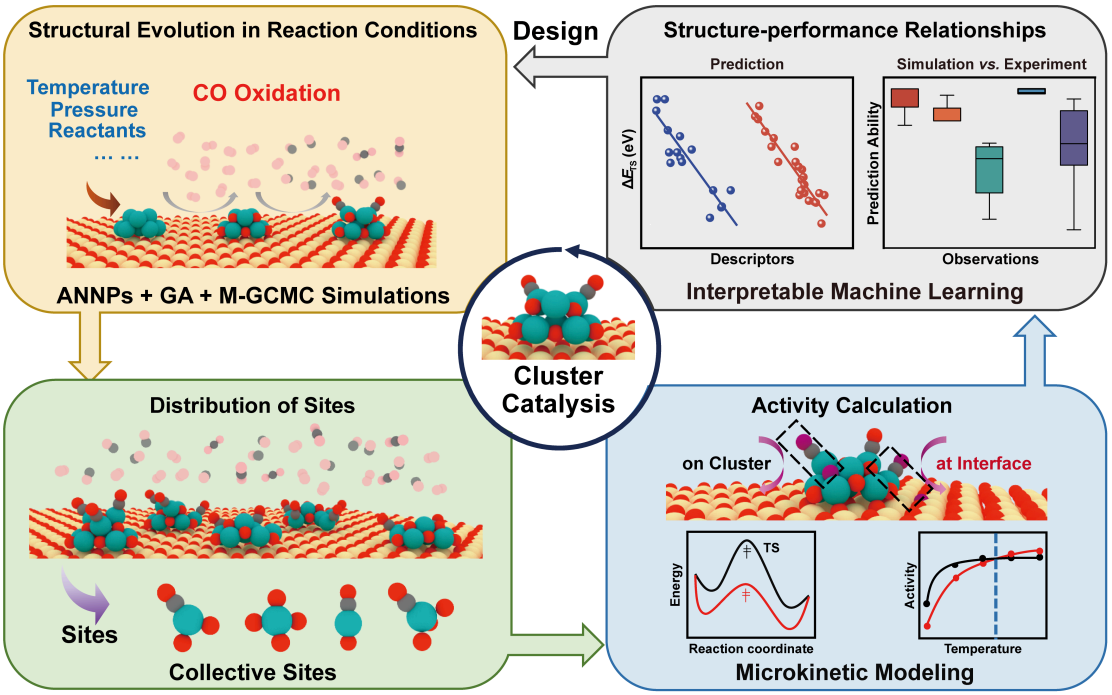
The Framework to Study Collectivity of Active Sites for Cluster Catalysis

This repository provides a computational framework to investigate the **collectivity of active sites** in cluster-based catalysis. It integrates several modules for modeling, simulation, and analysis of catalytic behavior using machine learning potentials and atomistic simulations.



📌 Key Components

1. **Artificial Neural Network Potentials (ANNPs) Training**: For accurate energy and force predictions.
2. **Genetic Algorithm (GA)**: Used for global optimization in catalyst design.
3. **Modified GCMC (M-GCMC) Simulations**: Simulates realistic catalytic environments under operational conditions.
4. **Reaction Mechanism Calculations**: For identifying transition states and elementary steps.
5. **Microkinetic Modeling (MKM)**: For simulating reaction kinetics and understanding the rate-determining steps.
6. **Catalytic Descriptors**: Uses compressed sensing methods (e.g., SISSO) to uncover key descriptors.

👥 Developers

Jia-Lan Chen (jlchen20@mail.ustc.edu.cn)

Advisors: Jin-Xun Liu (jxliu86@ustc.edu.cn) and Wei-Xue Li (wxli70@ustc.edu.cn)

🧪 Methods

- **EANN (Embedded Atom Neural Network)**: A simple yet efficient neural network framework for modeling atomic interactions.

- **Active learning to expand the configurations**: Iteratively expands configuration space to improve ANNP accuracy.
- **GA**: Integrated with ASE for global structure optimization.
- **M-GCMC simulations**: Modified GCMC (M-GCMC) simulations for simulating adsorption and reaction processes under *operational* conditions and search metastable structures.
- **SISSO (The sure independence screening and sparsifying operator)**: A dimensionality reduction method to identify key descriptors for catalytic performance prediction.

📁 Contents

Folder	Description
`nn/`	Scripts for training reliable ANNPs with active learning.
`str/`	Metastable structure search using GA and M-GCMC.
`sites/`	Identification and distribution analysis of active sites.
`ts/`	Transition state search using NEB and DyNEB with ANNPs.
`mkm/`	Microkinetic modeling input files.
`desp/`	SISSO input data and training sets for descriptor learning.

⚙️ Module Details

✅ Training Reliable ANNPs (`nn/`)

This process includes generating the initial dataset and expanding the configuration space using an active learning algorithm. The ASE interface for the Embedded Atom Neural Network (EANN) is employed via `eann.py`.

To train the model and handle uncertain or poorly predicted structures:

```
python3 train_forces.py # Handle uncertain regions in the configuration space
python3 get_atoms # Identify structures with poor predictive accuracy
```

✅ Searching metastable structures under *operational* conditions

These simulations are used to search metastable structures under *operational* conditions.

All relevant scripts and configuration files can be found in the **str** folder.

- **Genetic algorithm (GA)**

  Used to optimize catalyst structures via evolutionary principles.

```
python3 initial.py \
  --poscar POSCAR_base \
  --db gadb.db \
  --atoms 8x10 29x8 \
  --size 24 \
  --zvac 2.0 # Initialize the GA search

python3 ga_test.py  \
  --db gadb.db    \
  --processes 4 # Run the GA optimization
```

- **Modified GCMC (M-GCMC) simulations**

  A modified GCMC approach tailored for metastable structure database under varying chemical potentials of CO and O_2 .

```
python3 main.py # Execute the GCMC simulation in gcmc folder
python3 main.py  # Execute the M-GCMC simulation in m-gcmc folder
```

✅ Identification of Sites (`sites/`)

This step involves identifying metastable structures and analyzing the distribution of potential active sites.

The corresponding scripts and data are located in the **site** folder.

- **Metastable structures and their distribution**

  Identifies metastable structures.

```
python3 meta_min_energy_Cu3.py \
  --energy_files energy_meta energy_meta2 \
  --poscar_dirs dataall1 dataall2 \
  --temperature 400 \
  --energy_cutoff 0.5 # metastable structures
```

- **Site distribution**

  Determines the type and distribution of every site.

```
python3 meta_min_energy_next_sites.py \
--energy_files energy_meta energy_meta2 \
  --poscar_dirs dataall1 dataall2 \
  --temperature 400 \
  --energy_cutoff 0.5 # every site and its distribution
```

✅ Transition State Search (`ts/`)

ANNPs are employed as an initial guess for transition states in catalytic reactions.

The automated scripts for transition state searches are located in the ts folder.

- **Dynamic NEB (DyNEB)**
- Executes a dynamic NEB search to refine the transition state.

```
python3 dyneb.py --poscar1 POSCAR1 --poscar2 POSCAR2 \
  --pes EANN_PES_DOUBLE.pt --atomtype Cu O C \
  --n_images 3 --traj myneb.traj --dyneb  # Run DyNEB for more precise transition state search
```

`--poscar1`: Initial structure (POSCAR format)

`--poscar2`: Final structure (POSCAR format)

`--pes`: Path to the trained ANNPs file

`--atomtype`: List of atom types in the system

`--n_images`: Number of intermediate NEB images

`--traj`: Output trajectory file

`--dyneb`: Enables the dynamic NEB mode for improved TS localization

To visualize the NEB potential energy surface (PES), run:

```
python3 neb_pos.py
```

This will generate a file named `barrier.png`, which shows the energy profile along the reaction coordinate.

✅ Microkinetic Modeling (`mkm/`)

Microkinetic modeling is performed by MKMCXX to simulate reaction kinetics and identify key rate-determining steps.

The relevant scripts can be found in the **mkm** folder.

✅ Descriptor Discovery (`desp/`)

The transition state energies are predicted by SISSO. This helps in identifying key catalytic descriptors for further analysis.

The input files are located in the **desp** folder.

📦 Dependencies

Package	Purpose	Version
ASE	Structure handling & GA/M-GCMC interface	3.22.1
MKMCXX	Microkinetic modeling	—
PyTorch	ANNP model training	1.8.0
NumPy	Numerical operations	1.20.3
Matplotlib	Plotting	3.4.3

Package	Purpose	Version
Pandas	Data handling & export	1.3.3
SciPy	Polynomial fitting	1.7.1

🙏 Acknowledgements

We sincerely thank **Prof. Bin Jiang** and **Dr. Yaolong Zhang** for providing the EANN package and valuable guidance.

📚 References

1. Zhang, Y. L.; Hu, C.; Jiang, B. Embedded Atom Neural Network Potentials: Efficient and Accurate Machine Learning with a Physically Inspired Representation. *J. Phys. Chem. Lett.* **2019**, 10 (17), 4962-4967.
2. Lin, Q. D.; Zhang, L.; Zhang, Y. L.; Jiang, B. Searching Configurations in Uncertainty Space: Active Learning of High-Dimensional Neural Network Reactive Potentials. *J. Chem. Theory Comput.* **2021**, 17 (5).
3. Zhang, Y. Z.; Wang, H. D.; Chen, W. J.; Zeng, J. Z.; Zhang, L. F.; Wang, H.; Weinan, E. DP-GEN: A concurrent learning platform for the generation of reliable deep learning based potential energy models. *Comput. Phys. Commun.* **2020**, 253, 107206.
4. Vilhelmsen, L. B.; Hammer, B. A genetic algorithm for first principles global structure optimization of supported nano structures. *J. Chem. Phys.* **2014**, 141 (4), 044711.
5. Liu, J.-X.; Su, Y.; Filot, I. A. W.; Hensen, E. J. M. A Linear Scaling Relation for CO Oxidation on CeO2-Supported Pd. *J. Am. Chem. Soc.* **2018**, 140 (13), 4580-4587.
6. Lindgren, P.; Kastlunger, G.; Peterson, A. A. Scaled and Dynamic Optimizations of Nudged Elastic Bands. *J. Chem. Theory Comput.* **2019**, 15 (11), 5787-5793.
7. Kolsbjerg, E. L.; Groves, M. N.; Hammer, B. An automated nudged elastic band method. *J. Chem. Phys.* **2016**, 145 (9), 094107.
8. Filot, I. A. W.; van Santen, R. A.; Hensen, E. J. M. The Optimally Performing Fischer–Tropsch Catalyst. *Angew. Chem. Int. Ed.* **2014**, 53 (47), 12746-12750.
9. Ouyang, R.; Curtarolo, S.; Ahmetcik, E.; Scheffler, M.; Ghiringhelli, L. M. SISSO: A compressed-sensing method for identifying the best low-dimensional descriptor in an immensity of offered candidates. *Phys. Rev. Mater.* **2018**, 2 (8), 083802.