Modular Large Model Driven Catalyst Design For Polyolefin Elastomer Materials

Wentao Li[a], **Xiaonan Wang**[a]

[a] *Department of Chemical Engineering, Tsinghua University, Beijing 100084, China.* liwt24@mails.tsinghua.edu.cn wangxiaonan@tsinghua.edu.cn

* Presenting author

1. Introduction

In modern chemical engineering and materials science, the development and optimization of catalysts are at the core of chemical reaction engineering, directly determining the efficiency and economic viability of industrial processes. Catalysts not only enhance the rates of chemical reactions but also selectively produce target products and reduce the formation of by - products. In the fields of polyolefin elastomers (POE), a key material for new energy, metallocene catalysts have drawn much attention due to their unique catalytic efficiency and specific reaction mechanisms. The development and optimization of these catalysts require not only an in - depth understanding of their molecular - level reaction mechanisms but also the use of high - throughput computing and large - scale AI - based searches for rapid screening and performance evaluation.

To achieve high - throughput calculation and screening of POE catalysts, we have established a complete calculation flow from Density Function Theory (DFT) calculations to Kinetic Monte Carlo (KMC) simulations, which simulates the elementary reactions of polymer monomers to the formation of the entire polymer chain. We have innovatively proposed a full - process calculation and simulation framework for evaluating POE catalysts.

Although the high - throughput calculation process is efficient compared to catalyst preparation and evaluation experiments, it still consumes a large amount of computing power. Thus, we have proposed a modular expert model that integrates a large amount of general knowledge in the materials field and knowledge of the POE catalytic system to achieve efficient recommendation of POE catalysts. This modular processing of knowledge has not only achieved excellent results in the field of material property prediction but also provided good predictions for catalyst performance, such as activation energy, ΔH, etc.

In this work, we propose a framework for the rational design of POE catalysts called POECatDesign, aimed at high-throughput screening of POE catalysts to discover catalysts with better performance and address the performance bottleneck issues of industrial catalysts.

2. Overall framework

The POECatDesign framework is shown in Figure 1. The following are specific methods for high - throughput calculation and screening of catalysts.

Modular large-model training and inference Training process

Each module of the modular large - model is trained separately. A certain type of material property is selected for training in each module. The dataset is sourced from datasets in Matminer with more than 10,000 samples, and data such as single - point energy and activation energy of POE catalysts are integrated to obtain multiple expert modules. In property prediction for downstream tasks, we adaptively select k expert modules to jointly predict properties.

Activation energy and transition-state calculation

We use the ComplexGen library to automatically generate 3D structures of catalysts based on ligand SMILES and pair them with substrates. And we perform geometric optimization and transition-state calculations through the Gaussian 6 software. The optimized single-point energy and transition - state energy are used for the training of the modular large-model.

Catalyst recommendation by modular large-model

For the catalysts generated on a large scale by ComplexGen, we directly predict catalytic parameters (activation energy) through the modular large - model to obtain recommended candidate catalysts.

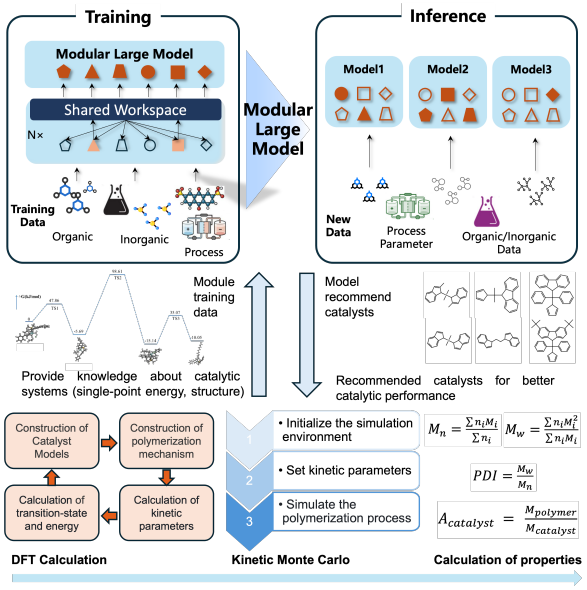

Fig. 1: POECatDesign framework diagram

Kinetic Monte Carlo (KMC) simulation

The polymerization process can be reflected through Kinetic Monte Carlo simulation, obtaining specific parameters of the final product, thereby evaluating the performance of the catalyst

3. Related work

There is limited research on AI-driven rational design of catalysts[1, 2, 3], but it has not addressed the complex multi-step reaction of catalytic polymerization. Therefore, it is essential to establish a high-throughput computational process for this system and to develop a modular AI model for better material property predictions.

4. Catalyst evaluation indicators and newly designed catalysts

$$M_n = \frac{\sum n_i M_i}{\sum n_i} \tag{1}$$

$$M_w = \frac{\sum n_i M_i^2}{\sum n_i M_i} \tag{2}$$

$$PDI = \frac{M_w}{M_n} \tag{3}$$

$$A_{catalyst} = \frac{M_{polymer}}{M_{catalyst}} \tag{4}$$

M represents mass, M_n represents the number average molecular weight, M_w represents the weight average molecular weight, PDI represents the polymer dispersity index, and $A_catalyst$ represents the catalyst activity.

References

[1] Dong Hyeon Mok, Hong Li, Guiru Zhang, Chaehyeon Lee, Kun Jiang, and Seoin Back. Data-driven discovery of electrocatalysts for co2 reduction using active motifs-based machine learning. *Nature Communications*, 14(1):7303, 2023.

[2] Zhilong Song, Linfeng Fan, Shuaihua Lu, Chongyi Ling, Qionghua Zhou, and Jinlan Wang. Inverse design of promising electrocatalysts for co2 reduction via generative models and bird swarm algorithm. *Nature Communications*, 16(1):1053, Jan 2025.

[3] Peng Yin, Xiangfu Niu, Shuo-Bin Li, Kai Chen, Xi Zhang, Ming Zuo, Liang Zhang, and Hai-Wei Liang. Machine-learning-accelerated design of high-performance platinum intermetallic nanoparticle fuel cell catalysts. *Nature Communications*, 15(1):415, 2024.