Replacing Chemical Intuition by Machine Learning: a Mixed Design of Experiments - Reinforcement Learning Approach to the Construction of Training Sets for Model Hamiltonians

Ruben Staub* and Stephan N. Steinmann*

*Univ Lyon, Ecole Normale Supérieure de Lyon, CNRS, Laboratoire de Chimie UMR 5182,
46 allée d'Italie, F-69364, LYON, France*

E-mail: ruben.staub@ens-lyon.fr; stephan.steinmann@ens-lyon.fr

Abstract

Model Hamiltonians based on the so-called cluster expansion (CE), which consist of a linear fit of parameters corresponding to geometric patterns, provide an efficient and rigorous means to quickly evaluate the energy of diverse arrangements of adsorbate mixtures on reactive surfaces as typically relevant for heterogeneous catalysis. However, establishing the model Hamiltonian is a tedious task, requiring the construction and optimization of many geometries. Today, most of these geometries are constructed by hand, based on chemical intuition or random choices. Hence, the quality of the training set is unlikely to be optimal and its construction is not reproducible. Herein, we propose a reformulation of the construction of the training set as a strategy-based game, aiming at an efficient exploration of the relevant patterns constituting the

model Hamiltonian. Based on this reformulation, we exploit a typical active learning solution for machine-learning such a strategy game: an upper confidence tree (UCT) based framework. However, in contrast to standard games, evaluating the true score is computationally expensive, as it requires a costly geometry optimization. Hence, we augment the UCT with a pre-exploration step inspired by the variance-based Design of Experiments (DoE) methods. This novel mixed UCT+DoE framework allows to automatically construct a well adapted training set, minimizing computational cost and user-intervention. As a proof of principle, we apply our UCT+DoE approach on the CO oxidation reaction on Pd(111), for which a relevant model Hamiltonian has been established previously. The results demonstrate the effectiveness of the custom built UCT and its significant benefits on a DoE-based approach.

1 Introduction

Many surface science and heterogeneous catalysis experiments benefit from complementary computational insights.^{1,2} Typically, the identification of the relevant atomistic surface state (which adsorbates are present and how are they arranged) is determined either via combined insights from theory and experiments (e.g., scanning-tunneling microscopy, interpreted by density functional theory (DFT) computations)³⁻⁶ or even by computations only.^{7,8} While DFT is the most general, widely applicable tool, it is computationally costly, limiting the number of geometry optimizations that are typically performed to the order of magnitude of hundreds. However, to explore the complex energy landscape of mixtures of adsorbates, as typical for heterogeneous catalytic processes, one to two orders of magnitude more energy evaluations are necessary. The most emblematic method to explore the atomistic details of catalytic reactions under realistic conditions is kinetic Monte Carlo (KMC).^{9,10} Even though the framework is not limited to reactions occurring on well-ordered surfaces, we implicitly assume throughout this article that the catalyst can be described by a graphical lattice, where geometries corresponding to energy minima are well defined and separated (i.e. the

adsorbates are preferentially located at active sites) and the relevant geometry space can therefore be discretized. Note, that this assumption in practice restricts the application to "rigid" adsorbates and that exploring the accessible phase-space of flexible adsorbates requires other algorithms.¹¹ KMC allows to solve differential equations describing the temporal evolution of a catalytic system at the atomistic level. This contrasts with traditional microkinetic models, such as the Langmuir-Hinshelwood model,^{12,13} which express the reaction rates in the spirit of a mean-field approximation, so that the analytical expressions can be integrated numerically. The atomistic level is, however, indispensable for accounting for the so called "lateral interactions", i.e., the coverage effects. Typically, DFT computations for a reaction pathway are carried out at "low" coverage, i.e., only one adsorbate is present in the unit cell, which is large enough that the periodic images barely influence each other. Under "real" reaction conditions, however, species tend to accumulate on the surface, leading to mostly repulsive interactions, depending on the precise arrangements.^{8,14–17} In specific cases, particularly relevant for water, hydrogen bonds can lead to stabilizing lateral interactions.¹⁸ These surface configurations are explored in KMC, leading to a more realistic description of the system,^{19,20} putting the lateral interactions on the forefront of interest. This situation is closely related to the phase diagrams of alloys.^{21–24} Indeed, in both fields, phase-space sampling heavily relies on the use of model Hamiltonians that are fitted to DFT data. In the present context, model Hamiltonians are particularly well suited to explore the phase-space of adsorbates on catalytic surfaces. In contrast to alloys, the "stoichiometry" of the surface state is evolving with time, so that one should not focus on the most stable arrangement but explore the potential energy landscape as exhaustively as possible.

A model Hamiltonian (or effective Hamiltonian) is a model linking the energy of a system to its geometrical features.^{8,23,24} The simplest yet rigorous model Hamiltonian for lattice-based surfaces is a linear model based on a Cluster Expansion^{25–28} (CE). The CE-based model Hamiltonian, is typically built as a linear relation between the total adsorption energy of a surface configuration and the number and identity of the adsorbates (called single-body

terms), supplemented by the number of occurring lateral interactions among the adsorbates, i.e., the many-body terms. Using a lattice-based representation of the surface, the types of lateral interactions become enumerable (since the positions are discretized) but infinitely many distinct types of lateral interactions can be defined for periodic systems. Therefore, a truncated CE is used in practice,²⁹ by considering only the lateral interactions that display significant energy contributions, typically short-ranged many-body patterns. The single-body terms, which are, of course, included as well, are usually directly extracted from DFT computations at low coverage. In other words, a CE-based model Hamiltonian is a linear model whose fitted parameters are the significant energy contributions of the corresponding relevant lateral interactions. Identifying and estimating these energy contributions from first-principle calculations is essential in order to set up the CE.

Like any empirical model, a CE-based model Hamiltonian must be trained on a set of empirical data. The training set of a CE-based model Hamiltonian must be composed of reaction or system specific geometries. Yet this training set is usually generated by the chemist,³⁰ selecting around a few hundreds hand-picked geometries using the personal intuition, and/or trying to sample geometries at pseudo-random/exhaustively, making this process time consuming and somewhat arbitrary.

In the recent years, some efforts were made toward optimizing the selection of relevant lateral interactions to include in the CE-based model Hamiltonian (i.e. optimizing the choice of parameters).^{31–35} Furthermore, optimizing the training set via variance-based Design of Experiments integrated in the dual optimization of both the terms and the geometries in a greedy approach has been proposed.^{30,36} These approaches were mainly designed for CEs that are updated on-the-fly for predicting the ground-state of solid solutions.³⁷ To our knowledge, the *a priori* selection of geometries for constructing an adequate training set for KMC or similar applications (for which non-equilibrium surface configurations are important) has not yet been properly discussed in view of the development of reinforcement learning based active learning schemes.^{38–40} These machine-learning frameworks open the possibility to optimize

the selection of geometries to include in the training set *before* the computationally expensive DFT computations.

In this contribution, we address the construction of an appropriate training set by proposing an active learning framework. Our UCT+DoE active learning algorithm combines Reinforcement Learning techniques (i.e. a Upper Confidence Tree) with classical variance-based Design of Experiments³⁸. Note that reinforcement learning is a sub-field of machine learning, along with supervised and unsupervised learning,⁴¹ and is getting progressively adopted by chemists, as illustrated by the recent application to approach setting up machine-learned potentials,⁴² the full configuration interaction problem⁴³ or discovering reaction paths.⁴⁴ Our UCT+DoE framework automates the construction of a CE-based model Hamiltonian training set from scratch, achieving a final training set that is consistently better designed than from a pseudo-random sampling, while reducing the computational cost.

The manuscript is organized as follows: First, we introduce the cluster expansion, together with the linear model Hamiltonian. Second, we reformulate the design of a training set as a strategy game, followed by the presentation of our active learning scheme. Finally, as a proof of principle, we apply our UCT+DoE framework to the CO oxidation model Hamiltonian over Pd(111).

2 Cluster Expansion Hamiltonian

The Cluster Expansion (CE) Hamiltonian²⁵⁻²⁸ is a linear model Hamiltonian based on lateral interactions, here for predicting the adsorption energy of a set of adsorbates. This linear model aims at reproducing the energy $E(g_i)$ (the target function, or regressand; here given by the DFT-based adsorption energy) of a geometry g_i , as a linear combination of the number of its input features Γ_i (called regressors):

$$E(g_i) = \Gamma_i^\top \beta + \epsilon_i, \tag{1}$$

where ϵ_i is the residual error. In a CE Hamiltonian, the regressors Γ are the number (or multiplicity) of times each possible lateral interaction is found in the geometry. Since the lateral interactions are simple graphs, the model Hamiltonian β provides with a simple and fast energy predictor. The parameters β of the model are, therefore, the energy contribution of each lateral interaction. These interactions and their contributions are seen as corrective terms of increasing order to the total energy:

$$E(g_i) = \underbrace{E^{(1)}(g_i)}_{\text{1-body model}} + \underbrace{\sum_{n=2}^{N_{\text{ads}}} E^{(n)}(g_i)}_{\text{many-body correction terms}}, \quad (2)$$

where N_{ads} is the number of adsorbates in g_i and $E^{(n)}(g_i)$ is the adsorption energy contributions of the n -body effects only (see section S1 in SI for further details of the n-body expansion).

In practice, only a truncated version of this model is considered,²⁹ typically up to the 2 or 3-body terms.²⁰ Fortunately, higher order corrections tend to fade out quickly. Besides, a cutoff is usually used to discard lateral interactions where the adsorbates are too far away to significantly "feel" each other, in agreement with Kohn's "nearsightedness of matter".⁴⁵ This truncation makes the number of terms to be determined tractable but is also at the origin of the residual error ϵ_i of Eq. 1.

Applying an ordinary least squares method to Eq. 1 allows to retrieve an unbiased, consistent and efficient estimator⁴⁶ $\hat{\beta}$ for the energy contributions of the included lateral interactions. Furthermore, the uncertainty on the parameters $\hat{\beta}$ is given by the covariance matrix $\text{Var}(\hat{\beta})$. Assuming the observation matrix $\mathbf{\Gamma}^\top$ has full column-rank, i.e., that the full regressor-space is spanned by the observations, $\text{Var}(\hat{\beta})$ can be estimated⁴⁶ via:

$$\text{Var}(\hat{\beta}) \approx s^2(\mathbf{\Gamma}^\top \mathbf{\Gamma})^{-1}, \quad (3)$$

where $s^2 = \frac{\|\mathbf{E} - \mathbf{\Gamma}^\top \hat{\beta}\|^2}{n-r}$ is an estimator of the variance of the noise σ^2 , with $\mathbf{E} = \begin{pmatrix} E_1 & \dots & E_n \end{pmatrix}^\top$,

n the number of observations and r the rank of $\mathbf{\Gamma}$ so that $n - r$ is the number of linearly redundant observations.

However, in the general case and especially during the training of a CE Hamiltonian, the number of observations is likely to be less than the number of regressors, so that one must deal with a underdetermined system. Since any linear system of equation can be decomposed into a purely non-overdetermined system and a purely non-underdetermined system,⁴⁷ it is nevertheless possible to define the unique maximum subspace $\text{Vect}(\mathbf{\Gamma}_{over})$ onto which a model Hamiltonian is overdetermined. As a consequence, it is possible to estimate a covariance matrix embedded unto that subspace $\text{Var}(\hat{\beta})_{empirical}$, using equation 3. Concerning the rest of the regressors space, the user-defined variance of the guessed parameters $\text{Var}(\beta_{guess})$ can be used onto that space. Combining these two terms (assuming that guessed parameters are uncorrelated with empirically estimated parameters) provides a covariance matrix estimator $\text{Var}(\hat{\beta}) = \text{Var}(\hat{\beta})_{empirical} + \text{Var}(\beta_{guess})$ that can be defined even on a non purely overdetermined model Hamiltonian.

Thus it is possible to estimate the uncertainty associated with the prediction of the energy E_j on a given input Γ_j (called prediction error):

$$\begin{aligned} \text{Var}(E_j - \Gamma_j^\top \hat{\beta}) &= \text{Var}(\Gamma_j^\top \beta + \epsilon_j - \Gamma_j^\top \hat{\beta}) \\ &= \sigma^2 + \Gamma_j^\top \text{Var}(\hat{\beta}) \Gamma_j \\ &\approx s^2 + \Gamma_j^\top \text{Var}(\hat{\beta}) \Gamma_j \end{aligned} \tag{4}$$

More generally, $\text{Var}(\hat{\beta})$ represents the inverse Fisher information matrix and thus informs on the quality of the CE Hamiltonian.

3 Presentation of the problem and the adopted overall strategy

Building a model Hamiltonian requires two main components: a functional form for the effective Hamiltonian and a training dataset to fit the model. Regarding the first component, we place our study in the context of a cluster expansion (CE) Hamiltonian for describing the many-body effects. The main focus of our study is related to the second ingredient, i.e., building the most appropriate training from scratch in an automated and reproducible way. The developed framework is general and is, thus, not strongly linked to the first component. In other words, instead of training a CE Hamiltonian, one could construct a training set for semi-empirical methods, neural networks or similar approaches⁴⁸ in an analogous way.

3.1 Definition of an appropriate training set

We define an appropriate training set as a set of geometries verifying three criteria:

- **Diverse:** The model Hamiltonian has not only to be able to describe the most likely events, but also rare events. Therefore, an appropriate training set should contain a diverse set of geometries. Ideally, such sampling should roughly cover the whole accessible chemical space. Mathematically, we require: any possible geometry should have a non-zero probability to be found in an adequate training set. It is up to the user to define the limits of the accessible chemical space. Typically, the user chooses a limited number of possible adsorbates and the use of a lattice-based model precludes surface reconstructions, multi-layer adsorbate films etc.
- **Chemically relevant:** The model Hamiltonian should better describe strong lateral interactions than negligible ones. This is translated mathematically as: the higher the absolute energy deviation from the 1-body model associated with a geometry, the more likely an adequate training set should contain this geometry. For a discussion of this

particular choice to measure the relevance of a geometry, see section S5 of the SI.

- **Reasonably-sized:** Including a new geometry in the training set typically requires to run a geometry optimization at a DFT level, with computational times in the order of hours. Hence, it is a computationally expensive task. Therefore, an appropriate training set should not contain geometries that can be removed without a major impact on the quality of the resulting model Hamiltonian.

In summary, building an appropriate training set requires a balance between focusing on groups of geometries that seem chemically relevant, exploring novel groups of geometries and keeping the overall computational cost down.

3.2 Reformulation of the training set construction as a game

In order to motivate our approach to the construction of an appropriate training set, let us describe the construction of a single geometry (also called configuration) to be included in the training set as a strategy 1-player turn-based board game "Build a configuration".

Figure 1 represents a synoptic illustration of the "Build a configuration" strategy-game features. The adsorbates are placed on the surface one at a time. The "board" is the lattice-representation of the surface, with the cells $\mathcal{S}_i$ representing the adsorption sites. Sites that are directly connected are called "adjacent" and, depending on the nature of the site (e.g., for a bridge site), relative orientation is encoded as well. Mathematically, a board is a node-colored and edge-colored graph whose nodes are sites. This game is played with a predefined set of pawns P_i representing the molecular adsorbate $\mathcal{M}_i$. Similarly to the sites, pawns can also have a relative orientation. Furthermore, pawns can span across multiple $\mathcal{S}_i$ in order to represent multidentate adsorbates. This game is played in N_{ads} turns, where N_{ads} is the target number of adsorbed molecular entities desired in the final configuration. Each turn the player decides which adsorbate is placed where and in which orientation on the board. At the end of the N_{ads} turns, the final configuration created is given a score. As mentioned

in section 3.1, the score represents the absolute total lateral interaction energy (i.e. the absolute deviation from the 1-body model) associated with the corresponding geometry. In other words, bonuses and penalties are granted for each many-body pattern present in the final configuration. Lateral interactions can be stabilizing or destabilizing and are *a priori* unknown to the player, so that they represent hidden score contributions whose sum is only revealed at the end of the game. Even though the scores are unknown to the player beforehand, playing the game over and over again, a strategy to maximize the final score can be developed.

Since evaluating the true, final score requires a DFT-based geometry optimization, "Build a configuration" is a peculiar game. Therefore, the UCT+DoE framework for model Hamiltonians introduced below relies on tweaks compared to established algorithms for learning strategy games to better fit the specific properties of the "Build a configuration" game.

3.3 Upper Confidence Trees for solving the adequate training set problem

"Build a configuration" and its trade-off between exploration (so far unseen lateral interactions) and exploitation (maximizing already identified lateral interactions) is reminiscent of the notorious multi-armed bandit problem.

3.3.1 Multi-armed bandit problem

The multi-armed bandit problem can be formulated as follows: at a casino, there are multiple armed bandit machines $(B_1, B_2, \dots, B_k)$. Each has its own average payout $\mathbb{E}(W_i)$ that is *a priori* unknown to the player. The aim is to find a playing strategy that has minimal cumulative regret (i.e. the total number of times a sub-optimal bandit has been played). Intuitively, the optimal strategy should maintain a balance between estimating the payout of bandits (i.e. exploration) and focusing on the apparently most profitable bandits (i.e. exploitation), with the latter evolving as a function of the exploration.

Figure 1: Synoptic illustration of the "Build a configuration" game. The parallel with heterocatalysis is highlighted by explicitly linking game features to surface science concepts. At each turn, the player chooses a pawn and places it on a free available cell. This is equivalent to the chemist placing adsorbates at non-occupied active sites on a surface. After N_{ads} turns, the occupied board (called configuration) represents an occupied surface with N_{ads} adsorbates. Each many-body pattern found on the final configuration earns points, in an amount that is based on the strength of the associated lateral interaction. These points can be positive or negative, but the individual details are not known to the player. The final score is the absolute value of the sum of these points (one could imagine some sort of game where the aim would be to be the best mythical creature, either "angel" by accumulating the most "good karma" points, or "demon" by accumulating the most "bad karma" points). This score of a final configuration is chosen to match the absolute adsorption energy deviation from the 1-body model.

For "Build a configuration" the bandit machines B_i are identified as possible groups G_i of similar geometries to be included in the training set. Hence, playing a bandit B_i means adding to the training set an arbitrarily chosen geometry $g_j \in G_i$ within the related group G_i . The associated payout W_i of playing B_i is taken as the absolute deviation from the 1-body model:

$$W_i = |E_j - E_j^{alone}| \quad (5)$$

where $E_j^{alone} = \sum_k E_k$ is the sum of the individual adsorption energies E_k for each adsorbate $\mathcal{M}_k$ present in g_j as if adsorbed alone at the surface.

3.3.2 The Upper Confidence Bound approach

The multi-armed bandit problem has been well studied in the literature, and the most efficient strategies are commonly based on Upper Confidence Bound (UCB) approaches.⁴⁹⁻⁵³

The general idea behind UCB-like approaches can seem intuitive: each bandit is given a score S_i :

$$S_i = \mathbb{E}(W_i)_{emp} + b_i(n_i, T) \quad (6)$$

where n_i is the number of times bandit B_i was played, T is the total number of plays (i.e. $\sum_i n_i = T$). $\mathbb{E}(W_i)_{emp}$ is the empirical estimator of the average payout

$$\mathbb{E}(W_i)_{emp} = \frac{\sum_j W_i}{n_i} \quad (7)$$

and $b_i(n_i, T)$ is called the exploration bias. The UCB strategy is then simply to repeatedly play the bandit B_i with the highest UCB score $S_i(T)$.

The main challenge lies in determining the exploration bias, which acts as an upper confidence interval for the empirical estimate of the payout. Finding an efficient functional form for the exploration bias is still an active field of research, even though many functions have been designed.⁵¹⁻⁵⁵ Probably the best known one comes from the UCB1 algorithm,⁵¹

with $b_i(n_i, T) = C\sqrt{\frac{\log(n_i)}{T}}$ where C is an input parameter acting as a confidence level. Since the organization of geometries into groups is not well-defined, a simple UCB approach is, however, not suitable for "Build a configuration".

3.3.3 Combining a Monte-Carlo Tree Search with the Upper Confidence Bound, leading to the Upper Confidence Tree construction of the training set

The number (and thus size) of the groups is critical for the interpolation power: the information gained from each play should be applicable to as large a group as possible, but might be completely irrelevant to some geometries, which therefore should be outside the optimal group. To achieve the concomitant exploration of the lateral interactions and the optimization of the corresponding groups, the UCB approach can be conveniently combined with a Monte-Carlo Tree Search (MCTS) scheme.⁴⁹

To be specific, instead of using a single arbitrary group size, a MCTS framework uses a recursive grouping. In our case, all geometries are organized in a tree structure: each single geometry is represented by a leaf, and each node $\mathcal{N}_i$ (a branch) represents a group of geometries G_i composed of all the geometries $g_j \in G_i$ associated with the leaves descending from that node. The more similar two geometries are, the closer their corresponding leaves in the tree structure are.

Selecting a leaf in a MCTS framework is done by a best-first search composed of the following steps:

- Selection/Expansion: Starting from the root node, the best child node is selected until an unexplored node (or a leaf corresponding to a single geometry) is reached. In a pure MCTS framework, each node $\mathcal{N}_i$ is evaluated using a score taken as the empirical average payout of its associated group of geometries $\mathbb{E}(W_i)_{emp}$ (see Eq. 7).
- Payout/rollout: Once an unexplored node $\mathcal{N}_i$ is selected, a geometry g_j is arbitrarily chosen from the associated group of geometries G_i .

- Backpropagation: The scores of all ancestors of $\mathcal{N}_i$ (i.e. all nodes selected during the selection/expansion step) are updated with the lateral interaction energy W_i (see Eq. 5) of the chosen geometry $g_j \in G_i$.

During the selection step, choosing the best child node can be seen as a typical multi-armed bandit problem (as described in section 3.3.1). Therefore, adding an upper confidence bound to the empirical average payout used as a score for choosing a best child node, i.e., $\mathbb{E}(W_i)_{emp} + b_i(n_i, T)$, is expected to be efficient and is known as an Upper Confidence Tree^{49,56} (UCT).

As a result, the selection step in MCTS can be seen as a successive selection over increasingly specific groups. In this manner, each selection step benefits in its early stages from coarse-grain sampling effects, while deeper in the tree the selection is performed over increasingly specific groups.

UCTs are a common tool of reinforcement learning used for designing optimal strategies and in particular for learning strategy-based games, where UCT-based approaches demonstrated spectacular performance.^{39,40,57} Hence, a UCT based approach seems well adapted to "Build a configuration".

In summary, the UCT framework is expected to be particularly relevant for solving our training set building problem: the diverse property is expected to be fulfilled since the whole geometry space is sampled, while the chemically relevant property is expected to be fulfilled by the enhanced sampling of regions associated with apparently larger many-body energy effects. Finally, the reasonably-sized property depends on the convergence rate of the chosen UCT framework. Given the successes of UCTs for strategy games, we expect an acceptable convergence rate for "Build a configuration".

4 Designing a UCT for model Hamiltonians: UCT+DoE

4.1 Patterns

Patterns have a crucial role when dealing with CE-based model Hamiltonians (where they are also sometimes called figures), as they represent lateral interactions. Their definition and identification are of utmost importance for building meaningful model Hamiltonians. In particular, careful considerations must be applied to recognize equivalent patterns, since lateral interactions are defined up to global translation, rotation and chirality (i.e. the energy of an isolated sub-system is invariant by translation, rotation and reflection). Our conventions and notations are extensively discussed in section S2 of the SI and graphically summarized in Fig. S1.

Patterns are rigorously defined by vertex-induced subgraphs of the abstract representation of the occupied surface, itself represented by a node-colored and edge-colored directed multigraph as detailed in the SI. However, such a definition is unpractical in terms of both manipulation/storage and comparison. Therefore, in practice, we rely on two simplifications: (1) when two occupied active sites are connected by multiple shortest paths, only one is considered per pattern (2) instead of considering the whole connectivity of a subgraph, a pattern is defined as a simple path (i.e. no node seen twice on the path), a simplification that only applies to patterns of 3 bodies and more. These simplifications reduce the number of patterns and, thus, extends the extrapolative power of the model Hamiltonian without loss of accuracy.

Our patterns are rigorously invariant to translations and rotations. To ensure that reflections, path reversals and potential symmetries inherited from the surface (e.g., hollow sites on a (111) fcc-metal surface have C_3 symmetry) and the adsorbates (e.g., CO has C_∞ symmetry) are correctly taken into account, the corresponding equivalent patterns are also generated. Similarly, as detailed in the SI, additional pattern multiplicities ought to be defined by the user (i.e. expert-based knowledge), so that all the equivalent patterns can be

merged.

From a technical point of view, a typical way to uniquely define equivalent patterns would be to consider only the minimum equivalent pattern, using a lexicographic order for example,⁵⁸ each time a pattern is identified. Such an approach becomes quickly computationally expensive, especially since pattern detection is constantly performed in a UCT+DoE run. Instead, we use a time-memory trade-off by defining a lookup table. In practice, we perform the exhaustive generation of equivalents only once per pattern. All of these equivalent patterns are then stored under the same unique identifier, leading to subsequent equivalency evaluations in $O(1)$ operations.

4.2 Basic UCT design

The UCT approach is highly generic, so applying it to a specific case requires to choose and tune implementation details.

4.2.1 Tree structure (modular organization)

One of the most crucial elements of an MCTS instance is undoubtedly the organization of the tree used. In other words, how the final configurations are organized in groups. Groups are meant to gather similar configurations, and the whole tree can be seen as a metric: the more similar two configurations are, the closer they should be on the tree (see section 3.3.3).

In the case of final configurations, the optimal *a priori* metric would be based on the sharing of similar lateral interactions. However, the partitioning of final configurations with respect to lateral interactions similarity is not a trivial task, since one would need to enumerate all possible configurations and identify all patterns beforehand, or at least privilege some lateral interactions to be more important. Instead, we implemented a partitioning inspired by game playing, where final configurations are partitioned by the first move (i.e. the first turn of a "Build a configuration" game), then sub-partitioned based on the second move, and so forth.

However, the number of possible moves per turn is quite large, so partitioning only on complete moves would lead to a large branching factor (i.e. the number of children per node) and a small maximal depth that would reduce the relevance of an MCTS approach. Inspired by move grouping,^{59,60} we subdivide each move into the following sub-partitions, in that order:

1. Chemical nature of the adsorbate (e.g. H₂O O-site, NH₃ N-site, ...).
2. Chemical nature of the active site (e.g. Pt, Au, ...).
3. Type of the active site (e.g. top, bridge, fcc, hcp, ...).
4. Site number (i.e. position on the surface).
5. Orientation of the adsorbate, with respect to the arbitrary axis (if applicable).
6. Chirality state of the adsorbate (if applicable).

These sub-partitions and this specific order were subjectively chosen based on chemical intuition in order to produce intermediate groups with a relative similarity in terms of lateral interactions. Figure 2 illustrates this partitioning on a simple example.

Let us consider a node $\mathcal{N}_i$ that is located at a depth corresponding to k turns of choices/partitioning. In other words, $\mathcal{N}_i$ corresponds to an incomplete configuration (with k adsorbates) generated after k turns. Therefore, all final configurations (with N_{ads} adsorbates) descending from that node $\mathcal{N}_i$ will result from the addition of $N_{\text{ads}} - k$ new adsorbates, corresponding to the remaining $N_{\text{ads}} - k$ turns. Since all adsorbates placed on the surface are not modified by subsequent moves, any final configuration or group/node descending from $\mathcal{N}_i$ is guaranteed to contain the incomplete k adsorbates based configuration represented by $\mathcal{N}_i$, along with all patterns present in this incomplete configuration. As a consequence, this partitioning ensures that, for any branch of the corresponding tree, the further down a node/group is in that branch, the more lateral interactions are shared among that group,

Figure 2: Illustration of the turn-based partitioning for constructing the full game tree. All possible moves during a turn are partitioned (see the Turn partition tree) according to properties of the new site(s) being occupied (at the end of that turn). This partitioning is fixed and based on chemical intuition (i.e. expert selection). The leaves (i.e. terminal nodes) of this turn partition tree represent acceptable moves for the current turn. In this illustration, the turn partition tree represents the first turn. Following turns are represented by their own altered partition tree (some moves might become unavailable, while others might finally become available). The full game tree is globally organized/partitioned in a turn-based manner. Indeed, each node representing a game state after a single turn is expanded into all 2-turns game states that can be played from that single turn game state. As a consequence, all nodes representing a k -turns game states are at the same depth in that tree, and the more turns are required to reach a certain game state, the deeper its corresponding node is located in the tree. Each turn is sub-partitioned according to the turn partition tree, leading to the full structure depicted in this figure.

since patterns are cumulative. Furthermore, this chemistry-based partitioning makes the definition of custom rules (for avoiding redundant or unphysical configurations) user-friendly.

This turn-based partitioning also provide highly modular trees that can easily be extended/truncated for generating final configurations with different numbers of adsorbates (i.e. changing the number of total turns, see figure 2). Indeed, sometimes one might want to re-use previous data that was generated from configurations with a different number of adsorbates instead of restarting from scratch, which we enable by storing the appropriate information.

Since adsorbates can be placed in any order to generate the same final configuration, this could cause a major source of redundancy. To avoid these redundancies, we enforce an arbitrary lexicographic order on the sequence of placed adsorbates, i.e., $\text{move}_{i-1} \leq \text{move}_i$ for the tuple associated with a move (adsorbate chemical nature, surface site nature, site type, site number, orientation, chirality). This lexicographic order ensures that for each branch in the tree no other permutation of the associated sequence of moves can be found in any other branch of the tree.

4.2.2 Normalized exploration rate

By default, a UCT approach will uniformly sample children that seem equally promising, independently of the number of geometries that they represent (i.e. independently of the number of accessible final configurations from them).

Unbalance can already result from the unequal representation of some properties within the 1-body terms. For example, if multiple distinct orientations (0° , 30° , 60°) of an NH_3 molecule are considered whereas CO does not have alternative orientations, then the number of moves placing an NH_3 molecule is roughly 3 times more than for placing a CO molecule. Furthermore, due to the lexicographic order, the children of a node are likely unbalanced in terms of number of accessible final configurations (e.g. all the other moves can be explored after the $\leq_{\text{move}}$ -smallest move whereas no move is accessible after the $\leq_{\text{move}}$ -largest one).

This contributes to create unbalanced branches, even when fully explored, leading to the over-representation of $\leq_{\text{move}}$ -small moves as first moves.

Uniformly sampling moves in such an unbalanced tree leads to a nonuniform sampling of configurations. Indeed, with equally promising branches, a move leading to more accessible final configurations should be sampled more than a move leading to just a few configurations.

A normalized exploration rate, inspired by Bosc *et al.*,⁶¹ can be used to take into account the number of accessible geometries and consider a uniform effective exploration ratio between unbalanced branches. The general idea behind this is intuitive: each exploration is weighted by the number of explorations that would be required on the most populated sibling to sample the same fraction of descendants. Back to our NH_3/CO example: each exploration of the CO node shall be counted as 3 explorations (i.e. the number of explorations required to equivalently sample the NH_3 node). Mathematically, this reads as:

$$n_i^{\text{norm}} = n_i \times \frac{\max_{j \in \text{Sbl}(i)} (d_j)}{d_i}, \quad (8)$$

where n_i the number of times node $\mathcal{N}_i$ was explored, n_i^{norm} is the normalized number of explorations of $\mathcal{N}_i$, d_j is the number of descendants of node $\mathcal{N}_j$ and $\text{Sbl}(i)$ are the siblings of node $\mathcal{N}_i$ (i.e. the children of the parent of node $\mathcal{N}_i$). Note that with this definition, the normalized number of explorations of the parent node T^{norm} (required for computing the UCB score, see eq. 6) is taken as the sum of the normalized number of explorations of all its children. Using these normalized number of explorations for computing the UCB score (through the exploration bias term, see section 3.3.2) should contribute to the reduction of the sampling bias introduced by unbalanced partitioning in the tree structure.

4.2.3 KL-UCB score

As already mentioned, the major difference between standard applications of UCTs and "Build a configuration" comes from the fact that evaluating the true score of a given con-

figuration is computationally costly. Hence, one would like to minimize the exploitation of barely relevant branches as much as possible. Among the many functional forms available for the exploration bias term of the UCB policy $b_i(n_i^{\text{norm}}, T^{\text{norm}})$ of Eq. 6, only the one of Garivier et al.⁵³ has been found to perform consistently better than the "naive" UCB1 policy for applications with a low number of samples. Relying on the Kullback-Leibler divergence, a metric for comparing probability distributions, the so called KL-UCB bias term is a bit different from the classical form of Eq. 6, so that the KL-UCB score of a node $\mathcal{N}_i$ is defined as:

$$S_i = \mu_i^{-1} \left(\max_q \left\{ q \in [0, 1] \mid KL_{div}(\mu_i(\mathbb{E}(W_i)_{emp}), q) \leq \frac{\log(T^{\text{norm}})}{n_i^{\text{norm}}} \right\} \right), \quad (9)$$

where μ_i is a normalization function for the target variable W_i , n_i^{norm} is the number of normalized explorations of node $\mathcal{N}_i$, T^{norm} is the number of normalized explorations of its parent node, and KL_{div} is the Bernoulli Kullback-Leibler divergence:

$$KL_{div}(p, q) = p \log \left(\frac{p}{q} \right) + (1 - p) \log \left(\frac{1 - p}{1 - q} \right) \quad (10)$$

Finally, KL-UCB requires only the empirical mean $\mathbb{E}(W_i)_{emp}$ and the bounds of the target function W_i of node $\mathcal{N}_i$, in comparison to other bias terms requiring additional information (e.g. the empirical variance⁵²).

Note that the target function W_i needs to be bounded in $[0, 1]$ for computing the KL-UCB score. However, the exact bounds of W_i are *a priori* unknown in our case: even though the absolute energy deviation is always positive, its effective lower and upper bound depend on the geometries within G_i and can thus only be known exactly by evaluating all the geometries in G_i . Therefore an estimate of the effective bounds $[\min(W_i), \max(W_i)]$ of W_i has to be computed. Here, it is reasonable to assume that the global extrema for all target functions, $\min(W) = \min_i(\min(W_i))$ and $\max(W) = \max_i(\max(W_i))$, can be found within any node (i.e. the geometry with the highest/lowest deviation can *a priori* be found anywhere). Therefore, for any node $\mathcal{N}_i$, $[\min(W_i), \max(W_i)]$ can be estimated from the empirical global minimum

$\min(W)_{emp} - \Delta_{margin}$ and maximum $\max(W)_{emp} + \Delta_{margin}$, where Δ_{margin} represents how much an unexplored node can perform better/worse than all currently seen nodes with respect to the target function (i.e. how much better/worse, terms of energy deviation, a geometry can still be found). Using these bounds, the normalization function μ_i is designed to rescale the W_i variable to $[0, 1]$, as detailed in section S4 of the SI.

4.3 Pre-exploration

Even if the UCT identifies the most promising branch, it has no information about which unexplored node is the most "useful" to choose, which is particularly relevant in the beginning of the exploration when the scores of the branches are largely unknown. Hence, in vanilla MCTS the choice among the equally promising nodes is made randomly. However, a rollout policy can be applied through heuristics^{62–64} or other exploration techniques⁶⁵ instead. Such policy is designed to accelerate convergence by excluding sub-optimal nodes early in the sampling process. This is especially important here, since only a few number of geometries can be evaluated due to the high computational cost associated.

In practice, after a pre-exploration step composed of P successive random pre-explorations, P novel (with respect to the current training set) final configurations are selected. These configurations are expected to display large absolute energy deviation, according to the UCT-based policy that selected them.

Instead of simply taking $P = 1$ and thus relying on random exploration, we actively select the configuration that is expected to benefit the model Hamiltonian the most in terms of diversity. Mathematically, the uncertainty associated with a linear model is properly defined by the variance of its parameters $\text{Var}(\hat{\beta})$. Each new observation reduces the variance matrix $\text{Var}(\hat{\beta})$ by a symmetric positive semi-definite matrix $\Delta \text{Var}(\hat{\beta})$ that roughly depends on how redundant the new observation is with respect to the previous ones. By estimating the changes in the variance matrix $\Delta \text{Var}(\hat{\beta})$, one can choose the configuration that benefits the most the model Hamiltonian. One should note that this is a typical variance-based

DoE approach.³⁸ A convenient quality estimator for the variance reduction is the trace $\text{Tr}(\Delta \text{Var}(\hat{\beta}))$. This estimator implicitly assumes that all lateral interactions are equivalently important. This choice of minimizing the trace of the inverse Fisher information matrix is a popular choice for active learning algorithms, and is known as A-optimal design.³⁸

We thus propose a rollout policy that consists in selecting, among the pre-explored configurations, the configuration expected to benefit the most the model Hamiltonian being learned. This policy can be seen as an additional compromise between exploration (choosing the configuration with *a priori* least redundant information) and exploitation (the UCT-based selection policy acts as a filter for predicted high-score configurations).

4.4 Implementation

The UCT+DoE framework described in this paper has been implemented in Python3, with a modular design and a relatively few number of required parameters. Our implementation heavily relies on the NetworkX library for all graph-related algorithms, and Numpy whenever applicable. For computing the KL-UCB score, our implementation uses the `scipy.optimize.brentq` solver and for updating the linear model associated with the internal model Hamiltonian, and estimating the covariance change, we use the Recursive Least Squares algorithm rank-Greville.⁴⁷ The code is available on GitHub, along with a rich documentation and examples.

The MCTS_RS module was designed as a versatile tool, with numerous flags/options for addressing most cases. For example, our implementation can easily be turned into a purely pseudorandom sampler, or even using a pre-exploration step on a pseudorandom sampling, leading to a typical variance-based active learning algorithm.

In section S3 we discuss the details of our implementation regarding choices that have to be made for our UCT+DoE framework applied to CE-based model Hamiltonians. In particular, we explain how we keep the node storage requirements down, how we normalize the exploration rate in practice and how we slightly bias the variance based score towards

the exploration of as many unseen patterns as possible. Finally, we also give the details for the normalization of the KL-UCB rewards.

5 Setup for proof of principles

We are interested in evaluating the performance of our UCT+DoE framework for generating a training set that is optimal for learning the relevant many-body energy contributions (i.e. training a relevant model Hamiltonian), using only a limited number of energy evaluations. As a proof of principle, the our UCT+DoE approach is applied to the CO oxidation on Pd(111), by learning the CE-based model Hamiltonian described by S. Piccinin and M. Stamatakis.²⁰ This model Hamiltonian defines which patterns are relevant and their associated energies. Hence, we do not have to worry about dimensionality reduction, i.e., the discrimination between relevant and irrelevant many-body terms, a topic that is beyond the scope of this study, but has been addressed in the dedicated literature.^{37,66–68} Of course, these contributions are not known beforehand by the framework which has only access to a limited number of predictions from the reference model Hamiltonian (with eventual Gaussian errors added), just as if the predictions were computed from first-principle calculations.

5.1 Relevant patterns

The model of ref 20 defines 52 relevant "figures" (which roughly correspond to what we call patterns) composed of up to 3 adsorbates/sites and exploring up to the second nearest neighbors; four of these figures correspond to steric clashes. The adsorbates (CO and O) were defined with no characteristic orientation (i.e. treated with D_∞ symmetry group on the surface). The active sites (fcc and hcp) rigorously have D_3 for symmetry group, but their symmetry is not considered in the reference model Hamiltonian, so for coherence purposes, we treat them with the D_∞ symmetry group.

We convert the 4 relevant patterns corresponding to steric clashes to a user-defined filter

that prevents any considered configuration from displaying these clashes. In summary, while the four single-body terms (CO and O on hcp and fcc sites) are taken to be known from the beginning, the remaining 24 2-body patterns and 20 3-body patterns are to be learned via an appropriate training set. Note that these 44 patterns are only a subset of the 70 distinct possible patterns that are generated based on the general rule of up to 3 adsorbates/sites at distances up to the second nearest neighbor.

5.2 UCT+DoE parameters

For our proof of principle study, we limit the number of predictions from the reference model Hamiltonian to 100, i.e., a final training set size of 100 configurations. This compares to the 92 DFT configurations used in ref 20.

Since our UCT+DoE framework aims at reducing the computational cost spent on first-principle calculations for creating a training set, it is natural to consider the smallest surface that can define all relevant patterns. In the present case, the smallest supercell where all patterns can be defined (due to periodic boundary conditions, i.e., after applying the minimal image convention) is the $p(4 \times 3)$ cell of the (111) surface, therefore composed of 12 unit cells each containing a one fcc and one hcp site, leading to 24 adsorption sites in total.

For each test, the application of our active learning framework consists of three steps that are dynamically selected:

- First, a pure DoE-based exploration is performed until all possible patterns have been explored (i.e. each possible pattern is present in at least one of the geometries selected so far). Indeed, in this initial phase the exploitation power of the UCT is not justified by an exploration/exploitation trade-off; hence, we choose to focus on exploring as many new patterns as possible. As a consequence, during this phase, the pre-exploration step is a simple random traversal of the UCT tree, while the final score is DoE-based and thus favoring configurations with a large number of yet unseen patterns.

- Then, the now pre-populated UCT framework is used (and updated), effectively applying a bias toward configurations that are predicted to feature large lateral interactions. At this stage, the score used for the final selection (i.e. the rollout policy) is now largely favoring configurations that are simply non-redundant (i.e. that cannot be written as a linear combination of previously selected configurations), until the corresponding linear system is no longer underdetermined.
- In the final stage, the UCT framework is still being updated and the exploration/exploitation trade-off is now fully operational. The DoE-based score used by the rollout policy now behaves in a typical A-optimal design fashion, targeting the reduction of $\text{Tr}(\Delta \text{Var}(\hat{\beta}))$.

5.3 Training set quality evaluation

In order to evaluate the performance of our active learning framework, the training set generated $\Gamma_1, \Gamma_2, \dots, \Gamma_n$ is evaluated with respect to its ability to accurately estimate the relevant energy contributions $\hat{\beta}_{rel}$ defined from the reference model Hamiltonian (4 single-body + 44 many-body contributions).

Since we are here focusing on efficiently estimating the hidden relevant energy contributions, the process of model selection is not treated here: the relevant energy terms are simply taken as the ones defined in the reference model Hamiltonian (i.e. the non-null energy contributions). It is important to note that our active learning framework does not have access to which many-body interactions have been expertly chosen as relevant. This is actually a key strength of our approach: despite the number and identity of the 44 relevant parameters being hidden from the active learning algorithm, the generated training set is expected to provide optimal information on the energy contributions expertly chosen *a posteriori* as non-negligible out of the 70 possible patterns (*vide supra*).

In order to evaluate the performance of the generated training set $\Gamma_1, \Gamma_2, \dots, \Gamma_n$ for reliably estimating the relevant energy contributions $\hat{\beta}_{rel}$, we rely on the covariance matrix $\text{Var}(\hat{\beta}_{rel})$ obtained when using this training set for fitting a model Hamiltonian containing

only the relevant terms. Assuming homoscedasticity, the covariance matrix $\text{Var}(\hat{\beta}_{rel})$ for a non-underdetermined linear system is given by:

$$\text{Var}(\hat{\beta}_{rel}) = \sigma^2(\mathbf{\Gamma}_{rel}^\top \mathbf{\Gamma}_{rel})^{-1}, \quad (11)$$

where σ^2 is the variance of the noise and $\mathbf{\Gamma}_{rel}$ is the observation matrix associated with the training set $\Gamma_1, \Gamma_2, \dots, \Gamma_n$, restricted to the 44 expert-chosen relevant terms.

One reason to consider the covariance matrix $\text{Var}(\hat{\beta}_{rel})$ is that we are interested in using the least susceptible metric available, with respect to the pseudorandom seed used. Indeed, limited variability with respect to the pseudorandom seed allows to obtain more meaningful statistics from a small amount of repeats. Compared to a simple root mean square error (RMSE) on the fitted parameters $\hat{\beta}_{rel}$, the covariance matrix $\text{Var}(\hat{\beta}_{rel})$ provides a much more consistent metric across repeats with different pseudorandom seeds, as the latter is roughly related to the asymptotic behavior of the former with respect to the number of repeats. See section S7.1 of the SI for more details.

The covariance matrix $\text{Var}(\hat{\beta}_{rel}) = \sigma^2(\mathbf{\Gamma}_{rel}^\top \mathbf{\Gamma}_{rel})^{-1}$ can be decomposed into σ^2 representing the noise in the input energies, and the inverse of a Gram matrix $(\mathbf{\Gamma}_{rel}^\top \mathbf{\Gamma}_{rel})^{-1}$ that can be seen as representing the susceptibility of the model with respect to the input noise. As a consequence, simply using $\text{Var}(\hat{\beta}_{rel})$ would yield different results for the same set of geometries, depending on the amount of noise used for the input energies. As we are solely interested in evaluating the quality of the set of geometries chosen by the active learning algorithm, we focus only on the $(\mathbf{\Gamma}_{rel}^\top \mathbf{\Gamma}_{rel})^{-1}$ term¹. This choice has the additional major benefit of being still meaningful in the ideal limit of exact input energies.

Finally, our quality estimator is reduced for convenience purposes to a single scalar by taking the trace of the $(\mathbf{\Gamma}_{rel}^\top \mathbf{\Gamma}_{rel})^{-1}$ matrix. This approach is therefore similar to the commonly used A-optimal design: treating each relevant contribution as equally important and,

¹Such choice does not mean that noise is not taken into account, as noise can mislead the algorithm into generating a sub-optimal training set.

most importantly, not including the empirical correlations between these terms.

6 Results and Discussion

In this section we present the performance of our active learning framework, compared with relevant baseline algorithms for an in-depth analysis. We also test how resilient our algorithm is with respect to what we believe could be the most common sub-optimal conditions of application, in order to give insights on the applicability domain of UCT+DoE.

6.1 Asymptotic performance

Here, we apply our active learning framework to the model Hamiltonian example in optimal conditions:

- Zero noise: the selected geometries are evaluated using the reference model Hamiltonian without any error added. This would correspond to exact energy computations and exact representability of the energies by the model Hamiltonian in practice. Under such condition, the UCT framework interpolates from exact and consistent values.
- Easy exploration: According to tests relying on random explorations, a coverage of $\frac{1}{3}$ of the available active sites, corresponding to 8 adsorbates, allows for the easiest sampling of a multitude of diverse patterns in a single geometry. Hence, we choose this coverage here.

During the pre-exploration step, $P = 10$ configurations were selected for the DoE-based rollout policy to choose from. We found this value to be a good compromise between quality (the higher P , the more efficient the DoE-based policy) and computational cost (the geometry selection scales linearly with respect to P). However, for real-life application, a larger number of pre-explorations can be considered, as the total computational cost will be dominated by

Figure 3: Evolution of the training set quality obtained during the sampling process, in the absence of input noise. 5 algorithms are compared: random sampling, random+DoE(A-optimal design on all terms) sampling, UCT+DoE(A-optimal design on all terms) sampling, random+DoE(A-optimal design on relevant terms only), UCT+DoE(A-optimal design on relevant terms only). The quality estimator $\text{Tr}((\mathbf{\Gamma}_{rel}^T \mathbf{\Gamma}_{rel})^{-1}) = \frac{\text{Tr}(\text{Var}(\hat{\beta}_{rel}))}{\sigma^2}$ represents the susceptibility of the fitted model to input noise (the smaller, the better). Results are reported at each step (i.e. each time a new geometry is added), starting from when the training set is able to determine all relevant terms (i.e when maximum rank of $\mathbf{\Gamma}_{rel}$ is achieved). Each sampling algorithm is repeated 100 times with a different pseudorandom seed. The resulting average is plotted along with a violin plot representing the distribution of results obtained at each step (the violin plots are rescaled so that their area are equal for a same step).

the energy evaluation (including geometry optimisation) at the DFT level. Nonetheless, one should note that the higher P , the more power is given to the DoE-based selection, which might reduce the expression of the UCT bias.

The evolution of the training set quality (from our estimator $\text{Tr}((\mathbf{\Gamma}_{rel}^\top \mathbf{\Gamma}_{rel})^{-1})$, see Figure 3) during the training is reported as soon as the relevant model is no longer underdetermined (i.e. $\mathbf{\Gamma}_{rel}^\top \mathbf{\Gamma}_{rel}$ is non-singular).

First of all, we note that a pure random sampling (purple line) is insufficient to properly learn our reference model Hamiltonian: not only the final training set quality is consistently quite low, but in 4 runs out of 100, the final training set did not even explore the full subspace of relevant energy contributions. Due to the high variability of the results provided by the purely random samplings, the corresponding violin plots are not represented for graphical convenience.

In comparison, a pure DoE-based sampling (blue) performs much better and far more consistently. This is expected, as despite the geometries being generated at random, the DoE-based selection is pushing towards the proper fitting of all 70 possible energy contributions. Therefore, the 44 relevant energy contributions are also properly sampled. Despite being non-specific to the relevant energy contributions, such general DoE-based algorithm provides a quite acceptable baseline in practice.

Despite not being aware of which terms are relevant, our mixed UCT+DoE framework (green) is performing significantly better than a pure DoE approach. This success can be explained by the natural correlation between a geometry displaying large lateral interactions and the corresponding configuration containing many-body patterns whose energy contributions are non-negligible. While it is clear that this correlation is essential for our UCT+DoE approach to be beneficial, it is interesting to notice that in the reference model Hamiltonian, a non-negligible set of 11 relevant terms have stabilizing energy contributions. Yet, despite this subset of terms being anticorrelated with a large energy deviation, these terms are them-

selves correlated with the presence of other, destabilizing, terms, so that the total lateral interaction is always destabilizing. In other words, the UCT policy allows an automatic, on the fly machine-learned, bias towards geometries that feature relevant patterns.

While the performance of our UCT+DoE approach shows that it is indeed possible to go beyond directionless optimization even when no prior information is available, effectively replacing chemical intuition with bias from reinforcement learning, it remains to be seen how efficient is this bias compared to an expertly focused optimization. Therefore, a pure DoE-based sampling is performed with prior expert knowledge of which contributions are relevant: for each step, $P = 10$ geometries are generated at random and an A-optimal design is applied to choose the most beneficial geometry focusing only on the relevant terms.

As expected, the expert knowledge supported DoE-based sampling (orange) is performing better than our UCT+DoE approach, as we are comparing expert knowledge against a guessed bias (from the UCT) toward an objective function (geometries displaying a large energy deviation) whose link to the true expert knowledge-based objective function (focusing on the efficient sampling of relevant terms) is only a correlation.

Incidentally, one might be interested in the benefits of our UCT framework in the case where chemical expertise is actually available. Therefore, our UCT+DoE framework is applied with the DoE part focusing only on the relevant terms. This UCT+DoE(relevant) approach (red) with access to prior expert knowledge is significantly outperforming all other tested approaches. Such performance is simply due to the sampling bias from the UCT being much better exploited by the DoE(relevant)-based rollout. In other words, the UCT part and the DoE(relevant) part are applying compatible biases. This concordance between both biases is yet another proof that our energy deviation-based score is appropriate, since it seems in practice coupled with a better sampling of expertly chosen patterns of relevance. This case, where UCT and DoE are optimally collaborating, leads to unmatched performance.

Nonetheless, the performance of our UCT+DoE approach compares admirably well against

the expert knowledge supported DoE-based sampling (compare green and orange in Figure 3). We believe that this excellent performance of UCT+DoE can be explained by two main factors:

- The carefully crafted UCT-based reinforcement learning framework described here is indeed able to effectively learn the geometries with largest energy deviation, as suggested by the swift improvement that can be observed in figure ?? (at the early stage after maximum rank of Γ is achieved), hinting that our metric (i.e. the turn-based tree structure described section 4.2.1) is acceptable.
- A pattern is expected to be expertly selected as relevant not only if its energy contribution is significant, but also if such pattern is commonly found in practice (or correlated with the presence of patterns with large energy contributions). The latter should then, by definition, be easily sampled even without being actively targeted. This means that the UCT policy is expected to select frequent but small terms are expected to be selected on the same foot as rare large terms.

In summary, these results are very promising and suggest that machine learning can be used as a complement or substitute when chemical intuition/expertise is limited or unavailable.

6.2 Challenging exploration

The performance of our UCT+DoE approach is directly linked to its ability to find a proper trade-off between exploration and exploitation. It is therefore meaningful to study the behavior of this approach when exploration is difficult and can therefore no longer be traded lightly. To investigate this point, we apply UCT+DoE to sub-optimal coverages.

A slightly sub-optimal regime (7 instead of 8 adsorbates) affects UCT+DoE only slightly more than pure DoE, so that UCT+DoE clearly remains the best performing algorithm (see

Figure 4: Evolution of the training set quality obtained during the sampling process, at a sub-optimal coverage. UCT+DoE sampling with zero noise, and random+DoE sampling are compared at a highly sub-optimal coverage ($N_{\text{ads}} = 5$ (optimal is 8)). The quality estimator $\text{Tr}((\mathbf{\Gamma}_{\text{rel}}^{\text{T}} \mathbf{\Gamma}_{\text{rel}})^{-1}) = \frac{\text{Tr}(\text{Var}(\hat{\beta}_{\text{rel}}))}{\sigma^2}$ represents the susceptibility of the fitted model to input noise (the smaller, the better). Results are reported at each step (i.e. each time a new geometry is added), starting from when the training set is able to determine all relevant terms (i.e. when maximum rank of $\mathbf{\Gamma}_{\text{rel}}$ is achieved). Each sampling algorithm is repeated 100 times with a different pseudorandom seed. The resulting average is plotted along with a violin plot representing the distribution of results obtained at each step (the violin plots are rescaled so that their area are equal for a same step).

Figure S6). However, when exploration becomes significantly more challenging (5 instead of 8 adsorbates), the performance of our mixed UCT+DoE approach was found to drastically decrease (see Figure 4). This result can be explained in terms of the trade-off between exploration and exploitation introduced by the UCT. Indeed, by focusing on apparently relevant geometries, the UCT framework is trading exploration for exploitation. If the exploration of various geometries is almost trivial, then the trade-off comes at virtually no cost. However, if the exploration is particularly challenging, trading some of it might come at a prohibitive cost where the DoE-based algorithm can no longer ensure an extensive sampling, as all pre-selected geometries are too similar.

Since the exploration/exploitation trade-off is at the heart of our mixed UCT+DoE approach, it is essential to ensure that such trade-off can be performed without compromising the sampling process. Therefore, the mixed UCT+DoE framework should only be applied to systems where exploration is not an issue. We suggest to run a simple random sampling (without evaluating energies) at different coverages in order to quickly estimate the optimal one.

6.3 Susceptibility to noise

In practice, energy evaluations and their representation by the model Hamiltonian are only approximate, leading to noisy inputs which might mislead the UCT framework into over-estimating or, even worse, under-estimating whole branches of the search tree, leading to medium-term stalled exploration of the affected branches. To test the resiliency of our UCT+DoE approach towards noisy input data, we test it with four (increasing) intensities of homoscedastic Gaussian noise added to the input energies during training (see Figures 5a, S4, 5b and S5).

Figure 5a demonstrates that UCT+DoE is barely affected by reasonably large errors ($\sigma \leq 0.5$ eV per geometry on the adsorption energy): indeed little to no notable effects except

(a) Effect of noise ($\sigma = 0.5$ eV).

(b) Effect of large noise ($\sigma = 2.0$ eV), corresponding to a $\approx 1:1$ ratio of signal/noise.

Figure 5: Evolution of the training set quality obtained during the sampling process. As a comparison, UCT+DoE sampling with zero noise, and random+DoE sampling (independent of input noise) is also shown.

for more dispersed results across repeats are obtained. Given that the scores W_i used in the UCT nodes are averages, the noise tends to cancel out, so that the misjudgements of the UCT are quickly corrected.

Even when the errors are comparable to the effective range of energy deviations ($\sigma = 2\text{eV} \approx \max(W) - \min(W)$), our UCT+DoE approach still yields a comparable performance to that of a pure DoE-based approach on average, but much less consistent across repeats, see Figure 5b. At this very high level of noise, the error cancellation is not able to mitigate the stalled exploration of relevant branches quickly enough. Note that for intermediate noise levels ($\sigma = 1\text{eV}$) UCT+DoE performs still significantly better than DoE (see S4).

It is only when the noise reaches unrealistically large errors ($\sigma = 5\text{ eV}$), that the substantial misjudgement of the UCT yields strong nonphysical biases, responsible for a worse performance than that of a pure DoE-based approach, see Figure S5.

The substantial resilience of our UCT+DoE framework with respect to noise is likely a combination of the UCT framework being intrinsically designed for handling random variables, and the conservative DoE-based rollout acting as a safeguard. In summary, UCT+DoE is a reliable approach even in the presence of very significant noise.

6.4 Conclusions

We have addressed the challenge of an automatic construction of a reliable training set for cluster expansions at minimal computational cost. Reformulating the problem as a strategy game, we propose a mixed UCT+DoE approach to solve it efficiently. This approach is finely adapted to the specific characteristic of the problem at hand: evaluating the final score (the DFT computed adsorption energy) is much more costly than the score evaluation in regular strategy games. As a proof of principle, we apply UCT+DoE to the description of lateral interactions on a graphical lattice, exemplified by the co-adsorption of CO and O on Pd(111). Numerical examples demonstrate the effectiveness of UCT+DoE even under sub-

optimal conditions such as the presence of many irrelevant patterns, sub-optimal coverages and noisy energies. These very promising results suggest that machine learning (here in the form of an upper confidence tree) can be used as a substitute when chemical intuition (for answering: which configurations/patterns should be included in the training set?) is unavailable. Therefore, we expect that UCT+DoE will be widely used for the training of (linear) model Hamiltonians, especially in the context of heterogeneous catalysis, where these models are used to drive kinetic Monte Carlo algorithms to determine the catalytic activity and selectivity.

Acknowledgement

The authors thank the SYSPROD project and AXELERA Pôle de Compétitivité for financial support (PSMN Data Center). We are grateful to A. Garivier, F. Hess and M. Stamatakis for stimulating discussions.

Supporting Information Available

The supporting information contains an extensive discussion of technical details and additional Figures.

The UCT+DoE code and examples are available under https://gitlab.com/lch_interfaces/reinforcementsampling.

References

- (1) Greeley, J.; Norskov, J. K.; Mavrikakis, M. Electronic Structure and Catalysis on Metal Surfaces. *Annu. Rev. Phys. Chem.* **2002**, *53*, 319.
- (2) van Santen, R. A.; Sautet, P. Conclusion: Challenges to Computational Catalysis. *Com-*

- putational Methods in Catalysis and Materials Science: An Introduction for Scientists and Engineers* **2009**, 441.
- (3) Nie, S.; Feibelman, P. J.; Bartelt, N. C.; Thürmer, K. Pentagons and Heptagons in the First Water Layer on Pt(111). *Phys. Rev. Lett.* **2010**, *105*, 026102.
 - (4) Mom, R. V.; Melissen, S. T. A. G.; Sautet, P.; Frenken, J. W. M.; Steinmann, S. N.; Groot, I. M. N. The Pressure Gap for Thiols: Methanethiol Self-Assembly on Au(111) from Vacuum to 1 bar. *J. Phys. Chem. C* **2019**, *123*, 12382–12389.
 - (5) Auer, A.; Andersen, M.; Wernig, E.-M.; Hörmann, N. G.; Buller, N.; Reuter, K.; Kunze-Liebhäuser, J. Self-activation of copper electrodes during CO electro-oxidation in alkaline electrolyte. *Nature Catalysis* **2020**, 1–7, Publisher: Nature Publishing Group.
 - (6) Fang, Y.; Ding, S.; Zhang, M.; Steinmann, S. N.; Hu, R.; Mao, B.; Feliu, J. M.; Tian, Z. Revisiting the atomistic structures at the interface of Au(111) electrode-sulfuric acid solution. *J. Am. Chem. Soc.* **2020**, *142*, 9439–9446, Publisher: American Chemical Society.
 - (7) Reuter, K.; Scheffler, M. Composition, structure, and stability of $\{\mathrm{RuO}\}_2(110)$ as a function of oxygen pressure. *Phys. Rev. B* **2001**, *65*, 035406, Publisher: American Physical Society.
 - (8) Vignola, E.; Steinmann, S. N.; Le Mapihan, K.; Vandegehuchte, B. D.; Curulla, D.; Sautet, P. Acetylene Adsorption on Pd–Ag Alloys: Evidence for Limited Island Formation and Strong Reverse Segregation from Monte Carlo Simulations. *J. Phys. Chem. C* **2018**, *122*, 15456–15463.
 - (9) Stamatakis, M. Kinetic modelling of heterogeneous catalytic systems. *Journal of Physics: Condensed Matter* **2014**, *27*, 013001.

- (10) Andersen, M.; Panosetti, C.; Reuter, K. A Practical Guide to Surface Kinetic Monte Carlo Simulations. *Frontiers in Chemistry* **2019**, *7*, 202.
- (11) Martí, C.; Blanck, S.; Staub, R.; Loehlé, S.; Michel, C.; Steinmann, S. N. DockOnSurf: A Python Code for the High-Throughput Screening of Flexible Molecules Adsorbed on Surfaces. *J. Chem. Inf. Model.* **2021**, *61*, 3386–3396, Publisher: American Chemical Society.
- (12) Kumar, K. V.; Porkodi, K.; Rocha, F. Langmuir–Hinshelwood kinetics—a theoretical study. *Catalysis Communications* **2008**, *9*, 82–84.
- (13) Baxter, R.; Hu, P. Insight into why the Langmuir–Hinshelwood mechanism is generally preferred. *The Journal of chemical physics* **2002**, *116*, 4379–4381.
- (14) Wu, C.; Schmidt, D. J.; Wolverton, C.; Schneider, W. F. Accurate coverage-dependence incorporated into first-principles kinetic models: Catalytic NO oxidation on Pt (111). *Journal of Catalysis* **2012**, *286*, 88–94.
- (15) Farkas, A.; Hess, F.; Over, H. Experiment-based kinetic monte carlo simulations: CO oxidation over RuO₂ (110). *The Journal of Physical Chemistry C* **2012**, *116*, 581–591.
- (16) Sutton, J. E.; Lorenzi, J. M.; Krogel, J. T.; Xiong, Q.; Pannala, S.; Matera, S.; Savara, A. Electrons to reactors multiscale modeling: catalytic CO oxidation over RuO₂. *ACS Catalysis* **2018**, *8*, 5002–5016.
- (17) Gautier, S.; Sautet, P. Coadsorption of Butadiene and Hydrogen on the (111) Surfaces of Pt and Pt₂Sn Surface Alloy: Understanding the Cohabitation from First-Principles Calculations. *J. Phys. Chem. C* **2017**, *121*, 25152–25163, Publisher: American Chemical Society.
- (18) Clabaut, P.; Staub, R.; Galiana, J.; Antonetti, E.; Steinmann, S. N. Water adlayers

- on noble metal surfaces: Insights from energy decomposition analysis. *J. Chem. Phys.* **2020**, *153*, 054703, Publisher: American Institute of Physics.
- (19) Hess, F.; Over, H. Rate-determining step or rate-determining configuration? the Deacon reaction over RuO₂ (110) studied by DFT-based KMC simulations. *ACS Catalysis* **2017**, *7*, 128–138.
 - (20) Piccinin, S.; Stamatakis, M. CO Oxidation on Pd(111): A First-Principles-Based Kinetic Monte Carlo Study. *ACS Catal.* **2014**, *4*, 2143–2152.
 - (21) Wolverton, C.; Zunger, A. Ising-like Description of Structurally Relaxed Ordered and Disordered Alloys. *Phys. Rev. Lett.* **1995**, *75*, 3162–3165.
 - (22) Muller, S. Bulk and surface ordering phenomena in binary metal alloys. *J. Phys.: Condens. Matter* **2003**, *15*, R1429–R1500, Publisher: IOP Publishing.
 - (23) Chen, W.; Schmidt, D.; Schneider, W. F.; Wolverton, C. Ordering and Oxygen Adsorption in Au–Pt/Pt(111) Surface Alloys. *J. Phys. Chem. C* **2011**, *115*, 17915–17924.
 - (24) Zhu, B.; Creuze, J.; Mottet, C.; Legrand, B.; Guesmi, H. CO Adsorption-Induced Surface Segregation and Formation of Pd Chains on AuPd(100) Alloy: Density Functional Theory Based Ising Model and Monte Carlo Simulations. *J. Phys. Chem. C* **2016**, *120*, 350–359.
 - (25) Sanchez, J. M.; Ducastelle, F.; Gratias, D. Generalized cluster description of multi-component systems. *Physica A: Statistical Mechanics and its Applications* **1984**, *128*, 334–350.
 - (26) Stampfl, C. Predicting surface phase transitions from ab initio based statistical mechanics and thermodynamics. *Phase Transitions* **2007**, *80*, 311–332.

- (27) Lerch, D.; Wieckhorst, O.; Hammer, L.; Heinz, K.; Müller, S. Adsorbate cluster expansion for an arbitrary number of inequivalent sites. *Physical Review B* **2008**, *78*, 121405.
- (28) Cao, L.; Li, C.; Mueller, T. The use of cluster expansions to predict the structures and properties of surfaces and nanostructured materials. *Journal of chemical information and modeling* **2018**, *58*, 2401–2413.
- (29) Schmidt, D. J.; Chen, W.; Wolverton, C.; Schneider, W. F. Performance of cluster expansions of coverage-dependent adsorption of atomic oxygen on Pt (111). *Journal of chemical theory and computation* **2012**, *8*, 264–273.
- (30) van de Walle, A.; Ceder, G. Automating first-principles phase diagram calculations. *Journal of Phase Equilibria* **2002**, *23*, 348.
- (31) Hess, F. Efficient Implementation of Cluster Expansion Models in Surface Kinetic Monte Carlo Simulations with Lateral Interactions: Subtraction Schemes, Supersites, and the Supercluster Contraction. *Journal of Computational Chemistry* **2019**, *40*, 2664–2676.
- (32) Vignola, E.; Steinmann, S. N.; Vandegehuchte, B. D.; Curulla, D.; Stamatakis, M.; Sautet, P. A machine learning approach to graph-theoretical cluster expansions of the energy of adsorbate layers. *The Journal of Chemical Physics* **2017**, *147*, 054106.
- (33) Blum, V.; Hart, G. L.; Walorski, M. J.; Zunger, A. Using genetic algorithms to map first-principles results to model Hamiltonians: Application to the generalized Ising model for alloys. *Physical Review B* **2005**, *72*, 165113.
- (34) Hart, G. L.; Blum, V.; Walorski, M. J.; Zunger, A. Evolutionary approach for determining first-principles hamiltonians. *Nature materials* **2005**, *4*, 391–394.

- (35) Bajpai, A.; Frey, K.; Schneider, W. F. Binary Approach to Ternary Cluster Expansions: NO–O–Vacancy System on Pt (111). *The Journal of Physical Chemistry C* **2017**, *121*, 7344–7354.
- (36) Seko, A.; Koyama, Y.; Tanaka, I. Cluster expansion method for multicomponent systems based on optimal selection of structures for density-functional theory calculations. *Physical Review B* **2009**, *80*, 165122.
- (37) Herder, L. M.; Bray, J. M.; Schneider, W. F. Comparison of cluster expansion fitting algorithms for interactions at surfaces. *Surface Science* **2015**, *640*, 104–111.
- (38) Settles, B. *Active learning literature survey*; 2009.
- (39) Silver, D.; Huang, A.; Maddison, C. J.; Guez, A.; Sifre, L.; Van Den Driessche, G.; Schrittwieser, J.; Antonoglou, I.; Panneershelvam, V.; Lanctot, M., et al. Mastering the game of Go with deep neural networks and tree search. *nature* **2016**, *529*, 484–489.
- (40) Silver, D.; Schrittwieser, J.; Simonyan, K.; Antonoglou, I.; Huang, A.; Guez, A.; Hubert, T.; Baker, L.; Lai, M.; Bolton, A., et al. Mastering the game of go without human knowledge. *nature* **2017**, *550*, 354–359.
- (41) Sutton, R. S.; Barto, A. G. *Reinforcement learning: An introduction*; MIT press, 2018.
- (42) Zhang, L.; Lin, D.-Y.; Wang, H.; Car, R.; E, W. Active learning of uniformly accurate interatomic potentials for materials simulation. *Phys. Rev. Materials* **2019**, *3*, 023804, Publisher: American Physical Society.
- (43) Goings, J. J.; Hu, H.; Yang, C.; Li, X. Reinforcement Learning Configuration Interaction. *J. Chem. Theory Comput.* **2021**, *17*, 5482–5491, Publisher: American Chemical Society.
- (44) Lan, T.; An, Q. Discovering Catalytic Reaction Networks Using Deep Reinforcement

- Learning from First-Principles. *J. Am. Chem. Soc.* **2021**, *143*, 16804–16812, Publisher: American Chemical Society.
- (45) Kohn, W. Density Functional and Density Matrix Method Scaling Linearly with the Number of Atoms. *Phys. Rev. Lett.* **1996**, *76*, 3168–3171.
 - (46) Greene, W. *Econometric Analysis*; Pearson, 2018.
 - (47) Staub, R.; Steinmann, S. N. Efficient recursive least squares solver for rank-deficient matrices. *Applied Mathematics and Computation* **2021**, *399*, 125996.
 - (48) Bartók, A. P.; De, S.; Poelking, C.; Bernstein, N.; Kermode, J. R.; Csányi, G.; Ceriotti, M. Machine learning unifies the modeling of materials and molecules. *Sci. Adv.* **2017**, *3*, e1701816.
 - (49) Browne, C. B.; Powley, E.; Whitehouse, D.; Lucas, S. M.; Cowling, P. I.; Rohlfshagen, P.; Tavener, S.; Perez, D.; Samothrakis, S.; Colton, S. A survey of monte carlo tree search methods. *IEEE Transactions on Computational Intelligence and AI in games* **2012**, *4*, 1–43.
 - (50) Lai, T.; Robbins, H. Asymptotically efficient adaptive allocation rules. *Advances in Applied Mathematics* **1985**, *6*, 4–22.
 - (51) Auer, P.; Cesa-Bianchi, N.; Fischer, P. Finite-time analysis of the multiarmed bandit problem. *Machine learning* **2002**, *47*, 235–256.
 - (52) Audibert, J.-Y.; Munos, R.; Szepesvári, C. Exploration-Exploitation Tradeoff Using Variance Estimates in Multi-Armed Bandits. *Theor. Comput. Sci.* **2009**, *410*, 1876–1902.
 - (53) Garivier, A.; Cappé, O. The KL-UCB algorithm for bounded stochastic bandits and beyond. Proceedings of the 24th annual conference on learning theory. 2011; pp 359–376.

- (54) Audibert, J.-Y.; Bubeck, S. Regret Bounds and Minimax Policies under Partial Monitoring. *Journal of Machine Learning Research* **2010**, *11*, 2785–2836.
- (55) Honda, J.; Takemura, A. An asymptotically optimal bandit algorithm for bounded support models. In Proceedings of the Twenty-third Conference on Learning Theory (COLT 2010. 2010; pp 67–79.
- (56) Kocsis, L.; Szepesvári, C.; Willemson, J. Improved monte-carlo search. *Univ. Tartu, Estonia, Tech. Rep* **2006**, *1*.
- (57) Schadd, M. P.; Winands, M. H.; Van Den Herik, H. J.; Chaslot, G. M.-B.; Uiterwijk, J. W. Single-player monte-carlo tree search. International Conference on Computers and Games. 2008; pp 1–12.
- (58) Barth, D.; David, O.; Quessette, F.; Reinhard, V.; Strozecki, Y.; Vial, S. Efficient Generation of Stable Planar Cages for Chemistry. 2015.
- (59) Saito, J.; Winands, M.; Uiterwijk, J.; Van Den Herik, H. Grouping Nodes for Monte-Carlo Tree Search. 19th Belgian-Dutch Conference on Artificial Intelligence. 2007; pp 276–283.
- (60) Childs, B. E.; Brodeur, J. H.; Kocsis, L. Transpositions and move groups in Monte Carlo tree search. 2008 IEEE Symposium On Computational Intelligence and Games. 2008; pp 389–395.
- (61) Bosc, G.; Boulicaut, J.-F.; Raïssi, C.; Kaytoue, M. Anytime Discovery of a Diverse Set of Patterns with Monte Carlo Tree Search. 2016.
- (62) Drake, P.; Uurtamo, S. Move ordering vs heavy playouts: Where should heuristics be applied in Monte Carlo Go,”. Proceedings of the 3rd North American Game-On Conference. 2007; pp 171–175.

- (63) James, S.; Rosman, B. S.; Konidakis, G. An investigation into the effectiveness of heavy rollouts in UCT. **2016**,
- (64) Toyoda, T.; Kotani, Y. Monte Carlo Go Using Previous Simulation Results. Proceedings of the 2010 International Conference on Technologies and Applications of Artificial Intelligence. USA, 2010; p 182–186.
- (65) Loth, M.; Sebag, M.; Hamadi, Y.; Schoenauer, M. Bandit-based search for constraint programming. International Conference on Principles and Practice of Constraint Programming. 2013; pp 464–480.
- (66) Leong, Z.; Tan, T. L. Robust cluster expansion of multicomponent systems using structured sparsity. *Phys. Rev. B* **2019**, *100*, 134108.
- (67) Nelson, L. J.; Ozoliņš, V.; Reese, C. S.; Zhou, F.; Hart, G. L. W. Cluster expansion made easy with Bayesian compressive sensing. *Phys. Rev. B* **2013**, *88*, 155105.
- (68) Nelson, L. J.; Hart, G. L. W.; Zhou, F.; Ozoliņš, V. Compressive sensing as a paradigm for building physics models. *Phys. Rev. B* **2013**, *87*, 035125.