Quantifying the Limits of Methane Activation in Cu-exchanged Zeolites using Reactive and Interpretable Machine Learning based Potentials

Jiawei Guo[1], Tyler Sours[1], Sam Holton[1], Chenghan Sun[1], Ambarish R. Kulkarni[1*]

[1]Department of Chemical Engineering, University of California, Davis, CA 95616, United States

*Correspondence and requests for materials should be addressed to A.R.K. (email: arkulkarni@ucdavis.edu)

Abstract

Natural gas remains an essential energy source for the industrial and residential sectors. However, selective valorization of methane (the main component of natural gas) into more mobile liquid energy carriers such as methanol remains challenging. Inspired by pMMO enzymes, many recent studies have examined Cu-exchanged zeolites as promising catalysts, specifically through $[CuOCu]^{2+}$ sites. These efforts, in part, have been motivated by the possibility of finding an elusive "Goldilocks" active site or topology that can outperform known catalysts while also maintaining selectivity towards methanol. As large-scale experiments with 1000s of material variations are impossible, theory will likely play an important role. Although computational screening studies are now routine for metals and alloys, similar studies for zeolites are not as straightforward due to the diversity of local chemical environments, and the aforementioned studies are not trivial using the traditional density functional theory (DFT)-based approach. Therefore, the overarching goal of this study is to leverage large-scale DFT calculations to develop a reactive machine learning based potential (rMLP) capable of systematically sampling the stability and reactivity of all $[CuOCu]^{2+}$ sites within a representative set of zeolites. Specifically, using methane activation as a prototypical example of an industrially relevant zeolite-catalyzed reaction, we have developed a

novel multistage active learning algorithm that preferentially samples the potential energy surface of the system near transition state of methane activation. We show that the resulting rMLP replaces the expensive DFT-based NEB calculations without any appreciable loss in accuracy (within 0.07 eV of the DFT computed energy barriers) – we evaluate C-H bond activation energies for 5,400 distinct sites across 52 zeolites and obtain 3,356 valid sites suitable for methane activation. By replacing the expensive DFT-based NEB calculations with rMLPs, we now report an exhaustive high-throughput screening study of thousands of $[CuOCu]^{2+}$ sites in zeolites, comparing the maximum rates of methane activation across 52 zeolite topologies and more than 3,000 sites. To the best of our knowledge, this work represents the first example of using reactive MLPs to identify the transition state geometries and screen the catalytic performance of thousands of zeolite-based active sites at DFT accuracies.

Introduction

Natural gas continues to be an important source of energy for both the industrial and residential sectors. While current approaches for transporting natural gas rely on transcontinental pipelines or specialized liquified natural gas (LNG) ships, these capital-intensive technologies are unsuitable for geographically dispersed sources such as shale gas oil fields[1]. In such cases, natural gas, which consists mainly of methane, is flared and converted to CO_2[1–7]. To mitigate the negative environmental impact of such wasteful approaches, several technologies have been explored for the valorization of methane into other liquid energy carriers. One possible solution is to directly convert methane into methanol[8–20] that is easier to transport and convert into other value-added products[21–25]. However, selective partial oxidation of methane is thermodynamically challenging due to the high symmetry and stability of the molecule[26].

Research related to partial oxidation of methane, particularly in the C-H bond activation, has considered many possible catalysts, such as zeolites[27–32], metal-organic frameworks (MOF)[33–36], metal oxides[37–39], and homogeneous[40–42] systems. Industrially relevant materials such as zeolites are particularly well-studied due to their favorable properties such as high thermal stability, chemical resistance, and relatively low cost. Inspired by the pMMO enzymes,[43–45] which oxidize methane under ambient conditions, many recent studies have examined Cu-exchanged zeolites as promising catalysts for this reaction[46,47]. Di-copper-oxo sites (i.e., $[CuOCu]^{2+}$) are among the most exciting motifs as there is good theoretical and experimental evidence for their excellent performance[48–53]. The use of zeolites for methane activation has been reviewed previously[54-56.]

Since the 1990s, several distinct zeolite frameworks have been experimentally and computationally studied for this reaction[47, 57-60]. Although significant attention has focused on industrially popular zeolites such as CHA, MOR, MFI, etc., recent studies have explored other frameworks such as BPH, EON, MEI, and HEU[61]. These efforts, in part, have been motivated by the possibility of finding an elusive "Goldilocks" active site or topology that can outperform known catalysts while also maintaining selectivity towards methanol[61–68]. For example, recent results from van Bokhoven and Sushkevich suggest that certain types of $[CuOCu]^{2+}$ sites (with 2.9 Å Cu--Cu separations) are more selective than others in MFI[69]. These insights are based on detailed kinetic studies and characterization experiments including UV/Vis, XAS, FTIR and etc.; the authors propose that a lower Cu/Al ratio (< 0.3, for CuMFI) facilitates the formation of more selective active sites over a wide temperature range. The question arises if other zeolite materials (topology, Si/Al ratio, Al distribution, etc.) could be designed to preferentially stabilize selective $[CuOCu]^{2+}$ sites. As large-scale experiments with 1000s of material variations are not possible,

theory is likely to play an important role in developing the next generation of intrinsically selective zeolite catalysts for methane activation.

Although computational screening studies are now routine for metals and alloys[70-73], similar studies for zeolites are not as straightforward due to the diversity of local chemical environments. For instance, while sampling high symmetry sites (e.g., ontop, bridge, fcc, hcp) is often sufficient to study the catalytic properties of metals and alloys, even a simple CHA unit cell consists of 31 unique $[CuOCu]^{2+}$ configurations. Exhaustive sampling of the entire International Zeolite Association (IZA) [74] database (including 231 uninterrupted zeolites where the T-sites are all tetrahedrally connected) will require investigating the performance of 15,255 unique $[CuOCu]^{2+}$ sites (obtained using the MAZE software)[75] with varying Cu-Cu distances (or Cu-O-Cu angles), Al-Al distances, and local confinement effects (**Fig. S3, S5, S6, Table S1**).

The aforementioned studies are not trivial using the traditional density functional theory (DFT)-based approach for several reasons. These include: (1) a lack of a centralized toolkit for the generation of active site structures, (2) the high computational cost of DFT-based nudged elastic band (NEB) calculations, and (3) challenges associated with using known descriptors (e.g., hydrogen binding energy) in situations where confinement effects may be important. Thus, it is not surprising that attempts to systematically sample all relevant active sites within a zeolite have been limited to topologies with small primitive unit cells.[76]

Another major factor that limits the efficacy of DFT-guided zeolite design is the lack of control over Al distributions during zeolite synthesis.[77] The Al distribution (which depends on the synthesis conditions) affects the Cu-speciation during the activation step and is thus correlated with the overall selectivity of the reaction. As precise characterization and identification of the active sites during the reaction remains challenging, these efforts would benefit from an

exploration of the catalytic performance of all possible Cu sites (and associated pairwise Al distributions) that can be formed within different zeolite topologies.[78]

In the context of the above discussion, the overarching goal of this study is to leverage large scale DFT calculations to develop a reactive machine learning based potential (rMLP) capable of systematically sampling the stability and reactivity of all $[CuOCu]^{2+}$ sites within a representative set of zeolites. Specifically, using methane activation as a prototypical example of an industrially relevant zeolite-catalyzed reaction, we have developed a novel multistage active learning algorithm that preferentially samples the potential energy surface of the system near transition state of methane activation. The resulting dataset, denoted as rCuZEO23 (detailed in the **SI, Section S2**), is iteratively used to train a deep neural network-based potential.[79] We show that the resulting rMLP replaces the expensive DFT-based NEB calculations without any appreciable loss in accuracy – we evaluate C-H bond activation energies for 5,400 distinct sites across 52 zeolites within a couple minutes (per calculation) and obtain 3,356 valid sites suitable for methane activation. Our predictions are within 0.07 eV of the DFT computed energy barriers.

As discussed later, our rMLP-based NEB calculations suggest that caution must be used while using simple adsorption energy-based descriptors[27] to predict reaction barriers in zeolites. More importantly, by replacing the expensive DFT-based NEB calculations with rMLPs, we now report an exhaustive high-throughput screening study of thousands of $[CuOCu]^{2+}$ sites in zeolites. Specifically, by comparing the maximum rates of methane activation across 52 zeolite topologies and more than 3,000 sites, we identify several less explored materials (e.g., MEI, CGF and USI) as promising candidates for C-H activation. These findings lay a foundation to develop a theory-guided approach for the experimental design of new zeolites for methane activation.

From a computational perspective, we recognize that force field-based methods (often accelerated by ML) are now routinely used to predict adsorption isotherms and diffusion properties across large libraries of zeolites.[79-81] Similarly, the computational catalysis community has reported several ML models to predict adsorption energy-based descriptors to approximate the kinetics of surface-mediated reactions.[72,73] However, to the best of our knowledge, this work represents the first example of using reactive MLPs to identify the transition state geometries and screen the catalytic performance of thousands of zeolite-based active sites at DFT accuracies.

Although we focus on a specific reaction (i.e., C-H bond activation) and one class of active sites (i.e., $[CuOCu]^{2+}$), we anticipate that the multistage workflow developed here will be broadly applicable to other zeolite-catalyzed reactions (e.g., NO_x chemistry, MTO reactions, etc.) and eventually, to other heterogeneous catalysts. Complementary to the recently reported studies focused on modeling gas-phase reactions, this work highlights the growing relevance of MLPs in predicting the catalytic performance for heterogeneous materials. We believe that the rCuZEO23 database, which includes 164k DFT-calculated geometries near the transition state and 16k Cu-related species optimized fully from DFT, is an important step towards achieving this goal. The resulting rMLP (denoted as rCuZEO2023/DMD) is available for use by the community.

Results and discussions

1. Multistage active learning algorithm for rMLP training.

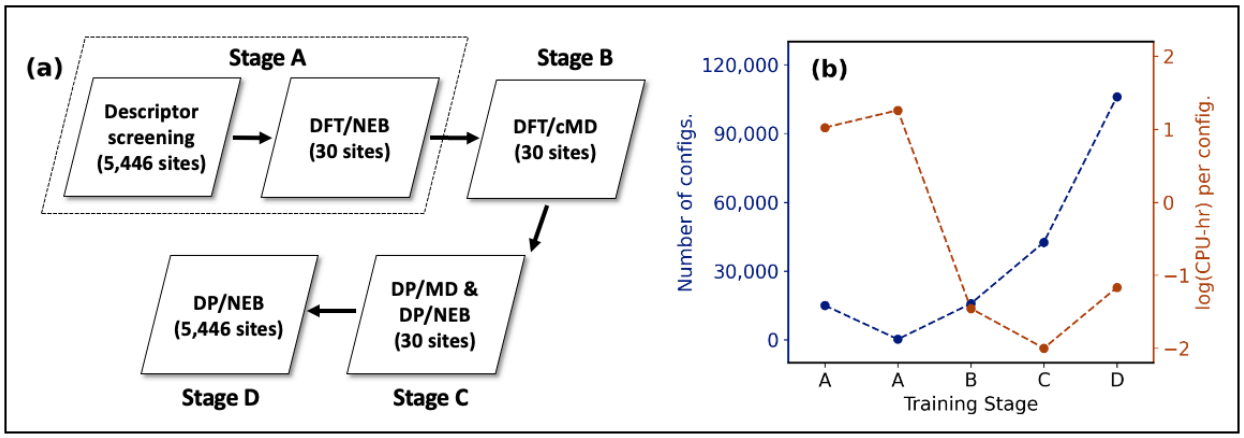

Figure 1. (a). Overview of the multistage active learning algorithm. **(b).** The number of configurations included at each training stage vs. computational cost associated with sampling each configuration.

We begin the discussion by outlining the multistage active learning algorithm. As discussed above, the overarching goal of this work is to predict the performance of zeolite catalysts for methane activation by explicitly calculating the C-H activation barriers for all possible $[CuOCu]^{2+}$ active sites that can be formed within each topology (**Fig. S8**). We choose to explore 52 zeolite topologies obtained from the IZA database (detailed in **Table S1**). Our analysis includes several experimentally well-studied materials (e.g., CHA, MOR, MFI, etc.) (**Fig. S4** showing the pie chart) and additional topologies chosen to efficiently and comprehensively explore the possible local confinements of the $[CuOCu]^{2+}$ sites. As described in **Fig. S5**, we choose zeolite topologies with the highest volumetric density of $[CuOCu]^{2+}$ sites ρ_{sites}; ρ_{sites} is defined as the ratio of the number of unique Al pairs that can be formed within 9Å separations (n_{sites}) and the volume of the primitive cell (V_{pc}). Taken together, we consider 3,912 1Al (i.e., Cu^{+}, $[CuOH]^{+}$) and 5,575 2Al sites (i.e., CuOCu, CuOHCu). More details are presented in the **SI**, **Fig. S8-S9**.

Following Latimer and others[27], the overall rate of methane activation depends on both the stability (quantified by active site formation energy, Eq (1)) and reactivity (i.e., activation energy, Eq (2)) of the $[CuOCu]^{2+}$ site.

$$E_f = E_{CuOCu} + E_{silica} - (E_{Z_1Cu} + E_{Z_2Cu} + \frac{1}{2}E_{O_2}) \qquad \text{Eq (1)}$$

$$E_{a,scaling} = 0.75E_h + 1.09 \qquad \text{Eq (2)}$$

The inverse correlation between these two quantities results in a volcano relationship between the turnover frequency (i.e., rate of C-H activation per site, Eq (3)) and the site's stability.

$$\text{TOF} = \frac{\sqrt{P_{O_2}}e^{-G_f/k_BT}}{1+\sqrt{P_{O_2}}e^{-G_f/k_BT}}\frac{k_BT}{h}e^{-G_a/k_BT} \qquad \text{Eq (3)}$$

Previous work has demonstrated that the hydrogen binding energy (calculated by Eq 4) serves as a useful descriptor to predict the methane activation energy.

$$E_h = E_{CuOHCu} - E_{CuOCu} - \frac{1}{2}E_{H_2O} + \frac{1}{4}E_{O_2} \qquad \text{Eq (4)}$$

More detailed descriptions can be found in the **SI, Section 2.4.**

While traditional DFT geometry optimizations can be (and were) used to calculate the formation energies of all $[CuOCu]^{2+}$ sites, brute-force calculation of the energy barriers using DFT-based NEB was not computationally feasible. Our strategy to overcome this limitation is to develop an MLP capable of faithfully reproducing the DFT potential energy surface (PES) along and near the reaction coordinate. As illustrated in **Fig. 1** and **2**, this requires obtaining sufficient DFT training data from the relevant regions of the PES. Although MLPs trained on molecular dynamics have been widely used to predict equilibrium properties and obtain geometry optimized structures[79], the development of reactive MLPs (denoted as rMLPs) for describing chemical reactions within configurationally complex materials (such as zeolites) has not yet been reported.

Thus, we introduce a novel protocol to develop rMLPs in zeolites. As detailed in **Fig. 2a-2d**, we use a multistage active learning approach that identifies promising $[CuOCu]^{2+}$ sites efficiently (i.e., using as few DFT calculations as possible) and accurately (i.e., by minimizing errors compared to DFT). While our workflow is demonstrated for the C-H bond activation reaction, we anticipate that the protocol illustrated in **Fig. 1a** can be generalized to other reactions.

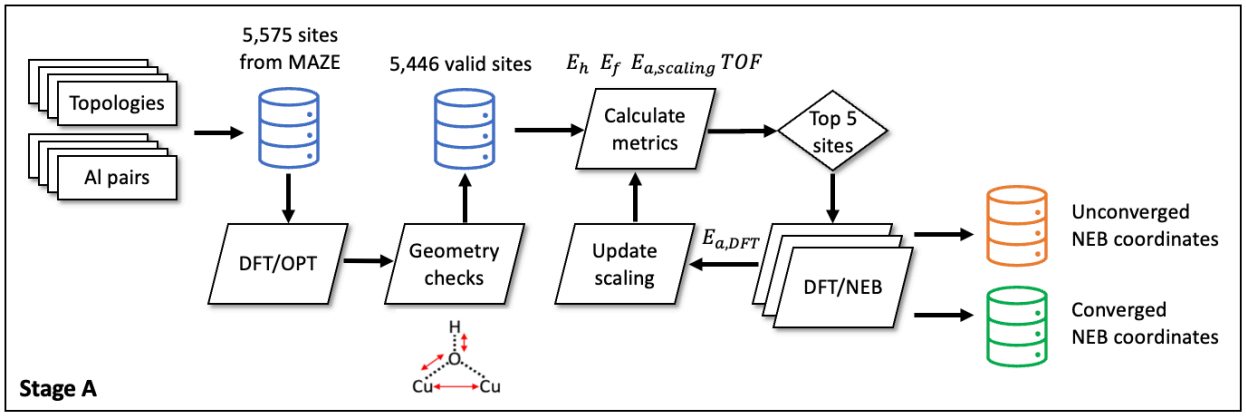

Figure 2(a). flowchart for training stage **A:** descriptor-based initial site sampling (30 sites from 27 topologies) and DFT/NEB calculations

Specifically, we use a multistage approach to obtain the relevant DFT data near the transition state region. In Stage **A (Fig. 2a)**, we use a DFT calculated energies of the various Cu species (obtained using brute force DFT calculations) to identify the initial set of promising sites for C-H activation. A ranking system that uses four metrics (i.e., E_f, E_h, $E_{a,scaling}$, and TOF) is used to select the top 5 sites from an initial set of 5,446 configurations. Then we perform traditional DFT/NEB calculations to obtain the true methane activation energy ($E_{a,DFT}$). The true DFT activation energy is used to develop a new scaling relationship, which is then used to re-evaluate the $E_{a,scaling}$ of all 5,446 sites. The new site ranking (since $E_{a,scaling}$ is updated) is used to identify

the next set of top 5 sites and DFT/NEB. This process is repeated 6 times until a set of 30 DFT-calculated reaction coordinates, transition state geometries and activation energies are identified. A fraction of unconverged reaction coordinates (~150 configurations per site), obtained during the NEB convergence steps, are saved for model development in Stage **B**.

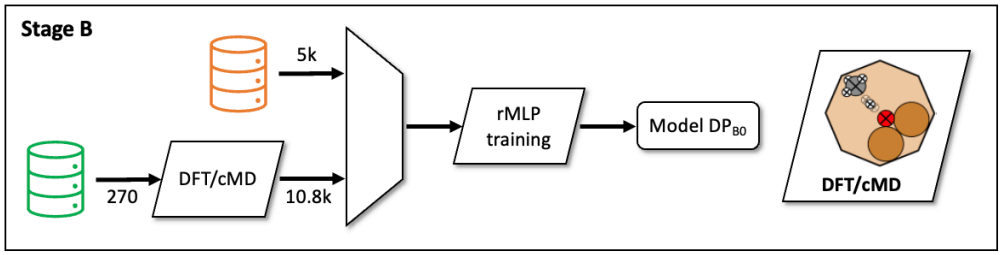

Figure 2(b). flowchart for training Stage **B**: initial rMLP model training using DFT/NEB and DFT/cMD sampled configurations.

At Stage **B** (**Fig. 2b**), the set of 30 promising $[CuOCu]^{2+}$ sites identified at Stage **A** (across 27 different topologies) are used to generate the necessary training data for rMLP development. Specifically, a set of 270 configurations obtained from the 30 converged NEB calculations (each NEB used 5 intermediate images) are used to run constrained MD simulations (0.5 fs time step, 200 fs total simulation time steps, 298 K). Here, we freeze the positions of the CH_4 (or $CH_3 - H$) and the oxygen of the $[CuOCu]^{2+}$ site; this ensures that the algorithm samples configurations in the neighborhood of the NEB images. A total of 10,800 configurations are obtained from these constrained MD. These are combined with the 5,000 configurations from unconverged DFT NEBs to train an initial rMLP using a deep neural network potential as implemented in DeepMD-kit. Details of the rMLP training are presented in the **SI, Section 4**. Beyond this stage, our algorithm only requires DFT-based single point energy calculations (denoted as DFT/SPE). No further

DFT/MDs or DFT/NEBs are used. We denote the model as DP_{B0}, where subscript indicates the zeroth iteration of the model obtained at Stage **B** of the overall protocol.

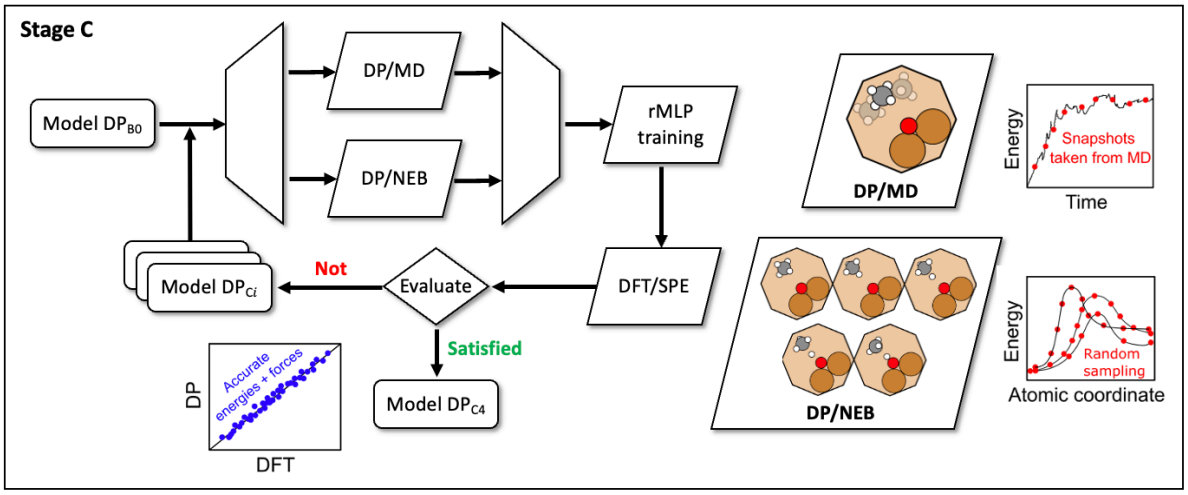

Figure 2(c). flowchart for training Stage **C**: rMLP model training for the 30 initial sites using configurations sampled from DP only (DP/MD & DP/NEB).

While Stage **B** used constrained DFT/MD and DFT/NEBs to generate the necessary training data, Stage **C (Fig. 2c)** uses the rMLP developed above (i.e., DP_{B0}) to obtain additional relevant configurations. Specifically, to ensure that the entire "length" of the reaction coordinate is well sampled, we use DP/MD and DP/NEB. These two approaches sample different regions of the PES. While DP/MD mainly samples the local minima of the PES (e.g., initial (CH_4) and final (CH_3—H) state of the reaction), the DP/NEBs sample configurations at various stages of the C-H bond breaking (and O-H bond formation) process.

The DP/MD is performed using LAMMPS using an approach similar to DP-GEN.[82] Briefly, an ensemble of four models is trained. Each model has an identical architecture but uses different seeds to initialize the neural net parameters. One of these four models is used to run a 50 ps NVT-MD (0.5 fs timestep), and the temperature ramped from 298 K to 500 K. This provides a

diverse set of configurations close to the initial (CH_4 + $[CuOCu]^{2+}$) and final state (CH_3 + $[CuOHCu]^{2+}$) of the system. The DP/NEB is performed using the Atomic Simulation Environment (ASE). Several strategies are used to induce randomness and to avoid biased sampling. For example, our DP/NEB iterations use different numbers of intermediate NEB images (ranging from 11 to 21) and total number of NEB iterations. Also, the initial and final states of the 30 DFT-derived NEB trajectories are "rattled" and then interpolated; this approach adds diversity to the initial guessed reaction coordinate used during DP/NEBs.

New training configurations sampled from both DP/MD and DP/NEB are selected by evaluating the uncertainty of the force predictions from the ensemble of models (denoted as ϵ_t, detailed in the **SI, Section 4.1, Fig. S11**). We perform DFT/SPE for configurations where the ϵ_t parameter lies between 0.05 and 0.4 eV/Å; this ensures that sampled configurations are physically relevant and are also sufficiently dissimilar from the existing data.[82] As ϵ_t measures the uncertainty of predictions between the four ensembles of models, this metric has been shown to be correlated with the errors between the DFT forces and the DP prediction.

The above two approaches are used iteratively to expand the training dataset while also improving the accuracy of the model. Across 4 iterations within Stage C, we added a total of 42,600 configurations – 36,400 near the initial and the final states (from DP/MD) and 6,100 sampled from the DP-NEBs (**Table S2**). The model refinement loop is stopped when the model-predicted energy and forces show near-DFT accuracy for the 30 sites identified in Stage **A**. The MSEs of the sequentially obtained models, which show systematic improvements, are presented in **Table S3**.

Although the training protocol may seem overly complicated, the actual implementation is relatively straightforward (as the individual steps are automated) and computationally efficient (as

the DFT/SPEs can be parallelized). We denote the resulting rMLP as DP_{C4}, which indicates that 4 iterations were performed at Stage **C** of the active learning process.

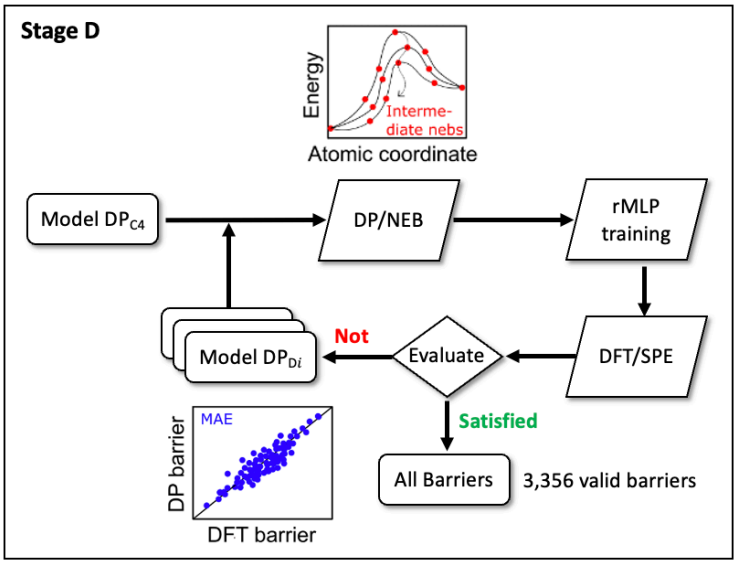

Figure 2(d). flowchart for training Stage **D**: extend the rMLP model training to all possible $[CuOCu]^{2+}$ sites using DP/NEB.

The final active learning step, Stage **D** shown in **Fig. 2d** is central to this study. In this training block, we expand the active site configuration space from the initial 30 promising sites to all possible 5,446 sites across 52 different topologies. Additionally, due to the training data obtained from DP/MD (Stage **C**), the model can now perform geometric optimization of the initial and final states for the previously unseen $[CuOCu]^{2+}$ sites. More importantly, the DP-optimized initial and final state geometries are sufficiently realistic to initiate a NEB calculation that relies entirely on the DP_{C4} PES. To address the increased diversity of the PES and achieve more efficient sampling considering the scope of the active site space (from 30 sites to 5,446), we adopt a new sampling scheme for the DP/NEBs. We extract unconverged reaction coordinates from different stages of the convergence (20%, 60%, and the converged coordinate). This methodology, used for

each active learning iteration in Stage **D**, requires one NEB run per site instead of multiple NEB runs per site (as in Stage **C**). In addition, we also include reaction coordinates at 40% and 80% from the final convergence into the training set for sites with more complicated PESs or where the initial guess transition state is further from the "true" transition state geometry, as these sites tend to require more NEB steps to converge. Similar to the strategy in Stage **C**, the new configurations are evaluated with DFT/SPE and selectively add into the training poll based on ϵ_t. The above approach is necessary since multiple DP/NEBs and DFT/SPEs (used for 30 sites in Stage **C**) are not feasible for 5,446 sites. Simply put, while Stage **C** repeatedly samples the DP/NEB converged reaction coordinate for 30 sites to improve the model (i.e., DP_{C1} to DP_{C4}), Stage **D** also samples the approach to the converged reaction coordinate for 5,446 sites.

In addition to comparing energies and forces obtained from the DP model to DFT, we also calculate the activation energy barrier of each $[CuOCu]^{2+}$ site using DFT. This is calculated using DFT/SPE of the initial and the transition state (i.e., $E_{TS} - E_{IS}$) (**Fig. 2c**), where the geometries are obtained from the current DP model. For block **D**, 106,000 new configurations are included in the training (**Table S2**). Details on the DeePMD-kit parameters used for training the final model, as well as test set energy and forces prediction errors for different hyperparameter combinations can be found in the **SI, Section 4.4, Table 4**.

Taken together, the entire model development procedure includes 180,700 DFT-calculated configurations obtained using a combination of scaling relationships, SPEs, MDs and NEBs. Specifically, the DFT PES (for 5,446 sites across 52 zeolites topologies) has been effectively sampled using surrogate models (DP_{B0}, $DP_{C1\text{-}C4}$, and $DP_{D1\text{-}D4}$) that show successive improvements – this has important implications related to the interpretability of the model (**Section 4**). The above multistage active learning algorithm overcomes the "chicken and egg" issue with developing

MLPs; sufficient number and diversity of relevant configurations near the transition state is required to train a rMLP capable of describing the bond breaking and bond formation steps.

2. Model performance evaluation and validation.

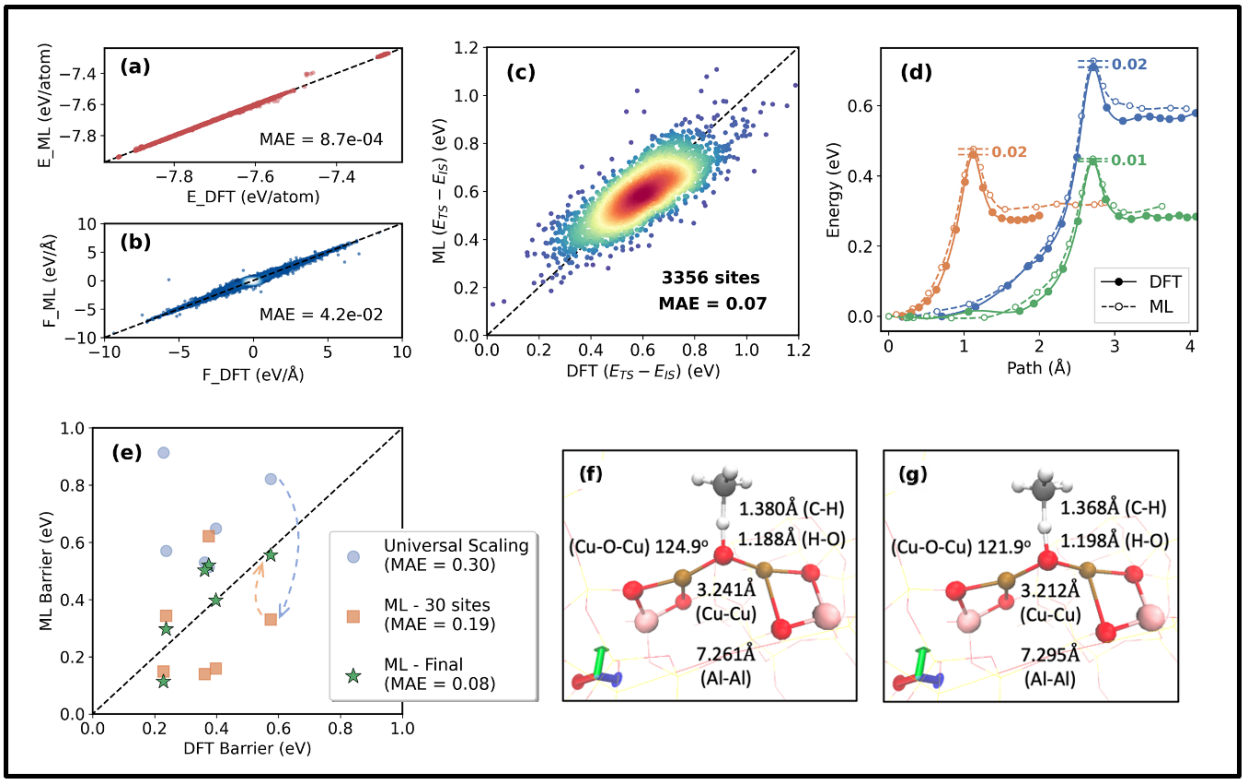

Figure 3. **(a-c).** Parity plots comparing ML-predicted energies per atom **(a)**, forces per dimensions **(b)**, and (E_TS – E_IS) **(c)**, with DFT-calculated values for datasets not included in the training pool. **(d).** ML-predicted minimum energy pathways (MEPs) vs. DFT-calculated MEPs (blue: HEU, green: MOR, orange: RUT) using 17 randomly generated images for both DFT and DP-NEBs including transition state. **(e).** Parity plot comparing the performance of universal scaling and the current DP model in predicting reaction barriers for six randomly selected validation sites from CHA, HEU, JSW, MFI, MOR, and RUT topologies, using 5 images for DFT NEBs and 17 images for DP NEBs. **(f-g).** Transition state geometries predicted by ML **(f)** vs. DFT **(g)**, taking CHA as an example. Coloring: Si on the background (yellow), Al (pink), Cu (brown), O (red), H (white), C (grey).

We show that the model trained on the multi-stage active learning protocols in **Fig. 1** and **2** demonstrates excellent agreement with DFT for predicting the energies and forces for configurations sampled near the reactive region of the potential energy surface. Specifically, we compare the model-predicted (DP_{D4}) per-atom energies and forces (x-, y-, and z-components) with

DFT-calculated values for a test configuration dataset. The test set is obtained from 10% of the overall configurations sampled from both the local minima and the reactive regions. The corresponding parity plots for energies and forces are shown in **Fig. 3a** and **b**, respectively, with an overall 0.87 meV/atom MAE for energies and an MAE of 42 meV/Å for forces considering all 52 different zeolite topologies.

As the single point energy and force study demonstrates the ability of the model to reproduce the DFT PES, we now show that the rMLP can be used to obtain the transition state for methane activation across 1000s of $[CuOCu]^{2+}$ sites. As the model was trained on the PES around the initial and final states of the reaction (from Stage **C**), the DP_{D4} model is used to optimize the initial and final states, and to initialize the NEB calculation. The DP_{D4} predicted barrier is compared with the DFT calculated value for 3,356 sites in **Fig. 3c** with an overall MAE of 0.07 eV. Model evaluations after each training iteration and the final MAE for all activation barriers within each topology can be found in the **SI, Fig. S12-S13**. We note that more than 2000 sites resulted in physically unrealistic geometries for the transition state search, and reasons will be detailed in **Section 3.3**.

The above calculations used DFT/SPE on the DP_{D4} optimized structures and reaction coordinates to calculate the C-H bond activation energies. A more stringent test of the model compares the entire reaction coordinate obtained using DFT and DP_{D4}. Specifically, instead of optimizing the initial and final states using DP_{D4} (in **Fig. 3c**), we perform DP_{D4}/NEB directly on DFT-optimized initial and final states. To ensure an unbiased comparison, we randomly select three $[CuOCu]^{2+}$ sites from three different topologies (**Fig. S14-S16**) and use identical stopping criteria (**SI, Section 3**) for both DFT/ and DP/NEB. As shown in **Fig. 3d** for one $[CuOCu]^{2+}$ site in MOR, almost identical minimum energy paths (MEPs) are obtained for DFT and DP_{D4}. Note

that sufficiently high density of intermediate images (> 9, **Fig. 3d** uses 17 images) is necessary to obtain this resolution (**Fig. S17**).

Encouraged by the accuracy (0.07 eV MAE, **Fig. 3c**) and computational efficiency (100x faster, **Fig. 1b** and **SI, Section 4.2**) of the DP_{D4} model, we now compare the DP_{D4}/NEB results with those obtained from the traditional descriptor approach. We randomly select six $[CuOCu]^{2+}$ sites from six different topologies (not part of the original 30) that were not included in the original 30 sites in Stage **C** of the model training. These are used to run DFT/NEB (with 5 images), DP_{C4}/NEB and DP_{D4}/NEB; the DP/NEBs are run with 17 intermediate images. **Fig. 3e** shows that the descriptor calculated value (using E_h) results in a significantly higher MAE (0.3 eV) compared to the perditions of the final model (0.08 eV MAE for DP_{D4}/NEB). Unsurprisingly, the intermediate model (DP_{C4}, trained on 30 sites) shows an MAE value between the descriptor estimate and the final DP model. These results suggest, as further discussed in **Section 3**, that caution must be used while using adsorption energy-based descriptors to predict reaction barriers in zeolites. In addition to predicting the reaction coordinates, we show that the transition state geometries (for CHA, used as an example) predicted by DP_{D4} (**Fig. 3f**) are nearly identical to DFT-calculated geometries in **Fig. 3g**. In all the above studies, vibrational analysis of the transition state is performed to ensure that only one non-trivial imaginary frequency is obtained.

3. Evaluating the relevant energetics through probability distributions.

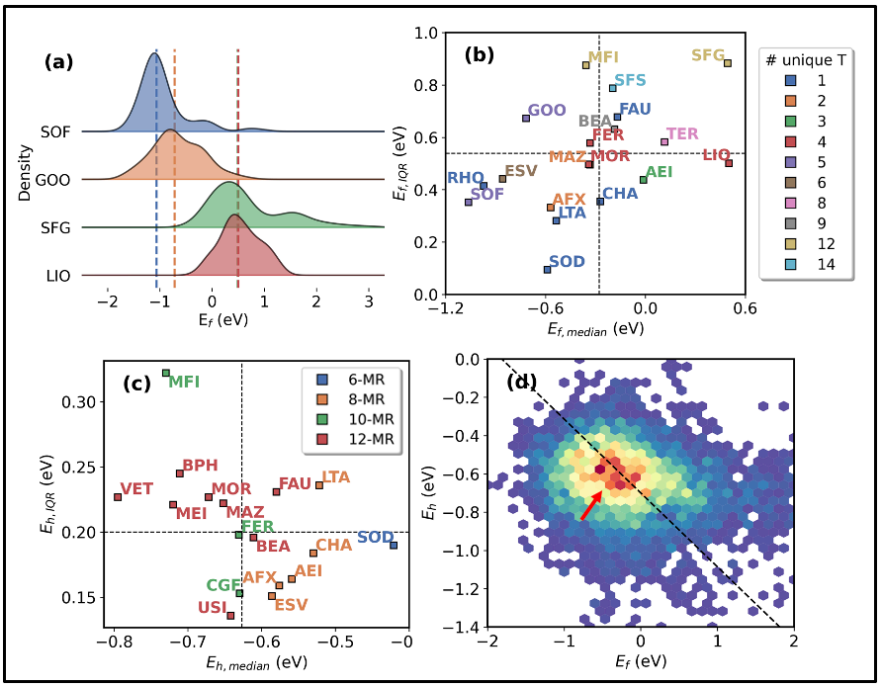

Figure 4 (a). E_f histograms of four zeolites with distinct shapes of distributions, including all valid $[CuOCu]^{2+}$ sites within each topology. Vertical dashed lines highlight the median of the distributions (SFG median overlaps with LIO). **(b).** Quadrant plot showing median and inter-quartile range (denoted as IQR) of the E_f distributions for $E_{f,median}$ below 0.6 eV, color-coded by the number of unique T-sites. Quadrants plot of E_h distributions in **(c)** is color-coded by maximum window sizes. Only the top 10 most studied and the few outlier topologies are highlighted. Quadrant plots for the rest of the topologies can be found in the **SI, Fig. S21-S25**. Statistics are summarized in **Table S5**, **Table S6**. The horizontal and vertical dashed lines in **(b)** and **(c)** indicates the average value of E_{median} and E_{IQR} of all topologies. **(d).** Probability distributions of E_h vs. E_f for all valid $[CuOCu]^{2+}$ sites (red: higher density, blue: lower density). Vertical dashed line indicates linear scaling extracted from previous work[27].

3.1. Impact of topology on the stability and reactivity of $[CuOCu]^{2+}$ sites

Having demonstrated that thousands of C-H activation energy barriers can be rapidly obtained using DP_{D4} with near-DFT accuracy, we are now one step closer to quantifying the limits of methane activation performance across different zeolite topologies. As the stability of the

$[CuOCu]^{2+}$ site is an important metric, we now focus on examining the formation energies (denoted as E_f) of all 5,446 geometrically possible sites in 52 zeolites. These data, plotted as histograms of E_f for SOF, GOO, SFG, and LIO, emphasize the role of the zeolite topology in controlling the stability of the $[CuOCu]^{2+}$ sites. For example, we observe that the SOF topology is more successful at stabilizing $[CuOCu]^{2+}$ sites than LIO. Although the E_f histogram for almost all topologies (**Fig. S18**) lies within -2.4 to 2.4 eV (calculated using Eq (1)), very few of them share identical shapes.

As the relationship between confinement, stability, and reactivity remains a core tenet of zeolite science, the rCuZEO23 database (including 16k DFT-optimized configurations) enables a quantitative exploration of this philosophy. For instance, we hypothesized that sites with similar formation energies would share similarities in the local bonding and confining environment. However, our analysis with simple structural descriptors (e.g., Al-Al distance, Al-pair separation, ring size, framework density, etc.) proved unsuccessful (importance factors ranking can be found in the **SI, Fig. S19**). While we recognize that the chosen zeolite dataset may induce bias (selected using the volumetric density of $[CuOCu]^{2+}$ sites), this problem may be well-suited for emerging ML algorithms.

Instead of searching for specific geometry-based descriptors to identify promising $[CuOCu]^{2+}$ sites, we propose an exhaustive ensemble-based approach that uses energy histograms to compare zeolite performance.[78] This is enabled by combing the MAZE codebase (which generates initial structures) with the accuracy and computational efficiency of the DP_{D4} ML model. Here, the DP_{D4} model (which is trained on the initial and the final states) serves as a valuable pre-optimizer that provides optimized geometries close to the DFT-based calculation.

In addition to differences in the "averaged" stability of the possible $[CuOCu]^{2+}$ sites (as measured by the median E_f, dotted vertical lines), **Fig. 4a** shows that the span of the E_f histograms also vary across the four topologies. This diversity in formation energies is quantified by the inter-quartile range (IQR). For example, the higher IQR for SFG (0.88 eV) suggests a great diversity in the stability of the $[CuOCu]^{2+}$ sites than LIO (IQR = 0.50 eV). We use median/IQR as our metrics of comparison as they are better suited for interpreting skewed distributions than the mean and standard deviation (widely used for normal probability distributions).

These metrics are directly quantified as a scatter plot in **Fig. 4b**. Using the two metrics that compare the overall stability of the possible $[CuOCu]^{+2}$ sites ($E_{f,median}$) and their diversity ($E_{f,IQR}$), we observe a weak trend where zeolites with a smaller number of unique T-sites are located in the bottom left region, while the ones with more unique T-sites tend to cluster towards the top right region of the plot. Similar plots for the hydrogen binding energy (E_h) in **Fig. 4c** show a smaller diversity in E_h across different sites from various topologies; distributions for all 52 zeolites are presented in **SI Fig. S20**. However, we observe that small 8-membered ring zeolites (e.g., CHA) show lower hydrogen binding energies for methane activation than other zeolites that have larger rings.

Contrasting **Fig. 4b** and **4c**, we conclude that the hydrogen binding ability of the individual $[CuOCu]^{2+}$ sites is less sensitive to their local environment than their stability. This is an important observation since E_h has been used as a descriptor for estimating the C-H activation energy of methane. As the data in **Fig. 4b** and **4c** includes many of the experimentally studied zeolites, we conclude that seeking topologies with lower E_h may yield limited success.

This database of DFT-calculated energies provides an opportunity to re-examine the relationship between E_f and E_h (**Fig. 4d**). Although majority of the sites (red arrow) follow the

previously reported linear relationship (dotted black line), several $[CuOCu]^{2+}$ deviate significantly from this scaling relationship. This suggests that explicit calculations of E_f and E_h is necessary to quantify the reactivity on the individual sites. Nonetheless, the 2D free energy volcano plot incorporating experimental conditions with TOF as background (similar to the Latimer et al.[27]) can be found in **SI Section 6, Fig. S26-S27**.

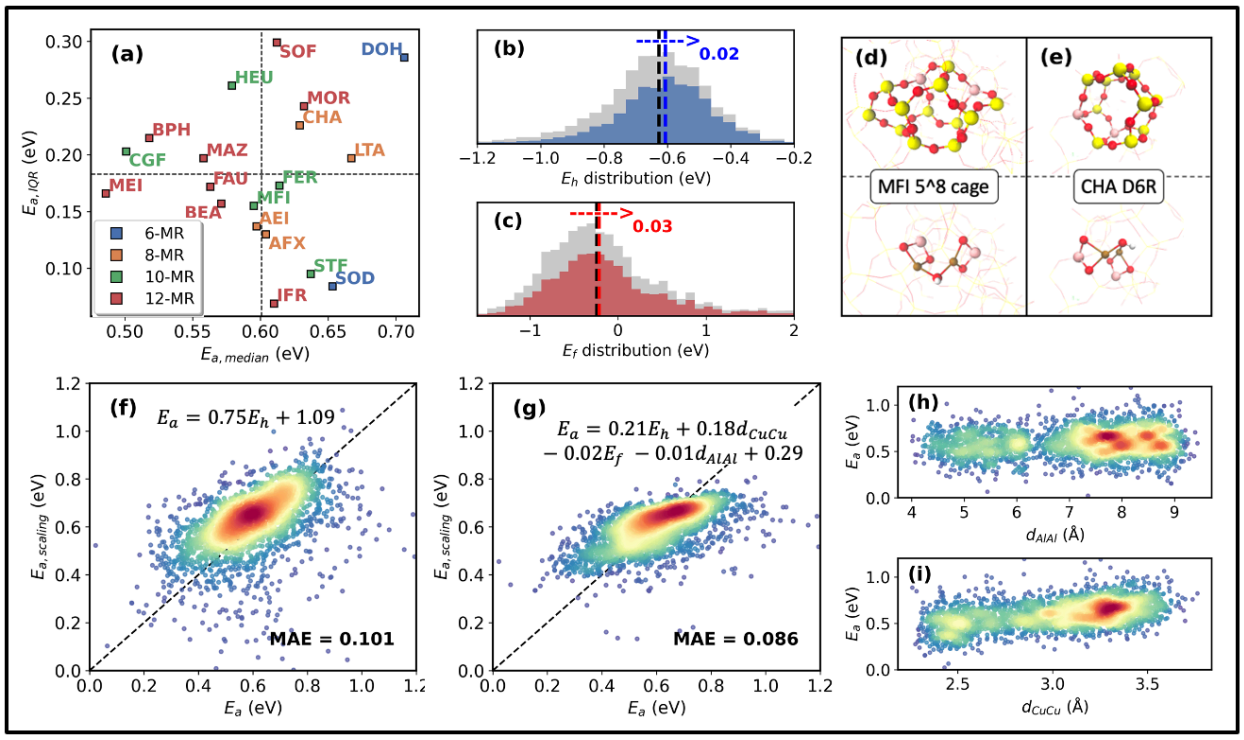

Figure 5 (a). quadrant plot of ML-predicted barriers (E_a), color-coded by the maximum window size of the corresponding topology as in **4(c).** Only the top 10 most studied and the outlier topologies are shown. **(b)-(c).** Distributions of E_h **(b)** and E_f **(c)**, including all 5,446 valid active sites (gray) vs. sites where a sensible transition state can be found using the DP_{D4} model (blue: E_h, red: E_f). Vertical dashed lines highlight the medians of the distributions. **(d)-(e).** Examples of $[CuOCu]^{2+}$ sites in MFI and CHA, accessible by a proton but inaccessible by methane. **(f).** Parity plot showing barrier predictions based on Latimer's scaling.[27] **(g).** Parity plot showing a new scaling relation more suited for Cu-exchanged zeolites, including geometric descriptors. **(h)-(i).** Distributions of geometric descriptors (d_{AlAl} in **(h)**, d_{CuCu} in **(i)**) regarding E_a.

3.2. ML-predicted activation barriers evaluating reactivity by active learning.

Although the above analysis in **Section 3.1** used E_h as a descriptor (as is common in the field), the DP_{D4} model developed in this work allows direct calculation of C-H activation barriers. **Fig. 5a** shows a quadrant plot for the DP_{D4} calculated methane activation energies for 3,356 sites across 52 zeolite topologies. Comparing these results with the analogous E_h plot in **Fig. 4c**, we observe significant re-ordering of the zeolite topologies. For instance, the $E_{a,IQR}$ (0.16 eV) is much lower than $E_{h,IQR}$ (0.32 eV) for MFI.

Moreover, compared to the number of valid E_h calculations (5,446 sites), we notice a significant reduction in the number of barriers obtained using DP_{D4}/NEBs (reduced to 3,356 sites). These data are summarized in **Fig. 5b** and **5c**, where the vertical dashed lines highlight the median of the distributions. Despite the decrease in the number of valid sites, negligible shifts in the medians are observed (0.02 eV for E_h and 0.03 eV for E_f). This indicates that most sites filtered out by the DP/NEB approach are rather "promising" (stable sites with strong hydrogen binding capabilities) and "common" (in terms of energies) based on the adsorption energy descriptors approach.

One key factor is responsible for this apparent inconsistency between the descriptor approach and the DP_{D4}/NEB results. Specifically, we observe that although hydrogen binding is feasible, the DP_{D4}/NEB calculation fails to converge as the site is sterically inaccessible to methane. Using MFI as an example, shown in **Fig. 5d,** one such stable site ($E_f = -1.15$ eV) lies within the 5^8 cage and shows a favorable E_h (-0.80 eV). Similar inaccessible sites were recently reported for Pt/MFI[83]. Another example of a sterically inaccessible $[CuOCu]^{2+}$ site (ring ($E_f = -0.53$ eV, $E_h = -0.39$ eV) within the CHA 6-membered ring is shown in **Fig. 5e**. As DP_{D4}

energetics are based on DFT, such sites are correctly excluded from further analysis. We conclude that the DP_{D4}/NEB approach provides more realistic estimates of reaction barriers as it accounts for the local steric and accessibility of the site.

Fig. 5f compares our explicitly calculated DP_{D4} barriers with previously reported universal scaling relationship. For a dataset of 3,356 sites, the Latimer model, which as trained using less than 50 calculations, shows remarkable agreement with the DP_{D4} results. **Fig. 5g** shows that these predictions for can be further improved by including geometric parameters that are directly relevant to Cu-exchanged zeolites. Our multilinear model for $[CuOCu]^{2+}$ active sites reduces the MAE to 0.086 eV. This new scaling relationship for methane activation incorporates not only energy descriptors measuring site stability and ability to bind hydrogen but also geometric descriptors such as Cu-Cu distance (denoted as d_{CuCu}) and Al-Al distance (denoted as d_{AlAl}).

The updated scaling relationship has important implications for predicting the overall reactivity of $[CuOCu]^{2+}$ sites towards methane. Consistent with the results in **Fig. 4c** and **4d**, we observe a lower contribution from site stability and the Al-Al separation (distribution shown in **Fig. 5h**) on E_a, indicating that the reactivity of a $[CuOCu]^{2+}$ site is not as sensitive to the local environment. Encouragingly, as discussed in **Section 4**, this hints that the DP_{D4}/NEB approach may be reasonably transferable to previously unseen topologies.

The new scaling relation in **Fig. 5g** and the d_{CuCu} distribution in **Fig. 5i** both reflect a non-trivial contribution of active site geometry (quantified by Cu-O-Cu angle, which can also be measured as d_{CuCu}) towards site reactivity. Since E_a is directly and positively related to d_{CuCu}, lower Cu-Cu distances result in lower E_a. Our conclusion agrees with experiment, where active sites with 2.9Å Cu-Cu separations are more selective than sites with 3.2Å separations.[69]

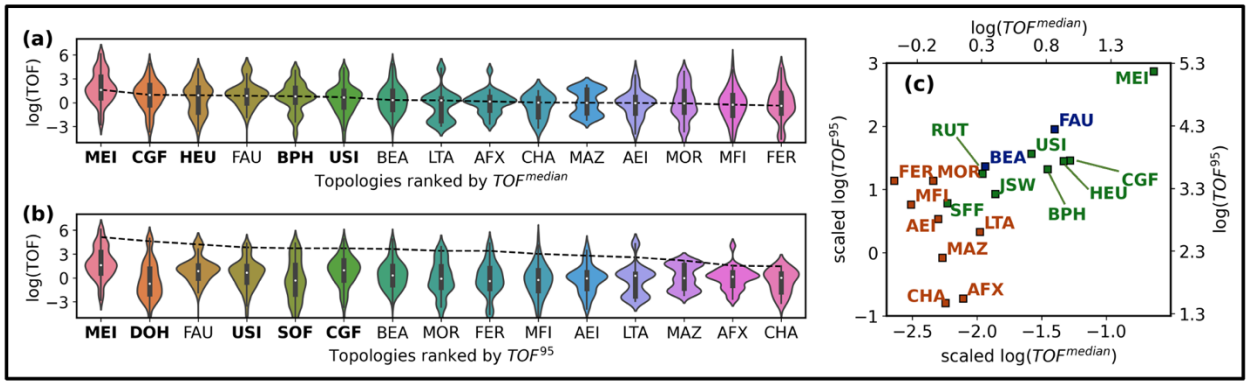

Figure 6. Violin plots ranking zeolite performance by **(a)**. median TOF (TOF^{median}) and **(b)**. TOF at the 95th percentile (TOF^{95}). Only the top 10 most studied topologies and the top 5 others are shown. (c). scatter plot of TOF^{median} and TOF^{95} on the top and right two axes, where the bottom and left axes are corresponding TOF values scaled from experimentally obtained rate constant.

3.3. Exploring the upper bond methane activation rates of each topology.

Having ensured that only sites that are accessible to methane are considered, we now turn to evaluating the methane activation performance of each topology. As a comparison, we focus on the stepwise methane activation process involving activation at 500 °C with 1 atm O_2, reaction at 80 °C, and extraction with water. Results for the higher reaction temperature at 200 °C (also seen in literature) are summarized in **SI. Fig. S31-S33**.

As the ensemble of all possible $[CuOCu]^{2+}$ sites are determined by choice of zeolite topology, we use two metrics, i.e., the median TOF (TOF^{median}) and TOF at the 95th percentile (TOF^{95}) to compare across topologies. We use these two metrics as the precise control of Al siting as well as the $[CuOCu]^{2+}$ location remain challenging experimentally. The TOFs for the individual sites are obtained according to Eq (3). These data are plotted as violin plots in **Fig. 6a** and **6b**, as this representation also captures the distribution of the TOFs for all viable $[CuOCu]^{2+}$ sites. Here, we only include 10 more experimentally explored zeolites and 5 other top performers (highlighted

in bold fonts); the complete dataset is available in the **Fig. S28-S30**. Our results show that MEI and FAU are top-5 ranked materials according to both metrics. This analysis also highlights the differences in the distributions of TOFs across the various topologies.

More importantly, since we now have a complete database of DFT calculated TOFs for all valid $[CuOCu]^{2+}$ sites in 52 zeolites, we can extend this analysis to make prediction about experimental rates Specifically, we scale our theory-predicted TOF^{median} for CHA using the experimentally measured rates for Cu-Chabazite (0.0057 s^{-1} at 80 °C).[63]

Fig. 6c summarizes the results of this analysis, where the bottom and left two axes are for scaled rates, and the top and right two axes are for theory-predicted rates. Among the well-studied topologies, zeolites such as FAU and BEA (colored in blue) are ranked in the top 20 by both metrics, while other topologies such as CHA and MAZ (colored in orange) are not. Interestingly, we notice several less explored frameworks, such as MEI (ranked first place by both metrics) and USI, can host promising $[CuOCu]^{2+}$ sites that are predicted to be more active than known materials (colored in green). While this conclusion motivates further experiments, the key assumptions and showstoppers are discussed in the **SI, Section 6**.

4. Model interpretation and configuration sampling explanation

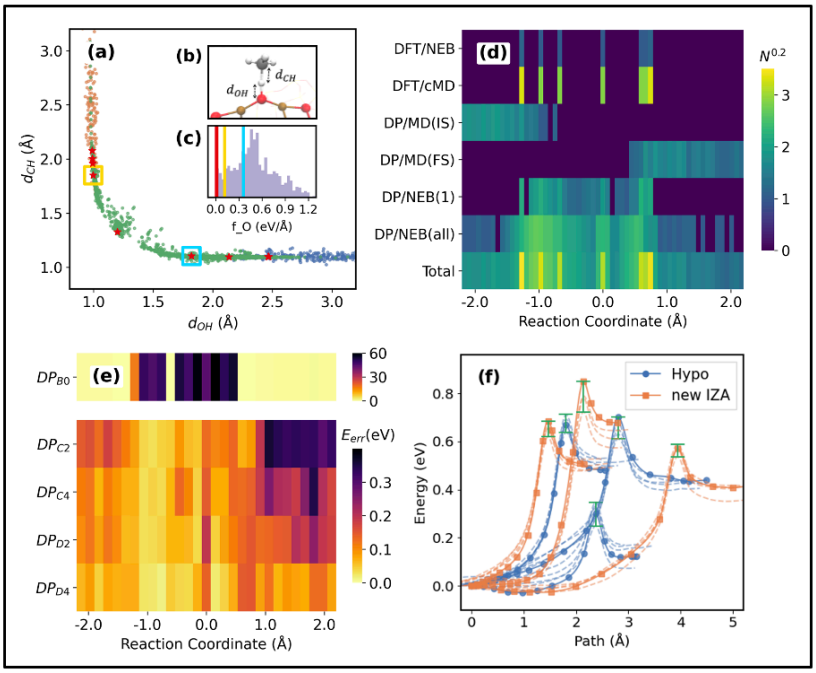

Figure 7. (a). Reaction coordinate described by d_{OH} and d_{CH} (distances labeled in **(b)**) including all MOR configurations sampled throughout entire the training protocol shown in **Fig. 1**. **(c).** distribution of forces on the oxygen atom of $[CuOCu]^{2+}$ sites from all MOR configurations sampled using DFT/cMD. Vertical lines highlight f_O sampled using DFT/NEB. **(d).** heatmap showing the number of MOR configurations sampled at each stage of the training. **(e).** heatmap showing the energy error across DP model at different training stages. **(f).** Model performance on previously unseen IZA topologies and hypothetical zeolites.

The above results provide an exhaustive analysis of the first step of the methane activation reaction in 3,356 sites across 52 zeolite topologies. As explicit DFT/NEB calculations were not possible, the aforementioned screening study has been enabled by the development of a rMLP (i.e., DP_{D4}). Thus, the reliability of our predictions depends largely on the accuracy of the trained model. Although we have previously shown (in **Fig. 3**) that DP_{D4} model faithfully represent the underlying DFT data, we now explore how the model learns and why it works.

These aspects are closely related to the interpretability of the rMLPs. Here we define interpretability as consisting of two questions: (1) what has the model "learned" from the training

data that it has seen and (2) for a system that the model has not seen before, can we trust the results of its predictions. As discussed in this Section, the multi-stage training protocol allows exploration of both these questions.

To address the first question, we explore what the model has "learned" from the training data seen throughout each stage (from **A** to **D**) using one site in MOR as an example. Here we use d_{OH} and d_{CH} to describe the reaction coordinate (**Fig. 7a**). At the beginning of the training workflow in Stage **A**, where 30 initial sites are selected based on the descriptor approach (including one site from MOR), the model has only seen the 9 images from converged DFT/NEBs (labeled as red stars). Starting from the DFT/NEB images, the "nearby" PES is then sampled using DFT/cMD at Stage **B.** Although the geometry overlaps with the overlapping with the original DFT/NEB images (as the d_{OH} and d_{CH} are constrained), the DFT/cMD is a critical initial step for stabilizing the model. Specifically, this step samples the forces on the oxygen from $[CuOCu]^{2+}$ (denoted as f_O) – the histogram in **Fig. 7b** shows that the model now has seen a much wider distribution of forces than what was available during the DFT/NEB. Additionally, the DFT/cMD also samples the zeolite backbone, which is shown previously by Sours et al[79].

The remaining steps are more obvious. From right to left (or bottom to top), we have the initial states sampled by DP/MD at stage **C1** (colored in blue), followed by the reactive region sampled by DP/NEB in stage **C2** for a single site and **D** for all other MOR sites (colored in green) and then ends with the orange region for the final states (also sampled by DP/MD at stage **C1**). Taken together, these results show that the model is exposed progressively to the region of the PES that is relevant for the C-H activation step.

In **Fig. 7b**, we highlight the uniqueness of each sampling approach (also hinted in **7a**) through a heatmap, showing the number of configurations (denoted as N) sampled by the different

approaches. Here, we define the reaction coordinate by the complete ensemble of all distances between the coordinate (i.e. [d_{OH}(i), d_{CH}(i)]) of each configuration in (a) and the coordinate of the transition state obtained from DFT/NEB (the red star in between the two boxed stars, denoted as ref_TS). To separate the two local minima regions, configurations with d_{OH} higher than that of the transition state are defined as negative in the reaction coordinate. So, from left to right of the reaction coordinate in **Fig. 7b**, we have the IS region, reactive region (ref_TS is the origin of the x-axis), and then the FS region.

As we include more configurations from different regions of the PES, the model becomes more reliable (evaluated by the difference between DFT and DP-predicted SPE). The heatmap in **Fig. 7c** shows the median errors of all configurations within a sub-section of the reaction coordinate evaluated using DP models from different training stages.

Since model DP_{B0} only saw 30 reaction coordinates through DFT/cMD, which mainly explores the zeolite backbone rather than the reactive region, we observe significantly high errors in the middle region of the reaction coordinate. On the other hand, the DP_{B0} model error at the initial and final state regions are much lower since 4 out of the 9 DFT/NEB images (where DFT/cMDs are done) are around the minima regions (also reflected in **Fig. 7a** where most red stars are actually closer to IS and FS than TS).

As we enter Stage **C** where DP/MD and DP/NEB are used to study the initial 30 sites, the errors drop immediately across the entire reaction coordinate. Since more configurations are sampled around the initial states due to the higher degree of rotational freedom of methane, errors on the left-hand side (IS) are lower than the right-hand side (FS) across all versions of the model. At Stage **D**, DP/NEBs are done for all possible sites instead of just 30, so we can see a further reduction in the error. Although the increment of error drop is getting smaller, it is because we are

getting closer to the end and by the end of the training in model DP_{D4}, the entire coordinate is below the 0.2 eV DFT error.

As the number of topologies and sites investigated in this work is already unprecedented, a significant fraction of topologies, including the hypothetical zeolite database, is yet to be explored. Therefore, the question arises of how confident the model is for a system it has not seen before and whether we can trust the results of its predictions.

Our results in **Fig. 7d** reveal that the model is not only transferable but also confident in its predictions. Specifically, we show the model transferability by performing DP/NEBs on three new IZA topologies and three hypothetical zeolites using the DP_{D4} model (with 17 intermediate images). The resulting reaction coordinates and transition state geometries are both sensible (detailed in **SI, Fig. S34, S35**). To illustrate the model confidence, we train an ensemble of models ($DP_{D3,\,i}$, dashed lines) and compare the discrepancy in prediction relative to the DP_{D4} predicted reaction coordinate (solid lines). The ensembles of models show high agreement near DFT accuracy, especially around the reactive regions.

Conclusion

In this work, we analyzed the reactivity, stability, and accessibility of complete ensembles of di-copper oxo sites that can be created within various zeolite frameworks and ranked their median and best-possible performance in activating methane. More importantly, we developed an interpretable reactive machine learning potential through a multistage active learning algorithm that preferably samples the reactive region of the potential energy surface. The quadrant plot provides a straightforward way of evaluating a material's catalytic performance through statistical interpretation of the entire energy distributions instead of concluding based on a few pre-selected

sites, enabling more objective comparisons across topologies. The database itself, combined with experimental techniques such as EXAFS, opens many opportunities and future directions for better understanding reaction mechanisms and identifying active site natures during catalytic processes. Even though we focus only on the CuOCu active site motif for direct methane to methanol conversion, our approach can be easily extended to other transition metals, motifs, and other reactions. This dataset is accessible through GitHub, which will contribute to the field by providing a database available for future ML force field design and enable quick screening of the catalyst space for other zeolite-catalyzed reactions.

References:

1. Mansoor, R., & Tahir, M. (2021). Recent Developments in Natural Gas Flaring Reduction and Reformation to Energy-Efficient Fuels: A Review. *Energy and Fuels*, *35*(5), 3675–3714.
2. Davoudi, M., Rahimpour, M. R., Jokar, S. M., Nikbakht, F., & Abbasfard, H. (2013). The major sources of gas flaring and air contamination in the natural gas processing plants: A case study. *Journal of Natural Gas Science and Engineering*, *13*, 7–19.
3. Rahimpour, M. R., & Jokar, S. M. (2012). Feasibility of flare gas reformation to practical energy in Farashband gas refinery: No gas flaring. *Journal of Hazardous Materials*, *209–210*, 204–217.
4. Elvidge, C. D., Bazilian, M. D., Zhizhin, M., Ghosh, T., Baugh, K., & Hsu, F. C. (2018). The potential role of natural gas flaring in meeting greenhouse gas mitigation targets. *Energy Strategy Reviews*, *20*, 156–162
5. Elvidge, C. D., Zhizhin, M., Baugh, K., Hsu, F. C., & Ghosh, T. (2016). Methods for global survey of natural gas flaring from visible infrared imaging radiometer suite data. *Energies*, *9*(1).
6. Thomas, S., & Dawe, R. A. (2003). Review of ways to transport natural gas energy from countries which do not need the gas for domestic use. Energy, 28(14), 1461–1477.
7. Elvidge, C. D., Ziskin, D., Baugh, K. E., Tuttle, B. T., Ghosh, T., Pack, D. W., Erwin, E. H., & Zhizhin, M. (2009). A fifteen year record of global natural gas flaring derived from satellite data. *Energies*, *2*(3), 595–622.
8. Hammond, C., Conrad, S., & Hermans, I. (2012). Oxidative methane upgrading. *ChemSusChem*, *5*(9), 1668–1686.
9. Yoshizawa, K., & Shiota, Y. (2006). Conversion of methane to methanol at the mononuclear and dinuclear copper sites of particulate methane monooxygenase (pMMO): A DFT and QM/MM study. *Journal of the American Chemical Society*, *128*(30), 9873–9881.
10. Osadchii, D. Y., Olivos-Suarez, A. I., Szécsényi, Á., Li, G., Nasalevich, M. A., Dugulan, I. A., Crespo, P. S., Hensen, E. J. M., Veber, S. L., Fedin, M. v., Sankar, G., Pidko, E. A., & Gascon, J. (2018). Isolated Fe sites in metal organic frameworks catalyze the direct conversion of methane to methanol. *ACS Catalysis*, *8*(6), 5542–5548.
11. Mae Alayon, E., Nachtegaal, M., Ranocchiari, M., & van Bokhoven, J. A. (2012). Catalytic conversion of methane to methanol over Cu-mordenite. *Chemical Communications*, *48*(3), 404–406.
12. Srivastava, R. K., Sarangi, P. K., Bhatia, L., Singh, A. K., & Shadangi, K. P. (2022). Conversion of methane to methanol: technologies and future challenges. In *Biomass Conversion and Biorefinery* (Vol. 12, Issue 5, pp. 1851–1875). Springer Science and Business Media Deutschland GmbH.
13. Tomkins, P., Ranocchiari, M., & van Bokhoven, J. A. (2017). Direct Conversion of Methane to Methanol under Mild Conditions over Cu-Zeolites and beyond. *Accounts of Chemical Research*, *50*(2), 418–425.
14. Wulfers, M. J., Teketel, S., Ipek, B., & Lobo, R. F. (2015). Conversion of methane to methanol on copper-containing small-pore zeolites and zeotypes. *Chemical Communications*, *51*(21), 4447–4450.

15. Zakaria, Z., & Kamarudin, S. K. (2016). Direct conversion technologies of methane to methanol: An overview. In *Renewable and Sustainable Energy Reviews* (Vol. 65, pp. 250–261). Elsevier Ltd.
16. Blanchette, C. D., Knipe, J. M., Stolaroff, J. K., Deotte, J. R., Oakdale, J. S., Maiti, A., Lenhardt, J. M., Sirajuddin, S., Rosenzweig, A. C., & Baker, S. E. (2016). Printable enzyme-embedded materials for methane to methanol conversion. *Nature Communications*, *7*.
17. Duan, C., Luo, M., & Xing, X. (2011). High-rate conversion of methane to methanol by Methylosinus trichosporium OB3b. *Bioresource Technology*, *102*(15), 7349–7353.
18. Arminio-Ravelo, J. A., & Escudero-Escribano, M. (2021). Strategies toward the sustainable electrochemical oxidation of methane to methanol. In *Current Opinion in Green and Sustainable Chemistry* (Vol. 30). Elsevier B.V.
19. Arminio-Ravelo, J. A., & Escudero-Escribano, M. (2021). Strategies toward the sustainable electrochemical oxidation of methane to methanol. In *Current Opinion in Green and Sustainable Chemistry* (Vol. 30). Elsevier B.V.
20. Nandy, A., Duan, C., Goffinet, C., & Kulik, H. J. (2022). New Strategies for Direct Methane-to-Methanol Conversion from Active Learning Exploration of 16 Million Catalysts. *Journal of the American Chemical Society*.
21. Conte, M., Lopez-Sanchez, J. A., He, Q., Morgan, D. J., Ryabenkova, Y., Bartley, J. K., Carley, A. F., Taylor, S. H., Kiely, C. J., Khalid, K., & Hutchings, G. J. (2012). Modified zeolite ZSM-5 for the methanol to aromatics reaction. *Catalysis Science and Technology*, *2*(1), 105–112.
22. Gao, P., Xu, J., Qi, G., Wang, C., Wang, Q., Zhao, Y., Zhang, Y., Feng, N., Zhao, X., Li, J., & Deng, F. (2018). A Mechanistic Study of Methanol-to-Aromatics Reaction over Ga-Modified ZSM-5 Zeolites: Understanding the Dehydrogenation Process. *ACS Catalysis*, *8*(10), 9809–9820.
23. Hemelsoet, K., van der Mynsbrugge, J., de Wispelaere, K., Waroquier, M., & van Speybroeck, V. (2013). Unraveling the reaction mechanisms governing methanol-to-olefins catalysis by theory and experiment. In *ChemPhysChem* (Vol. 14, Issue 8, pp. 1526–1545).
24. Bjørgen, M., Joensen, F., Spangsberg Holm, M., Olsbye, U., Lillerud, K. P., & Svelle, S. (2008). Methanol to gasoline over zeolite H-ZSM-5: Improved catalyst performance by treatment with NaOH. *Applied Catalysis A: General*, *345*(1), 43–50.
25. Tian, P., Wei, Y., Ye, M., & Liu, Z. (2015). Methanol to olefins (MTO): From fundamentals to commercialization. In *ACS Catalysis* (Vol. 5, Issue 3, pp. 1922–1938). American Chemical Society.
26. Hammond, C., Jenkins, R. L., Dimitratos, N., Lopez-Sanchez, J. A., Ab Rahim, M. H., Forde, M. M., Thetford, A., Murphy, D. M., Hagen, H., Stangland, E. E., Moulijn, J. M., Taylor, S. H., Willock, D. J., & Hutchings, G. J. (2012). Catalytic and mechanistic insights of the low-temperature selective oxidation of methane over Cu-promoted Fe-ZSM-5. *Chemistry - A European Journal*, *18*(49), 15735–15745
27. Latimer, A. A., Kulkarni, A. R., Aljama, H., Montoya, J. H., Yoo, J. S., Tsai, C., Abild-Pedersen, F., Studt, F., & Nørskov, J. K. (2017). Understanding trends in C-H bond activation in heterogeneous catalysis. *Nature Materials*, *16*(2), 225–229.

28. Kulkarni, A. R., Zhao, Z. J., Siahrostami, S., Nørskov, J. K., & Studt, F. (2016). Monocopper Active Site for Partial Methane Oxidation in Cu-Exchanged 8MR Zeolites. *ACS Catalysis*, *6*(10), 6531–6536.
29. Kurnaz, E., Fellah, M. F., & Onal, I. (2011). A density functional theory study of C-H bond activation of methane on a bridge site of M-O-M-ZSM-5 Clusters (M = Au, Ag, Fe and Cu). *Microporous and Mesoporous Materials*, *138*(1–3), 68–74.
30. Mahyuddin, M. H., Staykov, A., Saputro, A. G., Agusta, M. K., Dipojono, H. K., & Yoshizawa, K. (2020). Novel mechanistic insights into methane activation over Fe and Cu active sites in zeolites: A comparative DFT study using meta-GGA functionals. *Journal of Physical Chemistry C*, *124*(33), 18112–18125.
31. Wannakao, S., Warakulwit, C., Kongpatpanich, K., Probst, M., & Limtrakul, J. (2012). Methane activation in gold cation-exchanged zeolites: A DFT study. *ACS Catalysis*, *2*(6), 986–992.
32. Gabrienko, A. A., Arzumanov, S. S., Toktarev, A. v., Danilova, I. G., Prosvirin, I. P., Kriventsov, V. v., Zaikovskii, V. I., Freude, D., & Stepanov, A. G. (2017). Different Efficiency of Zn2+ and ZnO Species for Methane Activation on Zn-Modified Zeolite. *ACS Catalysis*, *7*(3), 1818–1830.
33. Xiao, D. J., Bloch, E. D., Mason, J. A., Queen, W. L., Hudson, M. R., Planas, N., Borycz, J., Dzubak, A. L., Verma, P., Lee, K., Bonino, F., Crocellà, V., Yano, J., Bordiga, S., Truhlar, D. G., Gagliardi, L., Brown, C. M., & Long, J. R. (2014). Oxidation of ethane to ethanol by N2 O in a metal-organic framework with coordinatively unsaturated iron(II) sites. *Nature Chemistry*, *6*(7), 590–595.
34. Verma, P., Vogiatzis, K. D., Planas, N., Borycz, J., Xiao, D. J., Long, J. R., Gagliardi, L., & Truhlar, D. G. (2015). Mechanism of oxidation of ethane to ethanol at Iron(IV)-oxo sites in magnesium-diluted Fe2(dobdc). *Journal of the American Chemical Society*, *137*(17), 5770–5781.
35. Rosen, A. S., Notestein, J. M., & Snurr, R. Q. (2019). Structure-Activity Relationships That Identify Metal-Organic Framework Catalysts for Methane Activation. *ACS Catalysis*, *9*(4), 3576–3587.
36. Pahls, D. R., Ortuño, M. A., Winegar, P. H., Cramer, C. J., & Gagliardi, L. (2017). Computational Screening of Bimetal-Functionalized Zr6O8 MOF Nodes for Methane C-H Bond Activation. *Inorganic Chemistry*, *56*(15), 8739–8743.
37. Wu, X. N., Li, J., Schlangen, M., Zhou, S., González-Navarrete, P., & Schwarz, H. (2017). Striking Doping Effects on Thermal Methane Activation Mediated by the Heteronuclear Metal Oxides [XAlO4].+ (X=V, Nb, and Ta). *Chemistry - A European Journal*, *23*(4), 788–792.
38. Xu, J., Cao, X. M., & Hu, P. (2019). Improved prediction for the methane activation mechanism on rutile metal oxides by a machine learning model with geometrical descriptors. *Journal of Physical Chemistry C*, *123*(47), 28802–28810.
39. Senanayake, S. D., Rodriguez, J. A., & Weaver, J. F. (2020). Low Temperature Activation of Methane on Metal-Oxides and Complex Interfaces: Insights from Surface Science. *Accounts of Chemical Research*, *53*(8), 1488–1497.
40. Yuan, Q., Deng, W., Zhang, Q., & Wang, Y. (2007). Osmium-catalyzed selective oxidations of methane and ethane with hydrogen peroxide in aqueous medium. *Advanced Synthesis and Catalysis*, *349*(7), 1199–1209.

41. Zimmermann, T., Soorholtz, M., Bilke, M., & Schüth, F. (2016). Selective Methane Oxidation Catalyzed by Platinum Salts in Oleum at Turnover Frequencies of Large-Scale Industrial Processes. *Journal of the American Chemical Society*, *138*(38), 12395–12400.
42. Raja, R., & Ratnasamy, P. (1997). Direct conversion of methane to methanol. *Appl. Catal,*158, L7–L15
43. Sirajuddin, S., & Rosenzweig, A. C. (2015). Enzymatic oxidation of methane. *Biochemistry*, *54*(14), 2283–2294.
44. Rosenzweig, A. C., Frederlckt, C. A., Lippard, S. J., & Nordlund, P. (1993). Crystal structure of a bacterial non-haem iron hydroxylase that catalyses the biological oxidation of methane. *Nature*, 366, 537 –543
45. Balasubramanian, R., Smith, S. M., Rawat, S., Yatsunyk, L. A., Stemmler, T. L., & Rosenzweig, A. C. (2010). Oxidation of methane by a biological dicopper centre. *Nature*, *465*(7294), 115–119.
46. Vanelderen, P., Hadt, R. G., Smeets, P. J., Solomon, E. I., Schoonheydt, R. A., & Sels, B. F. (2011). Cu-ZSM-5: A biomimetic inorganic model for methane oxidation. *Journal of Catalysis*, *284*(2), 157–164.
47. Groothaert, M. H., van Bokhoven, J. A., Battiston, A. A., Weckhuysen, B. M., & Schoonheydt, R. A. (2003). Bis(μ-oxo) dicopper in Cu-ZSM-5 and its role in the decomposition of NO: A combined in situ XAFS, UV-vis-near-IR, and kinetic study. *Journal of the American Chemical Society*, *125*(25), 7629–7640.
48. Woertink, J. S., Smeets, P. J., Groothaert, M. H., Vance, M. A., Sels, B. F., Schoonheydt, R. A., & Solomon, E. I. (2009). A [Cu2O]2 core in Cu-ZSM-5, the active site in the oxidation of methane to methanol. *Proc. Natl. Acad. Sci*.,106, 18908–18913.
49. Sushkevich, V. L., Palagin, D., Ranocchiari, M., & van Bokhoven, J. A. (2017). Selective anaerobic oxidation of methane enables direct synthesis of methanol. *Science*, 356, 523-527.
50. Zhao, Z. J., Kulkarni, A., Vilella, L., Nørskov, J. K., & Studt, F. (2016). Theoretical Insights into the Selective Oxidation of Methane to Methanol in Copper-Exchanged Mordenite. *ACS Catalysis*, *6*(6), 3760–3766.
51. Vanelderen, P., Snyder, B. E. R., Tsai, M. L., Hadt, R. G., Vancauwenbergh, J., Coussens, O., Schoonheydt, R. A., Sels, B. F., & Solomon, E. I. (2015). Spectroscopic definition of the copper active sites in mordenite: Selective methane oxidation. *Journal of the American Chemical Society*, *137*(19), 6383–6392.
52. Yumura, T., Hirose, Y., Wakasugi, T., Kuroda, Y., & Kobayashi, H. (2016). Roles of Water Molecules in Modulating the Reactivity of Dioxygen-Bound Cu-ZSM-5 toward Methane: A Theoretical Prediction. *ACS Catalysis*, *6*(4), 2487–2495.
53. Alayon, E. M. C., Nachtegaal, M., Bodi, A., Ranocchiari, M., & van Bokhoven, J. A. (2015). Bis(μ-oxo) versus mono(μ-oxo)dicopper cores in a zeolite for converting methane to methanol: An in situ XAS and DFT investigation. *Physical Chemistry Chemical Physics*, *17*(12), 7681–7693.
54. Newton, M. A., Knorpp, A. J., Sushkevich, V. L., Palagin, D., & van Bokhoven, J. A. (2020). Active sites and mechanisms in the direct conversion of methane to methanol using Cu in zeolitic hosts: A critical examination. In *Chemical Society Reviews* (Vol. 49, Issue 5, pp. 1449–1486). Royal Society of Chemistry.

55. Ravi, M., Ranocchiari, M., & van Bokhoven, J. A. (2017). The direct catalytic oxidation of methane to methanol-a critical assessment. *Angewandte Chemie International Edition*, *56*(52), 16464–16483. https://doi.org/10.1002/anie.201702550
56. Ravi, M., Sushkevich, V. L., Knorpp, A. J., Newton, M. A., Palagin, D., Pinar, A. B., Ranocchiari, M., & van Bokhoven, J. A. (2019). Misconceptions and challenges in methane-to-methanol over transition-metal-exchanged zeolites. *Nature Catalysis* (Vol. 2, Issue 6, pp. 485–494).
57. Raja, R., & Ratnasamy, P. (1997). Direct conversion of methane to methanol. *Applied Catalysis A: General*, *158*(1-2). https://doi.org/10.1016/s0926-860x(97)00105-1
58. Panov, G. I., Uriarte, A. K., Rodkin, M. A., & Sobolev, V. I. (1998). Generation of active oxygen species on solid surfaces. opportunity for novel oxidation technologies over zeolites. *Catalysis Today*, *41*(4), 365–385. https://doi.org/10.1016/s0920-5861(98)00026-1
59. Dubkov, K. A., Sobolev, V. I., Talsi, E. P., Rodkin, M. A., Watkins, N. H., Shteinman, A. A., & Panov, G. I. (1997). Kinetic isotope effects and mechanism of biomimetic oxidation of methane and benzene on FEZSM-5 zeolite. *Journal of Molecular Catalysis A: Chemical*, *123*(2-3), 155–161. https://doi.org/10.1016/s1381-1169(97)00051
60. Groothaert, M. H., Smeets, P. J., Sels, B. F., Jacobs, P. A., & Schoonheydt, R. A. (2005). Selective oxidation of methane by the bis(μ-oxo)dicopper core stabilized on ZSM-5 and mordenite zeolites. *Journal of the American Chemical Society*, *127*(5), 1394–1395. https://doi.org/10.1021/ja047158u
61. Park, M. B., Ahn, S. H., Mansouri, A., Ranocchiari, M., & van Bokhoven, J. A. (2017). Comparative Study of Diverse Copper Zeolites for the Conversion of Methane into Methanol. *ChemCatChem*, *9*(19), 3705–3713.
62. Snyder, B. E. R., Vanelderen, P., Schoonheydt, R. A., Sels, B. F., & Solomon, E. I. (2018). Second-Sphere Effects on Methane Hydroxylation in Cu-Zeolites. *Journal of the American Chemical Society*, *140*(29), 9236–9243.
63. Rhoda, H. M., Plessers, D., Heyer, A. J., Bols, M. L., Schoonheydt, R. A., Sels, B. F., & Solomon, E. I. (2021). Spectroscopic Definition of a Highly Reactive Site in Cu-CHA for Selective Methane Oxidation: Tuning a Mono-μ-Oxo Dicopper(II) Active Site for Reactivity. *Journal of the American Chemical Society*, *143*(19), 7531–7540.
64. Mahyuddin, M. H., Staykov, A., Shiota, Y., Miyanishi, M., & Yoshizawa, K. (2017). Roles of Zeolite Confinement and Cu-O-Cu Angle on the Direct Conversion of Methane to Methanol by [Cu2(μ-O)]2+-Exchanged AEI, CHA, AFX, and MFI Zeolites. *ACS Catalysis*, *7*(6), 3741–3751.
65. Mahyuddin, M. H., Saputro, A. G., Sukanli, R. P. P., Fathurrahman, F., Rizkiana, J., Nuruddin, A., & Dipojono, H. K. (2022). Molecular insight into the role of zeolite lattice constraints on methane activation over the Cu-O-Cu active site. *Physical Chemistry Chemical Physics*, *24*(7), 4196–4203.
66. Mahyuddin, M. H., Tanaka, T., Shiota, Y., Staykov, A., & Yoshizawa, K. (2018). Methane Partial Oxidation over [Cu2(μ-O)]2+ and [Cu3(μ-O)3]2+ Active Species in Large-Pore Zeolites. *ACS Catalysis*, *8*(2), 1500–1509.
67. Mahyuddin, M. H., Shiota, Y., & Yoshizawa, K. (2019). Methane selective oxidation to methanol by metal-exchanged zeolites: A review of active sites and their reactivity. In *Catalysis Science and Technology* (Vol. 9, Issue 8, pp. 1744–1768). Royal Society of Chemistry.

68. Pappas, D. K., Borfecchia, E., Dyballa, M., Pankin, I. A., Lomachenko, K. A., Martini, A., Signorile, M., Teketel, S., Arstad, B., Berlier, G., Lamberti, C., Bordiga, S., Olsbye, U., Lillerud, K. P., Svelle, S., & Beato, P. (2017). Methane to Methanol: Structure-Activity Relationships for Cu-CHA. *Journal of the American Chemical Society*, *139*(42), 14961–14975.
69. Artsiusheuski, M. A., van Bokhoven, J. A., & Sushkevich, V. L. (2022). Structure of selective and nonselective Dicopper (II) sites in CuMFI for methane oxidation to methanol. *ACS Catalysis*, *12*(24), 15626–15637. https://doi.org/10.1021/acscatal.2c05299
70. Li, Y., Bahamon, D., Sinnokrot, M., & Vega, L. F. (2022). Computational screening of transition metal-doped cds for photocatalytic hydrogen production. *Npj Computational Materials*, *8*(1). https://doi.org/10.1038/s41524-022-00922-4
71. Quaino, P., Juarez, F., Santos, E., & Schmickler, W. (2014). Volcano plots in hydrogen electrocatalysis – uses and abuses. *Beilstein Journal of Nanotechnology*, *5*, 846–854. https://doi.org/10.3762/bjnano.5.96
72. Chanussot, L., Das, A., Goyal, S., Lavril, T., Shuaibi, M., Riviere, M., Tran, K., Heras-Domingo, J., Ho, C., Hu, W., Palizhati, A., Sriram, A., Wood, B., Yoon, J., Parikh, D., Zitnick, C. L., & Ulissi, Z. (2021). Open catalyst 2020 (OC20) dataset and community challenges. *ACS Catalysis*, *11*(10), 6059–6072. https://doi.org/10.1021/acscatal.0c04525
73. Tran, R., Lan, J., Shuaibi, M., Wood, B. M., Goyal, S., Das, A., Heras-Domingo, J., Kolluru, A., Rizvi, A., Shoghi, N., Sriram, A., Therrien, F., Abed, J., Voznyy, O., Sargent, E. H., Ulissi, Z., & Zitnick, C. L. (2023). The Open Catalyst 2022 (OC22) dataset and challenges for oxide electrocatalysts. *ACS Catalysis*, *13*(5), 3066–3084. https://doi.org/10.1021/acscatal.2c05426
74. Baerlocher, C.; McCusker, L. B. Database of Zeolite Structures. http://www.iza-structure.org/databases/.
75. Antonio, D. D., Guo, J., Holton, S. J., & Kulkarni, A. R. (2021). Simplifying computational workflows with the Multiscale Atomic Zeolite Simulation Environment (MAZE). *SoftwareX*, *16*.
76. Mandal, K., Gu, Y., Westendorff, K. S., Li, S., Pihl, J. A., Grabow, L. C., Epling, W. S., & Paolucci, C. (2020). Condition-Dependent Pd Speciation and NO Adsorption in Pd/Zeolites. *ACS Catalysis*, *10*(21), 12801–12818.
77. Knott, B. C., Nimlos, C. T., Robichaud, D. J., Nimlos, M. R., Kim, S., & Gounder, R. (2018). Consideration of the Aluminum Distribution in Zeolites in Theoretical and Experimental Catalysis Research. In *ACS Catalysis* (Vol. 8, Issue 2, pp. 770–784). American Chemical Society. https://doi.org/10.1021/acscatal.7b03676
78. Xie, P., Pu, T., Aranovich, G., Guo, J., Donohue, M., Kulkarni, A., & Wang, C. (2021). Bridging adsorption analytics and catalytic kinetics for metal-exchanged zeolites. *Nature Catalysis*, *4*(2), 144–156.
79. Sours, T., & Kulkarni, A. (2022). Predicting structural properties of pure silica zeolites using deep neural network potentials. https://doi.org/10.26434/chemrxiv-2022-ssvwl
80. Verploegh, R. J., Kulkarni, A., Boulfelfel, S. E., Haydak, J. C., Tang, D., & Sholl, D. S. (2019). Screening diffusion of small molecules in flexible zeolitic imidazolate frameworks using a DFT-parameterized force field. The Journal of Physical Chemistry C, 123(14), 9153–9167. https://doi.org/10.1021/acs.jpcc.9b00733

81. Krishna, R., & van Baten, J. M. (2007). Using molecular simulations for screening of zeolites for separation of CO2/CH4 mixtures. Chemical Engineering Journal, 133(1-3), 121–131. https://doi.org/10.1016/j.cej.2007.02.011
82. Zhang, Y., Wang, H., Chen, W., Zeng, J., Zhang, L., Wang, H., & E, W. (2020). DP-gen: A concurrent learning platform for the generation of reliable deep learning based potential energy models. Computer Physics Communications, 253, 107206. https://doi.org/10.1016/j.cpc.2020.107206
83. Felvey, N., Guo, J., Rana, R., Xu, L., Bare, S. R., Gates, B. C., Katz, A., Kulkarni, A. R., Runnebaum, R. C., & Kronawitter, C. X. (2022). Interconversion of atomically dispersed platinum cations and platinum clusters in zeolite ZSM-5 and formation of platinum *gem*-dicarbonyls. *Journal of the American Chemical Society*, *144*(30), 13874–13887. https://doi.org/10.1021/jacs.2c05386