Pynta - An automated workflow for calculation of surface and gas-surface kinetics

Matthew S. Johnson,[†] Maciej Gierada [1] ,[†] Eric D. Hermes [2] ,[†] David H. Bross,[‡] Khachik Sargsyan,[†] Habib N. Najm,[†] and Judit Zádor[*,†]

†*Combustion Research Facility, Sandia National Laboratories, Livermore, California 94550, United States*

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

E-mail: jzador@sandia.gov

Phone: +1 (925) 294-3603

Abstract

Many important industrial processes rely on heterogeneous catalytic systems. However, given all possible catalysts and conditions of interest it is impractical to optimize most systems experimentally. Automatically generated microkinetic models can be used to efficiently consider many catalysts and conditions. However, these microkinetic models require accurate estimation of many thermochemical and kinetic parameters. Manually calculating these parameters is tedious and error prone involving many interconnected computations. We present Pynta, a workflow software for automating the calculation of surface and gas-surface reactions. Pynta takes the reactants, products

---

[1]Current affiliation: Maciej Gierada Technologies, Krukowskiego 1, 26-600 Radom, Poland
[2]Current affiliation: Quantum-Si, Guilford 06437, Connecticut, United States

and atom maps for the reactions of interest, generates sets of initial guesses for all species and saddle points, runs all optimizations, frequency, and IRC calculations and computes the associated thermochemistry and rate coefficients. It is able to consider all unique adsorption configurations for both adsorbates and saddle points allowing it to handle high index surfaces and bidentate species. Pynta implements a new saddle point guess generation method called harmonically forced saddle point searching (HFSP). HFSP defines harmonic potentials based on the optimized adsorbate geometries and which bonds are breaking and forming that allow initial placements to be optimized using the GFN1-xTB semiempirical method to create reliable saddle point guesses. This method is reaction class agnostic and fast, allowing Pynta to consider all possible adsorbate site placements efficiently. We demonstrate Pynta on 11 diverse reactions involving monodenate, bidentate and gas-phase species, many distinct reaction classes and both a low and a high index facet of Cu. Our results suggest that it is very important to consider reactions between adsorbates adsorbed in all unique configurations for inter-adsorbate group transfers and reactions on high index surfaces.

Introduction

Heterogeneous catalysis plays a key role in energy technology and in the chemical industry in general.[1] Catalysts are widely used to upgrade heavy fossil fuels,[2] enable the partial reduction of bio-derived feedstocks,[3,4] and convert small molecules (CO, CO_2, CH_4, etc.) into larger and more valuable compounds.[5–7] Besides synthesis, heterogeneous catalysis is also used in pyrolysis and exhaust emissions management[8,9] and heterogeneous processes are also crucial to battery technology.[10,11] Typical catalysts used in these processes are metals, their alloys, or metal oxides, often in a nanoparticle form. Catalytic reactions of a gas feedstock over these surfaces are fundamentally complex: they involve many reactions among gas-phase and adsorbed species and the yields of desired products depend on temperature, pressure, composition and the nature of the exposed catalyst surface. This makes

it incredibly impractical to optimize heterogeneous systems using experiments alone.

A large body of theoretical work describing catalytic processes on surfaces can be found in the literature[12,13] where the goal is to produce microkinetic models that can be used to interpret experimental findings as well as for optimizing conditions and catalysts to improve catalytic efficiency. However, manual exploration of the reaction mechanisms and characterization of elementary reaction steps are difficult and slow.[14] Automation, taking advantage of modern computing resources, offers a viable path towards accelerating catalysis research at the molecular scale.

Automatic microkinetic model generators such as the Reaction Mechanism Generator (RMG) tool[14–18] provide a way to rapidly test catalysts and conditions. Such tools use data science approaches to estimate unknown thermochemical and kinetic parameters based on existing information within their databases. However, the reliability of such estimates depends upon having large quantities of representative training data for generating estimators.[17,19] Furthermore, even when relatively accurate estimators are available to use in mechanism generation, the generation of reliable chemical mechanisms almost always requires the explicit calculation of a significant number of important reaction parameters.

Manually calculating microkinetic parameters involves many interdependent calculations and can be tedious and error-prone. Identifying the lowest energy structure of all reactants and products and then generating saddle point (SP) guesses that can reliably optimize to the lowest energy saddle points connecting reactants and products is a complex and delicate task. The most widely used approach in the literature is using chemical intuition and the energies of the separated reactants, to select a configuration of reactants and products. From that configuration, the researcher either manually guesses a transition state structure or uses a relatively expensive and more complex SP guess generation method, such as nudged elastic band (NEB).[20–23] Once the initial SP guess is generated, it must be optimized, hopefully to the desired SP. The investigator can determine if the optimized SP connects the correct reactants and products by analyzing its imaginary vibrational mode[24] or through the intrinsic

reaction coordinate (IRC) method.[25] A systematic approach to all of this, especially when the surface has multiple possible minima, demands meticulous care from the kineticist. In addition to all of the manual effort, they must correctly guess all of the unique placements of the adsorbates, of the SP, and either guess a good SP structure themselves or run what can often be a computationally expensive calculation to generate a good SP guess. Our goal is to develop a new computational workflow tool that is comprehensive, cheaper, and eliminates most of the manual labor from this process.

Similar challenges posed in the exploration of gas-phase systems have led to the development of many workflow tools that automate quantum chemistry calculations. Software for computing thermochemistry and rate coefficients for gas-phase kinetic systems include ARC,[26,27] Automech,[28,29] and KinBot.[30–32] The Automat[33,34] package can automate the calculation of the energies of adsorbed species. However, we are not aware of any widely distributed software able to automate kinetics calculations for surface reactions.

In this work we present Pynta (`https://github.com/zadorlab/pynta`), a workflow code that automatically calculates thermochemistry and rate coefficients of surface and gas-surface reactions from a simple 2D molecular graph representation of the reactants and products. Pynta implements a novel TS guess technique we call harmonically forced saddle point (HFSP) search to efficiently generate many saddle point guesses that can be filtered down to a tractable set of geometries to optimize. With this technique in hand, Pynta is able to generate initial guesses for adsorbates, to optimize adsorbate geometries, calculate frequencies for adsorbates, use the adsorbate geometries to generate saddle point guesses, optimize the best guesses and calculate frequencies and run intrinsic reaction coordinate (IRC) calculations for successful saddle point optimizations. The results can then be post-processed to give thermochemistry, barriers, and rate coefficients for each reaction. Pynta is designed to be able to take advantage of current petascale and upcoming exascale computing resources.
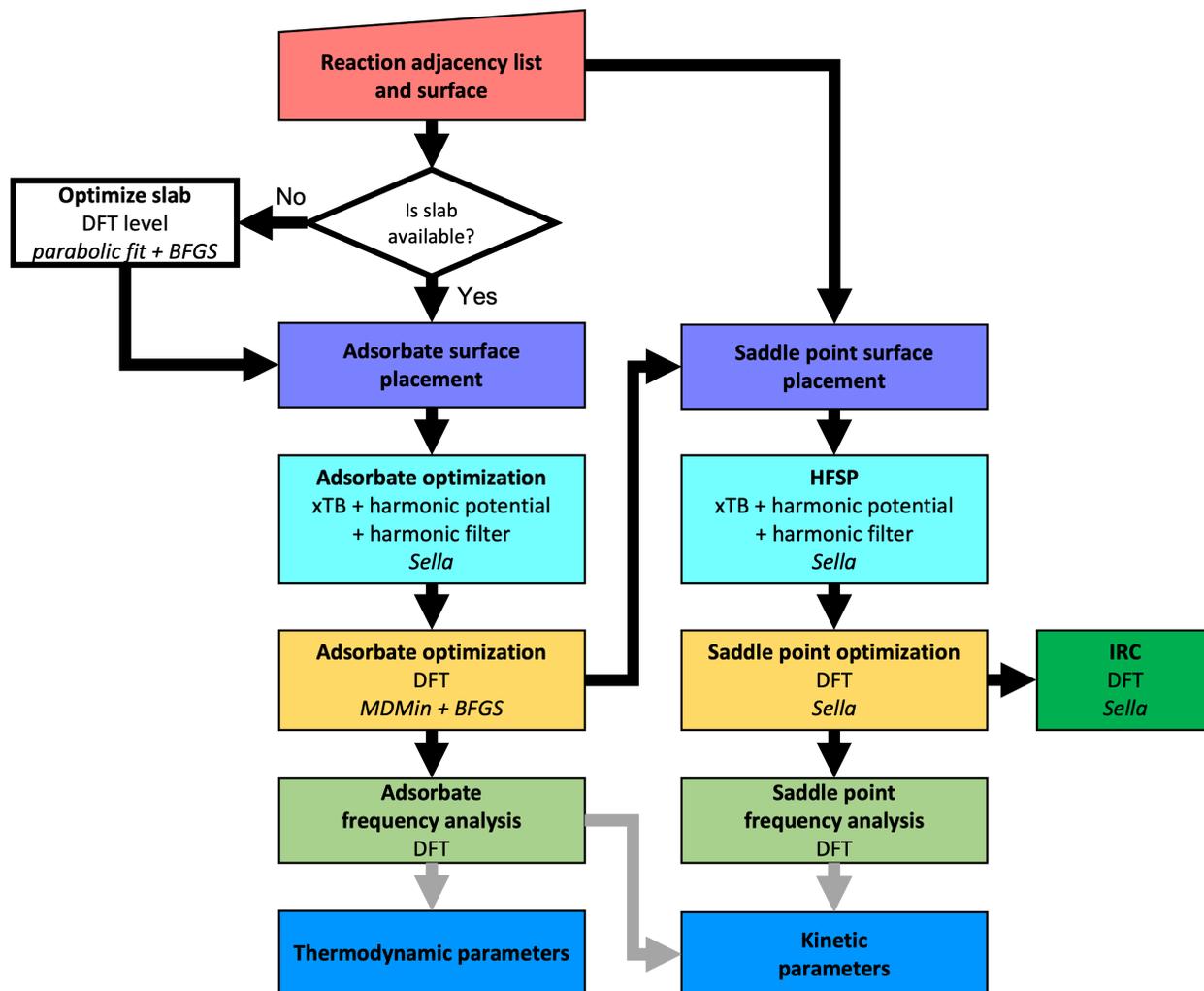
Workflow

Figure 1: Diagram of Pynta's workflow. Symmetrical geometries are filtered out after every optimization operation. Gray arrows represent postprocessing steps. For steps involving quantum chemistry calculations the level of theory and any optimizers used are noted.

A diagram of Pynta's full workflow is presented in Figure 1. We provide details on individual steps in later sections. Overall, provided a list of reactions and a surface, Pynta starts by extracting the unique adsorbates from the list. If the slab geometry is not already available, it is generated and optimized. Initial adsorbate geometries are generated, placed on the surface, and optimized using the GFN1-xTB semiempirical method[35] augmented with harmonic potentials to generate a set of adsorbate guesses. The unique filtered guesses are

optimized at the user-specified level of theory (usually DFT) first using ASE's MDMin[36] method (an adaptive gradient descent optimizer) until $f_{max} \leq 0.5$, and then ASE's BFGSLineSearch[37] implementation for the rest of the optimization. This two staged approach uses a weak gradient based optimizer followed with an aggressive Newton optimizer for robustness. Starting with the weak optimization helps ensure the adsorbate relaxes to the site it is placed on provided it can bind to that site.

Once the adsorbates are optimized using ASE, the workflow starts two processes. In one, we calculate frequencies by evaluating the Hessian of the energy with respect to the adsorbate atom positions for all unique successfully optimized structures. In the other, Pynta uses these adsorbate structures to generate placements of adsorbates for saddle points, run the HFSP search and then optimize the chosen saddle point guesses using Sella.[38–40] Frequency and IRC calculations[38–40] are then run for the unique successfully optimized saddle point structures. With optimized reactant, product, and saddle-point geometries and frequencies, thermodynamic and kinetic parameters can be calculated in our postprocessor provided with the code. The correctness of the saddle points can be verified based on the imaginary mode and/or IRC calculations also using the postprocessing tool. After each step redundant structures are identified and removed using ASE's SymmetryEquivalenceCheck tool.

The slab calculations and the adsorbate xTB optimizations are calculated serially at the start. The FireWorks workflow management software[41] is then used to handle all of the time-consuming DFT optimizations, frequency calculations, IRC calculations, and the SP placement, and HFSP operations.

The workflow takes advantage of already available computational tools. The Atomic Simulation Environment (ASE)[36] provides a way to run many kinds of quantum chemistry calculations using a wide range of quantum chemistry codes through a unified Python interface. Sella[38–40] provides a powerful ASE-compatible algorithm for saddle point optimization, constrained minimization and IRC calculations. We use the Alloy Catalysis Automated Toolkit (ACAT)[42] to identify and place adsorption sites on catalytic surfaces (an analogous tool to

CatKit[43]). Furthermore, for initial guess generation for both adsorbates and saddle points, Pynta makes significant use of the semiempirical GFN1-xTB method.[35,44,45] Lastly, we use FireWorks[41] to manage the large number of calculations on high-performance computing platforms.

Reaction Encoding

In order to initiate calculations for a surface reaction, Pynta needs to know the chemical nature of the reactants and the products and a mapping from reactants to products of all atoms participating in bonds that are formed or broken during the reaction. This mapping, called the atom mapping, indicates which atoms in the reactant correspond to which atoms in the product. This information is encoded as an adjacency list of the reactants and products as defined in the RMG package,[15–17] illustrated in Figure 2.

This reaction information is included in the `reactions.yml` file, and an example is given in Figure 2. All reactant and product connections are described within respective adjacency lists, and each atom that breaks or forms bonds in the reaction is assigned a unique label such as `*1` that identifies it in both the reactants and the product structures. Note that the validity of an adjacency list can be easily tested on the RMG website molecule tool `https://rmg.mit.edu/molecule_search`. Pynta's saddle point guess generation algorithm, HFSP, is reaction class agnostic, so within the workflow the `reaction` and `reaction_family` parameters are used for labeling only.

Slab Optimization

In order to calculate reaction properties on surfaces, Pynta needs a slab, characterized by the lattice constant, the facet type, and physical dimensions (i.e., number of layers and extent of each layer). Pynta can read in slab files prepared by the user corresponding to the desired level of theory provided the slab atoms are ordered from bottom layer to top layer. If not

```
- index: 0
  reaction: OH* + X <=> O* + H*
  reaction_family: Surface_Dissociation
  reactant: |
   1 *1 O u0 p2 c0 {2,S} {4,S}
   2 *2 H u0 p0 c0 {1,S}
   3 *3 X u0 p0 c0
   4    X u0 p0 c0 {1,S}
  product: |
   1 *1 O u0 p2 c0 {4,D}
   2 *2 H u0 p0 c0 {3,S}
   3 *3 X u0 p0 c0 {2,S}
   4    X u0 p0 c0 {1,D}
```

Figure 2: An example of a `reactions.yml` file containing adjacency lists representing the $OH^* + * \rightleftharpoons O^* + H^*$ reaction. The adjacency list format starts with an optional multiplicity line at the top that defaults to 1 if not present. All subsequent lines first include (1) the ordered index (starting from `1`) of an atom, (2) an optional label for that atom starting with `*`, (3) the element name for the atom (adsorption sites are represented as atoms with label `X`), then (4) `u` followed by the number of unpaired electrons on the atom, (5) `p` followed by the number of electron pairs on the atom, and (6) `c` followed by the charge of the atom. Lastly, (7) the bonds of that atom are specified in the format: `{index of the bonded atom, order of bond}`. The order is specified with a letter where `vdW` corresponds to hydrogen or other pseudo bonds, `S` corresponds to single bonds, `B` corresponds to benzene bonds, `D` corresponds to double bonds, `T` corresponds to triple bonds, and `Q` corresponds to quadruple bonds. The labels (*1,*2,*3) identify all atoms that form or break bonds in the reaction in both the reactants and products. Pynta assumes that the atom labeled as *1 in the reactants is the atom labeled by *1 in the products and so on.

provided, however, Pynta can optimize the lattice constant and generate and optimize the slab as part of the workflow.

Pynta first finds the corresponding experimental lattice constant value within ASE[36] and calculates the energy of the primitive cell between 0.1 Å above and below the experimental value at 0.01 Å increments. The seven lowest energies are fit to a parabola, and the lattice constant corresponding to the minimum of the parabola is found. Then the parabolic value is optimized further using SciPy's bounded method between 0.01 Å above and below the parabolic value. Once the lattice constant is known, Pynta will use ASE[36] to generate a slab corresponding to the lattice constant with the requested physical dimensions and facet specifications. This slab is then optimized while freezing a specified number of layers on the bottom of the slab to give the final slab geometry for use in the adsorbate and reaction calculations.

Harmonic Potentials

For initial guess generation for both adsorbates and saddle points Pynta makes significant use of the GFN1-xTB[35,44,45] method combined with specific harmonic potentials. This combination allows Pynta to efficiently optimize structures while softly enforcing certain constraints.

In Pynta, harmonic potentials can be defined between an atom and a point, or between two atoms:

$$E_{\mathrm{harm}} = \sum_b k_b (d_b - d_{\mathrm{eq},b})^2 \tag{1}$$

$$\mathbf{F}_{\mathrm{harm},j} = 2 \sum_{b \text{ if } j \in b} k_b \mathbf{n}_{b,j} (d_b - d_{\mathrm{eq},b}) \tag{2}$$

where b corresponds to a single harmonic potential/bond, j corresponds to an atom, $j \in b$ indicates that atom j is a part of harmonic potential b, E_{harm} is the total harmonic energy, k_b is the force constant of the bth potential, $d_{\mathrm{eq},b}$ is the equilibrium (i.e., desired) distance of the bth potential, $\mathbf{F}_{\mathrm{harm},j}$ is the harmonic force vector associated with the jth atom, and $\mathbf{n}_{b,j}$ is the component of the unit vector associated with atom j in the direction from atom j

along harmonic bond b to the other atom or point. These harmonic potentials allow Pynta to keep atoms approximately at a specified distance from each other or from a point during geometry optimization.

Adsorbate Initial Guess Generation

The use of harmonic potentials to generate initial adsorbate guess geometries for monodentate and bidentate species is illustrated in Figure 3. We first use the ACAT package[42] to analyze the surface and identify all unique sites and all unique pairs of sites. For each adsorbate, we use RDKit[46] to generate an initial desorbed gas-phase structure. Then, for each unique site or pair of sites, we define harmonic potentials with $k_b = 100$ eV/Å^2, which we will refer to as k_{hard}, and $d_{eq} = 0$ Å between the adsorbate atoms that are directly bonded to the surface and the estimated position for each of these atoms. The latter are defined as the nominal ACAT site location translated along the z coordinate by Δz above the surface. The distance, Δz, is determined based on the site type (bridge, hollow, etc.) and element identity of the binding atom. These potentials keep the adsorbates from desorbing from the surface when we optimize with the GFN1-xTB method. For monodentate adsorbates, we also add a harmonic potential between all atoms not bonded to the surface and a point 8.5 Å above the binding site with $k_{soft} = 0.1$ eV/Å^2 and $d_{eq} = 0$ Å. This potential prevents initial monodentate adsorbates from optimizing into bidentate structures in the initial guess-generation step. We optimize the geometries using the GFN1-xTB potential combined with the harmonic potentials while freezing the length of all bonds between atoms (as defined by the adjacency list) to generate an initial guess geometry. For bidentate adsorbates, we also perform a second optimization with the same combined potential, but the bond lengths are allowed to relax. The geometry minimizations for the guess generation stage are done using Sella with a maximum force (f_{max}) convergence criterion of 0.02 eV/Å .

However, in the bidentate case, this procedure can generate hundreds of adsorbate guesses, far more than it is usually practical or necessary to compute. This occurs because of
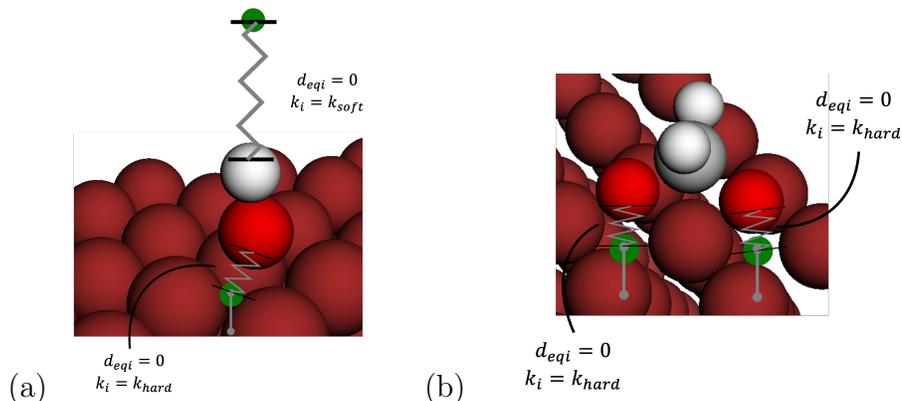
Figure 3: Schematic representation of initial guess optimizations for (a) monodentate and (b) bidentate adsorbates. The metal atoms are dark red, oxygen atoms are bright red, carbon atoms are dark grey and hydrogens are white. Green circles represent fixed points in space towards which atoms are pulled by harmonic potentials shown as springs.

the large number of site combinations, many of which represent unphysical scenarios: sites can be too close straining the ring the bidentate adsorbate forms with the surface, or sites can be too far apart, straining and eventually breaking bonds in the bidentate adsorbate. Note that this challenge also occurs for many saddle point guesses as will be discussed later, because many saddle points involve multiple sites. Unfortunately, it is difficult to reliably resolve this problem by restricting the pairs of sites considered using geometrical or similar arguments, because many bidentate adsorbates (and saddle points) involve non-adjacent sites in non-intuitive ways. In both cases we solve this problem using what we refer to as a harmonic filter. When a system has at least two (hard) harmonic potentials, the harmonic energy E_{harm} is a measure of how hard the springs are pulling against each other and how strained the system is. We, therefore, use E_{harm} to filter out overly strained systems at the guess generation stage using the following criteria. First, we demand that Pynta include all guesses such that

$$E_{\mathrm{harm}} < \tau_1 E_{\mathrm{harm,min}} \tag{3}$$

where $E_{\mathrm{harm,min}}$ is the energy of the guess with the lowest harmonic energy and τ_1 is a set

tolerance with a default value of 3.0. Second, we exclude all guesses such that

$$E_{\text{harm}} > \tau_2 \cdot E_{\text{harm,min}} \tag{4}$$

where τ_2 is a set tolerance with a default value of 30.0 – these are highly strained structures. Finally, we add the remaining lowest harmonic energy guesses until we have N_{TSmin} (defaults to 5) guesses or until we run out of guesses. A plot showing how this filter works for the $\text{OH}^* + * \rightleftarrows \text{O}^* + \text{H}^*$ reaction is shown in Figure 4. These tolerances τ_1 and τ_2 can be adjusted as desired to control how comprehensively the user wishes to search the space of bidentate adsorbates and saddle points.
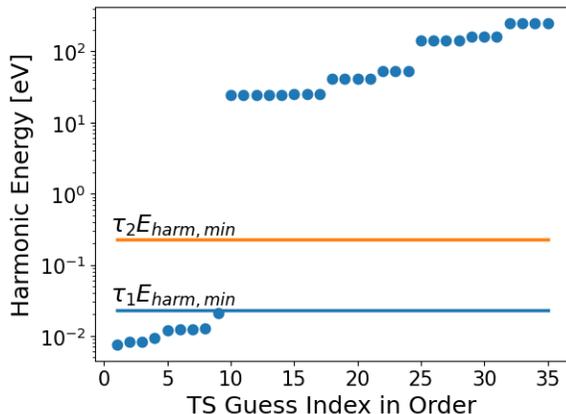
Figure 4: Illustration of the harmonic filter for the $\text{OH}^* + * \rightleftarrows \text{O}^* + \text{H}^*$ reaction. Below the blue (lower) line all guesses are selected and above the orange (upper) line no guesses are selected for DFT optimization. In this case, there are more than $N_{\text{TSmin}} = 5$ guesses already under the orange line, so no guesses are drawn from between the two limits. However, in cases where there are fewer than $N_{\text{TSmin}} = 5$ guesses under the blue line, guesses between the blue and orange lines are included up to $N_{\text{TSmin}} = 5$.

Harmonically Forced Saddle Point (HFSP) Guess Generation

Most reactions of interest on surfaces involve more than one site. As was the case with bidentate adsorbates, even on a 3×3 surface there are often hundreds of unique ways to position the reactants and products (including participating vacant sites) in the unit cell,

and it is far from trivial to determine *a priori* what placements can or cannot lead to saddle points for arbitrary reactions, especially considering the variety of catalytic surfaces.

Many popular methods for generating saddle point guesses typically require either significant computational expense such as in the NEB,[20–23] GSM[47–49] and AFIR[50–53] methods, or significant amounts of heuristics or data such as in KinBot,[30–32] AutoTST[54,55] or TS-GCN.[56] Naturally, we need to avoid the cost of running hundreds of NEB, GSM or AFIR calculations per reaction. At the same time, there are only a limited number of saddle point structures from databases available to build guess estimators like KinBot, AutoTST, and TS-GCN. For instance, the Catalysis Hub database[57] is able to store saddle point geometries for reactions, but only a small fraction of reactions have barriers heights and only a small fraction of those reactions include saddle point geometries. The Open Catalyst dataset[58] does not include any transition state structures.

In short, for Pynta to be an efficient kinetics code for heterogeneous catalysis, it requires a saddle point guess method that is fast and uses minimal data. Additionally, it would be desirable for this method to be reaction class agnostic and able to handle a wide range of surface reactions. To this end, we developed our Harmonically Forced Saddle Point (HFSP) search algorithm, a method with all of the aforementioned properties, which borrows and fuses ideas from KinBot and AFIR.

$$d_{eqi} = f_b d_{well}$$
$$k_i = k_{hard}$$
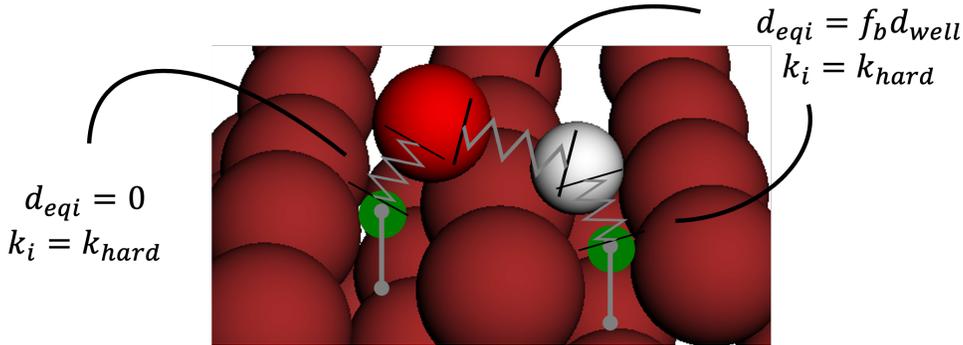
$$d_{eqi} = 0$$
$$k_i = k_{hard}$$

Figure 5: Harmonically forced saddle point search. The notation is the same as in Figure 3.

HFSP defines harmonic potentials for a saddle point based on the breaking and forming bonds as prescribed in the reaction's adjacency list (see Figure 2). These harmonic potentials

pull target structures into saddle point guess structures as illustrated in Figure 5. For a given set of site placements, as was done for adsorbates, we define harmonic potentials ($k_{\text{hard}} = 100 \text{ eV/Å}^2$, $d_{\text{eq}} = 0 \text{ Å}$) that keep surface-bound atoms bonded to the surface close to their associated site for each surface bond that does not break in the reaction. For bonds that break or form in the reaction we define harmonic potentials either between the adsorbate atoms or for surface bonds between an atom and a point above the site. Each of these harmonic potentials also has $k_{\text{hard}} = 100 \text{ eV/Å}^2$. Since we have the optimized structures for both the reactants and products, we are able to find the length of each broken/formed bond, $d_{\text{well}.i}$, in either the reactants or the products. The target point for site bonds is set so that $\Delta z_i = d_{\text{well},i}$ above the site location. Lastly, we define $d_{\text{eq},i}$ with

$$d_{\text{eq},i} = f_{\text{b}} d_{\text{well},i} \tag{5}$$

where f_{b} is a bond stretch factor, which is applied to all bonds that are broken or formed, relative to the bond length in the structure where this bond exists. We define f_{b} based on a small set of heuristics

$$f_{\text{b}} = \begin{cases} 0.2 \text{ if X} - \text{R} \\ 1.25 \text{ if R} - \text{R intramolecular} \\ 1.4 \text{ if R!H} - \text{R!H intermolecular} \\ 1.6 \text{ if R} - \text{H intermolecular} \end{cases} \tag{6}$$

where X denotes a surface site, R denotes any adsorbate atom, R!H denotes an adsorbate atom that is not a hydrogen, and H denotes a hydrogen. Intramolecular bonds are bonds that connect two atoms on the same adsorbate when formed. Bonds that are not intramolecular are considered intermolecular. Importantly, these heuristics are all reaction class independent. However, estimates of the bond stretch factor f_{b} likely can be improved in the future by learning based on data, for instance, using a machine-learning-based predictor based on

the subgraph isomorphic decision tree (SIDT) technique.[19] However, we will demonstrate in this work that these simple heuristics are sufficient by themselves to handle most reactions.

Similarly to the treatment of the adsorbates, GFN1-xTB augmented with the harmonic potentials is used for the optimization of these structures. The optimization is carried out by Sella with $f_{max} = 0.02$ eV/Å. Note that here we are using Sella as an energy minimizer rather than a saddle point optimizer. We then apply the harmonic filter (see Figure 4) to obtain a tractable set of saddle point guess geometries.

Statistical Mechanics

Pynta currently computes thermochemistry and kinetics based on the rigid rotor harmonic oscillator model for all gas-phase species and based on the 2D lattice gas model for adsorbed structures. Thermochemistry is calculated using ASE's[36] thermochemistry module. We assume that the located saddle point is a transition state.

We currently calculate rate coefficients based on the work in Campbell et al.[59] An alternative formulation is available in Hermes et al.[60] We start from the equation

$$\frac{R}{A} = \frac{k_B T}{h} \cdot \frac{N_{TS}}{A} \cdot \frac{1}{q_{TS,rc}} \tag{7}$$

where $\frac{R}{A}$ is the reactive flux per unit area, k_B is the Boltzmann constant, T is temperature, $\frac{N_{TS}}{A}$ is the number of transition state complexes present per unit area, and $q_{TS,rc}$ is the partition function associated with the reaction coordinate in the transition state. Assuming that the transition state is an ideal 2D lattice gas we have

$$\frac{N_{TS}}{A} = \theta_{TS}\frac{M}{A} = K_{eq,TS}(1 - \theta_{tot})\frac{M}{A}\prod_i a_i \tag{8}$$

where $\frac{M}{A}$ is the number of sites per unit area, θ_{TS} is the coverage fraction of the transition state on the surface, $K_{eq,TS}$ is the equilibrium constant between the reactants and the tran-

sition state, and $\prod_i a_i$ is the product of the activities of the reactants. For ideal 2D lattice gas adsorbates

$$a_i = \frac{c_i(1-\theta_0)}{\theta_0(1-\theta_{\text{tot}})}\frac{A}{M} \tag{9}$$

where $\theta_0 = 0.5$,[59] c_i is the concentration of reactant i, and θ_{tot} is the total coverage fraction of all the reactants. For gas-phase species we have

$$a_i = (c_i/c_0) \tag{10}$$

where $c_0 = 2.4282 * 10^{22}\frac{\text{molecules}}{L}$. With the partition functions and energies we can obtain $K_{\text{eq,TS}}$

$$K_{\text{eq,TS}} = \frac{q_{\text{TS}}}{\prod_i q_i}e^{-\frac{(E_{\text{TS}}-\sum_i E_i)}{RT}} \tag{11}$$

where q_{TS} is the partition function of the transition state, q_i is the partition function of the ith reactant, E_{TS} is the energy of the transition state, E_i is the energy of the ith adsorbate and R is the gas constant. Substituting Equation 11 into Equation 8, and then Equation 8 into Equation 7 gives

$$\frac{R}{A} = \frac{k_B T}{h}\frac{q_{\text{TS}}}{q_{\text{TS,rc}}\prod_i q_i}e^{-\frac{(E_{\text{TS}}-\sum_i E_i)}{RT}}(1-\theta_{\text{tot}})\frac{M}{A}\prod_i a_i \tag{12}$$

a tractable equation for computing rate coefficients. Plugging in Equations 9 and 10 we have

$$\frac{R}{A} = \frac{k_B T}{h}\frac{q_{\text{TS}}}{q_{\text{TS,rc}}\prod_i q_i}e^{-\frac{(E_{\text{TS}}-\sum_i E_i)}{RT}}\left(\frac{M}{A}\right)^{1-N_{\text{ads}}}\frac{1}{c_0^{N_{\text{gas}}}}\frac{\prod_i c_i}{(1-\theta_{\text{tot}})^{1-N_{\text{ad}}}} \tag{13}$$

and

$$k = \frac{k_B T}{h}\frac{q_{\text{TS}}}{q_{\text{TS,rc}}\prod_i q_i}e^{-\frac{(E_{\text{TS}}-\sum_i E_i)}{RT}}\left(\frac{M}{A}\right)^{1-N_{\text{ads}}}\frac{1}{c_0^{N_{\text{gas}}}} \tag{14}$$

where N_{ads} is the number of adsorbates in the reactants, N_{gas} is the number of gas-phase species in the reactants and k is the computed rate coefficient.

Results

Table 1: Reactions on copper computed with Pynta in this work. N_{SP} denotes the number of saddle points found by Pynta, $E_{a,min}$ denotes the lowest energy barrier in eV, $E_{a,max}$ denotes the highest energy barrier in eV and A (m,molecule,s), n and E_a (kJ/mol) denote the Arrhenius parameters for the forward rate coefficient in appropriate units.

Reaction	Reaction Image	Facet	N_{SP}	$E_{a,min}$	$E_{a,max}$	A	n	E_a	Fig.
$H^* + O^* \longleftrightarrow HO^* + *$		fcc111	4	0.882	0.912	2.452e-7	0.335	88.36	6
$HO^* + * \longleftrightarrow H^* + O^*$		fcc111	4	1.661	1.692	3.631e-9	0.638	159.22	6
$OCH^* + * \longleftrightarrow OC^* + H^*$		fcc111	2	0.158	0.168	3.557e-6	-0.381	18.26	7
$OC^* + H^* \longleftrightarrow OCH^* + *$		fcc111	2	0.969	0.979	2.402e-5	-0.793	97.60	7
$H^* + * \longleftrightarrow H^* + *$		fcc111	3	0.103	0.111	8.660e-7	0.019	12.44	8
$OC^* + O^* \longleftrightarrow CO_2 + 2*$		fcc111	2	0.441	0.441	6.289e-8	0.027	42.99	9
$CO_2 + 2* \longleftrightarrow OC^* + O^*$		fcc111	2	1.639	1.640	3.014e-46	1.907	155.47	9
$HO^* + H^* \longleftrightarrow H_2O + 2*$		fcc111	4	1.026	1.043	1.412e-4	-0.634	104.21	10
$H_2O + 2* \longleftrightarrow HO^* + H^*$		fcc111	4	1.095	1.111	4.183e-46	1.899	99.69	10
$HOCH_2{}^* + * \longleftrightarrow CH_3O^* + *$		fcc111	2	2.140	2.150	1.124e-4	-0.322	210.76	11
$CH_3O^* + * \longleftrightarrow HOCH_2{}^* + *$		fcc111	2	3.159	3.168	3.584e-10	1.991	307.60	11
$OCHO^* + * \longleftrightarrow CO_2 + H^* + *$		fcc111	3	0.522	0.539	1.608e-9	0.678	50.40	12
$CO_2 + H^* + * \longleftrightarrow OCHO^* + *$		fcc111	3	0.840	0.858	5.379e-43	0.753	80.39	12
$H^* + OCH_2O^{**} \longleftrightarrow HOCH_2O^* + 2*$		fcc111	3	0.740	0.762	1.315e-07	0.265	73.73	13
$HOCH_2O^* + 2* \longleftrightarrow H^* + OCH_2O^{**}$		fcc111	3	1.258	1.281	3.654e-30	0.948	120.34	13
$N^* + CH_3O^* \longleftrightarrow CH_2O + NH^* + *$		fcc111	11	0.576	0.802	5.218e-14	2.674	53.39	14
$CH_2O + NH^* + * \longleftrightarrow N^* + CH_3O^*$		fcc111	11	1.058	1.284	1.280e-48	2.326	95.09	14
$OCH^* + * \longleftrightarrow OC^* + H^*$		fcc211	1	0.382	0.382	2.603e-8	0.589	38.13	15
$OC^* + H^* \longleftrightarrow OCH^* + *$		fcc211	1	1.147	1.147	6.887e-10	1.059	110.47	15
$H^* + O^* \longleftrightarrow HO^* + *$		fcc211	4	1.195	1.556	4.730e-7	0.280	118.92	16
$HO^* + * \longleftrightarrow H^* + O^*$		fcc211	4	1.836	2.196	2.404e-9	0.740	177.18	16

In order to demonstrate Pynta's effectiveness, we ran Pynta on a diverse set of reactions. The eleven chosen examples shown in Table 1 in both forward and reverse directions, involve many different reaction classes, monodentate, bidentate, and gas-phase chemical

species (with elements C, O, N, and H), and low and high index facets (fcc111 and fcc211 respectively) of copper. Note that there are 4 symmetrically unique sites on the fcc111 facet and 14 on the fcc211 facet. All DFT calculations were run on a $3 \times 3 \times 4$ slab with a (3x3x1) k-point grid using the BEEF-vdW[61] functional with PBE-KJPAW pseudopotentials[62,63] and a kinetic energy cutoff of 40 Ry as implemented in Quantum Espresso.[64] All geometry optimizations were run until the maximum force on each atom was smaller than 0.02 eV/Å and all IRCs were run until the force was smaller than 0.1 eV/Å. The harmonic filter was run with default tolerances. Further details are available in the supporting information. All saddle points were validated visually using normal mode analysis and the IRCs through the postprocessing tool. In the following sections we will discuss each reaction individually. The accompanying saddle point images are positioned for best visibility, and the exact geometries are available in the supporting information. The barriers in Table 1 are always given relative to the lowest energy structure(s) of the reactant(s), and include the zero-point energy (ZPE). When there are more than one reactants we calculate the reactant energies as the sum of the energies of the isolated adsorbates rather than the coadsorbed structure. This is also how the postprocessing module determines barrier heights and calculates rate coefficients.

$\mathbf{H^* + O^* \longleftrightarrow HO^* + *}$ on Cu111

This is one of the simplest reactions one can imagine on a surface, a simple bond fission of a diatomic adsorbate. The four saddle points for $H^* + O^* \longleftrightarrow HO^* + *$ on Cu111 are available in Figure 6. All of the identified unique and valid barriers are similar for this reaction. Pynta also finds an additional, fifth higher energy saddle point that puts the hydrogen over a top site. The IRC indicates this supposed saddle point leads to the correct reactants and products, however, frequency analysis shows that it is a higher order saddle with a relatively large (219 cm^{-1}) second imaginary frequency.
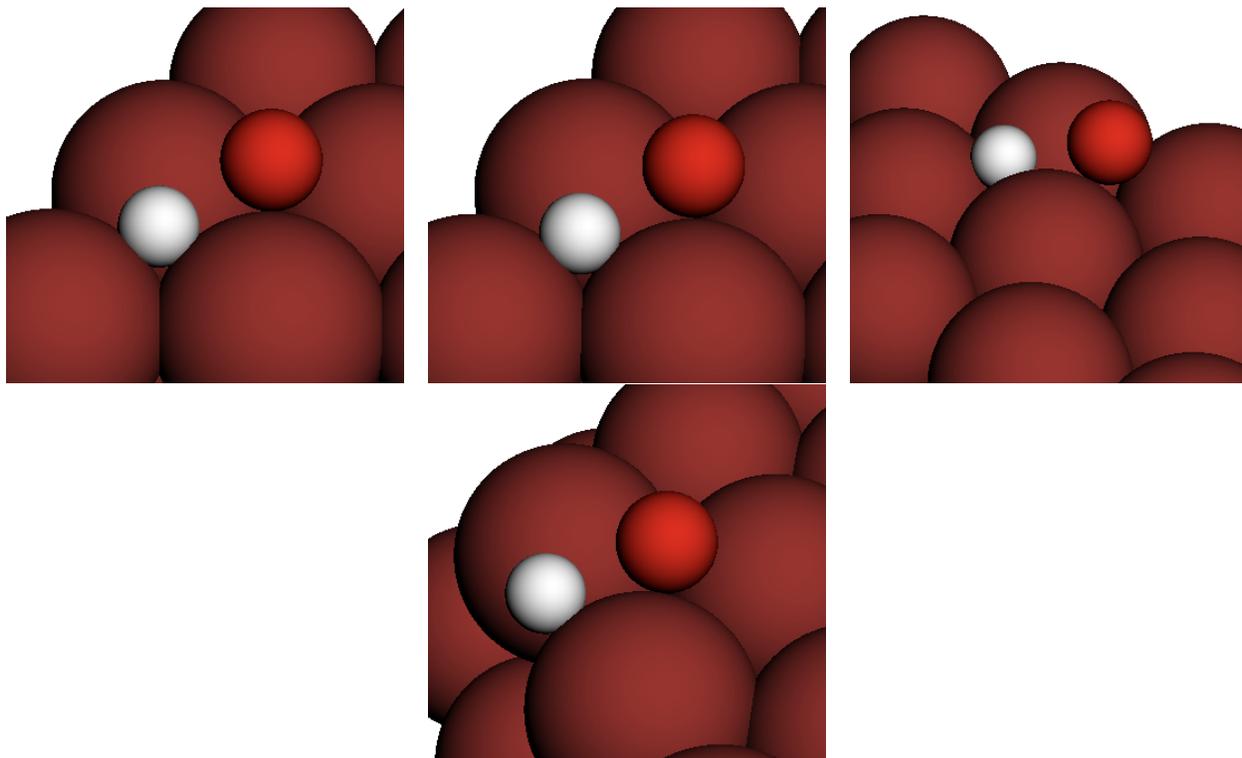
Figure 6: Saddle points for H* + O* ⟷ HO* + *. From left to right barriers are: Top: 0.909 eV, 0.912 eV and 0.884 eV and Bottom: 0.882 eV.

Figure 7: Saddle points for OCH* + * ⟷ OC* + H* on Cu111. From left to right barriers are: 0.969 eV and 0.979 eV.

OCH* + * ⟷ OC* + H* on Cu111

The two saddle points found by Pynta for the simple adsorbate dissociation reaction, OCH* + * ⟷ OC* + H*, are available in Figure 7. The two reactions have very similar barriers and primarily differ on what specific hollow site type the saddle point hovers over.

H* + * ⟷ H* + * Cu111
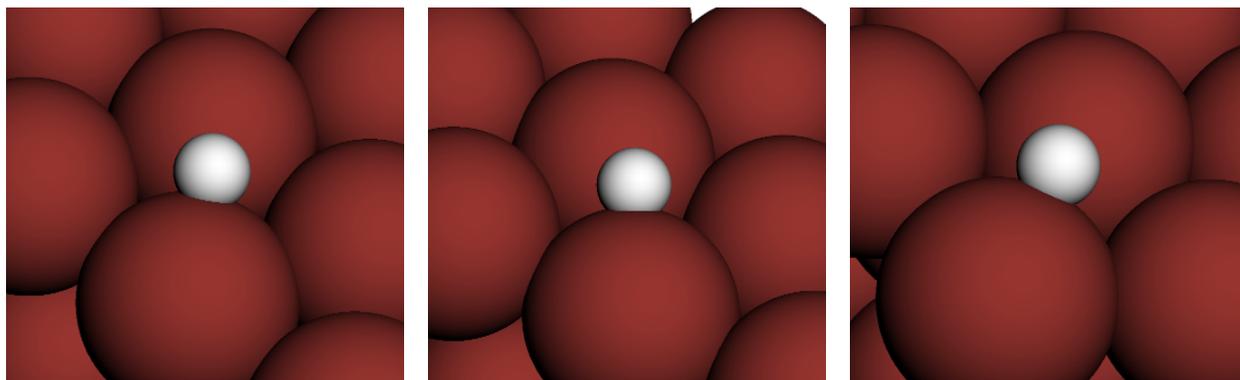
Figure 8: Saddle points for H* + * ⟷ H* + * on Cu111. From left to right barriers are: 0.104 eV, 0.103 eV and 0.111 eV.

The three saddle points for H* + * ⟷ H* + * diffusion on Cu111 are available in Figure 8. All of the barriers and geometries are similar and correspond to the same diffusion from an *fcc* hollow to an *hcp* hollow. Using a tighter or more sophisticated Sella convergence criteria would likely collapse these structures into a unique saddle. Since H* is only stable adsorbed to the hollow sites on Cu111, this is the only possible diffusion reaction for H* on the Cu111 surface. While we were able to calculate valid IRCs for this particular reaction, running IRCs for diffusions at $f_{\max} = 0.1$ does not usually generate useful IRCs. However, visualizing the imaginary mode is usually sufficient to validate diffusion saddle points.

OC* + O* ⟷ CO$_2$ + 2* on Cu111

The two saddle points for OC* + O* ⟷ CO$_2$ + 2* on Cu111 are available in Figure 9. These two saddle points look similar both in terms of barrier and geometry, however, they

Figure 9: Saddle points for $OC^* + O^* \longleftrightarrow CO_2 + 2*$ on Cu111. From left to right barriers are: 0.441 eV and 0.441 eV.

are different enough to be labeled as distinct: the breaking C-O bond length differs by about 0.04 Å between the two saddles.

$HO^* + H^* \longleftrightarrow H_2O + 2*$ on Cu111

The four saddle points for the dissociative adsorption reaction, $HO^* + H^* \longleftrightarrow H_2O + 2*$ on Cu111 are available in Figure 10. All of these saddle points are fairly similar in energy. Half of them put the HO^* on a *fcc* hollow and half on a *hcp* hollow. While three of them point the hydrogen directly up, one distinctly tilts the H to the side. In that particular reaction the HO^* adsorbate drifts away from the hollow it hovers over and the H^* relaxes to the hollow it is over. In all of the other cases the reverse happens along the reaction path.

$HOCH_2{}^* + * \longleftrightarrow CH_3O^* + *$ on Cu111

In this reaction, an internal hydrogen migration is accompanied by an overall migration on the surface and a change in the atom through which the adsorbate binds to the surface. The two saddle points for $HOCH_2{}^* + * \longleftrightarrow CH_3O^* + *$ on Cu111 are presented in Figure 11. The two pathways have very similar barriers and primarily differ on what specific hollow site types the saddle point hovers over.
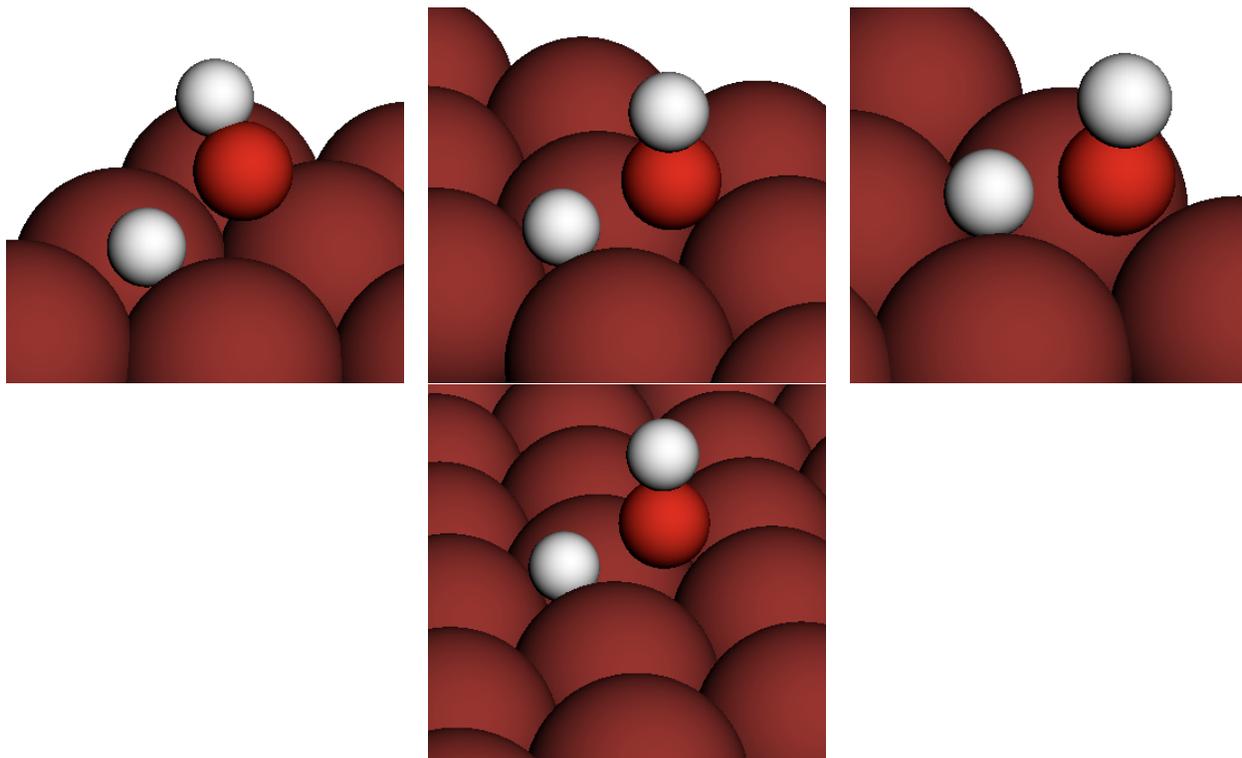
Figure 10: Saddle points for HO* + H* $\longleftrightarrow$ H$_2$O + 2* on Cu111. From left to right barriers are: Top: 1.043 eV, 1.026 eV, and 1.034 eV, Bottom: 1.033 eV.

Figure 11: Saddle points for HOCH$_2$* + * $\longleftrightarrow$ CH$_3$O* + * on Cu111. From left to right barriers are: 2.150 eV and 2.140 eV.
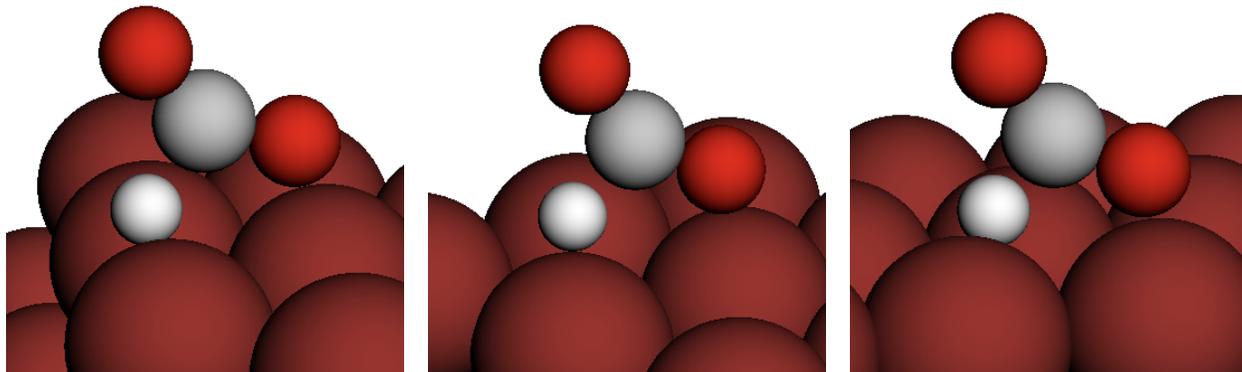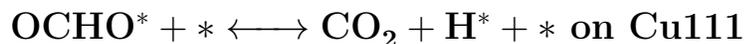
OCHO* + $*$ $\longleftrightarrow$ CO$_2$ + H* + $*$ on Cu111

Figure 12: Saddle points for OCHO* + $*$ $\longleftrightarrow$ CO$_2$ + H* + $*$ on Cu111. From left to right barriers are: 0.539 eV, 0.537 eV and 0.522 eV.

In this reaction the gas-phase CO$_2$ lands on an adsorbed H atom, binds to it, and the formed HCO$_2$ stays adsorbed to the surface through one of the oxygen atoms. The three saddle points for CO$_2$ + H* + $*$ $\longleftrightarrow$ OCHO* + $*$ on Cu111 are available in Figure 12. All of these saddle points have similar barriers, and geometries. While the structures are distinct geometrically, we believe they all represent the same saddle point. The optimization with Sella has difficulty optimizing these structures because in the saddle CO$_2$ has a very low frequency ($\sim$65 cm^{-1}) motion associated with CO$_2$ rotating relative to the the rest of the saddle. This very flat dimension of the potential energy surface likely contributes very little force to the structure making it difficult to optimize out when convergence is determined based on forces. Using a more sophisticated set of convergence criteria, a possibility through Sella's custom convergence mode, may collapse all of these saddles into one.

H* + OCH$_2$O** $\longleftrightarrow$ HOCH$_2$O* + 2 $*$ on Cu111

While this seems to be a relatively simple surface association/dissociation reaction, here, a bidentate adsorbate associates with a hydrogen atom in the forward direction producing two empty sites. Pynta found three unique and valid saddle points for H* + OCH$_2$O** $\longleftrightarrow$ HOCH$_2$O* + 2$*$ on Cu111, which are depicted in Figure 13. The associated barriers are all
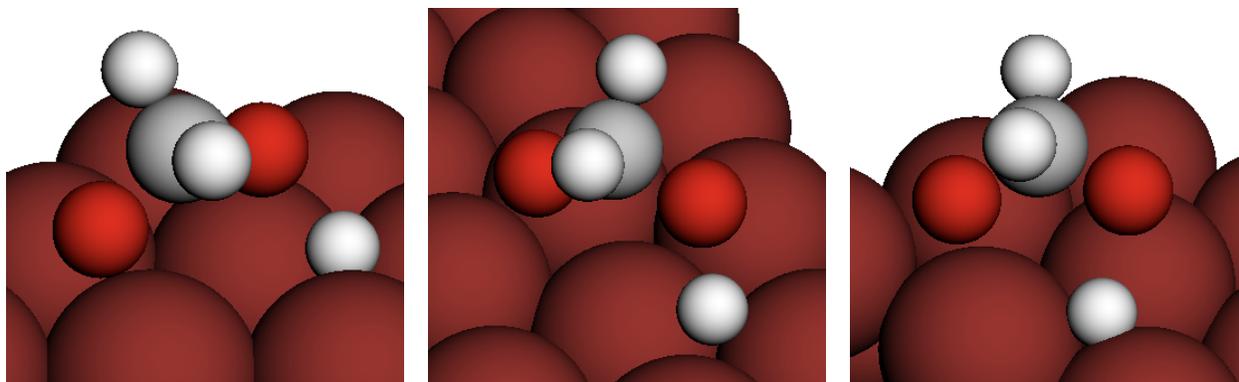
Figure 13: Saddle points for $H^* + OCH_2O^{**} \longleftrightarrow HOCH_2O^* + 2*$ on Cu111. From left to right barriers are: 0.762 eV, 0.761 eV and 0.740 eV.

relatively similar. Interestingly, the IRCs reveal that the hydrogen does not relax to the closest hollow site, usually drifting one site further out from the one it hovers over at the saddle point.

$N^* + CH_3O^* \longleftrightarrow CH_2O + NH^* + *$ on Cu111

This reaction involves a hydrogen transfer between two adsorbates, and the simultaneous desorption of the donor. The identified saddle points for $N^* + CH_3O^* \longleftrightarrow CH_2O + NH^* + *$ on Cu111 are presented in Figure 14. For this reaction, there were many similar harmonic energy saddle point guesses and the harmonic filter only limited the number of saddle point guesses to about 60. This is likely a result of the fact that this reaction is able to occur between farther apart sites. 25 of those guesses converged to a saddle point, out of which 11 were unique and valid. Unlike the previous example, the barrier heights differ by as much as 0.23 eV between saddle points, justifying the computational cost associated with the large number of trial calculations. We posit that a manual search is unlikely to yield the lowest energy structure in this case.
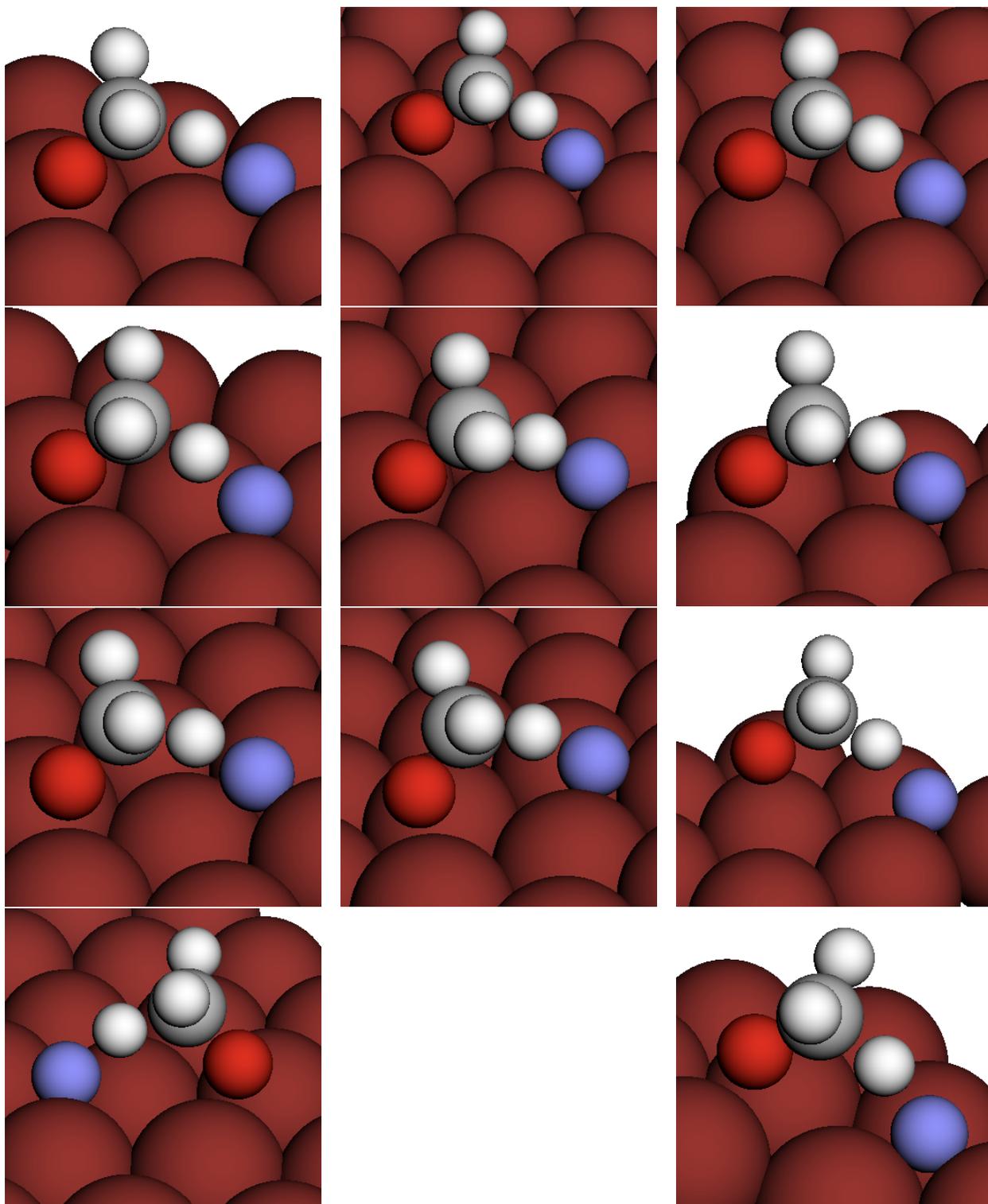
Figure 14: Saddle points for $N^* + CH_3O^* \longleftrightarrow CH_2O + NH^* + *$ on Cu111. From left to right barriers in eV are: first row: 0.728, 0.635 and 0.802, second row: 0.626, 0.735 and 0.582, third row: 0.768, 0.633 and 0.738, fourth row: 0.580 and 0.576.
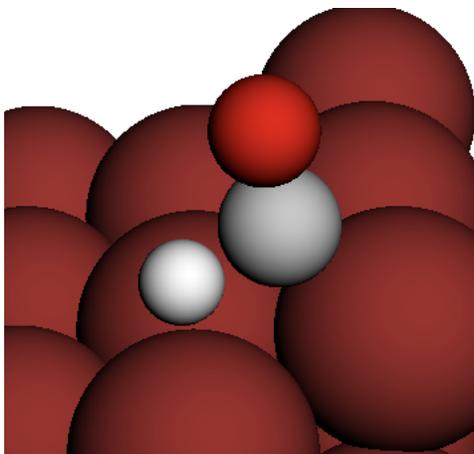
Figure 15: Saddle point for OCH* + $*$ $\longleftrightarrow$ OC* + H* on Cu211. The barrier is 0.382 eV.

OCH* + $*$ $\longleftrightarrow$ OC* + H* on Cu211

The saddle point for OCH* + $*$ $\longleftrightarrow$ OC* + H* on Cu211 is available in Figure 15. Following the IRC the hydrogen moves downhill on the surface, while the OC* drifts to a site uphill. For this reaction Pynta selected 6 saddle point guesses only one of which converged to a valid saddle point. The saddle point guess that converged was one of the two lowest energy saddle point guesses that had virtually the same harmonic energy. We suspect that improved estimates of the bond stretch factor taking into account the local surface structures may allow more of the chosen guesses to converge. However, one of the lowest harmonic energy guesses converged and the barrier Pynta calculates for this transition state is actually lower in energy than values in literature as we will note in detail later.

H* + O* $\longleftrightarrow$ HO* + $*$ on Cu211

The four saddle points for H* + O* $\longleftrightarrow$ HO* + $*$ on Cu211 are available in Figure 16. The barriers vary quite significantly by as much as 0.36 eV. Interestingly, the the lowest barrier saddle point has the hydrogen drifting along the contour of the "hill" of the facet. The highest barrier occurs when the hydrogen drifts up the wall at the bottom of the hill that occurs on the Cu211 facet.
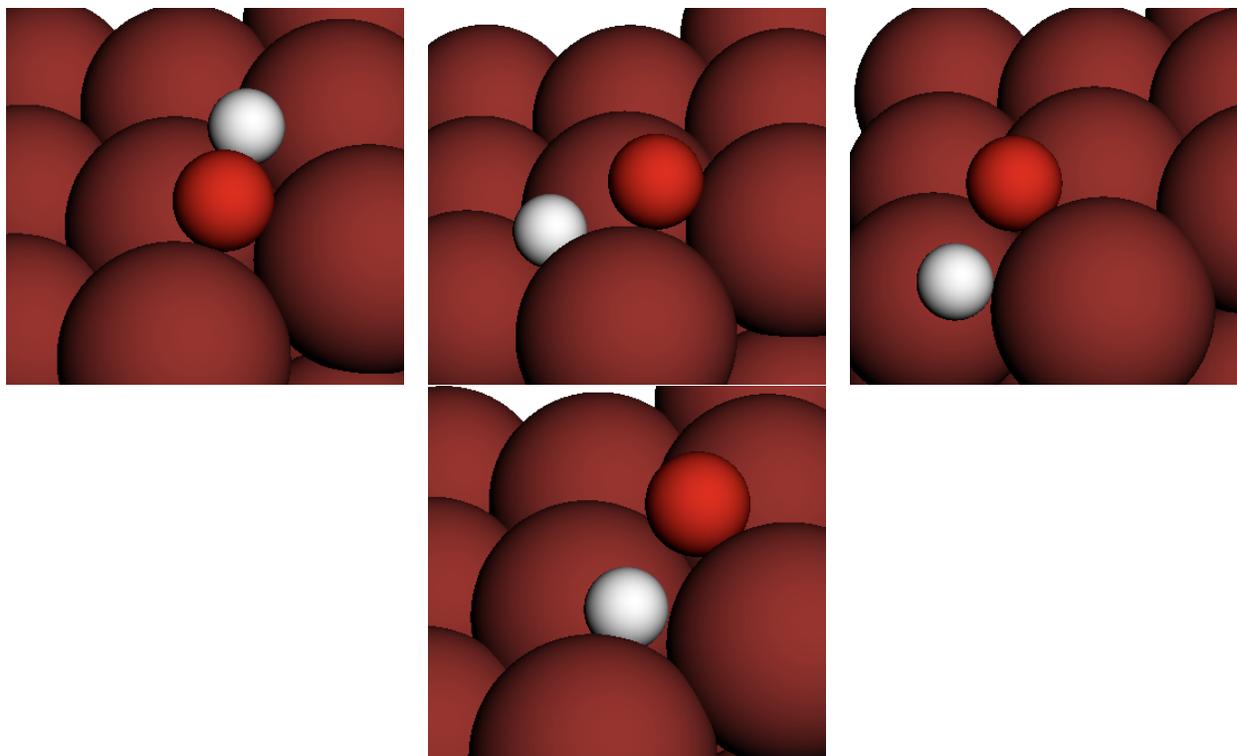
Figure 16: Saddle points for H* + O* ⟷ HO* + * on Cu211. From left to right barriers are: Top: 1.312 eV, 1.366 eV and 1.556 eV, Bottom: 1.195 eV.

Discussion

Comparison with Literature

Table 2: Comparison of Pynta calculated barriers with literature barriers. These reactions were identified by searching Catalysis Hub.[57] The monodentate OCHO* reaction calculated by Pynta was not within Catalysis Hub, but similar reactions from the Yoo et al. 2014[68] paper enabled us to find a calculated barrier within that paper. All energies are in eV.

Reaction	Reaction Image	Facet	Pynta $E_{a,\min}$	Lit. $E_{a,\min}$	Lit. DFT	Lit. Source
OCH* + * ⟷ OC* + H*		fcc111	0.158	0.29	BEEF-vdW[61]	Schumann et al. 2018[65]
HO* + * ⟷ H* + O*		fcc111	1.661	1.73	RPBE/DACAPO[66]	Wang et al. 2011[67]
OCHO* + * ⟷ CO$_2$ + H* + *		fcc111	0.522	0.67	RPBE/DACAPO	Yoo et al. 2014[68]
CO$_2$ + 2* ⟷ OC* + O*		fcc111	1.639	1.82	RPBE/DACAPO[66]	Falsig et al. 2008[69]
H$_2$O + 2* ⟷ HO* + H*		fcc111	1.095	1.33	BEEF-vdW[61]	Schumann et al. 2018[65]
HO* + * ⟷ H* + O*		fcc211	1.836	1.67	RPBE/DACAPO[66]	Wang et al. 2011[67]

In a literature search, aided by Catalysis Hub[57] we found reported barriers for six of the

reactions computed in this work. For each we selected the lowest energy barrier calculated with BEEF-vdW (the level of theory used in this work), or, if BEEF-vdW was not available, simply the lowest energy barrier. The compiled results are available in Table 2.

Comparing a limited number of quantum chemistry calculations, particularly at different levels of theory, requires a significant amount of care. One would first like to see that our calculations are not dramatically different from literature calculations. Second, one would like to see that the barriers are close to or lower than literature values, particularly when comparing with calculations at the same level of theory. It is worth recalling that we have confirmed each of our saddle points with IRC calculations, so we are confident that our saddles lead to the correct reactants and products.

All of the compared barriers agree within 0.24 eV, confirming the general validity of our approach. Of the six reactions, Pynta calculated appreciably lower barriers for five reactions, including both of the reactions calculated using the BEEF-vdW functional. The exception is the HO* dissociation reaction on Cu211, but because different density functionals were used, it is difficult to draw meaningful conclusions from the difference. Conversely, one should not read too much into other comparisons where Pynta found a lower barrier, but the used DFT functionals were different.

Interestingly, the largest difference between our calculations and the literature occurs for the $H_2O + 2* \longleftrightarrow HO^* + H^*$ reaction on fcc111. The Pynta barrier is 0.235 eV lower than the calculated value in Schumann et al.[65] Notably, the literature barrier was calculated using the same DFT method as in this work (BEEF-vdW). Catalysis Hub provides a saddle point geometry for this reaction. That saddle looks similar to the saddle point Pynta found with the OH* tilted (Figure 10 upper left), signaling that the underlying reaction path in the literature and in our work are the same. However, a closer inspection reveals a significant difference in the geometries: the breaking H*–O* bond distance is 0.114 shorter in the similar looking saddle point found by Pynta. It is possible that these differences occur because their work used an f_{max} of 0.05 eV/Å rather than 0.02 eV/Å used in this work.
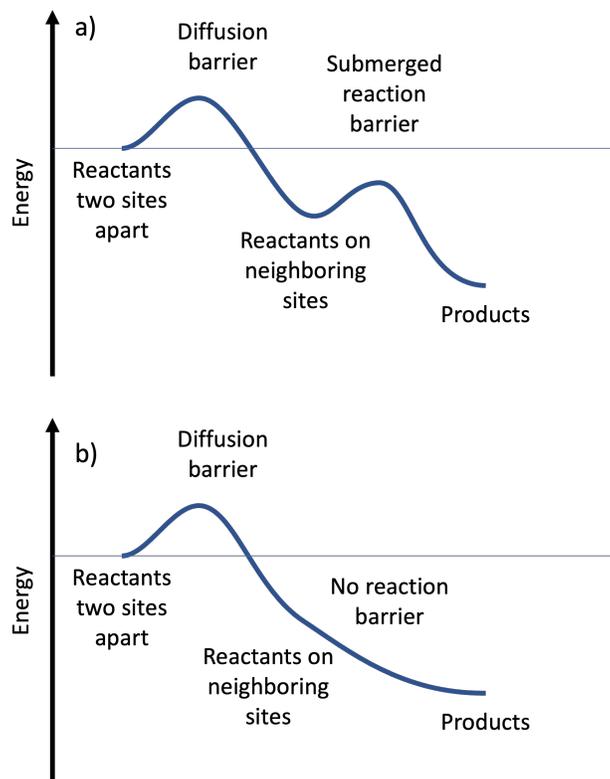
Figure 17: Schematic energy landscapes (a) of a reaction with a submerged barrier and (b) of a barrierless reaction on a catalyst surface.

Submerged Barrier Reactions

Formally, submerged barrier reactions can occur on surfaces when the lateral interactions of the reactants sufficiently lower their energy upon association before reacting. Since the reactants move to the saddle point from a lower energy well, it is possible to have a saddle point that has an energy below that of the fully separated reactants, resulting in a negative barrier, as shown schematically in Figure 17a. This is a well-understood phenomenon in the gas phase and has been studied in detail.[70,71] Between two adsorbed species in contact with the surface this process is likely to occur at thermal equilibrium and be well represented by the rate coefficient generated by Pynta. However, gas-phase reactants may not be thermalized with the surface. In this case it is debatable whether the process occurs closer to thermal equilibrium or more adiabatically, however, the Pynta rate coefficient should be a reasonable value especially at higher temperatures. In general, most negative barriers observed in Pynta

are associated with either invalid saddle points that do not connect to the correct reactants and products, or occasionally due to missed lower energy adsorbate configurations. A valid submerged barrier case can be identified by first confirming that the IRC connects to the correct reactants and products. Then, one should see that the reactants at the end of the IRC are lower in energy than the separated reactants. To confirm that the barrier is truly submerged due to strong lateral interactions, it is important to confirm that the individual reactant geometries at the end of the IRC were previously found in the associated adsorbate optimizations. If not, it is possible that the apparent negative barrier is present simply because the lowest energy conformer of the reactant was missed in the reactant calculation phase. Note that this problem is neither unique to submerged barriers nor to Pynta: if the lowest energy structure of the reactant is missed, the barriers appear lower. This can happen both in an automated and in a manual exploration. However, automated searches inherently reduce the chance of this happening.

Barrierless Reactions

Barrierless reactions are reactions where there is no saddle point separating the reactants from products. Such reactions cannot be characterized using conventional transition state theory, and require variational techniques to calculate rate coefficients.[72] Detecting that a reaction is barrierless is a nontrivial task. In the gas phase, barrierless reactions often involve the homolytic scission of a bond in closed shell species, which require multireference methods. Single reference methods often produce a spurious barrier along the reaction coordinate. When Pynta is run on a likely barrierless gas-surface reaction like $H + * \longleftrightarrow H^*$, we find that using our level of theory, no valid saddles are found. However, when a barrierless process occurs between adsorbates, such as in $O^* + HOCH^* \longleftrightarrow OCH^* + HO^*$, recognizing the process is barrierless requires significantly more care.

When Pynta was run on $O^* + HOCH^* \longleftrightarrow OCH^* + HO^*$ on Cu111, the default harmonic filter selected more than 70 saddle point guesses. 56 of those guesses converged to a first-order

saddle point, 28 of those were unique, and about 16 unique saddles were valid as confirmed by the IRCs. These 16 unique valid saddles had a range of energies differing by as much as 0.5 eV. In general, the results looked similar to those in $CH_2O + NH^* + * \longleftrightarrow N^* + CH_3O^*$. However, the IRC for one of the lowest energy saddles for $O^* + HOCH^* \longleftrightarrow OCH^* + HO^*$ clearly began with $HOCH^*$ diffusing to a new site and then relaxing into the reaction as illustrated schematically in Figure 17b. We calculated the diffusion barrier for $HOCH^*$, and although that species has some conformational challenges that complicate analysis, we found the diffusion saddle point energy to be approximately the same as the examined saddle.

The nuance is that unlike in gas-gas and gas-surface barrierless reactions, surface-surface barrierless reactions depend on the relative positions of the adsorbates on the surface, and the reaction is only barrierless for certain configurations. In other configurations there can be valid saddles for the same reaction. However, also unlike gas-gas and gas-surface barrierless reactions, surface-surface barrierless reactions require the reactants to overcome diffusional barriers in order to react as shown in Figure 17b. In this light we can argue the barrier for $O^* + HOCH^* \longleftrightarrow OCH^* + HO^*$ *is* the diffusional barrier for $HOCH^*$ (note that the diffusional barrier for O^* is significantly higher than that of $HOCH^*$).

Conformers

One challenging aspect of finding the lowest energy adsorbate geometry is that even the isolated molecule can have many different energy minima. While Pynta accounts for different minima associated with adsorbates bonding to different sites, it does not currently perform any conformer search that considers the internal rotors of the adsorbate. For simple adsorbates on simple surfaces there often are no additional unique conformers. However, for more complicated adsorbates on less symmetric surfaces there can be many additional unique conformers. In practice, primarily on Cu111, we have found this to be most problematic for adsorbates that have conformer-generating internal rotors when desorbed. There is no way to easily detect if one has the lowest energy conformer even in gas-phase systems. Failure to

find the lowest energy conformer may cause reactions to be falsely predicted to have negative energy barriers if the saddle point is lower in energy than the reactants. In that case one should follow the IRC from the saddle point to the reactants. If the reactant well does not experience intermolecular interactions, one should be able to find the lower energy reactant geometry by taking the adsorbate geometry from the end of the IRC and minimizing it. If the reactant well does involve intermolecular interactions, the reaction is likely to have a truly submerged barrier. This particular problem will be addressed in future work with an option to systematically search for the adsorbate conformers.

Importance of Considering All Possible Saddle Points

A major advantage Pynta has over common workflows for computing rate coefficients on surfaces is the ability to consider reactions between adsorbates adsorbed in all unique configurations on the surface. For most reactions calculated on Cu111 in this work, the differences in energies for saddle points belonging to a given reaction are between 0.01 and 0.03 eV. It is notable that these are differences between saddle points generated from the best guesses Pynta determined based on the harmonic filter, which means that the selection of evaluated saddles is strongly biased towards low energies. One would therefore expect to find additional higher barrier reactions if one were to attempt to optimize all of the guesses. While these differences are chemically meaningful (0.03 eV corresponds to about a factor of 3 error at 298 K) they are much smaller than typical errors inherent in DFT level calculations.

For $N^* + CH_3O^* \longleftrightarrow CH_2O + NH^* + *$, however, this was not the case. An energy diagram of this reaction is available in Figure 18. We see enormous energy differences in the TS, as large as 0.23 eV. We have also observed these large energy differences for other reactions involving a hydrogen transferred between two adsorbates. We also see differences of up to 0.36 eV for $H^* + O^* \longleftrightarrow HO^* + *$ on Cu211. We expect these higher energy differences to be the case in general for higher index facets where there are more unique and reasonable site configurations that can lead to more diverse saddle points. 0.24 eV at 298 K corresponds
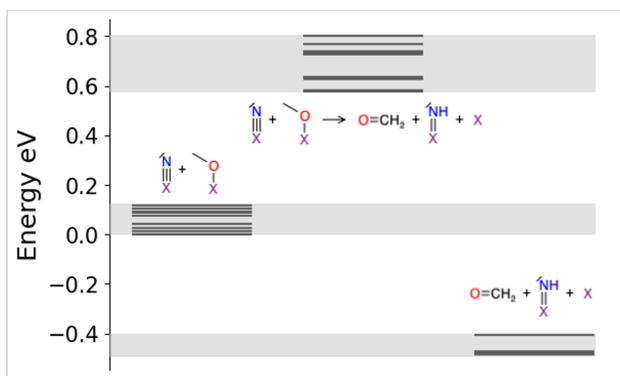
Figure 18: Energy diagram for $N^* + CH_3O^* \longleftrightarrow CH_2O + NH^* + *$. Each line represents a different valid saddle point or different combination of adsorbate geometries. There are 10 saddle point lines, 10 reactant lines and 4 product lines.

to an enormous factor error of about 10^4 in the rate coefficient, far more than is acceptable for parameters in a microkinetic model. This suggests that it is highly important to consider reactions between adsorbates adsorbed in all unique configurations for inter-adsorbate group transfers and reactions on high index surfaces.

Rate Coefficients for Kinetic Monte Carlo Simulations

In this work we have focused on computing mean-field/phenomenological rate coefficients for reactions on surfaces. These rate coefficients are suitable for traditional phenomenological simulators such as ReactionMechanismSimulator.jl,[73,74] TChem,[75] or Cantera.[76] However, Pynta's ability to search the entire space of possible saddle points can be used to generate rate coefficients for kinetic Monte Carlo (KMC) simulations that are able to explicitly resolve reaction processes in terms of the positioning of the reactants on the lattice. The comprehensiveness of Pynta's saddle point search can be easily increased by increasing the tolerances τ_1, τ_2, and N_{TSmin}. The IRC calculations already provide a good starting guess for the reactant configurations. In this view, calculating rate coefficients for KMC should be as simple as optimizing and calculating frequencies for the complexes at the end of the IRCs and calculating the unimolecular rate coefficients from those complexes to the saddle point.

Conclusions

We have presented Pynta, a workflow software for automatically computing the rate coefficients for surface and gas-surface reactions. Pynta uses ASE,[36] Sella,[38–40] and Fireworks[41] to automate generation and filtering of adsorbate initial guesses and saddle point guesses, optimization of adsorbates and saddle points, frequency calculations, IRC calculations, and statistical mechanics calculations. To efficiently generate saddle point guesses considering all unique adsorbate configurations on the surface, we developed a novel reaction class agnostic saddle point guess generation technique we call harmonically forced saddle point (HFSP) search. Pynta is designed to be able to take advantage of current petascale and upcoming exascale computing resources. We have demonstrated Pynta on 11 reactions including bidentate, monodentate, and gas-phase species, seven different reaction classes, and both a low index and a high index surface. Based on the saddle points found we are able to conclude that considering all possible site configurations when developing saddle point guesses is important for inter-adsorbate group transfers and reactions on high index surfaces.

Acknowledgement

References

(1) Huber, G. W.; Bond, J. Q. Catalysis: The Technology Enabler for New, Low Carbon Energy Technologies. *ChemCatChem* **2017**, *9*, 1521–1522.

(2) Song, C. An overview of new approaches to deep desulfurization for ultra-clean gasoline, diesel fuel and jet fuel. *Catalysis Today* **2003**, *86*, 211–263.

(3) Mathieu, Y.; Sauvanaud, L.; Humphreys, L.; Rowlands, W.; Maschmeyer, T.; Corma, A. Production of High Quality Syncrude from Lignocellulosic Biomass. *ChemCatChem* **2017**, *9*, 1574–1578.

(4) Asadieraghi, M.; Daud, W. M. A. W.; Abbas, H. F. Heterogeneous catalysts for advanced bio-fuel production through catalytic biomass pyrolysis vapor upgrading: a review. *RSC Advances* **2015**, *5*, 22234–22255.

(5) Hayward, J. S.; Smith, P. J.; Kondrat, S. A.; Bowker, M.; Hutchings, G. J. The Effects of Secondary Oxides on Copper-Based Catalysts for Green Methanol Synthesis. *ChemCatChem* **2017**, *9*, 1655–1662.

(6) Kattel, S.; Ramírez, P. J.; Chen, J. G.; Rodriguez, J. A.; Liu, P. Active sites for CO2 hydrogenation to methanol on Cu/ZnO catalysts. *Science* **2017**, *355*, 1296–1299.

(7) Liang, Z.; Li, T.; Kim, M.; Asthagiri, A.; Weaver, J. F. Low-Temperature activation of methane on the IrO2 (110) surface. *Science* **2017**, *356*, 299–303.

(8) Lu, Y.; Zhang, Z.; Lin, F.; Wang, H.; Wang, Y. Single-atom Automobile Exhaust Catalysts. *ChemNanoMat* **2020**, *6*, 1659–1682.

(9) Kritsanaviparkporn, E.; Baena-Moreno, F. M.; Reina, T. R. Catalytic Converters for Vehicle Exhaust: Fundamental Aspects and Technology Overview for Newcomers to the Field. *Chemistry 2021, Vol. 3, Pages 630-646* **2021**, *3*, 630–646.

(10) Li, H.; Zhu, H.; Zhuang, Z.; Lu, S.; Duan, F.; Du, M. Single-atom catalysts for electrochemical clean energy conversion: recent progress and perspectives. *Sustainable Energy Fuels* **2020**, *4*, 996–1011.

(11) Cheng, N.; Zhang, L.; Doyle-Davis, K.; Sun, X. Single-Atom Catalysts: From Design to Application. *Electrochemical Energy Reviews* **2019**, *2*, 539–573.

(12) Motagamwala, A. H.; Dumesic, J. A. Microkinetic Modeling: A Tool for Rational Catalyst Design. *Chemical Reviews* **2021**, *121*, 1049–1076.

(13) Chen, B. W.; Xu, L.; Mavrikakis, M. Computational Methods in Heterogeneous Catalysis. *Chemical Reviews* **2021**, *121*, 1007–1048.

(14) Goldsmith, C. F.; West, R. H. Automatic Generation of Microkinetic Mechanisms for Heterogeneous Catalysis. *Journal of Physical Chemistry C* **2017**, *121*, 9970–9981.

(15) Gao, C. W.; Allen, J. W.; Green, W. H.; West, R. H. Reaction Mechanism Generator: Automatic construction of chemical kinetic mechanisms. *Computer Physics Communications* **2016**, *203*, 212–225.

(16) 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.; Blondal, K.; West, R. H.; Goldsmith, C. F.; Green, W. H. Reaction Mechanism Generator v3.0: Advances in Automatic Mechanism Generation. *Journal of Chemical Information and Modeling* **2021**, *61*, 2686–2696.

(17) Johnson, M. S. et al. RMG Database for Chemical Property Prediction. *Journal of Chemical Information and Modeling* **2022**, *62*, 4906–4915.

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

(19) Johnson, M. S.; Green, W. H. A Machine Learning Based Approach to Reaction Rate Estimation. **2022**,

(20) Lindgren, P.; Kastlunger, G.; Peterson, A. A. Scaled and Dynamic Optimizations of Nudged Elastic Bands. *Journal of Chemical Theory and Computation* **2019**, *15*, 5787–5793.

(21) Henkelman, G.; Uberuaga, B. P.; Jónsson, H. Climbing image nudged elastic band method for finding saddle points and minimum energy paths. *Journal of Chemical Physics* **2000**, *113*, 9901–9904.

(22) Henkelman, G.; Uberuaga, B. P.; Jónsson, H. A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *The Journal of Chemical Physics* **2000**, *113*, 9901.

(23) Henkelman, G.; Jónsson, H. Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points. *The Journal of Chemical Physics* **2000**, *113*, 9978.

(24) Grabow, L. C.; Mavrikakis, M. Mechanism of methanol synthesis on cu through CO2

and CO hydrogenation. *ACS Catalysis* **2011**, *1*, 365–384.

(25) Zhao, Q.; Xu, Y.; Greeley, J.; Savoie, B. M. Deep reaction network exploration at a heterogeneous catalytic interface. *Nature Communications 2022 13:1* **2022**, *13*, 1–9.

(26) Dana, A. G.; Ranasinghe, D.; Wu, O. H.; Grambow, C.; Dong, X.; Johnson, M.; Goldman, M.; Liu, M.; Green, W. H. ReactionMechanismGenerator/ARC: ARC 1.1.0. **2019**,

(27) Dana, A. G. et al. Automated Reaction Kinetics and Network Exploration (Arkane): A Statistical Mechanics, Thermodynamics, Transition State Theory, and Master Equation Software. **2022**,

(28) 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.

(29) Elliott, S. N.; Moore, K. B.; Copan, A. V.; Georgievskii, Y.; Keçeli, M.; Somers, K. P.; Ghosh, M. K.; Curran, H. J.; Klippenstein, S. J. Systematically derived thermodynamic properties for alkane oxidation. *Combustion and Flame* **2023**, 112487.

(30) 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.

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

(32) Zádor, J.; Van De Vijver, R.; Martí, C.; Dewyer, A. L. KinBot 2.0. 2019; `https://github.com/zadorlab/KinBot`, Last accessed May 26, 2023.

(33) Annevelink, E. et al. AutoMat: Accelerated Computational Electrochemical systems Discovery. **2020**,

(34) Kavalsky, L.; Hegde, V. I.; Muckley, E.; Johnson, M. S.; Meredig, B.; Viswanathan, V. By how much can closed-loop frameworks accelerate computational materials discovery? **2022**,

(35) Bannwarth, C.; Caldeweyher, E.; Ehlert, S.; Hansen, A.; Pracht, P.; Seibert, J.; Spicher, S.; Grimme, S. Extended tight-binding quantum chemistry methods. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2021**, *11*, e1493.

(36) Larsen, H. et al. The atomic simulation environment—a Python library for working with atoms. *Journal of Physics: Condensed Matter* **2017**, *29*, 273002.

(37) Fletcher, R. Practical Methods of Optimization. *Practical Methods of Optimization* **2000**,

(38) 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**, 6536–6549.

(39) 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.

(40) 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**,

(41) Jain, A.; Ong, S. P.; Chen, W.; Medasani, B.; Qu, X.; Kocher, M.; Brafman, M.; Petretto, G.; Rignanese, G. M.; Hautier, G.; Gunter, D.; Persson, K. A. FireWorks: a dynamic workflow system designed for high-throughput applications. *Concurrency and Computation: Practice and Experience* **2015**, *27*, 5037–5059.

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

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

(43) GitHub - SUNCAT-Center/CatKit: General purpose tools for high-throughput catalysis. `https://github.com/SUNCAT-Center/CatKit`.

(44) Grimme, S.; Bannwarth, C.; Shushkov, P. A Robust and Accurate Tight-Binding Quantum Chemical Method for Structures, Vibrational Frequencies, and Noncovalent Interactions of Large Molecular Systems Parametrized for All spd-Block Elements (Z = 1-86). *Journal of Chemical Theory and Computation* **2017**, *13*, 1989–2009.

(45) Pracht, P.; Caldeweyher, E.; Ehlert, S.; Grimme, S. A Robust Non-Self-Consistent Tight-Binding Quantum Chemistry Method for large Molecules. **2019**,

(46) Landrum, G. RDKit. **2010**,

(47) Zimmerman, P. Reliable Transition State Searches Integrated with the Growing String Method. *Journal of Chemical Theory and Computation* **2013**, *9*, 3043–3050.

(48) Zimmerman, P. M. Single-ended transition state finding with the growing string method. *Journal of Computational Chemistry* **2015**, *36*, 601–611.

(49) Behn, A.; Zimmerman, P. M.; Bell, A. T.; Head-Gordon, M. Efficient exploration of reaction paths via a freezing string method. *The Journal of Chemical Physics* **2011**, *135*, 224108.

(50) Sameera, W. M.; Maeda, S.; Morokuma, K. Computational Catalysis Using the Artificial Force Induced Reaction Method. *Accounts of Chemical Research* **2016**, *49*, 763–773.

(51) Maeda, S.; Harabuchi, Y. Exploring paths of chemical transformations in molecular and periodic systems: An approach utilizing force. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2021**, *11*, e1538.

(52) Maeda, S.; Taketsugu, T.; Morokuma, K. Exploring transition state structures for intramolecular pathways by the artificial force induced reaction method. *Journal of Computational Chemistry* **2014**, *35*, 166–173.

(53) Hatanaka, M.; Yoshimura, T.; Maeda, S. Artificial Force-Induced Reaction Method for Systematic Elucidation of Mechanism and Selectivity in Organometallic Reactions. *Topics in Organometallic Chemistry* **2020**, *67*, 57–80.

(54) Bhoorasingh, P. L.; Slakman, B. L.; Khanshan, F. S.; Cain, J. Y.; West, R. H. Automated Transition State Theory Calculations for High-Throughput Kinetics. *Journal of Physical Chemistry A* **2017**, *121*.

(55) Harms, N. D.; Underkoffler, C. E.; West, R. H. Advances in Automated Transition State Theory Calculations: Improvements on the AutoTST Framework. **2020**,

(56) Pattanaik, L.; Dong, X.; Spiekermann, K.; Green, W. An Automated Workflow to Rapidly and Accurately Generate Transition State Structures Using Machine Learning. 2021.

(57) Winther, K. T.; Hoffmann, M. J.; Boes, J. R.; Mamun, O.; Bajdich, M.; Bligaard, T. Catalysis-Hub.org, an open electronic structure database for surface reactions. *Scientific Data 2019 6:1* **2019**, *6*, 1–10.

(58) Chanussot, L. et al. Open Catalyst 2020 (OC20) Dataset and Community Challenges. *ACS Catalysis* **2021**, *11*, 6059–6072.

(59) Campbell, C. T.; Sprowl, L. H.; Árnadóttir, L. Equilibrium Constants and Rate Constants for Adsorbates: Two-Dimensional (2D) Ideal Gas, 2D Ideal Lattice Gas, and Ideal Hindered Translator Models. *Journal of Physical Chemistry C* **2016**, *120*, 10283–10297.

(60) Hermes, E. D.; Janes, A. N.; Schmidt, J. R. Micki: A python-based object-oriented

microkinetic modeling code. *Journal of Chemical Physics* **2019**, *151*.

(61) 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.

(62) Blöchl, P. E. Projector augmented-wave method. *Physical Review B* **1994**, *50*, 17953.

(63) Corso, A. D. Pseudopotentials periodic table: From H to Pu. *Computational Materials Science* **2014**, *95*, 337–350.

(64) Giannozzi, P. et al. QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. *Journal of Physics: Condensed Matter* **2009**, *21*, 395502.

(65) Schumann, J.; Medford, A. J.; Yoo, J. S.; Zhao, Z. J.; Bothra, P.; Cao, A.; Studt, F.; Abild-Pedersen, F.; Nørskov, J. K. Selectivity of Synthesis Gas Conversion to C2+ Oxygenates on fcc(111) Transition-Metal Surfaces. *ACS Catalysis* **2018**, *8*, 3447–3453.

(66) Hammer, B.; Hansen, L. B.; Nørskov, J. K. Improved adsorption energetics within density-functional theory using revised Perdew-Burke-Ernzerhof functionals. *Physical Review B* **1999**, *59*, 7413.

(67) Wang, S.; Temel, B.; Shen, J.; Jones, G.; Grabow, L. C.; Studt, F.; Bligaard, T.; Abild-Pedersen, F.; Christensen, C. H.; Nørskov, J. K. Universal Brønsted-Evans-Polanyi relations for C-C, C-O, C-N, N-O, N-N, and O-O dissociation reactions. *Catalysis Letters* **2011**, *141*, 370–373.

(68) Yoo, J. S.; Abild-Pedersen, F.; Nørskov, J. K.; Studt, F. Theoretical analysis of transition-metal catalysts for formic acid decomposition. *ACS Catalysis* **2014**, *4*, 1226–1233.

(69) Falsig, H.; Hvolbaek, B.; Kristensen, I. S.; Jiang, T.; Bligaard, T.; Christensen, C. H.; Nørskov, J. K.; þ, C.; Hvolbaek, B.; Kristensen, I. S.; Jiang, T.; Bligaard, T.; Nørskov, J. K.; Falsig, H.; Christensen, C. H. Trends in the Catalytic CO Oxidation Activity of Nanoparticles. *Angewandte Chemie International Edition* **2008**, *47*, 4835–4839.

(70) Alvarez-Idaboy, J. R.; Mora-Diez, N.; Vivier-Bunge, A. A quantum chemical and classical transition state theory explanation of negative activation energies in OH addition to substituted ethenes. *Journal of the American Chemical Society* **2000**, *122*, 3715–3720.

(71) Greenwald, E. E.; North, S. W.; Georgievskii, Y.; Klippenstein, S. J. A two transition state model for radical-molecule reactions: A case study of the addition of OH to C 2H 4. *Journal of Physical Chemistry A* **2005**, *109*, 6031–6044.

(72) Georgievskii, Y.; Klippenstein, S. J. Variable reaction coordinate transition state theory: Analytic results and application to the C2H3 + H → C2H4 reaction. *J. Chem. Phys.* **2003**, *118*, 5442–5455.

(73) Johnson, M. S.; Pang, H.-W.; Payne, A. M.; Green, W. H. ReactionMechanismSimulator.jl: A Modern Approach to Chemical Kinetic Mechanism Simulation and Analysis. **2023**,

(74) Johnson, M. S.; McGill, C. J.; Green, W. H. Transitory Sensitivity in Automatic Chemical Kinetic Mechanism Analysis. **2022**,

(75) Safta, C.; Najm, H. N.; Knio, O. TChem - A Software Toolkit for the Analysis of Complex Kinetic Models. 2011; http://www.osti.gov/servlets/purl/1113874/.

(76) 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.

TOC Graphic

$$d_{eqi} = f_b d_{well}$$
$$k_i = k_{hard}$$

$$d_{eqi} = 0$$
$$k_i = k_{hard}$$