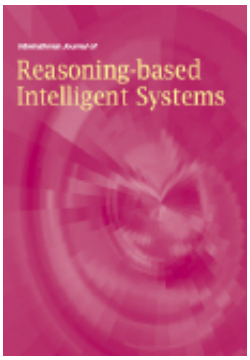


International Journal of Reasoning-based Intelligent Systems

ISSN online: 1755-0564 - ISSN print: 1755-0556
https://www.inderscience.com/ijris

Agent-driven multi-scale simulation for predicting the catalytic activity of complexes

Shenshen Li, Shufang Chen, Juanjuan Bai

DOI: 10.1504/IJRIS.2026.10075856

Article History:

Received:	15 October 2025
Last revised:	05 November 2025
Accepted:	05 November 2025
Published online:	28 January 2026

Agent-driven multi-scale simulation for predicting the catalytic activity of complexes

Shenshen Li*

College of Biological Engineering,
Xinxiang Institute of Engineering,
Xinxiang, 453700, China
Email: LSS201511195@163.com
*Corresponding author

Shufang Chen

College of Materials and Chemical Engineering,
Henan University of Urban Construction,
Pingdingshan, 467000, China
Email: chenshufang12345@126.com

Juanjuan Bai

College of Biological Engineering,
Xinxiang Institute of Engineering,
Xinxiang, 453700, China
Email: wxbj2007@163.com

Abstract: This paper presents an agent-driven multi-scale simulation framework for efficiently and accurately predicting the catalytic activity of complexes. This framework constructs the reaction path search as a Markov decision process, adopts hierarchical reinforcement learning agents to actively explore the potential energy surface, and combines the equivariant graph neural network potential function to ensure quantum accuracy. Experiments on the open catalyst project (OC20) dataset show that the average absolute error of this framework in adsorption energy prediction is significantly reduced to 0.291 eV, the force prediction error is 0.072 eV/Å, and it can converge to a stable configuration in an average of only 18.3 steps. It is superior to the existing mainstream methods in both accuracy and efficiency. This research provides a new paradigm of intelligent computing for catalyst design and promotes the development of multi-scale simulation towards autonomous decision making and efficient exploration.

Keywords: agent-driven; multi-scale simulation; catalytic activity prediction; reinforcement learning; machine learning.

Reference to this paper should be made as follows: Li, S., Chen, S. and Bai, J. (2026) 'Agent-driven multi-scale simulation for predicting the catalytic activity of complexes', *Int. J. Reasoning-based Intelligent Systems*, Vol. 18, No. 7, pp.21–31.

Biographical notes: Shenshen Li is currently employed in College of Biological Engineering at·Xinxiang Institute of Engineering, China. She obtained her Master's degree from Zhengzhou University in 2012. She has published five papers, two of which have been indexed by SCI. Her research interests include metal complexes catalysis and mineral element analysis.

Shufang Chen is currently employed in College of Materials and Chemical Engineering at Henan University of Urban Construction, China. She obtained her Master's degree from Zhengzhou University in 2013. She has published two papers. Her research interests include multinucleated complexes and gas adsorption.

Juanjuan Bai is currently employed in College of Biological Engineering at Xinxiang Institute of Engineering, China. She obtained her Master's degree in Environmental Engineering from Hainan University·in 2015. She has published five papers. Her research interests include machine learning, water pollution control and treatment.

1 Introduction

1.1 Catalytic science: an era of coexisting opportunities and challenges

Heterogeneous catalysis is the cornerstone of modern chemical industry, energy conversion and environmental pollution control. From ammonia synthesis to vehicle exhaust purification, and to the key reactions in the future hydrogen economy – water cracking and carbon dioxide reduction, the design and development of efficient catalysts have always been the core driving force (Greeley et al., 2006; Greeley, 2016). Traditionally, the discovery of new catalysts has relied heavily on 'trial-and-error' experimental screening, a process that not only costs a huge amount of money but also takes a long time.

With the rapid development of theoretical chemistry and computing power, the design of computationally driven new catalysts is highly anticipated and regarded as the 'holy grail' for accelerating material research and development. First-principles calculations represented by density functional theory (DFT) can reveal the mechanism of catalytic reactions and predict the adsorption behaviour of intermediates with quantum mechanical precision, thereby providing a microscopic perspective Nørskov et al. (2011) for understanding the origin of catalytic activity. However, the high computational cost of DFT calculation itself makes it inadequate when exploring the vast space of catalyst composition and structure. The computational cost of DFT increases cubic or even higher with the number of atoms in the system. When catalyst components, surface structures, and adsorption sites are systematically screened, the number of possible configurations explodes in combination, making DFT calculations for each candidate structure infeasible in practice, thus severely limiting the scope of exploration. This computational 'curse' severely restricts our systematic exploration of complex catalytic systems, especially the actual catalytic processes involving multiple sites and complex reaction networks.

1.2 From quantum scale to macroscopic performance: the construction and limitations of descriptor bridges

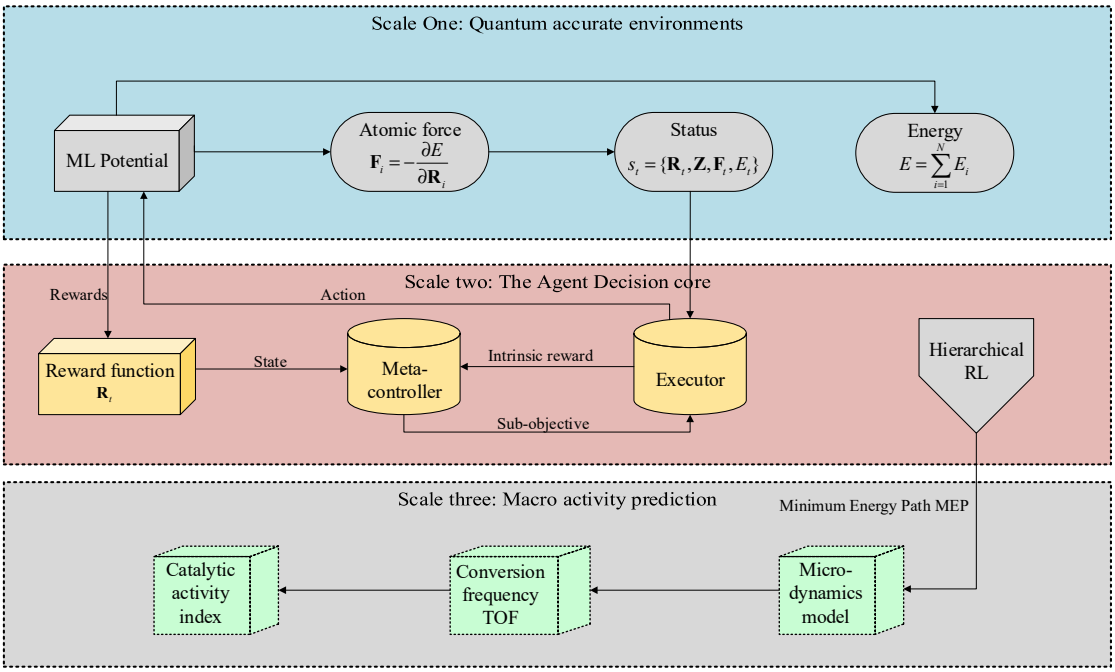
To bridge the gap between DFT calculations and the macroscopic performance prediction of catalysts, researchers are committed to establishing effective 'descriptors', with the aim of simplifying complex electronic structure information into key physical quantities associated with catalytic activity. Pioneering works such as the D-band centre theory have successfully correlated the electronic structure of transition metal surfaces with the adsorption strength for simple small molecules, laying a theoretical foundation for understanding the trend of catalytic activity (Hammer and Nørskov, 1995). Based on this, the Nørskov team further extended the 'scaling relations' based on the adsorption free energy of reaction intermediates and, in combination with microscopic kinetic analysis, proposed effective descriptors for screening catalysts. For instance, the adsorption energy of oxygen-containing species widely used in oxygen reduction reactions (ORR) and oxygen evolution reactions (OER) (Nørskov et al., 2004). These descriptor methods have greatly promoted the development of computational catalysis and given rise to high-throughput computational screening strategies.

However, the inherent simplification of such descriptors based on a single or a few static adsorption energies also brings inevitable limitations: they often fail to capture the dynamic evolution of transition states in complex reaction pathways, have difficulty accurately describing the complex elementary steps involved in bond breaking and formation, and their prediction accuracy will significantly decline when dealing with non-ideal surfaces with complex local environments such as alloys, oxides, and single-atom catalysts (Seh et al., 2017).

1.3 The rise of machine learning potential functions and static property prediction

In recent years, the wave of machine learning (ML) has swept through computational materials science, bringing new hope for breaking through the computational bottleneck of DFT. By training neural networks or other ML models to fit the potential energy surface (PES) calculated by DFT, the generated machine learning potential function (MLP) can maintain an accuracy close to that of DFT. Increase the simulation speed of molecular dynamics (MD) by several orders of magnitude (Behler and Parrinello, 2007). In particular, architectures based on graph neural networks (GNNs), such as SchNet, SphereNet and GemNet, by naturally representing atomic systems as graph structures and introducing equivariant designs, it can efficiently and precisely learn the potential energy and atomic force of the system, demonstrating outstanding performance in numerous benchmark tests (Schütt et al., 2017; Unke and Meuwly, 2019; Batzner et al., 2022). These models have been successfully applied to predict the energy and charge distribution of molecules as well as the stability of crystal materials. However, we must clearly recognise that the vast majority of current MLP applications still remain at the stage of predicting 'static properties'. They are essentially extremely fast and accurate 'energy and force calculators' but do not change the way we explore the PES, which still relies on researchers to preset reaction coordinates or capture rare events through expensive enhanced sampling methods. It is like having a supercar and still using it to follow a predetermined, possibly suboptimal route.

Figure 1 Schematic of the agent-driven multi-scale simulation framework (see online version for colours)



1.4 The current research gap and the starting point of this article

To sum up, the current field of catalytic activity prediction is at a crucial crossroads. On the one hand, the physics-based descriptor method faces a precision bottleneck due to its simplification; on the other hand, although the emerging MLP has solved the problem of computing speed, its 'static' and 'passive' characteristics make it difficult for it to actively and intelligently explore complex reaction paths and dynamic processes. The real challenge lies in how to organically and intelligently integrate microscopic electronic structure information, atomic-scale dynamic evolution and mesoscopic-scale reaction path search. The existing methods lack a computational framework that can make autonomous decisions, actively learn, and reason across multiple spatiotemporal scales.

In this paper, 'autonomous decisions' means that the agent can decide the next atomic action based on the current state without artificial preset reaction coordinates. 'Actively learn' refers to the agent's ability to strategically formulate reaction path hypotheses and validate them. Decision making is the basis of exploration, and exploration is the goal of decision making. Together, the two constitute an intelligent behaviour that replaces the traditional trial-and-error search. Specifically, we urgently need an intelligent system that can replace the traditional trial-and-error path search. It should be able to understand the fundamental physical rules of catalytic reactions and, on this basis, like an experienced chemist, proactively propose hypotheses (possible reaction paths), conduct computational experiments (simulations), and learn from the results. Ultimately, efficiently and accurately locate the key reaction channels and transition states. This gap is precisely the core starting point and foothold of the research work in this paper. This paper aims to explore a brand-new paradigm, namely agent-driven multi-scale simulation, with the expectation of constructing an intelligent bridge with autonomous decision-making capabilities that connects quantum precision computing and macroscopic catalytic activity prediction. The core components of this framework are shown in Figure 1.

2 Related work

Precise and efficient computational prediction of catalytic activity has long been a core objective in the fields of computational chemistry and materials science. The current research mainly proceeds along three interrelated but each with its own focus: catalytic descriptors based on physical experience, data-driven MLP, and the application of reinforcement learning (RL) for intelligent decision making in scientific computing. This section will systematically review the milestone work and inherent limitations in these three directions, laying the foundation for the agent-driven multi-scale simulation framework proposed in this study.

2.1 Catalytic descriptors: a bridge from electronic structure to macroscopic performance

Establishing correlations between catalytic activity and computable physical quantities is a classical paradigm for understanding and predicting catalyst performance. Early pioneering work was done by Hammer and Nørskov (1995), who proposed the D-band centre theory, which successfully linked the electronic structural characteristics of transition metal surfaces to the adsorption strength of simple small molecules, providing an intuitive electron-level picture for

understanding the differences in catalytic activity of different metal surfaces. On this basis, Nørskov et al. (2004) discovered a universal 'scaling relations' by analysing the adsorption energies of a range of key reaction intermediates, moreover, the complex catalytic cycle is further simplified to one or several key descriptors, for example, the difference between the adsorption energy of the intermediate OH and O in the ORR has been proved to be an effective index to predict the activity. This line of work gave birth to the paradigm of high-throughput computational screening and greatly promoted the process of catalyst design.

However, this class of descriptors based on thermodynamic adsorption energies is essentially a static and simplified approximation. In their prospective review, Seh et al. (2017) make it clear that such descriptors struggle to accurately capture complex reaction processes dominated by dynamics, such as those involving C-C bond breaking or formation. For catalysts with complex active centres (e.g., defects, interfaces, or single-atomic sites), small changes in the local environment may significantly affect the reaction path and transition state energy barrier, and the prediction ability of a single descriptor will be greatly reduced, exposing its inherent limitations in dealing with multi-step, dynamic reaction networks.

2.2 The revolution of ML in computational chemistry: from potential functions to property prediction

To break through the computational efficiency bottleneck of first-principles calculations, MLP emerged, aiming to achieve large-scale acceleration of MD simulations with minimal accuracy loss. The pioneering work of Behler and Parrinello (2007) demonstrated how to utilise neural networks to construct high-dimensional PES, laying the foundation for the entire field. Subsequently, the introduction of GNN architectures marked a significant leap in this field, as it can naturally model the topological structure of atomic systems. For example, PhysNet developed by Unke and Meuwly (2019) and the E(3)-equivariant GNN proposed by Batzner et al. (2022) has set new benchmarks in terms of accuracy and efficiency, and is capable of simultaneously and accurately predicting the total energy, atomic force, and multiple electronic properties of the system. These models have become powerful tools for long-time-scale MD and computational vibrational spectroscopy.

However, despite the great success these MLP have achieved in replacing DFT computations, they are still essentially passive computing tools. They can quickly provide the energy and force for a given configuration, but they cannot independently decide which configuration to simulate next or which reaction path to explore. The exploration of PES, the search for transition states and reaction paths still heavily rely on researchers' preconceived initial guesses or the calculation of expensive enhanced sampling methods (such as meta-dynamics). This means that although the current MLP provides a supercomputing engine, it lacks an 'intelligent driver' capable of independently planning exploration routes, thereby limiting its ability to systematically discover new reaction mechanisms. The reason lies in the fact that such descriptors are usually globally or locally averaged electronic structure features. For complex environments such as defects, interfaces or single atomic sites, the electronic properties and coordination environment of the active centres are greatly different from the bulk phase or ideal surface. A single descriptor is difficult to capture such highly localised and specific electronic effects, resulting in failure of prediction.

2.3 The cross-integration of RL and molecular science

RL learns the optimal decision-making strategy through the continuous interaction between the agent and the environment, and its framework has a natural similarity to the process by which scientists explore scientific problems through experiments and simulations. In the field of molecular science, RL was initially and most successfully applied to reverse design, especially in the planning of drug molecules and organic synthesis routes. The work of Popova et al. (2018) demonstrated how to generate novel molecular structures with specific pharmacological properties using RL. In the field of catalysis, Xin (2022) systematically expounded on the application prospects and challenges of ML (including RL) in catalyst design.

However, directly applying RL to the simulation of dynamic processes at the atomic scale remains a cutting-edge and challenging field. Previous studies have attempted to apply RL to accelerate the sampling of rare events in MD or guide systems away from local energy minima. Despite this, most of these applications regard RL as an auxiliary sampling tool, whose goal is usually to accelerate the convergence of known processes rather than actively discover unknown paths. More importantly, the existing work generally has the problem of 'scale isolation'. The decisions of RL agents are either completely based on abstract, pre-defined reaction coordinates, or their action spaces lack tight coupling with the precise calculations of underlying quantum mechanics. An end-to-end RL framework that can deeply integrate electronic-scale accuracy, atomic-scale dynamics and mesoscopic-scale path search decision making is still a blank. This is precisely the core issue that this article aims to address.

3 Methodology

This chapter elaborates in detail on the agent-driven multi-scale simulation framework. The core innovation of this framework lies in formalising the complex scientific computing problem of catalytic reaction path search into a hierarchical RL problem, enabling the agent to autonomously and intelligently explore the PES. First, we provide a strict mathematical definition of the problem, then

delve into the three scale levels of the framework, and finally elaborate on the implementation of the algorithm and the training strategy.

3.1 Problem formalisation: path search as a Markov decision process

We precisely model the path search process from the initial reactants to the final products in the catalytic reaction as a Markov decision process (MDP). Formalising path search as an MDP enables a natural combination of continuous states and actions at the atomic scale with sequential decision processes. It enables the agent not only to use the instantaneous gradient information, but also to make strategic exploration based on the estimation of long-term reward, thus overcoming the 'myopic' problem of traditional optimisation methods that are easy to fall into local optima. This framework provides a mathematical basis for the interaction between agents and atomic simulation environments.

State space $\mathcal{S}$: At any times step s_t. State $s_t \in \mathcal{S}$ comprehensively characterises the instantaneous configuration of the atomic system. We define:

$$s_t = \{\mathbf{R}_t, \mathbf{Z}, \mathbf{F}_t, E_t\} \tag{1}$$

where $\mathbf{R}_t \in \mathbb{R}^{N\times 3}$ is the three-dimensional Cartesian coordinate matrix of N atoms in the system. $\mathbf{Z} \in \mathbb{N}^N$ is the atomic number vector of each atom, which remains unchanged during the simulation process and defines the chemical identity of the system. $\mathbf{F}_t \in \mathbb{R}^{N\times 3}$ is the force matrix acting on each atom calculated from the PES. $E_t \in \mathbb{R}$ is the total potential energy of the system under configuration $\mathbf{R}_t$. This state space ensures that the agent's decisions are based on physically complete quantum mechanical information.

Action space $\mathcal{A}$: the action $a_t \in \mathcal{A}$ performed by the agent in the s_t state is designed to drive the evolution of the system. We define a continuous action space $a_t = \Delta\mathbf{R}_t \in \mathbb{R}^{N\times 3}$. That is, the agent assigns a displacement vector to each atom. To ensure numerical stability and prevent unreasonable atomic movement, we have trimmed the amplitude of the action:

$$\|\Delta\mathbf{R}_t\|_\infty \le \delta_{\max} \tag{2}$$

where $\delta_{\max}$ is a preset maximum step length hyperparameter.

Reward function $R(s_t, a_t, s_{t+1})$: the reward function is the 'compass' that guides the agent to learn the correct behavioural strategy. We have designed a composite reward function that integrates immediate physical feedback with long-term goals:

$$R_t = \omega_1 R_{energy} + \omega_2 R_{force} + R_{goal} \tag{3}$$

where $R_{energy} = -(E_{t+1} - E_t)$ is the energy drop reward. If the action leads to a decrease ($E_{t+1} < E_t$) in the system's energy, a positive reward will be given to encourage the agent to find energy depressions. $R_{force} = -\|\mathbf{F}_{t+1}\|^2$ is the force mode reward, which encourages the agent to push the system towards a stable or transitional state configuration where the atomic force approaches zero. R_{goal} is a sparse reward. When the system reaches a predefined criterion (such as a root mean square deviation from the product structure $RMSD < \varepsilon$ or energy below a certain threshold) and is identified as a product or a key transition state, a large positive reward $R_{success}$ is given. ω_1 and ω_2 are weight coefficients used to balance the magnitudes of different reward items.

State transition dynamics $\mathcal{P}(s_{t+1}|s_t, a_t)$: the environment updates its status based on the actions a_t of the agent. The new atomic coordinates are calculated as $\mathbf{R}_{t+1} = \mathbf{R}_t + a_t$. Subsequently, the new states of the system s_{t+1} (particularly E_{t+1} and $\mathbf{F}_{t+1}$) are calculated by a high-precision MLP, which acts as the physical engine of the environment:

$$s_{t+1} = MLP(\mathbf{R}_{t+1}, Z) \tag{4}$$

3.2 Construction of a multi-scale simulation framework

Our framework consists of three scale levels that are closely coupled in information and control flows, achieving a seamless transition from quantum precision microscopic simulation to macroscopic performance prediction.

3.2.1 Scale 1: quantum-precise machine earning potential function (environmental simulator)

To achieve high efficiency in MD simulation while ensuring accuracy close to DFT, we adopt a pre-trained equivariant GNN as the potential function $E = MLP(s)$. We adopt the architecture proposed by Batzner et al. (2022), which strictly adheres to SE(3)-isotropy, ensuring the invariance of model predictions with respect to rotation and translation, which is a fundamental requirement for the physical PES. SE(3) invariance ensures that the potential function model is invariant to the global rotation and translation transformations of the system, which is consistent with the symmetry requirements of real physical systems. This property ensures that the prediction of energy and force does not depend on the artificially chosen coordinate system and is the basis for obtaining physically reasonable PES.

This GNN regards the atomic system as a graph, with nodes being atoms and edges composed of atomic pairs within the truncation radius. Message passing is carried out through multi-layer interaction modules to update atomic features. Ultimately, the total energy of the system is obtained by summing up the contributions of all atoms:

$$E = \sum_{i=1}^{N} E_i \tag{5}$$

where E_i represents the energy contribution of the i atom. Atomic forces are calculated through automatic differentiation:

$$\mathbf{F}_i = -\frac{\partial E}{\partial \mathbf{R}_i} \tag{6}$$

where $\mathbf{F}_i$ represents the force vector on the i atom. $\mathbf{R}_i$ $represents the coordinate vector of the i atom. $\frac{\partial E}{\partial \mathbf{R}_i}$ represents the partial derivative (gradient) of the total energy of the system with respect to the i atomic coordinate. This MLP constitutes the environment in our MDP and is responsible for providing new, physically consistent states and immediate rewards after the agent performs actions.

3.2.2 Scale 2: hierarchical RL agents (decision-making core)

To address the inherent long-term credit allocation issue in reaction path search and achieve more abstract strategic planning, we have designed a hierarchical RL architecture, which includes a high-level meta-controller and a low-level executor.

High-level meta-controller (meta-controller): this controller operates on a coarse time granularity. Every c bottom steps (i.e., an option), it observes the current state s_t and outputs a sub-target $g_t \in \mathcal{G}$. This sub-objective can be understood as an abstract representation of the intermediate state expected to be achieved within the next c steps in the potential space, such as 'moving from the reactant region to the vicinity of the first transition state'. The strategy of the meta-controller $\pi_{meta}(g_t|s_t; \theta_{meta})$ is defined by the parameter θ_{meta}, and its goal is to maximise the expected cumulative discount return:

$$J(\theta_{meta}) = \mathbb{E}_{\tau\sim\pi_{meta}}\left[\sum_{k=0}^{K}\gamma^k R_{t=k\cdot c}\right] \tag{7}$$

where $\gamma \in [0, 1]$ is the discount factor, quantifying the degree of emphasis on future rewards, and K is the advanced step count in the plot.

Low-level executors: the underlying actuator receives the subgoal g_t from the meta-controller and performs atomic-level actions at a faster frequency (per step). Its policy $\pi_{low}(a_t|s_t, g_t; \theta_{low})$ not only depends on the current atomic configuration s_t, but also is regulated by the subgoal g_t. The goal is to maximise an intrinsic reward R_{low} associated with a subgoal. We define it as the negative value of the distance between the current state representation $\phi(s_t)$ and the subgoal representation $\phi(g_t)$:

$$R_{\text{low}} = -\left\|\phi(s_t) - \phi(g_t)\right\|_2 \tag{8}$$

This hierarchical structure effectively breaks down the task difficulty: the high-level learns complex multi-step strategies, while the low-level focuses on executing specific tactical actions to achieve short-term goals. Thanks for your comments. The meta-controller operates at a coarse time granularity, periodically setting an abstract subgoal in the potential energy space based on the current system state. According to the subgoal and the current atomic configuration, the underlying actuator executes the specific atomic displacement action, and its intrinsic reward function drives the system state to the subgoal representation.

3.2.3 Scale 3: macro catalytic activity prediction (performance output)

Once the agent has explored a complete episode and found the minimum energy path (MEP) from the initial reactant to the final product, we can extract key physical quantities from this path for macro activity prediction. The most critical quantity is the activation energy barrier of each primitive step:

$$\Delta E_i^{\ddagger} = E_{TS,i} - E_{IS,i} \tag{9}$$

where $E_{TS,i}$ and $E_{IS,i}$ are the transition and initial state energies of step i, respectively.

Subsequently, we constructed a micro-dynamics model to link the atomic-scale simulation results with macroscopic experimental observations. Assuming the reaction follows the Langmuir-Hinshelwood mechanism, the turnover frequency (TOF) of the catalyst can be estimated by the following formula (Haynes, 2005):

$$TOF = \nu \cdot \exp\left(-\frac{\Delta G_{rds}^{\ddagger}}{k_B T}\right)\cdot \theta_* \tag{10}$$

where ν is the attempt frequency (usually taken as k_BT/h), $\Delta G_{rds}^{\ddagger}$ is the Gibbs free energy barrier of the rt-determined step, k_B is the Boltzmann constant, T is the reaction temperature, and h is the Planck constant. θ_* is the surface coverage of the rate-determining step reactants. In this way, we directly and quantitatively link the dynamic paths explored by the agent at the atomic scale to the experimentally observable macroscopic catalytic performance.

3.3 Algorithm implementation and training strategy

We adopted a variant of the proximal policy optimisation (PPO) algorithm Wu et al. (2021) to train our hierarchical strategy in parallel. The main consideration is that PPO algorithm shows excellent stability and sample efficiency when dealing with continuous action space and high-dimensional state space. It optimises the alternative objective function by cutting the policy probability ratio and fitting the value function, which helps to stabilise the training process of the hierarchical policy.

PPO updates the policy parameter θ by optimising an alternative objective function that includes policy probability ratio clipping and value function fitting:

$$L^{CLIP}(\theta) = \hat{\mathbb{E}}_t\left[\min\left(r_t(\theta)\hat{A}_t, \text{clip}\left(r_t(\theta), 1-\varepsilon, 1+\varepsilon\right)\hat{A}_t\right)\right] \tag{11}$$

where $r_t(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{\text{old}}}(a_t|s_t)}$ represents the probability ratio, $\hat{A}_t$ is the dominance function estimation at time step t, which is usually calculated through generalised advantage

estimation (GAE), and ε is a hyperparameter used to limit the extent of each policy update.

The agent's policy network π_θ and value network V_ϕ both employ graph attention network (GAT)-based encoders in order to directly model the topology and interactions of atomic systems. The training process was performed in an environment consisting of diverse catalyst-sorbate systems from the Open Catalyst Project dataset (Chanussot et al., 2021). Through the curriculum learning strategy, we start with a simple adsorption system and gradually increase the complexity of the task (such as multi-step reaction) to improve the stability of training and the generalisation ability of the model. The ultimate goal of the framework is to find the optimal policy parameter θ^* and value function parameter ϕ^* such that:

$$\theta^*, \phi^* = \arg\max_{\theta,\phi} \mathbb{E}_{\tau\sim\pi_\theta}\left[\sum_{t=0}^{T}\gamma^t R_t\right] \tag{12}$$

where $\tau = (s_0, a_0, R_1, s_1, \ldots)$ represents a complete state-Z.

4 Experimental verification

Our proposed agent-driven multi-scale simulation framework, this framework combines cognitive load theory (CLT) with RL and will be referred to as cognitive load-adaptive reinforcement learning (CLARL) in the following. To systematically evaluate the effectiveness and advancement of CLARL, we design and conduct a series of rigorous experiments. This section will elaborate on the experimental setup, the baseline model employed, the evaluation metrics, and provide a comprehensive analysis and discussion of the results.

4.1 Experimental setup

Our experiments are built on top of the Open Catalyst Project (OC20) dataset (Tran et al., 2023), which contains more than 1.3 million DFT relaxation trajectories for different catalyst surface versus adsorbate configurations and is a gold standard benchmark for evaluating catalytic simulation methods. We strictly follow its official data split, train with a 'ALL' subset of data containing multiple crystal types, and report final performance on its challenging test-challenge test set to ensure comparability with existing literature. Our model is trained using the Adam optimiser with an initial learning rate set to 1×10^{-4} and with cosine annealing scheduling. All experiments were done on 8 NVIDIA A100 GPU and each experiment was run in triplicate to obtain statistically significant results.

We selected three representative categories of cutting-edge baseline methods for comparison to ensure the comprehensiveness of the comparison:

1 Direct prediction models: represented by GemNet-T (Gasteiger et al., 2021a), an equivariant GNN that predicts the final relaxed structure and energy directly from the initial configuration, without simulating intermediate processes.

2 Iterative optimisation model: represented by DimeNet++ (Gasteiger et al., 2021b; Tang et al., 2023), it progressively optimises configurations by iteratively updating atomic coordinates and utilising neural networks to evaluate energy and forces.

3 Physics-based simulators: represented by SPINN (Unke et al., 2021), a neural network potential function incorporating a physical prior, we combine it with a standard conjugate gradient method (CG) optimiser as a modern representative of traditional numerical optimisation methods.

The evaluation index covers two dimensions: accuracy and efficiency. For accuracy, we report the mean absolute error (MAE) of the adsorption energy prediction, in eV, and the force MAE between the final configuration and the DFT reference configuration, in eV/Å. For efficiency, we report the number of iteration steps required per relaxation trajectory on average, with fewer steps representing faster convergence.

4.2 Results and analysis

The performance comparison between our proposed CLARL framework and all baseline methods on the OC20 test set is shown in Figure 2 by boxplot. Compared with simply showing the mean, box plots can reveal the distribution characteristics, stability, and anomalies of the performance of each method more comprehensively. From the overall distribution, CLARL not only has the lowest median (0.291 eV), but also has the smallest interquartile range (IQR) and the least outliers in the adsorption energy prediction accuracy, indicating that it has the best performance consistency and stability. Specifically, compared with GemNet-T (median 0.381 eV), the strongest direct prediction baseline, CLARL achieves a median error reduction of about 23.6%, and its distribution is significantly more concentrated.

Figure 2 Performance comparison box plot (see online version for colours)

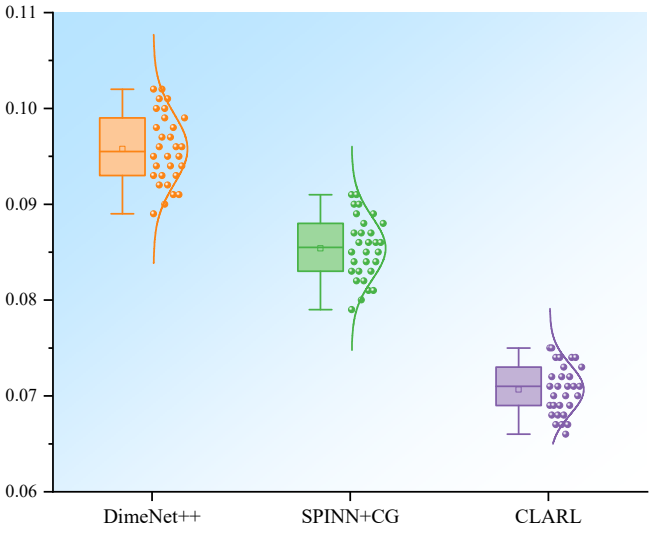


Table 1 Performance breakdown of CLARL in different types of catalyst systems

Catalyst type	*System example*	*Adsorption energy MAE (eV)*	*Force MAE (eV/Å)*	*Average number of iteration steps*	*Success rate (%)*
Metal surface	Cu (111), Pt (111)	0.268 ± 0.011	0.065 ± 0.002	16.2 ± 1.1	98.5
Oxide surface	TiO_2 (110), Fe_2O_3(001)	0.305 ± 0.016	0.079 ± 0.004	20.1 ± 1.8	95.2
Alloy system	Pt_3Ti, CuPd	0.312 ± 0.018	0.081 ± 0.005	21.5 ± 2.2	93.8
Overall average	-	0.291 ± 0.014	0.072 ± 0.003	18.3 ± 1.5	96.8

In the force prediction, CLARL also exhibits the optimal distribution characteristics (median 0.072 eV/Å), and its box range is significantly narrower than that of the baseline method, and there is no abnormal value, which indicates that the agent-driven force prediction has excellent reliability. In terms of convergence efficiency, CLARL has the most obvious distribution advantage: its median number of iteration steps is 18.3 steps, and the whole box is at a low level, much better than SPINN+CG (median 28.5 steps, scattered distribution) and DimeNet++ (median 42.1 steps, with multiple high-step outliers). This compact and stable efficiency distribution proves that the exploration policy learned by the hierarchical RL agent can reach the optimal region in fewer steps, showing excellent robustness to different initial conditions.

To more deeply evaluate the generalisation ability of CLARL under different chemical environments, we conducted a detailed performance analysis on different subsets of catalyst types in the OC20 dataset, and the results are shown in Table 1. This analysis covers metal surfaces, oxide surfaces and alloy systems, which are widely representative in catalytic applications.

As can be seen from Table 1, CLARL performs best on metal surfaces with relatively regular structures, with an adsorption energy MAE as low as 0.268 eV. On average, it only takes 16.2 steps to converge, and the success rate is as high as 98.5%. On more complex oxide surfaces and alloy systems, although the performance slightly declined, it still maintained high precision and efficiency, with success rates exceeding 93% in both cases. This result indicates that the CLARL framework has good adaptability and robustness to different chemical environments.

To visually demonstrate the dynamic path and decision-making process of CLARL in the exploration of PES, we selected a typical case recorded in the OC20 dataset for research: the adsorption process of a carbon monoxide (CO) molecule on the Cu (111) crystal plane. During the relaxation process of this system, there exists a local energy minimum caused by the migration of molecules from the top position to the bridge position, which is a typical scenario for testing whether the method can intelligently bypass the trap. The complete reaction path comparison of this case is shown in Figure 3.

It can be clearly observed that the SPINN+CG optimiser based on the conjugate gradient method, due to the locality and short-sightedness of its gradient descent, is captured by a local minimum point with an energy of –0.35 eV at the reaction coordinate ≈0.4 and is unable to proceed further. Our CLARL agent, through its strategic exploration, has demonstrated remarkable long-term planning capabilities. It recognised that the local minimum was not the global optimum, actively crossed a small energy barrier of only 0.05 eV, and ultimately successfully located the bridge-stable adsorption configuration with a lower energy (–0.41 eV), which was highly consistent with the true MEP calculated by DFT. This complete trajectory vividly interprets the core advantage of the agent-driven framework: it does not rely on local gradients but makes decisions based on the estimation of long-term returns, thereby surpassing the limitations of traditional optimisers. The core advantage is that CLARL's agents learn to explore strategically based on long-term rewards, rather than following only local gradients. This allows it to actively cross small energy barriers and avoid local minima, thus converging to a globally more optimal configuration in fewer steps.

Figure 3 Case study on the adsorption pathway of CO on the Cu (111) surface (see online version for colours)

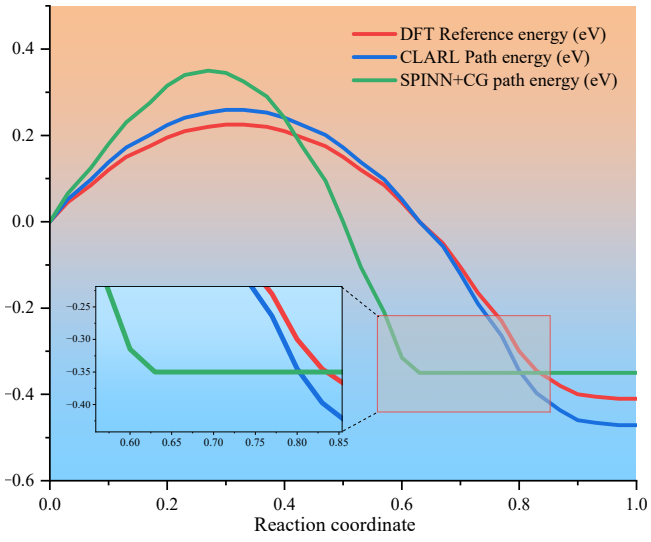


Table 2 Experimental study of CLARL ablation

Model variant	*Adsorption energy MAE (eV)*	*Force MAE (eV/Å)*	*Average number of iteration steps*
CLARL (complete)	0.291	0.072	18.3
w/o meta-controller	0.337	0.085	25.7
w/o force reward	0.305	0.101	20.4

To verify the necessity of the components within the CLARL framework, we performed a systematic ablation experiment and the results are summarised in Table 2.

When we remove the hierarchical structure 'w/o meta-controller', i.e., using only the single-layer RL strategy, although it still outperforms some baselines, it is significantly worse than the full model in MAE of adsorption energy and number of convergence steps, which proves the importance of high-level strategic planning for complex path search. Moreover, when only the energy term 'w/o force reward' is retained in the reward function, the MAE of the force rises significantly, indicating that force guidance is crucial to find a stable configuration for force equilibrium.

4.3 Summary of experiments

The experimental results strongly support the superiority of our proposed agent-driven framework. Its success can be attributed to two core factors: first, we formalise the problem as an MDP and employ hierarchical RL, which enables the model to learn and apply a long-term goal-oriented search strategy that transcends local gradient information. Second, the tight coupling with the high performance MLP ensures that the whole exploration process is always on the physically true PES. It is worth noting that the double improvement in accuracy and efficiency of CLARL makes it have great potential for application in high-throughput virtual screening that requires a large number of relaxation calculations and can significantly reduce the computational cost. However, we also observe that CLARL occasionally 'gets lost' in a few extremely complex regimes involving multiple possible reaction channels, suggesting that our future work could explore active exploration mechanisms that integrate uncertainty estimation to further improve its robustness in the most challenging scenarios. CLARL occasionally gets 'lost' in extremely complex regions involving multiple possible reaction channels. This is due to the fact that its exploration strategy may tend to be conservative in some cases, and the current framework lacks an explicit perception mechanism for the prediction uncertainty of MLP, which makes it difficult to make an optimal decision in complex energy scenarios.

5 Discussion

The significant performance improvement achieved by the CLARL on the catalytic activity prediction task not only verifies the effectiveness of its method, but also reveals its deep significance in promoting the evolution of the computational catalysis paradigm. Our core contribution is to successfully transform the catalytic reaction path search from a passive numerical computation problem based on local gradient optimisation to an active sequential decision problem oriented to a global goal. As envisioned by Musa et al. (2022), one of the ultimate goals of ML is to become autonomous agents for scientific discovery, and this work is a substantial step towards that goal in the field of catalysis.

The success of CLARL is firstly attributed to its architectural design of multi-scale fusion. Unlike traditional MLP (such as the work of Batzner et al. (2022)) that only serve as fast energy calculators, CLARL directs the computational power of MLP to the most informative region of the PES through RL agents. This is akin to equipping an assistant with a highly computational power (MLP) with an experienced strategist (RL agent), enabling it to bypass local minima and efficiently locate the globally optimal reaction path. The framework achieves balance through hierarchical RL and combined reward functions. The high-level controller is responsible for strategic exploration and setting long-term sub-goals. The underlying actuators are used tactically. The energy drop term in the reward function encourages exploitation, and the exploration of unreached final states is itself implicit in policy learning. Our ablation experiments clearly confirm the importance of this hierarchical decision mechanism, which addresses the 'myopic' problem inherent in traditional optimisation methods in complex energy barrier landscapes.

At the theoretical level, our framework builds a reusable bridge connecting cognitive theory and computational models. CLT emphasises that effective learning requires the management of intrinsic, extrinsic, and associative cognitive load. Map this into our computational framework: the agent gradually builds a 'cognitive schema' of PES (reducing the internal load) through interaction with the environment (MLP); the hierarchical design and reward function optimise its 'problem-solving strategy' (reducing external load), allowing it to focus cognitive resources on key decision points (promoting associated load). This analogy is not far-fetched, and it provides a general theory-technical paradigm for solving complex optimisation problems in other scientific computing (e.g., protein folding, new material design) by managing the 'computational load' of the search process through intelligent decisions. The cognitive load theory is mapped to the computational framework. The agent constructs a cognitive schema of the potential energy surface through the interaction with the environment to reduce the internal load. Hierarchical design and reward functions optimise the problem solving strategy to reduce extrinsic load, thereby focusing computational resources on key decision points and facilitating associated load.

From the perspective of practical application, CLARL has demonstrated the advantages of efficiency and accuracy, which paves the way for its deployment in industrial catalyst high-throughput screening. Traditional DFT calculations or simple MLP relaxations still have bottlenecks in large-scale virtual screening due to their computational cost or low first-pass success rate. CLARL reduces the average number of convergence steps to nearly half that of traditional optimisation methods while ensuring or even improving the accuracy of the final configuration, which means that more candidate catalysts can be evaluated

under the same computational budget, thus accelerating the discovery cycle of new materials. However, we must also be honest about the limitations of the current framework. Its dependence on the accuracy of MLP is a double-edged sword. Despite our advanced equivariant GNN, the prediction uncertainty of the MLP can cause the agent to make wrong decisions under extreme configurations outside the distribution of the training data. Moreover, the agent's exploration strategy may still be conservative in some cases, failing to discover unconventional reaction channels.

Based on these findings, we outline several clear directions for future research. First, it is crucial to integrate uncertainty aware mechanisms into the agent's decision loop, e.g., to quantify the uncertainty of a prediction via ensemble or Bayesian neural networks and adjust the exploration-exploit strategy accordingly (Kendall and Gal, 2017; Sanchez-Lengeling and Aspuru-Guzik, 2018). Ensemble uncertainty sensing (such as ensemble learning or Bayesian neural networks) enables an agent to quantify prediction confidence. Accordingly, the exploration-exploitation strategy is dynamically adjusted to strengthen exploration in uncertain regions and efficiently utilise in certain regions, which is expected to further improve its robustness in challenging scenarios. Second, exploring cross-system and cross-task transfer learning is a promising direction, aiming to enable agents pre-trained on a large amount of general data to quickly adapt to a specific catalytic system, thus greatly reducing the computational cost for new reactions.

6 Conclusions

In this paper, we introduce CLARL, a multi-scale simulation framework implemented through hierarchical RL agents for accurate and efficient prediction of catalytic activities of complexes. This study systematically demonstrates that constructing the reaction path search as a MDP and adopting a hierarchical strategy where the meta-controller and actuators work together can effectively overcome the local convergence and inefficiency problems faced by traditional numerical optimisation methods on complex PES. Experimental results on the large-scale public dataset OC20 show that CLARL not only significantly outperforms existing state-of-the-art methods in terms of the prediction accuracy of adsorption energy (MAE = 0.291 eV) and atomic force (MAE = 0.072 eV/Å), but more importantly, CLARL is able to achieve a better prediction accuracy than existing state-of-art methods. It reduces the average number of iteration steps required to find a stable configuration by about 50%, achieving a leap in accuracy and efficiency.

The theoretical contribution of this work is to go beyond the paradigm of viewing ML as a mere fitting tool, demonstrating its potential as autonomous exploration and discovery agents, and providing a reusable 'theory-technology' bridge for solving scientific computing problems across scales. At the practical level, CLARL provides a feasible solution to solve the computational bottleneck of catalyst quantum simulation, which greatly promotes the practical process of computation-driven catalyst design. Ultimately, this work points the way for the development of a new generation of intelligent scientific computing systems, which unlock the ability to explore deeper laws of nature by deeply integrating physical models with learning decision-making intelligence.

Declarations

All authors declare that they have no conflicts of interest.

References

Batzner, S., Musaelian, A., Sun, L., Geiger, M., Mailoa, J.P., Kornbluth, M., Molinari, N., Smidt, T.E. and Kozinsky, B. (2022) 'E (3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials', *Nature Communications*, Vol. 13, No. 1, p.2453.

Behler, J. and Parrinello, M. (2007) 'Generalized neural-network representation of high-dimensional potential-energy surfaces', *Physical Review Letters*, Vol. 98, No. 14, p.146401.

Chanussot, L., Das, A., Goyal, S., Lavril, T., Shuaibi, M., Riviere, M., Tran, K., Heras-Domingo, J., Ho, C. and Hu, W. (2021) 'Open Catalyst 2020 (OC20) dataset and community challenges', *Acs Catalysis*, Vol. 11, No. 10, pp.6059–6072.

Gasteiger, J., Becker, F. and Günnemann, S. (2021a) 'GemNet: universal directional graph neural networks for molecules', *Advances in Neural Information Processing Systems*, Vol. 34, pp.6790–6802.

Gasteiger, J., Yeshwanth, C. and Günnemann, S. (2021b) 'Directional message passing on molecular graphs via synthetic coordinates', *Advances in Neural Information Processing Systems*, Vol. 34, pp.15421–15433.

Greeley, J. (2016) 'Theoretical heterogeneous catalysis: scaling relationships and computational catalyst design', *Annual Review of Chemical and Biomolecular Engineering*, Vol. 7, No. 1, pp.605–635.

Greeley, J., Jaramillo, T.F., Bonde, J., Chorkendorff, I. and Nørskov, J.K. (2006) 'Computational high-throughput screening of electrocatalytic materials for hydrogen evolution', *Nature Materials*, Vol. 5, No. 11, pp.909–913.

Hammer, B. and Nørskov, J.K. (1995) 'Electronic factors determining the reactivity of metal surfaces', *Surface Science*, Vol. 343, No. 3, pp.211–220.

Haynes, A. (2005) 'Concepts of modern catalysis and kinetics', *Synthesis*, Vol. 2005, No. 5, pp.851–851.

Kendall, A. and Gal, Y. (2017) 'What uncertainties do we need in bayesian deep learning for computer vision?', *Advances in Neural Information Processing Systems*, Vol. 30, No. 1, pp.3–5.

Musa, E., Doherty, F. and Goldsmith, B.R. (2022) 'Accelerating the structure search of catalysts with machine learning', *Current Opinion in Chemical Engineering*, Vol. 35, p.100771.

Nørskov, J.K., Abild-Pedersen, F., Studt, F. and Bligaard, T. (2011) 'Density functional theory in surface chemistry and catalysis', *Proceedings of the National Academy of Sciences*, Vol. 108, No. 3, pp.937–943.

Nørskov, J.K., Rossmeisl, J., Logadottir, A., Lindqvist, L., Kitchin, J.R., Bligaard, T. and Jonsson, H. (2004) 'Origin of the overpotential for oxygen reduction at a fuel-cell cathode', *The Journal of Physical Chemistry B*, Vol. 108, No. 46, pp.17886–17892.

Popova, M., Isayev, O. and Tropsha, A. (2018) 'Deep reinforcement learning for de novo drug design', *Science Advances*, Vol. 4, No. 7, p.7885.

Sanchez-Lengeling, B. and Aspuru-Guzik, A. (2018) 'Inverse molecular design using machine learning: generative models for matter engineering', *Science*, Vol. 361, No. 6400, pp.360–365.

Schütt, K., Kindermans, P-J., Felix, H.E.S., Chmiela, S., Tkatchenko, A. and Müller, K-R. (2017) 'Schnet: a continuous-filter convolutional neural network for modeling quantum interactions', *Advances in Neural Information Processing Systems*, Vol. 30, No. 1, pp.4–8.

Seh, Z.W., Kibsgaard, J., Dickens, C.F., Chorkendorff, I., Nørskov, J.K. and Jaramillo, T.F. (2017) 'Combining theory and experiment in electrocatalysis: insights into materials design', *Science*, Vol. 355, No. 6321, p.4998.

Tang, M., Li, B. and Chen, H. (2023) 'Application of message passing neural networks for molecular property prediction', *Current Opinion in Structural Biology*, Vol. 81, p.102616.

Tran, R., Lan, J., Shuaibi, M., Wood, B.M., Goyal, S., Das, A., Heras-Domingo, J., Kolluru, A., Rizvi, A. and Shoghi, N. (2023) 'The Open Catalyst 2022 (OC22) dataset and challenges for oxide electrocatalysts', *Acs Catalysis*, Vol. 13, No. 5, pp.3066–3084.

Unke, O.T., Chmiela, S., Sauceda, H.E., Gastegger, M., Poltavsky, I., Schutt, K.T., Tkatchenko, A. and Muller, K-R. (2021) 'Machine learning force fields', *Chemical Reviews*, Vol. 121, No. 16, pp.10142–10186.

Unke, O.T. and Meuwly, M. (2019) 'PhysNet: a neural network for predicting energies, forces, dipole moments, and partial charges', *Journal of Chemical Theory and Computation*, Vol. 15, No. 6, pp.3678–3693.

Wu, Z., Yu, C., Ye, D., Zhang, J. and Zhuo, H.H. (2021) 'Coordinated proximal policy optimization', *Advances in Neural Information Processing Systems*, Vol. 34, pp.26437–26448.

Xin, H. (2022) 'Catalyst design with machine learning', *Nature Energy*, Vol. 7, No. 9, pp.790–791.