Table of Contents

Table of Contents 1
How to Train a Moment Tensor Potential in QuantumATK 2
- Prerequisites 2
- Procedure For Bulk HfO2 MTP Training 2
  - Create Workflow 3
  - Results 6
- Summary 9

How to Train a Moment Tensor Potential in QuantumATK

Version: U-2022.12

Downloads & Links

PDF version
`MomentTensorPotentialTraining`
crystalTrainingRandomDisplacements
MTP_tutorial_basic.hdf5

In this basic tutorial, you will learn how to train a moment tensor potential (MTP), a machine-learned forcefield (MLFF) typically used in molecular dynamics (MD) simulations, using the **Workflow Builder** tool in **QuantumATK** . MLFFs are trained to predict *ab initio* potential energy surfaces and vastly improves the accuracy of classical MD simulations while keeping the computational cost at the level of empirical forcefields. At the end of this tutorial, you will be familiar with the basics of training MTPs for bulk materials (e.g. HfO_2) in **QuantumATK**.

Prerequisites

- Basic familiarity with the **QuantumATK** interface, e.g. from the basic introduction tutorial for QuantumATK and the Workflow Builder introduction tutorial.
- Optionally, read the methodology paper for the MTP method [1]. Our independent implementation of MTP in **QuantumATK** is based on this paper.

Procedure For Bulk HfO_2 MTP Training

The procedure we will follow is listed here:

1. Choose reference unit cells describing different phases of HfO_2 .
2. Use built-in QuantumATK functionality to generate repeated and rattled configurations from the unit cells to use as reference configurations. We will generate reference structures of crystal HfO_2 using the crystalTrainingRandomDisplacements protocol.
3. Compute reference data which includes energy, forces and stresses for each of the reference

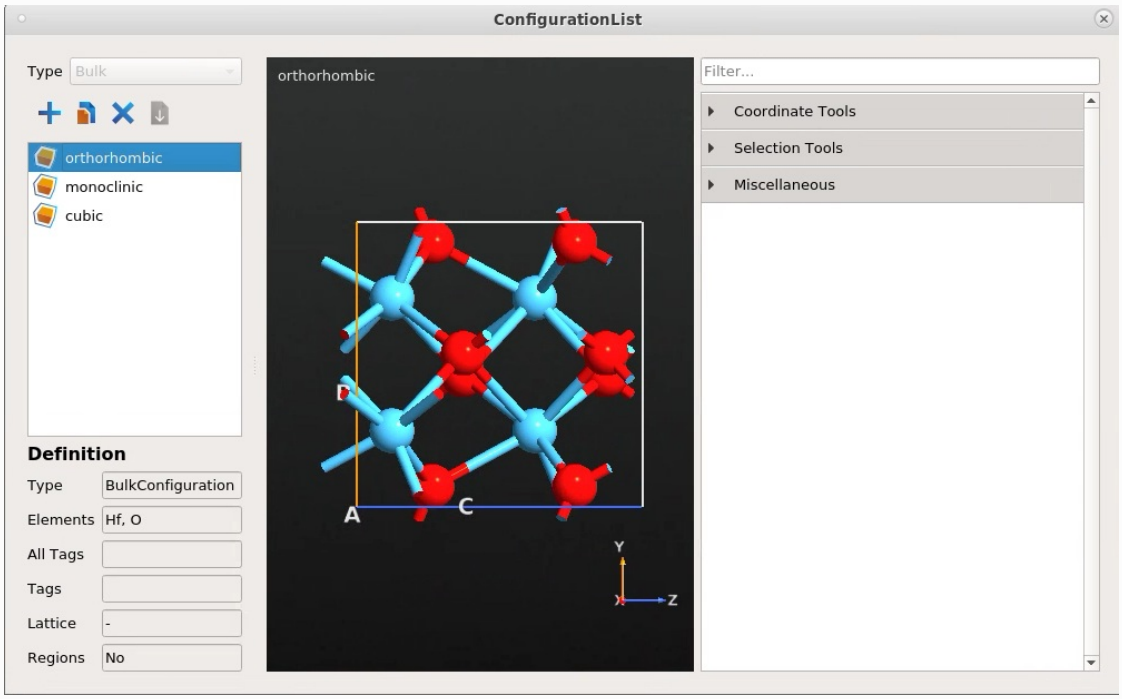
configurations using a chosen DFT reference calculator.
4. Split the reference data set into training set and test set.
5. Train many MTPs by fitting their parameters to the training set data by minimizing the error (energy, forces and stresses) with respect to the reference calculator.
6. Test the trained MTPs by applying them to the test set configurations and compute the error with respect to the reference calculator.
7. Choose the MTP that gives lower and comparable root mean squared error values for both the training and the test set.

The corresponding workflow will be set up using the **Workflow Builder** in the **NanoLab** GUI of **QuantumATK**.

Before creating the workflow, download the bulk geometries of different phases of HfO2 (cubic, monoclinic and orthorhombic) using the Materials Project database plugin in the **NanoLab** and store them on **Builder** stash. Alternatively, download this stash file (Builder_Stash.hdf5) to the new project folder and open the **Builder** tool to visualize the geometries. Notice that the configurations are already named after their phases such as “cubic”, “monoclinic” and “orthorhombic”.

Create Workflow

1. Select any HfO_2 configuration in **Builder** stash. This configuration is needed only to extract the element information for the calculator setup, therefore it can be any configuration containing Hf and O atoms.
2. Click on the icon and send the bulk configuration to the **Workflow Builder**. Right-click on the added configuration in the *Build* panel and rename the configuration to `HfO2`.
3. In the **Workflow Builder**, change the output **Filename** at the bottom of the *Build* panel to `MTP_basics_results.hdf5`.
4. From the **QuantumATK** tab on the right hand panel, expand the *Auxiliary* section and drag-and-drop the *ConfigurationList* block to the workflow and double-click to open its editor window. Press the + and select the three HfO_2 bulk phase geometries from the **Builder** stash. The *ConfigurationList* window should look like the below image before closing the window.

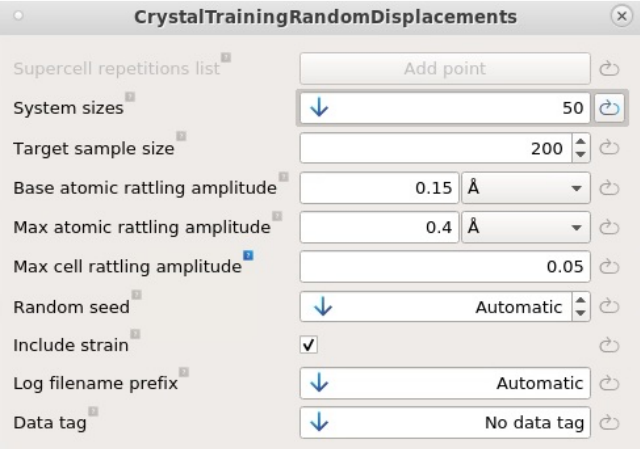


5. From the **QuantumATK** tab on the right hand panel, expand the *Calculators* section and drag-drop the *LCAOCalculator* block to the workflow. The DFT calculator settings must be adjusted for the

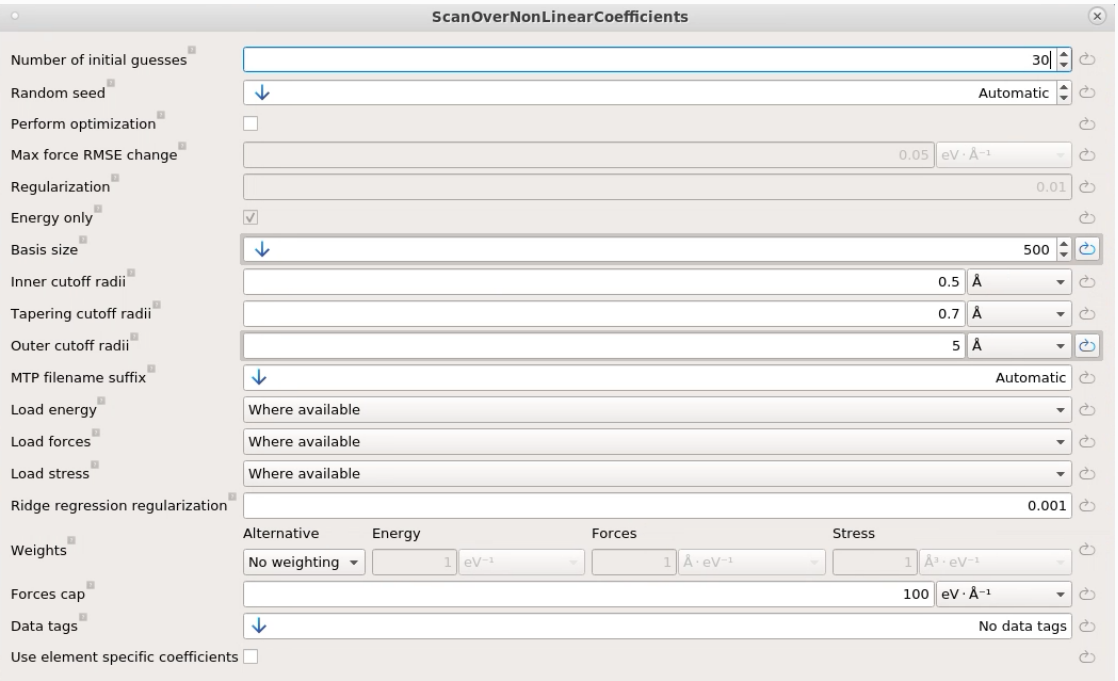
material under study. For this tutorial, we will use the default settings and no modification is necessary.

6. From the **QuantumATK** tab on the right hand panel, expand the *Moment Tensor Potential* section and

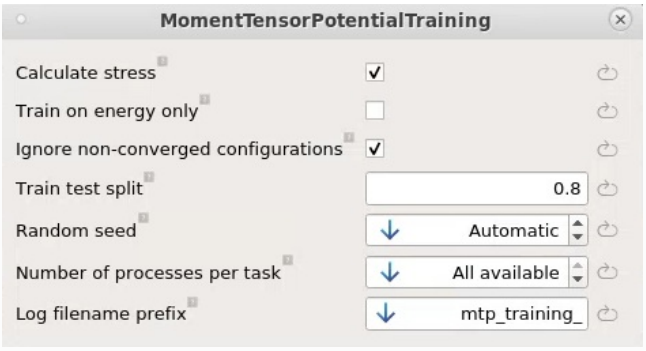
   - Drag-drop the *CrystalTrainingRandomDisplacements* block to the workflow. Double-click the block and set **system sizes** to *50* so that a minimum of 50 atoms will be present in the repeated geometries. All other parameters can have the default values as shown below. Make sure the edits are saved by pressing enter (the button will be lit next to the edited parameter values indicating a saved state) before closing the window.



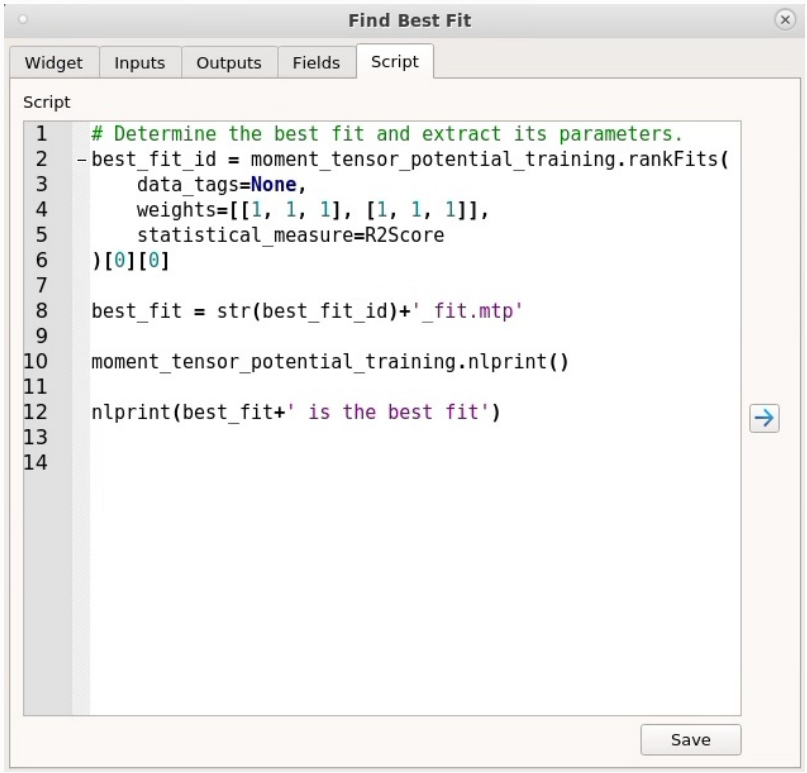
   - Drag-drop the *ScanOverNonLinearCoefficients* block to the workflow. Double-click the block in the workflow and set **Basis size** to *500* and a **outer cutoff radii** of *5 Angstrom*. All other parameters can have the default values as shown below. A total of 30 MTPs will be generated using randomly generated initial coefficients. Do not forget to press enter before closing the window to ensure all edits are saved.



   - Drag-drop the *MomentTensorPotentialTraining* block with default settings.

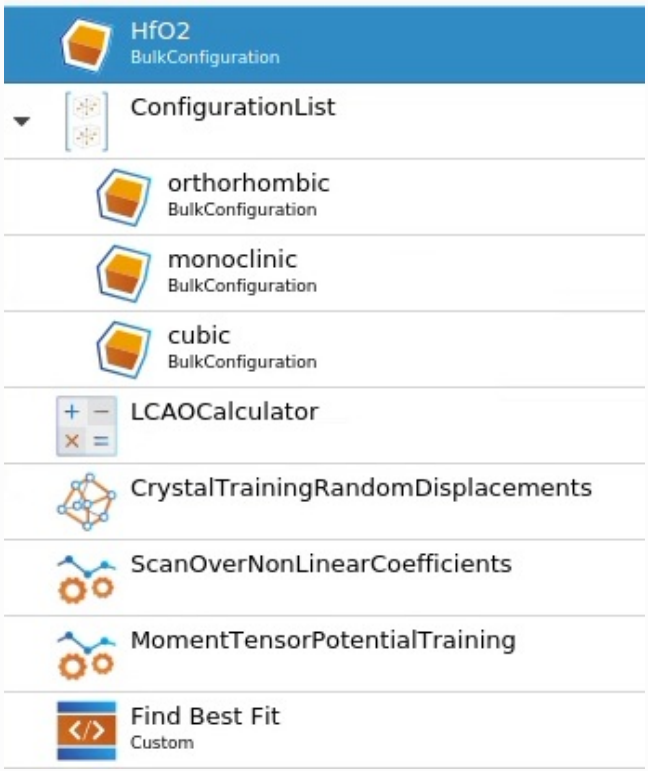


7. From the **QuantumATK** tab on the right hand panel, expand the *Algorithm Blocks* section and drag-drop a *custom* block to the workflow. Rename it to **Find Best Fit**, double-click the block and copy-paste the lines from the attached file custom-best-fit.py into the **script** tab and click the "save" button. The custom block should look like the below image before closing.



After the training has concluded, the fits are ranked by calling the `rankFits()` method on the MomentTensorPotentialTraining instance. By default the r^2 score between reference data and predicted data is used for ranking. This custom code goes through the 30 MTPs and finds the MTP with the lowest and comparable training and test set r^2 score with respect to the reference calculator. Accuracy report of the 30 MTPs are printed to the log file along with the best MTP filename. We will look at the results further below.

Now the workflow is complete and it should look like the below image



You can also download the workflow MTP_tutorial_basic.hdf5 to the workflows folder of the project and open it in the **Workflow Builder**.

Export the workflow as script using the icon and send it to the **Jobs** tool and submit it. Since a DFT calculator is used, this calculation took around 9h 35m on 20 MPI cores.

Results

Select the tutorial folder in the **Data Tool** and you will find the following files:

- `mtp_training_fit_{0,...,29}_fit.log` - contain the accuracy report for the 30 MTPs that were trained.
- `mtp_training_{Hf,O}.log` - contain the atomic reference energy calculation using the reference calculator.
- `{0,...,29}_fit.mtp` - are the 30 fitted MTPs. These are encrypted files containing the coefficients required to construct the basic set for descriptor calculation and the linear fitting parameters for inference.
- `MTP_basics_results.hdf5` - contain the `MomentTensorPotentialTraining` object which includes the reference dataset. It can be opened with the **Movie Tool** to reveal the geometries and their energies, forces and stresses.
- `MTP_basics_results.log` - is the main log file and the best fit is printed at the end of this file.

We use random numbers to initialize the coefficients for the MTP training, rattling the configurations and training-test set splitting. Therefore, reproducibility of the results to the numerical accuracy is only ensured if you use the same random seed while re-running the script. Random numbers are your friend while training MLFFs since the configuration space to sample the geometries and the hyper-parameter space of the MTP coefficients are often much too big to make use of grid based searches. Many different sets of MTP parameters could result in similar results as there could be many degenerate minima in the hyper-parameter surface. So, it is not a cause for concern if you get different sets of MTP fitting parameters with similar accuracy.

Warning

The MTP generated using this workflow can accurately predict the energetics of geometries similar to the reference configurations, but it will fail to describe non-similar geometries. For that, active learning

is needed, please see our tutorial on this topic: Generating A Moment Tensor Potential for HfO2 Using Active Learning

Opening the log file of the best fit, e.g. `mtp_training_fit_1.log` in the **Data Tool**, reveals the following information:

```
+--------------------------------------------------------------------------+
|                                                                          |
| Task FitMomentTensorPotential [Started Tue Jan  3 21:19:02 2023]         |
|                                                                          |
+--------------------------------------------------------------------------+
| fitting error:                                                           |
|     energy:                                                              |
|         mean absolute error: 0.07477469208922793                         |
|         mean squared error: 0.013199517989078                            |
|         root mean square error: 0.1148891552283243                       |
|         maximum absolute error: 0.5262835045068641                       |
|         median absolute error: 0.040108389945089584                      |
|         standard deviation: 0.11488915522832431                          |
|         variance: 0.013199517989078                                      |
|     forces:                                                              |
|         mean absolute error: 0.10306716480208633                         |
|         mean squared error: 0.022906967522765696                         |
|         root mean square error: 0.1513504790965846                       |
|         maximum absolute error: 1.1690371492046867                       |
|         median absolute error: 0.0682895880512062                        |
|         standard deviation: 0.15135047907573454                          |
|         variance: 0.022906967516454362                                   |
|     stress:                                                              |
|         mean absolute error: 0.004365365688971586                        |
|         mean squared error: 5.343125066189659e-05                        |
|         root mean square error: 0.007309668300401639                     |
|         maximum absolute error: 0.03508444345429798                      |
|         median absolute error: 0.0015038598584613903                     |
|         standard deviation: 0.006905206535375291                         |
|         variance: 4.768187729618962e-05                                  |
| regularization: 0.001                                                    |
+--------------------------------------------------------------------------+
|                                                                          |
| Task FitMomentTensorPotential [Finished Tue Jan  3 21:19:23 2023]        |
|                                                                          |
+--------------------------------------------------------------------------+
```

The above log file lists the accuracy of MTP in terms of prediction errors when compared to the reference calculator. They are listed for energy, forces and stress. For each of them, the distribution of the errors including its range (min, max), mean, mean squared value, root mean squared value, standard deviation and variance are listed. Note that for energies, the error is listed in eV per structure. The force errors are given in eV/Angstrom and the stress errors are given in eV/Angstrom^3 .

To get an overview of the accuracy report of all 30 MTPs, take a look at the bottom of the main log file `MTP_basics_results.log` to find the table below. Please note that only a few selected lines from the file are shown here.

```
+-------------------------------------------------------------------------+
| Moment Tensor Potential Training Report                                 |
+-------------------------------------------------------------------------+
+-------------------------------------------------------------------------+
| RMSE (training):                                                        |
|                                                                         |
| Index             Filename     Energy/atom           Force          Stress   |
|                                       (eV)          (eV/Å)        (eV/Å^3)   |
|     0            0_fit.mtp    0.0028585778    0.1791727688    0.0121666313   |
|     1            1_fit.mtp    0.0031880065    0.1789248914    0.0089993819   |
|     2            2_fit.mtp    0.0034210373    0.1727509155    0.0116840178   |
|    19           19_fit.mtp    0.0023482618    0.1513504791    0.0073096683   |
|    28           28_fit.mtp    0.0045982846    0.2725017935    0.0142426258   |
|    29           29_fit.mtp    0.0032051499    0.1893583455    0.0097620363   |
+-------------------------------------------------------------------------+
| RMSE (testing):                                                         |
|                                                                         |
| Index             Filename     Energy/atom           Force          Stress   |
|                                       (eV)          (eV/Å)        (eV/Å^3)   |
|     0            0_fit.mtp    0.0032278151    0.1710122149    0.0120115590   |
|     1            1_fit.mtp    0.0039635338    0.1727359298    0.0090392157   |
|     2            2_fit.mtp    0.0025221608    0.1611050877    0.0112292069   |
|    19           19_fit.mtp    0.0018212650    0.1487881743    0.0066584542   |
|    28           28_fit.mtp    0.0058463322    0.2555467061    0.0142605236   |
|    29           29_fit.mtp    0.0022653063    0.1804519035    0.0089915593   |
+-------------------------------------------------------------------------+
| r^2 (training):                                                         |
|                                                                         |
| Index             Filename     Energy/atom           Force          Stress   |
|     0            0_fit.mtp    0.9997630055    0.9933546725    0.8905795910   |
|     1            1_fit.mtp    0.9997052346    0.9933730468    0.9401336198   |
|     2            2_fit.mtp    0.9996605672    0.9938224952    0.8990881765   |
|    19           19_fit.mtp    0.9998400695    0.9952582355    0.9605040125   |
|    28           28_fit.mtp    0.9993867608    0.9846286820    0.8500530052   |
|    29           29_fit.mtp    0.9997020558    0.9925776526    0.9295568979   |
+-------------------------------------------------------------------------+
| r^2 (testing):                                                          |
|                                                                         |
| Index             Filename     Energy/atom           Force          Stress   |
|     0            0_fit.mtp    0.9996299880    0.9925899380    0.8529427566   |
|     1            1_fit.mtp    0.9994420907    0.9924398060    0.9167183803   |
|     2            2_fit.mtp    0.9997740856    0.9934236323    0.8714755267   |
|    19           19_fit.mtp    0.9998822002    0.9943907550    0.9548108170   |
|    28           28_fit.mtp    0.9987861489    0.9834534235    0.7927194246   |
|    29           29_fit.mtp    0.9998177564    0.9917493054    0.9175942168   |
+-------------------------------------------------------------------------+
```

In this report, you find specifically the energy, forces and stress RMSEs of the training and test sets produced by the 30 MTPs (only a few are shown above). Note that the energy error in this report is listed **per atom** as opposed to the **per structure** values given in the individual log files. Additionally r^2 values, coefficient of determination, are also reported. These values indicate how good is the MTP in predicting the trained properties in the dataset, the larger the better. We can see that most of the trained MTPs have an r^2 value of over 0.999 indicating that they are all very accurate within the configurational space sampled in the training data set.

The above report can also be obtained in NanoLab on demand by opening the MomentTensorPotentialTraining object from `MTP_basics_results.hdf5` in the **Editor** or in **Text Representation**.

The best MTP fit is also mentioned at the end of the log file `MTP_basics_results.log` as

```
19_fit.mtp is the best fit
```

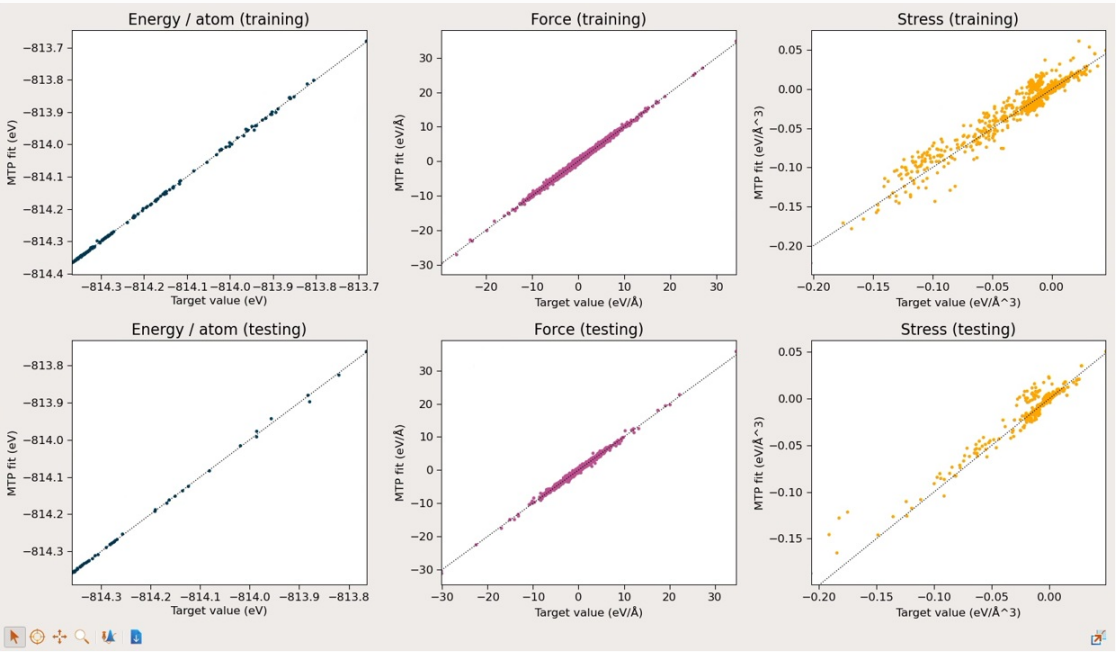
You can also plot the error distribution of any MTP from the list of 30 MTPs. To do that, open a terminal, navigate to the tutorial folder containing `MTP_basics_results.hdf5` and type

```
atkpython
```

to open the atkpython console. Now, enter the following commands

```
moment_tensor_potential_training=nlread('MTP_basics_results.hdf5',MomentTensorPotentialTraining)[0]
moment_tensor_potential_training._nlplotscatter(fit_index=1)
```

The value of **fit_index** can be changed to plot the results of the desired MTP (0 - 29). A plot example is shown below:



The above scatter plot compares the MTP predicted data (y-axis) and reference DFT data (x-axis). The energy, forces and stress values are compared for both training and test sets. The data points for energy and forces (training and test sets) lie along the 45 degree line indicating a high accuracy of prediction. The data points in the stress plot appear to be scattered away from the 45 degree line, however, this is mainly due to the small range of values of the stress tensor components and the error values are very low. Thus, the stress prediction is also very accurate.

Summary

You have now trained your first MLFF, an MTP, to describe crystalline HfO_2 using **NanoLab** GUI tools in **QuantumATK**. We used a random displacements protocol to repeat and rattle unitcells of different HfO_2 phases and used those for training. The MTP generated in this tutorial need to be further improved using the active learning approach (refer to our tutorial Generating A Moment Tensor Potential for HfO2 Using Active Learning) before deploying in the production simulation.

Tip

There are many other protocols to add training data in **QuantumATK** as mentioned in the manual (link: MTP manual page) with all of them accessible via scripting. Find links for some of these protocols below.

- RandomDisplacementsParameters
- MolecularDynamicsSnapshotsParameters
- TrainingSet

- CrystalInterfaceTrainingParameters
- MolecularConfigurationsParameters
- AlloyTrainingParameters

[1]
Alexander V Shapeev. Moment tensor potentials: a class of systematically improvable interatomic potentials. *Multiscale Modeling & Simulation*, 14(3):1153–1173, 2016.

Previous Next