Breaking scaling relations with inverse catalysts: a machine learning exploration of trends in CO_2 hydrogenation energy barriers

Luuk H. E. Kempen[1], Marius Juul Nielsen[1], and Mie Andersen[1,2,*]

[1]Center for Interstellar Catalysis, Department of Physics and Astronomy, Aarhus University, Aarhus C, Denmark
[2]Aarhus Institute of Advanced Studies, Aarhus University, Aarhus C, Denmark
[*]Corresponding author: mie@phys.au.dk

April 24, 2025

The conversion of CO_2 into useful products such as methanol is a key strategy for abating climate change and our dependence on fossil fuels. Developing new catalysts for this process is costly and time-consuming and can thus benefit from computational exploration of possible active sites. However, this is complicated by the complexity of the materials and reaction networks. Here, we present a workflow for exploring transition states of elementary reaction steps at inverse catalysts, which is based on the training of a neural network-based machine learning interatomic potential. We focus on the crucial formate intermediate and its formation over nanoclusters of indium oxide supported on Cu(111). The speedup compared to an approach purely based on density functional theory allows us to probe a wide variety of active sites found at nanoclusters of different sizes and stoichiometries. Analysis of the obtained set of transition state geometries reveals different structure–activity trends at the edge or interior of the nanoclusters. Furthermore, the identified geometries allow for the breaking of linear scaling relations, which could be a key underlying reason for the excellent catalytic performance of inverse catalysts observed in experiments.

1 Introduction

In industry, the $Cu/ZnO/Al_2O_3$ catalyst is currently employed in the conventional method of methanol production from CO and CO_2 [1]. Recently, ZnO_x overlayers have been observed on the Cu nanoparticles [2], possibly hinting to metal oxides playing an important role in the catalytic process. Furthermore, inverse catalysts—metal oxide nanoparticles on a metal support—have been shown to be more active than their respective conventional counterparts for CO_2 hydrogenation [3–5]. Hence, inverse catalysts are a promising class of catalyst materials.

In computational screening of catalysts, microkinetic modeling is an invaluable technique for predicting catalytic activity [6]. Part of constructing such a model is collecting adsorption energies and activation energies of elementary reaction steps on the catalyst surface, which in turn requires knowledge of the surface geometry and active sites. These inputs are typically obtained from density functional theory (DFT) calculations. Moreover, for metal surfaces, linear scaling relations often allow for cheap estimation of activation energies from reaction energies through Brønsted–Evans–Polanyi (BEP) relations [7–10].

Inverse catalysts are much more difficult to model than low-index metal facets due to the additional complexity in possible compositions and structures [11, 12]. Furthermore, the structures can be highly asymmetric, leading to many possible active sites to consider in the modelling (see Figure 1). And to make matters worse, previous work on oxide-supported metal catalysts have revealed that linear scaling relations may not generally hold at metal–oxide interfaces [13, 14], which necessitates the development of other computationally tractable approaches for obtaining activation energies.

Activation energies can be calculated by finding saddle points on the potential energy surface. Common methods for this task are interpolation methods, e.g., the nudged elastic band (NEB) method [15], that optimize a chain of images along the reaction path, or local methods, e.g., the dimer method [16], that optimize one or two images near the transition state (TS). However, all these methods can be hard to converge and often require many potential energy evaluations. Combined with the immense search space of possible active sites, this renders a pure DFT approach to TS search at inverse catalysts intractible.

In recent years, machine learning interatomic potentials (MLIPs) have become viable for obtaining approximations of the potential energy surface at near-DFT-level accuracy for a fraction of the computational cost. This unlocks the possibility of high-throughput screening of elementary reaction steps across cluster compositions, structures and active sites. In this work, we develop a ML-driven workflow for obtaining accurate TS guesses for elementary reaction steps. The employed MLIPs are based on the Gaussian moment neural net-
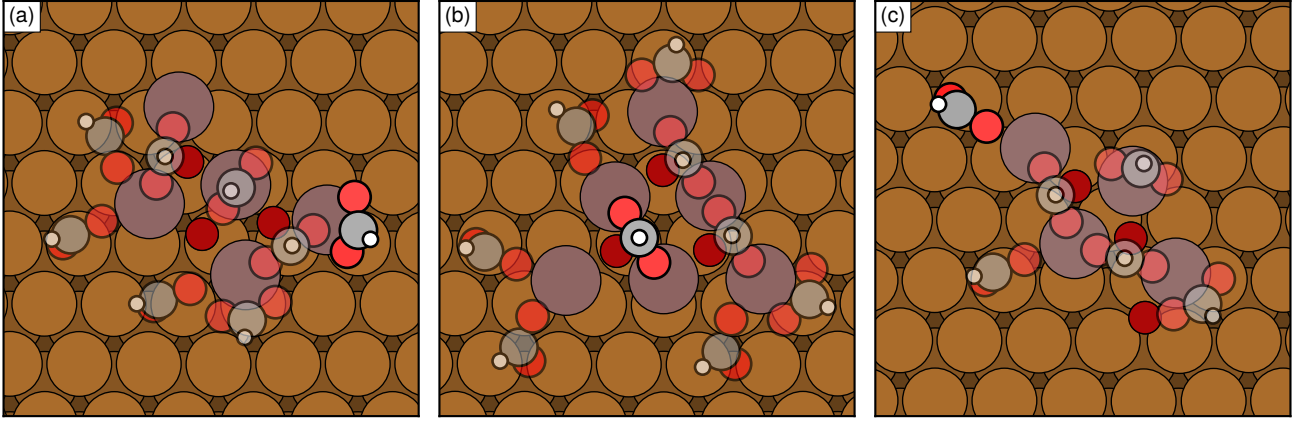
Figure 1: A subset of (unoptimized) possible formate binding sites on three $In_yO_x/Cu(111)$ inverse nanocluster geometries. Only non-overlapping binding sites are shown for ease of visualization. On each geometry, one binding site is highlighted to indicate three different types of binding sites: (a) on-top; (b) bridging between two cluster atoms; (c) bridging at the metal–oxide interface. Color legend: Cu (brown), In (purple), O in oxide (dark red), O in formate (light red), C (gray), H (white).

work (GM-NN) architecture [17, 18]. We focus here on the $In_yO_x/Cu(111)$ inverse catalyst system, for which stable cluster compositions and structures were identified in our previous work [12]. The rationale for exploring this system in detail is (i) that Cu and indium oxide are interesting catalysts for CO_2 conversion in their own right [19, 20] and (ii) that much less attention has been devoted to this system compared to other similar inverse catalyst systems such as $Zn_yO_x/Cu(111)$ [21, 22]. And we focus on the formation of formate (HCOO) from CO_2 and H, as formate has been found to be a key intermediate in CO_2 hydrogenation to methanol [20, 21, 23].

We show that our developed workflow can not only predict accurate TS structures for active sites considered in the training, but also generalizes well to new active site motifs. This capability allows us to assemble a large data set of TS structures across inverse catalyst compositions and structures at a fraction of the computational cost required for traditional DFT-based approaches. Analysis of this TS data set reveals siginificant structure-activity trends. In particular, we show that the active sites at inverse catalysts can break linear scaling relations, which may be a key reason for the high catalytic activity reported in experiments [3–5, 24].

2 Methods

2.1 Density functional theory calculations

All training and validation data for the machine learning models was obtained with density functional theory (DFT). The DFT calculations were carried out using the GPAW package [25–27], which implements the projector augmented-wave method [28]. GPAW was used in the plane-wave mode and interfaced with the atomic simulation environment (ASE) package [29]. The exchange–correlation functional was approximated using the Perdew–Burke–Ernzerhof (PBE) functional [30].

All calculations were further corrected by a DFT-D4 dispersion correction [31–33]. The metal slab model was not modified from our previous work [12]; it was constructed using a $7\times7\times4$ supercell with a vacuum of 18 Å between the metal layers in the direction perpendicular to the surface. Orbital occupancies were set using the Fermi–Dirac distribution with a width of 0.10 eV.

To speed up data collection, optimization performed to obtain structures for training data was done with low-level DFT settings: a plane-wave cutoff of 400 eV was employed and the Brillouin zone was sampled using the Γ point. Structure optimization was performed using the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm with an atomic force convergence of $F_{\max} = 0.05 \, \text{eV} \, \text{Å}^{-1}$, where the topmost layer of the metal surface, the nanocluster, and the adsorbate atoms were allowed to relax. DFT validation of TSs was performed using a local bond climbing method, beginning from a suitable ML-optimized guess structure, as outlined in Section 2.4. For this we employed a plane-wave cutoff of 600 eV and Γ point sampling of the Brillouin zone. Structures selected to be part of the training data as well as optimized TS structures were finally refined with a single-point higher-level DFT calculation employing a plane-wave cutoff of 600 eV and a $2 \times 2 \times 1$ Monkhorst–Pack k-point grid [34] for the Brillouin zone sampling.

Formation energies, E_f, of adsorbates (formate or the TS for its formation) were calculated as

$$E_f = E_{\text{cluster+ads}} - E_{\text{cluster}} - E_{CO_2 \text{(g)}} - \frac{1}{2}E_{H_2 \text{(g)}} \quad (1)$$

where $E_{\text{cluster+ads}}$ is the energy of the nanocluster with the adsorbate, E_{cluster} is the energy of the clean nanocluster, and $E_{CO_2 \text{(g)}}$ and $E_{H_2 \text{(g)}}$ are the energies of gas-phase CO_2 and H_2, respectively.

2.2 Machine learning interatomic potential

In this work, we use the Gaussian moment neural network (GM-NN) architecture [17, 18] as machine learn-

ing interatomic potential, as implemented in the Apax package [35]. In this architecture, the total energy of a system S can be decomposed into a sum of atomic energies [36]

$$\hat{E}(S) \approx \sum_{i=1}^{N_{\text{at}}} \hat{E}_i(\mathbf{G}_i), \tag{2}$$

where $\mathbf{G}_i$ is the local atomic representation of atom i. This representation is constructed from the pair distance vectors $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$ for all atomic pairs within a predefined cutoff radius. These pair distance vectors are used in a translationally and permutationally invariant tensor-valued function, which is converted into features $\mathbf{G}_i$ that are also rotationally invariant through full tensor contractions. These features are used as input to a fully connected feedforward neural network to give atomic energy and force predictions.

The GM-NNs were generally trained in two steps. First, a single network was trained using mean squared error (MSE) on energies and forces as loss function, where the force error was weighted by a factor of $3N_{\text{at}}$ (three times the number of atoms) relative to the energy error. With the exception of last-layer weights, the parameters of this network were then transferred to a shallow ensemble with 10 output heads for further training with a negative log-likelihood (NLL) loss function. In this network, the last layer is split into an ensemble of models, following the direct propagation of shallow ensembles (DPOSE) architecture [37]. Each model head makes its own predictions, and the model output is given by the mean of the ensemble predictions. This also allows for evaluation of an ensemble uncertainty *via* the variance of the individual predictions.

Hyperparameters for the GM-NN were obtained by optimization using the Optuna framework [38]. After setting up a space of hyperparameters for the GM-NN models, Optuna dynamically selects suggestions for hyperparameter values based on earlier results. With these suggestions, a GM-NN model is trained and the resulting loss value on validation data is used to feed hyperparameter choices for the next trials. The set of hyperparameters resulting in the optimal validation loss was then chosen for further GM-NN model training. Learning rates were reduced for training the shallow ensemble. An overview of the optimized hyperparameters for the used networks is given in Table S1.

In total, 10 separate shallow-ensemble GM-NN models were trained fully independently, with the goal of obtaining models to individually predict slightly different TS geometries.

2.3 Training data acquisition

To train the GM-NN models for TS prediction, a data set containing structures relevant for the reaction pathway was constructed. Here we focused on the CO_2 + H $\rightleftharpoons$ HCOO reaction on In_yO_x/Cu(111). The geometries of formate adsorbed on inverse catalysts from our earlier work [12, 39] form the basis for the training data in this work. We included In_yO_x nanoclusters

with 3 to 6 indium atoms and 2 to 3 oxygen atoms, including all nanoclusters within 0.1 eV of the global minimum energy. From these formate-adsorbed nanoclusters and corresponding clean nanoclusters, structures approximating the reaction pathway were constructed using the following procedure:

1. Sample formate-adsorbed structures from our earlier work [39];

2. For each formate-adsorbed structure, generate a corresponding CO_2-adsorbed structure by removing the hydrogen atom from formate and relaxing the resulting structure;

3. Add an H atom to the metal surface on the three-fold site closest to the CO_2 adsorbate and relax the resulting structure;

4. Generate an initial guess of a reaction path by interpolating 10 images between the initial state (CO_2* + H*) and the corresponding final state (HCOO*) using an image dependent pair potential (IDPP) path [40] with a force convergence of $F_{\max} = 0.1\,\text{eV}\,\text{Å}^{-1}$, and perform single-point DFT evaluations for these interpolated images;

5. With the highest-energy image from the interpolated path, carry out two short (i.e., not converged) local TS optimizations in parallel; one using the ClimbFixInternals methods from ASE [29, 41] and one using the bond climbing method outlined in Section 2.4.

Using the above approach, we sampled 48 pairs of initial and final states of the CO_2 + H $\rightleftharpoons$ HCOO reaction. Structures from all the steps were combined into a single data set, where only a small subset of the structures from step 5 were chosen due to their high relative similarity. This resulted in a total of 1120 structures. Due to practical memory constraints, training the GM-NN models was only done on a randomly-selected subset of 694 structures; this was further split into training and validation sets using a 90 : 10 split. The remaining 426 structures were used as testing data.

The 48 pairs of initial and final states used for training data acquisition were also used for TS search (outlined below), forming an in-sample TS data set for searching purposes. Additionally, steps 1 through 3 of the procedure above were repeated for 76 new pairs of initial and final states, ensuring that the sampled adsorption geometries are distinct from the ones included in the training dataset, to generate a fully out-of-sample TS data set.

2.4 Transition state search

TS searches were performed on the in-sample TS data set consisting of 48 pairs of initial and final states of the CO_2 + H $\rightleftharpoons$ HCOO reaction, as well as on the 76 pairs of initial and final states forming the out-of-sample TS data set. Using the set of 10 independently trained GM-NN models (see Section 2.2), 10 TS searches were performed for each pair of initial and final states using a two-step procedure.

The first step consists of a nudged elastic band (NEB) calculation [15]. To begin, 40 images are interpolated between the initial and the final state using the IDPP interpolation method with a force convergence of $F_{\mathrm{max}} = 0.01\,\mathrm{eV\,\mathring{A}^{-1}}$. The NEB calculation is then performed with the IDPP path as a starting guess, and the minimum energy path is evolved using the `ode12r` solver [42] until either convergence ($F_{\mathrm{max}} < 0.05\,\mathrm{eV\,\mathring{A}^{-1}}$) or stagnation of the solver. The NEB force evaluation is switched to the climbing image method [43] if $F_{\mathrm{max}} < 0.10\,\mathrm{eV\,\mathring{A}^{-1}}$. The highest-energy image, if not yet optimized until convergence, is then selected for further refinement with a local bond climbing method [44]. In this method, the forces acting on the carbon and hydrogen atoms—between which a bond is to be formed in the formation of formate—are reflected with respect to the plane perpendicular to the bond vector. This path is then similarly evolved using the `ode12r` solver until convergence or stagnation. During TS search, only the nanocluster and adsorbate atoms are allowed to move.

Validation of TS guesses produced by the GM-NN model was performed by continuing the bond climbing method outlined above with DFT instead of with the GM-NN model until convergence. This allows for energetic and geometric comparison between the GM-NN-converged TS structure and the DFT-converged TS structure.

2.5 Active learning

An active learning workflow was created to attempt to iteratively improve the GM-NN models. The IPSuite framework [45, 46] was utilized to set up the workflow, benefitting from its native integration with Apax.

First, with the training and validation data sets outlined in Section 2.3 and hyperparameters determined as outlined in Section 2.2, a GM-NN model was trained.

Second, the model was used as a potential for performing NEB calculations for all 48 binding sites with initial–final pairs. These calculations were initialized using IDPP-interpolated paths with 10 interpolated images and an IDPP force convergence of $F_{\mathrm{max}} = 0.1\,\mathrm{eV\,\mathring{A}^{-1}}$. The minimum energy path was evolved using the `ode12r` solver either until convergence ($F_{\mathrm{max}} < 0.05\,\mathrm{eV\,\mathring{A}^{-1}}$) or until the uncertainty on the forces predicted by the shallow ensemble exceeded the magnitude of the force on any atom in the interpolated images.

Third, the 10 structures from the full NEB trajectories predicted to most improve model training were selected. The selection was done by evaluating the gradient of the last-layer weights of the network with respect to each candidate structure [47]. A gradient kernel was then established by considering all combinations of pairs of structures, and farthest point sampling (greedy distance maximization) was applied to this gradient kernel to select the structures. This corresponds to the method referred to as MaxDist + FEAT(LL) by Zaverkin *et al.* [47].

Fourth, the 10 selected structures were evaluated using single-point DFT calculations as outlined in Sec-
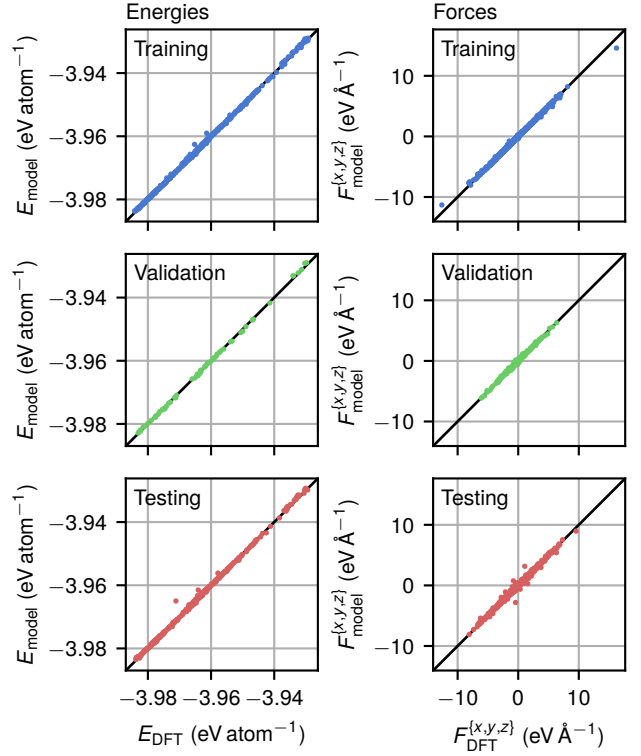

Figure 2: Parity plots showing prediction performance of one of the trained GM-NN models on the full training data. The force predictions shown include forces on the nanocluster and formate atoms.

tion 2.1 and added to the training data set.

After expanding the training data set and retraining the model, the workflow returns to the second step where NEB calculations are again initialized from IDPP-interpolated paths. Every third iteration of the workflow, the GM-NN model is retrained from scratch in two steps: first an MSE model, then a shallow ensemble. The other iterations, only the shallow ensemble model is retrained, initialized from the latest available MSE model. This approach balances time spent on training with model prediction quality.

3 Results

3.1 Model training

We begin by assessing the performance of the 10 separate GM-NN models all trained on the same initial training data set. Figure 2 shows parity plots for one of the models, which achieves a root mean square error (RMSE) of $0.43\,\mathrm{meV\,atom^{-1}}$ for energy predictions and $62\,\mathrm{meV\,\mathring{A}^{-1}}$ for force component predictions (on the nanocluster and formate atoms) on the testing set. On average, all 10 models achieve an RMSE of $0.54\,\mathrm{meV\,atom^{-1}}$ for energies and $69\,\mathrm{meV\,\mathring{A}^{-1}}$ for force components on the testing set. RMSEs for all 10 models on all data subsets are provided in Section S1. Due to stochastic model training, not all models perform equally well; two models in particular stand out for performing worse than the average.

3.2 In-sample TS search

With the 10 trained GM-NN models, we subsequently performed TS search on the in-sample set of 48 pairs of initial and final states of the $CO_2 + H \rightleftharpoons HCOO$ reaction. Since some of the models produced unreasonable TS guesses, we applied the following filtering steps. First, a check of atomic distances was performed and structures with atoms too close to each other were filtered out. The remaining structures were evaluated with a single-point DFT calculation to obtain accurate energy values. Secondly, we discarded all structures that were more than 1 eV higher in energy than the lowest-energy structure. Finally, to validate the TS guesses, we randomly selected 20 out of the 48 reaction paths and continued the TS optimization by running the bond climbing method with DFT for all ML-optimized TS guesses remaining after the filtering step.

In Figure 3, we show how the formation energies associated with the in-sample TS guess structures produced by the GM-NN models compare to those of the validated TS structures obtained with DFT optimization. 84 % of the ML-optimized TS structures retained after filtering are within 0.1 eV of the DFT-optimized TS structures and the average energy error is 52 meV. Note that the energies for both the ML- and DFT-optimized TS structures (as well as for the initial and final state structures) are obtained from a single-point DFT evaluation using high-level settings, as described in Section 2.1. The TSs for which the error is large are generally also characterized by the DFT validation step displacing the reactant atoms by a relatively large amount, as shown in Figure S2.

It should be noted that the trained GM-NN models achieve this accuracy mostly due to them having learned the general *shape* of the potential energy surface well; the activation energies predicted by the GM-NN for the TS guesses can be quite far off the DFT-evaluated energies for the same geometries. As shown in Figure S3, the mean absolute error of GM-NN activation energy predictions for model-optimized TSs is about 0.2 eV. Therefore, a single-point DFT evaluation of the model-generated TS guess is still required.

Using the procedure described above, the achieved average energy error of 52 meV is already excellent and far below intrinsic errors of DFT with generalized gradient approximation functionals for surface adsorption energies, which are typically about 0.2–0.3 eV [48]. Still, we investigated whether it would be possible to improve on either TS guess structures or energies by acquiring further training data in an active learning framework. However, this did not turn out to meaningfully improve either model prediction performance or TS search performance. We discuss these active learning results in more detail in Section S3.

Nevertheless, our results highlight the power of ML-based TS optimization. The GM-NN models have learned the potential energy surface well enough that TS geometries can be guessed with high accuracy, requiring only a single-point DFT evaluation to obtain the corresponding potential energy. Continuing bond climbing with DFT, as done here for validation pur-
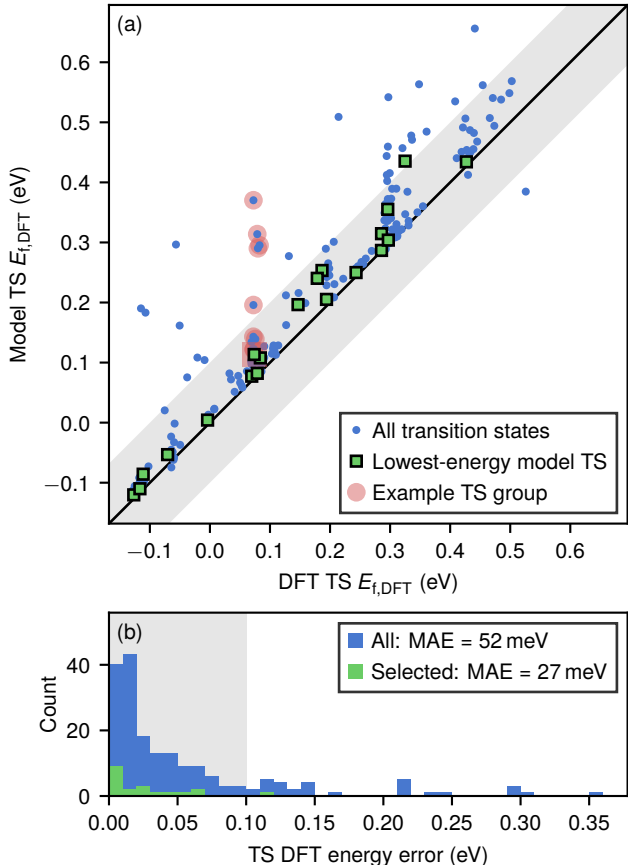

Figure 3: Comparison of energies of in-sample TS guesses produced by the GM-NN models to those of validated TSs produced by continuing bond climbing with DFT. (a) Parity plot of formation energies. The TS guesses highlighted in red correspond to one group of TSs from the same pair of initial and final states. (b) Histogram of errors. The highlighted band in both plots indicates an error below 0.1 eV. The data points highlighted in green correspond to those selected using the selection method described in the text.

poses, took a median of 106 single-point DFT evaluations per TS guess, for an average energy accuracy improvement of only 52 meV.

Interestingly, the DFT validation process as applied in this work tends to lower the TS energies, i.e., TS guesses produced by the GM-NN models are only rarely underpredicting the corresponding formation energy. This is most likely due to the bond climbing method being equivalent to a regular relaxation for all non-reacting atoms, and the energy decrease from relaxing these atoms into their DFT local minimum dominates any energy increase from reacting atoms moving against the force direction. As we have 10 independently-generated TS candidates for every pair of initial and final states (as exemplified with the highlighted points in Figure 3a), we can use this observation to guide TS selection from this set of candidates. Under the assumption that TS guesses are never underpredicted, the model-generated TS guess with the lowest energy must be closest in energy to a DFT-validated TS. While this assumption is not always valid, it is valid often enough
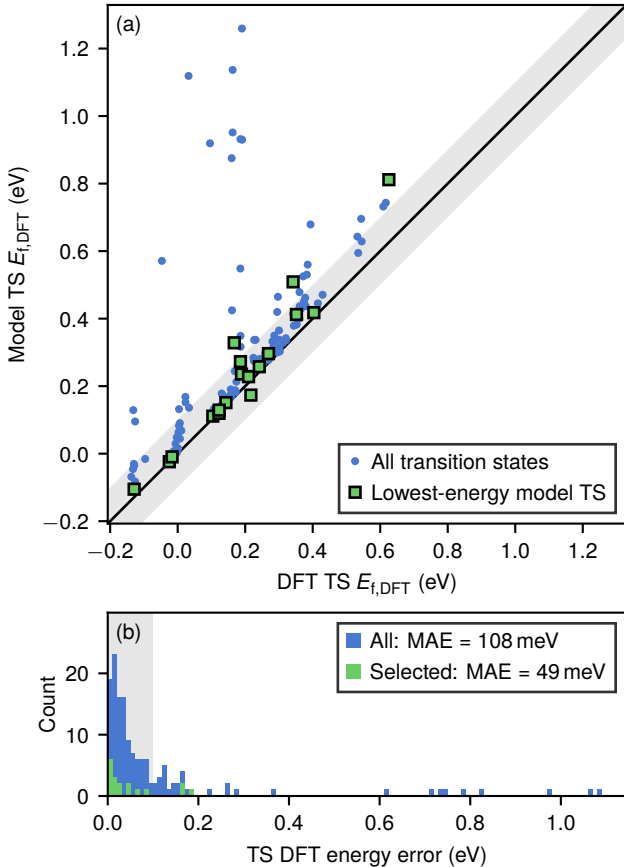
Figure 4: Comparison of energies of out-of-sample TS guesses produced by the GM-NN models to those of validated TSs obtained by continuing bond climbing with DFT. A detailed description is given in the caption of Figure 3.

that selecting TSs with this criterion increases accuracy even further. The TSs selected using this method are indicated in green in Figure 3. With this selection, the average energy error between the model-generated TSs and DFT-validated TSs is reduced from 52 meV to 27 meV, and all but one of the selected TSs have an error below 0.1 eV.

3.3 Out-of-sample TS search

Encouraged by the excellent accuracy for the in-sample TS search, we next test the accuracy of our approach for the out-of-sample set of 76 new pairs of initial and final states, applying the same filtering procedure as for the in-sample data set, and continuing bond climbing with DFT for another random subset of 20 TSs. The activation energies for these out-of-sample TS guesses are compared to the DFT-validated TSs in Figure 4. For the out-of-sample data set, prediction is slightly worse than for the in-sample data set, with 76 % of the TS guesses accurate within 0.1 eV to the DFT-validated TSs.

Analogous to the in-sample data set, applying the TS selection criterion outlined in Section 3.2 improves accuracy on the out-of-sample data set, reducing the average energy error from 108 meV across the full data

set to 49 meV for the selected data set. Direct activation energy prediction by the GM-NN is also similarly performant for the in-sample and out-of-sample data sets, with the latter having a mean absolute error of 0.22 eV (see Figure S4).

Overall, these results highlight that the trained GM-NNs are not only performant on the pairs of initial and final states they have been trained on, but also on geometries not encountered during training. This makes the outlined approach particularly suitable for screening a large data set of nanocluster and adsorbate geometries.

3.4 TS structure–activity trends

With the extensive set of TSs identified, made possible by our computationally efficient workflow relying on TS optimization in the GM-NN model, we next analyze structure–activity trends. For this we employ the combined set of in-sample and out-of-sample model-optimized TS structures, selecting always the TS structure with the lowest DFT energy for each pair of initial and final state, leading to a total number of 124 TSs. Figure 5 shows that the TSs discovered in this work can be categorized into two distinct classes based on the closest H–In distance in the TS. Figure 5c clearly shows a bimodal distribution with a relatively narrow peak for H near In, and a much broader peak for H further away from In. The case where H is near In describes TSs where the hydrogen atom reacts with the CO_2 adsorbate *via* a nanocluster indium atom, as shown in Figure 5a. The case where H is far from In describes TSs where the hydrogen atom more directly binds to CO_2 from the metal surface, as exemplified in Figure 5b. The preference for either of these classes is mostly dictated by geometry; indium atoms present near the hydrogen atom and the CO_2 compound can both offer a site *via* which hydrogen can react, but also hinder the hydrogen atom from approaching CO_2 directly.

The formation energy distributions for these two cases are shown in Figure 5d. The division in H–In distance somewhat correlates with formation energies; those where H is transferred *via* In are on average higher than those for reactions where this does not occur. This is explained by the fact that the adsorption of H on an In nanocluster atom is approximately 0.4 eV less stable than that of H on the Cu surface.

3.5 Scaling relations

Linear scaling relations have in the past been found to be good, simple descriptors of catalytic processes on metal surfaces [9, 10]. The data collected in this work allows for validating whether these correlations also apply to the inverse catalysts explored here. Two types of scaling relations have been evaluated (see Figure 6). The first is the BEP relation, relating activation energy E_{act} to reaction energy ΔE_r, the reference state being H and CO_2 adsorbed to the surface. The second is a relation between the formation energy of the TS and that of formate, the reference state being H_2 and CO_2
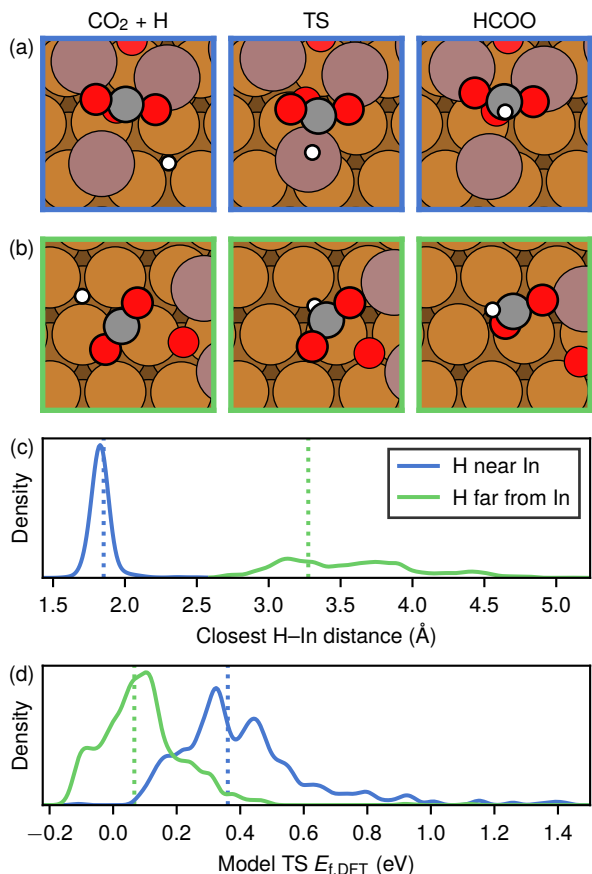
Figure 5: Two distinct classes of TSs. (a–b) Top-down visualizations of (left-to-right) initial state, TS, and final state. (a) A TS where hydrogen is located near indium; (b) a TS where hydrogen is located far from indium. (c) Closest H–In distance distribution for all TSs, showing the distinction in two classes. (d) Formation energy distributions for these two classes of TSs. The dotted lines in (c–d) indicate the distances and formation energies for the examples shown in (a–b).
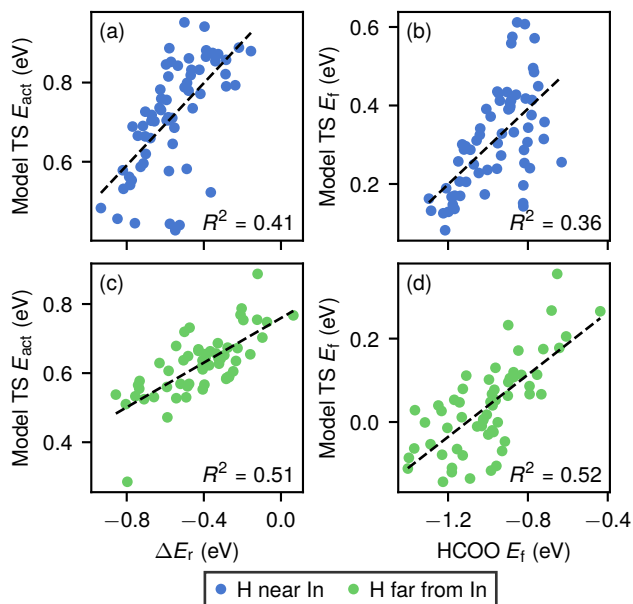

Figure 6: Two types of scaling relations for the formate formation step, distinguishing the two separate TS classes discussed in the text. (a, c) BEP relation; (b, d) relationship between formation energies of the TS and of formate.

in the gas phase.

Since we have found two classes of TSs in the previous section, we evaluate the scaling relations for these classes separately. As visible, the linear relation best describing the data points is a quite poor predictor of the activation energy; prediction errors can be as large as half the full range of activation energies. The relations are better for the class of TSs where hydrogen is not near indium. This is likely because the geometries of these TSs are more similar to that of the final state, formate, than the geometries of TSs where hydrogen reacts *via* an indium atom. The poor fits observed underscore the need for more sophisticated methods of determining activation energies of reaction pathways on these more complex catalyst surfaces.

4   Conclusions

We have developed a ML-driven workflow for exploring energy barriers of elementary reaction steps at inverse catalysts, which captures the complexity associated with the possible cluster sizes and compositions of these catalysts. By acquiring a suitable training data set consisting mainly of interpolated structures along the reaction pathway of the $CO_2 + H \rightleftharpoons HCOO$ step at a diverse subset of cluster geometries, we show that it is possible to train a MLIP (a GM-NN model) that gives accurate TS guess structures for both in-sample cluster sites considered in the training data as well as out-of-sample sites. While model-predicted activation energies based on these structures lack predictive accuracy with a MAE of approximately $0.2\,\text{eV}$, a single-point DFT evaluation of the TS structure guess brings down the TS energy error to a MAE of $52\,\text{meV}$ for in-sample TSs. Furthermore, we introduce a TS guess structure selection method that brings down the error even further: Realizing that TS guess structures predicted by independently trained GM-NN models all tend to be higher in energy than the DFT-optimized TS structure, selecting the model TS structure guess with the lowest DFT energy further lowers the MAE to $27\,\text{meV}$. This procedure is also performant for out-of-sample cluster sites, where we achieve an only slightly higher MAE of $49\,\text{meV}$.

Finally, we make use of the combined data set of 124 model-optimized TS structures to analyze structure–activity trends. We find that the TS structures fall into two categories depending on whether migration of the H atom occurs over the Cu surface next to the cluster edge or over an In atom, where the former structures tend to be lower in energies due to favorable Cu–H interactions. Scaling relations are obeyed for none of the two TS structure categories. In fact, we observe several TS structures that exhibit catalytically interesting deviations from scaling relations, i.e., it is possible to have a low TS formation energy despite a moderate formate

formation energy. Such sites could make it possible to overcome limitations posed by the Sabatier principle (as quantified by scaling relation-based volcano plots in heterogeneous catalysis [49]) by allowing facile activation and adsorption of the reactants (i.e., CO_2 and H_2) without poisoning the surface with too-strongly bound intermediates, which would tend to make product formation steps (e.g., methanol, methane, or carbon monoxide formation) difficult.

In this connection it is noteworthy that a recent experimental study has found that the $In_yO_x/Au(111)$ inverse catalyst is particularly active and selective for methanol formation, where the reaction pathway is expected to occur via the crucial formate intermediate [24]. Similarly, previous studies of other inverse catalyst systems such as $Ce_yO_x/Cu(111)$ and $Zn_yO_x/Cu(111)$ have highlighted their exceptional catalytic properties. Based on our results, it seems likely that these catalytic properties could be related to the ability of the active site motifs found at inverse catalyst to break scaling relations.

Data availability

The structure datasets will be made available on Zenodo upon publication under a CC BY 4.0 license.

Code availability

The workflow scripts will be made available upon publication under a GNU GPLv3 license.

Supporting Information

Additional information on model training, TS guess energies, and active learning (PDF)

Acknowledgments

Author contributions

Luuk H. E. Kempen: Conceptualization (equal); Investigation (lead); Methodology (lead); Software (lead); Visualization (lead), Writing – original draft (lead); Writing – review & editing (equal). **Marius Juul Nielsen**: Investigation (supporting); Methodology (supporting); Visualization (supporting); Writing – review & editing (equal). **Mie Andersen**: Conceptualization (equal); Investigation (supporting); Writing – review & editing (equal); Supervision (lead); Resources (lead); Funding acquisition (lead).

Competing interests

The authors declare no competing interests.

References

1. Guil-López, R. *et al.* Methanol Synthesis from CO_2: A Review of the Latest Developments in Heterogeneous Catalysis. *Materials* **12**, 3902. ISSN: 1996-1944. http://dx.doi.org/10.3390/ma12233902 (2019).

2. Lunkenbein, T., Schumann, J., Behrens, M., Schlögl, R. & Willinger, M. G. Formation of a ZnO Overlayer in Industrial $Cu/ZnO/Al_2O_3$ Catalysts Induced by Strong Metal–Support Interactions. *Angewandte Chemie International Edition* **54**, 4544–4548. ISSN: 1521-3773. http://dx.doi.org/10.1002/anie.201411581 (2015).

3. Graciani, J. *et al.* Highly active copper–ceria and copper–ceria–titania catalysts for methanol synthesis from CO_2. *Science* **345**, 546–550. ISSN: 1095-9203. http://dx.doi.org/10.1126/science.1253057 (2014).

4. Senanayake, S. D. *et al.* Hydrogenation of CO_2 to Methanol on $CeO_x/Cu(111)$ and $ZnO/Cu(111)$ Catalysts: Role of the Metal–Oxide Interface and Importance of Ce^{3+} Sites. *The Journal of Physical Chemistry C* **120**, 1778–1784. ISSN: 1932-7455. http://dx.doi.org/10.1021/acs.jpcc.5b12012 (2016).

5. Wu, C. *et al.* Inverse ZrO_2/Cu as a highly efficient methanol synthesis catalyst from CO_2 hydrogenation. *Nature Communications* **11**, 5767. ISSN: 2041-1723. http://dx.doi.org/10.1038/s41467-020-19634-8 (2020).

6. Motagamwala, A. H. & Dumesic, J. A. Microkinetic Modeling: A Tool for Rational Catalyst Design. *Chemical Reviews* **121**, 1049–1076. ISSN: 1520-6890. http://dx.doi.org/10.1021/acs.chemrev.0c00394 (2020).

7. Brønsted, J. N. Acid and Basic Catalysis. *Chemical Reviews* **5**, 231–338. ISSN: 1520-6890. http://dx.doi.org/10.1021/cr60019a001 (1928).

8. Evans, M. G. & Polanyi, M. Further considerations on the thermodynamics of chemical equilibria and reaction rates. *Transactions of the Faraday Society* **32**, 1333–1360. ISSN: 0014-7672. http://dx.doi.org/10.1039/TF9363201333 (1936).

9. Nørskov, J. K. *et al.* Universality in Heterogeneous Catalysis. *Journal of Catalysis* **209**, 275–278. ISSN: 0021-9517. http://dx.doi.org/10.1006/jcat.2002.3615 (2002).

10. Michaelides, A. *et al.* Identification of General Linear Relationships between Activation Energies and Enthalpy Changes for Dissociation Reactions at Surfaces. *Journal of the American Chemical Society* **125,** 3704–3705. ISSN: 1520-5126. http://dx.doi.org/10.1021/ja027366r (2003).

11. Reichenbach, T., Walter, M., Moseler, M., Hammer, B. & Bruix, A. Effects of Gas-Phase Conditions and Particle Size on the Properties of Cu(111)-Supported Zn_yO_x Particles Revealed by Global Optimization and Ab Initio Thermodynamics. *The Journal of Physical Chemistry C* **123,** 30903–30916. ISSN: 1932-7455. http://dx.doi.org/10.1021/acs.jpcc.9b07715 (2019).

12. Kempen, L. H. E. & Andersen, M. Inverse catalysts: tuning the composition and structure of oxide clusters through the metal support. *npj Computational Materials* **11,** 8. ISSN: 2057-3960. http://dx.doi.org/10.1038/s41524-024-01507-z (2025).

13. Mehta, P., Greeley, J., Delgass, W. N. & Schneider, W. F. Adsorption Energy Correlations at the Metal–Support Boundary. *ACS Catalysis* **7,** 4707–4715. ISSN: 2155-5435. http://dx.doi.org/10.1021/acscatal.7b00979 (2017).

14. Lustemberg, P. G. *et al.* Breaking Simple Scaling Relations through Metal–Oxide Interactions: Understanding Room-Temperature Activation of Methane on M/CeO_2 (M = Pt, Ni, or Co) Interfaces. *The Journal of Physical Chemistry Letters* **11,** 9131–9137. ISSN: 1948-7185. http://dx.doi.org/10.1021/acs.jpclett.0c02109 (2020).

15. Jónsson, H., Mills, G. & Jacobsen, K. W. *Nudged elastic band method for finding minimum energy paths of transitions* in *Classical and Quantum Dynamics in Condensed Phase Simulations* (World Scientific, 1998), 9978–9985. http://dx.doi.org/10.1142/9789812839664_0016.

16. Henkelman, G. & Jónsson, H. A dimer method for finding saddle points on high dimensional potential surfaces using only first derivatives. *The Journal of Chemical Physics* **111,** 7010–7022. ISSN: 1089-7690. http://dx.doi.org/10.1063/1.480097 (1999).

17. Zaverkin, V. & Kästner, J. Gaussian Moments as Physically Inspired Molecular Descriptors for Accurate and Scalable Machine Learning Potentials. *Journal of Chemical Theory and Computation* **16,** 5410–5421. ISSN: 1549-9626. http://dx.doi.org/10.1021/acs.jctc.0c00347 (2020).

18. Zaverkin, V., Holzmüller, D., Steinwart, I. & Kästner, J. Fast and Sample-Efficient Interatomic Neural Network Potentials for Molecules and Materials Based on Gaussian Moments. *Journal of Chemical Theory and Computation* **17,** 6658–6670. ISSN: 1549-9626. http://dx.doi.org/10.1021/acs.jctc.1c00527 (2021).

19. Grabow, L. C. & Mavrikakis, M. Mechanism of Methanol Synthesis on Cu through CO_2 and CO Hydrogenation. *ACS Catalysis* **1,** 365–384. ISSN: 2155-5435. http://dx.doi.org/10.1021/cs200055d (2011).

20. Frei, M. S. *et al.* Mechanism and microkinetics of methanol synthesis via CO_2 hydrogenation on indium oxide. *Journal of Catalysis* **361,** 313–321. ISSN: 0021-9517. http://dx.doi.org/10.1016/j.jcat.2018.03.014 (2018).

21. Kattel, S., Ramírez, P. J., Chen, J. G., Rodriguez, J. A. & Liu, P. Active sites for CO_2 hydrogenation to methanol on Cu/ZnO catalysts. *Science* **355,** 1296–1299. ISSN: 1095-9203. http://dx.doi.org/10.1126/science.aal3573 (2017).

22. Reichenbach, T. *et al.* Ab initio study of CO_2 hydrogenation mechanisms on inverse ZnO/Cu catalysts. *Journal of Catalysis* **360,** 168–174. ISSN: 0021-9517. http://dx.doi.org/10.1016/j.jcat.2018.01.035 (2018).

23. Cheula, R., Tran, T. A. M. Q. & Andersen, M. Unraveling the Effect of Dopants in Zirconia-Based Catalysts for CO_2 Hydrogenation to Methanol. *ACS Catalysis* **14,** 13126–13135. ISSN: 2155-5435. http://dx.doi.org/10.1021/acscatal.4c03206 (2024).

24. Prabhakar Reddy, K. *et al.* Insights into the Surface Electronic Structure and Catalytic Activity of $InO_x/Au(111)$ Inverse Catalysts for CO_2 Hydrogenation to Methanol. *ACS Catalysis* **14,** 17148–17158. ISSN: 2155-5435. http://dx.doi.org/10.1021/acscatal.4c05837 (2024).

25. Mortensen, J. J., Hansen, L. B. & Jacobsen, K. W. Real-space grid implementation of the projector augmented wave method. *Physical Review B* **71,** 035109. ISSN: 1550-235X. http://dx.doi.org/10.1103/physrevb.71.035109 (2005).

26. Enkovaara, J. *et al.* Electronic structure calculations with GPAW: a real-space implementation of the projector augmented-wave method. *Journal of Physics: Condensed Matter* **22,** 253202. ISSN: 1361-648X. http://dx.doi.org/10.1088/0953-8984/22/25/253202 (2010).

27. Mortensen, J. J. *et al.* GPAW: An open Python package for electronic structure calculations. *The Journal of Chemical Physics* **160,** 092503. ISSN: 1089-7690. http://dx.doi.org/10.1063/5.0182685 (2024).

28. Blöchl, P. E. Projector augmented-wave method. *Physical Review B* **50,** 17953–17979. ISSN: 1095-3795. http://dx.doi.org/10.1103/physrevb.50.17953 (1994).

29. Hjorth Larsen, A. *et al.* The atomic simulation environment—a Python library for working with atoms. *Journal of Physics: Condensed Matter* **29,** 273002. ISSN: 1361-648X. http://dx.doi.org/10.1088/1361-648x/aa680e (2017).

30. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Physical Review Letters* **77,** 3865–3868. ISSN: 1079-7114. http://dx.doi.org/10.1103/physrevlett.77.3865 (1996).

31. Caldeweyher, E., Bannwarth, C. & Grimme, S. Extension of the D3 dispersion coefficient model. *The Journal of Chemical Physics* **147,** 034112. ISSN: 1089-7690. http://dx.doi.org/10.1063/1.4993215 (2017).

32. Caldeweyher, E. *et al.* A generally applicable atomic-charge dependent London dispersion correction. *The Journal of Chemical Physics* **150,** 154122. ISSN: 1089-7690. http://dx.doi.org/10.1063/1.5090222 (2019).

33. Caldeweyher, E., Mewes, J.-M., Ehlert, S. & Grimme, S. Extension and evaluation of the D4 London-dispersion model for periodic systems. *Physical Chemistry Chemical Physics* **22,** 8499–8512. ISSN: 1463-9084. http://dx.doi.org/10.1039/d0cp00502a (2020).

34. Monkhorst, H. J. & Pack, J. D. Special points for Brillouin-zone integrations. *Physical Review B* **13,** 5188–5192. ISSN: 0556-2805. http://dx.doi.org/10.1103/physrevb.13.5188 (1976).

35. Schäfer, M. R., Segreto, N. & Zills, F. *apax-hub/apax: v0.7.0* 2024. https://zenodo.org/records/14002846.

36. Behler, J. & Parrinello, M. Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces. *Physical Review Letters* **98,** 146401. ISSN: 1079-7114. http://dx.doi.org/10.1103/PhysRevLett.98.146401 (2007).

37. Kellner, M. & Ceriotti, M. Uncertainty quantification by direct propagation of shallow ensembles. *Machine Learning: Science and Technology* **5,** 035006. ISSN: 2632-2153. http://dx.doi.org/10.1088/2632-2153/ad594a (2024).

38. Akiba, T., Sano, S., Yanase, T., Ohta, T. & Koyama, M. *Optuna: A Next-generation Hyperparameter Optimization Framework* in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (ACM, 2019), 2623–2631. http://dx.doi.org/10.1145/3292500.3330701.

39. Nielsen, M. J., Kempen, L. H. E. & Andersen, M. Interpretable machine learned predictions of adsorption energies at the metal–oxide interface. In preparation.

40. Smidstrup, S., Pedersen, A., Stokbro, K. & Jónsson, H. Improved initial guess for minimum energy path calculations. *The Journal of Chemical Physics* **140,** 214106. ISSN: 1089-7690. http://dx.doi.org/10.1063/1.4878664 (2014).

41. Plessow, P. N. Efficient Transition State Optimization of Periodic Structures through Automated Relaxed Potential Energy Surface Scans. *Journal of Chemical Theory and Computation* **14,** 981–990. ISSN: 1549-9626. http://dx.doi.org/10.1021/acs.jctc.7b01070 (2018).

42. Makri, S., Ortner, C. & Kermode, J. R. A preconditioning scheme for minimum energy path finding methods. *The Journal of Chemical Physics* **150,** 094109. ISSN: 1089-7690. http://dx.doi.org/10.1063/1.5064465 (2019).

43. 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* **113,** 9901–9904. ISSN: 1089-7690. http://dx.doi.org/10.1063/1.1329672 (2000).

44. Cheula, R., Andersen, M. & Kitchin, J. Transition state search with fine-tuned machine learning potentials. In preparation.

45. Zills, F. *et al.* Collaboration on Machine-Learned Potentials with IPSuite: A Modular Framework for Learning-on-the-Fly. *The Journal of Physical Chemistry B* **128,** 3662–3676. ISSN: 1520-5207. http://dx.doi.org/10.1021/acs.jpcb.3c07187 (2024).

46. Zills, F., Schäfer, M., Tovey, S., Kästner, J. & Holm, C. *ZnTrack – Data as Code* 2024. arXiv: 2401.10603 [cs.SE]. https://arxiv.org/abs/2401.10603.

47. Zaverkin, V., Holzmüller, D., Steinwart, I. & Kästner, J. Exploring chemical and conformational spaces by batch mode deep active learning. *Digital Discovery* **1,** 605–620. ISSN: 2635-098X. http://dx.doi.org/10.1039/d2dd00034b (2022).

48. Wellendorff, J. *et al.* A benchmark database for adsorption bond energies to transition metal surfaces and comparison to selected DFT functionals. *Surface Science* **640,** 36–44. ISSN: 0039-6028. http://dx.doi.org/10.1016/j.susc.2015.03.023 (2015).

49. Medford, A. J. *et al.* From the Sabatier principle to a predictive theory of transition-metal heterogeneous catalysis. *Journal of Catalysis* **328,** 36–42. ISSN: 0021-9517. http://dx.doi.org/10.1016/j.jcat.2014.12.033 (2015).

Supporting Information: "Breaking scaling relations with inverse catalysts: a machine learning exploration of trends in CO_2 hydrogenation energy barriers"

Luuk H. E. Kempen[1], Marius Juul Nielsen[1], and Mie Andersen[1,2,*]

[1]Center for Interstellar Catalysis, Department of Physics and Astronomy, Aarhus University, Aarhus C, Denmark
[2]Aarhus Institute of Advanced Studies, Aarhus University, Aarhus C, Denmark
[*]Corresponding author: `mie@phys.au.dk`

S1 Model training

Table S1: GM-NN model hyperparameters used for the models in this work. If no value is listed in the 'shallow ensemble' column, the value is the same as for the single network.

Hyperparameter	Value	
	Single network	Shallow ensemble
Radial basis function	Gaussian	
Number of Gaussians	13	
Basis function $R_{\min}$	$0.72\,\text{Å}$	
Basis function $R_{\max}$	$7.0\,\text{Å}$	
Number of contractions	7	
Number of basis functions	11	
Empirical correction	exponential	
Empirical correction $R_{\max}$	$1.22\,\text{Å}$	
Number of hidden layers	4	
Number of nodes in hidden layers	51, 71, 82, 35	
Embedding initialization	uniform in $[-1, 1]$	
Weight initialization	LeCun [1]	
Bias initialization	0	
Optimizer	AdamW [2]	
Embedding contraction coefficient learning rate	2.41×10^{-3}	2.41×10^{-5}
Neural network parameter learning rate	1.17×10^{-3}	1.17×10^{-5}
Elemental output scaling factor learning rate	1.03×10^{-3}	1.03×10^{-5}
Elemental output shift learning rate	2.87×10^{-3}	3.87×10^{-5}
Empirical correction scaling factor learning rate	0.507	0.01
Empirical correction prefactor learning rate	0.444	0.01
Per-element gradient clipping	5.53	
Weight decay	3.91×10^{-4}	
Learning rate schedule	linear	

Table S2: Model prediction RMSEs for all 10 models on the data generated for model training.

Model index	Energies (meV atom^{-1})			Force components (meV Å^{-1})		
	Training	Validation	Testing	Training	Validation	Testing
1	0.30	0.36	0.43	30	58	62
2	0.90	0.87	0.91	80	92	91
3	0.31	0.29	0.45	43	61	58
4	0.76	0.75	0.75	83	94	98
5	0.42	0.38	0.40	31	62	62
6	0.26	0.31	0.31	19	57	64
7	0.43	0.38	0.51	30	56	56
8	0.53	0.61	0.59	41	66	66
9	0.46	0.49	0.50	36	64	65
10	0.45	0.48	0.51	28	56	68
Average	0.48	0.49	0.54	42	67	69
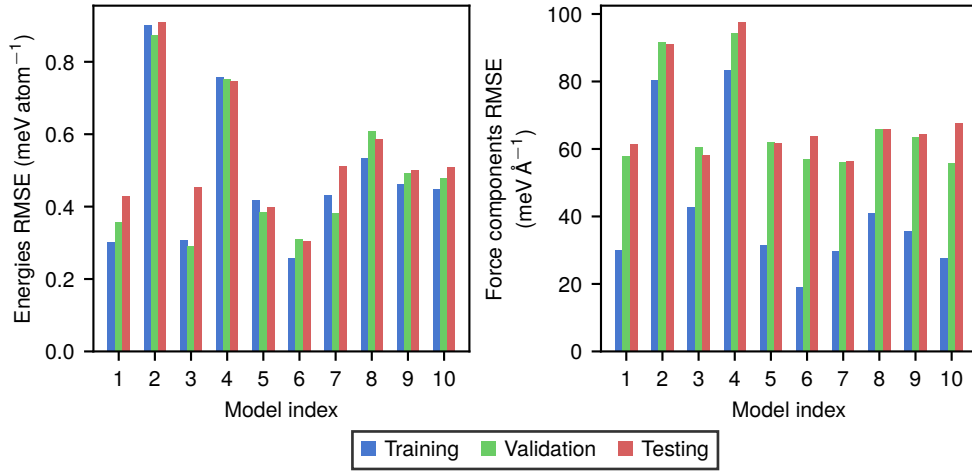
Figure S1: Model prediction RMSEs for all 10 models on the data generated for model training.
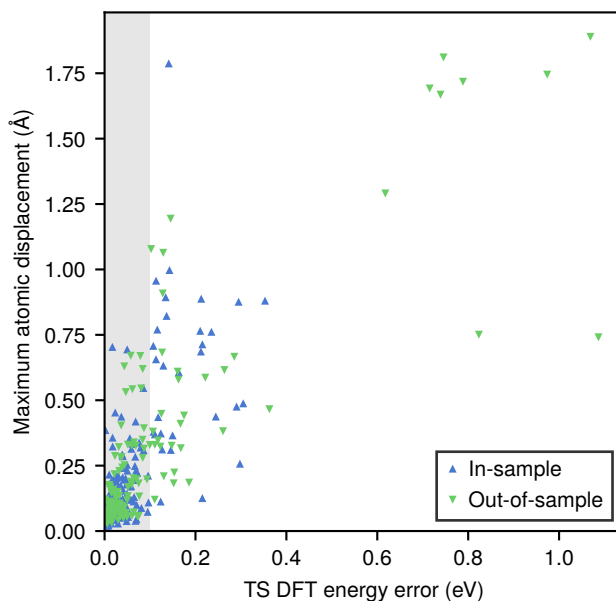
S2 Transition state search

Figure S2: Correlation between DFT energy differences of TS guesses produced by the GM-NN models to those of validated TSs obtained by continuing bond climbing with DFT and the maximum displacement of reactant atoms as a result of continued bond climbing. The highlighted band indicates an error below $0.1\,\mathrm{eV}$.
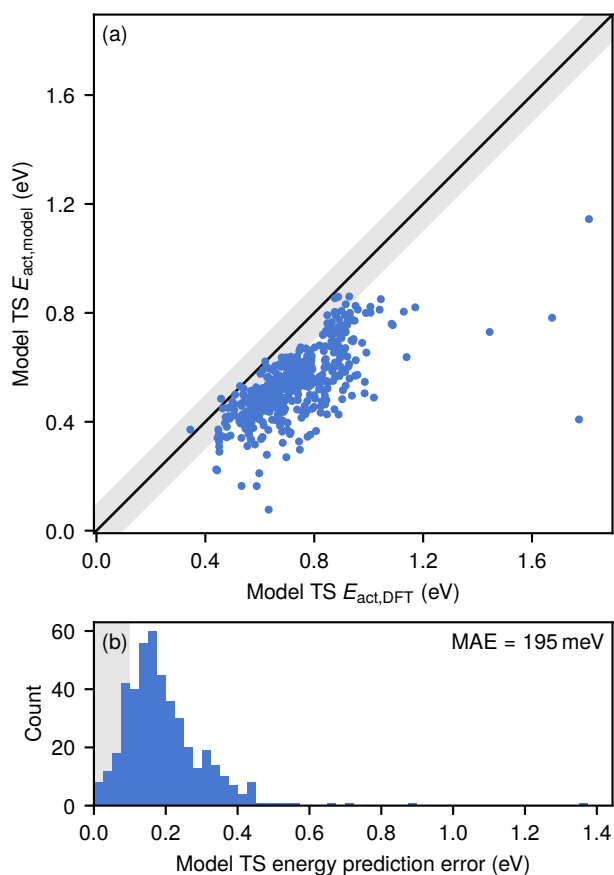

Figure S3: Comparison of model-predicted energies of in-sample TS guesses produced by the GM-NN models to DFT-evaluated energies of these geometries. (a) Parity plot of activation energies. (b) Histogram of prediction errors. The highlighted band in both plots indicates an error below $0.1\,\mathrm{eV}$.
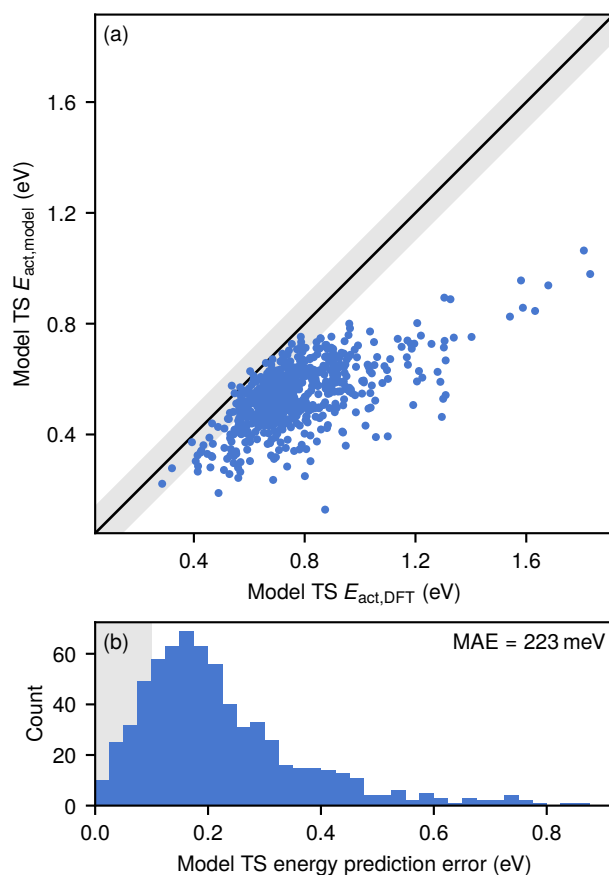
Figure S4: Comparison of model-predicted energies of out-of-sample TS guesses produced by the GM-NN models to DFT-evaluated energies of these geometries. (a) Parity plot of activation energies. (b) Histogram of prediction errors. The highlighted band in both plots indicates an error below $0.1\,\mathrm{eV}$.

S3   Active learning

Figure S5 shows the prediction performance of the GM-NN models as data points sampled from NEB TS searches were added to the training set, as described in Section 2.5. Noteworthy is that these data points are far enough removed from the data present in the initial training set for them to be poorly predicted at first, even after retraining the shallow ensemble model. This is evidenced by their significantly higher RMSE values at iterations 1 and 2 of the model, compared to those of the initial sets. At iteration 3, full retraining of the model from scratch did successfully incorporate this new class of data points into the model, and subsequent data iteratively added continued to be predicted with similar quality.
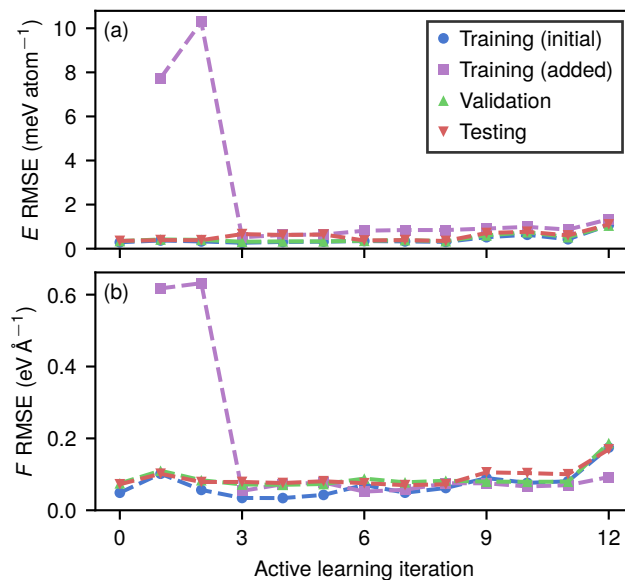

Figure S5: GM-NN model prediction performance in the active learning workflow. (a) Energies; (b) forces on the nanocluster and formate atoms.

More importantly, the active learning framework did not lead to any improvement in the quality of the guess TS structures. In Figure S6, model-generated TS guesses are compared with DFT-validated ones, and the average energy error stays relatively constant throughout active learning iterations. Continuing bond climbing with DFT also does not become notably faster throughout the iterations. Similarly, direct prediction of activation energies by the GM-NN also does not greatly improve during active learning, as shown in Figure S7. The reason for the ineffectiveness of the employed active learning strategy is currently unclear and requires further investigation.
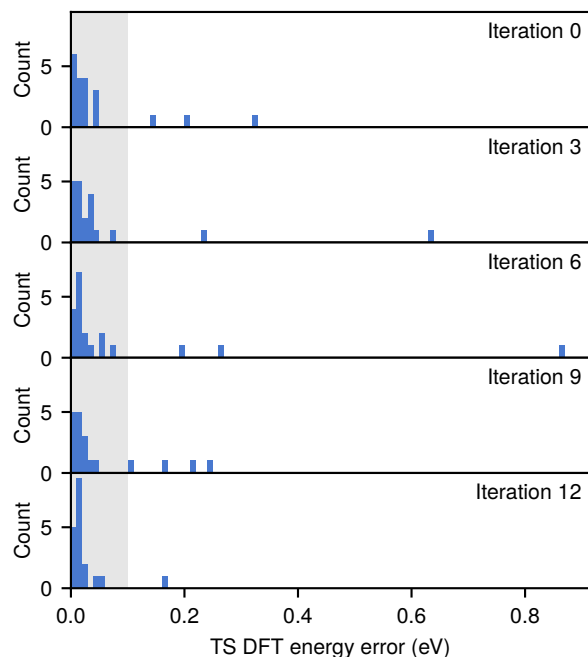
Figure S6: Comparison of energies of TS guesses produced by the iteratively trained GM-NN models to those of validated TSs by continuing bond climbing with DFT. The highlighted band indicates an error below 0.1 eV.
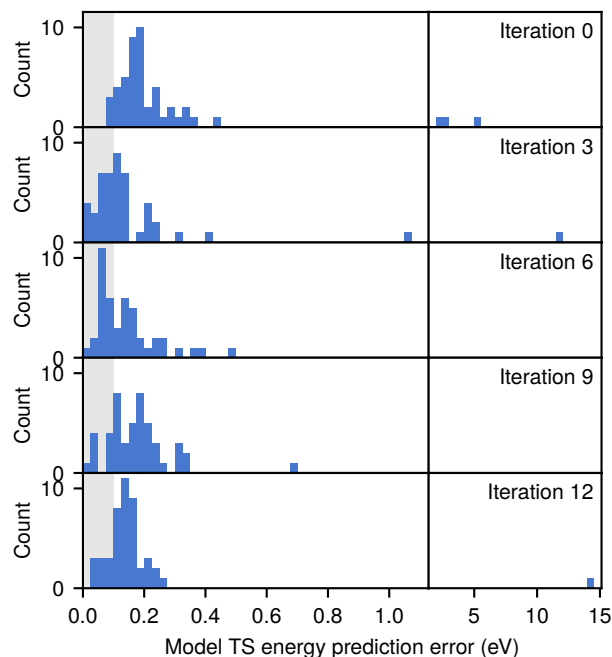
Figure S7: Comparison of model-predicted energies of TS guesses produced by the iteratively trained GM-NN models in the active learning workflow to DFT-evaluated energies of the same geometries. The highlighted band indicates an error below 0.1 eV.

References

1. LeCun, Y., Bottou, L., Orr, G. B. & Müller, K.-R. in *Neural Networks: Tricks of the Trade* (eds Orr, G. B. & Müller, K.-R.) 9–50 (Springer Berlin Heidelberg, 1998). ISBN: 9783540494300. `http://dx.doi.org/10.1007/3-540-49430-8_2`.

2. Loschchilov, I. & Hutter, F. *Decoupled Weight Decay Regularization* 2019. arXiv: 1711.05101 [cs.LG]. `https://arxiv.org/abs/1711.05101`.