Chemical Science

ROYAL SOCIETY OF CHEMISTRY

EDGE ARTICLE

View Article Online
View Journal | View Issue

Cite this: *Chem. Sci.*, 2023, **14**, 8777

All publication charges for this article have been paid for by the Royal Society of Chemistry

Received 12th June 2023
Accepted 13th July 2023

DOI: 10.1039/d3sc02974c

rsc.li/chemical-science

Accessing complex reconstructed material structures with hybrid global optimization accelerated *via* on-the-fly machine learning†

Xiangcheng Shi,[abdf] Dongfang Cheng,[abd] Ran Zhao,[abd] Gong Zhang,[abd] Shican Wu,[abd] Shiyu Zhen,[abd] Zhi-Jian Zhao *[abcd] and Jinlong Gong *[abcde]

The complex reconstructed structure of materials can be revealed by global optimization. This paper describes a hybrid evolutionary algorithm (HEA) that combines differential evolution and genetic algorithms with a multi-tribe framework. An on-the-fly machine learning calculator is adopted to expedite the identification of low-lying structures. With a superior performance to other well-established methods, we further demonstrate its efficacy by optimizing the complex oxidized surface of Pt/Pd/Cu with different facets under (4 × 4) periodicity. The obtained structures are consistent with experimental results and are energetically lower than the previously presented model.

Introduction

The rational design of highly efficient materials requires an atom-level understanding of their structure–performance relationship.[1–5] However, under working conditions, most materials undergo a structural reconstruction accompanied by an unpredictable performance.[6–9] For example, some bimetallic catalysts like Au–Ag alloys can exhibit dynamic geometrical and compositional reconstruction during the reaction, which generates active sites to boost performance.[6,10–14] In contrast, under high voltage, metal catalysts can be partially oxidized, which results in destabilization and higher dissolution of the active species.[15,16] Surface reconstruction sensitively varies with the nature of unreconstructed surfaces and the compositions/concentrations of adsorbent.[17,18] Due to the difficulty to model a reconstructed surface, it is a generally adopted, but improper, practice to oversimplify a complex surface when modeling. For example, to model a partially oxidized surface, some studies place adsorbed oxygen,[19] or only a layer of metal oxides,[20] upon a metal surface without considering reconstruction.[21] To achieve the rational design of highly efficient materials, it is essential to properly model their reconstructed structures to understand their real structure–performance relationship.

Reconstructed surfaces tend to adopt the most thermodynamically stable structures,[22] which is the lowest point on the potential energy surface (PES), the so-called global minimum (GM).[18] Finding the GM of a working material constitutes a global optimization (GO) problem. There has been significant progress in developing GO algorithms for chemical structure optimization. However, most of them are mainly developed for isolated particles like crystals, clusters, or supported clusters, with simplified models deemed to be sufficient for most investigations.[22–24] The optimization of the surface system is more geometrically restricted than that of isolated particles, owing to the periodicity and the presence of strong covalent bonds between surface atoms and the underlying support,[25] and generally more atoms are required for reliable modeling.

Two major difficulties should be considered for globally optimizing surface structures: how to efficiently explore the highly complex PES, and how to reduce high computational costs caused by massive local relaxations used to describe the PES.[26] Indeed, some GO algorithms that were originally developed for isolated particles have been applied to surface structures.[27,28] However, their efficiency has not yet been examined systematically. Previous reports have expressed concern about the efficiency of the genetic algorithm (GA) that two good parent structures may produce poor candidates with high energy.[29,30] The insufficient efficiency also limits the model size for optimizing a surface system, as most studies are conducted generally with no more than (2 × 2) periodicity.[31–37] It is even unreliable since no known criterion guarantees that the "best

[a]*School of Chemical Engineering and Technology, Key Laboratory for Green Chemical Technology of Ministry of Education, Tianjin University, Tianjin 300072, China. E-mail: zjzhao@tju.edu.cn; jlgong@tju.edu.cn*

[b]*Collaborative Innovation Center of Chemical Science and Engineering (Tianjin), Tianjin 300072, China*

[c]*Haihe Laboratory of Sustainable Chemical Transformations, Tianjin 300192, China*

[d]*National Industry-Education Platform of Energy Storage, Tianjin University, 135 Yaguan Road, Tianjin 300350, China*

[e]*Joint School of National University of Singapore and Tianjin University, International Campus of Tianjin University, Binhai New City, Fuzhou 350207, Fujian, China*

[f]*Department of Chemistry, National University of Singapore, 3 Science Drive 3, Singapore, 117543, Republic of Singapore*

† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d3sc02974c

structure" encountered by a GO algorithm is the GM that reflects the reconstructed structures.[38]

The absence of a criterion is primarily attributed to the inherent limitations in the spatiotemporal resolution of characterization techniques, resulting in a dearth of prior knowledge regarding the precise atomic-level structure of a reconstructed surface.[39] This situation poses a significant challenge to the integration of machine learning (ML) aimed at mitigating the computational burden associated with GO. It is dangerous to rely solely on a pre-trained ML calculator that suffers extrapolation problems, as the GM can correspond to a very narrow basin of the PES and can hardly be involved in the training set.[40,41] Previous research on ML-involved structural searches, using techniques without extrapolation design, such as neural networks, has generally been randomly generated or an already built database,[42] or adjusting the atomic position is very constrained during the search.[43–45] Even for on-the-fly ML frameworks with sampling design such as Bayesian optimization,[46] if they are applied solely without incorporating near-GM structural features into the training set, their effectiveness can be compromised.[47]

This paper describes a new strategy for the global optimization of complex catalytic surfaces using a hybrid evolutionary algorithm (HEA) that combines differential evolution (DE) and genetic algorithms with a co-evolution framework. An on-the-fly machine learning calculator based on Gaussian processes is adopted to complement local evaluations and expedite the identification of low-lying structures. We demonstrate the HEA method in obtaining the complex surface oxide structure of different facets of transition metals like Pt, Pd, and Cu using a (4 × 4) supercell. The globally optimized structures are lower than previously reported theoretical modeling and are consistent with experimental observation, providing important clues for the rational design of catalysts.

HEA methods

A flowchart of the HEA program is shown in Fig. 1. A "tribe" framework is adopted in the HEA program to simulate the real evolutionary process in nature. Specifically, several optimization processes (each is considered as a "tribe") are concurrently run with a periodic exchange of the most stable members among tribes. Firstly, an initial set of structures is generated at random within appropriate limits for a (small) cell shape and interatomic distances, which is then enlarged to the required periodicity and relaxed at the DFT level. Thus, the generated structure has high symmetry and is more likely to be stable.[48,49] An ML calculator based on a Gaussian process (GP) regressor is trained on-the-fly using relaxed structures to expedite the identification of low-lying structures, separately for each tribe.

For the offspring population in each tribe, excessive offspring candidates will be generated by a hybrid evolutionary operator: one is to produce offspring candidates from two-parent structures by the GA operator introduced by Deaven and Ho.[29] Note that mutations (permutation, rattle, and mirror) are applied with pre-set probability for a newly generated

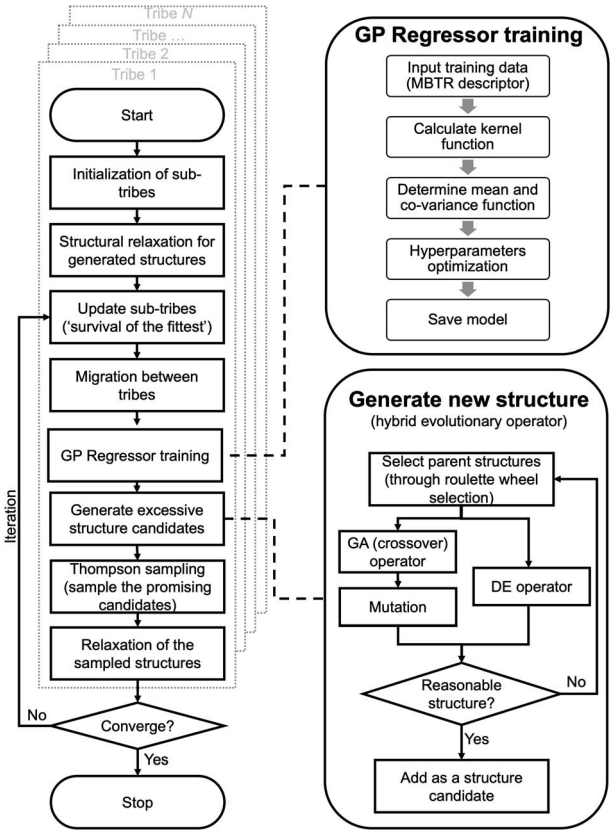

Fig. 1 Overview of the workflow in the HEA program.

structure.[51] Another is to perform the DE operator, and three strategies are introduced:[52]

$$X_{\mathrm{de}} = X_{r_3} + F \times (X_{r_1} - X_{r_2}) \quad (1)$$

$$X_{\mathrm{de}} = X_{\mathrm{best}} + F \times (X_{r_1} - X_{r_2}) \quad (2)$$

$$X_{\mathrm{de}} = X_{r_3} + F \times (X_{\mathrm{best}} - X_{r_3}) + F \times (X_{r_1} - X_{r_2}) \quad (3)$$

where the new structure X_{de} is generated by a linear combination between several randomly selected parent structures (X_{r_1}, X_{r_2}, X_{r_3},…) and a scaled difference (controlled by scaling factor $F \in [0, 2]$) between other donor structures. X_{best} represents the most stable structures in the parent population. Eqn (1)–(3) can be denoted as "DE/rand/1", "DE/best/1" and "DE/rand-to-best/1" respectively.

The generated candidates will first be evaluated by the GP regressor. In practice, the GP regressor needs to deal with an unrelaxed structure. Considering time consumption and memory problems for force prediction, the GP regressor is directly trained and performed using unrelaxed structures with their relaxed energy.[47,53,54]

The GP regressor was implemented using the GPyTorch Python library.[55] A Gaussian process uses Bayesian inference and assumes that the prior distribution for the data can be given by a multivariate normal distribution, while its task is to infer the Gaussian posterior distribution $p(E_*|X,y,X_*)$ for the unseen datapoint X_* (waiting for exploration) based on the

observed training set (X, E_θ). X is taken to be the feature vector rather than the Cartesian coordinates, where the many-body tensor representation (MBTR) descriptor is adopted by the DSCRIBE software package.[50,56] MBTR descriptors provide whole-system representations of periodic systems with the locality of chemical interactions being exploited. To reduce complexity for representing large surface systems, only the top three layers (that are more prominent to reconstruct) are selected for training, while the bottom layer (representing the bulk structure) is excluded. For more realistic modeling, the target value y differs from E_θ by adding an independent identically distributed Gaussian noise, $\varepsilon \sim \mathcal{N}(0, {\sigma_n}^2)$. The key predictive equations for the GP regressor are[57–59]

$$E_*|X, y, X_* \sim N\left(\overline{E_*}, \mathrm{cov}(E_*)\right) \quad (4)$$

where

$$\overline{E_*} = K(X_*, X)\left[K(X, X) + {\sigma_n}^2 I\right]^{-1} y \quad (5)$$

$$\mathrm{cov}(E_*) = K(X_*,X) - K(X_*,X)[K(X,X) + {\sigma_n}^2 I]^{-1} K(X_*,X) \quad (6)$$

The predictive mean $\overline{E_*}$ of the distribution is the estimation of the potential energy, while the variance $\mathrm{cov}(E_*)$ can be an uncertainty quantification. $K(\cdot,\cdot)$ denotes the covariance function (or kernel) matrix that is used to characterize the similarity between samples, which is the very heart of the GP regressor. In our study, the covariance function was chosen to be a sum of a Matern kernel and a linear kernel as

$$k(x_i, x_j) = \frac{2^{1-\nu}}{\Gamma(\nu)}\left[\sqrt{2\nu}\, d(x_i, x_j)\right]^{\nu} K_\nu\left[\sqrt{2\nu}\, (x_i, x_j)\right] + a {x_i}^T x_j \quad (7)$$

where Γ is the gamma function, $d(x_i,x_j)$ is the scaled distance between x_i and x_j, ν is a smoothness parameter where smaller values are less smooth, K_ν is a modified Bessel function and a variance parameter. The Matern kernel is a generalization of the Gaussian kernel that has proved to have an advantage in high dimensional inputs,[60,61] and its synergy effects with the MBTR descriptor have been proven previously.[50]

The optimal hyperparameters Θ^* are determined by the log marginal likelihood as

$$\Theta^* = \arg\max \log p(y|X,\Theta) \quad (8)$$

The Thompson sampling (TS) method is used as the acquisition function that leverages the uncertainty in the posterior to guide exploration, a randomized strategy that samples a reward function (that is relative to potential energy) from the posterior and queries the structure x_{n+1} with the highest reward.[62,63] Fig. S1† shows an illustrative example of a TS-guided on-the-fly structure search in a non-convex search space.

$$x_{n+1} = \arg\max f_w(a) \quad \text{where } w \sim p(y|X,\Theta) \quad (9)$$

Equipped with this trained GP regressor, we choose a batch of multiple candidates, namely the batch Thompson sampling (B-TS) method. Different from the lower confidence bound (UCB) function used in similar studies,[47,59,64] the B-TS method can naturally trade-off between the exploration and exploitation of the PES with no free parameters, thus avoiding the damage of efficiency caused by an inappropriate parameters setting of the UCB function,[41] the effectiveness of which in searching chemical space has been demonstrated and reported before.[63]

Only these "most promising" structures are evaluated at the DFT level. The population is then updated under the 'survival of the fittest': a certain number of the most stable structures from the current (parent + offspring) population are kept, while others are eliminated.[22] Nonetheless, all DFT-evaluated structures are added to the training dataset, and the GP regressor is re-trained on-the-fly.

Results and discussion

Performance of the HEA method

Fig. 2(a) shows the optimizing performance of (4 × 4) 0.75 ML O–Pt(111) as a function of the number of local evaluations, among HEA (with different settings) and other well-established methods, such as GOFEE[59] and SSW,[65] for surface systems, where the HEA program (with and without the GP regressor) achieves the highest performance. With the on-the-fly GP regressor, the HEA program can be further accelerated: the number of local evaluations decreased almost threefold (which saves around 3600 local evaluations) to reach the same energy level, and the GM found eventually is much lower. A dimensionally-reduced visualization of this accelerated performance is

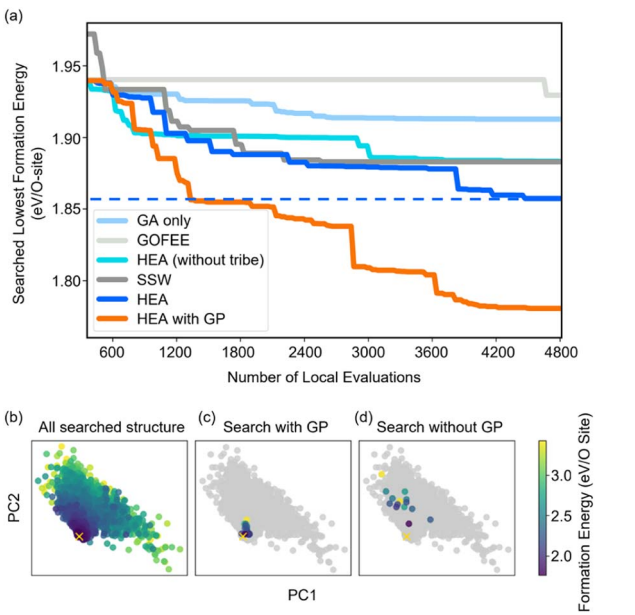

Fig. 2 (a) Structural optimization performance of (4 × 4) 0.75 ML O–Pt(111) among HEA and other well-established methods. All results are repeated three times, and the plotted line represents the mean value. (b) Principal component analysis (PCA) visualization of the MBTR descriptor of the searched structures along the first two PCs (preserving 87% of the dataset variance) visited in the 10th generation in independent searches with/without the GP regressor. The yellow 'x' represents the GM the HEA program finally found.

presented in Fig. 2(b), showing that the search with the GP regressor is closer to the GM region than that without the GP regressor after a certain number of generations. The superior performance of the HEA program is also presented in Fig. S3(a)† for the optimization of the (2 × 2) surface. After the supercell, the obtained (2 × 2) structure is 0.28 eV per O-site less stable than the optimized (4 × 4) structure, shown by the green dot of Fig. 4(a), highlighting the necessity of directly optimizing in a bigger periodicity. As the structural search for a (4 × 4) surface requires even 10 times more local relaxation than that of a (2 × 2) surface, an accelerated module like the GP regressor is desired.

The high efficiency of the HEA program stems from three features: firstly, in Fig. 3a and b, the newly introduced DE operators are much more effective and stable at generating lower energy configurations compared with the GA operator. Mirror and permutation mutation, which is widely used in the optimization of isolated particles, performs poorly when dealing with the surface system. As a result, in Fig. 2(a), GA only cannot efficiently optimize the (4 × 4) surface, which is exceeded by the HEA program combining both GA and DE operators. Secondly, the (on-the-fly) accuracy of the GP regressor produces reliable prediction of (unrelaxed) candidates. In Fig. 3(b), despite the accuracy loss, the MAE for both the models trained using relaxed and unrelaxed structures is below 10 meV per atom after around 10 generations (where the number of data points for the training set is 750).

A relatively correct ranking of the offspring candidates is achieved, as shown in Fig. 3(d) through Kendall's τ coefficient, demonstrating a reliable sampling during the search. Fig. 2(b) also reflects that the MBTR descriptor contains the relevant structural information for approximating the energy, which is the prerequisite for an accurate GP regressor. Finally, introducing the "tribe" framework not only helps maintain a high structural diversity for each tribe,[48] but also enhances the efficiency of the GP regressor. Fig. 3(b and d) show that the accuracy is improved, and the sampling is more targeted. While the sparse GP regressor does not provide enough accuracy (Fig. S5†), here a GP regressor with a full covariance matrix is used that requires $o(N^3)$ computational time and $o(N^2)$ memory space for Cholesky decomposition. Thus, it becomes expensive and its numerical stability is degraded for a large dataset.[66,67] Dividing the training dataset using the "tribe" framework naturally avoids these problems as shown in Fig. 3(c). All the above features contribute to the enhanced structural searchability of the HEA program.

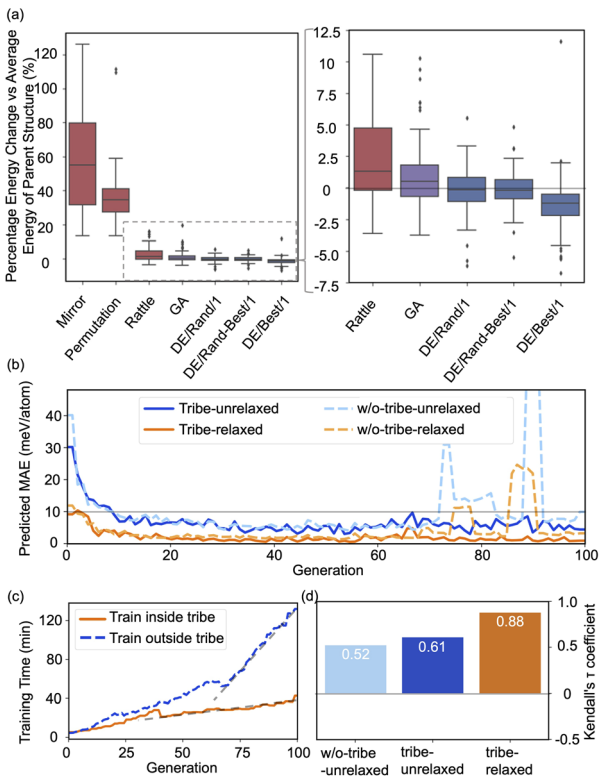

Fig. 3 (a) Boxplots of energy changes compared to the average energy of the parent structures for different offspring operators in the HEA program. Details of the equation are provided in the ESI.† The lines inside the boxplots show the median energy change, the boxplots extend from the lower to the upper quartile of the data, whiskers extend from the 5th to the 95th percentile of the data, and black dots show data points outside of the whisker range. (b) The on-the-fly predict accuracy in optimizing O–Pt(111). "Tribe-unrelaxed": train the GP regressor using unrelaxed structures with their relaxed energy separately inside each tribe, "-relaxed": using relaxed structures, "w/o-tribe": train only one GP regressor outside using structures from all tribes. The number of data points for the training set for 0, 20, 40, 60, and 80 generations is 360, 1260, 2160, 3060, and 3960, respectively, of which one-third are for "tribe" as three tribes are adopted. (c) Time spent for training the GP regressor inside or outside the tribe. (d) Kendall's τ coefficient of the ranking of all predicted energy and its relaxed energy collected from a HEA search. Details of Kendall's τ coefficient are provided in the ESI.†

Application in reconstructed oxide structures

The efficacy of the HEA program is further demonstrated in modeling the reconstruction of metal oxides, knowledge of which has been limited because of their complexity and the scarcity of surface-sensitive characterization techniques.[18,68] Global structural optimization is thus necessary to be applied. As a proof of concept, the complex surface oxides of different metals (Pt, Pd, and Cu) are studied here using the HEA program.

Pt can undergo irreversible restructuring under reaction conditions, due to the surface oxidation and subsequent Pt dissolution, which is thought to decrease catalytic efficiency and durability.[19,69–72] However, the atomic-level modeling of the Pt oxidation remains uncertain.[73]

The optimized structures of 0.75 ML O–Pt(111), which are thought to exist at around 1.0–1.2 V_{RHE} during electro-oxidation,[73,76] are shown in Fig. 4(d and e). The structures consist of two interconnected, protruding square planar PtO_4 units that are 1.7 Å in height. It is consistent with the scanning tunneling microscopy (STM) in Fig. 4(b and c) that the oxidized Pt(111) surface consists of a network of mono-atom-high (1.7 Å), worm-shaped islands.[75,77] The surface oxidation state of PtO_4 units is between that of PtO (Pt^{2+}) and Pt_3O_4 ($Pt^{2.7+}$), which exactly fits the *in situ* XANES showing that the oxidized Pt surface formed at >1.0 V presents square-planar PtO_4 units with Pt in a slightly higher oxidation state than in PtO.[78] With a similar oxidation state, the PtO_4 units also resemble the PtO and Pt_3O_4 bulk oxide shown in Fig. S6.†

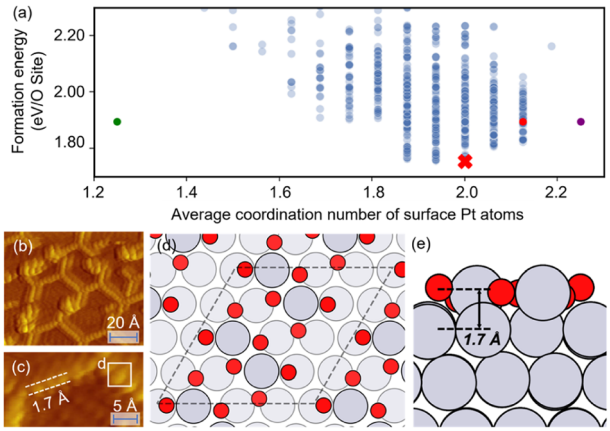

Fig. 4 (a) The scatterplot of the energy of the structure search during a HEA search and the average coordination number of surface Pt atoms. Red cross: the globally optimized structure. Purple dot: the structure proposed by Hawkins *et al.*[74] Red dot: the structure searched by LASP. Green dot: the (2 × 2) optimized structure (after supercell). (b and c) STM images of oxidized Pt(111) (reproduced with permission from ref. 75, Copyright 2008 Elsevier). (d and e) The globally optimized structures of (4 × 4) 0.75 ML O–Pt(111) through the HEA program.

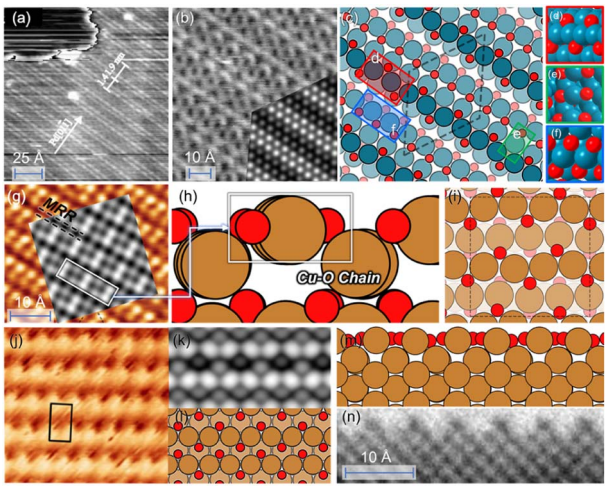

Fig. 5 (a and b) STM image of oxidized Pd(111). (Reproduced with permission from ref. 79, Copyright 2000 IOP Publishing.) Inset: simulated STM images based on the optimized structure. (c–f) Optimized structures of 1 ML O–Pd(111). (g) STM images of O–Cu(100). (Reproduced with permission from ref. 80, Copyright 2013 American Chemical Society.) Inset: simulated STM images based on the optimized structure; (h and i) optimized structures of 1 ML O–Cu(100); (j) experimental STM images of Cu(110). (Reproduced with permission from ref. 81, Copyright 2014 Elsevier.); (k) the simulated STM images based on the globally optimized 1 ML O–Cu(110). (i–m) The obtained globally optimized surface structures of 1 ML O–Cu(110) through the HEA approach. (n) HRTEM image of reconstructed Cu(110) layers under O_2 pressure (reproduced with permission from ref. 82, Copyright 2022 American Chemical Society).

In Fig. 4(a), compared with the previously proposed O–Pt(111) model by Hawkins *et al.*,[74] our optimized structure is 0.14 eV per O-site lower. Although Hawkins *et al.*'s model (Fig. S3†) also contained PtO_4 units, the chain structures that are linked with each PtO_4 unit are not as stable as the separated PtO_4–PtO_4 structures we obtained. Note there is no known criterion guaranteeing that the "best structure" encountered by a GO algorithm is the "true" global optimum.[38] Nevertheless, a lower energy allows our obtained structure to have a greater possibility to be the most abundant and representative phase under reaction conditions.[22]

Pd and Cu are widely used catalysts but both suffer from severe reconstruction under the reaction conditions.[83–85] A typical characteristic for oxidized Pd(111) is the "Persian-carpet" pattern observed in STM, with which the simulated STM image based on the optimized structure is well consistent, as shown in Fig. 5b. Fig. 5c–f show that the optimized structure consists of several parallel chains with different features from the PdO_2 to the PdO_4 unit. Such multiple co-existing oxygen species have been observed by *in situ* XPS studies.[86] Our optimized structure is 0.35 eV per O-site lower than the structure searched by USPEX as shown in Fig. S4,† and 0.61 eV per O-site lower than CALYPSO as reported previously.[87] Compared with the CALYPSO model, we both presented the existence of subsurface O atoms that have been experimentally reported,[88] while the CALYPSO model failed to further contain parallel chains that form a "Persian-carpet" pattern.

Fig. 5(h and i) show that the oxidized Cu(100) forms *via* the creation of the Cu–O chain that resembles the bulk Cu_2O.[89] Experimentally, a Cu_2O signal has been observed during the initial oxide growth of Cu(100)[90] along with the missing-row reconstruction (MRR) in Fig. 5(g),[82,91,92] which is consistent with the simulated STM image based on our optimized structure. Such MRR is reported widely for Cu oxidation, which is caused by the increasing surface stress and has been previously proven to be energetically favorable.[82,91] Subsurface oxygen is also contained in our structure, linking with the Cu–O chain through O–Cu–O units, which is believed to form above 0.5 ML O coverage experimentally,[92,93] and to contribute to the increased stability of the MRR structure.[94] Our optimized structure is 0.31 eV per atom more stable than the well-known $(2\sqrt{2} \times \sqrt{2})R45°$ reconstructed Cu(100) model[80,95] that also presents an MRR structure but fails to reflect the experimentally observed formation of Cu_2O.[96,97] Similarly, the oxidized Cu(110) also consists of the parallel, added-row Cu–O chains, whose simulated images are consistent with both experimental STM and TEM observation as shown in Fig. 5(j–n).[81,82,98]

Conclusions

In summary, we present a new strategy for global structural optimization using a HEA that combines DE and GA with a "tribe" framework. This algorithm combines the ability of the GA to explore the PES and the ability of the DE to exploit the PES. In practice, the HEA program performs better than well-established methods for optimizing surface systems. The high efficiency stems from the newly introduced DE operators that are effective in generating lower energy configurations, an efficient GP regressor that expedites the identification of low-lying structures, and a multi-tribe framework that maintains a high structural diversity. We demonstrate the efficacy of the HEA

method in obtaining the complex surface oxide structure of different facets of Pt, Pd, and Cu. The optimized structures are lower than previously reported models and are consistent with experimental observation. The newly proposed HEA program may open a new avenue for the study of the complex reconstruction of heterogeneous catalysts under reaction conditions.

Data availability

The data that support the findings of this study are available within the article and its ESI,† or from the corresponding author on reasonable request.

Author contributions

Xiangcheng Shi: methodology; investigation; visualization; formal analysis; validation; writing – original draft. Dongfang Cheng: methodology; investigation; formal analysis; writing – original draft. Ran Zhao: methodology; formal analysis. Gong Zhang: investigation; formal analysis. Shican Wu: methodology; formal analysis; Shiyu Zhen: visualization; formal analysis; Zhi-Jian Zhao: supervision; writing – review & editing; resources and funding acquisition. Jinlong Gong: supervision; formal analysis; writing – review & editing; resources and funding acquisition.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

We acknowledge the National Key R&D Program of China (2021YFA1500704), the National Natural Science Foundation of China (No. 22121004, U22A20409), the Haihe Laboratory of Sustainable Chemical Transformations, the Program of Introducing Talents of Discipline to Universities (BP0618007) and the XPLORER PRIZE for financial support. We also acknowledge the generous computing resources at the High Performance Computing Center of Tianjin University.

References

1 J. Zhang, Q. Xu, J. Wang, Y. Hu, H. Jiang and C. Li, *Sci. China Mater.*, 2022, **66**, 634–640.
2 X.-M. Lin, X.-T. Yang, H.-N. Chen, Y.-L. Deng, W.-H. Chen, J.-C. Dong, Y.-M. Wei and J.-F. Li, *J. Energy Chem.*, 2023, **76**, 146–164.
3 R. Ge, J. Li and H. Duan, *Sci. China Mater.*, 2022, **65**, 3273–3301.
4 B. Liu, G. Liu, Y. Tang and H.-M. Cheng, *Sci. China Mater.*, 2022, **65**, 3187–3189.
5 K. Qian, Z. Yu, Y. Liu, D. J. Gosztola, R. E. Winans, L. Cheng and T. Li, *J. Energy Chem.*, 2022, **70**, 340–346.
6 B. Zugic, L. Wang, C. Heine, D. N. Zakharov, B. A. J. Lechner, E. A. Stach, J. Biener, M. Salmeron, R. J. Madix and C. M. Friend, *Nat. Mater.*, 2016, **16**, 558–564.
7 Z. Ma, J. Li and T. Ling, *Trans. Tianjin Univ.*, 2022, **28**, 193–198.
8 W. Guo, H. Luo, D. Fang, Z. Jiang, J. Chi and W. Shangguan, *J. Energy Chem.*, 2022, **70**, 373–381.
9 J. Cai, Z. Yang, X. Zhou, B. Wang, A. Suzana, J. Bai, C. Liao, Y. Liu, Y. Chen, S. Song, X. Zhang, L. Wang, X. He, X. Meng, N. Karami, B. Ali Shaik Sulaiman, N. A. Chernova, S. Upreti, B. Prevel, F. Wang and Z. Chen, *J. Energy Chem.*, 2023, **85**, 126–136.
10 M. Behrens, F. Studt, I. Kasatkin, S. Kühl, M. Hävecker, F. Abild-Pedersen, S. Zander, F. Girgsdies, P. Kurr, B.-L. Kniep, M. Tovar, R. W. Fischer, J. K. Nørskov and R. Schlögl, *Science*, 2012, **336**, 893–897.
11 F. Tao, M. E. Grass, Y. Zhang, D. R. Butcher, J. R. Renzas, Z. Liu, J. Y. Chung, B. S. Mun, M. Salmeron and G. A. Somorjai, *Science*, 2008, **322**, 932–934.
12 S. Liu, C. Yang, S. Zha, D. Sharapa, F. Studt, Z. J. Zhao and J. Gong, *Angew. Chem., Int. Ed.*, 2021, **61**, e202109027.
13 Z. Tan, Y. Li, X. Xi, S. Jiang, X. Li, X. Shen, P. Zhang and Z. He, *Nano Res.*, 2023, **16**, 4950–4960.
14 Q. Gan, N. Qin, Z. Li, S. Gu, K. Liao, K. Zhang, L. Lu, Z. Xu and Z. Lu, *Nano Res.*, 2022, **16**, 513–520.
15 A. Grimaud, A. Demortière, M. Saubanère, W. Dachraoui, M. Duchamp, M.-L. Doublet and J.-M. Tarascon, *Nat. Energy*, 2016, **2**, 16189.
16 Y.-R. Zheng, J. Vernieres, Z. Wang, K. Zhang, D. Hochfilzer, K. Krempl, T.-W. Liao, F. Presel, T. Altantzis, J. Fatermans, S. B. Scott, N. M. Secher, C. Moon, P. Liu, S. Bals, S. Van Aert, A. Cao, M. Anand, J. K. Nørskov, J. Kibsgaard and I. Chorkendorff, *Nat. Energy*, 2021, **7**, 55–64.
17 A. F. Lee, C. V. Ellis, J. N. Naughton, M. A. Newton, C. M. A. Parlett and K. Wilson, *J. Am. Chem. Soc.*, 2011, **133**, 5724–5727.
18 F. Polo-Garzon, Z. Bao, X. Zhang, W. Huang and Z. Wu, *ACS Catal.*, 2019, **9**, 5692–5707.
19 T. Fuchs, J. Drnec, F. Calle-Vallejo, N. Stubb, D. J. S. Sandbeck, M. Ruge, S. Cherevko, D. A. Harrington and O. M. Magnussen, *Nat. Catal.*, 2020, **3**, 754–761.
20 D. J. Miller, H. Öberg, S. Kaya, H. Sanchez Casalongue, D. Friebel, T. Anniyev, H. Ogasawara, H. Bluhm, L. G. M. Pettersson and A. Nilsson, *Phys. Rev. Lett.*, 2011, **107**, 195502.
21 J. Hao, S. Xie, Q. Huang, Z. Ding, H. Sheng, C. Zhang and J. Yao, *CCS Chem.*, 2022, 1–13.
22 J. Zhang and V. A. Glezakou, *Int. J. Quantum Chem.*, 2020, **121**, e26553.
23 M. Jäger, R. Schäfer and R. L. Johnston, *Adv. Phys.: X*, 2018, **3**, 1516514.
24 M. Khatun, R. S. Majumdar and A. Anoop, *Front. Chem.*, 2019, **7**, 644.
25 L. Grajciar, C. J. Heard, A. A. Bondarenko, M. V. Polynski, J. Meeprasert, E. A. Pidko and P. Nachtigall, *Chem. Soc. Rev.*, 2018, **47**, 8307–8348.
26 E. Musa, F. Doherty and B. R. Goldsmith, *Curr. Opin. Chem. Eng.*, 2022, **35**, 100771.
27 M. Sierka, *Prog. Surf. Sci.*, 2010, **85**, 398–434.

28 L. R. Merte, M. K. Bisbo, I. Sokolović, M. Setvín, B. Hagman, M. Shipilin, M. Schmid, U. Diebold, E. Lundgren and B. Hammer, *Angew. Chem., Int. Ed.*, 2022, **61**, e202204244.
29 D. M. Deaven and K. M. Ho, *Phys. Rev. Lett.*, 1995, **75**, 288–291.
30 A. O. Lyakhov, A. R. Oganov and M. Valle, *Comput. Phys. Commun.*, 2010, **181**, 1623–1632.
31 D. V. Gruznev, L. V. Bondarenko, A. V. Matetskiy, A. Y. Tupchaya, E. N. Chukurov, C. R. Hsing, C. M. Wei, S. V. Eremeev, A. V. Zotov and A. A. Saranin, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2015, **92**, 245407.
32 J. P. Chou, C. M. Wei, Y. L. Wang, D. V. Gruznev, L. V. Bondarenko, A. V. Matetskiy, A. Y. Tupchaya, A. V. Zotov and A. A. Saranin, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2014, **89**, 155310.
33 D. V. Gruznev, L. V. Bondarenko, A. Y. Tupchaya, A. A. Yakovlev, A. N. Mihalyuk, A. V. Zotov and A. A. Saranin, *Surf. Sci.*, 2018, **669**, 183–188.
34 D. V. Gruznev, S. V. Eremeev, L. V. Bondarenko, A. Y. Tupchaya, A. A. Yakovlev, A. N. Mihalyuk, J.-P. Chou, A. V. Zotov and A. A. Saranin, *Nano Lett.*, 2018, **18**, 4338–4345.
35 Q. Wang, A. R. Oganov, Q. Zhu and X.-F. Zhou, *Phys. Rev. Lett.*, 2014, **113**, 266101.
36 H. A. Zakaryan, A. G. Kvashnin and A. R. Oganov, *Sci. Rep.*, 2017, **7**, 10357.
37 A. G. Kvashnin, D. G. Kvashnin and A. R. Oganov, *Sci. Rep.*, 2019, **9**, 14267.
38 B. C. Revard, W. W. Tipton and R. G. Hennig, in *Prediction and Calculation of Crystal Structures*, 2014, ch. 489, pp. 181–222, DOI: **10.1007/128_2013_489**.
39 X. Shi, X. Lin, R. Luo, S. Wu, L. Li, Z.-J. Zhao and J. Gong, *JACS Au*, 2021, **1**, 2100–2120.
40 A. O. Lyakhov, A. R. Oganov, H. T. Stokes and Q. Zhu, *Comput. Phys. Commun.*, 2013, **184**, 1172–1182.
41 M. S. Jørgensen, U. F. Larsen, K. W. Jacobsen and B. Hammer, *J. Phys. Chem. A*, 2018, **122**, 1504–1509.
42 G. Cheng, X.-G. Gong and W.-J. Yin, *Nat. Commun.*, 2022, **13**, 1492.
43 Y. Zuo, M. Qin, C. Chen, W. Ye, X. Li, J. Luo and S. P. Ong, *Mater. Today*, 2021, **51**, 126–135.
44 A. Palizhati, W. Zhong, K. Tran, S. Back and Z. W. Ulissi, *J. Chem. Inf. Model.*, 2019, **59**, 4742–4749.
45 M. Yao, J. Ji, X. Li, Z. Zhu, J.-Y. Ge, D. J. Singh, J. Xi, J. Yang and W. Zhang, *Sci. China Mater.*, 2023, **66**, 2768–2776.
46 H. Li, Y. Jiao, K. Davey and S.-Z. Qiao, *Angew. Chem., Int. Ed.*, 2023, **62**, e202216383.
47 S. Kaappa, E. G. del Río and K. W. Jacobsen, *Phys. Rev. B*, 2021, **103**, 174114.
48 S. Hajinazar, E. D. Sandoval, A. J. Cullo and A. N. Kolmogorov, *Phys. Chem. Chem. Phys.*, 2019, **21**, 8729–8742.
49 C. J. Pickard and R. J. Needs, *J. Phys.: Condens. Matter*, 2011, **23**, 053201.
50 H. Huo and M. Rupp, *Mach. learn.: sci. technol.*, 2022, **3**, 045017.
51 L. B. Vilhelmsen and B. Hammer, *J. Chem. Phys.*, 2014, **141**, 044711.
52 T.-E. Fan, G.-F. Shao, Q.-S. Ji, J.-W. Zheng, T.-d. Liu and Y.-H. Wen, *Comput. Phys. Commun.*, 2016, **208**, 64–72.
53 L. Chanussot, A. Das, S. Goyal, T. Lavril, M. Shuaibi, M. Riviere, K. Tran, J. Heras-Domingo, C. Ho, W. Hu, A. Palizhati, A. Sriram, B. Wood, J. Yoon, D. Parikh, C. L. Zitnick and Z. Ulissi, *ACS Catal.*, 2021, **11**, 6059–6072.
54 A. Sriram, A. Das, B. M. Wood, S. Goyal and C. L. Zitnick, *arXiv*, 2022, preprint, arXiv:2203.09697, DOI: **10.48550/arXiv.2203.09697**.
55 J. R. Gardner, G. Pleiss, D. Bindel, K. Q. Weinberger and A. G. Wilson, *arXiv*, 2018, preprint, arXiv:1809.11165, DOI: **10.48550/arXiv.1809.11165**.
56 L. Himanen, M. O. J. Jäger, E. V. Morooka, F. Federici Canova, Y. S. Ranawat, D. Z. Gao, P. Rinke and A. S. Foster, *Comput. Phys. Commun.*, 2020, **247**, 106949.
57 C. E. Rasmussen and C. K. I. Williams, *Gaussian processes for machine learning*, MIT Press, Cambridge, Mass, 2006.
58 G. N. Simm and M. Reiher, *J. Chem. Theory Comput.*, 2018, **14**, 5238–5248.
59 M. K. Bisbo and B. Hammer, *Phys. Rev. Lett.*, 2020, **124**, 086102.
60 R. Moriconi, K. S. S. Kumar and M. P. Deisenroth, *Optim. Lett.*, 2019, **14**, 51–64.
61 S. Manzhos and M. Ihara, *J. Chem. Phys.*, 2023, **158**, 044111.
62 B. Shahriari, K. Swersky, Z. Wang, R. P. Adams and N. de Freitas, *Proc. IEEE*, 2016, **104**, 148–175.
63 J. M. Hernandez-Lobato, J. Requeima, E. O. Pyzer-Knapp and A. Aspuru-Guzik, *Proc. Mach. Learn. Res.*, 2017, **70**, 1470–1479.
64 M. Arrigoni and G. K. H. Madsen, *Npj Comput. Mater.*, 2021, **7**, 71.
65 S.-D. Huang, C. Shang, X.-J. Zhang and Z.-P. Liu, *Chem. Sci.*, 2017, **8**, 6327–6337.
66 V. L. Deringer, A. P. Bartók, N. Bernstein, D. M. Wilkins, M. Ceriotti and G. Csányi, *Chem. Rev.*, 2021, **121**, 10073–10141.
67 A. Banerjee, D. B. Dunson and S. T. Tokdar, *Biometrika*, 2012, **100**, 75–89.
68 L. Grajciar, C. J. Heard, A. A. Bondarenko, M. V. Polynski, J. Meeprasert, E. A. Pidko and P. Nachtigall, *Chem. Soc. Rev.*, 2018, **47**, 8307–8348.
69 P. Zhang, T. Wang and J. Gong, *CCS Chem.*, 2023, **5**, 1028–1042.
70 S. Liu, J. Zong, Z.-J. Zhao and J. Gong, *GreenChE*, 2020, **1**, 56–62.
71 L. Li, T. Liu, Z. Zhou, P. Guo, X. Li and S. Wu, *Sci. China Mater.*, 2022, **65**, 3033–3042.
72 H. Kang, Y. Zhang, Y. Wu, S. Hu, J. Li, Z. Chen, Y. Sui, S. Wang, S. Zhao, R. Xiao, G. Yu, S. Peng, Z. Jin and X. Liu, *Sci. China Mater.*, 2022, **65**, 2763–2770.
73 Z. Duan and G. Henkelman, *ACS Catal.*, 2021, **11**, 14439–14447.
74 J. M. Hawkins, J. F. Weaver and A. Asthagiri, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2009, **79**, 125434.

75 S. P. Devarajan, J. A. Hinojosa and J. F. Weaver, *Surf. Sci.*, 2008, **602**, 3116–3124.
76 E. F. Holby, J. Greeley and D. Morgan, *J. Phys. Chem. C*, 2012, **116**, 9942–9946.
77 M. A. van Spronsen, J. W. M. Frenken and I. M. N. Groot, *Nat. Commun.*, 2017, **8**, 429.
78 D. Friebel, D. J. Miller, C. P. O'Grady, T. Anniyev, J. Bargar, U. Bergmann, H. Ogasawara, K. T. Wikfeldt, L. G. M. Pettersson and A. Nilsson, *Phys. Chem. Chem. Phys.*, 2011, **13**, 262–266.
79 G. Zheng and E. I. Altman, *Surf. Sci.*, 2000, **462**, 151–168.
80 H. Mönig, M. Todorović, M. Z. Baykara, T. C. Schwendemann, L. Rodrigo, E. I. Altman, R. Pérez and U. D. Schwarz, *ACS Nano*, 2013, **7**, 10233–10244.
81 Q. Liu, L. Li, N. Cai, W. A. Saidi and G. Zhou, *Surf. Sci.*, 2014, **627**, 75–84.
82 M. Li, M. T. Curnan, W. A. Saidi and J. C. Yang, *Nano Lett.*, 2022, **22**, 1075–1082.
83 H. Over and A. P. Seitsonen, *Science*, 2002, **297**, 2003–2005.
84 G. Jiang, D. Han, Z. Han, J. Gao, X. Wang, Z. Weng and Q.-H. Yang, *Transactions of Tianjin University*, 2022, **28**, 265–291.
85 B. Deng, X. Zhao, Y. Li, M. Huang, S. Zhang and F. Dong, *Sci. China Chem.*, 2022, **66**, 78–95.
86 D. Zemlyanov, B. Aszalos-Kiss, E. Kleimenov, D. Teschner, S. Zafeiratos, M. Hävecker, A. Knop-Gericke, R. Schlögl, H. Gabasch, W. Unterberger, K. Hayek and B. Klötzer, *Surf. Sci.*, 2006, **600**, 983–994.
87 T. Jin, F. Chen, L. Guo, Q. Tang, J. Wang, B. Pan, Y. Wu and S. Yu, *J. Phys. Chem. C*, 2021, **125**, 19497–19508.
88 D. L. Weissman-Wenocur, M. L. Shek, P. M. Stefan, I. Lindau and W. E. Spicer, *Surf. Sci.*, 1983, **127**, 513–525.
89 L. Li, X. Mi, Y. Shi and G. Zhou, *Phys. Rev. Lett.*, 2012, **108**, 176101.
90 S. Kunze, L. C. Tănase, M. J. Prieto, P. Grosse, F. Scholten, L. de Souza Caldas, D. van Vörden, T. Schmidt and B. R. Cuenya, *Chem. Sci.*, 2021, **12**, 14241–14253.
91 T. Kangas and K. Laasonen, *Surf. Sci.*, 2012, **606**, 192–201.
92 K. Lahtonen, M. Hirsimäki, M. Lampimäki and M. Valden, *J. Chem. Phys.*, 2008, **129**, 124703.
93 M. Lampimäki, K. Lahtonen, M. Hirsimäki and M. Valden, *J. Chem. Phys.*, 2007, **126**, 034703.
94 W. A. Saidi, M. Lee, L. Li, G. Zhou and A. J. H. McGaughey, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2012, **86**, 245429.
95 S. J. Tjung, Q. Zhang, J. J. Repicky, S. F. Yuk, X. Nie, N. M. Santagata, A. Asthagiri and J. A. Gupta, *Surf. Sci.*, 2019, **679**, 50–55.
96 M. Lee and A. J. H. McGaughey, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2011, **83**, 165447.
97 J. Cao, A. Rinaldi, M. Plodinec, X. Huang, E. Willinger, A. Hammud, S. Hieke, S. Beeg, L. Gregoratti, C. Colbea, R. Schlögl, M. Antonietti, M. Greiner and M. Willinger, *Nat. Commun.*, 2020, **11**, 3554.
98 Y. Li, H. Chen, W. Wang, W. Huang, Y. Ning, Q. Liu, Y. Cui, Y. Han, Z. Liu, F. Yang and X. Bao, *Nano Res.*, 2020, **13**, 1677–1685.