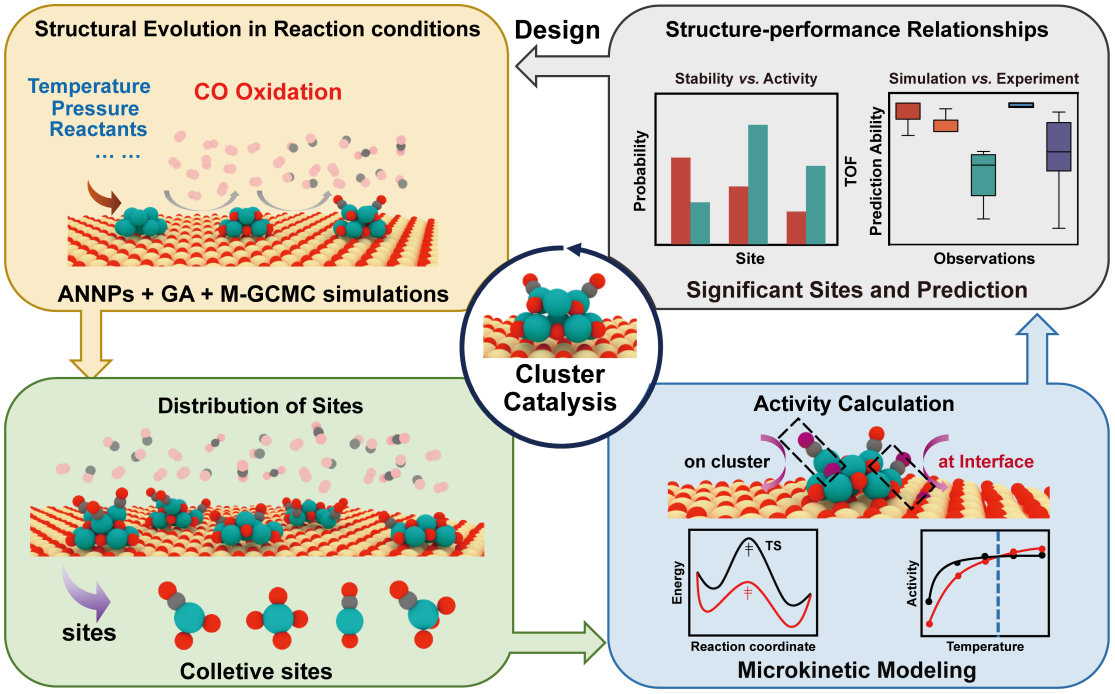
The Framework to Study Collectivity of Active Sites for Cluster Catalysis

This repository provides a framework for the study of collective behavior in cluster catalysis. It includes the following components:

1. **Artificial Neural Network Potentials (ANNPs) Training**: For accurate energy and force predictions.
2. **Genetic Algorithm (GA)**: Applied for global optimization in catalyst design.
3. **Modified GCMC (M-GCMC) Simulations**: An enhanced version of GCMC to simulate complex catalytic environments.
4. **Reaction Mechanism Calculations**: To investigate pathways and intermediates under catalytic reactions.
5. **Microkinetic Modeling (MKM)**: For simulating reaction kinetics and understanding the rate-determining steps.
6. **Catalytic Descriptors**: Predictive metrics for catalytic activity, aiding in catalyst screening and optimization.



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**: Utilized to iteratively expand the configuration space, improving model accuracy.

- **GA**: Genetic algorithm integrated with the Atomic Simulation Environment (ASE) for optimization of catalyst structures.
- **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

This repository includes several folders that implement the framework:

1. **nn**: Contains scripts for training reliable ANNPs, including training parameters and the active learning algorithm.
2. **str**: Focuses on searching metastable structures under *operational* conditions using GA and M-GCMC simulations.
3. **sites**: Provides tools for the identification of sites and their distributions.
4. **ts**: Includes scripts for searching transition states with ANNPs, utilizing auto NEB and dyNEB methods.
5. **mkm**: Contains input files for microkinetic modeling.
6. **desp**: Provides input data and training sets for predicting transition state energies using the SISSO method.

Training reliable ANNPs

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`.

See more details in the nn folder.

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; # Initialize the GA search
python3 ga_test.py # Run the GA optimization
```

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

  A modified GCMC approach tailored for metastable structure database under varying chemical potentials.

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

Identification of sites

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

The corresponding scripts and data are located in the sites folder.

- **Metastable structures and their distribution**

  Identifies metastable structures.

```
python3 meta_min_energy_Cu3.py # metastable structures
```

- **Site distribution**

  Determines the type and distribution of every site.

```
python3 meta_min_energy_next_sites.py # every site and its distribution
```

Transition State Search
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.

- **Automated NEB**

  Performs an automated Nudged Elastic Band (NEB) calculation to identify the transition state.

```
python3 a_neb.py  # Run automated NEB for transition state search
```

- **Dynamic NEB (DyNEB)**

  Executes a dynamic NEB search to refine the transition state.

```
python3 dyneb.py  # Run DyNEB for a more precise transition state search
```

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.

Prediction of Transition State Energies
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 and Softwares

- **ASE**: Used for structure search through GA and M-GCMC simulations.
  Version: `ASE=3.22.1`
- **MKMCXX**: A software suite for microkinetic modeling.

- **PyTorch**: Utilized for training ANNPs.
  Version: `torch=1.8.0`
- **NumPy**: For vector and matrix operations.
  Version: `numpy=1.20.3`
- **Matplotlib**: For generating plots and visualizations.
  Version: `matplotlib=3.4.3`
- **Pandas**: To export outputs into Excel files for data handling and analysis.
  Version: `pandas=1.3.3`
- **SciPy**: Applied for polynomial fitting in active learning processes.
  Version: `scipy=1.7.1`

Related Publication

J.-L. Chen, X.-C. Jiang, L. Feng, J.-Z. Zhu, J.-W. Zhao, J.-X. Liu* and W.-X. Li*, Collectivity of Active Sites for Cluster Catalysis under Operational Conditions, ***Nature Chemistry*** (2024). (submitted)

Acknowledgement

We genuinely thank Professor Bin Jiang and Dr. Yaolong Zhang for providing the EANN package.

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.