D. R. Entem and R. Machleidt, **Accurate charge-dependent nucleon-nucleon potential at fourth order of chiral perturbation theory**, *Physical Review C* **68**, 041001 (2003), DOI: 10.1103/PhysRevC.68.041001.

[68] T. Hüther, K. Vobig, K. Hebeler, R. Machleidt, and R. Roth, **Family of chiral two- plus three-nucleon interactions for accurate nuclear structure studies**, *Physics Letters B* **808**, 135651 (2020), DOI: 10.1016/j.physletb.2020.135651.

[69] Hüther, Thomas, **Families of Chiral Two- plus Three-Nucleon Interactions for Accurate Nuclear Structure Studies**, Technische Universität Darmstadt Darmstadt (2021), DOI: 10.26083/TUPRINTS-00018588.

[70] E. Epelbaum, H. Krebs, and U. .-. Meißner, **Improved chiral nucleon-nucleon potential up to next-to-next-to-next-to-leading order**, *The European Physical Journal A* **51**, 53 (2015), DOI: 10.1140/epja/i2015-15053-8.

[71] P. Reinert, H. Krebs, and E. Epelbaum, **Semilocal momentum-space regularized chiral two-nucleon potentials up to fifth order**, *The European Physical Journal A* **54**, 86 (2018), DOI: 10.1140/epja/i2018-12516-4.

[72] E. Epelbaum, J. Golak, K. Hebeler, T. Hüther, H. Kamada, H. Krebs, P. Maris, U.-G. Meißner, A. Nogga, R. Roth, R. Skibiński, K. Topolnicki, J. P. Vary, K. Vobig, and H. Witała (LENPIC Collaboration), **Few- and many-nucleon systems with semilocal coordinate-space regularized chiral two- and three-body forces**, *Physical Review C* **99**, 024313 2 (2019), DOI: 10.1103/PhysRevC.99.024313.

[73] P. Maris, R. Roth, E. Epelbaum, R. J. Furnstahl, J. Golak, K. Hebeler, T. Hüther, H. Kamada, H. Krebs, H. Le, U.-G. Meißner, J. A. Melendez, A. Nogga, P. Reinert, R. Skibiński, J. P. Vary, H. Witała, and T. Wolfgruber (LENPIC Collaboration), **Nuclear properties with semilocal momentum-space regularized chiral interactions beyond N 2 LO**, *Physical Review C* **106**, 064002 (2022), DOI: 10.1103/PhysRevC.106.064002.

[74] P. Maris, E. Epelbaum, R. J. Furnstahl, J. Golak, K. Hebeler, T. Hüther, H. Kamada, H. Krebs, U.-G. Meißner, J. A. Melendez, A. Nogga, P. Reinert, R. Roth, R. Skibiński, V. Soloviov, K. Topolnicki, J. P. Vary, Yu. Volkotrub, H. Witała, and T. Wolfgruber (LENPIC Collaboration), **Light nuclei with semilocal momentum-space regularized chiral interactions up to third order**, *Physical Review C* **103**, 054001 5 (2021), DOI: 10.1103/PhysRevC.103.054001.

[75] S. K. Bogner, R. J. Furnstahl, and R. J. Perry, **Similarity renormalization group for nucleon-nucleon interactions**, *Physical Review C* **75**, 061001 (2007), DOI: 10.1103/PhysRevC.75.061001.

[76] S. Bogner, R. Furnstahl, R. Perry, and A. Schwenk, **Are low-energy nuclear observables sensitive to high-energy phase shifts?**, *Physics Letters B* **649**, 488–493 (2007), DOI: 10.1016/j.physletb.2007.04.048.

[77] H. Hergert and R. Roth, **Unitary correlation operator method from a similarity renormalization group perspective**, *Physical Review C* **75**, 051001 (2007), DOI: 10.1103/PhysRevC.75.051001.

[78] S. Bogner, R. Furnstahl, S. Ramanan, and A. Schwenk, **Low-momentum interactions with smooth cutoffs**, *Nuclear Physics A* **784**, 79–103 (2007), DOI: 10.1016/j.nuclphysa.2006.11.123.

[79] S. Bogner, R. Furnstahl, P. Maris, R. Perry, A. Schwenk, and J. Vary, **Convergence in the no-core shell model with low-momentum two-nucleon interactions**, *Nuclear Physics A* **801**, 21–42 (2008), DOI: 10.1016/j.nuclphysa.2007.12.008.

[80] A. Calci, **Evolved Chiral Hamiltonians at the Three-Body Level and Beyond**, Technische Universität Darmstadt (2014).: https://tuprints.ulb.tu-darmstadt.de/4069/.

[81] A. Antognini, F. Hagelstein, and V. Pascalutsa, **The Proton Structure in and out of Muonic Hydrogen**, *Annual Review of Nuclear and Particle Science* **72**, 389–418 (2022), DOI: 10.1146/annurev-nucl-101920-024709.

[82] H. Gao and M. Vanderhaeghen, **The proton charge radius**, *Reviews of Modern Physics* **94**, 015002 (2022), DOI: 10.1103/RevModPhys.94.015002.

[83] G. A. Miller, **Defining the proton radius: A unified treatment**, *Physical Review C* **99**, 035202 (2019), DOI: 10.1103/PhysRevC.99.035202.

[84] H.-W. Hammer and U.-G. Meißner, **The proton radius: from a puzzle to precision**, *Science Bulletin* **65**, 257–258 (2020), DOI: 10.1016/j.scib.2019.12.012.

[85] C. Peset, A. Pineda, and O. Tomalak, **The proton radius (puzzle?) and its relatives**, *Progress in Particle and Nuclear Physics* **121**, 103901 (2021), DOI: 10.1016/j.ppnp.2021.103901.

[86] J. Zhou, V. Khachatryan, H. Gao, D. W. Higinbotham, A. Parker, X. Bai, D. Dutta, A. Gasparian, K. Gnanvo, M. Khandaker, N. Liyanage, E. Pasyuk, C. Peng, and W. Xiong, **Advanced extraction of the deuteron charge radius from electron-deuteron scattering data**, *Physical Review C* **103**, 024002 (2021), DOI: 10.1103/PhysRevC.103.024002.

[87] A. D. Brandt, S. F. Cooper, C. Rasor, Z. Burkley, A. Matveev, and D. C. Yost, **Measurement of the 2 S 1 / 2 − 8 D 5 / 2 Transition in Hydrogen**, *Physical Review Letters* **128**, 023001 (2022), DOI: 10.1103/PhysRevLett.128.023001.

[88] Y.-H. Lin, F.-K. Guo, and U.-G. Meißner, **The proton charge radius from dimuon photoproduction off the proton**, *Physics Letters B* **858**, 139023 (2024), DOI: 10.1016/j.physletb.2024.139023.

[89] P. Mohr, D. Newell, B. Taylor, and E. Tiesinga. **CODATA Recommended Values of the Fundamental Physical Constants: 2022**, arXiv: hep-ph/2409.03787 (2024), pre-published.

[90] S. Navas et al., **Review of Particle Physics**, *Physical Review D* **110**, 030001 (2024), DOI: 10.1103/PhysRevD.110.030001.

[91] J. L. Friar, J. Martorell, and D. W. L. Sprung, **Nuclear sizes and the isotope shift**, *Physical Review A* **56**, 4579–4586 (1997), DOI: 10.1103/PhysRevA.56.4579.

[92] L. L. Foldy and S. A. Wouthuysen, **On the Dirac Theory of Spin 1/2 Particles and Its Non-Relativistic Limit**, *Physical Review* **78**, 29–36 (1950), DOI: 10.1103/PhysRev.78.29.

[93] G. W. Drake, W. Nörtershäuser, and Z. .-. Yan, **Isotope shifts and nuclear radius measurements for helium and lithium**, *Canadian Journal of Physics* **83**, 311–325 (2005), DOI: 10.1139/p05-020.

[94] B. Maaß, P. Müller, W. Nörtershäuser, J. Clark, C. Gorges, S. Kaufmann, K. König, J. Krämer, A. Levand, R. Orford, R. Sánchez, G. Savard, and F. Sommer, **Towards laser spectroscopy of the proton-halo candidate boron-8**, *Hyperfine Interactions* **238**, 25 (2017), DOI: 10.1007/s10751-017-1399-5.

[95] B. Maaß, T. Hüther, K. König, J. Krämer, J. Krause, A. Lovato, P. Müller, K. Pachucki, M. Puchalski, R. Roth, R. Sánchez, F. Sommer, R. B. Wiringa, and W. Nörtershäuser, **Nuclear Charge Radii of B 10 , 11**, *Physical Review Letters* **122**, 182501 (2019), DOI: 10.1103/PhysRevLett.122.182501.

[96] P. M. Müller, **Laserspectroscopic determination of the nuclear charge radius of ^{13}C**, (2024), DOI: 10.26083/TUPRINTS-00026746.

[97] M. Wang, W. Huang, F. Kondev, G. Audi, and S. Naimi, **The AME 2020 atomic mass evaluation (II). Tables, graphs and references***, *Chinese Physics C* **45**, 030003 (2021), DOI: 10.1088/1674-1137/abddaf.

[98] P. Navrátil and E. Caurier, **Nuclear structure with accurate chiral perturbation theory nucleon-nucleon potential: Application to Li6 and B10**, *Physical Review C* **69**, 014311 (2004), DOI: 10.1103/PhysRevC.69.014311.

[99] P. Maris, J. P. Vary, and A. M. Shirokov, ***Ab initio* no-core full configuration calculations of light nuclei**, *Physical Review C* **79**, 014308 (2009), DOI: 10.1103/PhysRevC.79.014308.

[100] P. Maris, H. Le, A. Nogga, R. Roth, and J. P. Vary, **Uncertainties in ab initio nuclear structure calculations with chiral interactions**, *Frontiers in Physics* **11**, 1098262 (2023), DOI: 10.3389/fphy.2023.1098262.

[101] C. Forssén, J. P. Vary, E. Caurier, and P. Navrátil, **Converging sequences in the *ab initio* no-core shell model**, *Physical Review C* **77**, 024301 (2008), DOI: 10.1103/PhysRevC.77.024301.

[102] R. J. Furnstahl, G. Hagen, and T. Papenbrock, **Corrections to nuclear energies and radii in finite oscillator spaces**, *Physical Review C* **86**, 031301 (2012), DOI: 10.1103/PhysRevC.86.031301.

[103] R. J. Furnstahl, G. Hagen, T. Papenbrock, and K. A. Wendt, **Infrared extrapolations for atomic nuclei**, *Journal of Physics G: Nuclear and Particle Physics* **42**, 034032 (2015), DOI: 10.1088/0954-3899/42/3/034032.

[104] K. A. Wendt, C. Forssén, T. Papenbrock, and D. Sääf, **Infrared length scale and extrapolations for the no-core shell model**, *Physical Review C* **91**, 061301 (2015), DOI: 10.1103/PhysRevC.91.061301.

[105] S. A. Coon, M. I. Avetian, M. K. G. Kruse, U. Van Kolck, P. Maris, and J. P. Vary, **Convergence properties of *ab initio* calculations of light nuclei in a harmonic oscillator basis**, *Physical Review C* **86**, 054002 (2012), DOI: 10.1103/PhysRevC.86.054002.

[106] F. Rosenblatt, **The perceptron: A probabilistic model for information storage and organization in the brain.**, *Psychological Review* **65**, 386–408 (1958), DOI: 10.1037/h0042519.

[107] F. Rosenblatt, **Principles of neurodynamics: Perceptions and the theory of brain mechanism**. Washington, DC: Spartan Books (1961).

[108] A. G. Ivakhnenko, **Polynomial Theory of Complex Systems**, *IEEE Transactions on Systems, Man, and Cybernetics* **SMC-1**, 364–378 (1971), DOI: 10.1109/TSMC.1971.4308320.

[109] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Y. Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng, *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*, (2015).: https://www.tensorflow.org/.

[110] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala. **PyTorch: An Imperative Style, High-Performance Deep Learning Library**, arXiv: cs/1912.01703 (2019), pre-published.

[111] J. Ansel, E. Yang, H. He, N. Gimelshein, A. Jain, M. Voznesensky, B. Bao, P. Bell, D. Berard, E. Burovski, G. Chauhan, A. Chourdia, W. Constable, A. Desmaison, Z. DeVito, E. Ellison, W. Feng, J. Gong, M. Gschwind, B. Hirsh, S. Huang, K. Kalambarkar, L. Kirsch, M. Lazos, M. Lezcano, Y. Liang, J. Liang, Y. Lu, C. K. Luk, B. Maher, Y. Pan, C. Puhrsch, M. Reso, M. Saroufim, M. Y. Siraichi, H. Suk, M. Suo, P. Tillet, E. Wang, X. Wang, W. Wen, S. Zhang, X. Zhao, K. Zhou, R. Zou, A. Mathews, G. Chanan, P. Wu, and S. Chintala, **PyTorch 2: Faster Machine Learning Through Dynamic Python Bytecode Transformation and Graph Compilation**, in *29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 2 (ASPLOS '24)*, ACM (2024), DOI: 10.1145/3620665.3640366.

[112] K. Hornik, M. Stinchcombe, and H. White, **Multilayer feedforward networks are universal approximators**, *Neural Networks* **2**, 359–366 (1989), DOI: 10.1016/0893-6080(89)90020-8.

[113] R. P. Lippmann, **An introduction to computing with neural nets**, *ACM SIGARCH Computer Architecture News* **16**, 7–25 (1988), DOI: 10.1145/44571.44572.

[114] A. Laudani, G. M. Lozito, F. Riganti Fulginei, and A. Salvini, **On Training Efficiency and Computational Costs of a Feed Forward Neural Network: A Review**, *Computational Intelligence and Neuroscience* **2015**, 1–13 (2015), DOI: 10.1155/2015/818243.

[115] Y. LeCun, B. Boser, J. Denker, D. Henderson, R. Howard, W. Hubbard, and L. Jackel, **Handwritten Digit Recognition with a Back-Propagation Network**, in *Advances in Neural Information Processing Systems*, D. Touretzky, Ed., 2, Morgan-Kaufmann (1989).

[116] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, **Gradient-based learning applied to document recognition**, *Proceedings of the IEEE* **86**, 2278–2324 (1998), DOI: 10.1109/5.726791.

[117] Y. LeCun, Y. Bengio, and G. Hinton, **Deep learning**, *Nature* **521**, 436–444 (2015), DOI: 10.1038/nature14539.

[118] A. Krizhevsky, I. Sutskever, and G. E. Hinton, **ImageNet classification with deep convolutional neural networks**, *Communications of the ACM* **60**, 84–90 (2017), DOI: 10.1145/3065386.

[119] J. L. Elman, **Finding Structure in Time**, *Cognitive Science* **14**, 179–211 (1990), DOI: 10.1207/s15516709cog1402_1.

[120] I. D. Mienye, T. G. Swart, and G. Obaido, **Recurrent Neural Networks: A Comprehensive Review of Architectures, Variants, and Applications**, *Information* **15**, 517 (2024), DOI: 10.3390/info15090517.

[121] Y. Yu, X. Si, C. Hu, and J. Zhang, **A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures**, *Neural Computation* **31**, 1235–1270 (2019), DOI: 10.1162/neco_a_01199.

[122] Z. C. Lipton, J. Berkowitz, and C. Elkan. **A Critical Review of Recurrent Neural Networks for Sequence Learning**, arXiv: cs/1506.00019 (2015), pre-published.

[123] S. Hochreiter and J. Schmidhuber, **Long Short-Term Memory**, *Neural Computation* **9**, 1735–1780 (1997), DOI: 10.1162/neco.1997.9.8.1735.

[124] G. A. Negoita, G. R. Luecke, J. P. Vary, P. Maris, A. M. Shirokov, I. J. Shin, Y. Kim, E. G. Ng, and C. Yang, **Deep Learning: A Tool for Computational Nuclear Physics**, in *The Ninth International Conference on Computational Logics, Algebras, Programming, Tools, and Benchmarking*, G. Luecke and

J. L. Mauri, Eds., USA: IARIA (2018), pp. 20–28, ISBN: 978-1-61208-613-2.

[125] 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, and G. M. Prabhu, **Deep learning: Extrapolation tool for *ab initio* nuclear theory**, *Physical Review C* **99**, 054308 (2019), DOI: 10.1103/PhysRevC.99.054308.

[126] D. Freedman and P. Diaconis, **On the histogram as a density estimator:L 2 theory**, *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **57**, 453–476 (1981), DOI: 10.1007/BF01025868.

[127] M. Knöll, T. Wolfgruber, M. L. Agel, C. Wenz, and R. Roth, **Machine learning for the prediction of converged energies from ab initio nuclear structure calculations**, *Physics Letters B* 137781 (2023), DOI: 10.1016/j.physletb.2023.137781.

[128] T. Wolfgruber, M. Knöll, and R. Roth, **Precise neural network predictions of energies and radii from the no-core shell model**, *Physical Review C* **110**, 014327 (2024), DOI: 10.1103/PhysRevC.110.014327.

[129] T. Wolfgruber, T. Gesser, M. Knöll, P. Maris, and R. Roth, **High-precision *ab initio* radius calculations of boron isotopes**, *Physical Review C* **112**, (2025), DOI: 10.1103/8mfb-wc36.

[130] M. Knöll, M. L. Agel, T. Wolfgruber, P. Maris, and R. Roth, **Machine learning for correlations of electromagnetic properties in *ab initio* calculations**, *Physical Review C* **112**, 034306 (2025), DOI: 10.1103/frfs-xz86.

[131] V. Nair and G. E. Hinton, **Rectified Linear Units Improve Restricted Boltzmann Machines**, in *Proceedings of the 27th International Conference on Machine Learning*, Omnipress (2010), pp. 807–814, ISBN: 978-1-60558-907-7.

[132] I. Loshchilov and F. Hutter. **Decoupled Weight Decay Regularization**. version 3, arXiv: cs/1711.05101 (2017), pre-published.

[133] L. Huang, J. Qin, Y. Zhou, F. Zhu, L. Liu, and L. Shao. **Normalization Techniques in Training DNNs: Methodology, Analysis and Application**, arXiv: cs/2009.12836 (2020), pre-published.

[134] A. Shirokov, I. Shin, Y. Kim, M. Sosonkina, P. Maris, and J. Vary, **N3LO NN interaction adjusted to light nuclei in ab exitu approach**, *Physics Letters B* **761**, 87–91 (2016), DOI: 10.1016/j.physletb.2016.08.006.

[135] R. McCarty, **Ab initio nuclear structure of lithium isotopes with applications of deep learning**, Ph.D. dissertation, Iowa State University Ames, Iowa (2024), 106 pp., ISBN: 979-8-38-449295-5 DOI: 10.31274/td-20250502-321.

[136] P. Maris, *Private Communication*, (2025).

[137] H. M. Aktulga, C. Yang, E. G. Ng, P. Maris, and J. P. Vary, **Improving the scalability of a symmetric iterative eigensolver for multi-core platforms**, *Concurrency and Computation: Practice and Experience* **26**, 2631–2651 (2014), DOI: 10.1002/cpe.3129.

[138] M. Shao, H. M. Aktulga, C. Yang, E. G. Ng, P. Maris, and J. P. Vary, **Accelerating nuclear configuration interaction calculations through a preconditioned block iterative eigensolver**, *Computer Physics Communications* **222**, 1–13 (2018), DOI: 10.1016/j.cpc.2017.09.004.

[139] I. Angeli and K. Marinova, **Table of experimental nuclear ground state charge radii: An update**, *Atomic Data and Nuclear Data Tables* **99**, 69–95 (2013), DOI: 10.1016/j.adt.2011.12.006.

[140] M. Knöll and R. Roth, **Hyperon-nucleon interaction constrained by light hypernuclei**, *Physics Letters B* **846**, 138258 (2023), DOI: 10.1016/j.physletb.2023.138258.

[141] M. Knöll, **Ab Initio Theory and Machine Learning for Hypernuclei**, (2024), DOI: 10.26083/TUPRINTS-00027654.

[142] M. A. Caprio, P. J. Fasano, and P. Maris, **Robust *ab initio* prediction of nuclear electric quadrupole observables by scaling to the charge radius**, *Physical Review C* **105**, L061302 (2022), DOI: 10.1103/PhysRevC.105.L061302.

[143] M. A. Caprio, P. Maris, and P. J. Fasano, **Robust *ab initio* predictions for dimensionless ratios of E 2 and radius observables. I. Electric quadrupole moments and deformation**, *Physical Review C* **112**, 044318 (2025), DOI: 10.1103/6zk6-1sy6.

Publications

M. Knöll, M. L. Agel, T. Wolfgruber, P. Maris, and R. Roth, **Machine learning for correlations of electromagnetic properties in ab initio calculations**, *Physical Review C* **112**, 034306 (2025), DOI: 10.1103/frfs-xz86.

T. Wolfgruber, T. Gesser, M. Knöll, P. Maris, and R. Roth, **High-precision ab initio radius calculations of boron isotopes**, *Physical Review C* **112**, (2025), DOI: 10.1103/8mfb-wc36.

M. Knöll, M. Lockner, P. Maris, R. J. McCarty, R. Roth, J. P. Vary, and T. Wolfgruber, **Benchmarking artificial neural network extrapolations of the ground-state energies and radii of Li isotopes**, *Physical Review C* **111**, (2025), DOI: 10.1103/physrevc.111.064304.

T. Wolfgruber, M. Knöll, and R. Roth, **Precise neural network predictions of energies and radii from the no-core shell model**, *Physical Review C* **110**, 014327 (2024), DOI: 10.1103/PhysRevC.110.014327.

M. Knöll, T. Wolfgruber, M. L. Agel, C. Wenz, and R. Roth, **Machine learning for the prediction of converged energies from ab initio nuclear structure calculations**, *Physics Letters B* 137781 (2023), DOI: 10.1016/j.physletb.2023.137781.

P. Maris et al. (LENPIC Collaboration), **Nuclear properties with semilocal momentum-space regularized chiral interactions beyond N 2 LO**, *Physical Review C* **106**, 064002 (2022), DOI: 10.1103/PhysRevC.106.064002.

P. Maris et al. (LENPIC Collaboration), **Light nuclei with semilocal momentum-space regularized chiral interactions up to third order**, *Physical Review C* **103**, 054001 5 (2021), DOI: 10.1103/PhysRevC.103.054001.

C. Stumpf, T. Wolfgruber, and R. Roth, **Electromagnetic Strength Distributions from the Ab Initio No-Core Shell Model**, arXiv: nucl-th/1709.06840 (2017), pre-published.

Danksagung

Zuerst möchte ich meinem Doktorvater Robert Roth ganz herzlich danken. Nicht nur für die Möglichkeit, diese Arbeit anzufertigen, sondern vor allem für die tolle Betreuung, die vielen anregenden Diskussionen und seinen unerschöpflichen Optimismus. Ohne das alles wäre diese Arbeit so nicht zustande gekommen.

Ebenso danke ich Kai Hebeler für die Übernahme des Korreferats und die sehr gute Zusammenarbeit.

Ein großer Dank geht außerdem an alle aktuellen und ehemaligen Mitglieder der TNP++ Gruppe, die ich während meiner Zeit hier kennenlernen durfte und die meinen Alltag bereichert haben: Alex, Carl, Cedric, Christina, Daniel, Julius, Katrin, Klaus, Laura, Lisa, Luis, Marc, Marco, Michelle, Roland, Stefan, Thomas, Tobi, Tobi, Tobi, Vanessa und alle, die ich eventuell vergessen habe. Danke insbesondere an Cedric, Laura, Lisa, Michelle und Tobi für die vielen schönen Stunden im Büro, die Spaziergänge im Herrngarten, die gemeinsamen Konferenzen und vieles mehr. Ein ganz besonderes Dankeschön geht außerdem an Marco für die großartige Zusammenarbeit, die spannenden Diskussionen und das Korrekturlesen dieser Arbeit, aber auch dafür, dass du ein so verlässlicher und unterhaltsamer Reise- und vor allem Roadtrip-Begleiter bist (und natürlich für deinen hervorragenden Musikgeschmack).

Genauso möchte ich mich bei meinen Freunden bedanken, die mich in dieser Zeit begleitet, unterstützt und auch das ein oder andere Mal ertragen haben. Ihr macht mein Leben deutlich besser. Besonders danke ich Mark, Laura, Carl, Netti, Daniel und Jan für die unglaubliche Unterstützung, die vielen gemeinsamen Erlebnisse und unvergesslichen Momente.

Zum Schluss möchte ich noch herzlich meiner Familie danken. Meinem Vater danke ich von Herzen für seinen bedingungslosen Rückhalt. Meiner Mutter, die meinen akademischen Weg leider nicht mehr miterleben konnte, bin ich unendlich dankbar für ihre Unterstützung, die Förderung meiner Interessen und meiner Neugier und für all die schönen gemeinsamen Zeiten. Diese Arbeit ist ihr gewidmet.