TECHNISCHE UNIVERSITÄT DARMSTADT

Machine Learning for the Prediction of Converged Observables from NCSM Calculations

Tobias Wolfgruber

LENPIC Meeting 2022

Motivation

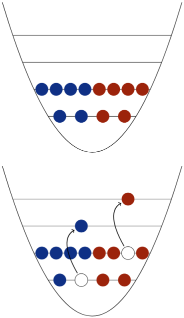
TECHNISCHE UNIVERSITÄT DARMSTADT

- *ab initio* basis expansion methods initially have infinite Hilbert spaces
- truncate to finite model spaces
  - introduces truncation error
  - systematic improvement of results with increasing model-space size
- rapid growth of underlying bases
  - only access to limited model spaces
- need prediction with uncertainty for full Hilbert space
- classical extrapolations are phenomenological

No-Core Shell Model

- stationary Schrödinger equation as matrix eigenvalue problem

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

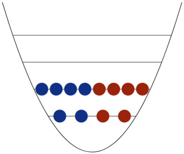


- Slater determinants $|\phi_i\rangle$ constructed from HO basis
  - dependency on HO frequency $\hbar\Omega$
- truncate model space by number of excitation quanta N_{max} w.r.t. the lowest-energy Slater determinant

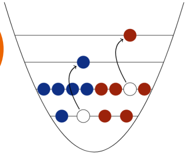
No-Core Shell Model

- stationary Schrödinger equation as matrix eigenvalue problem

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

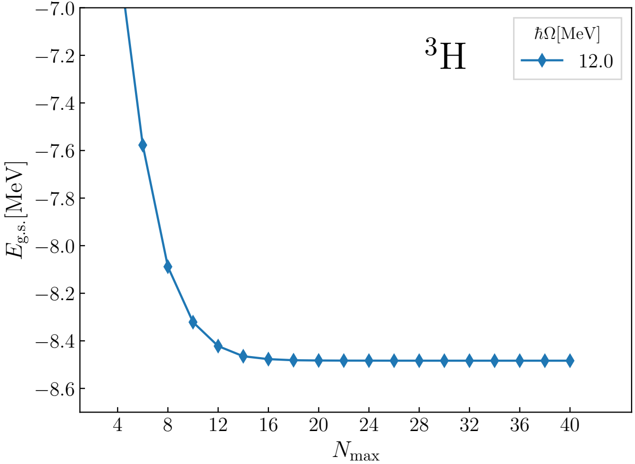


- Slater determinants $|\phi_i\rangle$ constructed from HO basis
  - dependency on HO frequency $\hbar\Omega$
- truncate model space by number of excitation quanta N_{max} w.r.t. the lowest-energy Slater determinant



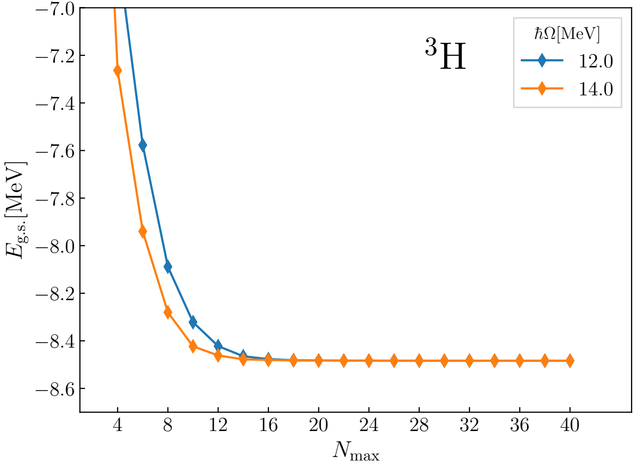
- **convergence controlled by two parameters**

Convergence Behavior



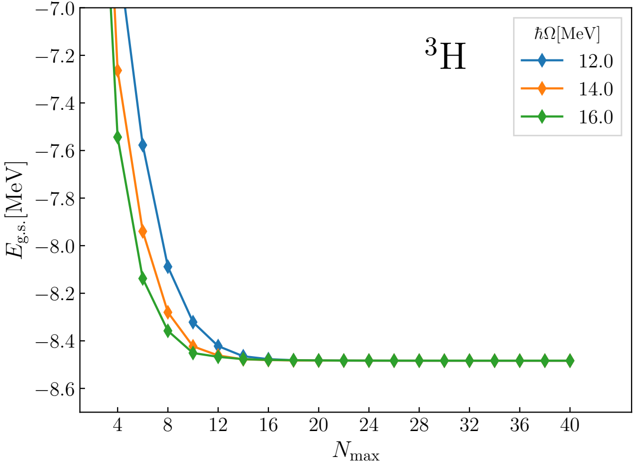
- monotonously decreasing with N_{max}

Convergence Behavior



- monotonously decreasing with $N_{\max}$
- different rates of convergence for different HO frequencies

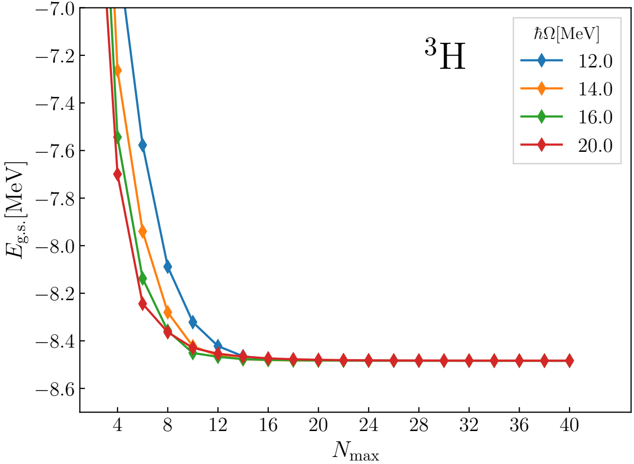
Convergence Behavior



- monotonously decreasing with N_{max}
- different rates of convergence for different HO frequencies

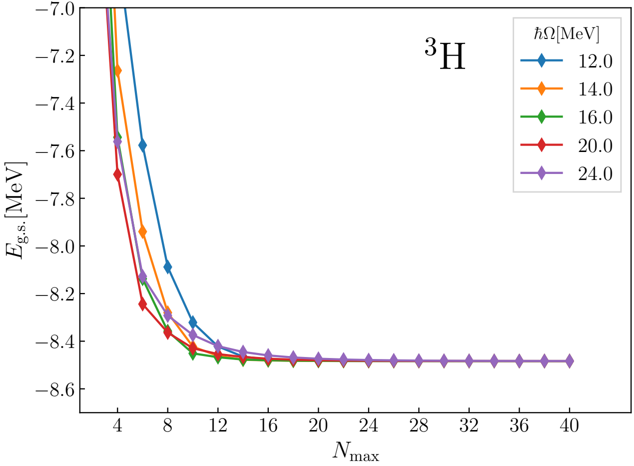
Convergence Behavior

TECHNISCHE
UNIVERSITÄT
DARMSTADT



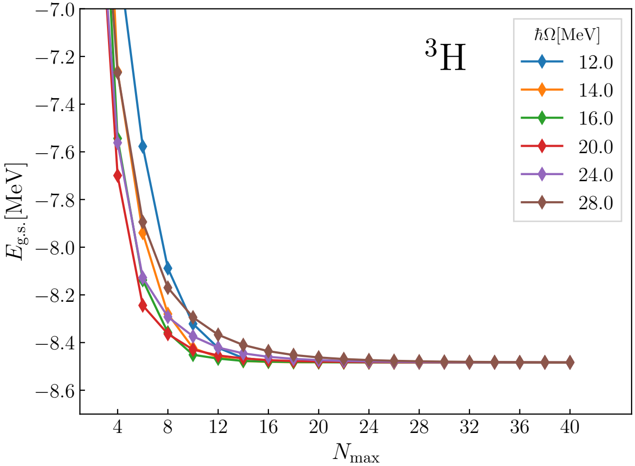
- monotonously decreasing with N_{max}
- different rates of convergence for different HO frequencies

Convergence Behavior



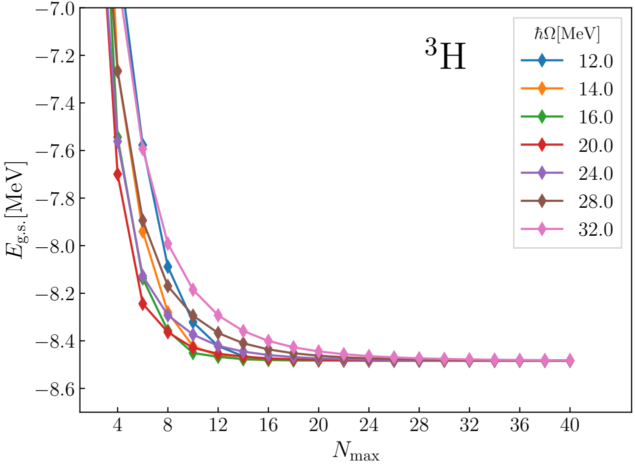
- monotonously decreasing with N_{max}
- different rates of convergence for different HO frequencies

Convergence Behavior



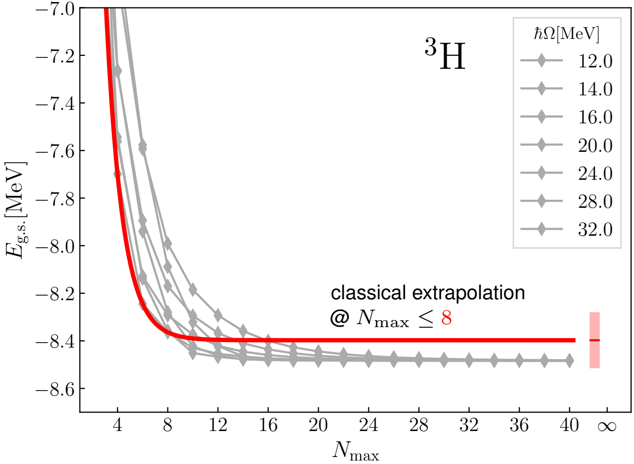
- monotonously decreasing with N_{max}
- different rates of convergence for different HO frequencies

Convergence Behavior



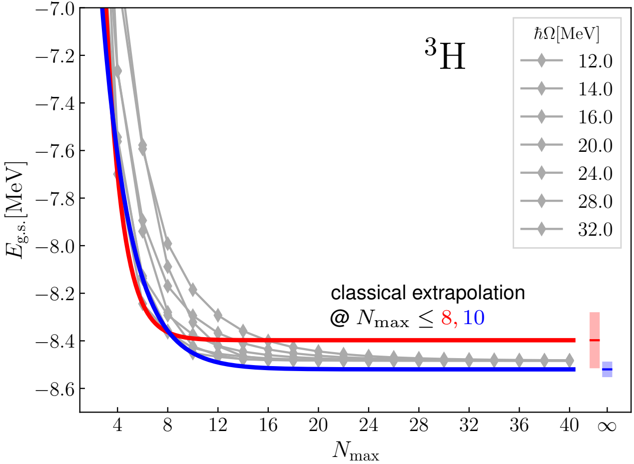
- monotonously decreasing with N_{max}
- different rates of convergence for different HO frequencies
- all sequences share the same limit

Convergence Behavior



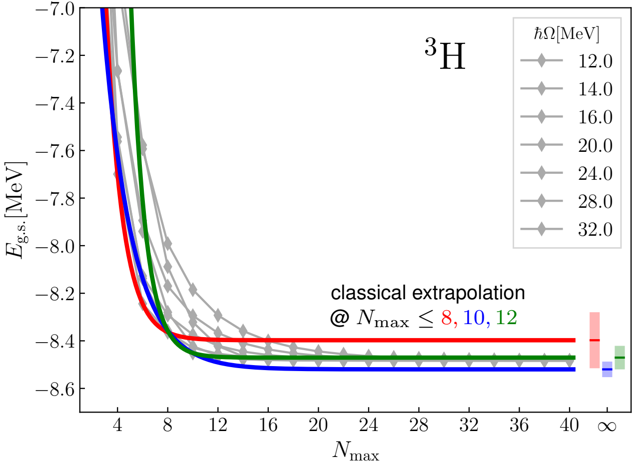
- monotonously decreasing with $N_{\max}$
- different rates of convergence for different HO frequencies
- all sequences share the same limit

Convergence Behavior



- monotonously decreasing with $N_{\max}$
- different rates of convergence for different HO frequencies
- all sequences share the same limit

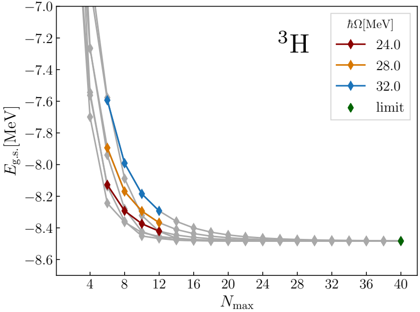
Convergence Behavior

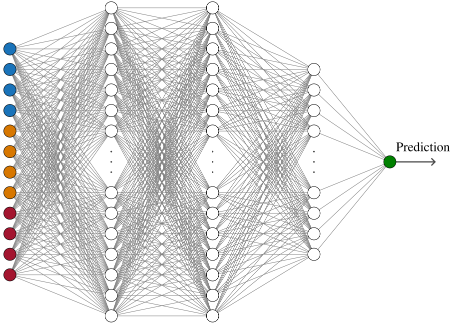


- monotonously decreasing with $N_{\max}$
- different rates of convergence for different HO frequencies
- all sequences share the same limit

Machine Learning Approach

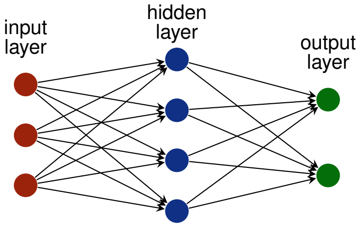
- previous applications: capture $f(N_{\max}, \hbar\Omega)$ — Negoita et al. PR C 99, 054308 (2019); Jiang et al. PR C 100, 054326 (2019)
- now: directly predict converged value from available calculations
  - include information of multiple frequencies

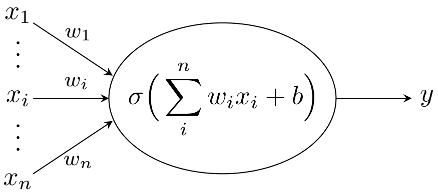




Artificial Neural Networks (ANNs)

- feedforward network
- neurons organized in layers
- resembles functional dependency $y^{\mathrm{out}} = f_{\mathrm{ANN}}(\mathbf{x}^{\mathrm{in}})$





- weights w_i and biases b as adjustable parameters
- nonlinearity through activation function σ

How to Train ANNs

- supervised learning: fit to labeled data
- loss function quantifies deviation between network's outputs and labels

$$L(\mathbf{x}^{\text{in}}, \mathbf{w}, \mathbf{b}) = \frac{1}{N} \sum_{i}^{N} ||f_{\text{ANN}}(\mathbf{x}_i^{\text{in}}, \mathbf{w}, \mathbf{b}) - l_i||^2$$

- backpropagation: adjust parameters through gradient descend (similar to linear regression)

$$w_i \to w_i' = w_i - \eta \frac{\partial L}{\partial w_i}\,, \quad b_i \to b_i' = b_i - \eta \frac{\partial L}{\partial b_i}$$

- control step size with learning rate η

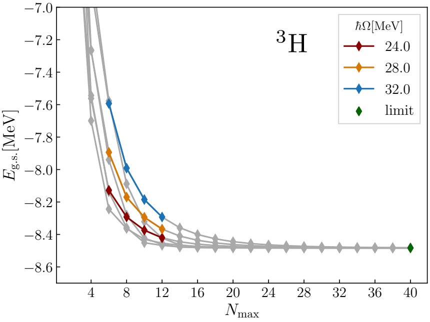
ANN Input Modes

ABS

- energy values:
  $E_{\hbar\Omega}^{N_{\max}}$
- sample shape:

$$\Big(E_{\hbar\Omega_1}^{\mathcal{N}_{\max}-6}, E_{\hbar\Omega_1}^{\mathcal{N}_{\max}-4}, E_{\hbar\Omega_1}^{\mathcal{N}_{\max}-2}, E_{\hbar\Omega_1}^{\mathcal{N}_{\max}},$$
$$E_{\hbar\Omega_2}^{\mathcal{N}_{\max}-6}, E_{\hbar\Omega_2}^{\mathcal{N}_{\max}-4}, E_{\hbar\Omega_2}^{\mathcal{N}_{\max}-2}, E_{\hbar\Omega_2}^{\mathcal{N}_{\max}},$$
$$E_{\hbar\Omega_3}^{\mathcal{N}_{\max}-6}, E_{\hbar\Omega_3}^{\mathcal{N}_{\max}-4}, E_{\hbar\Omega_3}^{\mathcal{N}_{\max}-2}, E_{\hbar\Omega_3}^{\mathcal{N}_{\max}}\Big)$$

- target:
  E^{∞}



ANN Input Modes

TECHNISCHE UNIVERSITÄT DARMSTADT

ABS

- energy values:
  $E_{\hbar\Omega}^{N_{\max}}$
- sample shape:

$$\Big(E_{\hbar\Omega_1}^{\mathcal{N}_{\max}-6}, E_{\hbar\Omega_1}^{\mathcal{N}_{\max}-4}, E_{\hbar\Omega_1}^{\mathcal{N}_{\max}-2}, E_{\hbar\Omega_1}^{\mathcal{N}_{\max}}, E_{\hbar\Omega_2}^{\mathcal{N}_{\max}-6}, E_{\hbar\Omega_2}^{\mathcal{N}_{\max}-4}, E_{\hbar\Omega_2}^{\mathcal{N}_{\max}-2}, E_{\hbar\Omega_2}^{\mathcal{N}_{\max}}, E_{\hbar\Omega_3}^{\mathcal{N}_{\max}-6}, E_{\hbar\Omega_3}^{\mathcal{N}_{\max}-4}, E_{\hbar\Omega_3}^{\mathcal{N}_{\max}-2}, E_{\hbar\Omega_3}^{\mathcal{N}_{\max}}\Big)$$

- target:
  E^{∞}

DIFF

- energy differences:
  $\Delta_{\hbar\Omega}^{N_{\max}} = E_{\hbar\Omega}^{N_{\max}} - E_{\hbar\Omega}^{N_{\max}-2}$
- sample shape:

$$\Big(\Delta_{\hbar\Omega_1}^{\mathcal{N}_{\max}-4}, \Delta_{\hbar\Omega_1}^{\mathcal{N}_{\max}-2}, \Delta_{\hbar\Omega_1}^{\mathcal{N}_{\max}}, \Delta_{\hbar\Omega_2}^{\mathcal{N}_{\max}-4}, \Delta_{\hbar\Omega_2}^{\mathcal{N}_{\max}-2}, \Delta_{\hbar\Omega_2}^{\mathcal{N}_{\max}}, \Delta_{\hbar\Omega_3}^{\mathcal{N}_{\max}-4}, \Delta_{\hbar\Omega_3}^{\mathcal{N}_{\max}-2}, \Delta_{\hbar\Omega_3}^{\mathcal{N}_{\max}}\Big)$$

- target:
  $\Delta^{\infty} = E^{\infty} - \min(E_{\hbar\Omega_i}^{\mathcal{N}_{\max}})$

Training Data & Data Preparation

TECHNISCHE UNIVERSITÄT DARMSTADT

1. few-body calculations

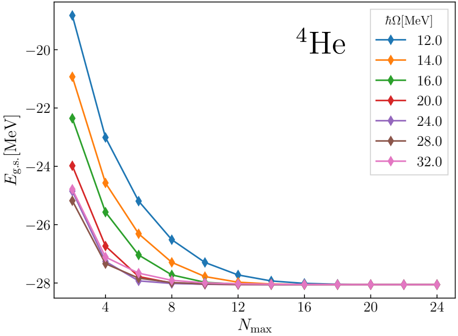
^{2}H, ^{3}H and ^{4}He
36 chiral interactions
7 HO frequencies

2. combinatorics

subsets & permutations

3. random scaling

$$a \cdot S + b$$
$$a \in (0.5, 2), b \in (-10, 10)$$



- 108 training sets
- family of realistic NN + 3N chiral interactions
  Hüther et al. PL B 808, 135651 (2020)

Training Data & Data Preparation

TECHNISCHE UNIVERSITÄT DARMSTADT

1. few-body calculations

^{2}H, ^{3}H and ^{4}He
36 chiral interactions
7 HO frequencies

2. combinatorics

subsets & permutations

3. random scaling

$a \cdot S + b$

$a \in (0.5, 2), b \in (-10, 10)$

3 nuclei
$\times$
36 interactions
$\times$
35 subsets of three $\hbar\Omega$
$\times$
16 subsets of four consecutive $N_{\max}$
$\times$
6 input permutations

=

362.880 samples

Training Data & Data Preparation

1. few-body calculations

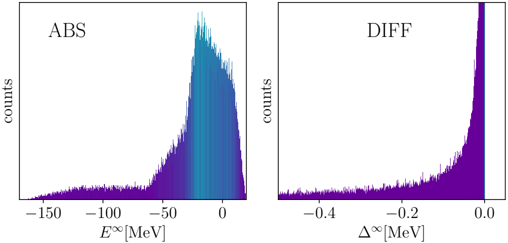
^{2}H, ^{3}H and ^{4}He
36 chiral interactions
7 HO frequencies

2. combinatorics

subsets & permutations

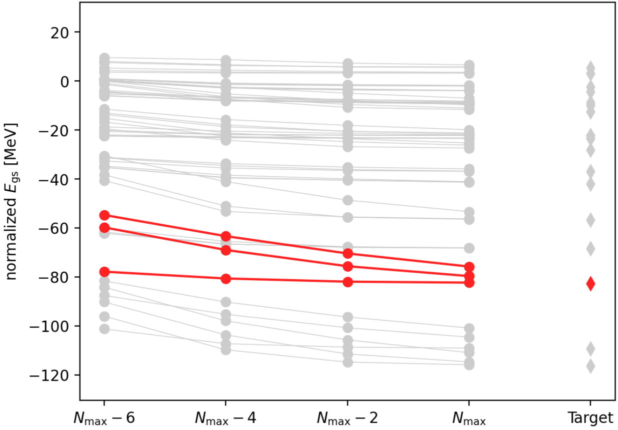
3. random scaling

$$a \cdot S + b$$
$$a \in (0.5, 2), b \in (-10, 10)$$



- training: 1.000.000 samples
- validation: 50.000 samples

ABS Samples


20
0
−20
−40
−60
−80
−100
−120
normalized E_{gs} [MeV]
$N_{max} - 6$
$N_{max} - 4$
$N_{max} - 2$
N_{max}
Target

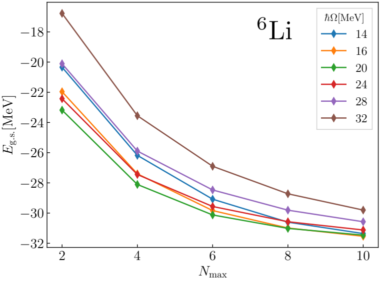
Finalizing the Machine Learning Tool

TECHNISCHE
UNIVERSITÄT
DARMSTADT

- train 1000 different networks
- accuracy of 60 keV (ABS) or 6 keV (DIFF) on the validation data
- store trained networks ready for application

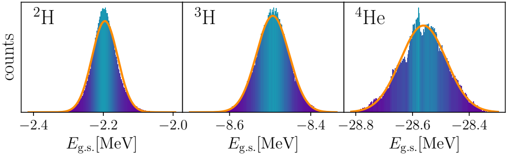
universal tool directly applicable to any nucleus and interaction

Statistical Evaluation

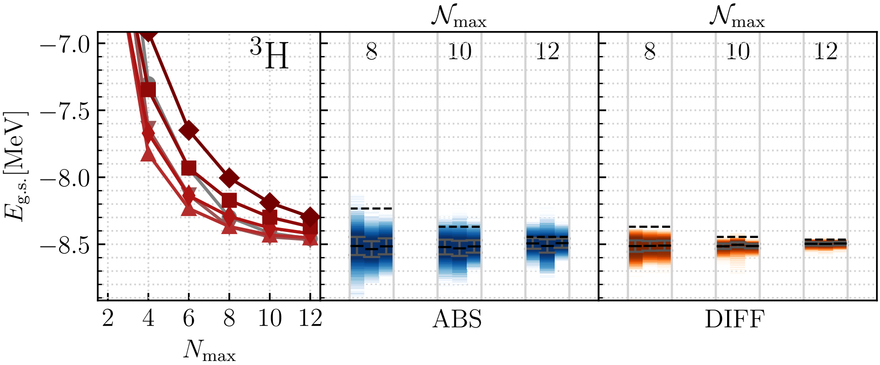


- different family of interactions
  Maris et al. PR C 103, 054001 (2021)
- construction of evaluation samples analogoulsy to training samples
- different predictions from one ANN
- turn to statistical approach

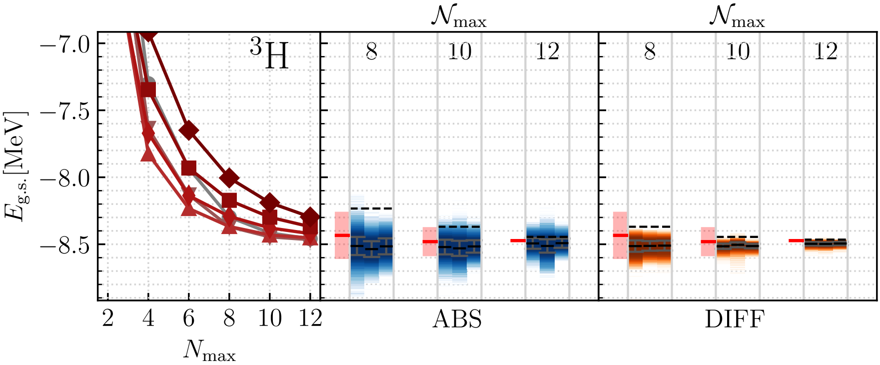
- apply 1000 ANN
- prediction and uncertainty from Gaussian fit



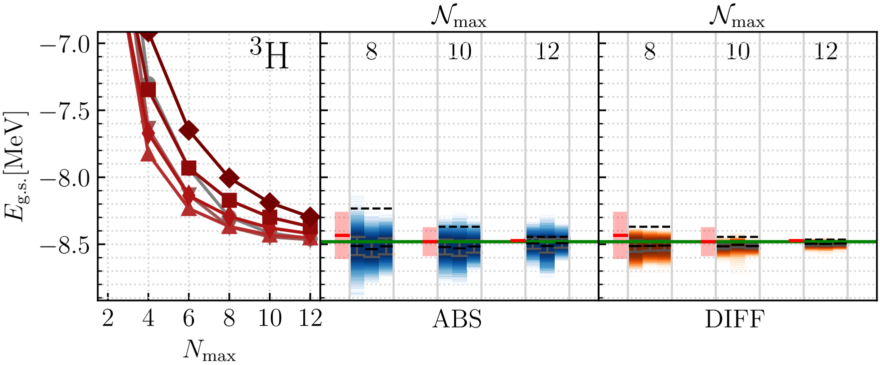
Application to Few-Body Systems



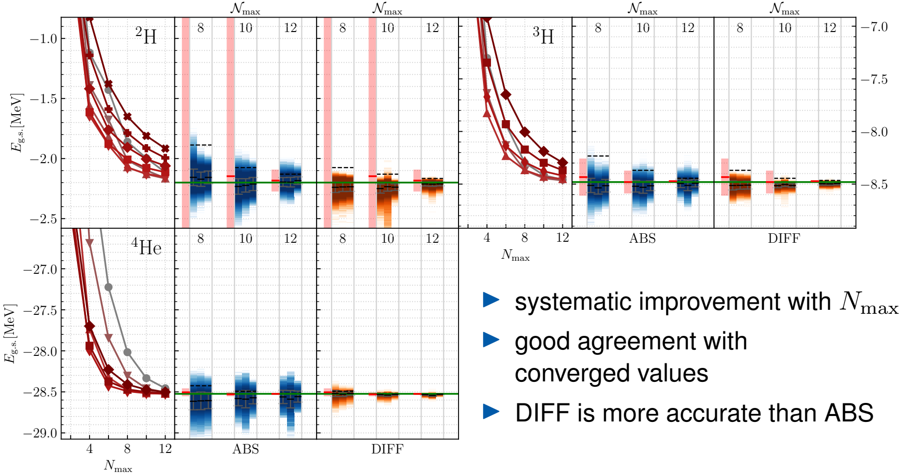
Application to Few-Body Systems



Application to Few-Body Systems

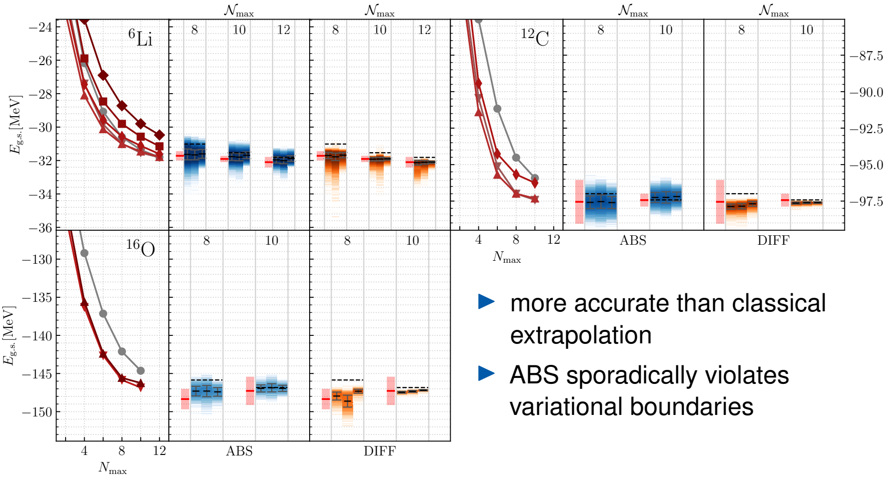


Application to Few-Body Systems



- systematic improvement with $N_{\max}$
- good agreement with converged values
- DIFF is more accurate than ABS

Application Beyond Few-Body Systems



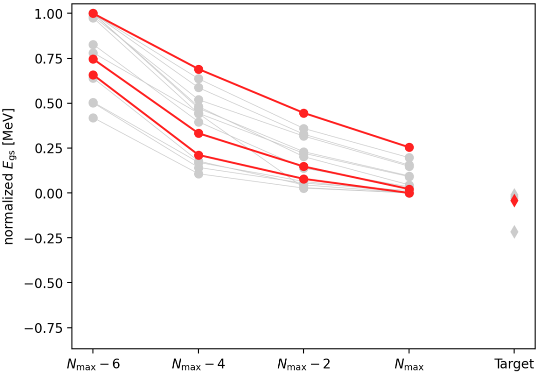
- more accurate than classical extrapolation
- ABS sporadically violates variational boundaries

New Developments
ANN Input Normalization

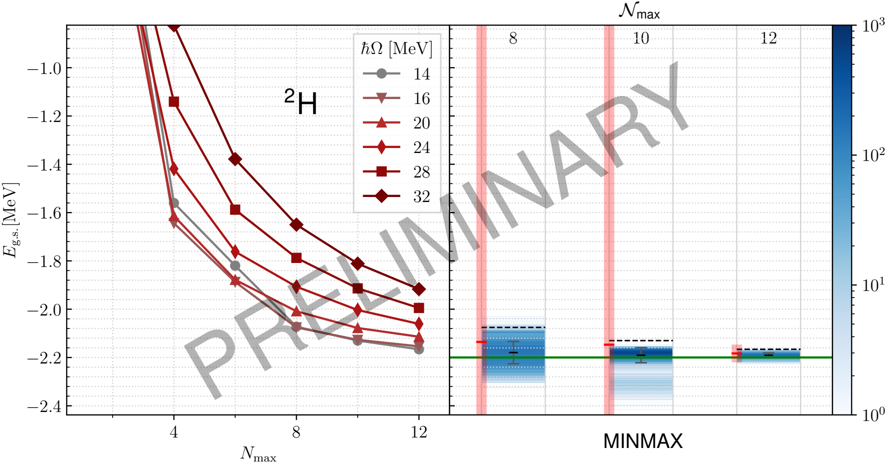
TECHNISCHE UNIVERSITÄT DARMSTADT

MINMAX

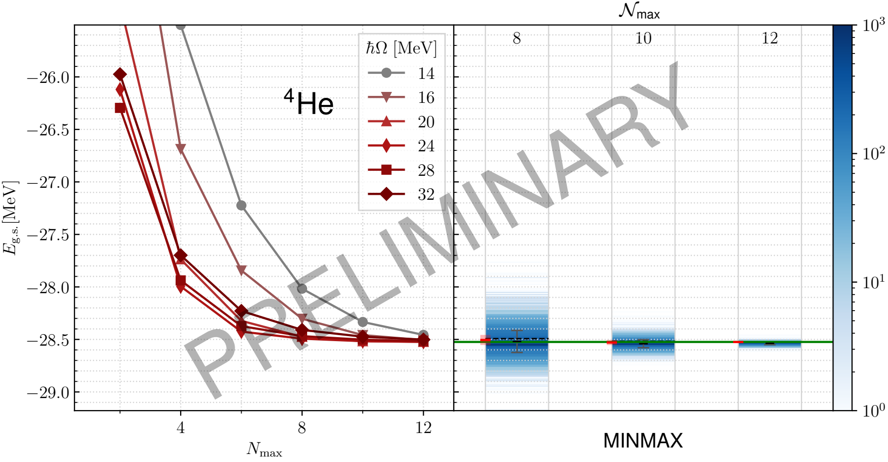
- normalization alleviates scale dependence
- normalize every input sample individually
- scale and shift training sample so that all data points are between 0 and 1



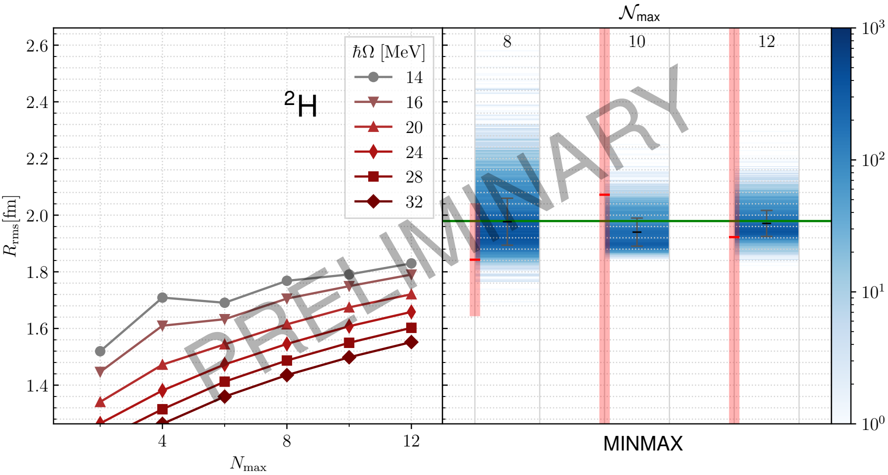
New Developments
MINMAX



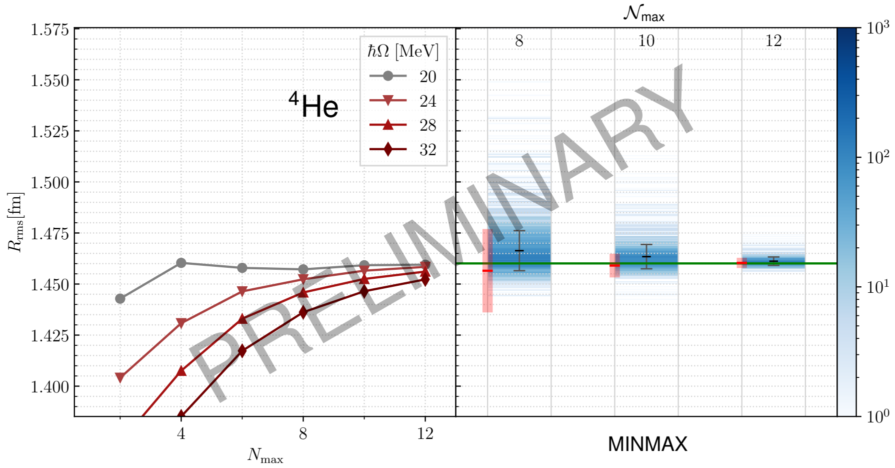
New Developments
MINMAX



New Developments
Radii



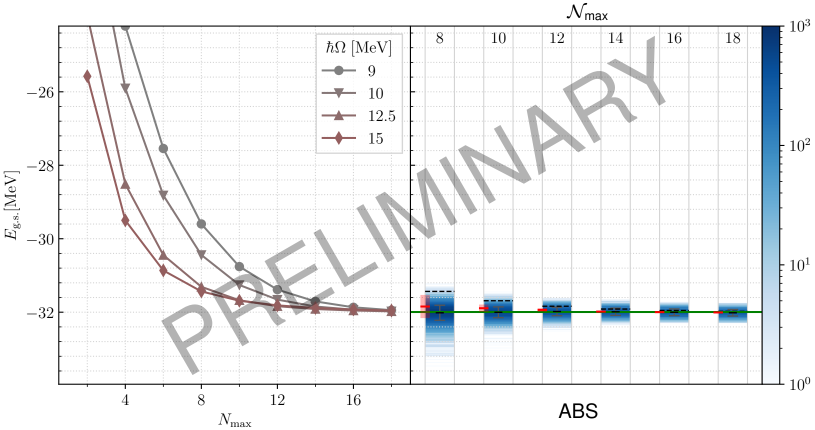
New Developments
Radii



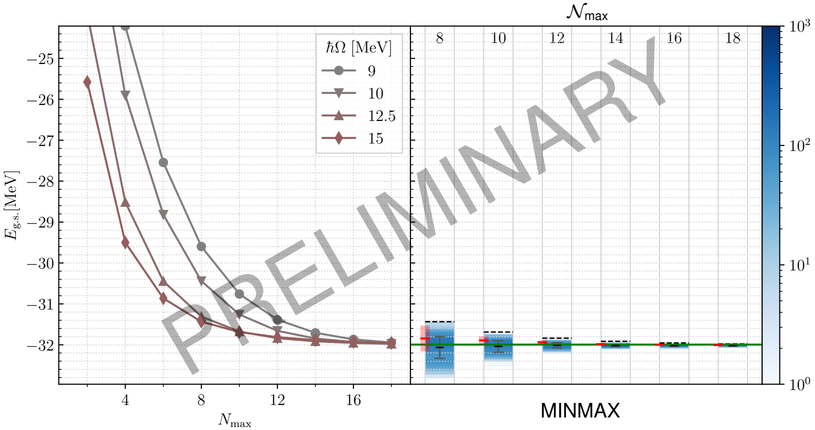
New Developments
Synthetic data

- idea: modify Hamiltonian to produce more varied training data
- introduce scaling factor in front of potential: $H_\lambda = T + \lambda \cdot V$
- improve training data quality
  - ABS/DIFF: allows to forgo scaling and shifting
    $\Rightarrow$ more truely unique training data
  - Minmax: adds variety, since scaling and shifting samples has no effect due to normalization
- might enable the calculation of otherwise inaccessible observables
  e.g. $B(\mathrm{E2})$ in few-body systems by binding first 2^+ state

Benchmark
^{6}Li



Benchmark
^{6}Li



Benchmark
^{6}Li

N_{max}	ABS	DIFF	MINMAX
8	-32.02(21)	-31.67(10)	-32.06(27)
10	-32.00(15)	-31.86(6)	-32.04(14)
12	-31.98(11)	-31.94(3)	-32.01(6)
14	-31.99(10)	-31.97(2)	-32.02(3)
16	-32.00(9)	-31.99(1)	-32.02(2)
18	-32.02(9)	-32.00(1)	-32.02(1)

Variational boundary: -31.977

Conclusion & Outlook

TECHNISCHE UNIVERSITÄT DARMSTADT

- ANNs provide robust predictions with reliable uncertainty estimates
  - more accurate than classical extrapolations
  - robust w.r.t. changes in the training data
- applicable to any nucleus accessible via NCSM
- extension to radii (work in progress) and other observables
  - challenge: more complex convergence patterns
- great potential for optimization:
  - normalization of training data
  - adjustment of topology and hyperparameters
- more: Knöll, TW, Agel, Wenz, Roth: arXiv:2207.03828

Thank you for your attention!

TECHNISCHE UNIVERSITÄT DARMSTADT

- **thanks to my group and collaborators**

M. L. Agel, M. Knöll, L. Mertes, T. Mongelli, J. Müller, D. Rodriguez, **R. Roth**, L. Wagner, C. Wenz, N. Zimmermann

Bundesministerium für Bildung und Forschung

SFB 1245

computing time

Hessisches Kompetenzzentrum für Hochleistungsrechnen

DFG