

PAPER • OPEN ACCESS

## Deep reinforcement learning for predicting kinetic pathways to surface reconstruction in a ternary alloy

To cite this article: Junwoong Yoon *et al* 2021 *Mach. Learn.: Sci. Technol.* **2** 045018

View the [article online](#) for updates and enhancements.



## PAPER

## OPEN ACCESS

RECEIVED  
15 June 2021REVISED  
18 July 2021ACCEPTED FOR PUBLICATION  
29 July 2021PUBLISHED  
27 August 2021

Original Content from  
this work may be used  
under the terms of the  
[Creative Commons  
Attribution 4.0 licence](#).

Any further distribution  
of this work must  
maintain attribution to  
the author(s) and the title  
of the work, journal  
citation and DOI.



# Deep reinforcement learning for predicting kinetic pathways to surface reconstruction in a ternary alloy

Junwoong Yoon<sup>1,4</sup>, Zhonglin Cao<sup>2,4</sup> , Rajesh K Raju<sup>1,4</sup> , Yuyang Wang<sup>2</sup> , Robert Burnley<sup>1</sup> , Andrew J Gellman<sup>1,3</sup> , Amir Barati Farimani<sup>1,2,\*</sup> and Zachary W Ulissi<sup>1,\*</sup>

<sup>1</sup> Chemical Engineering Department, Carnegie Mellon University, Pittsburgh, PA 15217, United States of America

<sup>2</sup> Mechanical Engineering Department, Carnegie Mellon University, Pittsburgh, PA 15213, United States of America

<sup>3</sup> W.E. Scott Institute for Energy Innovation, Carnegie Mellon University, Pittsburgh, PA 15213, United States of America

<sup>4</sup> These authors contributed equally to this work.

\* Authors to whom any correspondence should be addressed.

E-mail: [barati@cmu.edu](mailto:barati@cmu.edu) and [zulissi@andrew.cmu.edu](mailto:zulissi@andrew.cmu.edu)

**Keywords:** catalysis, reinforcement learning, machine learning, reconstruction, segregation

Supplementary material for this article is available [online](#)

## Abstract

The majority of computational catalyst design focuses on the screening of material components and alloy composition to optimize selectivity and activity for a given reaction. However, predicting the metastability of the alloy catalyst surface at realistic operating conditions requires an extensive sampling of possible surface reconstructions and their associated kinetic pathways. We present CatGym, a deep reinforcement learning (DRL) environment for predicting the thermal surface reconstruction pathways and their associated kinetic barriers in crystalline solids under reaction conditions. The DRL agent iteratively changes the positions of atoms in the near-surface region to generate kinetic pathways to accessible local minima involving changes in the surface compositions. We showcase our agent by predicting the surface reconstruction pathways of a ternary  $\text{Ni}_3\text{Pd}_3\text{Au}_2(111)$  alloy catalyst. Our results show that the DRL agent can not only explore more diverse surface compositions than the conventional minima hopping method, but also generate the kinetic surface reconstruction pathways. We further demonstrate that the kinetic pathway to a global minimum energy surface composition and its associated transition state predicted by our agent is in good agreement with the minimum energy path predicted by nudged elastic band calculations.

## 1. Introduction

The performance of a heterogeneous catalyst depends on the catalyst's surface composition and structure. The discovery of novel robust catalysts for a given reaction is often achieved via surface engineering of existing catalysts. Significant attention has been drawn towards the design and development of alloy catalysts, as the synergistic effects of alloying two or more metals can provide catalytic activity, selectivity, and stability superior to their pure component counterparts [1–3]. Furthermore, alloying noble metal catalysts (Pt, Pd, Ag, Au, etc) with low cost, highly abundant metals (Ni, Cu, Sn, Co, etc) can function to reduce the catalyst cost in scaling up industrial level operations. The design of these catalysts is complicated by dynamic transformations of the multi-metallic atomic environment. Often, the reconstruction of a catalyst's surface causes the real surface structure to differ from that predicted simply by cleaving the bulk crystals along a given plane. In alloy catalysts, segregation can result in the surface enrichment or depletion of one or more components in an effort to minimize the surface free energy and reduce lattice strain. Lateral rearrangements can cause a change in the surface layer periodicity. The simultaneous compositional evolution and lateral rearrangement of an alloy catalyst's surface play a critical role in determining the performance of a catalyst.

Predictions of the surface composition of alloy catalysts commonly employ thermodynamic arguments, however, kinetic information is a prerequisite for predicting metastable states that have no

kinetically-feasible nearby local minima due to the high kinetic barriers or high transition state energies. Under reaction conditions, changes in temperature or pressure can cause a change in the stability of the initial thermodynamic equilibrium. The ability of a catalyst to realize the new equilibrium or new local minimum surface composition under these conditions is dictated by the barriers along the kinetic reconstruction pathways. A study of the pathway that describes the evolution of the multi-metallic catalyst surface at reaction conditions is thus necessary to understand the catalytic mechanism at the atomic scale as well as to tune the catalyst's activity and selectivity by controlling the surface reconstruction processes.

The structural transformation of catalyst surfaces has been widely studied from *in-situ* spectroscopic techniques such as *in-situ* x-ray diffraction (XRD), *in-situ* x-ray absorption spectroscopy (XAS), *in-situ* x-ray photoelectron spectroscopy (XPS) and *in-situ* infrared spectroscopy [4–11]. Other experimental techniques include *in-situ* scanning probe microscopy (SPM), *in-situ* scanning tunneling microscopy (STM), *in-situ* atomic force microscopy (AFM), and transmission electron microscopy (TEM) [12–20]. Due to the prohibitive size of the design space, physical limitations, and experimental costs many studies have focused on few discrete temperatures and alloy compositions, limiting the size of the design space explorable by these techniques.

Rapid growth in the design of high performance supercomputing resources coupled with the development of modern computational methodologies primarily based on density functional theory (DFT) now complement the traditional trial and error experimental approaches to the discovery of novel materials for various catalytic applications. Computational approaches can accelerate the catalyst discovery efforts by screening component and composition spaces to eliminate material candidates with low activity and selectivity before performing experimental measurements. However, the search for both stable and high-efficient multi-metallic catalysts for a specific reaction through computational catalytic design is still a Herculean task dramatically increasing in complexity and computational cost with the number of alloy components. Moreover, material screening calculations relying on equilibrium properties ignore the possibility of surface reconstruction. The prediction of a specific surface reconstruction is a complicated process involving many structural evaluations along a hypothesized reconstruction pathway. Such an exercise may require hundreds to thousands of cpu-hours to evaluate only a single pathway of all possible surface reconstructions. Therefore, a robust exploration of the reconstruction pathways requires a technique capable of creatively generating reconstruction pathways because most of the current computational efforts in heterogeneous catalysis avoid the comprehensive exploration of reconstruction pathways required to identify metastable states.

With recent advances in machine learning (ML), researchers have successfully applied ML techniques in the prediction of materials properties [21–25], relaxed structures [26], and alloy properties [27–29], but have yet to predict phenomena on the timescales of surface reconstruction pathways needed to identify metastable states. This is still an extremely challenging problem for the supervised ML models that are commonly used for the direct prediction of the structure/property relationships as the identification of the metastable structures require not only the property labels, but also the information about various kinetic pathways that lead to the changes in the alloy catalyst structure and properties. ML has also established atomistic force field methods using a potential energy surface (PES) fitted to a set of DFT samples in the materials search space [30–35]. However, the residual force errors in the fitted force field often result in structures differing from the true equilibrium states, or worse, lead to non-physical configurations. Furthermore, generating a huge amount of DFT samples throughout the energy pathways for accurate force field predictions while considering all possible catalyst surface reconstructions at a specific reaction condition is infeasible. Despite the considerable excitement about these ML methods, identifying metastable catalyst surfaces still remains untapped due to its highly complex design space.

To deal with the complexity issue, an alternative approach called deep reinforcement learning (DRL) that is capable of traversing the compositional space more efficiently has been employed. Reinforcement learning (RL) is a subfield of machine learning where a decision maker or an agent learns strategies to solve an optimization problem by iteratively interacting with an environment. DRL involves applying deep neural network in RL frameworks to generalize complex problems. Recent applications of RL in optimizing molecular structures [36–39] and chemical reaction pathways [40] have demonstrated its ability to tackle complex molecular design problems.

In this work, we introduce a new framework for predicting the kinetic energy pathways to the catalyst surface segregation under reaction conditions by combining DRL with the domain knowledge of catalysis. For a given catalyst surface, a DRL agent repeatedly attempts to explore nearby local minima with changed surface compositions, and the exploration process is evaluated based on the properties of the local minima and the transition states the agent discovered. In this way, the agent iteratively builds up the knowledge that can generate the kinetic pathways to the accessible surface transformations at a given reaction temperature,

and thus it can be further used to identify metastable catalyst surfaces. The new framework is called CatGym, the first open RL environment developed for reconstruction kinetics.

We showcase our framework with a demonstration of surface segregation kinetics for a ternary  $\text{Ni}_3\text{Pd}_3\text{Au}_2(111)$  alloy catalyst. Pd and Pd-based catalysts are of great industrial importance, being used as hydrogen purification membranes and as a catalyst for the hydrogen evolution reaction (HER), oxygen reduction reaction (ORR), and ethanol oxidation reaction (EOR). In particular, NiPdAu catalysts are useful in hydrogen generation for fuel cells via the catalytic dehydrogenation of formic acid, as well as in the EOR in direct ethanol fuel cells [41, 42]. Researchers have found that the addition of one or more components can increase the activity of Pd-based catalysts, improve resistance to poisoning, and prevent the hydrogen embrittlement brought on by the metal/hydride phase transition [43–49]. Ternary Pd-based alloys have attracted significant attention as the addition of two components allows greater tuning of electrical and structural properties as compared to their binary counterparts. By predicting surface reconstructions and their associated activation barriers, CatGym can aid in the discovery of complex and novel multicomponent catalysts.

## 2. Methods

### 2.1. Reinforcement learning background

RL is a class of artificial intelligence which aims at training an artificial agent by actively interacting with the environment [50]. Following the Markov decision process (MDP), at each timestep  $t$ , the RL agent chooses an action  $a_t$  from the action space  $\mathcal{A}$ , given the current state  $s_t$  from state space  $\mathcal{S}$ . The environment returns a reward,  $r_t$ , as the feedback for the state action pair  $(s_t, a_t)$ . The policy  $\pi_t$  is a mapping from the state space  $\mathcal{S}$  to the action space  $\mathcal{A}$ , namely how the agent decides the action at each timestep. An episode will be terminated once the agent achieves the goal state  $s_g$  or the number of steps reaches the set maximal. The goal of the RL agent is to learn the policy which maximizes the cumulative return  $R_t = \sum_{t=0}^T \gamma^t r_t$ , where  $\gamma$  is the discount factor which adjust the importance of future reward and  $T$  is the length of the episode. During training, the RL agent optimizes the policy  $\pi_t$  sequentially as it actively interacts with the environment. Recently, combination of RL and Deep Learning has made it possible to tackle the problems in very high dimensional space. One of the main drawbacks of RL is its inability to resolve the curse of dimensionality. Deep neural networks, can be used to approximate the value function in reinforcement learning. Benefiting from both reinforcement learning and deep learning, DRL has seen astonishing advancement in various fields, including superhuman-level video game control [51, 52], GO playing [53], robotics control [54–56], and chemical compound design [57–59].

### 2.2. Actor-critic TRPO

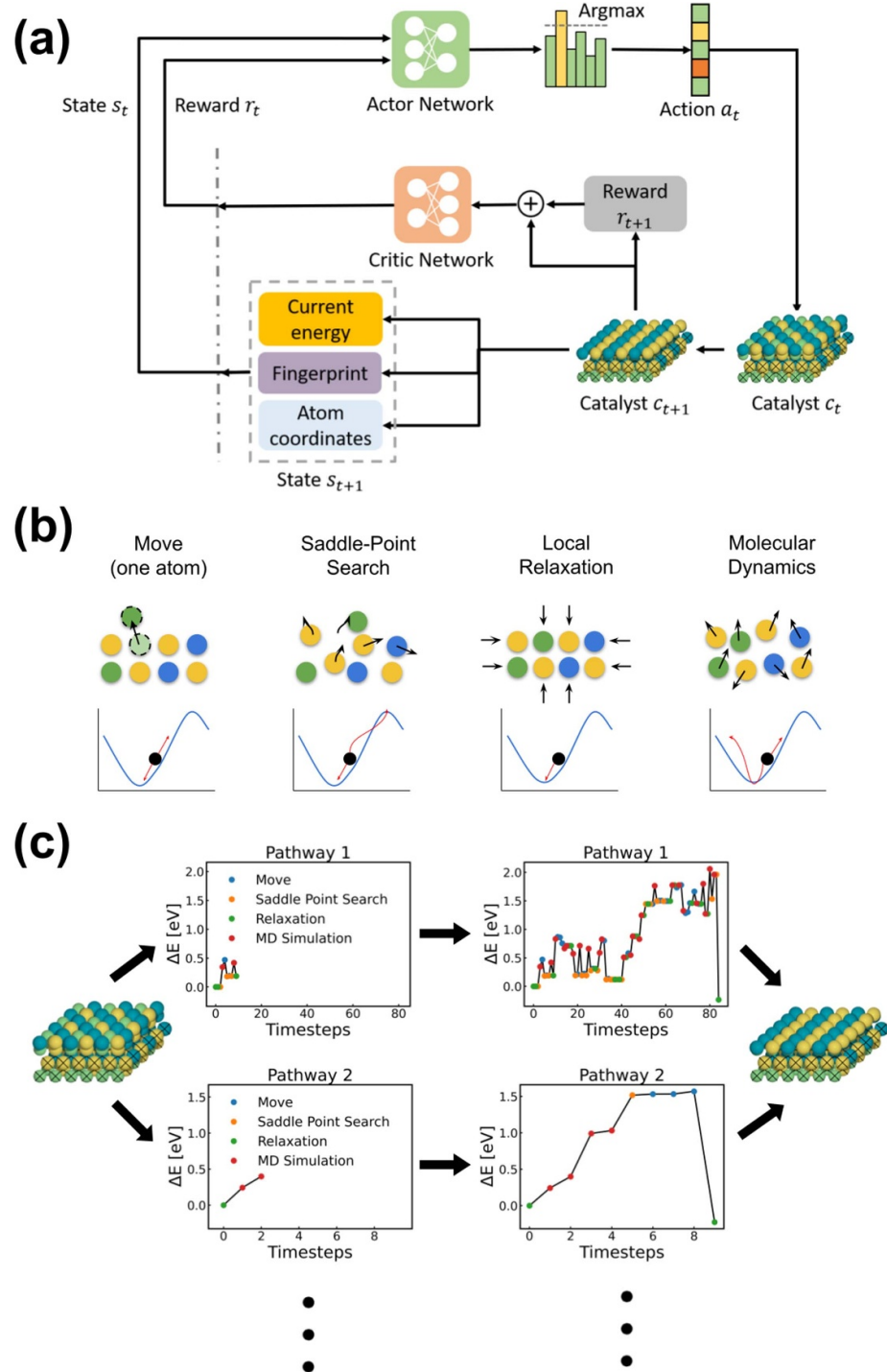
To efficiently optimize the policy in DRL, an actor-critic algorithm [60–62] is proposed which contains two networks: an actor network which determines actions based on a given state embedding, and a critic network which evaluates how good the chosen action is (by estimating the cumulative return). In our work, we train an actor-critic network within the customized framework (figure 1(a)). We implement trust region policy optimization (TRPO)[55], an actor-critic network which has been prevalent in various RL applications. TRPO is effective for optimizing neural network based policies and guarantees monotonic improvement on each update. To this end, we utilize parameterized action space [63, 64], which consists of hierarchical sub-actor networks to decompose the action space.

### 2.3. State representation

The state space  $\mathcal{S}$  is a set of state observations that describes quantified abstraction of the perceived environment. Training an agent that discovers possible surface segregation requires representations for each of the atoms in the near-surface region that can capture dynamically changing local chemical environment. To meet such a requirement, we employ atomic symmetry functions [65] and atomic positions of the free atoms near the surface to efficiently encode the elemental and spatial atomic information. In addition, we provide properties of the surface structure such as energies and forces, so that the agent can understand how the changes in the configurations affect their properties. Finally, we also encode a binary vector that tells whether the agent has found transition states in each episode. These five state observations are then processed through the distinct multi-layer perceptrons and combined to create an embedding of states that are fed to the actor network.

### 2.4. Action space

The action space  $\mathcal{A}$  is a set of actions that the agent can perform to interact with the environment. In general, the action space is designed to be either discrete or continuous based on the reinforcement learning problem



**Figure 1.** (a) Overview of CatGym framework. At each timestep  $t$ , the state space  $s_t$  consists of features of surface structure  $c_t$ , including its current energy, fingerprint, and atom coordinates. The state space along with reward  $r_t$ , which is calculated by critic network based on current energy, are fed into the actor network, and the actor network decides which action to take. The action will modify  $c_t$  to new surface structure  $c_{t+1}$ . This process repeats until the episode is over. (b) Four types of action the actor can choose and their corresponding effects on the energy. If the agent choose to move an atom, it has to further pick which atom to move and the distance (within the range of  $-0.1 \text{ \AA}$  to  $0.1 \text{ \AA}$ ) in the  $x$ ,  $y$  and  $z$  direction. (c) Evolution of the reconstruction pathways to a minimum configuration generated by DRL agent trained with CatGym.  $\Delta E$  represents the relative energy of the catalyst structure at each timestep with respect to the initial state energy.

and the algorithm used for solving it. In this work, we formulate a hierarchical hybrid discrete-continuous action space (figure 1(b)), in which each action  $a_t = (a_t^{(1)}, a_t^{(2)}, a_t^{(3)})$  comprises

- $a_t^{(1)} \in \{1, 2, 3, 4\}$  selects an action from the four different possible actions,

(a) moving individual atoms by a fixed distance towards a direction,

- (b) finding a nearby transition state with a saddle point solver,
  - (c) triggering a local energy relaxation to minimize the current structure,
  - (d) performing short molecular dynamics simulations to perturb the free atoms at a specified reaction temperature.
- $a^{(2)} \in \{1, \dots, N\}$  chooses one of the  $N$  free atoms on the top two surface layers to move if  $a^{(1)} = 1$ .
  - $a^{(3)} \in \{(x, y, z) | -0.1 \leq x, y, z \leq 0.1\}$  that specifies the distances to move the free atom chosen by  $a^{(2)}$  in  $x$ ,  $y$ , and  $z$  directions, respectively.

We use Sella [66] as a saddle point solver for finding nearby first order saddle points ( $a^{(1)} = 2$ ). The energy relaxation uses the BFGS optimizer implemented in atomic simulation environment (ASE) [67] to find nearby local minima ( $a^{(1)} = 3$ ), and the MD simulations ( $a^{(1)} = 4$ ) use Langevin dynamics at constant temperature implemented in ASE as well.

At each timestep, the agent decides how to change the structure of the catalyst surface in order to eventually generate a sequence of actions that leads to the exploration of new local minima with different surface compositions. Figure 1(c) illustrates how the sequence of actions can become the reconstruction pathway to a specific minimum configuration. We note that distinct configurations at different timesteps may end up with the same transition states or the same local minima after performing the saddle point solver or the relaxation. The relative energies ( $\Delta E$ ) between timesteps 0 and 40 in ‘Pathway 1’ plot in figure 1(c) shows that several saddle point search (orange) and relaxation (green) actions lead to the local minima previously visited. This greatly slows down the exploration process by repeatedly leading the agent to the same states. To mitigate the issue, we introduce short MD simulations in our action space so that the agent can more effectively escape from the current minimum or transition state and traverse the potential energy surface. Long enough MD simulations alone might be able to explore new minimum states, but the limited time scale and computing resources make it infeasible. Instead, the agent in our method can decide when one of the other actions is better to explore nearby local minima by learning from its interaction with the CatGym environment.

## 2.5. Reward function

The reward function is designed on the basis of chemical properties of the surface structure. If the agent identifies a transition state at any time during each episode, the agent is rewarded by,

$$r(s_t, a_t) = \frac{1}{\Delta E_{\text{trans}}}, \quad (1)$$

where  $\Delta E_{\text{trans}}$  is the relative energy of the transition state to the energy of the initial state. In addition, after the completion of each episode ( $t = T$ ), the agent is rewarded based on whether it discovers a pathway to a state with a changed surface composition. When the agent successfully observes kinetically feasible surface segregation from the initial surface, it is rewarded by,

$$r(s_T, a_T) = \exp\left(-\frac{\Delta E}{k_B T_r}\right), \quad (2)$$

where  $k_B$  denotes the Boltzmann constant,  $T_r$  is a reaction temperature, and  $\Delta E$  is the potential energy of the final ( $t = T$ ) state relative to the initial state ( $t = 0$ ). These rewards aim at encouraging the agent to explore both the lower transition states and the nearby local minima resulted from surface segregation. The traversability of the transition state is also taken into consideration when evaluating the generated pathway. The higher the transition state energy is, the more difficult the surface segregation to be realized at a given temperature. Therefore, when the agent yields a high energy states that exceeds a predefined upper energy bound ( $3k_B T_r$  in our case), the episode is terminated and it is penalized by,

$$r(s_T, a_T) = \frac{\Delta E}{k_B T_r}. \quad (3)$$

Intuitively, the agent is instructed to find a reconstruction pathway, ideally one leads to a new state with thermodynamically more stable surface composition ( $\Delta E < 0$ ) through a low energy barrier ( $\Delta E_{\text{trans}}$ ). The potential energies can be calculated using accurate quantum chemistry simulations such as DFT, however, evaluating a large number of trial configurations is computationally expensive. Instead, we avoid this cost by employing fast effective medium theory (EMT) [68–70] to estimate the energies.



## 2.6. Experiments

To demonstrate the capability of our proposed DRL framework, we conduct experiments to generate surface reconstruction pathways for a  $2 \times 2 \times 4$  Ni<sub>3</sub>Pd<sub>3</sub>Au<sub>2</sub>(111) alloy catalyst at a specific reaction temperature (1200 K). The top surface layer in the initial Ni-Pd-Au catalyst has a Ni:Pd:Au composition of 1:2:1 and the second layer has a composition of 2:1:1 corresponding to a total of 3 Ni, 3 Pd, and 2 Au in eight lattice positions in the unit cell as shown in figure 2(a). Atoms in the top two layers are free to move, while atoms in the bottom two layers are considered as bulk atoms and are fixed.

We set up the DRL environment under the OpenAI Gym framework [71] and use Tensorforce [72] DRL package to run the experiments. We utilize parallel environment execution to perform multiple experiments running with the same initial conditions. These parallel experiments share the agent and all other model parameters. In each episode, the agent is asked to generate a sequence of at most 500 actions with the aim of exploring the nearby local and global minima with changed surface compositions while finding the minimum energy pathways to these minima. We utilize the EMT potentials for energy and force evaluations for all kinds of actions.

## 2.7. Surface optimization baseline

We use a brute force method and a minima hopping (MH) simulation method [73] for locating local and global minimum surface configurations for the Ni-Pd-Au ternary system. In the brute force method, we considered all the possible arrangements of lattice atoms in the unit cell by performing a distinct permutational method in the top two surface layers. Permutation of 3 Ni, 3 Pd, and 2 Au atoms in the eight lattice positions in the top layers leads to a total of 560 distinct arrangements. We assume that this approach would span all possible local minima that arises from the movement of free atoms in the top two surface layers.

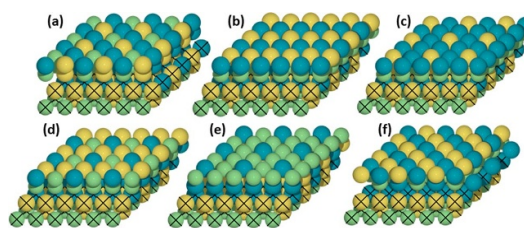
MH is an efficient search method employed for locating the global minimum for systems with highly complex PES. For instance, MH method has been applied to determine the structure of the reconstructed chalcopyrite surfaces [74] and metastable decorate borospherene B<sub>40</sub> [75]. The MH method avoids revisiting known parts of the configuration space by utilizing a feedback mechanism based on simulation history which in turn accelerates and enforces the exploration of the new regions in the configurational space. The MH algorithm consists of an inner part that performs the moves on the PES employing molecular dynamics (MD) followed by the relaxation of the current minimum and an outer part which determines the acceptance or rejection of a new minimum. See figure S1 in supplementary information (available online at [stacks.iop.org/MLST/2/045018/mmedia](https://stacks.iop.org/MLST/2/045018/mmedia)) for more details about MH simulation used in the present study.

## 3. Results and discussion

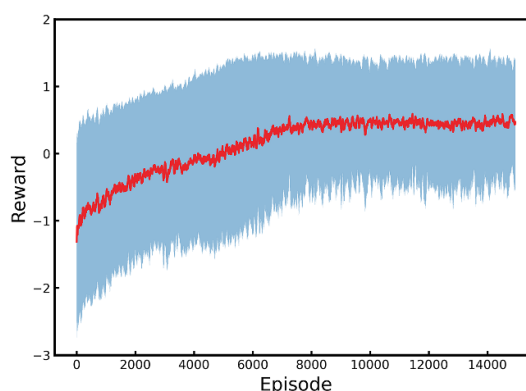
### 3.1. Baseline surface configurations

Using the brute force approach, we have found the global minimum at a relative energy of  $-0.23$  eV with respect to the initial configuration. The minimum has a 0:2:2 composition of Ni, Pd and Au on the top surface layer in the unit cell as shown in figure 2(b). Moreover, several configurations with the same compositions; however, with slightly different arrangement of atoms in the top two layers were identified within the relative energy range of  $-0.23$  to  $-0.19$  eV and can be considered as a group of global minima energy configurations. Followed by this, we have identified a group of the second-lowest energy configurations with Ni:Pd:Au composition of 0:3:1 in the top surface layer (figure 2(c)) with different configurational orientations in the relative energy range of  $-0.18$  to  $-0.16$  eV. A group of the third-lowest energy configurations starts at  $-0.11$  to  $-0.07$  eV energy range and consists of Ni:Pd:Au composition of 1:1:2 on the surface layer as illustrated in figure 2(d). The minima configurations with negative relative energies consistently find surface enrichment of Au and/or Pd with respect to the initial configuration. This observation can be rationalized by arguments for segregation primarily driven by differences in the pure component surface energies. Estimates of the surface energies ( $\gamma_i$ ) for each of Au, Pd, and Ni elements indicate  $\gamma_{\text{Au}} < \gamma_{\text{Pd}} < \gamma_{\text{Ni}}$  [76]. Thus we find a global minimum in which all free (non-fixed) Au atoms segregate to the surface and a nearby local minimum in which Pd fully replaces Ni in the surface. These findings are qualitatively in agreement with experimental studies of segregation in binary alloys: Au is favored over Pd [77–79], Pd is favored over Ni [80–82], and Au is favored over Ni [83–85] in the surface layer. To reaffirm, we also located the configuration with the highest energy (i.e global maximum) which has a Ni:Pd:Au composition of 3:1:0 in the surface layer (figure 2(e)) with an relative energy of  $0.81$  eV. In fact, the global maximum configuration corresponds to the reversal of the top two layers in the global minimum configuration.

For a direct comparison with our DRL model, we have used MH as a standard baseline approach for the local and global minimum search. We found the global minimum configuration with an relative energy



**Figure 2.** Configurations of the Ni(green)-Pd(blue)-Au(yellow) ternary catalyst. (a) Initial catalyst structure, (b)–(e) global minimum, second and third lowest energy configurations, and global maximum configuration obtained via brute-force method, respectively; and (f) global minimum configuration obtained via minima hopping (MH) method. Atoms in the bottom two layers with crosses represent fixed atoms that are constrained from moving.



**Figure 3.** Average reward vs. episode across ten different seeds. Red line represents the average reward, and the blue shadow represents the standard deviation.

of  $-0.23$  eV (figure 2(f)) in the MH simulation. The global minimum configuration obtained via MH simulation shows similar segregation of Pd and Au onto the top layer with 0:2:2 composition of Ni, Pd, and Au on the surface layer. However, they differ in their configurational orientations of the atoms in the top two layers. Like brute-force permutation method, MH method also found many similar configurations of 0:2:2 composition of Ni, Pd, and Au in the top layer with different arrangements with similar energies and can be collectively grouped into a cluster of global minimum configurations.

### 3.2. DRL training summary

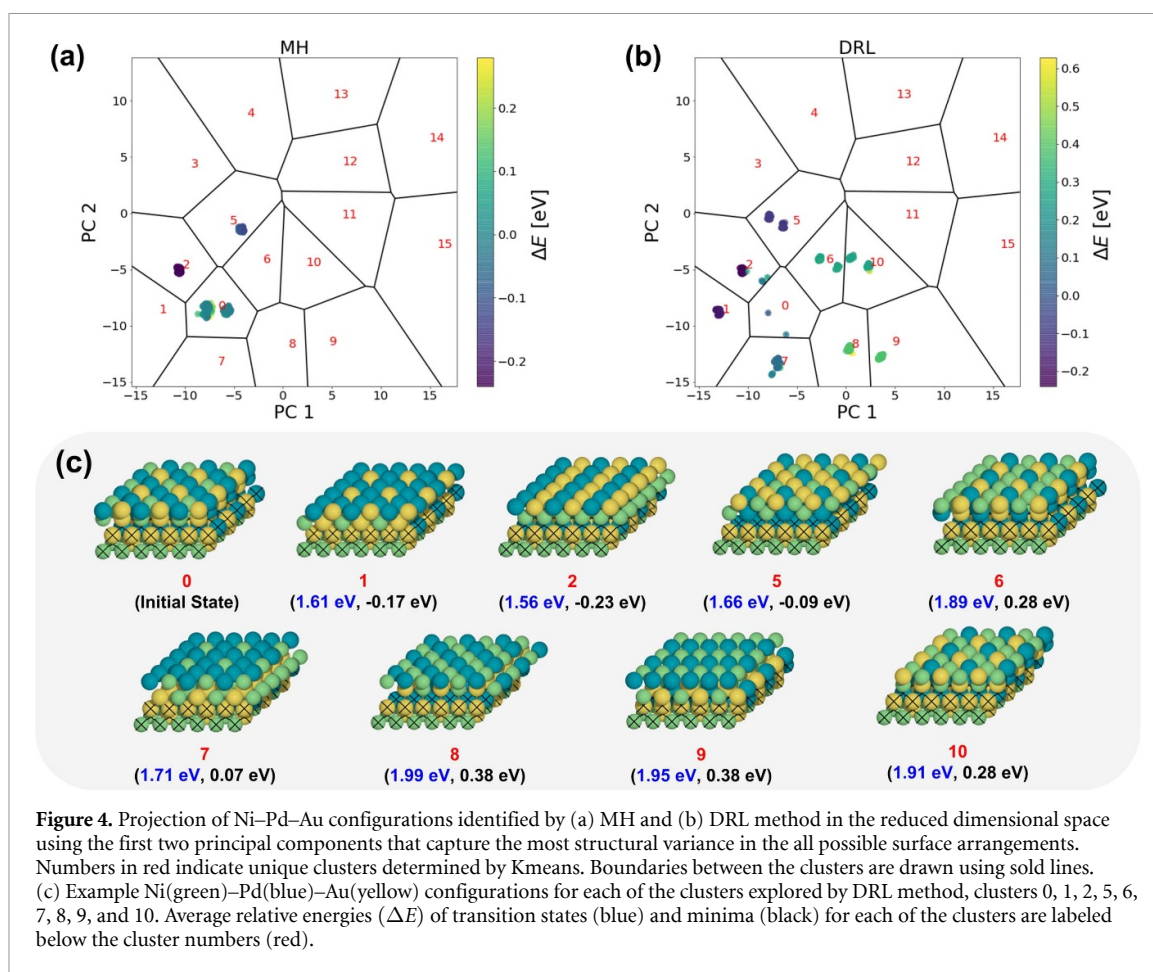
The DRL agent was trained with 10 different seeds by interacting with the CatGym environment. In figure 3, the averaged episodic reward collected during the training process is represented by the red curve, and the standard deviation of the rewards in ten training is represented by blue shadow. The reward at the initial stage of training ( $<5000$  episodes) is low and the policy is noisy because the agent is instructed to explore and has not yet learned to discover surface reconstruction pathway without exceeding the upper energy bound that results in negative rewards. After approximately 8000 episodes of training, the agent starts to receive stable positive reward between 0 and 1. The converged positive reward indicate that the agent has learned a stable policy to discover surface reconstruction pathways that lead to different surface compositions while ensuring the traversability of transition state by not exceeding the upper energy bound.

### 3.3. Surface configuration exploration

First, we performed principal component analysis (PCA) on the minima configurations generated by the brute-force permutation method to determine the reduced dimensional space that captures the most structural variance in the all possible minima configurations. Each of the configurations is represented by a combined atomic fingerprints of the atoms in the top two layers. We then projected the minima configurations onto the reduced dimensional space using the first two principal components as shown in figure S2 in the supplementary information. On the same plot, we also drew the boundaries of clusters determined by the K-means algorithm. We used 16 clusters to partition the configurations plotted in the reduced dimensional space based on their structural representations and properties.

To compare the diversity of the Ni–Pd–Au catalyst surface configurations of local minima explored by our DRL method and the baseline MH methods, we projected these configurations onto the pre-determined

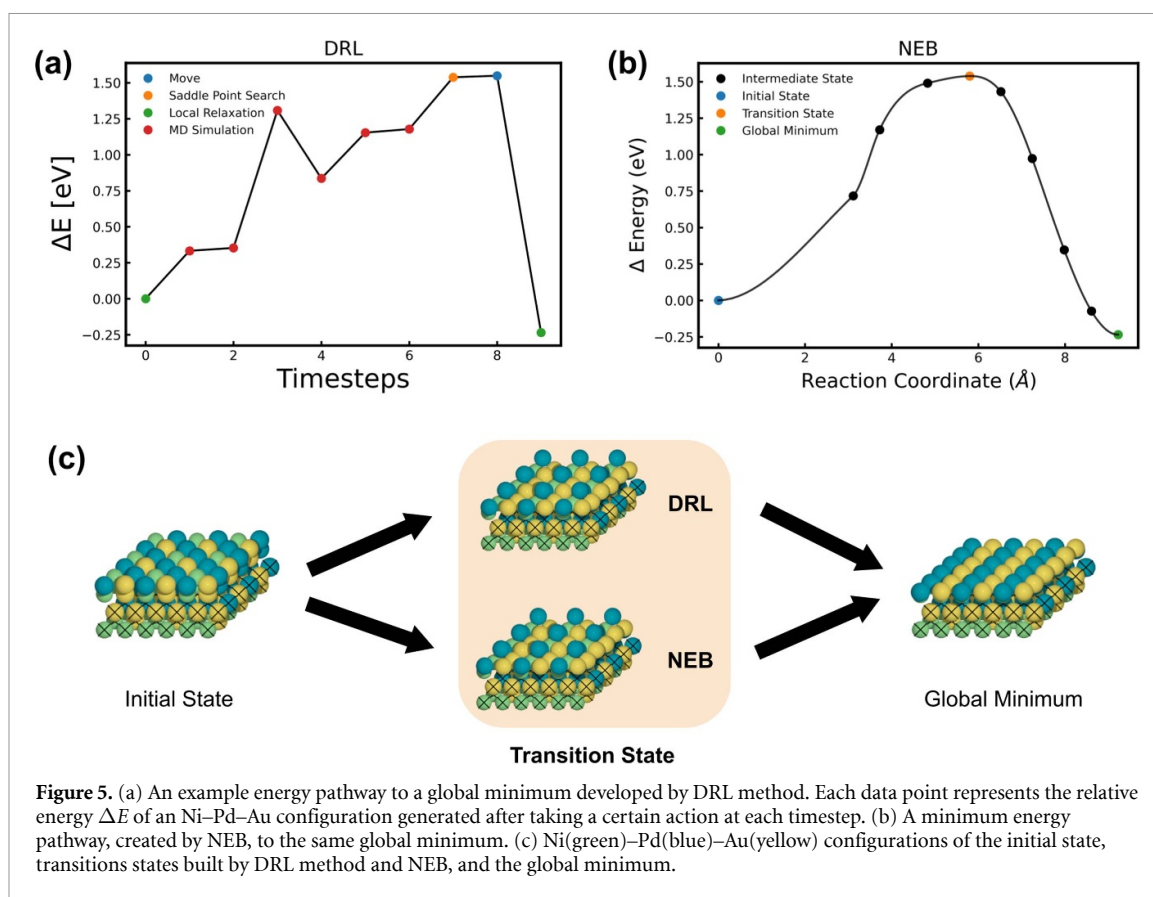




reduced dimensional space mentioned above. Figure 4(a) shows that the MH method could explore surface configurations corresponding to only two different minima clusters, clusters 2 and 5, where Ni:Pd:Au compositions of 0:2:2 and 1:1:2 on the top surface layer were found, respectively. Based on the brute-force permutational analysis, cluster 2 is one of the global minima clusters (figure 2(a)) and cluster 5 is one of the clusters with configurations in the third-lowest energy group (figure 2(c)). On the other hand, CatGym could explore more diverse surface configurations found in eight different minima clusters as shown in figure 4(b). Figure 4(c) shows example Ni-Pd-Au configurations in the eight different clusters visited by the DRL agent with diverse Ni:Pd:Au compositions: 0:2:2, 1:1:2, 0:3:1, 1:2:1, 2:1:1, 1:3:0, and 2:2:0. We note that the same Ni:Pd:Au composition can be found in different nearby clusters with different surface orientations, such as clusters 8 and 9 with 2:2:0 compositions, and clusters 6 and 10 with 2:1:1 compositions. We also notice that our DRL agent was unable to explore other local minima that are not nearby. In figure 4(b), clusters distant from the initial structure in cluster 0 were not identified by the DRL method in this experiment. Those undiscovered clusters include different surface orientations of global minima (cluster 3), second-lowest minima (cluster 4), global maxima (cluster 15), and several other configurations with surface Ni:Pd:Au compositions of 2:1:1 (cluster 11 and 12), 2:0:2 (cluster 13), 3:0:1 (cluster 14) based on figure S2.

### 3.4. Reconstruction pathway and transition state

In addition to the diverse local minima configurations, our DRL method can generate reconstruction pathways to the explored minima with different surface compositions and can provide the structure and properties of not only the minima configurations, but also the transition states, thereby enabling the identification of metastable catalyst surfaces. We note that MH cannot determine the metastability of the catalyst surface because this method does not provide kinetic information about the pathways to the minima it explores. A reconstruction pathway from the initial state to the global minimum state in figure 5(a) demonstrates that the DRL method can creatively generate a sequence of actions to explore the global minimum. The transition state energy or the highest energy barrier in this pathway was determined as 1.56 eV above the energy of the initial state. In figure 4(c), using the same analysis, we show the average transition state energies for all local minima within each of the clusters in the PCA plot (figure 4(b)). We noticed that the transition state energies systematically increase as the relative energies of their final



**Figure 5.** (a) An example energy pathway to a global minimum developed by DRL method. Each data point represents the relative energy  $\Delta E$  of an Ni-Pd-Au configuration generated after taking a certain action at each timestep. (b) A minimum energy pathway, created by NEB, to the same global minimum. (c) Ni(green)-Pd(blue)-Au(yellow) configurations of the initial state, transitions states built by DRL method and NEB, and the global minimum.

minimum states increase. For a cluster of global minimum configurations (cluster 2), the average transition state energy is determined as 1.56 eV while the average transition state energies for clusters of the highest energy states (clusters 8 and 9) explored by our DRL method are in a range of 1.95 eV to 1.99 eV. More example pathways to other local minima can be found in figure S3 in the supplementary information.

### 3.5. Transition state verification

We further performed nudged elastic band (NEB) [86, 87] calculations to verify the structure and energy of the transition state in the path to the global minimum. NEB is a method for finding saddle points and minimum energy paths between known initial and final states, which are in this work the same initial structure and the global minimum state discovered by our DRL method. We extracted the intermediate configurations from the reconstruction pathway developed by the DRL method (configurations at timesteps from 1 to 8 in figure 5(a)), and used them as the intermediate configurations along the minimum energy path in the NEB calculation. In figure 5(a), the agent discovered a transition state to the global minimum at timestep 7, and then performed a local relaxation to achieve the global minimum at timestep 9. Figure 5(b) shows the minimum energy path constructed by NEB between the same initial state and the same global minimum state. The energy barrier in this minimum path is 1.54 eV, which is very close to the one (1.56 eV) estimated by the DRL method. Further, figure 5(c) visually verifies that the structures of the transition states identified by both the DRL method and NEB are close to each other. Both DRL and NEB find transition states in which one of the Pd (blue) atoms vacate the surface layer. This vacancy then facilitates the swapping of a surface Ni (green) and a subsurface Au (yellow) atom, followed by the return of Pd atom down to the surface to fill the vacancy, thereby resulting in the global minimum configuration with Ni: Pd: Au composition of 0:2:2 for the top surface layer.

## 4. Conclusion

In this work, we present CatGym DRL environment for studying kinetic arguments of catalyst surface segregation under reaction conditions. We aim at exploring possible surface segregation phenomena and the associated transition states to address the challenge in predicting catalyst metastability. For a given catalyst surface, the DRL agent iteratively alters the positions of atoms and learns strategies for generating kinetic pathways to nearby local minima with different surface compositions resulted from surface segregation.

Trained with the TRPO algorithm and a ternary Ni<sub>3</sub>Pd<sub>3</sub>Au<sub>2</sub>(111) alloy catalyst, the agent in our CatGym environment not only explores more diverse local and global minima configurations compared to the baseline MH method, but also generates kinetic pathways to those configurations. We also verify that the reconstruction pathway to the global minimum surface configuration generated by the DRL agent is in a good agreement with the minimum energy path calculated using NEB. CatGym is the first general DRL approach towards the design of metastable catalysts under reaction conditions. This approach can be extended to other systems of interest possibly containing different catalyst surfaces with varying unit cell sizes, metals, oxides, and adsorbates with only a few minor modifications.








## Data availability statement

The data and code that support the findings of this study are openly available at <https://github.com/ulissigroup/catgym>.

## Acknowledgments

The information, data, or work presented herein was funded in part by the Advanced Research Projects Agency—Energy (ARPA-E), U.S. Department of Energy, under Award No. DE-AR0001221.

## ORCID iDs

Zhonglin Cao  <https://orcid.org/0000-0003-2096-1178>  
Rajesh K Raju  <https://orcid.org/0000-0002-0030-3242>  
Yuyang Wang  <https://orcid.org/0000-0003-0723-6246>  
Robert Burnley  <https://orcid.org/0000-0003-3346-6432>  
Andrew J Gellman  <https://orcid.org/0000-0001-6618-7427>  
Amir Barati Farimani  <https://orcid.org/0000-0002-2952-8576>  
Zachary W Ulissi  <https://orcid.org/0000-0002-9401-4918>

## References

- [1] Poncet V 2001 Alloy catalysts: the concepts *Appl. Catal. A* **222** 31–45
- [2] Bertolini J, Miegge P, Hermann P, Rousset J and Tardy B 1995 On the reactivity of 2d Pd surface alloys obtained by surface segregation or deposition technique *Surf. Sci.* **331**–333 651–8
- [3] Fu J, Yang X, Menning C A, Chen J G and Koel B E 2016 Composition, structure and stability of surfaces formed by Ni deposition on Pd(111) *Surf. Sci.* **646** 56–64
- [4] Zhang S, Nguyen L, Zhu Y, Zhan S, Tsung C K F and Tao F F 2013 In-situ studies of nanocatalysis *Acc. Chem. Res.* **46** 1731–9
- [5] Tao F and Crozier P A 2016 Atomic-scale observations of catalyst structures under reaction conditions and during catalysis *Chem. Rev.* **116** 3487–539
- [6] Bergmann A and Roldan Cuenya B 2019 Operando insights into nanoparticle transformations during catalysis *ACS Catal.* **9** 10020–43
- [7] Salmeron M and Schlögl R 2008 Ambient pressure photoelectron spectroscopy: a new tool for surface science and nanotechnology *Surf. Sci. Rep.* **63** 169–99
- [8] Green I X, Tang W, Neurock M and Yates J T 2011 Spectroscopic observation of dual catalytic sites during oxidation of CO on a Au/TiO<sub>2</sub> catalyst *Science* **333** 736–9
- [9] Herranz T, Deng X, Cabot A, Liu Z and Salmeron M 2011 In situ XPS study of the adsorption and reactions of NO and O<sub>2</sub> on gold nanoparticles deposited on TiO<sub>2</sub> and SiO<sub>2</sub> *J. Catal.* **283** 119–23
- [10] Wu C Y, Wolf W J, Levartovsky Y, Bechtel H A, Martin M C, Toste F D and Gross E 2017 High-spatial-resolution mapping of catalytic reactions on single particles *Nature* **541** 511–5
- [11] Tao F et al 2008 Reaction-driven restructuring of Rh-Pd and Pt-Pd core-shell nanoparticles *Science* **322** 932–4
- [12] Pfisterer J H, Liang Y, Schneider O and Bandarenka A S 2017 Direct instrumental identification of catalytically active surface sites *Nature* **549** 74–7
- [13] Tao F, Tang D, Salmeron M and Somorjai G A 2008 A new scanning tunneling microscope reactor used for high-pressure and high-temperature catalysis studies *Rev. Sci. Instrum.* **79** 084101
- [14] Tao F, Dag S, Wang L-W, Liu Z, Butcher D R, Bluhm H, Salmeron M and Somorjai G A 2010 Break-up of stepped platinum catalyst surfaces by high CO coverage *Science* **327** 850–3
- [15] Subramanian A and Marks L 2004 Surface crystallography via electron microscopy *Ultramicroscopy* **98** 151–7
- [16] Su D S, Zhang B and Schlögl R 2015 Electron microscopy of solid catalysts—transforming from a challenge to a toolbox *Chem. Rev.* **115** 2818–82
- [17] Nakamura E 2017 Atomic-resolution transmission electron microscopic movies for study of organic molecules, assemblies and reactions: the first 10 years of development *Acc. Chem. Res.* **50** 1281–92
- [18] Ortalan V, Uzun A, Gates B C and Browning N D 2010 Towards full-structure determination of bimetallic nanoparticles with an aberration-corrected electron microscope *Nat. Nanotechnol.* **5** 843–7
- [19] Wendt S et al 2008 The role of interstitial sites in the Ti 3d defect state in the band gap of titania *Science* **320** 1755–9
- [20] Tang M, Yuan W, Ou Y, Li G, You R, Li S, Yang H, Zhang Z and Wang Y 2020 Recent progresses on structural reconstruction of nanosized metal catalysts via controlled-atmosphere transmission electron microscopy: a review *ACS Catal.* **10** 14419–50

- [21] Xie T and Grossman J C 2018 Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties *Phys. Rev. Lett.* **120** 145301
- [22] Back S, Yoon J, Tian N, Zhong W, Tran K and Ulissi Z W 2019 Convolutional neural network of atomic surface structures to predict binding energies for high-throughput screening of catalysts *J. Phys. Chem. Lett.* **10** 4401–8
- [23] Tran K and Ulissi Z W 2018 Active learning across intermetallics to guide discovery of electrocatalysts for CO<sub>2</sub> reduction and H<sub>2</sub> evolution *Nat. Catal.* **1** 696–703
- [24] Li Z, Wang S, Chin W S, Achenie L E and Xin H 2017 High-throughput screening of bimetallic catalysts enabled by machine learning *J. Mater. Chem. A* **5** 24131–8
- [25] Ma X, Li Z, Achenie L E K and Xin H 2015 Machine-learning-augmented chemisorption model for CO<sub>2</sub> electroreduction catalyst screening *J. Phys. Chem. Lett.* **6** 3528–33
- [26] Yoon J and Ulissi Z W 2020 Differentiable optimization for the prediction of ground state structures (DOGSS) *Phys. Rev. Lett.* **125** 173001
- [27] Hu M, Tan Q, Knibbe R, Wang S, Li X, Wu T, Jarin S and Zhang M-X 2021 Prediction of mechanical properties of wrought aluminium alloys using feature engineering assisted machine learning approach *Metallurgical Mater. Trans. A* **52** 2873–84
- [28] Bhandari U, Rafi M R, Zhang C and Yang S 2021 Yield strength prediction of high-entropy alloys using machine learning *Mater. Today Commun.* **26** 101871
- [29] Xiong J, Shi S-Q and Zhang T-Y 2020 A machine-learning approach to predicting and understanding the properties of amorphous metallic alloys *Mater. Des.* **187** 108378
- [30] Deringer V L, Pickard C J and Csányi G 2018 Data-driven learning of total and local energies in elemental boron *Phys. Rev. Lett.* **120** 156001
- [31] Deringer V L, Caro M A and Csányi G 2019 Machine learning interatomic potentials as emerging tools for materials science *Adv. Mater.* **31** 1902765
- [32] Behler J 2015 Constructing high-dimensional neural network potentials: a tutorial review *Int. J. Quantum Chem.* **115** 1032–50
- [33] Podryabinkin E V, Tikhonov E V, Shapeev A V and Oganov A R 2019 Accelerating crystal structure prediction by machine-learning interatomic potentials with active learning *Phys. Rev. B* **99** 064114
- [34] Schütt K T, Kessel P, Gastegger M, Nicoli K A, Tkatchenko A and Müller K-R 2019 Schnetpack: a deep learning toolbox for atomistic systems *J. Chem. Theory Comput.* **15** 448–55
- [35] Schütt K T, Kindermans P-J, Sauceda H E, Chmiela S, Tkatchenko A and Müller K-R 2017 Schnet: a continuous-filter convolutional neural network for modeling quantum interactions (arXiv:1706.08566 [stat.ML])
- [36] Christiansen M-P V, Mortensen H L, Meldgaard S A and Hammer B 2020 Gaussian representation for image recognition and reinforcement learning of atomistic structure *J. Chem. Phys.* **153** 044107
- [37] Simm G, Pinsler R and Hernandez-Lobato J M 2020 Reinforcement learning for molecular design guided by quantum mechanics *Proc. 37th Int. Conf. on Machine Learning (PMLR)* vol 119, ed H D III and A Singh pp 8959–69
- [38] Meldgaard S A, Mortensen H L, Jorgensen M S and Hammer B 2020 Structure prediction of surface reconstructions by deep reinforcement learning *J. Phys.: Condens. Matter.* **32** 404005
- [39] Jorgensen M S, Mortensen H L, Meldgaard S A, Kolsbjerg E L, Jacobsen T L, Sørensen K H and Hammer B 2019 Atomistic structure learning *J. Chem. Phys.* **151** 054111
- [40] Zhou Z, Li X and Zare R N 2017 Optimizing chemical reactions with deep reinforcement learning *ACS Cent. Sci.* **3** 1337–44
- [41] Wang Z-L, Ping Y, Yan J-M, Wang H-L and Jiang Q 2014 Hydrogen generation from formic acid decomposition at room temperature using a NiAuPd alloy nanocatalyst *Int. J. Hydrog. Energy* **39** 4850–6
- [42] Dutta A and Datta J 2012 Outstanding catalyst performance of PdAuNi nanoparticles for the anodic reaction in an alkaline direct ethanol (with anion-exchange membrane) fuel cell *J. Phys. Chem. C* **116** 25677–88
- [43] Honrado Guerreiro B, Martin M H, Roué L and Guay D 2016 Hydrogen solubility of magnetron co-sputtered FCC and BCC PdCuAu thin films *J. Phys. Chem. C* **120** 5297–307
- [44] Zhao M, Brouwer J, Sloof W G and Bottger A 2020 Surface segregation of ternary alloys: effect of the interaction between solute elements *Adv. Mater. Interfaces* **7** 1901784
- [45] Tarditi A M, Imhoff C, Miller J B and Cornaglia L 2015 Surface composition of PdCuAu ternary alloys: a combined LEIS and XPS study *Surf. Interface Anal.* **47** 745–54
- [46] Pan B et al 2020 Unexpectedly high stability and surface reconstruction of PdAuAg nanoparticles for formate oxidation electrocatalysis *Nanoscale* **12** 11659–71
- [47] Tarditi A M and Cornaglia L M 2011 Novel PdAgCu ternary alloy as promising materials for hydrogen separation membranes: synthesis and characterization *Surf. Sci.* **605** 62–71
- [48] Luo L-M, Zhan W, Zhang R-H, Chen D, Hu Q-Y, Guo Y-F and Zhou X-W 2019 Ternary CoAuPd and binary AuPd electrocatalysts for methanol oxidation and oxygen reduction reaction: enhanced catalytic performance by surface reconstruction *J. Power Sources* **412** 142–52
- [49] Liu S, Zhang H, Mu X and Chen C 2019 Surface reconstruction engineering of twinned Pd<sub>2</sub>CoAg nanocrystals by atomic vacancy inducement for hydrogen evolution and oxygen reduction reactions *Appl. Catal. B* **241** 424–9
- [50] Sutton R S and Barto A G 2018 *Reinforcement Learning: An Introduction* (Cambridge, MA: MIT press)
- [51] Mnih V, Kavukcuoglu K, Silver D, Graves A, Antonoglou I, Wierstra D and Riedmiller M 2013 Playing atari with deep reinforcement learning (arXiv:1312.5602)
- [52] Mnih V et al 2015 Human-level control through deep reinforcement learning *Nature* **518** 529–33
- [53] Silver D et al 2016 Mastering the game of go with deep neural networks and tree search *Nature* **529** 484–9
- [54] Lillicrap T, Hunt J, Pritzel A, Heess N, Erez T, Tassa Y, Silver D and Wierstra D 2016 Continuous control with deep reinforcement learning (arXiv:1509.02971)
- [55] Schulman J, Levine S, Abbeel P, Jordan M and Moritz P 2015 Trust region policy optimization *Int. Conf. on Machine Learning* pp 1889–97
- [56] Schulman J, Wolski F, Dhariwal P, Radford A and Klimov O 2017 Proximal policy optimization algorithms (arXiv:1707.06347)
- [57] Popova M, Isayev O and Tropsha A 2018 Deep reinforcement learning for de novo drug design *Sci. Adv.* **4** eaa7885
- [58] Zhou Z, Kearnes S, Li L, Zare R N and Riley P 2019 Optimization of molecules via deep reinforcement learning *Sci. Rep.* **9** 1–10
- [59] Wang Y, Cao Z and Barati Farimani A 2021 Efficient water desalination with graphene nanopores obtained using artificial intelligence *npj 2D Mater. Appl.* **5** 1
- [60] Sutton R S et al. 1999 Policy gradient methods for reinforcement learning with function approximation *Proc. of the 12th Int. Conf. on Neural Information Processing Systems* vol 99 (Cambridge, MA: MIT Press) pp 1057–63



- [61] Mnih V, Badia A P, Mirza M, Graves A, Lillicrap T, Harley T, Silver D and Kavukcuoglu K 2016 Asynchronous methods for deep reinforcement learning *Int. Conf. on Machine Learning* (PMLR) pp 1928–37
- [62] Haarnoja T et al 2018 Soft actor-critic algorithms and applications (arXiv:1812.05905)
- [63] Neunert M et al 2020 Continuous-discrete reinforcement learning for hybrid control in robotics *Conf. on Robot Learning* (PMLR) pp 735–51
- [64] Fan Z, Su R, Zhang W and Yu Y 2019 Hybrid actor-critic reinforcement learning in parameterized action space (arXiv:1903.01344)
- [65] Behler J and Parrinello M 2007 Generalized neural-network representation of high-dimensional potential-energy surfaces *Phys. Rev. Lett.* **98** 146401
- [66] Hermes E D, Sargsyan K, Najm H N and Zádor J 2019 Accelerated saddle point refinement through full exploitation of partial Hessian diagonalization *J. Chem. Theory Comput.* **15** 6536–49
- [67] Larsen A H et al 2017 The atomic simulation environment—a python library for working with atoms *J. Phys.: Condens. Matter.* **29** 273002
- [68] Tadmor E B, Elliott R S, Sethna J P, Miller R E and Becker C A 2011 The potential of atomistic simulations and the knowledgebase of interatomic models *JOM* **63** 17
- [69] Jacobsen K W, Norskov J K and Puska M J 1987 Interatomic interactions in the effective-medium theory *Phys. Rev. B* **35** 7423–42
- [70] Jacobsen K, Stoltze P and Norskov J 1996 A semi-empirical effective medium theory for metals and alloys *Surf. Sci.* **366** 394–402
- [71] Brockman G, Cheung V, Pettersson L, Schneider J, Schulman J, Tang J and Zaremba W 2016 Openai gym (arXiv:1606.01540 [cs.LG])
- [72] Kuhnle A, Schaarschmidt M and Fricke K 2017 Tensorforce: a tensorflow library for applied reinforcement learning (available at: <https://github.com/tensorforce/tensorforce>)
- [73] Goedecker S 2004 Minima hopping: an efficient search method for the global minimum of the potential energy surface of complex molecular systems *J. Chem. Phys.* **120** 9911–17
- [74] Thinius S, Islam M M and Bredow T 2018 The structure of reconstructed chalcopyrite surfaces *Surf. Sci.* **669** 1–9
- [75] Saha S, Genovese L and Goedecker S 2017 Metastable exohedrally decorated borospherene B<sub>40</sub> *Sci. Rep.* **7** 7618
- [76] Vitos L, Ruban A, Skriver H and Kollár J 1998 The surface energy of metals *Surf. Sci.* **411** 186–202
- [77] Creuze J, Guesmi H, Mottet C, Zhu B and Legrand B 2015 Surface segregation in AuPd alloys: ab initio analysis of the driving forces *Surf. Sci.* **639** 48–53
- [78] Yin C, Guo Z and Gellman A J 2020 Surface segregation across ternary alloy composition space: Cu<sub>x</sub>Au<sub>y</sub>Pd<sub>1-x-y</sub> *J. Phys. Chem. C* **124** 10605–14
- [79] Yi C W, Luo K, Wei T and Goodman D W 2005 The composition and structure of Pd–Au surfaces *J. Phys. Chem. B* **109** 18535–40
- [80] Stoddart C, Moss R and Pope D 1975 Determination of the surface composition of palladium-nickel alloy film catalysts using Auger electron spectroscopy *Surf. Sci.* **53** 241–56
- [81] Derry G N, Wan R, Strauch F and English C 2011 Segregation and interlayer relaxation at the NiPd(111) surface *J. Vac. Sci. Technol. A* **29** 011015
- [82] Abel M, Robach Y, Bertolini J-C and Porte L 2000 STM comparative study of the Pd<sub>8</sub>Ni<sub>92</sub>(110) alloy surface and the Pd/Ni(110) surface alloy *Surf. Sci.* **454-456** 1–5
- [83] Williams F L and Boudart M 1973 Surface composition of nickel–gold alloys *J. Catal.* **30** 438–43
- [84] Burton J J, Helms C R and Polizzotti R S 1976 Surface segregation in alloys: LEED, Auger, and gas adsorption study of segregation of Au to the (111) surface of Ni *J. Chem. Phys.* **65** 1089–100
- [85] Krawczyk M, Zommer L, Sobczak J, Jablonski A, Petit M, Robert-Goumet C and Gruzza B 2004 IMFP measurements near Au–Ni alloy surfaces by EPES: indirect evidence of submonolayer Au surface enrichment *Surf. Sci.* **566-568** 856–61
- [86] Henkelman G and Jónsson H 2000 Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points *J. Chem. Phys.* **113** 9978–85
- [87] Henkelman G, Uberuaga B P and Jónsson H 2000 A climbing image nudged elastic band method for finding saddle points and minimum energy paths *J. Chem. Phys.* **113** 9901–4