

# Python Rapid Artificial Intelligence Ab Initio Molecular Dynamics

## User Manual



Version 2.3a

Apr 24, 2023

**Jingbai Li**

2022 – present Hoffmann Institute of Advanced Materials  
Shenzhen Polytechnic, China

2020 – 2022 Northeastern University, Boston, U.S.A

Project co-founders:

Steven A. Lopez	Northeastern University, Boston, U.S.A
Pascal Friederich	Karlsruhe Institute of Technology, Germany

Collaborators:

Patrick Reiser	Karlsruhe Institute of Technology, Germany
----------------	--

**Citation:**

1. Jingbai Li, Patrick Reiser, Benjamin R. Boswell, André Eberhard, Noah Z. Burns, Pascal Friederich, and Steven A. Lopez, "Automatic discovery of photoisomerization mechanisms with nanosecond machine learning photodynamics simulations", *Chem. Sci.* **2021**, 12, 5302-5314. DOI:10.1039/D0SC05610C
2. Jingbai Li, Rachel Stein, Daniel Adrion, Steven A. Lopez, "Machine-learning photodynamics simulations uncover the role of substituent effects on the photochemical formation of cubanes", *J. Am. Chem. Soc.* **2021**, 143, 48, 20166–20175. DOI:10.1021/jacs.1c07725
3. Jingbai Li, Steven A. Lopez, "Excited-state distortions promote the reactivities and regioselectivities of photochemical 4 $\pi$ -electrocyclizations of fluorobenzenes", *Chem. A Eur J.* **2022**, 28, e202200651. DOI:10.1002/chem.202200651
4. Jingbai Li, Steven A. Lopez, "A Look Inside the Black Box of Machine Learning Photodynamics Simulations", *Acc. Chem. Res.*, **2022**, 55, 1972–1984. DOI:10.1021/acs.accounts.2c00288

# Contents

1. What is PyRAI <sup>2</sup> MD	4
2. Features	5
2.1. Nonadiabatic molecular dynamics	5
2.2. Machine-learning models	5
2.3. External quantum chemical programs	5
3. Installation	6
4. Getting started with PyRAI <sup>2</sup> MD	7
4.1. Input structure	7
5. Keyword sections	10
5.1. CONTROL	10
5.2. MOLECULE	15
5.3. MOLCAS	20
5.4. BAGEL	22
5.5. ORCA	24
5.6. XTB	25
5.7. MD	26
5.8. NN (MLP, SCHNET, E2N2, DIMENET)	32
5.9. SEARCH	35
5.10. EG and EG2	37
5.11. NAC and NAC2	40
5.13. SOC and SOC2	42
5.14. SCH_EG	45
5.15. SCH_SOC	47
5.16. E2N2_EG	49
5.17. E2N2_NAC	52
5.18. E2N2_SOC	52
5.19. DIME_NAC	54
5.20. FILE	55
5. Nonadiabatic molecular dynamics	56
5.1. Fewest switches surface hopping	56
5.2. Zhu-Nakamura surface hopping	56

6. Machine learning models	56
6.1. Preparing training data	56
6.2. Creating a neural network	56
6.3. Training a neural network	57
6.5. Adaptive sampling	57
7. External quantum chemical program	57
7.1. Molcas	57
7.2. BAGEL	57
7.3. ORCA	57
7.4. GFN-xTB	57
7.5. MNDO	57

# 1. What is PyRAI<sup>2</sup>MD

Python Rapid Artificial Intelligence Ab Initio Molecular Dynamics (PyRAI<sup>2</sup>MD) is a suite of Python scripts for nonadiabatic molecular dynamics simulation using machine-learning (ML) potentials. The primary aim of this project is to leverage the present nonadiabatic molecular dynamics (NAMD) techniques enabling nanosecond-scale simulations for medium-size molecular systems at high-level quantum chemical methods e.g., complete active space self-consistent field (CASSCF) with extended multistate second-order perturbative corrections (XMS-CASPT2).

PyRAI<sup>2</sup>MD is designed as a user-friendly platform that integrate the trajectory surface hopping algorithms, and the state-of-the-art Neural Networks (NNs) models. PyRAI<sup>2</sup>MD aims to simplify the job preparation procedures for newcomers of ML and NAMD.

PyRAI<sup>2</sup>MD integrates a NAMD kernel and an ML kernel via an internal communication in memory. In turn, new features in NAMD simulations and ML models can be developed simultaneously.

## 2. Features

### 2.1. Nonadiabatic molecular dynamics

NVE, NVT, center of mass velocity removal, excessive kinetic energy  
FSSH, ZNSH, NOSH

### 2.2. Machine-learning models

NNs  
Model selection

### 2.3. External quantum chemical programs

Molcas  
Local, slurm, customized basis set

BAGEL  
Local, slurm

ORCA  
Local, slurm

GFN-xTB  
Local, slurm

MNDO  
In the future

### 3. Installation

PyRAI<sup>2</sup>MD is tested on Python 3.7–3.9.

First, download the codes.

```
git clone https://github.com/mlcclab/PyRAI2MD-hiam.git
```

Go to the PyRAI<sup>2</sup>MD folder and install. After installation, it creates a command pyrai2md to run calculations.

```
cd ./PyRAI2MD-hiam  
pip install .
```

Compile fssh library using pyrai2md command.

```
pyrai2md update
```

To run PyRAI<sup>2</sup>MD, simply use the command following by the input file.

```
pyrai2md input
```

PyRAI<sup>2</sup>MD contains some test calculations to verify the code and dependencies. Go to the test folder.

```
cd ./test
```

Edit test\_case.py and choose the test job by setting test\_\$job = 1. Modify the environment variables in the run script file, run\_test.sh. Then run the script.

```
bash run_test.sh
```

## 4. Getting started with PyRAI<sup>2</sup>MD

### 4.1. Input structure

PyRAI<sup>2</sup>MD reads a plain text file and does not require a specific extension. An input file looks like below:

```
&CONTROL
title      test
jobtype    train

&NN
train_data data.json
```

The content is case insensitive, but each keyword (*blue*) must take one to read the input value (*red*) properly. The ‘&’ defines a keyword section (*black*) and the empty line will be automatically skipped. Current available keyword sections include:

<b>CONTROL</b>	This section reads general information to set up calculations. It also controls the parameters used in adaptive sampling for the neural network active learning.
<b>MOLECULE</b>	This section reads molecular specifications including configuration interaction space, spin multiplicities. It also defines the interstate couplings, multiscale regions, periodic conditions, and external constraints.
<b>MOLCAS</b>	This section reads environment variables for setting up Molcas calculations.
<b>BAGEL</b>	This section reads environment variables for setting up BAGEL calculations.
<b>ORCA</b>	This section reads environment variables for setting up ORCA calculations.
<b>XTB</b>	This section reads environment variables for setting up GFN2-xTB calculations.
<b>MD</b>	This section reads (nonadiabatic) molecular dynamics parameters. It controls the cutoff of the trajectories for the neural network active learning.
<b>NN</b>	This section reads the model information of neural networks. It trains PyRAI <sup>2</sup> MD native MLP models.



<b>MLP</b>	This section reads the model information of neural networks. It trains MLP models using pyNNsMD library.
<b>SCHNET</b>	This section reads the model information of neural networks. It trains SchNet models using pyNNsMD library.
<b>E2N2</b>	This section reads the model information of neural networks. It trains E2N2 models using GCNNP library (E2N2 is currently under development and not available yet).
<b>SEARCH</b>	This section reads the parameters used in grid search for optimizing neural network hyperparameters. Currently, it only support PyRAI <sup>2</sup> MD native MLP models.
<b>EG</b>	This section reads the hyperparameters for energy+gradient model. It is required when <b>NN</b> or <b>MLP</b> is set.
<b>NAC</b>	This section reads the hyperparameters for nonadiabatic coupling model. It is required when <b>NN</b> or <b>MLP</b> is set.
<b>SOC</b>	This section reads the hyperparameters for spin-orbit coupling model. It is required when <b>NN</b> or <b>MLP</b> is set.
<b>EG2</b>	This section reads the hyperparameters for the second energy+gradient model. It is required when <b>NN</b> or <b>MLP</b> is set.
<b>NAC2</b>	This section reads the hyperparameters for the second nonadiabatic coupling model. It is required when <b>NN</b> or <b>MLP</b> is set.
<b>SOC2</b>	This section reads the hyperparameters for the second spin-orbit coupling model. It is required when <b>NN</b> or <b>MLP</b> is set.
<b>SCH_EG</b>	This section reads the hyperparameters for energy+gradient model. It is required when <b>SCHNET</b> is set. SchNet models do not have many parameters to tune, thus the second set of hyperparameters are not used.
<b>SCH_NAC</b>	The current SchNet model does not support NAC prediction
<b>SCH_SOC</b>	This section reads the hyperparameters for spin-orbit coupling model. It is required when <b>SCHNET</b> is set. SchNet models do not have many parameters to tune, thus the second set of hyperparameters are not used.
<b>E2N2_EG</b>	This section reads the hyperparameters for energy+gradient model. It is required when <b>E2N2</b> is set. E2N2 models do not have many parameters to tune, thus the second set of hyperparameters are not used.

<b>E2N2_NAC</b>	This section reads the hyperparameters for nonadiabatic coupling model. It is required when <b>E2N2</b> is set. E2N2 models do not have many parameters to tune, thus the second set of hyperparameters are not used.
<b>E2N2_SOC</b>	This section reads the hyperparameters for spin-orbit coupling model. It is required when <b>E2N2</b> is set. E2N2 models do not have many parameters to tune, thus the second set of hyperparameters are not used.
<b>FILE</b>	This section reads molecular information to use PyRAI <sup>2</sup> MD tool for training data extraction.

## 5. Keyword sections

### 5.1. CONTROL

The keywords, default values, and short descriptions are listed below.

\$CONTROL		
title	None	name for the output, user defined
ml_ncpu	1	number of cpu used for ml jobs
qc_ncpu	1	number of cpu used for qc jobs
gl_seed	1	random number seed
jobtype	sp	type of PyRAI <sup>2</sup> MD job
qm	nn	neural networks as the electronic property calculator
----- keywords below are used for adaptive sampling -----		
abinit	molcas	molcas as the ab initio calculator
load	1	load existing model for adaptive sampling
pop_step	200	save average population for the first 200 steps
refine	0	refine data collected near the state-crossing region, the default value skips this procedure
refine_num	4	number of data collected near the state-crossing region for refinement
refine_end	200	the range to search structures near the state-crossing region, the default searches the first 200 steps in the recorded structures in MD
refine_gap	0.3	the energy gap to select the crossing region structures
maxiter	1	maximum number of iterations in the adaptive sampling
maxsample	1	Maximum number of sampled structures per trajectory
dynsample	0	use dynamically weighted thresholds, the default value uses constant thresholds to sample structures
maxdiscard	0	maximum discarded snapshots before adjusting thresholds
maxenergy	0.05	maximum energy threshold to stop trajectories, the unit is Hartree

minenergy	0.02	minimum energy threshold to record snapshots of a trajectory
dynenergy	0.1	weights to increase or decrease the current energy threshold according to the distance between the minimum and maximum energy threshold
inienergy	0.3	initial value of the maximum energy threshold
fwdenergy	1	number of iterations delayed before increasing the current energy threshold
bckenergy	1	number of iterations delayed before decreasing the current energy threshold
maxgrad	0.15	maximum gradient threshold to stop trajectories, the unit is Hartree·Bohr <sup>-1</sup>
mingrad	0.06	minimum gradient threshold to record snapshots of a trajectory
dyngrad	0.1	weights to increase or decrease the current gradient threshold according to the distance between the minimum and maximum gradient threshold
inigrad	0.3	initial value of the maximum gradient threshold
fwdgrad	1	number of iterations delayed before increasing the current gradient threshold
bckgrad	1	number of iterations delayed before decreasing the current gradient threshold
maxnac	0.15	maximum nac threshold to stop trajectories, the unit is Bohr <sup>-1</sup>
minnac	0.06	minimum nac threshold to record snapshots of a trajectory
dynnac	0.1	weights to increase or decrease the current nac threshold according to the distance between the minimum and maximum nac threshold
ininac	0.3	initial value of the maximum nac threshold
fwdnac	1	number of iterations delayed before increasing the current nac threshold
bcknac	1	number of iterations delayed before decreasing the current nac threshold
maxsoc	50	maximum soc threshold to stop trajectories, the unit is cm <sup>-1</sup>
minsoc	20	minimum soc threshold to record snapshots of a trajectory
dynsoc	0.1	weights to increase or decrease the current soc threshold according to the distance between the minimum and maximum soc threshold

inisoc	0.3	initial value of the maximum soc threshold
fwdsoc	1	number of iterations delayed before increasing the soc energy threshold
bcksoc	1	number of iterations delayed before decreasing the soc energy threshold

Full descriptions for all available keywords are summarized below.

**title** sets the name of the calculation, all temporary and logfiles will be named according to this value.

**ml\_ncpu** sets the number of cpu that will be used to run ML-related jobs using python multiprocessing. ML-related **jobtype** are **train**, **adaptive**, **search**.

**qc\_ncpu** sets the number of cpu that will be used to run QC-related jobs using python multiprocessing. QC-related **jobtype** is **adaptive**.

**ms\_ncpu** sets the number of cpu that will be used to run multiscale calculations using python multiprocessing. Currently disabled.

**gl\_seed** sets the global seed for random number generator. It affects the reproducibility of the surface hopping calculations during NAMD and adaptive sampling.

**jobtype** sets the type of PyRAI<sup>2</sup>MD job. Available options are:

- sp** single-point calculations,
- md** NAMD simulation,
- hop** surface hopping calculation,
- adaptive** adaptive sampling,
- train** training NNs,
- prediction** predicting electronic properties using trained NNs,
- search** NN hyperparameter optimization with grid search.

**qm** chooses the electronic property calculator. Available options are:

- nn** uses PyRAI2MD native MLP model,
- mlp** uses pyNNsMD MLP model,
- schNet** uses pyNNsMD SchNet model,
- e2n2** uses GCNNP E2N2 model,
- molcas** uses OpenMolcas for CASSCF calculations,
- mlctkr** uses OpenMolcas/Tinker for QM/MM calculations,
- bagel** uses BAGEL, for CASSCF and XMS-CASPT2 calculations
- orca** uses ORCA for DTF (only ground-state), TD-DFT, or Spin-flip TDDFT calculations
- xtb** uses GFN2-xTB for ground-state calculations

specifying a method followed with **xtb** will enable ONIOM-type QM/QM2 calculation. e.g, **qm molcas xtb**. The QM region is defined in **&MOLECULE** section. Three-layer ONIOM calculation can be invoked by **qm molcas xtb xtb**. Besides the QM and QM2 calculations, It uses GFN-FF for the third layer (MM) calculation without further keywords.

<b>abinit</b>	chooses the reference QC electronic property calculator. Available options are the same as <b>qm</b> except for <b>nn</b> . The chosen program will be used to recompute the QC-data for the collected structures during adaptive sampling.
<b>load</b>	reads a pretrained NNs for adaptive sampling. When it is set to 0, it will first training NNs before running the adaptive sampling.
<b>pop_step</b>	sets the number of MD steps to compute the average population over all trajectories propagated during adaptive sampling. Note that the step size depends on both the timestep and checkpointing frequency, which can be specified by <b>size</b> in <b>&amp;MD</b> section.
<b>refine</b>	turns on additional structural sampling around the state-crossing region during adaptive sampling. Available options are: <b>0</b> skip additional sampling, which is the default; <b>1</b> always do additional sampling, this will search structure even trajectories complete with no other uncertain structures; <b>2</b> do additional sampling until trajectories complete without uncertain structures; <b>3</b> do additional sampling only if trajectories do not complete and do not find uncertain structures.
<b>refine_num</b>	sets the number of structures that will be collected near the state-crossing regions during adaptive sampling. The selection start from the structure with the smallest energy gap.
<b>refine_end</b>	sets the range to search the structures near the state-crossing region in the recorded trajectories. Note that the adaptive sampling only records the last a few MD steps to reduce the memory usage. The number of recorded MD steps can be adjusted by <b>record_step</b> in <b>&amp;MD</b> section.
<b>refine_gap</b>	sets the energy gap threshold to select the structures near the state-crossing region. The default value is 0.3 eV.

<b>maxiter</b>	sets the maximum number of iterations for adaptive sampling. The adaptive sampling will stop when it reach the maximum value or no longer find new structures.
<b>maxsample</b>	set the number of structures to be collected during the adaptive sampling. Note that this number does not include the number of structure refinement from <b>refine_num</b> .
<b>dynsample</b>	turns on the dynamically weighted adaptive sampling. The threshold values will be dynamically adjusted according to the numerical distance between the minimum and the maximum value. It is turn off in default.
<b>maxdiscard</b>	set the maximum number of discard structures in a trajectory. A structure will be discarded if it contains a non-physical bond length shorter than the sum of the van der Waals radius of each atom multiplied by 0.7. When the number of discarded structures exceed <b>maxdiscard</b> , the current threshold will be decreased to limit the exploration region of adaptive sampling. Otherwise, the current threshold will be increased to expand the exploration region of adaptive sampling. Note that the threshold adjustment can be delayed by <b>fwd*</b> and <b>bck*</b> keywords for the forward and backward direction.
<b>maxenergy</b>	sets the maximum value of the energy threshold to stop a trajectory.
<b>minenergy</b>	sets the minimum value of the energy threshold to record a trajectory.
<b>dynenergy</b>	sets the weights of the to increase or decrease the current energy threshold according to the distance between the minimum and maximum energy threshold. The adjustment is weights * (max - min) but the adjusted values will not exceed the minimum or maximum values.
<b>inienergy</b>	set the initial value of the energy threshold to be dynamically adjusted.
<b>fwdenergy</b>	set the number of delayed iterations to increase the current threshold.
<b>bckenergy</b>	set the number of delayed iterations to decrease the current threshold.
<b>maxgrad</b>	sets the maximum value of the gradient threshold to stop a trajectory.
<b>mingrad</b>	sets the minimum value of the gradient threshold to record a trajectory.
<b>dyngrad</b>	sets the weights of the to increase or decrease the current gradient threshold according to the distance between the minimum and maximum gradient threshold. The adjustment is weights * (max - min) but the adjusted values will not exceed the minimum or maximum values.

<b>inigrad</b>	set the initial value of the gradient threshold to be dynamically adjusted.
<b>fwdgrad</b>	set the number of delayed iterations to increase the current threshold.
<b>bckgrad</b>	set the number of delayed iterations to decrease the current threshold.
<b>maxnac</b>	sets the maximum value of the nac threshold to stop a trajectory.
<b>minnac</b>	sets the minimum value of the nac threshold to record a trajectory.
<b>dynnac</b>	sets the weights of the to increase or decrease the current nac threshold according to the distance between the minimum and maximum nac threshold. The adjustment is $\text{weights} * (\text{max} - \text{min})$ but the adjusted values will not exceed the minimum or maximum values.
<b>ininac</b>	set the initial value of the nac threshold to be dynamically adjusted.
<b>fwdnac</b>	set the number of delayed iterations to increase the current threshold.
<b>bcknac</b>	set the number of delayed iterations to decrease the current threshold.
<b>maxsoc</b>	sets the maximum value of the soc threshold to stop a trajectory.
<b>minsoc</b>	sets the minimum value of the soc threshold to record a trajectory.
<b>dynsoc</b>	sets the weights of the to increase or decrease the current soc threshold according to the distance between the minimum and maximum soc threshold. The adjustment is $\text{weights} * (\text{max} - \text{min})$ but the adjusted values will not exceed the minimum or maximum values.
<b>inisoc</b>	set the initial value of the nac threshold to be dynamically adjusted.
<b>fwdsoc</b>	set the number of delayed iterations to increase the current threshold.
<b>bcksoc</b>	set the number of delayed iterations to decrease the current threshold.

## 5.2. MOLECULE

The keywords, default values, and short descriptions are listed below.

<b>&amp;MOLECULE</b>		
<b>ci</b>	<b>1</b>	definition of the configuration interaction space for



		each spin state
spin	0	definition of the spin multiplicity for each spin state
coupling	None	definition of the interstate couplings
highlevel	None	definition of the high level atoms
midlevel	None	definition of the middle level atoms
embedding	False	embed surrounding charge in high level region
read_charge	False	read charge from a .charge file
freeze	None	definition of frozen atoms
constrain	None	definition of constrained atoms
cbond	None	definition of restrained bonds
cangle	None	definition of restrained angles
cdihedral	None	definition of restrained dihedrals
tbond	None	target value of restrained bonds
tangle	None	target value of restrained angles
tdihedral	None	target value of restrained dihedrals
fbond	10.0	prefactor of the biasing potential on bond
fangle	0.005	prefactor of the biasing potential on angle
fdihedral	1e-8	prefactor of the biasing potential on dihedral
shape	ellipsoid	definition of constraining potential
factor	40	exponential factor of the constraining potential
scale	1.0	prefactor of the exponential potential
cavity	None	constraining radius along x, y, and z-axis
center	None	center of the constraining potential
center_type	xyz	set type of center for constraining potential
groups	None	define molecule groups to apply constrain
compress	None	compress shape of potential
track_type	None	track geometric changes in given type of parameter
track_index	None	atom indices to compute geometrical parameters
track_thrhd	None	threshold of geometrical changes to stop MD

track_stop	0	task to stop when the tracking threshold is met
------------	---	---

Full descriptions for all available keywords are summarized below.

- ci** sets configuration interaction space for each spin state, i.e., the number of states in each spin multiplicity, **2** means two states of the first spin, i.e., S0, S1. It can take multiple integers if multiple spin states are involved, e.g. **2 2** means two states in spin 1 and two states in spin 2. the spin multiplicities are defined by **spin**.
- spin** sets the total spin number for each spin state, 0 is singlet, 1 is triplet. It follows the same order as **ci**.
- coupling** reads pairwise indices to define the coupling between two states. Each pair should be separated by ','. The following example,  
**ci**                **2 2**  
**spin**            **0 1**  
**coupling**       **1 2, 2 3, 2 4, 3 4**  
defines that state 1 and 2 are singlet and state 3 and 4 are triplet. It includes the nac between state 1 and 2 (singlet) and state 3 and 4 (triplet) as well as the soc between state 2 and 3 (singlet-triplet) and state 2 and 4 (singlet-triplet). The order of index pairs does not matter and the coupling of the non-defined pairs (e.g, state 1 and 4) will be treated as zero.
- highlevel** reads the atom indices in QM region. The indices can be written individually, or in a range, e.g., 1 2 3 5 6, 1-3 5-6 or 1-2 3 5-6.
- midlevel** reads the atom indices in QM2 region. The indices can be written individually, or in a range, e.g., 1 2 3 5 6, 1-3 5-6 or 1-2 3 5-6. The rest of atoms will be placed in the outer region (MM).
- embedding** embed middle-level surrounding charge in the high-level region if set **true**.
- read\_charge** read middle level surrounding charge from a .charge file if set **true**. This option will keep the same charge throughout the dynamics, suitable for rigid crystal environment. Turn it **false** to dynamically update the embedding charge during dynamics for flexible environment like solvent. Note the ML models in PyRAI<sup>2</sup>MD currently do not have charge embedding function. You must set it **true** to use the same embedding charge throughout the NN training. Otherwise, the energies and forces are not learnable.
- freeze** reads the indices to freeze atoms during dynamics

<b>constrain</b>	reads the indices to apply constraints on atoms during dynamics. All atoms will be included If no indices are provided.
<b>cbond</b>	reads the indices to apply biasing potential on bonds during dynamics. No bond will be restrained If the keyword is omitted. <b>1 2, 3 4</b> represent two bonds, one between atom 1 and 2 and the other between atom 3 and 4.
<b>cangle</b>	reads the indices to apply biasing potential on angles during dynamics. No angle will be restrained If the keyword is omitted. <b>1 2 3, 4 5 6</b> represent two angles, one between atom 1, 2, and 3, centered on atom 2 and the other between atom 4, 5, and 6, centered on atom 5.
<b>cdihedral</b>	reads the indices to apply biasing potential on dihedrals during dynamics. No dihedral will be restrained If the keyword is omitted. <b>1 2 3 4, 5 6 7 8</b> represent two dihedrals, one over atom 1, 2, 3, and 4, where the axis is defined by atom 2 and 3; the other over atom 5, 6, 7 and 8, where the axis is defined by atom 6 and 7.
<b>tbond</b>	define the target value for the restrained bonds. Multiple values are acceptable for multiple restrained bonds. For example, <b>1.5 1.6</b> will restrain the two bonds to 1.5 and 1.6 Angstrom. If only one value is given, it will be applied to all restrained bonds. If this keyword is omitted, the original values of the restrained bonds will be used as target values.
<b>tangle</b>	define the target value for the restrained angles. Multiple values are acceptable for multiple restrained angles. For example, <b>100 90</b> will restrain the two angles to 100 and 90 Degree. If only one value is given, it will be applied to all restrained angles. If this keyword is omitted, the original values of the restrained angles will be used as target values.
<b>tdihedral</b>	define the target value for the restrained dihedrals. Multiple values are acceptable for multiple restrained dihedrals. For example, <b>100 90</b> will restrain the two dihedrals to 100 and 90 Degree. If only one value is given, it will be applied to all restrained dihedrals. If this keyword is omitted, the original values of the restrained dihedrals will be used as target values.
<b>fbond</b>	define the prefactor for the biasing potential on bonds. Default is <b>10.0</b> Hartree.
<b>fangle</b>	define the prefactor for the biasing potential on angles. Default is <b>0.005</b> Hartree.
<b>fdihedral</b>	define the prefactor for the biasing potential on dihedrals. Default is <b>1e-8</b> Hartree.

<b>shape</b>	define the shape of the constraining potential. Available options are ellipsoid and cuboid.
<b>factor</b>	define the exponential factor of the constraining potential. The larger the value is, the shaper the potential wall is. Default is 40.
<b>scale</b>	define the prefactor for the exponential potential. Default is 1.0 Hartree.
<b>cavity</b>	reads constraining radius along x, y, and z-axis. If no value is provided, the constraining potential will be turned off.
<b>center</b>	reads the atom indices to define the center of the constraining potential.
<b>center_type</b>	choose the type of center to apply the constraining potential. mass use the center of mass xyz use geometrical center origin set the center to (0, 0, 0)
<b>groups</b>	define the molecule groups to apply the constraining potential. In default, each atom is considered in the individual group, so the constraining potential will be applied to each of them according to their distance to the center. Alternatively, we can define a group of atoms that their constraining potential will be determined by the distance between the center of mass and potential center. For example, 8 3, 4 5 represent the first group has 8 molecules with 3 atoms, the second group has 4 molecules with 5 atoms. Note the total number of atoms ( $8*3+4*5=44$ ) must equal to the number of constrain atoms.
<b>compress</b>	reads the target ratio and step to compress the shape of the constraining potential. For example, 0.75 1000 will compress the constraining potential from the original shape defined by cavity to 0.75 of them along x, y, and z-axis in 1000 steps. This option is useful to tune the density of the system.
<b>track_type</b>	set the type of geometrical parameter used to early stop the trajectories. Available options are: frag track the distance between two fragments. dist track the distance between two atoms.
<b>track_index</b>	reads the atom indices to define the fragments of interatomic distances.

To define fragments, the format follows as **1 2 3 4, 5 6 7 8**, where the first and second four indices, separated by a ',' punctuation, correspond to the atoms defining the first and second fragment. For tracking interatomic distances, the format follows as **1 2, 3 4, 5 6, 7 8**, where the four pairs of indices are separated by ','. Each of them corresponds to a distance between two atoms.

#### track\_thrhd

reads the threshold to early stop the trajectories if the distances exceed the thresholds. For tracking fragments, only one value is needed. For tracking interatomic distances, more values are supported. If only one value is given, it will be used for all distances. If multiple values are provided, each of them will be used to check the distance defined by [track\\_index](#) accordingly. In this case, the number of values should match the number of tracked distances. The unit is Angstrom.


#### track\_stop

decide what task to be stopped when the tracking geometric parameters meet the threshold.

Available options are:

- 0** has no effects, which is the default.
- 1** stop the trajectory immediately
- 2** stop surface hopping and continue the trajectory at the current state until the tracking parameters become smaller than the threshold.

## 5.3. MOLCAS

The Molcas calculation also needs an input template and guess orbital named with .StrOrb in the current folder. See  for examples of running Molcas calculations.

The keywords, default values, and short descriptions are listed below.

&MOLCAS		
<a href="#">molcas</a>	<b>None</b>	path to Molcas executable
<a href="#">molcas_nproc</a>	<b>1</b>	number of cpu for OpenMP parallelization
<a href="#">molcas_mem</a>	<b>2000</b>	number of memories for calculation
<a href="#">molcas_print</a>	<b>2</b>	logfile printing level
<a href="#">molcas_project</a>	<b>None</b>	project name
<a href="#">molcas_calcdir</a>	<b>\$PWD</b>	path to the temporary calculation folder
<a href="#">molcas_workdir</a>	<b>None</b>	path to Molcas scratch folder

basis	2	additional basis set information
omp_num_threads	1	number of threads for OpenMP parallelization
use_hpc	0	submit calculation to remote cluster
keep_tmp	1	keep the temporary calculation folder

Full descriptions for all available keywords are summarized below.

- molcas** sets the path to Molcas executable.
- molcas\_nproc** sets \$MOLCAS\_NPROC environment variable, the default value is 1.
- molcas\_mem** sets \$MOLCAS\_MEM environment variable, the default value is 2000 MB.
- molcas\_print** sets \$MOLCAS\_PRINT environment variable, the default value is 2.
- molcas\_project** sets \$MOLCAS\_PROJECT environment variable, the default value is taken from **title** in &CONTROL section
- molcas\_calcdir** sets the path to a temporary folder for Molcas calculation. The temporary folder will be named as tmp\_MOLCAS. If no path is provided, the tmp\_MOLCAS will be created in the current folder. Note this is the folder to run Molcas calculations, but not necessary to be the Molcas scratch folder, which is set by **molcas\_workdir**.
- molcas\_workdir** sets \$MOLCAS\_WORKDIR environment variable. If no path is provided, it will be the same path as the tmp\_MOLCAS folder set by **molcas\_calc**. Note that Molcas is input/output intensive, the temporary files could be large and the calculation running in SLURM's /scratch could be slower than in a local disk. It is recommended to use a local folder such as /tmp or /srv/tmp. If you are not sure which folder to use, a shortcut is **AUTO**, which needs to be upper-case.
- basis** reads atom annotation to use different basis sets if it is set to 1. It is turned off in default (2). To use different basis sets, you need to prepare a xyz file following the same atom order and annotate the atom with '\_', e.g. "C\_ X Y Z". The coordinates can be random. Then add the basis set in &GATEWAY in the Molcas input template, e.g. "ANO-S-MB, C\_ ANO-S-VDZP", which will use ANO-S-VDZP for annotated atoms but ANO-S-MB for others.
- omp\_num\_threads** sets OpenMP parallel threads for OpenMolcas, the default value is 1. Note that not all Molcas functions are parallelized.


### use\_hpc

submits the Molcas calculation to the job scheduler. It is turned off in default, thus the calculation is running as a subprocess in the current machine. For single calculation, it is recommended to run the Molcas calculation without **use\_hpc** because it does not have to wait in the queue. However, if there are more Molcas calculations than available cpus or the disk space for all calculations is not enough, e.g. in adaptive sampling, it is better to use **use\_hpc** to distribute the calculations to all available nodes via a job scheduler. To use this function, you need to prepare a submission script template with the same name as **title** in **&CONTROL** section, e.g. **job\_title.slurm** and specify the all necessary **#SBATCH** variables.

### keep\_tmp

keep the temporary Molcas calculation folder. It is turned on in default. Set to **0** to turned off.

## 5.4. BAGEL

The BAGEL calculation also needs an input template and orbital archive in the present folder. See  for examples of running BAGEL calculations.

The keywords, default values, and short descriptions are listed below.

&BAGEL		
<b>bagel</b>	<b>None</b>	path to BAGEL executable
<b>bagel_nproc</b>	<b>1</b>	number of cpu for BAGEL parallelization
<b>bagel_project</b>	<b>Npne</b>	project name
<b>bagel_workdir</b>	<b>\$PWD</b>	path to BAGEL calculation folder
<b>bagel_archive</b>	<b>None</b>	name of BAGEL orbital archive
<b>mpi</b>	<b>None</b>	path to the MPI library
<b>blas</b>	<b>None</b>	path to BLAS library
<b>lapack</b>	<b>None</b>	path to LAPACK library
<b>boost</b>	<b>None</b>	path to BOOST library
<b>mkl</b>	<b>None</b>	path to MKL library
<b>arch</b>	<b>None</b>	cpu architecture
<b>omp_num_threads</b>	<b>None</b>	number of threads for OpenMP parallelization
<b>use_mpi</b>	<b>0</b>	use MPI for parallelization

use_hpc	0	submit calculation to remote cluster
keep_tmp	1	keep the temporary calculation folder

Full descriptions for all available keywords are summarized below.


<b>bagel</b>	sets the path to BAGEL executable.
<b>bagel_nproc</b>	sets the number of cpu for BAGEL calculation with OpenMP parallelization
<b>bagel_project</b>	sets the name of BAGEL calculation, the default value is taken from <b>title</b> in &CONTROL section
<b>bagel_workdir</b>	sets the path to a temporary folder. It creates a sub folder tmp_BAGEL for BAGEL calculation. BAGEL is mainly running in memory. Therefore, it does not suffer from the input/output overhead issue.
<b>bagel_archive</b>	sets the name of BAGEL orbital archive if the orbital archive has a different name from <b>title</b> in &CONTROL section. In default, the name is taken from <b>title</b> in &CONTROL section.
<b>mpi</b>	sets the path to MPI. For the latest (2022) Intel's OneAPI, the environment variables of mkl and mpi can be initialized together by sourcing the setvar.sh in the OneAPI's folder. PyRAI2MD will use <b>mkl</b> to find the source file. and this keyword can be left to empty.
<b>blas</b>	sets the path to BLAS library.
<b>lapack</b>	sets the path to LAPACK library.
<b>boost</b>	sets the path to BOOST library.
<b>mkl</b>	sets the path to Intel MKL library. For the latest (2022) Intel's OneAPI, the environment variables of mkl and mpi can be initialized together by sourcing the setvar.sh in the OneAPI's folder. Thus, this keyword needs to be set to the OneAPI's folder that contains the setvar.sh.
<b>arch</b>	specifies the cpu architecture, the previous default value is <b>intel64</b> . For the latest (2022) Intel's OneAPI, the environment variables of mkl and mpi can be initialized together by sourcing the setvar.sh in the OneAPI's folder. Thus, this keyword needs to be left empty.
<b>omp_num_threads</b>	sets OpenMP parallel threads for BAGEL, the default value is 1.



**use\_hpc** submits the BAGEL calculation to the job scheduler. It is turned off in default, thus the calculation is running as a subprocess in the current machine. For single calculation, it is recommended to run the BAGEL calculation without **use\_hpc** because it does not have to wait in the queue. However, if there are more BAGEL calculations than available cpus or the disk space for all calculations is not enough, e.g. in adaptive sampling, it is better to use **use\_hpc** to distribute the calculations to all available nodes via a job scheduler. To use this function, you need to prepare a submission script template with the same name as **title** in &CONTROL section, e.g. job\_title.slurm and specify the all necessary #SBATCH variables.

**keep\_tmp** keep the temporary BAGEL calculation folder. It is turned on in default. Set to **0** to turned off.

## 5.5. ORCA

The ORCA calculation only needs an input template the present folder. See  for examples of running ORCA calculations.

The keywords, default values, and short descriptions are listed below.

&BAGEL		
<b>orca</b>	<b>None</b>	path to ORCA executable
<b>orca_project</b>	<b>None</b>	project name
<b>orca_workdir</b>	<b>\$PWD</b>	path to ORCA calculation folder
<b>dft_type</b>	<b>tddft</b>	type of DFT calculation
<b>mpi</b>	<b>\$PWD</b>	path to the OpenMPI library
<b>use_hpc</b>	<b>0</b>	submit calculation to remote cluster
<b>keep_tmp</b>	<b>1</b>	keep the temporary calculation folder

Full descriptions for all available keywords are summarized below.


**orca** sets the path to ORCA executable. It only supports ORCA 5.0

**orca\_project** sets the name of ORCA calculation, the default value is taken from **title** in &CONTROL section

**orca\_workdir** sets the path to a temporary folder. It creates a sub folder tmp\_ORCA for ORCA calculation.

<b>dft_type</b>	sets the type of DFT calculation.
<b>dft</b>	ground-state DFT calculation.
<b>tddft</b>	TDDFT calculation.
<b>sf_tddft</b>	Spin-flip TDDFT calculation. It only supports 1-particle-1-hole operator, it could be hard to converge more than 3 singlet states. Must be used with cautions.
<b>mpi</b>	sets the path to OpenMPI
<b>use_hpc</b>	submits the ORCA calculation to the job scheduler. It is turned off in default, thus the calculation is running as a subprocess in the current machine. For single calculation, it is recommended to run the ORCA calculation without <b>use_hpc</b> because it does not have to wait in the queue. However, if there are more ORCA calculations than available cpus or the disk space for all calculations is not enough, e.g. in adaptive sampling, it is better to use <b>use_hpc</b> to distribute the calculations to all available nodes via a job scheduler. To use this function, you need to prepare a submission script template with the same name as <b>title</b> in &CONTROL section, e.g. job_title.slurm and specify the all necessary #SBATCH variables.
<b>keep_tmp</b>	keep the temporary ORCA calculation folder. It is turned on in default. Set to <b>0</b> to turned off.

## 5.6. XTB

The GFN2-xTB calculation does not needs any input template in the present folder. See  for examples of running GFN2-xTB calculations.

The keywords, default values, and short descriptions are listed below.

&XTB		
<b>xtb</b>	<b>None</b>	path to xTB executable
<b>xtb_project</b>	<b>None</b>	project name
<b>xtb_workdir</b>	<b>\$PWD</b>	path to xTB calculation folder
<b>xtb_nproc</b>	<b>1</b>	Number of OMP threads
<b>gnfver</b>	<b>-2</b>	version of GFN-xTB
<b>mem</b>	<b>1000</b>	Memory for OMP stack size
<b>use_hpc</b>	<b>0</b>	submit calculation to remote cluster

keep_tmp	1	keep the temporary calculation folder
----------	---	---------------------------------------

Full descriptions for all available keywords are summarized below.

<b>xtb</b>	sets the path to GFN-xTB executable.
<b>xtb_project</b>	sets the name of GFN-xTB calculation, the default value is taken from <b>title</b> in &CONTROL section
<b>xtb_workdir</b>	sets the path to a temporary folder. It creates a sub folder tmp_XTB for GFN-xTB calculation.
<b>xtb_nproc</b>	sets the number of threads for parallel GFN-xTB calculation
<b>gnfver</b>	sets the version of GFN-xTB calculation. Available options are: -2 default GFN version of the installed GFN-xTB -1 use GFN_FF 0 use GFN0 1 use GFN1 2 use GFN2
<b>xmem</b>	sets the memory for OMP_STACKSIZE in MB.
<b>use_hpc</b>	submits the GFN2-xTB calculation to the job scheduler. It is turned off in default, thus the calculation is running as a subprocess in the current machine. For single calculation, it is recommended to run the GFN2-xTB calculation without <b>use_hpc</b> because it does not have to wait in the queue. However, if there are more ORCA calculations than available cpus or the disk space for all calculations is not enough, e.g. in adaptive sampling, it is better to use <b>use_hpc</b> to distribute the calculations to all available nodes via a job scheduler. To use this function, you need to prepare a submission script template with the same name as <b>title</b> in &CONTROL section, e.g. job_title.slurm and specify the all necessary #SBATCH variables.
<b>keep_tmp</b>	keep the temporary ORCA calculation folder. It is turned on in default. Set to 0 to turned off.

## 5.7. MD

The keywords, default values, and short descriptions are listed below.

&MD		
-----	--	--

initcond	0	sample initial condition
excess	0	excess kinetic energy in Hartree
scale	1	scale kinetic energy by a factor
target	0	set a target kinetic energy in Hartree
graddesc	0	gradient descent mode (zero velocity)
reset	0	remove center of mass velocity
resetstep	0	center of mass velocity reset interval
ninitcond	20	number of sampled initial conditions
method	wigner	initial condition sampling method
format	molden	frequency file format
randvelo	0	Initialize random velocity
temp	300	temperature in Kelvin
step	10	number of threads for OpenMP parallelization
size	20.67	step size in the atomic unit of time
root	1	initial state
activestate	0	only compute gradients of the current state
sfhp	nosh	surface hopping algorithm
nactype	ktdc	type of nac
phasecheck	0	apply phase correction to nac
gap	0.5	energy gap threshold to compute Zhu-Nakamura surface hopping between the same spin states
gapsoc	0.5	energy gap threshold to compute Zhu-Nakamura surface hopping between the different spin states
substep	20	number of substep in wave function integration in FSSH calculation
integrate	0	accumulate the nuclear amplitude transfer in FSSH calculation <b>*This is only for debug purpose*</b>
deco	0.1	energy-based decoherence correction in Hartree
adjust	1	adjust velocity at surface hopping
reflect	1	reflect velocity at frustrated hopping
maxh	10	Maximum number of allowed surface hoppings

dosoc	0	compute Zhu-Nakamura surface hopping between the different spin states
thermo	off	apply a thermostat for NVT ensemble
thermodelay	200	delay time for applying a thermostat in the ground-state
silent	1	no output prints on screen
verbose	0	logfile printing level
direct	2000	number of MD steps that will be written in output
buffer	500	number of MD steps that will be skipped in output
record	0	part of the data that will be recorded for adaptive sampling
record_step	0	number of the last MD snapshots that will be recorded for adaptive sampling
checkpoint	0	checkpoint a trajectory for a given number of MD steps
restart	0	restart calculation
addstep	0	add MD steps in a restart calculation

Full descriptions for all available keywords are summarized below.

**initcond** generates initial conditions from a frequency file. It is turned off in default. Thus, it reads coordinates and velocities from .xyz and .velo files. In adaptive sampling, the initial conditions are always generated from a frequency file, no matter it is set to 1 or 0.

**excess** adds extra kinetic energy beyond the initial kinetic energy then scales the initial velocity isotopically. It is sometimes useful to accelerate the MD and drive the trajectory uphill. The unit is Hartree. This option is the first adjustment to the kinetic energy.

**scale** scales the initial kinetic energy isotropically by a factor. It is sometimes useful to accelerate the MD and drive the trajectory uphill. This option is the second adjustment to kinetic energy.

**target** sets a target kinetic energy to scale the initial velocity isotopically. It is sometimes useful to accelerate the MD and drive the trajectory uphill. This option is the last adjustment to the kinetic energy.

**graddesc** propagates a trajectory following the gradient descent by setting the velocities to zero during the MD. It is turned off in default.

<b>reset</b>	removes translation and rotation velocity at the center of mass. It is turned off in default. It helps avoid the “flying ice” artifact, which results from the draining of vibration energy to translation and rotation energy when velocity rescaling (e.g., thermostat) is frequently used.
<b>resetstep</b>	sets the interval of removing translation and rotation velocity at the center of mass. It is usually recommended to reset velocity every 2000 steps with a timestep of 0.5 fs. If it is set to 0, it only reset the initial velocity. This keyword must be used together with <b>reset</b> .
<b>ninitcond</b>	sets the number of initial conditions in sampling. The last condition is used in MD if the value is greater than 1. In adaptive sampling, this value determines the number of trajectories to collect new structures.
<b>method</b>	chooses the method to do initial condition sampling. It is recommended to do Wigner sampling using <b>wigner</b> . The Boltzmann sampling is also available with <b>boltzmann</b> .
<b>format</b>	sets the frequency file format. It supports the Molcas’ molden file (\$xxx.freq.molden), BAGEL frequency calculation output file (need to rename as \$xxx.freq.bagel), ORCA frequency calculation output file (need to rename as \$xxx.freq.orca), Gaussian frequency calculation output file and fchk file with “Freq=SaveNormalModes” (need to rename as \$xxx.freq.log and \$xxx.freq.fchk).
<b>temp</b>	sets the temperature in Kelvin for initial condition sampling and thermostat. It is not used in microcanonical ensemble (i.e., NVE).
<b>randvelo</b>	initialize random atomic velocity according to the input temperature.
<b>step</b>	sets the number of MD steps.
<b>size</b>	sets the step size in the atomic unit of time. 1 au = 0.02418884254 fs.
<b>root</b>	sets the initial state in NAMD. It should not be larger than the total number of states defined by <b>ci</b> in &MOLECULE.
<b>activestate</b>	only computes the gradients of current state with QC calculations. It is turned off in default. It reduces the cost of FSSH dynamics because the gradients of other states are not used. However, the gradients of all states are needed in Zhu-Nakamura surface hopping. This keyword is not used in ML-NAMD as NNs predict gradients of all states.
<b>sfhp</b>	chooses the surface hopping algorithm. Available options are:

	<b>fssh</b> Tully's the fewest switches surface hopping with explicit nac, <b>gsh</b> Zhu-Nakamura surface hopping, <b>nosh</b> turn off the surface hopping calculation.
<b>nactype</b>	chooses the type of nac for <b>fssh</b> calculation. Available options are: <b>nac</b> nonadiabatic coupling vectors, non-weighted by the state energy gap <b>ktdc</b> curvature driven time-dependent coupling, which approximates nonadiabatic coupling by the first-order derivative of energy in two adjacent MD step.
<b>phasecheck</b>	apply phase correction to nonadiabatic coupling by the overlap of nac vectors at two adjacent MD step. It is turned off in default. It is only used when <b>sfhp</b> is set to <b>fssh</b> and <b>nactype</b> is set to <b>nac</b> .
<b>gap</b>	sets the energy gap threshold (in eV) to compute Zhu-Nakamura surface hopping between two states with same spin multiplicity. The Zhu-Nakamura surface hopping calculations are skipped when the energy gap is larger than this value. When <b>sfhp</b> is set to <b>fssh</b> and <b>nactype</b> <b>ktdc</b> , the NACs are considered as 0 if the energy gap is larger than this value.
<b>gapsoc</b>	sets the energy gap threshold (in eV) to compute Zhu-Nakamura surface hopping between two states with different spin multiplicities. The surface hopping calculations are skipped when the energy gap is larger than this value. This keyword is not used when <b>sfhp</b> is set to <b>fssh</b> .
<b>substep</b>	sets the number of substeps to integrate the electronic wave function in <b>fssh</b> calculation. It is not used when <b>sfhp</b> is set to <b>gsh</b> .
<b>integrate</b>	accumulate the nuclear amplitude in <b>fssh</b> calculation. <b><i>This is only used for debug purpose and must not be used to produce results for publication.</i></b>
<b>deco</b>	applies the energy-based decoherence correction in <b>fssh</b> calculation. The unit is in Hartree. It is not used when <b>sfhp</b> is set to <b>gsh</b> .
<b>adjust</b>	scales the velocity at surface hopping events. Available options are: <b>0</b> do not scale velocity, <b>1</b> scale velocity isotropically, <b>2</b> scale velocity along the NAC direction.
<b>reflect</b>	changes the velocity direction when frustrated hopping happens. Available options are: <b>1</b> directly reflect velocity

	<b>2</b> reflect the velocity component along the NAC vectors.
<b>maxh</b>	sets the maximum number of allowed surface hopping events.
<b>dosoc</b>	computes Zhu-Nakamura surface hopping between two states with different spin multiplicities. It requires additional calculations of spin-orbit coupling and is turned off in default.
<b>thermo</b>	controls the ensemble of trajectory. Available options are: <b>off</b> do not rescale velocity (NVE) <b>0</b> rescale velocity to conserve total energy (forced to NVE ensemble) <b>1</b> rescale velocity using N��se-Hoover thermostat (NVT ensemble) <b>2</b> rescale velocity to conserve total energy in the excited state then applying N��se-Hoover thermostat in the ground-state.
<b>thermodelay</b>	sets the number of MD step delayed for applying a thermostat in the ground-state. It is only used when set <b>thermo</b> is set to <b>2</b> .
<b>silent</b>	turns off printing output on screen. It is turned on in default.
<b>verbose</b>	controls the printing level. <b>0</b> only prints energy and state populations, <b>1</b> prints coordinates, velocities, gradients, and NACs, <b>2</b> prints more calculations information (screen output only).
<b>direct</b>	sets the number of MD steps to be written in the output file. It starts from the first step.
<b>buffer</b>	sets the number of MD steps to be skipped in output file after <b>direct</b> writing steps.
<b>record</b>	choose the part of data to be recorded in a trajectory during adaptive sampling. Available options are: <b>whole</b> all data will be recorded; <b>qm</b> only data in qm region will be recorded.
<b>record_step</b>	sets the number of the latest MD steps in a trajectory to be cached in memory. The cached trajectories are used to sample uncertain data in adaptive sampling. Reduce this number if the molecular dynamics have a huge number of steps or the adaptive sampling does not have enough memory to proceed.
<b>checkpoint</b>	sets the number of MD steps to checkpoint a trajectory. The trajectory is



stored in python pickle file (.pkl) and can be used to restart the calculation. It is turned off in default.

#### restart

reads the .pkl file to restart a calculation. It is turned off in default.

#### addstep

adds additional MD steps in the restarted calculation. Use this if you want to continue to propagate a completed trajectory.

## 5.8. NN (MLP, SCHNET, E2N2, DIMENET)

The neural networks in PyRAI<sup>2</sup>MD are implemented with TensorFlow/Keras API and pyTorch. The neural network is built upon fully connected feedforward multilayer perceptron and graph convolutional neural networks. They consist of an input layer, several hidden layers, and an output layer. Each layer is connected by multiple neurons with activation functions. The connection between layers is a linear function including weights and bias.

PyRAI<sup>2</sup>MD offers a convenient interface to train a neural network and load a trained model for the prediction of energies, forces, non-adiabatic couplings, and spin-orbit couplings. PyRAI<sup>2</sup>MD always trains two sets of neural networks, which can have completely different architectures or only different initial weights. This is useful to measure the prediction uncertainty when predicting data out of the training set. The energies and forces are combined in one model and the non-adiabatic couplings and spin-orbit couplings use an independent model. Users can choose to train either one or all of them.


The keywords, default values, and short descriptions are listed below. All types of neural networks share the same keywords in their sections. Here we use &NN section as an example.

&NN (MLP, SCHNET, E2N2, and DIMENET)		
modeldir	\$PWD	path to save or load NN
train_data	None	path to load training data
pred_data	None	path to load prediction data
nsplits	10	number of folds to split training data
shuffle	False	shuffle training data every epoch
nn_eg_type	1	number of energy+gradient model
nn_nac_type	0	number of nac model
nn_soc_type	0	number of soc model
eg_unit	si	unit of energy+gradient model
nac_unit	si	unit of nac model

soc_unit	si	unit of soc model
select_eg_out	None	select the output of energy+gradient model
select_nac_out	None	select the output of nac model
select_soc_out	None	select the output of soc model
permute_map	No	path to permutation map for data augmentation
silent	1	no output prints on screen
gpu	0	Use GPU for training, only support E2N2

Full descriptions for all available keywords are summarized below.

**modeldir** sets a path to save or load a NN model. The default location is the present folder. The model is saved in a folder named as “NN- $\$xxx$ ”.

**train\_data** sets a path to load the training data from a JSON file. See  for the information of data format. If a file name is provided, it assumes that the file is in the current folder.

**pred\_data** sets a path to load the prediction data from a JSON file. If a file name is provided, it assumes that the file is in the current folder. It is only used when **jobtype** is set to **prediction**.

**nsplits** sets the number of folds to split the training data. The first fold will be used for validation of the first model, and the second fold will be used for validation of the second model. The rest of the data will be used for training model accordingly.

**shuffle** shuffle the training data every epoch. It helps accelerate the training.

**nn\_eg\_type** defines the number of energy+force models with different architectures. Available options are:

- 1** build two neural networks with the same architecture but being initialized with different weights. The hyperparameters are read from **&EG**.
- 2** build two neural networks with different architecture being initialized with different weights. The hyperparameters are read from **&EG** and **&EG2**, respectively.

**nn\_nac\_type** defines the number of nac models with different architectures. Available options are:

- 0 skip the nac model.
- 1 build two neural networks with the same architecture but being initialized with different weights. The hyperparameters are read from **&NAC**.
- 2 build two neural networks with different architecture being initialized with different weights. The hyperparameters are read from **&NAC** and **&NAC2**, respectively.

#### nn\_soc\_type

defines the number of nac models with different architectures. Available options are:

- 0 skip the soc model.
- 1 build two neural networks with the same architecture but being initialized with different weights. The hyperparameters are read from **&SOC**.
- 2 build two neural networks with different architecture being initialized with different weights. The hyperparameters are read from **&SOC** and **&SOC2**, respectively.

#### eg\_unit

set the unit of energy and gradients used in training. Available options are:

- au energy in Hartree and gradient in Hartree·Bohr<sup>-1</sup>,
- si energy in eV and gradients in eV·Å<sup>-1</sup>.

#### nac\_unit

set the unit of nac used in training. Available options are:

- au nac in Hartree·Bohr<sup>-1</sup>,
- si nac in eV·Å<sup>-1</sup>.

#### soc\_unit

set the unit of nac used in training. Available options are:

- si soc in cm<sup>-1</sup>.

#### select\_eg\_unit

select the output of the energy+gradient model for ML-NAMD simulations. The default option is to use all predicted energy and gradients. The number of states is determined by the training data. If you wish only include the energy and gradients of a few states instead of all states, you can specify them as the following.

**1 2 3 4** will use the data of the first 4 states if more than 4 states are trained. If different spin states are included, for instant, 3 singlet states followed by 3 triplet states, you can use **1 2 4 5** to select the first two singlet states and two triplet states.

#### select\_nac\_unit

select the output of the nac model for ML-NAMD simulations. The default option is to use all predicted nac. You can set the indices of the nac in the training data to use specific nac values. For instance, if the nac training data are organized for state pair 1 2, 2 3, 3 4, and 4 5, setting **select\_nac\_unit**

to **1 2 3** will select the nac of the state pair 1 2, 2 3, and 3 4 (skip the state pair 4 5) for ML-NAMD simulation.

**select\_soc\_unit** select the output of the soc model for ML-NAMD simulations. The default option is to use all predicted soc. You can set the indices of the soc in the training data to use specific soc values. For instance, if the soc training data are organized for state pair 1 3, 1 4, 2 3, and 2 4, setting **select\_soc\_unit** to **1 3** will select the soc of the state pair 1 3 and 2 3 (skip the state pair 1 4 and 2 4) for ML-NAMD simulation.

**permute\_map** read a text file that defined the permutations of atom indexing. Each line should only include one set of permutation. "1 5 3 2 4 6" means first switch the index of atom 2 and atom 5 then switch the index of atom 4 and the atom 2.

**silent** turns off printing output on screen. It is turned on in default.

**gpu** use GPU to train E2N2 models. This option does not work for other NN models for the moment.

## 5.9. SEARCH

The keywords, default values, and short descriptions are listed below.

&SEARCH		
----- keywords below are available for nn -----		
depth	<b>1</b>	a list to search number of hidden layers
nn_size	<b>20</b>	a list to search number of neurons per hidden layer
batch_size	<b>32</b>	a list to search batch size
reg_l1	<b>1e-8</b>	a list to search l1 factor
reg_l2	<b>1e-8</b>	a list to search l2 factor
dropout	<b>0.005</b>	a list to search dropout ratio
----- keywords below are available for e2n2 -----		
n_features	<b>16</b>	a list to search number of features
n_blocks	<b>3</b>	a list to search number of interaction blocks
l_max	<b>1</b>	a list to search number of rotation order
n_rbf	<b>8</b>	a list to search number of radial basis

rbf_layer	2	a list to search number of RBF layers
rbf_neurons	32	a list to search number of RBF neurons
use_hpc	1	unit of energy+gradient model
retrieve	0	read results from training logfiles

Full descriptions for all available keywords are summarized below.

<b>depth</b>	searches a list of parameters for hidden layers, e.g., 2 3 4 5. This keyword only works for nn.
<b>nn_size</b>	searches a list of parameters for number of neurons per hidden layer, e.g., 100 200 300. This keyword only works for nn.
<b>batch_size</b>	searches a list of parameters for batch size, e.g., 64 128. This keyword only works for nn.
<b>reg_l1</b>	searches a list of parameters for l1 factor, e.g., 1e-5 1e-6 1e-7. It is used when use_reg_activ, use_reg_weight, or use_reg_bias is set to l1 or l1_l2 in &EG, &EG2, &NAC, &NAC2, &SOC, and &SOC2 sections. This keyword only works for nn.
<b>reg_l2</b>	searches a list of parameters for l1 factor, e.g., 1e-5 1e-6 1e-7. It is used when use_reg_activ, use_reg_weight, or use_reg_bias is set to l2 or l1_l2 in &EG, &EG2, &NAC, &NAC2, &SOC, and &SOC2 sections. This keyword only works for nn.
<b>dropout</b>	searches a list of parameters for dropout ratio, e.g., 0.001 0.002 0.003. This keyword only works for nn.
<b>n_features</b>	searches a list of parameters for number of features, e.g., 8 16. This keyword only works for e2n2.
<b>n_blocks</b>	searches a list of parameters for number of blocks, e.g., 3 4. This keyword only works for e2n2.
<b>l_max</b>	searches a list of parameters for number of rotation order, e.g., 1 2. In general 1 is good. This keyword only works for e2n2.
<b>n_rbf</b>	searches a list of parameters for number of radia basis, e.g., 16 32. This keyword only works for e2n2.

<b>rbf_layers</b>	searches a list of parameters for number of RBF layers, e.g, <b>2 3</b> . This keyword only works for e2n2.
<b>rbf_neurons</b>	searches a list of parameters for number of RBF neurons, e.g, <b>32 64</b> . This keyword only works for e2n2.
<b>use_hpc</b>	submits the NN training to the job scheduler. It is turned on in default, thus the training will be submitted to SLURM as a subprocess in the current machine. For training a few NNs on a node with many cpu, it is not recommended to use <b>use_hpc</b> because the job will have to wait in the queue while the current machine is idle. However, if there are hundreds of training in a grid search, it is better to use <b>use_hpc</b> to distribute the calculations to all available nodes via SLURM. To use this function, you need to prepare a SLURM template with the same name as <b>title</b> in &CONTROL section, e.g. job_title.slurm and specify the all necessary #SBATCH variables. If <b>gpu</b> is used (for using e2n2 mode), the SLURM tempalte need to be named as job_title.gres. Note the difference in the file extension.
<b>retrieve</b>	reads the logfiles of NN trainings in a completed grid-search and regenerate a logfile containing a summary of training results. No training calculation is performed. It is used when the grid search completed normally but the failed to print results. It is turned off in default.

## 5.10. EG and EG2

The keywords, default values, and short descriptions are listed below.

&EG and &EG2		
<b>invd_index</b>	<b>None</b>	path to inverse distance indices file
<b>depth</b>	<b>4</b>	number of hidden layers
<b>nn_size</b>	<b>100</b>	number of neurons per hidden layer
<b>batch_size</b>	<b>64</b>	number of data in one batch
<b>activ</b>	<b>leaky_softplus</b>	activation function
<b>activ_alpha</b>	<b>0.03</b>	activation function coefficient alpha
<b>loss_weights</b>	<b>1 1</b>	weights of energy and gradient loss
<b>use_dropout</b>	<b>False</b>	turn on dropout
<b>dropout</b>	<b>0.005</b>	dropout ratio

use_reg_activ	None	turn on regularization on activation function
use_reg_weight	None	turn on regularization on weights
use_reg_bias	None	turn on regularization on bias
reg_l1	1e-5	l1 factor
reg_l2	1e-5	l2 factor
use_step_callback	True	turn on stepwise learning rate scheduler
scale_x_mean	False	shift x values to mean
scale_x_std	False	scale x values to std
scale_y_mean	True	shift y values to mean
scale_y_std	True	scale y values to std
normalization_mode	1	normalize hidden layer weights
epo	2000	number of epochs
epostep	10	number of epochs for validation
learning_rate	1e-3	initial learning rate
learning_rate_step	1e-3 1e-4 1e-5 1e-6	stepwise learning rates
epoch_step_reduction	500 500 500 500	number of epochs for stepwise learning rate reduction

Full descriptions for all available keywords are summarized below.

**invd\_index** sets a path to a file containing the pairwise indices for counting inverse distance. Each line should contain a pair of atom indices. If it is not used, all pairwise distances will be included.

**depth** sets the number of hidden layers.

**nn\_size** sets the number of neurons per hidden layer.

**batch\_size** sets the number of training data in one batch.

**activ** sets the activation function. **leaky\_softplus** is used in default.

**activ\_alpha** sets the alpha coefficient in **leaky\_softplus** activation function.

**loss\_weights** sets the weights of energy and gradient loss in the total loss function. It reads two values, e.g., **1 1**

<b>use_dropout</b>	turn on dropout during the training.
<b>dropout</b>	sets the dropout ratio. Note that dropout should not be used together with <b>use_reg_activ</b> , <b>use_reg_weight</b> , or <b>use_reg_bias</b> .
<b>use_reg_activ</b>	turn on regularization on activation function. Available options are: <b>l1</b> l1 regularization, <b>l2</b> l2 regularization, <b>l1_l2</b> l1 and l2 regularization.
<b>use_reg_weight</b>	turn on regularization on hidden layer weights. Available options are: <b>l1</b> l1 regularization <b>l2</b> l2 regularization <b>l1_l2</b> l1 and l2 regularization
<b>use_reg_bias</b>	turn on regularization on hidden layer bias. Available options are: <b>l1</b> l1 regularization <b>l2</b> l2 regularization <b>l1_l2</b> l1 and l2 regularization
<b>reg_l1</b>	sets a l1 factor. It is used when <b>use_reg_activ</b> , <b>use_reg_weight</b> , or <b>use_reg_bias</b> is set to <b>l1</b> or <b>l1_l2</b> .
<b>reg_l2</b>	sets a l2 factor. It is used when <b>use_reg_activ</b> , <b>use_reg_weight</b> , or <b>use_reg_bias</b> is set to <b>l2</b> or <b>l1_l2</b> .
<b>use_step_callback</b>	turn on the stepwise learning rate scheduler. It is turned on in default.
<b>scale_x_mean</b>	shift x values to their mean value. It is not recommended because x values are inverse distances.
<b>scale_x_std</b>	shift x values to their standard deviation. It is not recommended because x values are inverse distances.
<b>scale_y_mean</b>	shift y values to their mean value. It is used in default to standardize the target data.
<b>scale_y_std</b>	shift y values to their standard deviation. It is used in default to standardize the target data.
<b>normalization_mode</b>	normalize the weights of hidden layer to avoid gradient explosion during the training.



**learning\_rate** sets the initial learning rate.

**epo** sets the number of epochs.

**epostep** sets the number of epochs to validate the model.

**learning\_rate\_step** sets the stepwise reduced learning rates for each portion of epochs.

**epoch\_step\_reduction** sets the number of epochs for each portion of learning rates reduction.

## 5.11. NAC and NAC2

The keywords, default values, and short descriptions are listed below.

&NAC and &NAC2		
invd_index	None	path to inverse distance indices file
depth	4	number of hidden layers
nn_size	100	number of neurons per hidden layer
batch_size	64	number of data in one batch
activ	leaky_softplus	activation function
activ_alpha	0.03	activation function coefficient alpha
phase_less_loss	False	use phaseless loss for nac
use_dropout	False	turn on dropout
dropout	0.005	dropout ratio
use_reg_activ	None	turn on regularization on activation function
use_reg_weight	None	turn on regularization on weights
use_reg_bias	None	turn on regularization on bias
reg_l1	1e-5	l1 factor
reg_l2	1e-5	l2 factor
use_step_callback	True	turn on stepwise learning rate scheduler
scale_x_mean	False	shift x values to mean
scale_x_std	False	scale x values to std
scale_y_mean	True	shift y values to mean

scale_y_std	True	scale y values to std
normalization_mode	1	normalize hidden layer weights
epo	2000	number of epochs
epostep	10	number of epochs for validation
learning_rate	1e-3	initial learning rate
learning_rate_step	1e-3 1e-4 1e-5 1e-6	stepwise learning rates
epoch_step_reduction	500 500 500 500	number of epochs for stepwise learning rate reduction

Full descriptions for all available keywords are summarized below.

**invd\_index** sets a path to a file containing the pairwise indices for counting inverse distance. Each line should contain a pair of atom indices. If it is not used, all pairwise distances will be included.

**depth** sets the number of hidden layers.

**nn\_size** sets the number of neurons per hidden layer.

**batch\_size** sets the number of training data in one batch.

**activ** sets the activation function. **leaky\_softplus** is used in default.

**activ\_alpha** sets the alpha coefficient in **leaky\_softplus** activation function.

**phase\_less\_loss** use phaseless loss for nac.

**use\_dropout** turn on dropout during the training.

**dropout** sets the dropout ratio. Note that dropout should not be used together with **use\_reg\_activ**, **use\_reg\_weight**, or **use\_reg\_bias**.

**use\_reg\_activ** turn on regularization on activation function. Available options are:

**l1** l1 regularization,  
**l2** l2 regularization,  
**l1\_l2** l1 and l2 regularization.

**use\_reg\_weight** turn on regularization on hidden layer weights. Available options are:

**l1** l1 regularization  
**l2** l2 regularization  
**l1\_l2** l1 and l2 regularization

<b>use_reg_bias</b>	turn on regularization on hidden layer bias. Available options are: <b>l1</b> l1 regularization <b>l2</b> l2 regularization <b>l1_l2</b> l1 and l2 regularization
<b>reg_l1</b>	sets a l1 factor. It is used when <b>use_reg_activ</b> , <b>use_reg_weight</b> , or <b>use_reg_bias</b> is set to <b>l1</b> or <b>l1_l2</b> .
<b>reg_l2</b>	sets a l2 factor. It is used when <b>use_reg_activ</b> , <b>use_reg_weight</b> , or <b>use_reg_bias</b> is set to <b>l2</b> or <b>l1_l2</b> .
<b>use_step_callback</b>	turn on the stepwise learning rate scheduler. It is turned on in default.
<b>scale_x_mean</b>	shift x values to their mean value. It is not recommended because x values are inverse distances.
<b>scale_x_std</b>	shift x values to their standard deviation. It is not recommended because x values are inverse distances.
<b>scale_y_mean</b>	shift y values to their mean value. It is used in default to standardize the target data.
<b>scale_y_std</b>	shift y values to their standard deviation. It is used in default to standardize the target data.
<b>normalization_mode</b>	normalize the weights of hidden layer to avoid gradient explosion during the training.
<b>learning_rate</b>	sets the initial learning rate.
<b>epo</b>	sets the number of epochs.
<b>epostep</b>	sets the number of epochs to validate the model.
<b>learning_rate_step</b>	sets the stepwise reduced learning rates for each portion of epochs.
<b>epoch_step_reduction</b>	sets the number of epochs for each portion of learning rates reduction.

## 5.13. SOC and SOC2

The keywords, default values, and short descriptions are listed below.

&EG and &EG2		
invd_index	None	path to inverse distance indices file
depth	4	number of hidden layers
nn_size	100	number of neurons per hidden layer
batch_size	64	number of data in one batch
activ	leaky_softplus	activation function
activ_alpha	0.03	activation function coefficient alpha
use_dropout	False	turn on dropout
dropout	0.005	dropout ratio
use_reg_activ	None	turn on regularization on activation function
use_reg_weight	None	turn on regularization on weights
use_reg_bias	None	turn on regularization on bias
reg_l1	1e−5	l1 factor
reg_l2	1e−5	l2 factor
use_step_callback	True	turn on stepwise learning rate scheduler
scale_x_mean	False	shift x values to mean
scale_x_std	False	scale x values to std
scale_y_mean	True	shift y values to mean
scale_y_std	True	scale y values to std
normalization_mode	1	normalize hidden layer weights
epo	2000	number of epochs
epostep	10	number of epochs for validation
learning_rate	1e−3	initial learning rate
learning_rate_step	1e−3 1e−4 1e−5 1e−6	stepwise learning rates
epoch_step_reduction	500 500 500 500	number of epochs for stepwise learning rate reduction

Full descriptions for all available keywords are summarized below.

<b>invd_index</b>	sets a path to a file containing the pairwise indices for counting inverse distance. Each line should contain a pair of atom indices. If it is not used, all pairwise distances will be included.
<b>depth</b>	sets the number of hidden layers.
<b>nn_size</b>	sets the number of neurons per hidden layer.
<b>batch_size</b>	sets the number of training data in one batch.
<b>activ</b>	sets the activation function. <b>leaky_softplus</b> is used in default.
<b>activ_alpha</b>	sets the alpha coefficient in <b>leaky_softplus</b> activation function.
<b>use_dropout</b>	turn on dropout during the training.
<b>dropout</b>	sets the dropout ratio. Note that dropout should not be used together with <b>use_reg_activ</b> , <b>use_reg_weight</b> , or <b>use_reg_bias</b> .
<b>use_reg_activ</b>	turn on regularization on activation function. Available options are: <b>l1</b> l1 regularization, <b>l2</b> l2 regularization, <b>l1_l2</b> l1 and l2 regularization.
<b>use_reg_weight</b>	turn on regularization on hidden layer weights. Available options are: <b>l1</b> l1 regularization <b>l2</b> l2 regularization <b>l1_l2</b> l1 and l2 regularization
<b>use_reg_bias</b>	turn on regularization on hidden layer bias. Available options are: <b>l1</b> l1 regularization <b>l2</b> l2 regularization <b>l1_l2</b> l1 and l2 regularization
<b>reg_l1</b>	sets a l1 factor. It is used when <b>use_reg_activ</b> , <b>use_reg_weight</b> , or <b>use_reg_bias</b> is set to <b>l1</b> or <b>l1_l2</b> .
<b>reg_l2</b>	sets a l2 factor. It is used when <b>use_reg_activ</b> , <b>use_reg_weight</b> , or <b>use_reg_bias</b> is set to <b>l2</b> or <b>l1_l2</b> .
<b>use_step_callback</b>	turn on the stepwise learning rate scheduler. It is turned on in default.
<b>scale_x_mean</b>	shift x values to their mean value. It is not recommended because x values are inverse distances.

<b>scale_x_std</b>	shift x values to their standard deviation. It is not recommended because x values are inverse distances.
<b>scale_y_mean</b>	shift y values to their mean value. It is used in default to standardize the target data.
<b>scale_y_std</b>	shift y values to their standard deviation. It is used in default to standardize the target data.
<b>normalization_mode</b>	normalize the weights of hidden layer to avoid gradient explosion during the training.
<b>learning_rate</b>	sets the initial learning rate.
<b>epo</b>	sets the number of epochs.
<b>epostep</b>	sets the number of epochs to validate the model.
<b>learning_rate_step</b>	sets the stepwise reduced learning rates for each portion of epochs.
<b>epoch_step_reduction</b>	sets the number of epochs for each portion of learning rates reduction.

## 5.14. SCH\_EG

The keywords, default values, and short descriptions are listed below.

&SCH_EG		
<b>node_features</b>	128	number of node-embedding feature
<b>n_features</b>	64	number of trainable node features
<b>n_edges</b>	10	maximum number of neighbors
<b>n_filters</b>	64	number of Gaussian filters
<b>use_filter_bias</b>	True	add filter bias
<b>cfc_activ</b>	shifted_softplus	activation function for the filters
<b>n_blocks</b>	3	number of interaction blocks
<b>maxradius</b>	4	maximum radius cutoff
<b>offset</b>	0.0	offset of Gaussian filter centers
<b>sigma</b>	0.4	width of Gaussian filters

mlp	64	neurons per layer in the output MLP
use_mlp_bias	True	add bias to the output MLP
mlp_activ	shifted_softplus	activation function for the MLP
use_output_bias	True	add bias to the output layer
use_step_callback	True	turn on stepwise learning rate scheduler
loss_weights	1 1	weights of energy and gradient loss
epo	2000	number of epochs
epostep	10	number of epochs for validation
learning_rate	1e-3	initial learning rate
learning_rate_step	1e-3 1e-4 1e-5 1e-6	stepwise learning rates
epoch_step_reduction	500 500 500 500	number of epochs for stepwise learning rate reduction

Full descriptions for all available keywords are summarized below.

**node\_features**      number of features for node embedding. It needs to be larger than the largest atomic number in the training data.

**n\_features**          number of trainable node feature for graph convolution.

**n\_edges**             maximum number of neighboring atoms within the radius cutoff.

**n\_filters**            number of trainable Gaussian filters to extract the edge features.

**use\_filter\_bias**      add bias to the Gaussian filters.

**cfc\_activ**            sets the activation function for Gaussian filters. shifted\_softplus is only option.

**n\_blocks**            number of interaction blocks. Larger number will increase the training time. 3–5 usually works well.

**maxradius**          sets a radius in Angstrom to cut a spheric atomic environment.

**offset**                apply an offset to the center of the Gaussiann filters.

**sigma**                sets the width of the Gaussian filters. Narrower Gaussian filter requires a greater number of filter

<b>mlp</b>	specifies the neurons per hidden layers in the output MLP, e.g., 64 64 64 will build three hidden layers and each contains 64 neurons.
<b>use_mlp_bias</b>	add bias to the output MLP layers.
<b>mlp_activ</b>	sets the activation function for the output MLP layers. shifted_softplus is only option.
<b>use_step_callback</b>	turn on the stepwise learning rate scheduler. It is turned on in default.
<b>use_output_bias</b>	add bias to the last output layer.
<b>use_step_callback</b>	turn on the stepwise learning rate scheduler. It is turned on in default.
<b>loss_weights</b>	sets the weights of energy and gradient loss in the total loss function. It reads two values, e.g., 1 1
<b>learning_rate</b>	sets the initial learning rate.
<b>epo</b>	sets the number of epochs.
<b>epostep</b>	sets the number of epochs to validate the model.
<b>learning_rate_step</b>	sets the stepwise reduced learning rates for each portion of epochs.
<b>epoch_step_reduction</b>	sets the number of epochs for each portion of learning rates reduction.

## 5.15. SCH\_SOC

The keywords, default values, and short descriptions are listed below.

<b>&amp;SCH_SOC</b>		
<b>node_features</b>	128	number of node-embedding feature
<b>n_features</b>	64	number of trainable node features
<b>n_edges</b>	10	maximum number of neighbors
<b>n_filters</b>	64	number of Gaussian filters
<b>use_filter_bias</b>	True	add filter bias
<b>cfc_activ</b>	shifted_softplus	activation function for the filters
<b>n_blocks</b>	3	number of interaction blocks



maxradius	4	maximum radius cutoff
offset	0.0	offset of Gaussian filter centers
sigma	0.4	width of Gaussian filters
mlp	64	neurons per layer in the output MLP
use_mlp_bias	True	add bias to the output MLP
mlp_activ	shifted_softplus	activation function for the MLP
use_output_bias	True	add bias to the output layer
use_step_callback	True	turn on stepwise learning rate scheduler
epo	2000	number of epochs
epostep	10	number of epochs for validation
learning_rate	1e-3	initial learning rate
learning_rate_step	1e-3 1e-4 1e-5 1e-6	stepwise learning rates
epoch_step_reduction	500 500 500 500	number of epochs for stepwise learning rate reduction

Full descriptions for all available keywords are summarized below.

**node\_features**      number of features for node embedding. It needs to be larger than the largest atomic number in the training data.

**n\_features**      number of trainable node feature for graph convolution.

**n\_edges**      maximum number of neighboring atoms within the radius cutoff.

**n\_filters**      number of trainable Gaussian filters to extract the edge features.

**use\_filter\_bias**      add bias to the Gaussian filters.

**cfc\_activ**      sets the activation function for Gaussian filters. shifted\_softplus is only option.

**n\_blocks**      number of interaction blocks. Larger number will increase the training time. 3–5 usually works well.

**maxradius**      sets a radius in Angstrom to cut a spheric atomic environment.

**offset**      apply an offset to the center of the Gaussiann filters.

<b>sigma</b>	sets the width of the Gaussian filters. Narrower Gaussian filter requires a greater number of filter
<b>mlp</b>	specifies the neurons per hidden layers in the output MLP, e.g., 64 64 64 will build three hidden layers and each contains 64 neurons.
<b>use_mlp_bias</b>	add bias to the output MLP layers.
<b>mlp_activ</b>	sets the activation function for the output MLP layers. shifted_softplus is only option.
<b>use_step_callback</b>	turn on the stepwise learning rate scheduler. It is turned on in default.
<b>use_output_bias</b>	add bias to the last output layer.
<b>use_step_callback</b>	turn on the stepwise learning rate scheduler. It is turned on in default.
<b>learning_rate</b>	sets the initial learning rate.
<b>epo</b>	sets the number of epochs.
<b>epostep</b>	sets the number of epochs to validate the model.
<b>learning_rate_step</b>	sets the stepwise reduced learning rates for each portion of epochs.
<b>epoch_step_reduction</b>	sets the number of epochs for each portion of learning rates reduction.

## 5.16. E2N2\_EG

The keywords, default values, and short descriptions are listed below.

<b>&amp;E2N2_EG</b>		
<b>n_edges</b>	10	maximum number of neighbors
<b>maxradius</b>	4	maximum radius cutoff
<b>n_features</b>	64	number of trainable node features
<b>n_blocks</b>	3	number of interaction blocks
<b>l_max</b>	1	rotation order
<b>parity</b>	True	Use tensor parity

n_rbf	20	number of radial basis functions
trainable_rbf	True	trainable rbf weights
rbf_cutoff	6	exponential of the rbf cutoff function
rbf_layer	2	number of radial net hidden layer
rbf_neurons	64	number of radial net neurons/layer
rbf_act	silu	activation function for the radial net
normalization_y	component	spheric harmonic normalization scheme
normalize_y	True	Normalize spheric harmonic vectors
self_connection	True	add self-connection contribution
gate	True	use gated activation
loss_weights	1 1	weights of energy and gradient loss
epo	400	number of epochs
epostep	10	number of epochs for validation
subset	0	use part of training data
batch_size	64	batch size
val_batch_size	0	validation batch size
nbatch	0	number of batch
learning_rate	1e-3	initial learning rate
learning_rate_step	1e-3 1e-4 1e-5 1e-6	stepwise learning rates
epoch_step_reduction	100 100 100 100	number of epochs for stepwise learning rate reduction

Full descriptions for all available keywords are summarized below.

**n\_edges** maximum number of neighboring atoms within the radius cutoff.

**maxradius** sets a radius in Angstrom to cut a spheric atomic environment.

**n\_features** number of trainable node feature for graph convolution.

**n\_blocks** number of interaction blocks. Larger number will increase the training time. 3–5 usually works well.

**l\_max** the largest rotation order that will be kept in tensor products

<b>parity</b>	include the parity of tensors
<b>n_rfb</b>	number of Bessel radial basis function
<b>trainable_rbf</b>	update the weights of the radial basis function during training.
<b>rbf_cutoff</b>	exponential of the cutoff function used in DimeNet.
<b>rbf_layers</b>	number of hidden layers in the radial basis network
<b>rbf_neurons</b>	number of neurons per hidden layer in the radial basis network
<b>rbf_act</b>	activation function in the radial basis network. SiLU is recommended. Another option is <b>shifted_softplus</b> .
<b>normalization_y</b>	chooses the normalization scheme for spheric harmonic vectors.
<b>normalize_y</b>	normalizes the spheric harmonic vectors.
<b>self_connection</b>	include self-connection when updating the node feature
<b>gate</b>	use gated activation for tensor convolution. The activation functions are <b>silu</b> for even scaler and even tensor and <b>tanh</b> for old scaler and old tensor. The keywords are: <b>act_scalars_e silu</b> <b>act_scalars_o tanh</b> <b>act_gates_e silu</b> <b>act_gates_e tanh</b>
<b>loss_weights</b>	sets the weights of energy and gradient loss in the total loss function. It reads two values, e.g., <b>1 1</b>
<b>epo</b>	sets the number of epochs.
<b>epostep</b>	sets the number of epochs to validate the model.
<b>subset</b>	use a portion of the training data for training.
<b>batch_size</b>	specify the batch size of training data. It is recommended to use <b>5</b> .
<b>val_batch_size</b>	specify the batch size of validation data. It is recommended to use <b>5</b> .

**nbatch** specify the number of batches in training data. A value greater than 0 will overwrite the **batch\_size**, the batch size will be automatically determined by the number of training data. The default value is 0. It could be useful to keep the same number of batches during adaptive sampling as the total number of training data is increasing.

**learning\_rate** sets the initial learning rate.

**learning\_rate\_step** sets the stepwise reduced learning rates for each portion of epochs.

**epoch\_step\_reduction** sets the number of epochs for each portion of learning rates reduction.

## 5.17. E2N2\_NAC

(not available yet)

## 5.18. E2N2\_SOC

The keywords, default values, and short descriptions are listed below.

&E2N2_SOC		
n_edges	10	maximum number of neighbors
maxradius	4	maximum radius cutoff
n_features	64	number of trainable node features
n_blocks	3	number of interaction blocks
l_max	1	rotation order
parity	True	Use tensor parity
n_rbf	20	number of radial basis functions
trainable_rbf	True	trainable rbf weights
rbf_cutoff	6	exponential of the rbf cutoff function
rbf_layer	2	number of radial net hidden layer
rbf_neurons	64	number of radial net neurons/layer
rbf_act	silu	activation function for the radial net
normalization_y	component	spheric harmonic normalization scheme
normalize_y	True	Normalize spheric harmonic vectors

self_connection	True	add self-connection contribution
gate	True	use gated activation
epo	400	number of epochs
epostep	10	number of epochs for validation
subset	0	use part of training data
batch_size	64	batch size
nbatch	0	number of batch
learning_rate	1e-3	initial learning rate
learning_rate_step	1e-3 1e-4 1e-5 1e-6	stepwise learning rates
epoch_step_reduction	100 100 100 100	number of epochs for stepwise learning rate reduction

Full descriptions for all available keywords are summarized below.

<b>n_edges</b>	maximum number of neighboring atoms within the radius cutoff.
<b>maxradius</b>	sets a radius in Angstrom to cut a spheric atomic environment.
<b>n_features</b>	number of trainable node feature for graph convolution.
<b>n_blocks</b>	number of interaction blocks. Larger number will increase the training time. 3–5 usually works well.
<b>l_max</b>	the largest rotation order that will be kept in tensor products
<b>parity</b>	include the parity of tensors
<b>n_rfb</b>	number of Bessel radial basis function
<b>trainable_rbf</b>	update the weights of the radial basis function during training.
<b>rbf_cutoff</b>	exponential of the cutoff function used in DimeNet.
<b>rbf_layers</b>	number of hidden layers in the radial basis network
<b>rbf_neurons</b>	number of neurons per hidden layer in the radial basis network
<b>rbf_act</b>	activation function in the radial basis network. SiLU is recommended. Another option is <b>shifted_softplus</b> .

<b>normalization_y</b>	chooses the normalization scheme for spheric harmonic vectors.
<b>normalize_y</b>	normalizes the spheric harmonic vectors.
<b>self_connection</b>	include self-connection when updating the node feature
<b>gate</b>	use gated activation for tensor convolution. The activation functions are <b>silu</b> for even scaler and even tensor and <b>tanh</b> for old scaler and old tensor. The keywords are: <b>act_scalars_e silu</b> <b>act_scalars_o tanh</b> <b>act_gates_e silu</b> <b>act_gates_e tanh</b>
<b>epo</b>	sets the number of epochs.
<b>epostep</b>	sets the number of epochs to validate the model.
<b>subset</b>	use a portion of the training data for training.
<b>batch_size</b>	specify the batch size of training data.
<b>nbatch</b>	specify the number of batches in training data. A value greater than 0 will overwrite the <b>batch_size</b> , the batch size will be automatically determined by the number of training data. The default value is 0. It is useful to keep the same number of batches during adaptive sampling as the total number of training data is increasing.
<b>learning_rate</b>	sets the initial learning rate.
<b>learning_rate_step</b>	sets the stepwise reduced learning rates for each portion of epochs.
<b>epoch_step_reduction</b>	sets the number of epochs for each portion of learning rates reduction.

## 5.19. DIME\_NAC

The keywords, default values, and short descriptions are listed below.

<b>&amp;DIME_NAC</b>		
<b>model_type</b>	<b>None</b>	Choose DimeNet model
<b>batch_size</b>	<b>64</b>	batch size
<b>val_size</b>	<b>64</b>	validation size

hidden_channels		
blocks	3	number of interaction blocks
bilinear	True	Use tensor parity
spherical	20	number of radial basis functions
radial	True	trainable rbf weights
lr	1e-3	initial learning rate
epo	400	number of epochs

Full descriptions for all available keywords are summarized below.

**model\_type** choose DimeNet model. Available options are:  
**None** DimeNet model. This is default.  
**pp** DimeNet++ model.

**batch\_size** specify the batch size of training data.

**val\_size** specify the size of validation data.

**hidden\_channels** specify the number of hidden channels.

**blocks** specify the number of interaction blocks.

**bilinear** specify the number of bilinear functions.

**spherical** specify the number of spherical functions.

**radial** specify the number of radial functions.

**lr** specify the starting learning rate.

**epo** specify the number of training epochs

## 5.20. FILE

The keywords, default values, and short descriptions are listed below.

&FILE		
natom	0	number of atoms
file	None	path to a list file to read QC calculation results



Full descriptions for all available keywords are summarized below.

**natom** sets the number of atoms for reading the coordinates from the QC calculation logfiles.

**file** read the path to a list file for extracting the QC data from the calculation logfiles. If a file name is provided, it assumes that the list file is in the current folder. In the list file, each line should contain a path to a QC calculation folder.

## 5. Nonadiabatic molecular dynamics

### 5.1. Fewest switches surface hopping

### 5.2. Zhu-Nakamura surface hopping

## 6. Machine learning models

### 6.1. Preparing training data

### 6.2. Creating a neural network

First, we create a model to predict energies, forces, and non-adiabatic couplings (if requested). The input example below shows the frequently used keywords for creating models.

**jobtype** determines the type of calculation. It takes 'train' for training neural networks, 'prediction' for predicting energies, forces, and non-adiabatic couplings, 'adaptive' for adaptive sampling of conformational space using molecular dynamics trajectories, and 'md' for molecular dynamics simulation.

PyRAI<sup>2</sup>MD has a flexible training scheme depending on the available computing resources. When **ml\_ncpu = 1**, all models will be trained sequentially. When **ml\_ncpu <=4**, all models will be

trained in subprocess so they can use all given numbers of CPUs. If **ml\_ncpu > 4**, the extra CPU resources will be used to parallelize the training, which is automatically managed by TensorFlow.

## 6.3. Training a neural network

## 6.5. Adaptive sampling

# 7. External quantum chemical program

## 7.1. Molcas

## 7.2. BAGEL

## 7.3. ORCA

## 7.4. GFN-xTB

## 7.5. MNDO

