Non-parametric Bayesian approach to the extrapolation in configuration interaction methods

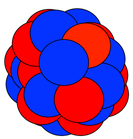
Progress in Ab Initio Techniques in Nuclear Physics,
Feb.25 - Mar.1, 2019, TRIUMF Vancouver BC Canada

Sota Yoshida

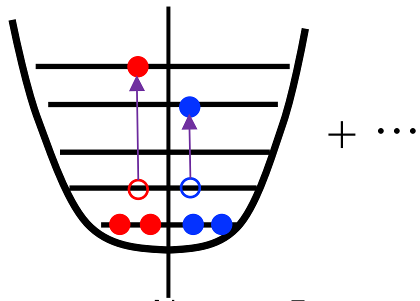
THE UNIVERSITY OF TOKYO

Extrapolations in NCFC(NCSM)/MCSM

No-Core Full Configuration (NCFC)



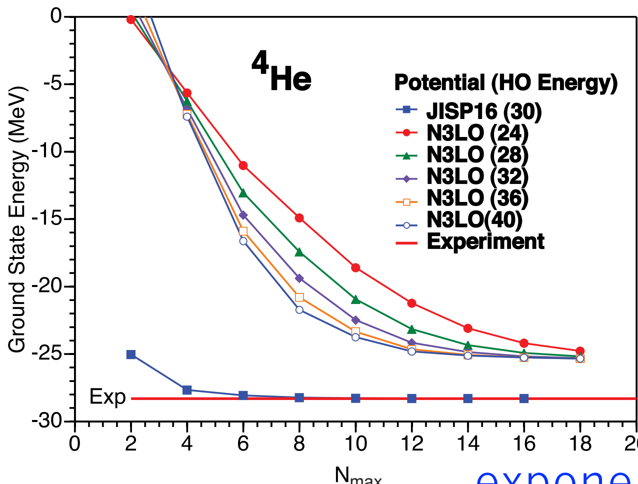
=



e.g. Nmax =5

N_{max}: maximum excitation from lowest config.

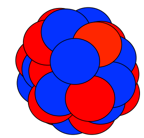
exact results ~ $N_{max} \to \infty$



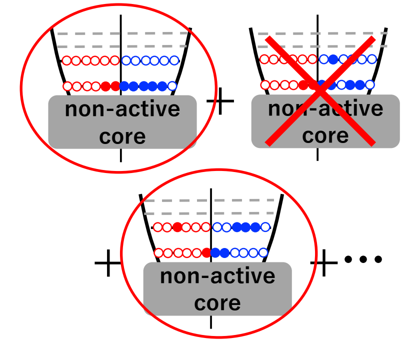
exponential ?

P. Maris et al., PRC **79**, 014308 (2009).

Monte Carlo Shell Model (MCSM)



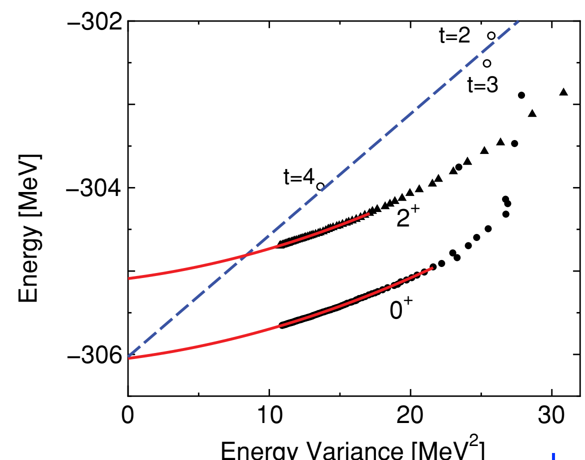
=



Importance Truncation

# of MCSM basis→ original dim. ~∞

or Energy variance $<\Delta H^2> \to 0$



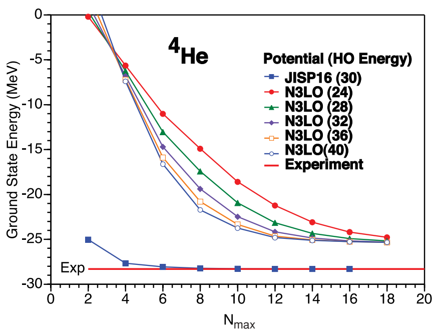
polynomial ?

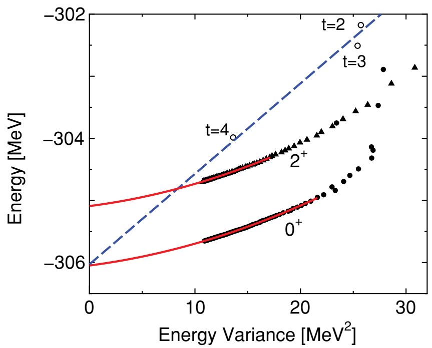
N. Shimizu et al., PRC **82**, 061305(R) (2010).

parametric vs non-parametric

parametric (conventional)

exponential or (low-rank) polynomial





risk of overfitting

(exception: there is “underlying mechanism”)

☒ not robust to outliers
(need to get rid of outliers)

☒ point estimation of “parameter” for
exponential/polynomial

→ difficult to see uncertainty
coming from “fit”

✔ **strong conclusions** under **“assumptions”**

non-parametric (proposed)

Gaussian Processes(**GPs**)
resolve these problems

Bayes GP model

☑ “Regularization term” is naturally introduced

→ avoid overfitting

(GP is mathematically equivalent to
ANN with 1 hidden layer & # of node →∞)

☑ Robust against outliers

☑ It says “I’m not quite sure”
in the region very far from data

☑ proper UQ
predictions are given as prob. distribution

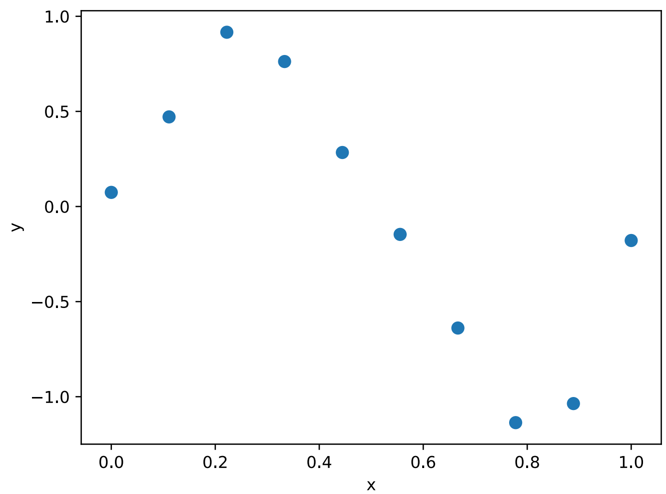
✔ **weak(=safer) conclusions** under
data+ **physical principle**

What are GPs?

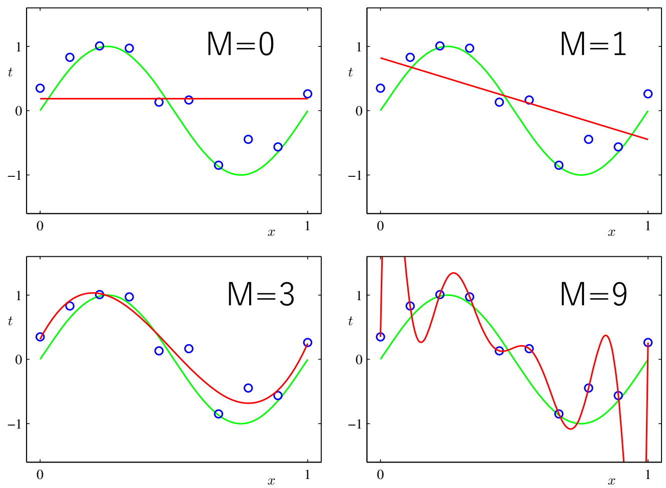
Toy Problem

10 data points
$D = \{(x_i, y_i) | i=1,2,\cdots,10\}$

y^* at unobserved points x^*



polynomial? Artificial Neural Network?



red: prediction, blue: data, green: true function

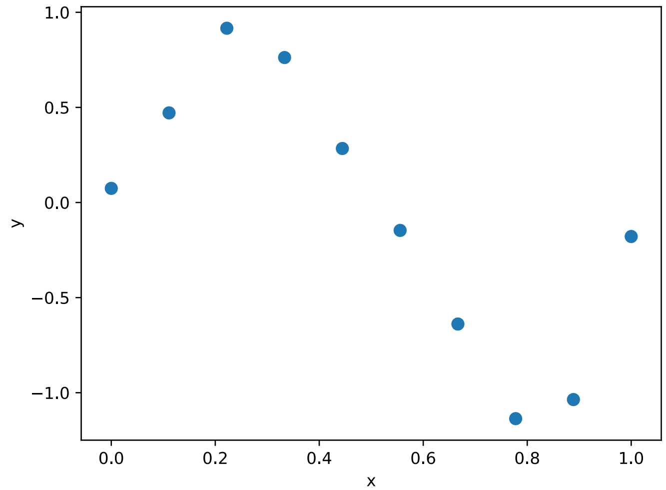
Figs. from PRML (Springer, 2006)

What are GPs?

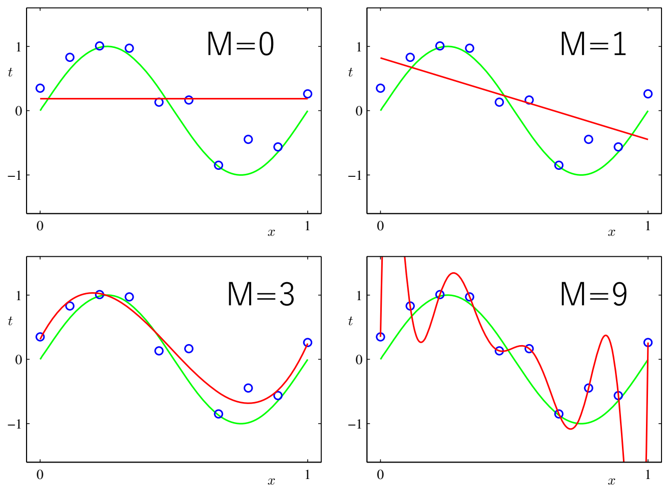
Toy Problem

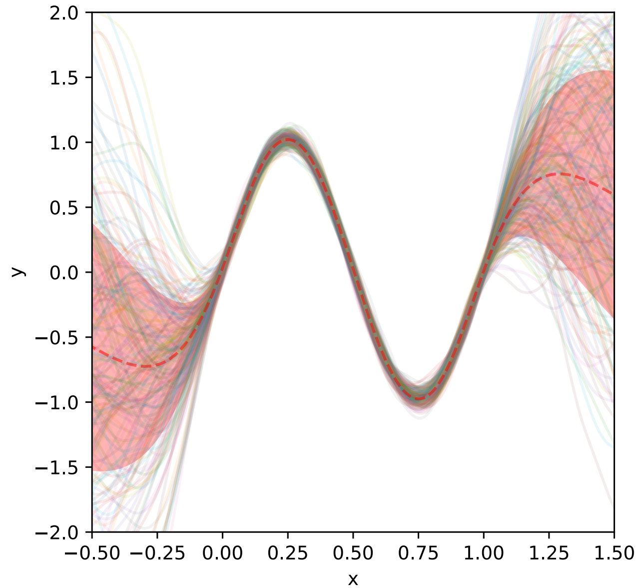
10 data points
$D = \{(x_i, y_i) | i = 1, 2, \cdots, 10\}$

y^* at unobserved points x^*



polynomial? Artificial Neural Network?





red: prediction, blue: data, green: true function

Figs. from PRML (Springer, 2006)

What are GPs?

Essential idea of GPs:

1.

2.

3.

Essential idea of GPs:

1. two target values y_i, y_j at near two points x_i, x_j must be “similar”

2.

3.

What are GPs?

Essential idea of GPs:

1. two target values y_i, y_j at near two points x_i, x_j must be “similar”

2. This “similarity” is defined by Kernel, a function of distance $|\mathbf{x}_i - \mathbf{x}_j|$

e.g. RBF Kernel $$k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \tau \exp\left(-\frac{||\boldsymbol{x}_i - \boldsymbol{x}_j||^2}{2\sigma^2}\right)$$

τ,σ: hyperparameters

3.

What are GPs?

Essential idea of GPs:

1. two target values y_i, y_j at near two points x_i, x_j must be “similar”

2. This “similarity” is defined by Kernel, a function of distance $|\mathbf{x}_i-\mathbf{x}_j|$

e.g. RBF Kernel $$k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \tau \exp\left(-\frac{||\boldsymbol{x}_i - \boldsymbol{x}_j||^2}{2\sigma^2}\right)$$

τ,σ: hyperparameters

3. Assume that <u>**y** and **y*** are obeying Multi-dimensional Gaussian</u>

whose covariance matrix is defined by Kernel function

correlation among data points(N × N)

corr. b/w data & prediction points (N × M)

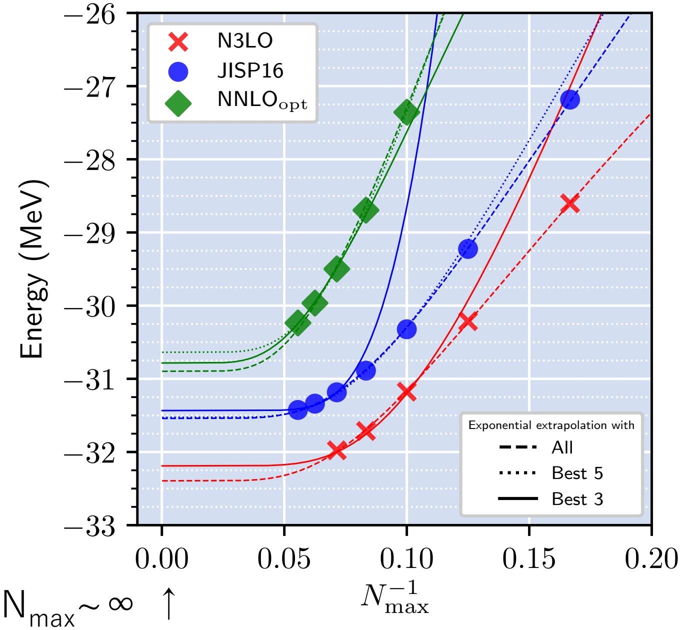
$$P(\boldsymbol{y}, \boldsymbol{y}^*) \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu} \\ \boldsymbol{\mu}^* \end{bmatrix}, \begin{bmatrix} K_{XX} & K_{XX^*} \\ K_{XX^*}^T & K_{X^*X^*} \end{bmatrix}\right)$$

corr. among prediction points (M × M)

$X = \{x_i \mid i=1,\ldots,N\}$
$X^* = \{x_j^* \mid j=1,\ldots,M\}$

* Usually, data points are normalized, i.e. $\mu=0$, $\mu^*=0$

NCFC results of ^{6}Li g.s. energy



$N_{max} \sim \infty$ ↑

× : N3LO (N_{max}=6-14)

M. Kruse et al., PRC **87**, 044301(2013).

● : JISP16 (N_{max}=2-18)

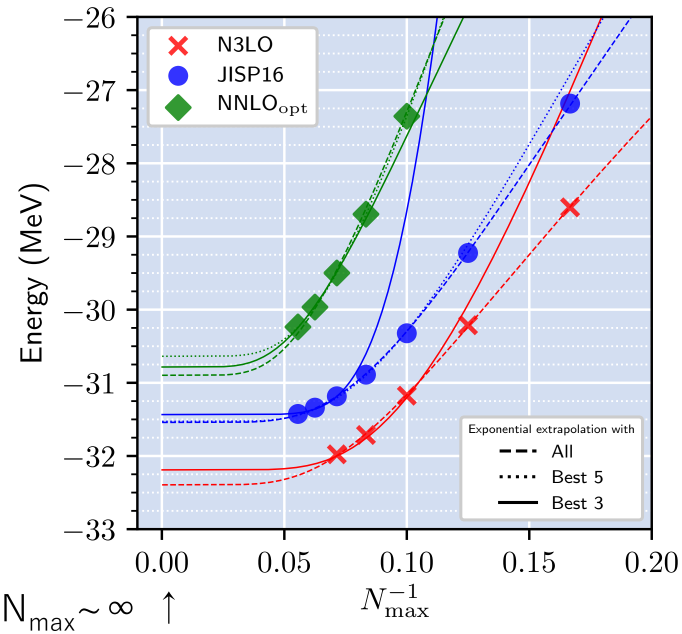
◆ : NNLOopt(N_{max}=2-18)

I.J.Shin et al., J. Phys. G: Nucl. Part. Phys. 44, 075103 (2017).

Lines: exponential fit using

- All data (dashed)
- Best 5 data (dotted)
- Best 3 data (solid)

NCFC results of ^{6}Li g.s. energy



$N_{max}\sim\infty$ ↑

✕ : N3LO (N_{max}=6-14)

M. Kruse et al., PRC **87**, 044301(2013).

● : JISP16 (N_{max}=2-18)

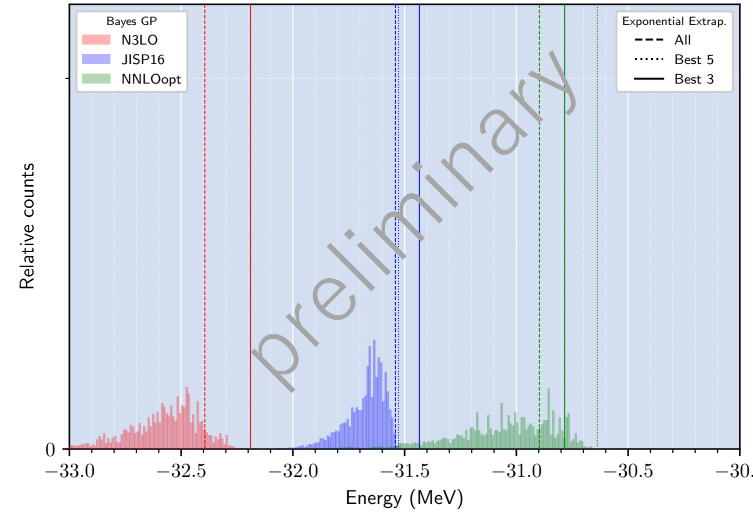
◆ : NNLOopt(N_{max}=2-18)

I.J.Shin et al., J. Phys. G: Nucl. Part. Phys. 44, 075103 (2017).

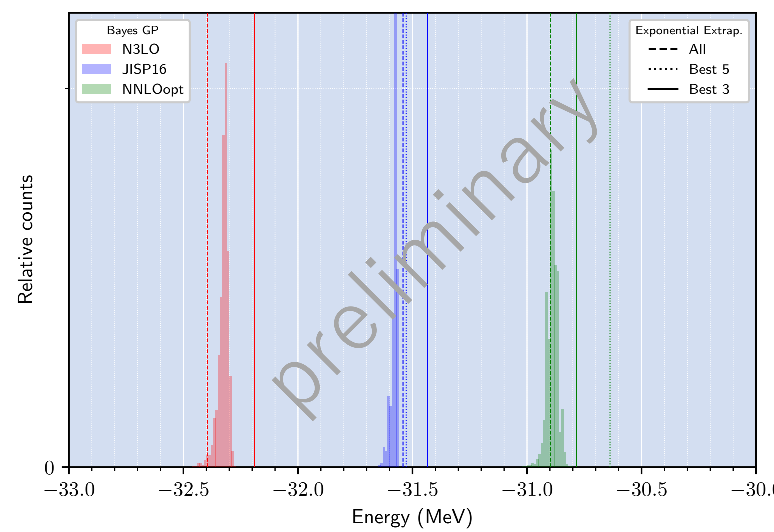
Lines: exponential fit using

- All data (dashed)
- Best 5 data (dotted)
- Best 3 data (solid)

Bayes GP (with monotonicity)

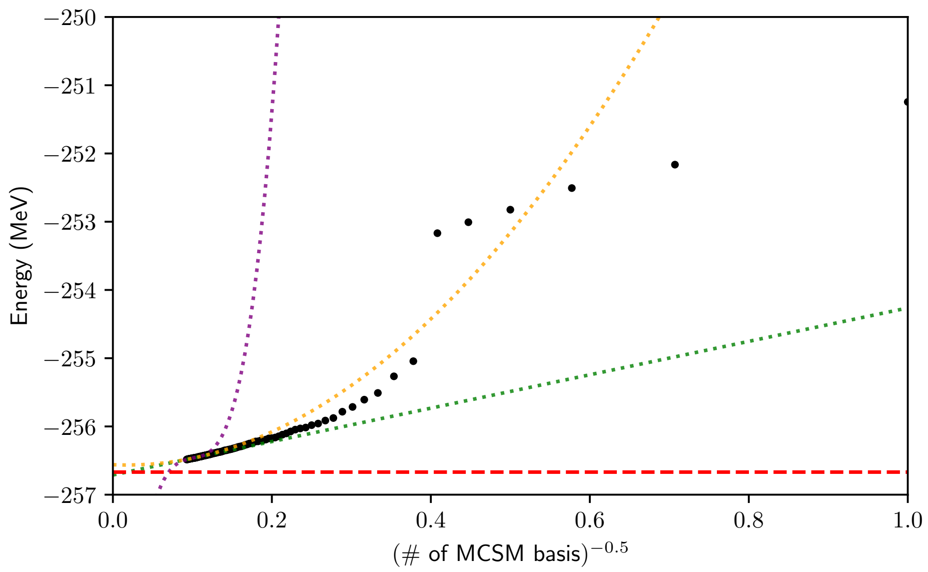


+convexity

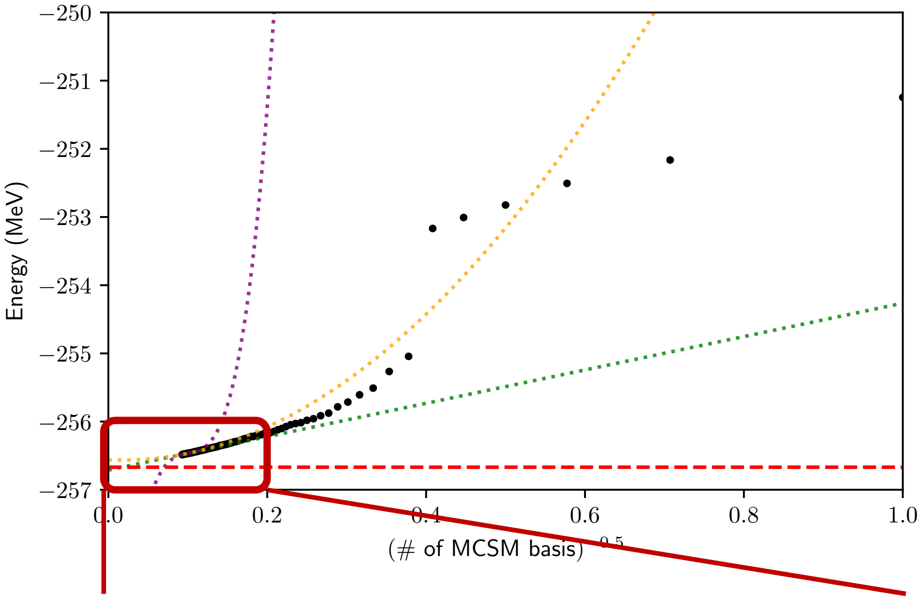


consistent with “Extrap B”

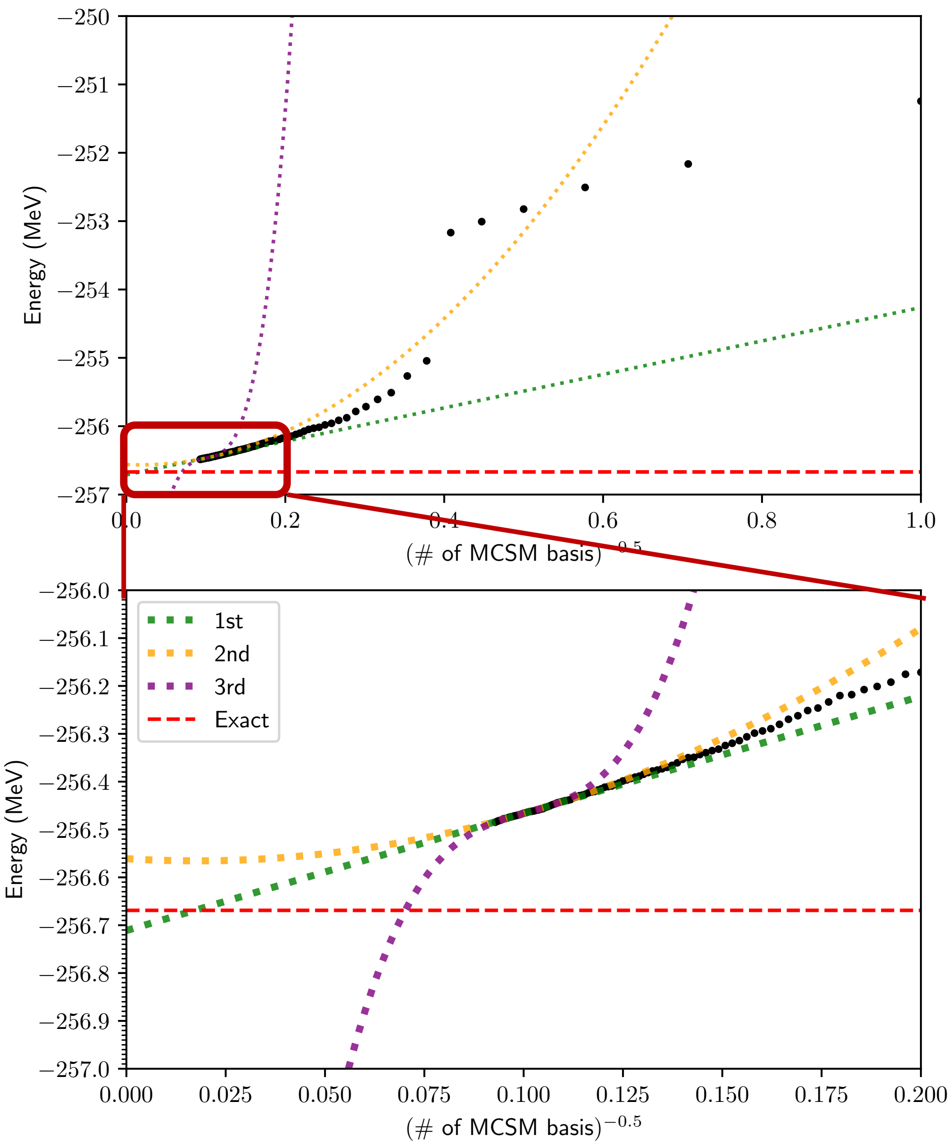
MCSM extrapolation: ^{76}Sr g.s.



MCSM extrapolation: ^{76}Sr g.s.

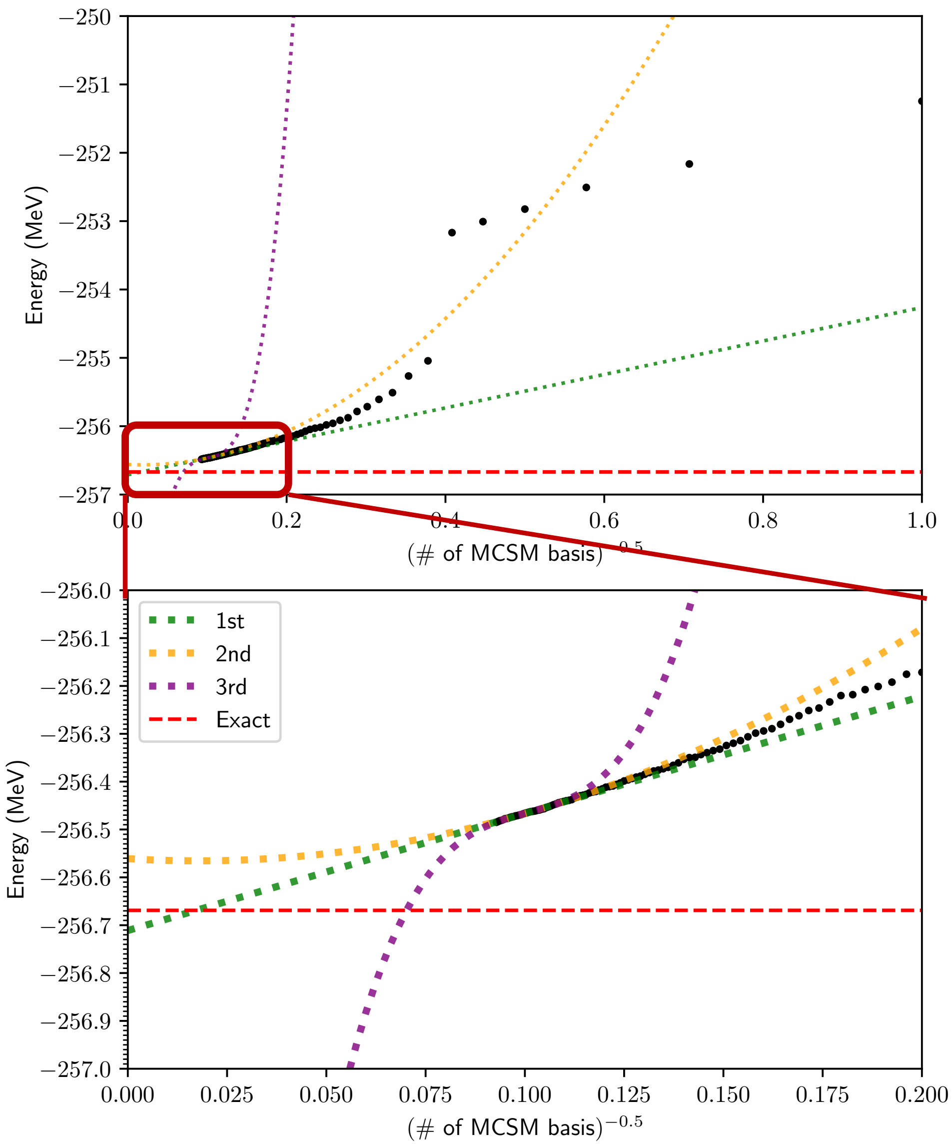


MCSM extrapolation: ^{76}Sr g.s.



←Scattered predictions due to the limited expression power of each polynomial

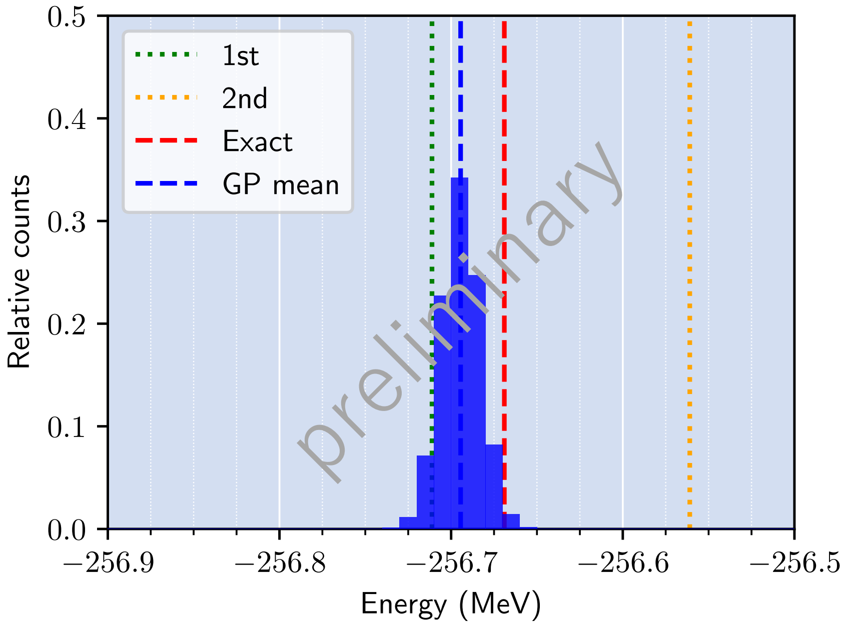
MCSM extrapolation: ^{76}Sr g.s.



←Scattered predictions due to the limited expression power of each polynomial

Bayes GP model may tell safer and reasonable predictions

More validity checks are needed!!



Summary

Bayes GP model

☑ Flexible non-parametric extrapolation method

applicable to weird shape function

☑ Easy to incorporate *"physics"* (monotonicity, etc.)

☑ Proper uncertainty quantification

Future perspectives

- Open the core part of code
- Application to NCMCSM, IT-(NC)SM, other observables,

  strongly correlated electrons system, etc.

More details

- Poster
- Upcoming paper
- (free) Gaussian Process for Machine Learning
  by Carl Edward Rasmussen and Christopher K. I. Williams