Resolving the Coverage Dependence of Surface Reaction Kinetics with Machine Learning and Automated Quantum Chemistry Workflows

Matthew S. Johnson,[†] David H. Bross,[‡] and Judit Zádor[*,†]

[†]*Combustion Research Facility, Sandia National Laboratories, California 94551-0969, United States*

[‡]*Chemical Sciences and Engineering Division, Argonne National Laboratory, Lemont, Illinois 60439, United States*

E-mail: jzador@sandia.gov

https://doi.org/10.26434/chemrxiv-2024-36w9w **ORCID:** https://orcid.org/0000-0002-4624-2852 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

Abstract

Microkinetic models for catalytic systems require estimation of many thermodynamic and kinetic parameters that can be calculated for isolated species and transition states using ab initio methods. However, the presence of nearby co-adsorbates on the surface can dramatically alter these thermodynamic and kinetic parameters causing them to be dependent on species coverage fractions. As there are combinatorially many co-adsorbed configurations on the surface, computing the coverage dependence of these parameters is far less straightforward.

We present a framework for generating and applying machine learning models to predict coverage dependent parameters for microkinetic models. Our toolkit enables automatic calculation and evaluation of co-adsorbed configurations allowing us to sample 2000 co-adsorbed adsorbates and transition states (TSs) for a diverse set of 9 reactions on Cu111, a challenging surface, with four possible co-adsorbates. This dataset was then used to train subgraph isomorphic decision trees (SIDTs) to predict the stability and association energy of configurations. With which we were able to achieve mean absolute errors (MAEs) of 0.106 eV on adsorbates, 0.172 eV on TSs, and due to natural error cancellation in SIDTs for relative properties 0.130 eV on reaction energies and 0.180 eV on activation barriers. We then explain how to use these models to predict coverage dependent corrections for arbitrary adsorbates and TSs and demonstrate on H^*, HO^* and O^* comparing the generated SIDT model with an iteratively refined version.

Introduction

Heterogeneous catalysis plays an incredibly important role in energy technologies and chemical manufacturing. Catalytic systems involve many elementary reactions and are sensitive to temperature, pressure, and the nature of the catalyst. In order to predict the behavior of these systems at a range of conditions usually one would build and simulate a microkinetic (MKM) or a kinetic Monte Carlo (KMC) model.

https://doi.org/10.26434/chemrxiv-2024-36w9w **ORCID:** https://orcid.org/0000-0002-4624-2852 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

Under low-coverage conditions it is often adequate to treat adsorbates and reactions as if they are isolated on the surface, assuming an otherwise empty surface. However, in many systems and under many conditions the surface has many adsorbates packed closely together. In these co-adsorbed systems the lateral interactions between co-adsorbed species can very significantly alter the thermochemistry and kinetics.[1–9] However, brute force computation of the minimum association energy of an adsorbate and co-adsorbates at a given coverage requires one to compute the average binding energy of every conceivable configuration at that coverage. Even just computing the minimum association energy of an adsorbate in the presence of a single type of co-adsorbate is computationally very expensive. Consideration of several co-adsorbates is even less computationally feasible, therefore, the effects are often either ignored or approximated crudely. Furthermore, due to the additional challenges most researchers do not attempt to compute the coverage dependence of transition state properties, and instead use Brønstedt-Evans-Polanyi (BEP) relations.[10,11]

To tackle these challenges researchers have developed a variety of approaches. In one approach, the dimensions of the periodic system are adjusted to achieve the desired coverage with the smallest system possible and reduce the computational cost.[12] Additionally, researchers have built models for predicting the energies of co-adsorbed configurations like the cluster expansion (CE) model[13] that decomposes the overall energy into contributions from groups of N specific adsorbates to a given N.

However, these approaches have a number of weaknesses. Size reduced periodic cells inherently assume that the minimum energy coverage pattern is periodic in the size of the unit cell. In many cases lowest energy configurations cannot be captured with reduced size unit cells.[14] In fact, such reduced size cell configurations may not even be minima on the potential energy surface of a larger cell. Consider a 1×1 cell with one O*. The energy for the reaction sequence 2 O$^* \rightarrow$ O$_2^* \rightarrow$ O$_2 + *$ can be comparable to predicted lateral interactions between two O*.[12,15] Given the highest barrier in this reaction sequence is usually not much higher in energy than O$_2 + *$, it is very believable that at high coverage 2 O$^* \rightarrow$ O$_2^*$ or O$_2 + *$

https://doi.org/10.26434/chemrxiv-2024-36w9w **ORCID:** https://orcid.org/0000-0002-4624-2852 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

could occur barrierlessly spontaneously, with the 2 O^* state not being stable. However, in the 1×1 periodic cell the distance between O^*'s is fixed making the stable configuration with O's bonded pairwise with each other unachievable, causing an optimization to find a configuration that is unstable on any realistically sized slab.

CE parameters are particular to the adsorbates and interaction terms they were computed for.[16] Thus these models are primarily only useful for set interactions of adsorbates and co-adsorbates they have been explicitly fit to and cannot be used to predict fit interactions with new adsorbates or co-adsorbates. Additionally, due to the combinatorics it is rare to fit them out to more than clusters of $N = 3$. So in practice this method may have some difficulties at higher coverages.

In contrast to CE, machine learning techniques do not have the limitations of cluster expansion methods. They can learn from diverse datasets and predict on interactions not included in the training data. However, popular deep neural network (DNNs) based machine learning requires large amounts of training data that can be computationally very expensive to obtain. Additionally, DNNs are not interpretable making them difficult to analyze and improve.

The subgraph isomorphic secision tree (SIDT) machine learning method provides an alternative to DNNs and is free of the above weaknesses.[17,18] SIDTs are decision trees made up of nodes associated with molecular substructures represented as molecular subgraphs. They are evaluated by descending a target graph structure down the tree to the nodes with subgraphs it matches until it reaches the most specific matching node and making a prediction based on either the nodes matched or the final node. SIDTs can be applied to datasets too small for DNNs, they are straightforward to extend and retrain, and the substructures in the tree are inherently visualizable making SIDTs easy to analyze and interpret, and thus modify the generation process and tree itself to achieve desired outcomes. Additionally, SIDTs have a property that enables unique inherent error cancellation on many important chemical problems that DNNs do not. One such important case involves kinetics,

https://doi.org/10.26434/chemrxiv-2024-36w9w **ORCID:** https://orcid.org/0000-0002-4624-2852 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

where rate coefficients are not dependent on the absolute energy of any given configuration, only on the energy differences between reactants and the transition state and reactants and products which are used to compute the forward and reverse rate coefficients for a given reaction. When the SIDT predicts the association energy of a reactant configuration and that of a transition state or product configuration, most of the matched subgraphs associated with interactions between adsorbates are unchanged, because only interactions close to the reaction center are different in these structures. As a result, predictions on unchanged interactions will cancel exactly in the SIDT and our error is only associated with the subgraphs that are modified by the reaction, i.e., bond breaking and forming. This property makes SIDTs significantly more accurate on the relative properties that actually matter for kinetics than one would expect from a given level of absolute accuracy.

In this work we present a framework for automatically computing co-adsorbed configurations and using machine learning to predict the coverage dependent energetics of adsorbates and transition states and thus the coverage dependence of rate coefficients and thermochemistry. We automatically generate a training dataset of co-adsorbed adsorbates and transition states on a 3x3x4 Cu111 slab, a challenging surface. The size of the slab is chosen to minimize interaction between periodic. We use the dataset to train a sequence of SIDTs to predict whether a configuration is stable or not and to predict the association energy of the configuration for both co-adsorbed adsorbates and transition states. We show the effectiveness of the SIDT predictors and demonstrate the error-cancellation property discussed above. We also show how to go from association energy predictions to coverage dependent rate coefficients.

https://doi.org/10.26434/chemrxiv-2024-36w9w **ORCID:** https://orcid.org/0000-0002-4624-2852 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

Methods

Dataset Generation

We started our co-adsorbed calculations from a set of isolated calculations. We took the lowest energy configurations for 12 adsorbates and 9 transition states on Cu111 calculated by Johnson et al.[19] using our software, Pynta. This set, listed in Table 1, includes transition states for a range of different reaction classes, adsorbates consisting of H, C, O and N atoms and one bidentate adsorbate. In this work we selected four adsorbates to be co-adsorbates for purposes of sampling: H, N, and O atoms, and OH.

Table 1: Reactions on Cu111 considered in this work for lateral interactions.

Reaction
$H^* + O^* \longleftrightarrow HO^* + *$
$OCH^* + * \longleftrightarrow OC^* + H^*$
$H^* + * \longleftrightarrow H^* + *$
$OC^* + O^* \longleftrightarrow CO_2 + 2*$
$HO^* + H^* \longleftrightarrow H_2O + 2*$
$HOCH_2{}^* + * \longleftrightarrow CH_3O^* + *$
$OCHO^* + * \longleftrightarrow CO_2 + H^* + *$
$H^* + OCH_2O^{**} \longleftrightarrow HOCH_2O^* + 2*$
$N^* + CH_3O^* \longleftrightarrow CH_2O + NH^* + *$

All of our calculations were done using the same software, methods and parameters as in Johnson et al. We used the BEEF-vdW functional[20] with PBE-KJPAW pseudopotentials and an energy cutoff of 40 Ry as implemented in Quantum Espresso[21,22] for a 3x3x4 Cu111 slab with a 3x3x1 k-point grid. All geometry optimizations targeting wells were done in two stages: first using the MDMin method implemented in the Atomic Simulation Environment (ASE) until $f_{\max} \leq 0.5$ eV/A and then using ASE's BFGSLineSearch algorithm until $f_{\max} \leq 0.02$ eV/A.[23] Saddle point optimizations and intrinsic reaction coordinate (IRC) calculations were done using Sella until $f_{\max} \leq 0.02$ eV/A and $f_{\max} \leq 0.1$ eV/A respectively.[24–26] Vibrational calculations were run using ASE's vibrations module.[23]

Unless otherwise specified, all placements for optimization on the surface were done by

https://doi.org/10.26434/chemrxiv-2024-36w9w ORCID: https://orcid.org/0000-0002-4624-2852 Content not peer-reviewed by ChemRxiv. License: CC BY-NC-ND 4.0

copying the isolated configuration for the adsorbate or transition state and then placing co-adsorbates on the selected sites using Pynta's placement algorithm.[19] 3D configurations are converted to 2D graphs automatically using a set of algorithms contained within Pynta. These 2D representations are key to employing SIDT. For adsorbates this process is relatively straightforward. ASE's analysis tool was used to identify covalent bonds within the adsorbates; possible surface bonds are identified by searching for the closest site within 2.5 Å of the relevant adsorbate atom, and only considering sites on which the associated adsorbate is stable under isolated conditions. We then complete the 2D description (the graph of the system) by incrementing bond multiplicities to satisfy octet rule, with surface bonds having the lowest priority.

For TSs we considered a number of additional factors. Reaction bonds that break/form in the reaction are identified based on the original reaction template associated with the TS. Since transition state atoms may be close to the surface without being properly associated with a site, we included both reactant and product sites in the set of valid sites to determine the 2D structure. Since the reaction bonds in the 2D representation do not have a well defined order we cannot always satisfy the octet rule when we increment bonds. As a result, we can end up with extra bonds to the surface that are artifacts of the process. To handle this we removed surface bonds if they were a single bond and the associated atom already had two reaction bonds. This covers most common cases and all reactions considered here.

TSs for diffusion type reactions, where one atom has two reaction bonds to different sites, pose a further challenge for generating 2D representations, because we cannot simply form the reaction bonds with the closest stable site. Instead, we seek a pair of sites that capture the origin and destination of the diffusion process. To identify the right pair of sites we considered five criteria. First, we required that the vector connecting the two sites and the normal mode vector associated with the imaginary frequency are aligned well

$$| < \mathbf{v}_{\text{sites}}, \mathbf{v}_{\text{imagfreq}} > | \geq 0.95 \tag{1}$$

https://doi.org/10.26434/chemrxiv-2024-36w9w **ORCID:** https://orcid.org/0000-0002-4624-2852 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

where $\mathbf{v}$ denotes the associated vector or motion. Second, we required that the distance between the sites is less than 3 Å. Third, we required that the distance between the adsorbing atom of the adsorbate and any site, d_{site} be less than 2 Å. Fourth, we only considered sites where the reactants or products were stable in the isolated calculations. Finally, we also defined a measure, h, of how closely the atom is positioned to the halfway point between the sites

$$h = \frac{||\mathbf{u}_1 + \mathbf{u}_2||}{||\mathbf{u}_1|| + ||\mathbf{u}_2||} \tag{2}$$

where $\mathbf{u}_i$ is the vector from site i to the atom. When more than one pair of sites fulfilled the previous criteria, we chose the pair that maximizes $1/(d_{\text{site}} \times h)$.

The techniques discussed above are sufficient to generate the 2D representations of TSs. However, for TSs we also need to validate that the saddle point optimized in the presence of the co-adsorbates still connects the reactants and products of the original reaction. We found that two criteria were sufficient to separate correct from incorrect TSs in a 100 TS subset of our data. We first required that the normal mode corresponding to the imaginary frequency in the co-adsorbed case, $\mathbf{v}_{coad}$, aligns with the isolated mode $\mathbf{v}_{isolated}$

$$| < \mathbf{v}_{\text{isolated}}, \mathbf{v}_{\text{coad}} > | \geq 0.7 \tag{3}$$

We additionally required that any co-adsorbate atoms be more than 1.1 times the covalent bond cutoff threshold from ASE away from any TS atoms that are involved in breaking/forming bonds.

One major challenge of sampling the co-adsorbed space is that many configurations one might propose are not stable. Lateral interactions may prevent two adsorbates from being placed next to each other, surface restructuring may affect the stability of old sites or create new sites, and reactions that normally have a barrier may occur spontaneously at higher coverages. We simultaneously mitigated this challenge and provided a useful base set of samples by first calculating every unique and valid placement pairing between every adsorbate and

https://doi.org/10.26434/chemrxiv-2024-36w9w **ORCID:** https://orcid.org/0000-0002-4624-2852 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

every selected co-adsorbate that put them within a maximum distance of 3 Å. We analyzed the results comparing the initial 2D graph and the 2D graph after optimization to determine which pairwise configurations are stable.

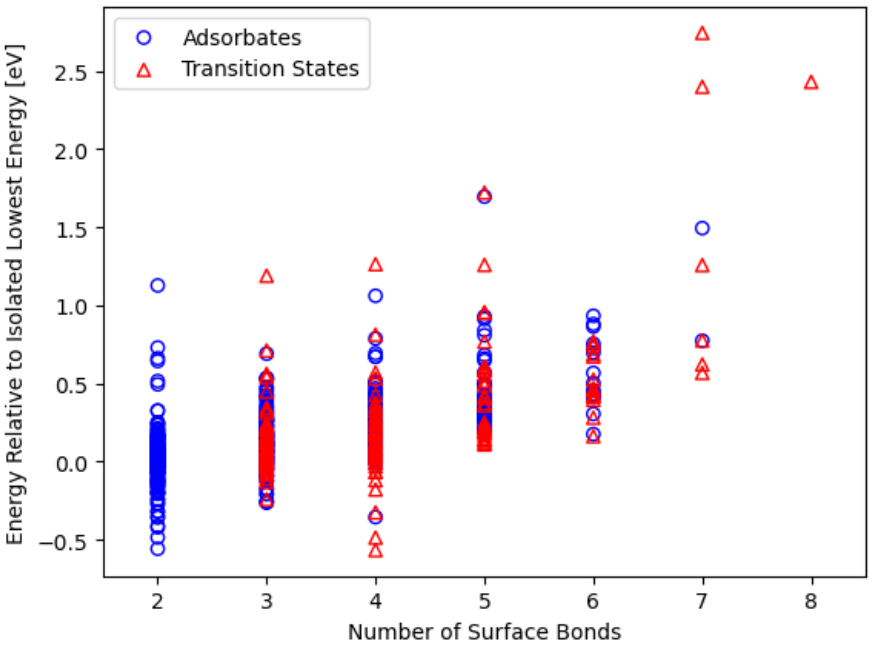


Figure 1: Scatter of the dataset energies and number of surface bonds.

To generate a random, yet balanced set, we first chose either to sample an adsorbate or a TS with equal probability. We then chose a single random co-adsorbate at 90% probability and a random sample of mixed co-adsorbates at 10% probability. To determine the sampled coverage fraction we drew a uniformly distributed sample from [0,1], which was then rounded to an integer number of co-adsorbates to put on the slab. The co-adsorbates were distributed randomly over the stable sites. Samples that involved sub-configurations that matched a failed pairwise optimization were rejected. For all unique successful TS optimizations we ran an IRC to find the reactant and product configurations and optimized and computed frequencies for both. This enabled explicit sampling of activation barriers and reaction energies.

Drawing and running calculations for 2,000 samples gave us 477 unique and valid co-adsorbed adsorbate configurations and 207 unique and valid co-adsorbed TS configurations.

https://doi.org/10.26434/chemrxiv-2024-36w9w ORCID: https://orcid.org/0000-0002-4624-2852 Content not peer-reviewed by ChemRxiv. License: CC BY-NC-ND 4.0

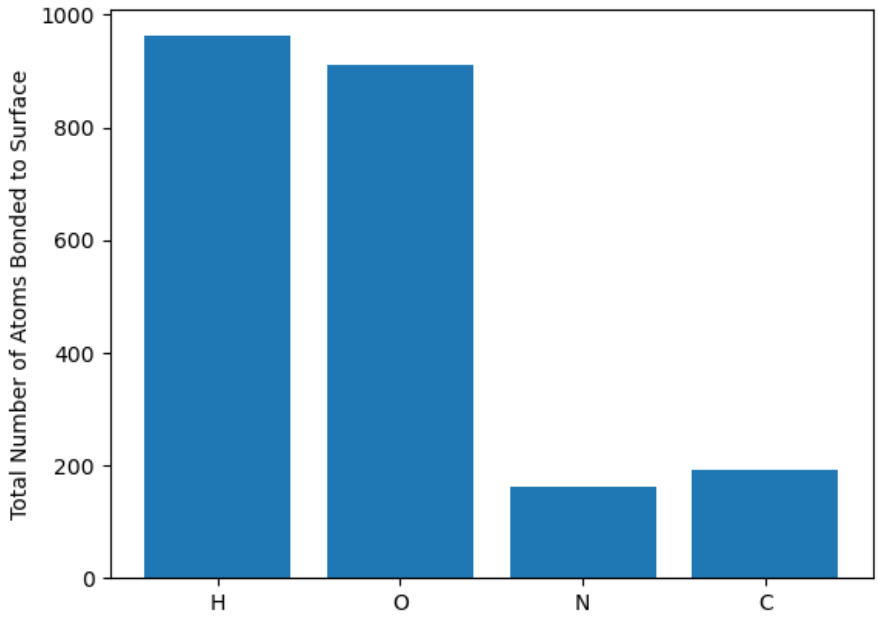


Figure 2: Plot of the number of atoms bonded to the surface in dataset configurations by element.

Figure 1 shows the energy of these configurations as a function of the number of surface bonds, while Figure 2 shows the distribution of the elements bonded to the surface across the whole dataset, i.e., including the adsorbing atoms of the central adsorbate or TS and of the co-adsorbates. In general, H atoms tend to have weaker inter-adsorbate interactions owing to its smaller size, making co-adsorbed configurations with H more likely to be stable and optimize successfully. For this reason it is unsurprising that H has so many occurrences in the dataset. The similar number of O occurences is unsurprising, given O and OH are co-adsorbates and many of the adsorbate and TSs targeted involve bonds between O and the surface. Given none of the co-adsorbates involve a bond between the surface and C the significantly lower number of occurrences is unsurprising. However, the low number of N occurences is unexpected. N does not occur in many of the isolated adsorbate and TSs considered (see Table 1, but it is a co-adsorbate, so one would expect it to exist in at least a quarter of proposed sample configurations. This seems to suggest that configurations involving N in general have lower optimization success rates than O and OH on this surface.

In Figure 3 we examine the distribution of the differences between the activation barrier in

https://doi.org/10.26434/chemrxiv-2024-36w9w ORCID: https://orcid.org/0000-0002-4624-2852 Content not peer-reviewed by ChemRxiv. License: CC BY-NC-ND 4.0

co-adsorbed configurations and the corresponding isolated configuration for configurations in the dataset. Differences can be quite large in both negative and positive directions and range from approximately -0.8 to 1.0 eV. The center of the distribution for individual reactions (Figure 3b) appears to differ significantly from zero in at least a few cases. The range for individual reactions seems to always span at least about 0.5 eV, and HO* →H*+ O* spans approximately 1.5 eV.

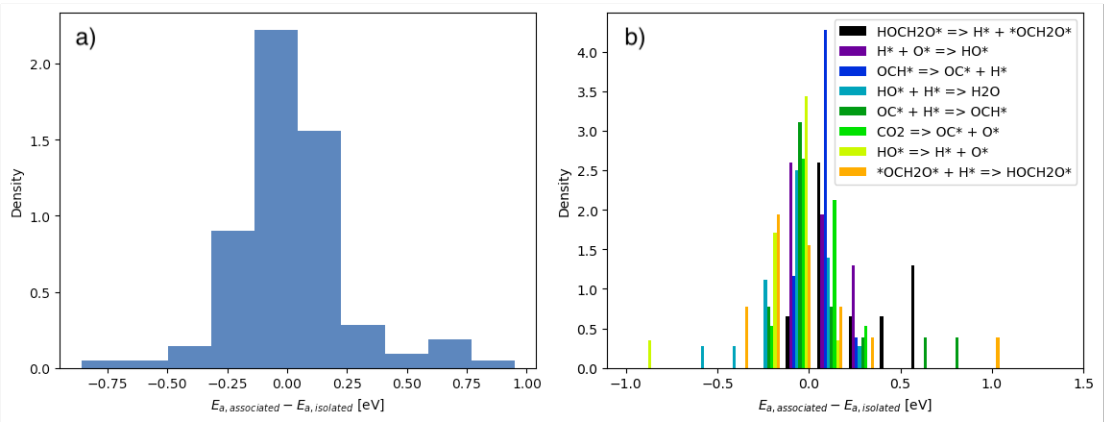


Figure 3: Distribution of differences between the activation barrier of co-adsorbed configurations and the activation barrier of the corresponding isolated reaction (a) also shown separated by reaction (b).

Machine Learning

Our goal is to predict the change in energetics of stable and TS surface configurations relative to their isolated energetics due to the presence of co-adsorbates. We divided the prediction process into two primary steps: 1) decide whether a proposed surface configuration is stable and if so 2) predict the association energy of the configuration. For the former we use multi-evaluation SIDT binary classifiers, and for the latter we use multi-evaluation SIDT regressors as implemented in our software, PySIDT.[27] Both of these SIDT algorithms decompose a 2D graph representation of the chemical configuration into a set of chemical substructures contributing to the prediction. In the case of the classifier, each substructure is predicted to either be locally stable or unstable, and in this application if any prediction is unstable the

https://doi.org/10.26434/chemrxiv-2024-36w9w ORCID: https://orcid.org/0000-0002-4624-2852 Content not peer-reviewed by ChemRxiv. License: CC BY-NC-ND 4.0

configuration is classified as unstable. In the case of the regressor, an energetic contribution is predicted from each substructure and summing across all substructures gives the associated energy.

For our 2D representation we represented the periodic slab in its entirety resolving each site and adsorbate atom as nodes. Edges were included between all covalently bonded atoms, each site and adsorbate atom bound to that site, and between neighboring sites as defined by the ACAT software.[28] We used RMG's cheminformatics engine and molecular representation software[29] within PySIDT[27] for all operations on this representation.

When selecting decompositions of the 2D graphs into subgraphs there are two primary considerations: (i) we would like the set of decompositions to involve minimum redundancy in chemical information and (ii) we would like the decompositions to locally include all chemical information we need to make the prediction. The simplest decomposition one might think of is to look locally at each individual atom and site in the configuration. However, we are examining inter-adsorbate interactions so we would not necessarily expect every atom or site to have a significant unique contribution to the energetics. Looking at surface bonds should significantly reduce redundancy, however, since we study interactions between adsorbates that may be far apart in the 2D representation, we need to consider at least pairs of surface bonds. While pairs of surface bonds are sufficient to encode the interactions between co-adsorbates we are interested in resolving, unlike many prior studies, we have allowed co-adsorbates to be adsorbed at sites that do not correspond to their lowest energy isolated configuration. Therefore, we also need to predict the energetics of moving adsorbates between sites. For this reason our overall representation included both surface bonds alone (to account for energetics of adsorbates binding to individual sites) and pairs of surface bonds (to account for the lateral interactions). Triads of surface bonds (three-body interactions) can be used as well. However, while we have found including triad-wise interactions to be beneficial for learning coverage dependence in some cases, it did not improve performance in this work. It should be noted that SIDT does not need triad-wise

https://doi.org/10.26434/chemrxiv-2024-36w9w **ORCID:** https://orcid.org/0000-0002-4624-2852 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

decompositions to learn triad-wise interactions since the pair-wise decompositions can be grown to resolve more than two adsorbates in the learning processes.

To simplify and compartmentalize the training processes, we first train trees to predict the single surface-bond interactions on isolated data from Johnson et al.,[?] including not just the lowest energy, but all valid isolated structures. We then train a second tree on the co-adsorbed datasets generated in this work to add the contribution from the interactions of pairs of surface-bonds. For the classifier we only train the second tree on configurations the single-surface-bond tree predicts to be stable, and for the regressor we do delta learning, subtracting the single-surface-bond-tree prediction from the dataset before training. This architecture is shown in Figure 4.

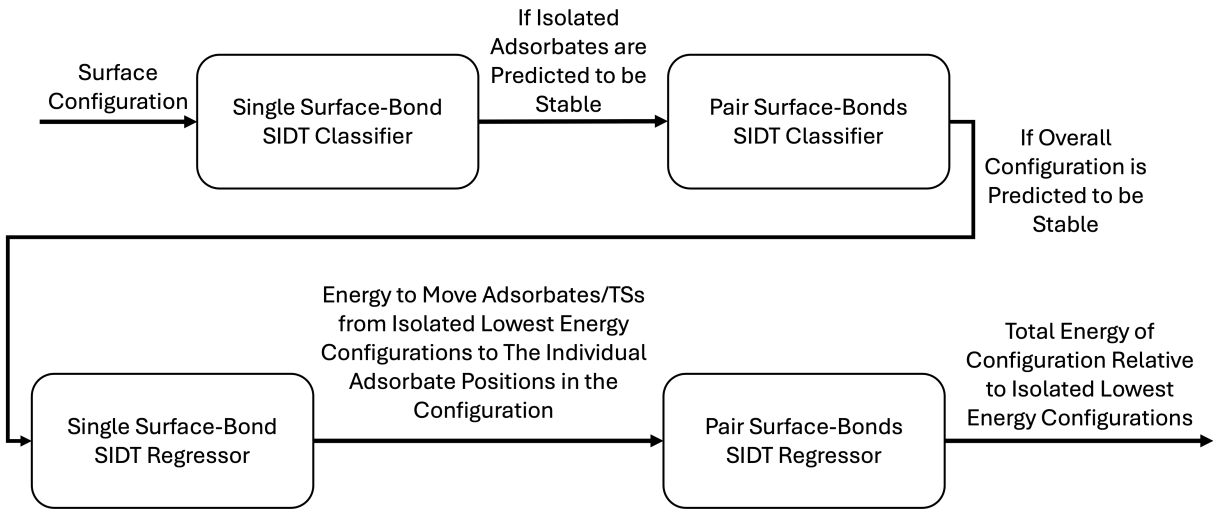


Figure 4: Diagram for computing the stability and energy of a given input surface configuration.

Results

Stability Predictions for Co-Adsorbed Systems

Our dataset provided stability labels for 1720 unique and otherwise valid configurations. The configurations included unstable configurations that were proposed and failed to preserve the

https://doi.org/10.26434/chemrxiv-2024-36w9w ORCID: https://orcid.org/0000-0002-4624-2852 Content not peer-reviewed by ChemRxiv. License: CC BY-NC-ND 4.0

original graph representation upon optimization, stable configurations that were proposed and successfully optimized, and stable configurations that were found in failed optimizations. Of these configurations, 91 were predicted to be unstable by the single surface-bond classifier, 75 of which incorrectly predicted to be unstable and 16 of which were correctly predicted to be unstable. The pair surface-bonds classifier was trained using a 8:1:1 train:validation:test split on the other 1629 configurations. Each iteration the SIDT computed the accuracy on the validation set and at the end of the run it reverted to the tree from the iteration with the best validation accuracy. The confusion matrix for the test set of the pair-surface-bonds classifier is available in Table 2. This implies an accuracy (fraction of classifications that are right) of 86% and a precision (fraction of predicted Trues that are correct) of 84%. Given the inherent challenge of determining the substructures that underlie all of the many different kinds of stability in these systems this is good performance from a machine learning perspective. From a more practical and problem oriented perspective, we do not need extremely high levels of accuracy for the stability classifier because high association energy predictions and instability are inherently correlated. A co-adsorbed configuration that is unstable and cannot be occupied and a co-adsorbed configuration that is high in energy relative to other configurations at identical coverage and thus has very low occupation are both unimportant kinetically.

Table 2: Test set confusion matrix for stability of interactions between co-adsorbates.

	Predict True	Predict False
Value True	91	9
Value False	14	49

Association Energy Predictions for Co-Adsorbed Systems

For association energy prediction, as noted earlier, we only have $477 + 207 = 684$ valid configurations. For this smaller dataset we trained on the full dataset and computed leave-one-out errors for every training point. We first trained the tree out to 152 nodes only on

https://doi.org/10.26434/chemrxiv-2024-36w9w **ORCID:** https://orcid.org/0000-0002-4624-2852 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

configurations with two co-adsorbates and no TSs before training out to 475 nodes (targeting 470) on the full dataset.

The resulting comparison plot is available in Figure 5a and the associated uncertainty calibration plot is available in Figure 5b. For a more detailed view, Figures 5c and 5d show the parity plot for the adsorbates and the TSs separately. The error analysis gives an overall MAE of 0.126 eV, a MAE on adsorbates of 0.106 eV and a MAE on TSs of 0.172 eV. The uncertainty calibration shows that the model is slightly underconfident at small confidence intervals and slightly overconfidence at large confidence intervals, but in general, the model uncertainties appear to be a good representation of actual uncertainties. This is especially encouraging given that it is challenging to predict accurate uncertainties for some of the more unusual configurations in this dataset, especially for TSs.

Activation Barrier and Reaction Energy Correction Predictions for Co-Adsorbed Systems

While lateral interactions are often discussed in terms of the energies of specific configurations, this is not the most relevant quantity for kinetics, which, as discussed earlier, are only sensitive to relative configuration energies: activation barriers and reaction energies rather than the absolute energies. Our dataset offers a unique opportunity, allowing us to look directly at the relevant properties using the optimized endpoint configurations from the IRCs for each unique TS. We present parity plots for activation barriers and reaction energies in Figures 6a and 6b respectively. We achieve an MAE of 0.180 eV for activation barriers and 0.130 eV for reaction energy. Noting the MAEs of 0.106 eV and 0.172 eV for adsorbate and TS energy predictions, the low MAEs of the relative quantities suggest significant error cancellation. Assuming no correlation and normally distributed errors, one would expect that

$$\sigma_{\mathrm{Ea}}^2 \approx \sigma_{\mathrm{TS}}^2 + \sigma_{\mathrm{Ad}}^2 \tag{4}$$

https://doi.org/10.26434/chemrxiv-2024-36w9w **ORCID:** https://orcid.org/0000-0002-4624-2852 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

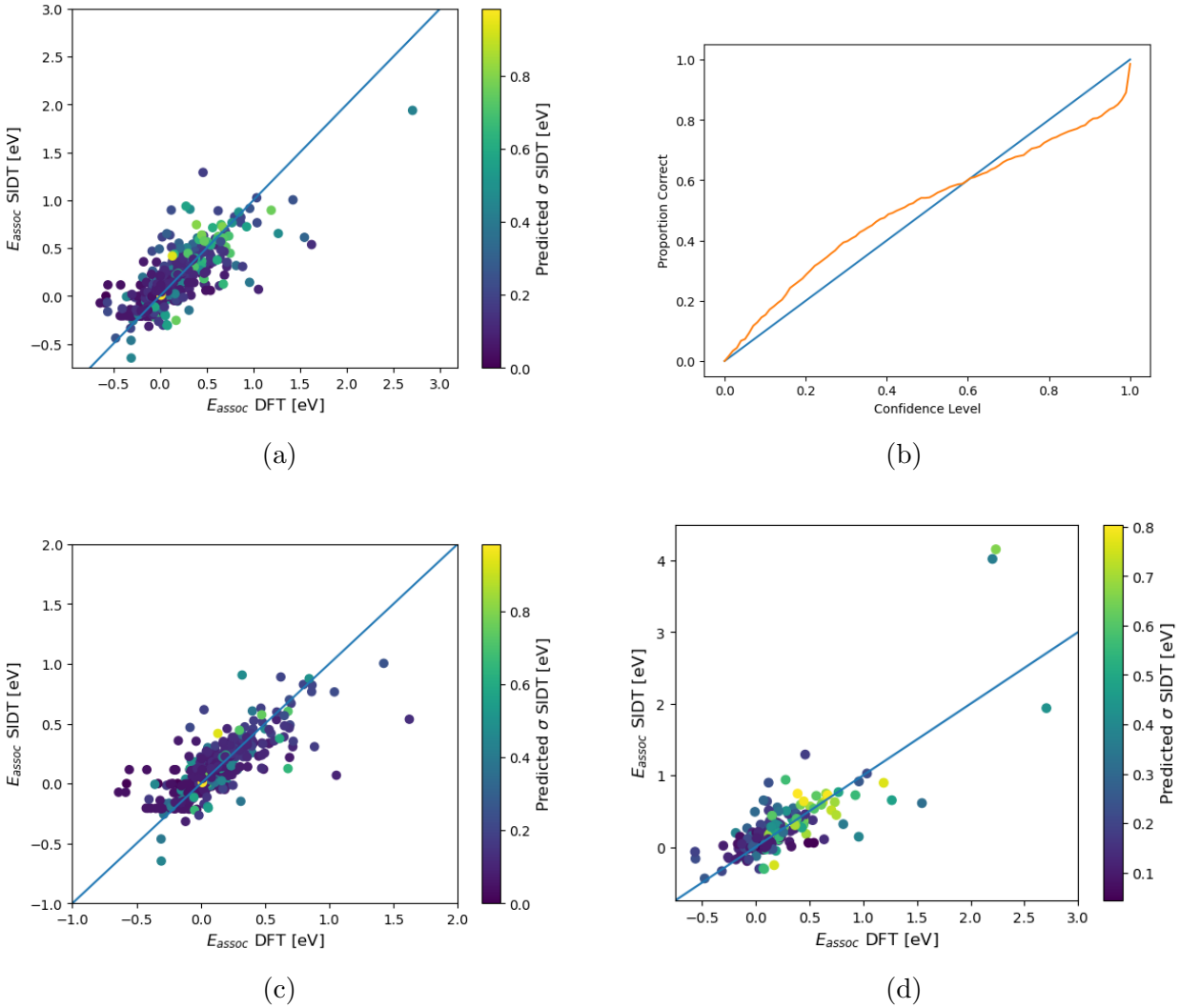
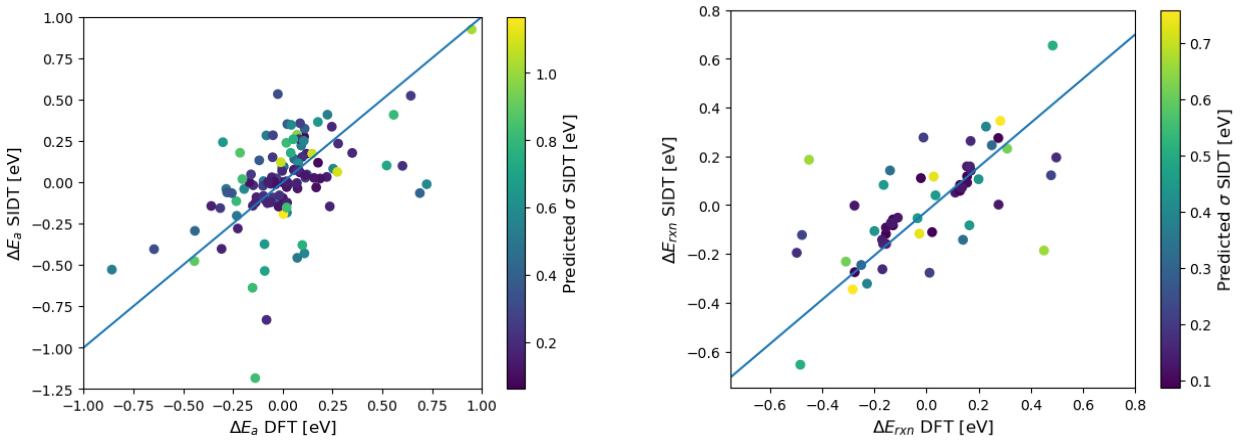


Figure 5: SIDT performance on absolute energies. (a) Parity plot for leave-one-out errors in association energy for all configurations. (b) Uncertainty calibration plot for leave-one-out and estimated errors in association energy for all configurations. (c) Parity plot for leave-one-out errors in association energy for adsorbate configurations. (d) Parity plot for leave-one-out errors in association energy for TS configurations.

https://doi.org/10.26434/chemrxiv-2024-36w9w **ORCID:** https://orcid.org/0000-0002-4624-2852 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0



(a) (b)

Figure 6: SIDT performance on relative energies. (a) Parity plot for leave-one-out errors in activation barrier corrections. (b) Parity plot for leave-one-out errors in reaction energy corrections.

and

$$\sigma^2_{\Delta E_{\text{rxn}}} \approx 2\sigma^2_{\text{Ad}} \tag{5}$$

where σ denotes the standard deviation of the property, TS denotes the transition state energy correction, Ad denotes the adsorbate energy correction, Ea denotes the activation barrier energy correction and ΔE_{rxn} denotes the reaction energy correction. The above equations imply that the error in the relative properties should be significantly larger than the error in the absolute properties. However, for our model $\sigma^2_{\text{Ea}}/(\sigma^2_{\text{TS}} + \sigma^2_{\text{Ad}}) = 0.561$ and $\sigma^2_{\Delta E_{\text{rxn}}}/(2\sigma^2_{\text{Ad}}) = 0.694$, implying that for the relative energies our model is significantly more accurate that one would expect from the the accuracy of the absolute predictions. This demonstrates the power of natural error cancellation inherent in the structure of the SIDT predictor.

In order to explain how the error cancellation occurs, let us consider inference for the pair surface-bonds SIDT regressor used above. As discussed earlier, inference occurs by finding all pairs of surface bonds, tagging the surface bonds in each, descending each down the SIDT summing the contribution from each node touched in the descent and then summing the contribution from each pair of surface bonds. With this in mind let us consider computing

https://doi.org/10.26434/chemrxiv-2024-36w9w ORCID: https://orcid.org/0000-0002-4624-2852 Content not peer-reviewed by ChemRxiv. License: CC BY-NC-ND 4.0

the difference in energy between a given reactant configuration and the corresponding TS or product configuration. Only a handful of bonds are created or formed during a reaction so the two configurations are usually not very different. Especially at higher coverages many of the interactions may be unchanged and thus their contributions to the two configurations are identical and cancel exactly removing any contribution to the variance from that interaction. In many other cases the interaction is only slightly changed resulting in an SIDT descend that only differs deep into the tree. In these cases the contributions from the upper nodes, before the descents diverge, cancel exactly and only the nodes farther down the tree, where the energy contributions should be much smaller, contribute to the variance. This natural error cancellation allows our SIDT predictor to be much more accurate on these relative properties that are actually important for simulations than one would expect from a given level of absolute accuracy.

Coverage Dependence Corrections in Microkinetic Models

Simply considering the difference in energy between reactant, TS, and product configurations is sufficient for KMC simulations. However, mean-field kinetics simulators[30–32] do not resolve the exact configurations of the co-adsorbates. For mean-field parameterizations we need to predict energies as a function of average coverage. This is typically done for a given species or TS by taking the lowest energy configuration at each coverage (corresponding to an integer number of co-adsorbates).

To find the lowest energy configuration, we generate all unique stable 2D representations at a given coverage, and then make energy predictions on them. We start with a list of all stable isolated configurations of the target adsorbate or TS. For a given co-adsorbate we iterate through all of the sites on the surface (here we always used a 3×3 slab). For each configuration in our list, if the site is free and placing the co-adsorbate on the site results in a configuration that our SIDT stability classifiers predict to be stable and is unique compared to the configurations in our list, we add the new configuration to the list. Once we have

https://doi.org/10.26434/chemrxiv-2024-36w9w **ORCID:** https://orcid.org/0000-0002-4624-2852 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

iterated through all sites, we have a list of all feasible configurations. We then use our SIDTs to predict the association energy of each configuration and find the lowest energy corresponding to each integer number of co-adsorbates. One can also integrate association energy predictions into the feasibility search and include an energy-based criterion to reduce the number of feasible configurations generated, but we did not need to do so for the cases discussed here.

For a given adsorbate or TS this algorithm gives us a minimum association energy at a sequence of coverage values for the given co-adsorbate. We can then calculate the coverage-dependent corrections based on

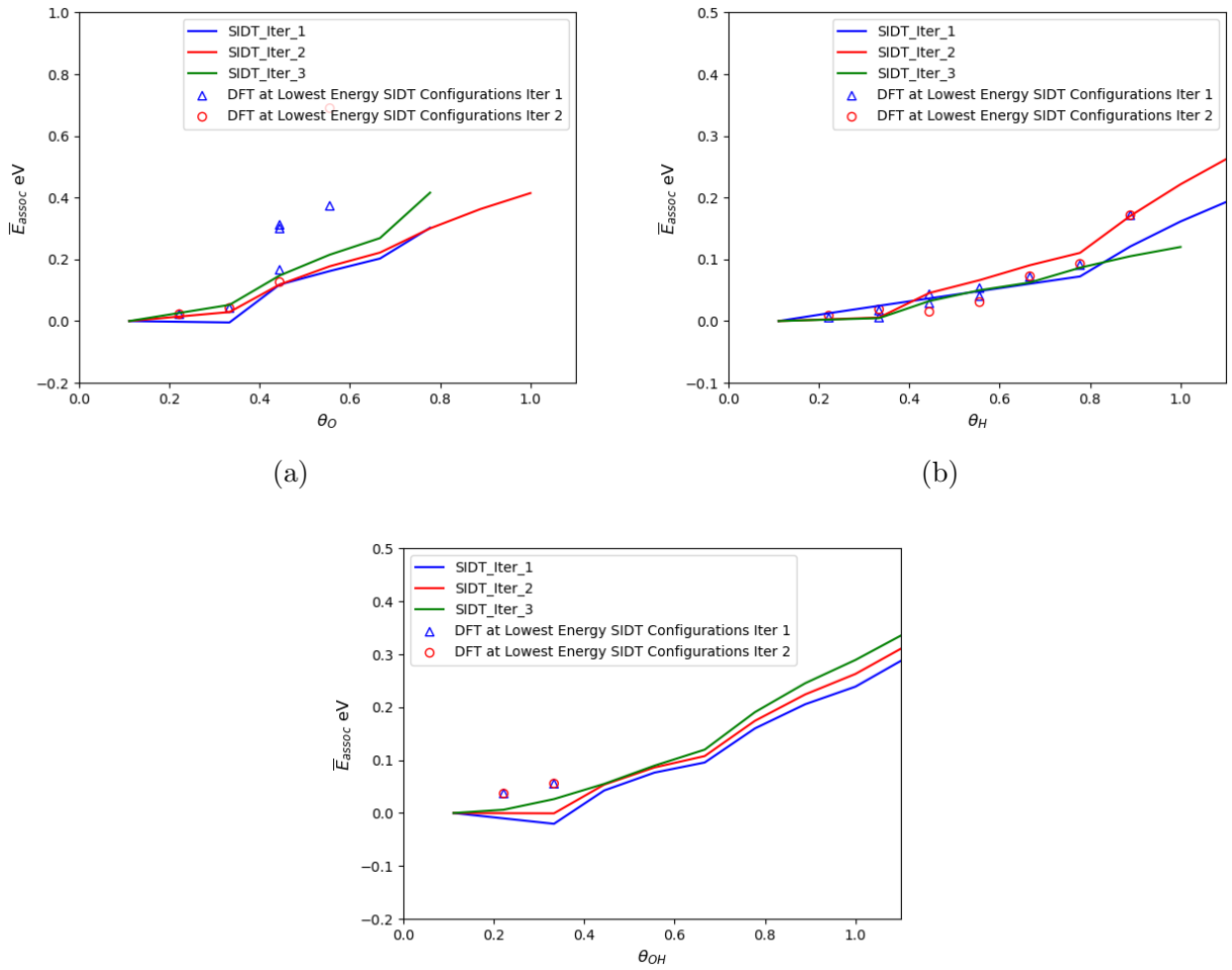
$$\Delta E_{\mathrm{spc},N_{\mathrm{coad}}} = E_{\mathrm{spc},N_{\mathrm{coad}}} - E_{N_{\mathrm{coad}}} \tag{6}$$

where $E_{\mathrm{spc},N_{coad}}$ is the predicted lowest association energy of the the adsorbate or TS species and N_{coad} co-adsorbates, $E_{N_{\mathrm{coad}}}$ is the predicted lowest association energy of N_{coad} co-adsorbates on the surface and $\Delta E_{\mathrm{spc},N_{\mathrm{coad}}}$ is the energy correction for the adsorbate or TS energy with N_{coad} co-adsorbates on the surface.

Given the combinatorial nature of the configurational space, computing the exact DFT correction for a single adsorbate or TS with respect to one co-adsorbate species would have required computational expense on par with generating the entire dataset in this work, making it computationally too expensive to present parity plots. Instead, here we used a simple iterative refinement procedure to analyze the accuracy of our predictions. In each iteration we retrained the SIDTs and predicted the lowest energy stable configurations at each coverage level according to the above procedure. We then took the predicted lowest energy configurations and attempted to calculate their energy using DFT and added them to the dataset for training the next iteration of the SIDT. We ran two refinement iterations for each case presented here.

We show the average association energy, $\frac{E_{N_{coad}}}{N_{coad}}$, for H^*, O^*, and HO^* in Figures 7a, 7b,

https://doi.org/10.26434/chemrxiv-2024-36w9w **ORCID:** https://orcid.org/0000-0002-4624-2852 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0



(a)

(b)

(c)

Figure 7: Average association energies for (a) O*, (b) H*, and (c) HO* as a function of coverage of (a) O*, (b) H*, and (c) HO* based on an iterative refinement process along with DFT calculations for the predicted lowest energy structures. Each SIDT refinement iteration learns from the original dataset and all DFT calculations run for prior iterations. Iter 1 corresponds to the original model. Coverages with no SIDT predictions indicate that SIDT did not find any stable configurations. Coverages without DFT points indicate that the lowest energy predicted configurations proposed by SIDT at that coverage were all found to be unstable.

https://doi.org/10.26434/chemrxiv-2024-36w9w **ORCID:** https://orcid.org/0000-0002-4624-2852 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

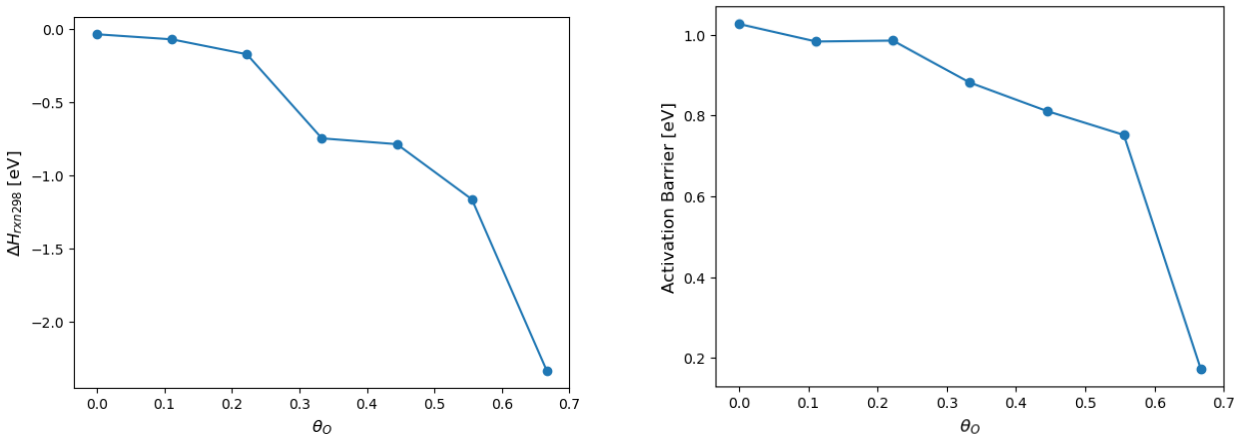
and 7c respectively computed using the SIDT model at each iteration and from DFT at the lowest energy configurations from the specified SIDT iterations. It should be emphasized that the DFT calculations in these plots are the DFT calculations of the configurations the associated SIDT predicts are the lowest in energy, not the true lowest energy DFT configurations, which are, in fact, unknown for these systems.

To understand these plots it is useful to first look closely at the the predictions just above 0.4 coverage for Figure 7a. At this coverage there are three DFT calculations from the first iteration, one of which is close to the SIDT energy predictions and the DFT calculation from the second iteration and two of which are about 0.2 eV higher in energy. We know the true DFT lowest energy configuration for this point is at or below the energy of the DFT configuration from the second iteration (since it is the lowest energy DFT calculation at this coverage) which agrees well with all of the SIDT predictions including the first iteration. This suggests that SIDT is significantly better at predicting the energy of the lowest energy configuration than it is at predicting the energy of a single given configuration.

Considering the entirety of Figure 7a we can see that the SIDT models all agree well with the DFT at lowest energy from the second iteration within 0.05 eV. Interestingly the second iteration model agrees exceptionally well with its lowest energy configuration DFT calculations. In Figure 7b we benefit significantly from the fact that H^* has weaker lateral interactions, greatly increasing our success rate at optimizing higher coverage configurations proposed by SIDT. While the second iteration SIDT seems to predict a steeper coverage dependence allowing it to match the lowest energy DFT point from the first iteration at about 0.9 coverage, the third iteration SIDT believes there is a lower energy configuration at that coverage more inline with the earlier DFT points and the first iteration SIDT. From the two second iteration DFT calculations between 0.4 and 0.6 coverage below the SIDT lines we are able to tell that the SIDTs are slightly overpredicting in this range. However, apart from the 0.9 coverage point, all lowest energy DFT calculations at each coverage agree within about 0.02 eV with the first and third SIDT iterations. The SIDT predictions for

https://doi.org/10.26434/chemrxiv-2024-36w9w **ORCID:** https://orcid.org/0000-0002-4624-2852 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

the HO^* case shown in Figure 7c look very similar to those for the O^*, likely because of the similar size and chemical composition. While the predicted overall behaviour seems physically plausible, especially compared with that of O^*, the DFT calculations in Figure 7c are appreciably above the SIDT predictions. It is unlikely these are the lowest energy points as there should be analogous configurations to the lowest energy DFT calculations from Figure 7a that offer comparable energies. However, it does seem likely that SIDT is having significant difficulty predicting which configuration is the lowest energy in this case. This could be a result of conformational effects associated with the orientation of the HO^* adsorbates that are not possible for O^*.



(a) (b)

Figure 8: SIDT predicted O* coverage dependence of the (a) enthalpy of reaction at 298 K and (b) activation barrier for the HO*+H* $\rightarrow$ H_2O +2* reaction on Cu111.

In Figure 8 we apply our model to estimate coverage-dependent properties of the HO*+ H* $\rightarrow$ H_2O +2* reaction on Cu111 using Equation 6 and the isolated properties of the reaction from Johnson et al.[19] In Figure 8a we examine the enthalpy of reaction at 298 K. This reaction removes two adsorbates from the surface and the associated lateral interactions so naturally we expect that a higher coverages where lateral interactions are more significant the reactants will be higher in energy and thus the enthalpy of reaction will decrease, which is in agreement with Figure 8a. The activation barrier shown in Figure 8b also decreases with coverage as a result of the stronger lateral interactions in the reactants at higher coverages,

https://doi.org/10.26434/chemrxiv-2024-36w9w ORCID: https://orcid.org/0000-0002-4624-2852 Content not peer-reviewed by ChemRxiv. License: CC BY-NC-ND 4.0

however, it does so at a significantly slower rate because unlike the gas phase product, the transition state for the reaction does have its own lateral interactions that cause the transition state to increase in energy at higher coverages, albeit at a slower rate than that of the reactants that are closer to the surface.

Discussion

The Cu111 Models

For general coverage-dependent property estimation on Cu111 we believe the presented Cu111 models are sufficient for arbitrary adsorbates and transition states composed of H, C, O and N atoms and co-adsorbed with H*, O* or HO*. While we did sample with N^* as a co-adsorbate, we had too few valid samples and did not find the model able to make good predictions with N^* as the co-adsorbate.

Calculation of Co-Adsorbed Configurations

We believe our presented approach to calculate and analyze co-adsorbed configurations is a highly effective and efficient way to examine the co-adsorbed configurational space based on combining direct DFT calculations with a low-data ML approach. One caveat of our current approach is that it relies on the assumption that configurations that fail to optimize (to a minimum or to a saddle) are unstable and do not exist. Failure to optimize a configuration to a target well does not imply in general that the target well does not exist. If an initial guess is too far from the target well a configuration may optimize to a different well. However, configurations that fail to optimize, but do in fact exist are likely to be shallow, high energy wells. Since these configurations are high in energy and we are in general interested in and/or sensitive to the lowest energy configurations, the distinction between these edge case configurations being stable and unstable is unlikely to be very important.

https://doi.org/10.26434/chemrxiv-2024-36w9w **ORCID:** https://orcid.org/0000-0002-4624-2852 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

Advantages of Using Subgraph Isomorphic Decision Trees

As discussed in the introduction and demonstrated here the SIDT approach to coverage dependence presented here is significantly more flexible, automatic, and powerful than current state of the art cluster expansion techniques. SIDT is able to predict on arbitrary co-adsorbed configurations not just those adsorbates and co-adsorbates a CE scheme is fit for, and is able to learn interactions CE has to be explicitly told how to resolve.

Simultaneously, SIDT is easier to apply, more flexible, and easier to interpret than possible DNN based approaches. Crucially, SIDT can be applied to much smaller datasets than is feasible for DNNs and because of its interpretability it is much easier analyze results to improve performance.

Conclusions

Coverage dependence of chemical reactions is a key, but often ignored aspect of microkinetic model construction because of the computational expense and complexity that it requires to determine the necessary parameters.[14] Comparable challenges such as rate coefficient pressure dependence[33,34] in gas phase have readily available tools[35–37] that are able to fully automate high accuracy computations using ab inito methods. The framework and tools presented here open the door for decreasing the barrier to include coverage dependence routinely in future microkinetic models. Here we used SIDT approach to construct a predictor based on a fixed pre-generated dataset, however, it is easy to imagine using the workflow within an active learning scheme that automatically identifies what configurations should be calculated to improve the SIDT predictors. Moreover, our entire workflow is built to be automatic using Pynta, running the necessary calculations and post-processing them.

However, higher-level generalizations, such as the ones built into software such as RMG[29,38,39] allow for efficiently approximating kinetic parameters without any ab initio calculations for instance for pressure-dependent reactions in the gas-phase.[40,41] It is possible to imagine

https://doi.org/10.26434/chemrxiv-2024-36w9w ORCID: https://orcid.org/0000-0002-4624-2852 Content not peer-reviewed by ChemRxiv. License: CC BY-NC-ND 4.0

a similar, generalized approach, for instance for Cu111 using the model presented in this work. However, in general it may be impractical to evaluate the SIDTs at every possible 2D configuration as done in this work. Doing so is unlikely to be strictly necessary, but the scheme by which configurations are sampled must be considered carefully. Approximations across arbitrary or even a range of metals, however, is much more challenging. Constructing such a scheme might be best done by training a foundational SIDT model on one surface across a wide range of adsorbates, transition states, and co-adsorbates and then (applying delta learning) training correction SIDTs on much smaller datasets to predict the difference between the foundational SIDT model and particular surfaces.

We have presented a framework for generating machine learning models and applying them to predict coverage dependent kinetic parameters for microkinetic models. Our toolkit enables automatic ab initio computation of co-adsorbed configurations and automatic post-processing including identification of the optimized configuration and TS validity evaluation for TSs. We demonstrate the training of SIDT on the generated dataset to predict the stability and association energy of co-adsorbed configurations. Lastly, we explain how to use the SIDTs to compute mean-field coverage dependent energy corrections for adsorbates thermochemistry and reaction activation barriers.

On Cu111, a challenging surface, we are able to achieve association energy MAEs of 0.106 eV on adsorbates and 0.172 eV on transition states and due to natural error cancellation in SIDTs on relative properties MAEs of 0.130 eV on reaction energies and 0.180 eV on activation barriers. We hope to extend these techniques to enable high accuracy and efficient calculation of coverage dependent kinetic parameters.

Acknowledgement

This work was done within the Exascale Catalytic Chemistry (ECC) Project, which is supported by the U.S. Department of Energy, Office of Science, Basic Energy Sciences, Chemical

https://doi.org/10.26434/chemrxiv-2024-36w9w **ORCID:** https://orcid.org/0000-0002-4624-2852 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

Sciences, Geosciences and Biosciences Division, as part of the Computational Chemistry Sciences Program.

This research used resources of the National Energy Research Scientific Computing Center (NERSC), a Department of Energy Office of Science User Facility using NERSC award BES-ERCAP0026789.

This article has been authored by employees of National Technology & Engineering Solutions of Sandia, LLC under Contract No. DE-NA0003525 with the U.S. Department of Energy (DOE). The employees co-own right, title and interest in and to the article and are responsible for its contents. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this article or allow others to do so, for United States Government purposes. The DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan https://www.energy.gov/downloads/doe-public-access-plan.

References

(1) Kitchin, J. R. Correlations in coverage-dependent atomic adsorption energies on Pd(111). *Phys. Rev. B* **2009**, *79*, 205412.

(2) Getman, R. B.; Schneider, W. F.; Smeltz, A. D.; Delgass, W. N.; Ribeiro, F. H. Oxygen-Coverage Effects on Molecular Dissociations at a Pt Metal Surface. *Phys. Rev. Lett.* **2009**, *102*, 076101.

(3) Grabow, L. C.; Hvolbæk, B.; Nørskov, J. K. Understanding Trends in Catalytic Activity: The Effect of Adsorbate–Adsorbate Interactions for CO Oxidation Over Transition Metals. *Topics in Catalysis* **2010**, *53*, 298–310.

https://doi.org/10.26434/chemrxiv-2024-36w9w **ORCID:** https://orcid.org/0000-0002-4624-2852 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

(4) İnoğlu, N.; Kitchin, J. R. Simple model explaining and predicting coverage-dependent atomic adsorption energies on transition metal surfaces. *Phys. Rev. B* **2010**, *82*, 045414.

(5) Lausche, A. C.; Medford, A. J.; Khan, T. S.; Xu, Y.; Bligaard, T.; Abild-Pedersen, F.; Nørskov, J. K.; Studt, F. On the effect of coverage-dependent adsorbate–adsorbate interactions for CO methanation on transition metal surfaces. *Journal of Catalysis* **2013**, *307*, 275–282.

(6) Lu, J.; Behtash, S.; Faheem, M.; Heyden, A. Microkinetic modeling of the decarboxylation and decarbonylation of propanoic acid over Pd(111) model surfaces based on parameters obtained from first principles. *Journal of Catalysis* **2013**, *305*, 56–66.

(7) Majumdar, P.; Greeley, J. Generalized scaling relationships on transition metals: Influence of adsorbate-coadsorbate interactions. *Phys. Rev. Mater.* **2018**, *2*, 045801.

(8) Wu, C.; Schmidt, D.; Wolverton, C.; Schneider, W. Accurate coverage-dependence incorporated into first-principles kinetic models: Catalytic NO oxidation on Pt (111). *Journal of Catalysis* **2012**, *286*, 88–94.

(9) Xu, Z.; Kitchin, J. R. Probing the Coverage Dependence of Site and Adsorbate Configurational Correlations on (111) Surfaces of Late Transition Metals. *The Journal of Physical Chemistry C* **2014**, *118*, 25597–25602.

(10) Bronsted, J. N.; Sandved, K. H.; Lamer, V. K. Acid and Basic Catalysis. *Chemical Reviews* **1928**, *5*, 231–338.

(11) Evans, M. G.; Polanyi, M. Inertia and driving force of chemical reactions. *Transactions of the Faraday Society* **1938**, *34*, 11–24.

(12) Miller, S. D.; Kitchin, J. R. Relating the coverage dependence of oxygen adsorption on Au and Pt fcc(1 1 1) surfaces through adsorbate-induced surface electronic structure effects. *Surface Science* **2009**, *603*, 794–801.

https://doi.org/10.26434/chemrxiv-2024-36w9w **ORCID:** https://orcid.org/0000-0002-4624-2852 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

(13) Lerch, D.; Wieckhorst, O.; Hammer, L.; Heinz, K.; Müller, S. Adsorbate cluster expansion for an arbitrary number of inequivalent sites. *Physical Review B - Condensed Matter and Materials Physics* **2008**, *78*, 121405.

(14) Nolen, M. A.; Farberow, C. A.; Kwon, S. Incorporating Coverage-Dependent Reaction Barriers into First-Principles-Based Microkinetic Models: Approaches and Challenges. *ACS Catalysis* **2024**, 14206–14218.

(15) Shan, B.; Kapur, N.; Hyun, J.; Wang, L.; Nicholas, J. B.; Cho, K. CO-coverage-dependent oxygen dissociation on Pt(111) surface. *Journal of Physical Chemistry C* **2009**, *113*, 710–715.

(16) Frey, K.; Schmidt, D. J.; Wolverton, C.; Schneider, W. F. Implications of coverage-dependent O adsorption for catalytic NO oxidation on the late transition metals. *Catalysis Science Technology* **2014**, *4*, 4356–4365.

(17) Johnson, M. S.; Green, W. H. A machine learning based approach to reaction rate estimation. *Reaction Chemistry Engineering* **2024**, *9*, 1364–1380.

(18) Pang, H. W.; Dong, X.; Johnson, M. S.; Green, W. H. Subgraph Isomorphic Decision Tree to Predict Radical Thermochemistry with Bounded Uncertainty Estimation. *Journal of Physical Chemistry A* **2024**,

(19) Johnson, M. S.; Gierada, M.; Hermes, E. D.; Bross, D. H.; Sargsyan, K.; Najm, H. N.; Zádor, J. Pynta-An Automated Workflow for Calculation of Surface and Gas-Surface Kinetics. *Journal of Chemical Information and Modeling* **2023**, *63*, 5168.

(20) Wellendorff, J.; Lundgaard, K. T.; Møgelhøj, A.; Petzold, V.; Landis, D. D.; Nørskov, J. K.; Bligaard, T.; Jacobsen, K. W. Density functionals for surface science: Exchange-correlation model development with Bayesian error estimation. *Physical Review B - Condensed Matter and Materials Physics* **2012**, *85*, 235149.

https://doi.org/10.26434/chemrxiv-2024-36w9w **ORCID:** https://orcid.org/0000-0002-4624-2852 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

(21) Giannozzi, P.; Baroni, S.; Bonini, N.; Calandra, M.; Car, R.; Cavazzoni, C.; Ceresoli, D.; Chiarotti, G. L.; Cococcioni, M.; Dabo, I. et al. QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. *Journal of Physics: Condensed Matter* **2009**, *21*, 395502.

(22) Giannozzi, P.; Andreussi, O.; Brumme, T.; Bunau, O.; Nardelli, M. B.; Calandra, M.; Car, R.; Cavazzoni, C.; Ceresoli, D.; Cococcioni, M. et al. Advanced capabilities for materials modelling with Quantum ESPRESSO. *Journal of Physics: Condensed Matter* **2017**, *29*, 465901.

(23) Larsen, A. H.; Mortensen, J. J.; Blomqvist, J.; Castelli, I. E.; Christensen, R.; Dułak, M.; Friis, J.; Groves, M. N.; Hammer, B.; Hargus, C. et al. The atomic simulation environment—a Python library for working with atoms. *Journal of Physics: Condensed Matter* **2017**, *29*, 273002.

(24) Hermes, E. D.; Sargsyan, K.; Najm, H. N.; Zádor, J. Sella, an Open-Source Automation-Friendly Molecular Saddle Point Optimizer. *Journal of Chemical Theory and Computation* **2022**,

(25) Hermes, E. D.; Sargsyan, K.; Najm, H. N.; Zádor, J. Accelerated Saddle Point Refinement through Full Exploitation of Partial Hessian Diagonalization. *Journal of Chemical Theory and Computation* **2019**, *15*, 6536–6549.

(26) Hermes, E. D.; Sargsyan, K.; Najm, H. N.; Zádor, J. Geometry optimization speedup through a geodesic approach to internal coordinates. *The Journal of Chemical Physics* **2021**, *155*, 094105.

(27) Johnson, M. S.; Pang, H.-W. zadorlab/PySIDT. `https://github.com/zadorlab/PySIDT`.

(28) Han, S.; Lysgaard, S.; Vegge, T.; Hansen, H. A. Rapid and accurate mapping of reac-

https://doi.org/10.26434/chemrxiv-2024-36w9w **ORCID:** https://orcid.org/0000-0002-4624-2852 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

tion condition-dependent alloy phase diagrams via Bayesian evolutionary multitasking. **2022**,

(29) Liu, M.; Dana, A. G.; Johnson, M. S.; Goldman, M. J.; Jocher, A.; Payne, A. M.; Grambow, C. A.; Han, K.; Yee, N. W.; Mazeau, E. J. et al. Reaction Mechanism Generator v3.0: Advances in Automatic Mechanism Generation. *Journal of Chemical Information and Modeling* **2021**, *61*, 2686–2696.

(30) Johnson, M. S.; Pang, H. W.; Payne, A. M.; Green, W. H. ReactionMechanismSimulator.jl: A modern approach to chemical kinetic mechanism simulation and analysis. *International Journal of Chemical Kinetics* **2024**,

(31) Goodwin, D. G.; Speth, R. L.; Moffat, H. K.; Weber, B. W. Cantera: An object-oriented software toolkit for chemical kinetics, thermodynamics, and transport processes. 2021; `https://www.cantera.org`.

(32) Ansys Chemkin-Pro — Chemical Kinetics Simulation Software. `https://www.ansys.com/products/fluids/ansys-chemkin-pro`.

(33) Klippenstein, S. J. From theoretical reaction dynamics to chemical modeling of combustion. *Proceedings of the Combustion Institute* **2017**, *36*, 77–111.

(34) Johnson, M. S.; Green, W. H. Examining the accuracy of methods for obtaining pressure dependent rate coefficients. *Faraday Discussions* **2022**, *238*, 380–404.

(35) de Vijver, R. V.; Zádor, J. KinBot: Automated stationary point search on potential energy surfaces. *Computer Physics Communications* **2020**, *248*, 106947.

(36) Zádor, J.; Martí, C.; Van de Vijver, R.; Johansen, S. L.; Yang, Y.; Michelsen, H. A.; Najm, H. N. Automated reaction kinetics of gas-phase organic species over multiwell potential energy surfaces. *J. Phys. Chem. A* **2023**, *127*, 565–588.

https://doi.org/10.26434/chemrxiv-2024-36w9w **ORCID:** https://orcid.org/0000-0002-4624-2852 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

(37) Elliott, S. N.; Moore, K. B.; Copan, A. V.; Keçeli, M.; Cavallotti, C.; Georgievskii, Y.; Schaefer, H. F.; Klippenstein, S. J. Automated theoretical chemical kinetics: Predicting the kinetics for the initial stages of pyrolysis. *Proceedings of the Combustion Institute* **2021**, *38*, 375–384.

(38) Johnson, M. S.; Dong, X.; Dana, A. G.; Chung, Y.; Farina, J. D.; Gillis, R. J.; Liu, M.; Yee, N. W.; Blondal, K.; Mazeau, E. et al. RMG Database for Chemical Property Prediction. *Journal of Chemical Information and Modeling* **2022**, *62*, 4906–4915.

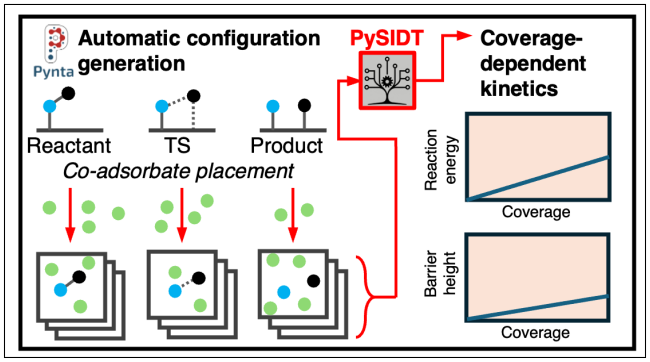
(39) Johnson, M. S.; Pang, H.-W.; Liu, M.; Green, W. H. Species Selection for Automatic Chemical Kinetic Mechanism Generation. **2023**,

(40) Matheu, D. M.; Lada, T. A.; Green, W. H.; Dean, A. M.; Grenda, J. M. Rate-based screening of pressure-dependent reaction networks. *Computer Physics Communications* **2001**, *138*, 237–249.

(41) Johnson, M. S.; Dana, A. G.; Green, W. H. A workflow for automatic generation and efficient refinement of individual pressure-dependent networks. *Combustion and Flame* **2023**, *257*, 112516.

https://doi.org/10.26434/chemrxiv-2024-36w9w **ORCID:** https://orcid.org/0000-0002-4624-2852 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

TOC Graphic



https://doi.org/10.26434/chemrxiv-2024-36w9w **ORCID:** https://orcid.org/0000-0002-4624-2852 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0