Neural Network Fitting of Total Energy and Atomic Forces of Bulk Hexagonal Boron Nitride under Area Elements Representation

Man Kit CHAN
Department of Physics, The Chinese University of Hong Kong

This project is an extension of PHYS4610, we present a Neural Network (NN) approach for fitting total energy and atomic forces based on an area element method, using randomly perturbed structures with Density Functional Theory (DFT) calculated total energy and atom forces as training data. Our method is tested on bulk hexagonal Boron Nitride (h-BN), and the fitting results show similar accuracy as DFT calculations with a small extent of extrapolation ability. This approach can possibly be generalized to other types of crystal structures and speed up structure properties predictions, serving as a potential substitute of DFT with less computational cost.

I. INTRODUCTION

A. Background and Motivation

In DFT, the total energy of a many-body system can be expressed as kinetic term, Hartree energy term, exchange and correlation (XC) term and external potential term in functionals of electron density.

$$(-\frac{\hbar^2}{2m}\nabla^2 + v_{eff})\psi_i = \epsilon_i\psi_i \tag{1}$$

$$v_{eff}(\boldsymbol{r}) = v_{ext}(\boldsymbol{r}) + v_{Hatree}[\rho(\boldsymbol{r})] + v_{XC}[\rho(\boldsymbol{r})] \tag{2}$$

Eq. (1) shows the Kohn Sham Equations where v_{eff} is the effective potential, ψ_i and ϵ_i are the Kohn-Sham orbital and the respective eigenenergy. Eq. (2) shows the expression of v_{eff}, including external potential, Hatree potential, XC potential, where $\rho(\boldsymbol{r})$ is electron charge density. In practical cases, external potential and XC potential require approximations.

Iterative optimization algorithms are able to solve Eq. (1) and Eq. (2), below shows a simplified self-consistent scheme: 1, start with an initial guess for electron density; 2, calculate the effective potential; 3, solve the Kohn-Sham equations for the Kohn-Sham orbitals and orbitals energy; 4, update the electron density then repeat from step 2 until convergence.

The tradeoff of the high accuracy DFT calculations is computational time and resources. Machine learning has been a potential candidate to give near DFT-quality predictions with much less computational cost and power. Many previous attempts have used different representations of many-body systems to fit physical properties.

B. Prior Progress

In the previous project PHYS4610, we adapted a similar neural network approach from Behler and Parrinello, which involves using neural network to fit the sub-energy of each element in the structure, such that the sum of sub-energies yields to total energy of the structure.

In contrast to Behler and Parrinello's proposal of using atom-based symmetry functions as representation. We used a completely different approach to represent the structure, we propose to define area element based on local configurations of the center, first and second neighbor atoms, such that the geometry of element should captures the deformation around the center atom and gives distinct information about the local electron density .

The definition of area element is based on two principles. The first is the consistency principle, all elements should cover the entire structure and each element should have the same atomic count as the empirical chemical formula. The second is the symmetry principle, the atom under study is near the element's center and the element should be symmetrical about the center atom.

To build our dataset, we apply random small perturbations on all atoms of a relaxed 60-atom h-BN cell to sample the phase space around the minimum configuration. A total of 4450 perturbed structures are generated by random sampling, and their total energies and net forces on atoms are numerically calculated by The Vienna Ab initio Simulation Package (VASP) using DFT. These calculations serve as the ground truth of each structures and forms our training and testing data.

In the previous report, there are detailed documentation about preparations of training and testing data, along with primary fitting results of total energy, comparisons and discussions on our area element method and Behler and Parrinello's symmetry functions method.

C. Current State of Progress

In this report, we first slightly reformulate and update the method for defining area element. Then, we discuss and compare the approaches that are experimented to fit the net forces on center atom of area element. Finally, we show results of the final NN model in simultaneously fitting both total energy and net forces on atoms of bulk h-BN structures.

II. METHOD

A. Area Element

We provide a systematic procedure of defining area element, it is the same for both Boron-centered or Nitrogen-centered element, the following example demonstrates the procedure for Boron-centered element. Choose a Boron atom as center (O), and the 3 first nearest neighboring atoms (A, E, I) are treated as the atom vertices of the area element. For the non-atom vertices (X, Y, Z), the surrounding hexagonal ring of atoms will be considered and their averaged position is computed as the non-atom vertex. Figure (1) illustrate such approach.

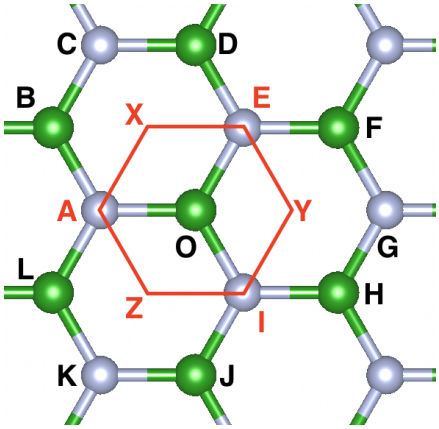

FIG. 1: Boron-centered area element outlined in red, Boron atoms are in green and Nitrogen atoms are in silver. Select one center Boron atom (O), the first Nitrogen atom neighbors (A, E, I) are used as the atom vertices of the element. Then, the average coordinates of surrounding hexagonal ring are computed for non-atom vertices. Ring (A, B, C, D, E, O) corresponds to vertex (X), ring (E, F, G, H, I, O) corresponds to vertex (Y), ring (I, J, K, L, A, O) corresponds to vertex (Z).

Following this definition, each area element encloses one Boron and one Nitrogen exactly, satisfying the electron counting and the empirical chemical formula of BN. Also, the hexagonal area element will still be relatively symmetrical about the center under slight perturbation.

B. Higher Order Representation of Area Element

The hexagonal area element mentioned above is the simplest form of area element, it is a basis which spans through the whole structure without overlaps. However, the calculation of non-atom sites involves averaging the coordinates of nearest neighbours which loses the actual positions of some nearby atoms. We can view this as a low order approximation of geometry of such area element.

We can also define area element without using non-atom vertices, which the three hexagonal rings around the center atom can directly be used as vertices. The high order area element is illustrated in Figure (2).

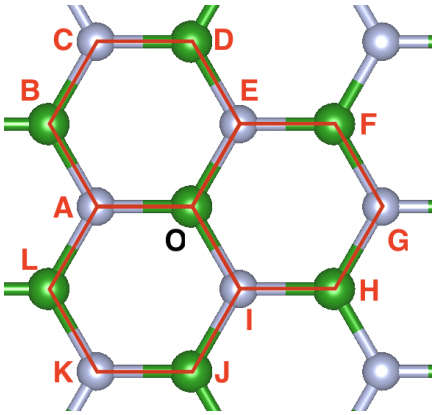

FIG. 2: High order representation of Boron-centered area element outlined in red, all atoms of the three hexagonal rings (A-L) near the center Boron (O) are used as vertices.

This high order representation encloses three Boron and three Nitrogen exactly, satisfying the electron counting and the empirical chemical ratio of BN, it is also relatively symmetric under slight perturbation. Moreover, it includes positional information of all nearby atoms, which is more suitable for force fitting which requires precise local configuration of neighbours.

It is also possible even further expand the scope of area element by including more neighbours while satisfying electron counting and empirical chemical ratio. But our results have shows that representation in Figure (2) already gives highly accurate fitting of both total energy and forces.

C. Neural Network

The NN used in the project is a fully connected network of 3 hidden layers of 32, 64, 32 neurons with Leaky ReLU activations. The inputs are the vertices of area element, the outputs are the sub energy and net forces on center atom of area element. The sum of sub energies predictions yields the total energy prediction of a structure. The loss is computed as Mean Squared Error (MSE) of both total energy and net forces on atoms compared to DFT calculations, and is backward propagated for optimization. The model is implemented in PyTorch and trained for 50 epochs.

Equation (3) states the definition of MSE, $\hat{y}$ is the prediction, y is the truth, n is the total number of samples.

$$MSE = \sum_{i=0}^{n} (\hat{y} - y)^2 / n \tag{3}$$

III. RESULTS

We have experimented with both low-order and high-order area elements and fitted the total energy and forces on center atom of area elements of h-BN structures using NN. This section shows the testing set performances of using Nitrogen-centered area element. The error metric used is Root Mean Squared Error (RMSE), which is simply the square root of MSE. For detailed supplementary information, Appendix A includes the full results for Nitrogen-centered and Boron-centered elements.

A. Testing Set Performance

Table (I) shows the overall performance of on testing set. Each method is fitted for 3 trials and the averaged values are taken to reduce the randomness due to NN initialization. It is found that the performance between training set and testing set are similar, thus no significant sign of overfitting. To further minimize overfitting, better sampling from the phase space and larger dataset are needed for training. Also, regularization or weight decay can be added during the training process.

TABLE I: Overall fitting performance comparison using low order and high order Nitrogen-centered area element

Methods	Low Order	High Order
Num of Vertices	6	12
Num of Inputs	12	24
RMSE of Total Energy (eV)	0.029873	0.029015
RMSE of Energy per Atom (meV)	0.497883	0.483578
RMSE of x-component Force on Center Atom (eV/Å)	0.160153	0.071714
RMSE of y-component Force on Center Atom (eV/Å)	0.160673	0.072221
RMSE of Force Magnitude on Center Atom (eV/Å)	0.159276	0.071915
Mean Angular Error of Atomic Force (Degree)	7.788658	3.491130

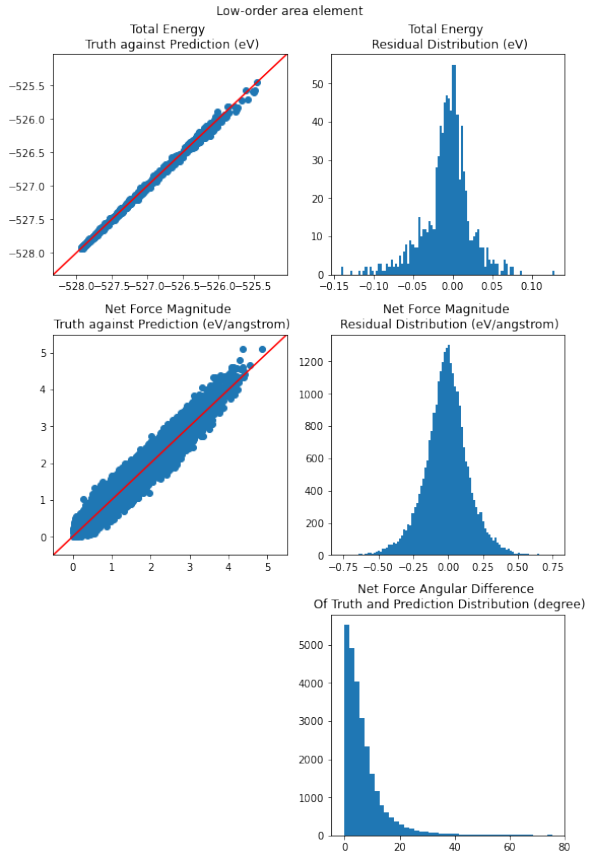

FIG. 3: Low order Nitrogen-centered area element fitting result.

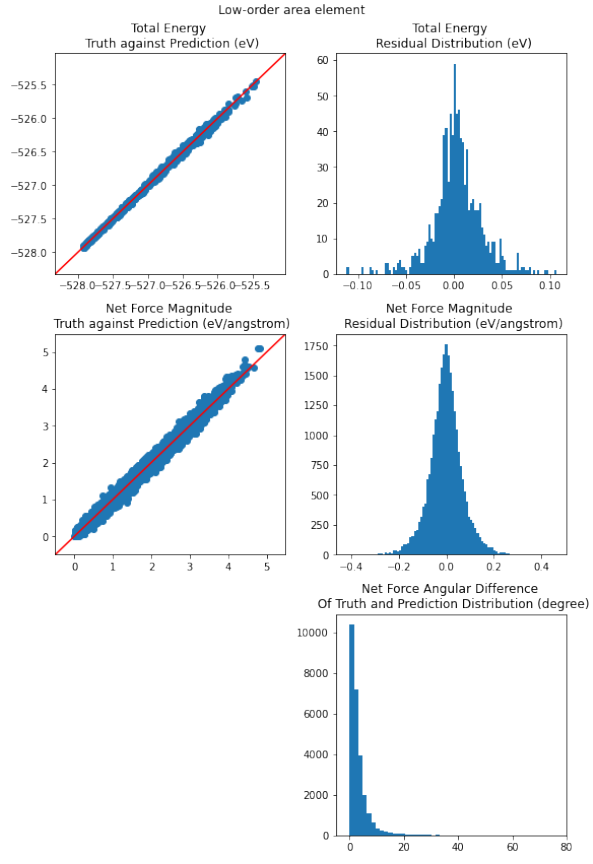

FIG. 4: High order Nitrogen-centered area element fitting result.

B. Residual Analysis

From Figure (3) and (4), the range of error of total energy predictions is around $-0.1 \sim 0.1eV$ for both low-order and high-order area element, which is equivalent as an error range of $-1.67 \sim 1.67$meV for energy per atom. The error range of atomic force magnitude predictions is around $-0.5 \sim 0.5$eV/Å for low-order area element, while the range for high-order area element is $-0.2 \sim 0.2$eV/Å. The range of angular error of atomic force predictions is around $0° \sim 20°$ for low-order area element, while the range for high-order area element is $0° \sim 10°$.

C. Extrapolation

As the perturb strengths of training set are ranged from 1% to 5%, we can investigate the extrapolation performances for unseen data of lower or higher perturb strengths. The following extrapolation experiments are done by predicting 2 unseen samples for each out-of-range perturb strength (1 sample for 0% perturb strength, i.e. the relaxed structure), the final results are averaged over 3 trials of predictions, as to reduce randomness due to NN initialization. Still, the extrapolation results may be unstable due to insufficient sampling of testing data.

For 0% perturb strength, i.e. the relaxed structure. For both low order and high order representation, the predicted sub-energies and atomic forces are about the same for each element, which is consistent as all area elements have roughly the same geometry. The sum of sub-energies is very close to ground truth, within a precision of ±0.05eV. The predicted net forces on center atom are very close to zero, within a precision of ±0.018eV/Å, which is physical as the unperturbed structure is of zero potential energy gradient.

For total energy predictions of varying perturb strength, both low order and high order area element representation have a similar trend transiting from interpolation range and extrapolation range. At interpolation range of 1.2% to 5% perturb strength, the RMSE rises from $0.01 \sim 0.05$eV. At extrapolation range of 5.5% to 6.5% perturb strength, the RMSE is about $0.1 \sim 0.2$eV. As the range further increase from 6.5% to 7.5%, the RMSE jumps to above 0.4eV.

For forces on atom predictions in varying perturb strength, high order area element representation outperforms the low order. At interpolation range of 1.2% to 5% perturb strength, the high order RMSE rises from $0.03 \sim 0.12$eV/Å while $0.06 \sim 0.25$eV/Å for low order. At extrapolation range, the high order RMSE remains stable at ~ 0.12eV/Å for perturb strength 5.5% to 6.5% while the low order RMSE remains stable at ~ 0.25eV/Å for perturb strength 5.5% to 6%. The high order RMSE jumps to ~ 0.25eV/Å for perturb strength above 6.5% while the low order RMSE jumps to ~ 0.4eV/Å for perturb strength above 6%. For force direction, the error remains mostly constant for interpolation but strong fluctuations for extrapolation, at mean level about 7.8° for low order while 3.5° for high order.

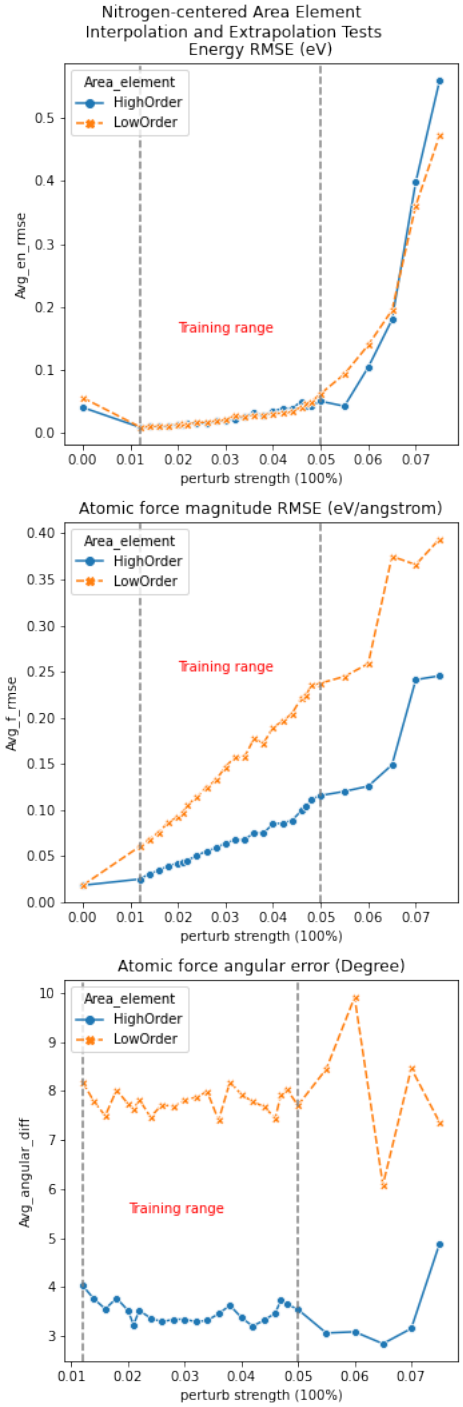

FIG. 5: Interpolation and extrapolation performance of different perturb strength of low order and high order Nitrogen-centered area element.

IV. DISCUSSION

A. Comparison to Symmetry Functions NN Fitting

From Behler and Parrinello's publication on "Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces", they had attempted to fit the Potential Energy Surface (PES) of bulk silicon using symmetry functions, where 8200 samples of training data were taken from Silicon structures at high-pressure phases and from MD simulations of different pressures and temperature. It was reported that the RMSE of energy per atom was $5 \sim 6$meV while atomic forces was 0.2eV/Å in the testing set.

Our fitting results using high order area elements on h-BN have shown significantly improved performance that the RMSE of energy per atom is about 10 times smaller and the RMSE of atomic forces is about 3 times smaller than Behler and Parrinello's results.

Still, the material choice of Behler and Parrinello was 3D bulk Silicon while our material choice is 2D bulk hexagonal Boron Nitride. Therefore, the fitting comparsion may not be on the same ground in terms of material properties and dimensionality.

B. Extrapolation Abilities

In terms of energy fitting, both low order and high order area element have similar performances within and out of the training range, and are able to predict the minimum configuration with acceptable error. The effective extrapolation range is +1% of perturbation strength out of the training range.

In terms of force fitting, high order area element performs about twice better than low order area element within and out of the training range, except that both methods are able to precisely predict the minimum configuration. The effective extrapolation range for force magnitude is +1.5% of perturbation strength out of the training range. For force direction, the mean error is still roughy same as in training range, but the fluctuation goes up as perturbation strength becomes higher. Meanwhile the force magnitude becomes more inaccurate for further extrapolation, the force direction is still reliable.

It is important to note that a smooth sampling of increasing perturbation strength is crucial for good fitting and extrapolation ability, for example a step size of $0.1\% \sim 0.2\%$ in our case. Also, starting from a smaller perturbation strength for example 1% in our case is needed in order to better sample around the minimum of phase space, thus enabling the model to extrapolate the minimum configuration.

C. Order of Area Element Representations

The main feature of area element is to capture the geometry of the local atoms configurations around a specified center atom. We believe that the geometry of local configurations can reflect the underlying geometry of electron density around the center atom, enabling to fit the local energy and atomic forces due to the electron density's geometry and interactions.

Due to the neighbour atoms cutoff for defining the vertices area element, the area element is simply an approximation that does not capture the full environment of around the center atom. Therefore, the "order" of area element depends on the size of cutoff for vertices. The further the cutoff is, the more positional details about local configurations of center atom can be encapsulated. In principle, the perfect area element should includes all neighboring atoms from the specified center atom.

The fact that the lowest order area element is sufficient to fit the total energy indicates that energy is a rather localized quantity which can be easily captured by low order element. In contrast, the atomic force requires higher order area element, indicating that atomic forces are easily affected by long-ranged interactions and correlation effects of electron density.

D. Future Directions

There are more rooms for complicated cases in studying area element representation such as edges and surfaces of material, and 3D structures, i.e. volume elements.

In addition, the current NN method is not permutation and rotational invariant that preprocessing algorithms for data are required, a general fitting method with permutation and rotational invariant can be another interesting direction.

Finally, it is possible that the electron density can be experimented and fitted using the area element method, thus strengthening our argument that geometry of local configurations about the center atom encapsulates the geometry of its electron density.

V. CONCLUSION

Our fitting results on h-BN justified that area element representation is a highly possible candidate for fitting total energy and atomic forces with DFT-like quality. The resulting model is also robust towards over-fitting and is able to extrapolate for a small extent, given the training data is well sampled from the phase space of small, smooth varying perturbations.

ACKNOWLEDGMENTS

I would like to acknowledge the support from my supervisor Prof. Junyi Zhu, who provides references along with useful advice and ideas for the project, he also raises interesting and inspiring questions which leads to further development of the project. My grateful appreciation also extends to research group member Mr. Bin Xi being the major contribution to this project, for data generation of perturbed structures and DFT calculations. He leads discussion, gives detailed analysis and draws insights for the group, he is also preparing for the paper writing. Furthermore, I wish to mention the help by research group member Mr. Chan Ho Ming Tommy in the early stage of data preparation and data processing algorithms design for the project. Finally, I sincerely appreciate Kejie, Reo, Timothy, Calvin and other research group members for their active discussions and questions regarding this topic.

Appendix A: Details of testing set results

Jargons:

- Low Order Area Element (LO)
- High Order Area Element (HO)
- Root Mean Squared Error (RMSE)
- Mean Absolute Percentage Error (MAPE)

TABLE II: Overall testset performance comparison using low order and high order Nitrogen-centered area element

Methods	LO	HO
Num of Vertices	6	12
Num of Inputs	12	24
RMSE of Total Energy (eV)	0.029873	0.029015
RMSE of Energy per Atom (meV)	0.497883	0.483578
MAPE of Total Energy (100%)	0.000040	0.000039
RMSE of x-component Force on Center Atom (eV/Å)	0.160153	0.071714
MAPE of x-component Force on Center Atom (100%)	0.119936	0.053304
RMSE of y-component Force on Center Atom (eV/Å)	0.160673	0.072221
MAPE of y-component Force on Center Atom (100%)	0.120606	0.053917
RMSE of Force Magnitude on Center Atom (eV/Å)	0.159276	0.071915
MAPE of Force Magnitude on Center Atom (100%)	0.119415	0.053303
Mean Absolute Angular Error (Degree)	7.788658	3.491130

TABLE III: Overall testset performance comparison using low order and high order Boron-centered area element

Methods	LO	HO
Num of Vertices	6	12
Num of Inputs	12	24
RMSE of Total Energy (eV)	0.031926	0.027477
RMSE of Energy per Atom (meV)	0.532105	0.457945
MAPE of Total Energy (100%)	0.000043	0.000037
RMSE of x-component Force on Center Atom (eV/Å)	0.175519	0.070212
MAPE of x-component Force on Center Atom (100%)	0.154635	0.067202
RMSE of y-component Force on Center Atom (eV/Å)	0.175200	0.071835
MAPE of y-component Force on Center Atom (100%)	0.156375	0.064003
RMSE of Force Magnitude on Center Atom (eV/Å)	0.172564	0.071591
MAPE of Force Magnitude on Center Atom (100%)	0.097022	0.039900
Mean Absolute Angular Error (Degree)	9.223182	3.655632

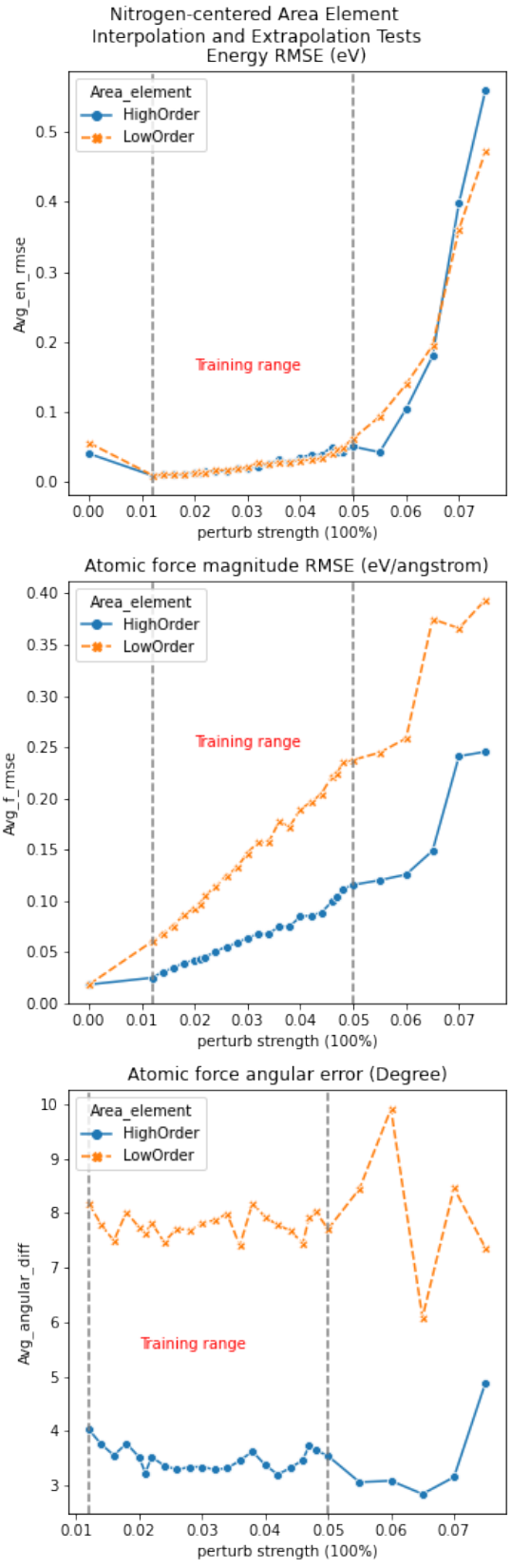

FIG. 6: Interpolation and extrapolation performance of different perturb strength of low order and high order Nitrogen-centered area element.

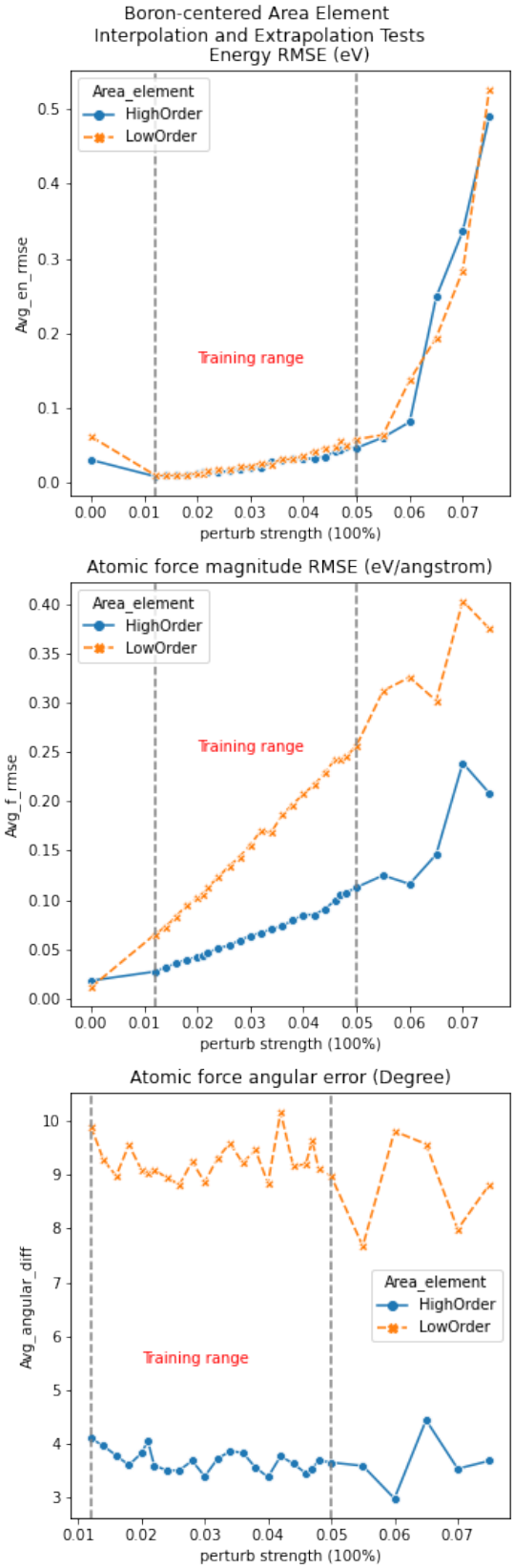

FIG. 7: Interpolation and extrapolation performance of different perturb strength of low order and high order Boron-centered area element.

[1] W. Kohn and L. J. Sham, Self-consistent equations including exchange and correlation effects, Phys. Rev. **140**, A1133 (1965).
[2] J. Behler and M. Parrinello, Generalized neural-network representation of high-dimensional potential-energy surfaces, Phys. Rev. Lett. **98**, 146401 (2007).
[3] L. Hu, B. Huang, and F. Liu, Atomistic mechanism underlying the Si(111)$-(7 \times 7)$ surface reconstruction revealed by artificial neural-network potential, Phys. Rev. Lett. **126**, 176101 (2021).
[4] G. Kresse and J. Furthmüller, Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set, Computational Materials Science **6**, 15 (1996).
[5] G. Kresse and J. Furthmüller, Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set, Phys. Rev. B **54**, 11169 (1996).
[6] V. Wang, N. Xu, J.-C. Liu, G. Tang, and W.-T. Geng, Vaspkit: A user-friendly interface facilitating high-throughput computing and analysis using vasp code, Computer Physics Communications **267**, 108033 (2021).