Data-efficient modeling of catalytic reactions via enhanced sampling and on-the-fly learning of machine learning potentials

Simone Perego[1, 2] and Luigi Bonati[1, a)]

[1)] *Atomistic Simulations, Italian Institute of Technology, Genova, Italy*

[2)] *Department of Materials Science, University of Milano-Bicocca, Milano, Italy*

Simulating catalytic reactivity under operative conditions poses a significant challenge due to the dynamic nature of the catalysts and the high computational cost of electronic structure calculations. Machine learning potentials offer a promising avenue to simulate dynamics at a fraction of the cost, but they require datasets containing all relevant configurations, particularly reactive ones. Here we present a scheme to construct reactive potentials in a data-efficient manner. This is achieved by combining enhanced sampling methods first with Gaussian processes to discover transition paths and then with graph neural networks to obtain a uniformly accurate description. The necessary configurations are extracted via an active learning procedure based on local environment uncertainty. We validated our approach by studying several reactions related to the decomposition of ammonia on iron-cobalt alloy catalysts. Our scheme proved efficient, requiring only $\sim 1,000$ DFT calculations per reaction, and robust, sampling reactive configurations from the different accessible pathways. Using this potential, we calculated free energy profiles and characterized reaction mechanisms, showing the ability to provide microscopic insights into complex processes under dynamic conditions.

Dynamics has long been recognized as a key ingredient in chemical reactivity, particularly in heterogeneous catalysis where active sites continuously evolve according to reaction conditions[1,2]. High temperatures and pressures induce intricate transformations in the microscopic structure of the catalyst, ranging from surface diffusivity[3] to reactant- or adsorbate-induced[4] reconstructions. If capturing these dynamic phenomena experimentally is a formidable challenge, their computational modeling is no less so. Atomistic simulations, particularly molecular dynamics (MD), are an ideal candidate for providing microscopic insights into their workings[5–7]. However, to fully describe these dynamical effects we need both accurate quantum-mechanical (QM) models of the potential energy surface and extended time- and length-scale simulations[6], two often conflicting requirements.

Machine learning (ML) potentials[8] have emerged in recent years as promising tools to address the accuracy-efficiency trade-off. They are optimized to reproduce energies and forces from a dataset of reference calculations, typically performed at the Density Functional Theory (DFT) level. Hence, their effectiveness depends on the quality of the training dataset, which must include not only the equilibrium structures but also the relevant high-energy ones. This is especially crucial for transition state (TS) geometries, whose energies are connected to reaction rates through an exponential relationship. However, the identification of these structures in a complex and dynamic environment remains elusive, especially under operating conditions where an ensemble of TS configurations often exists[9–11]. The quest is made even more challenging by the cost of *ab initio* calculations, which can rapidly become prohibitive for large systems or when an accurate electronic description is required. Therefore, the ability to efficiently construct reactive potentials with a

minimum number of QM calculations is critical to enable their widespread use. In this regard, recent developments in data-efficient architectures can help in reducing the number of points needed to train a robust model. These advances include equivariant graph neural networks[12] or transfer learning approaches[13] leveraging foundational models[14,15]. However, these techniques still do not solve the crucial issue of identifying the few relevant configurations to be included in the reference dataset set.

The training set construction typically involves active learning procedures[16–18], in which an ML model is trained on an initial dataset and used to generate new structures (e.g., via molecular dynamics). A subset of these configurations is then labeled with single-point DFT calculations and added to the training set, proceeding iteratively until convergence. In the field of computational catalysis, such schemes have been also employed in combination with nudged elastic band[19] and minima hopping[20] to accelerate the calculation of energetic barriers and adsorbate geometries. However, we cannot rely on these static approaches to simulate the dynamics under operating conditions, since the mechanism and the relevant environments may deviate significantly from the calculations performed at T=0 K[9,21–23]. Adaptive enhanced sampling techniques, such as metadynamics[24], are better suited to sample the reactive landscape at finite temperatures and to collect a diverse set of atomic environments[25,26]. These methods introduce an external bias potential that increases the fluctuations of selected collective variables (CVs), allowing to sample the transitions between metastable states[27]. By integrating these sampling techniques with active learning strategies, it has been possible to construct ML potentials for a wide variety of rare events, from phase transitions [28–30] to chemical reactions[31–34], and catalytic processes[11,23,35–37]. Recently, these techniques have been also used to enhance the model uncertainty to explore high-uncertainty configurations[38–41]. Despite the success of enhanced sampling-based active learning, these studies
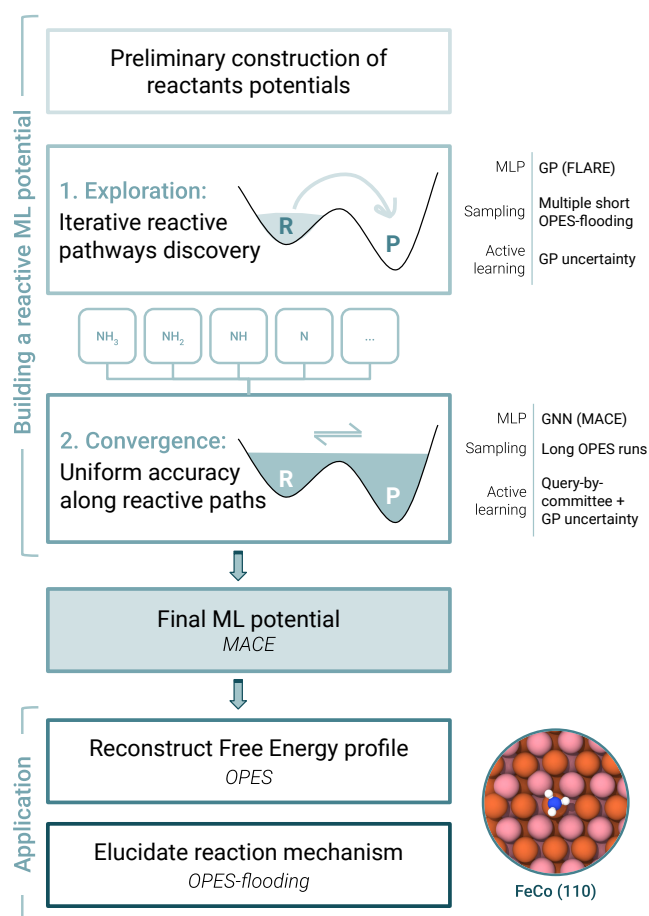
a)luigi.bonati@iit.it

Figure 1. **Diagram of the protocol** used to obtain uniformly accurate and data-efficient reactive potentials. On the right, we have reported some details regarding the ML potential, enhanced sampling scheme, and selection criterion used in the different stages.

col). As an enhanced sampling method, we used OPES[43], a recent evolution of metadynamics that offers greater flexibility and comes with different variants that can be used to explore or converge the free energy landscape. Furthermore, another important ingredient are Gaussian processes (GPs)[44], which we used first to learn the potential energy surface on-the-fly and then to identify novel local environments to build a minimal data set within our active learning scheme.
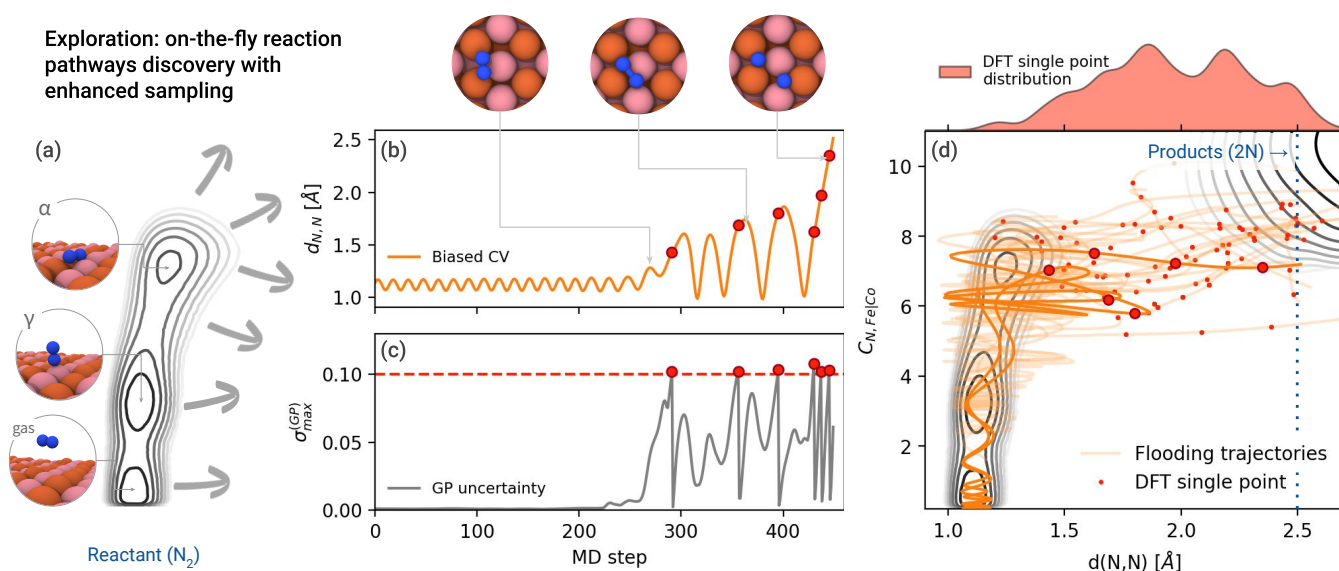
We illustrate these methodological advances by studying several reactions related to ammonia decomposition over iron-cobalt alloy catalysts. This process is the reverse of ammonia synthesis, carried out through the famous Haber-Bosch process, and is crucial to enable ammonia to be used as a hydrogen carrier[45]. On metal surfaces, cracking of ammonia typically occurs through three steps of dehydrogenation followed by nitrogen desorption via recombination[46]. In addition, the ferromagnetic and metallic properties of the surface make the description of the electronic structure particularly computationally expensive. Experimentally, an improvement in catalytic performance over pure Fe has been observed[47,48], but a microscopic understanding of the mechanism is still lacking. Moreover, the crystal structure is the same as that of α-iron (bcc), on which we have extensively investigated both ammonia synthesis [23,36] and decomposition[11]. All these features thus make it an ideal candidate for showcasing our highly data-efficient protocol in a real-world scenario.

RESULTS

Preliminary construction of reactant potentials.

In our pursuit to model the catalytic process of ammonia decomposition FeCo surfaces, we initiated our investigation by constructing ML potentials for the reactants. This involved gathering configurations to characterize the pristine 110 surface (see Fig. S1) and the different intermediate species adsorbed on it (NH_3, NH_2, NH, N, N_2, H,...). To efficiently accomplish this task, we employed GPs to learn on-the-fly the potential energy surface, with the sparse implementation of FLARE[49] using the Atomic Cluster Expansion (ACE) descriptors[50]. Recognizing the limitations of GPs with large training datasets[44], we trained separate models for each intermediate species.

Initially, we generated a dataset by conducting a set of uncertainty-aware molecular dynamics (MD) simulations based on GPs at the *operando* temperature of T=700 K. Subsequently, we performed simulations at higher temperatures to diversify our configurations and capture surface dynamics. In addition, to obtain an exhaustive coverage of the reactant space, we carried out enhanced sampling simulations to explore the various adsorption sites and the diffusion of the molecules on the surface. This preparatory stage produced about 2500 configurations for all the different intermediates.

often required many iterations and numerous DFT calculations, leaving room for improvement. Two aspects should be considered to make the process more efficient. At first, when there is little data available on the reactive regions, incremental learning[42] approaches should be used in which the potential is frequently updated. Otherwise, by pushing the system out of equilibrium with enhanced sampling, we risk extrapolating poorly and sampling the wrong configurational regions. On the other hand, we also need an efficient criterion for selecting structures for DFT calculations from those obtained in the active learning process.

In this manuscript, we introduce a new scheme to efficiently construct reactive potentials, leveraging both enhanced sampling methods and on-the-fly selection of relevant structures. This is achieved following a two-step protocol: an exploratory phase to harvest an initial pool of reactive configurations and a second stage in which we obtain a uniformly accurate description of the transition pathways (see Fig. 1 for a diagram of the proto-

Figure 2. **Exploration stage: iterative discovery of reaction paths through flooding simulations and GP-based MD.** We report a schematic presentation for the case of N_2 dissociation. a) The reactant free energy landscape constructed at the end of the preliminary phase, highlighting the metastable minima (gas phase and two adsorption states). Arrows indicate that the reaction paths toward the product state are unknown *a priori*. Several iterations are then initiated from the reactant basin. The time evolution of the distance between N atoms $d_{N,N}$ (b) and the GP maximum uncertainty on the local environments (c) are shown for a specific iteration. Every time the uncertainty of the GP exceeds the threshold of 0.1 (red dotted line), energy and forces are recomputed at the DFT level (red dots), and the GP is updated. Panel (d) shows all iterations in the 2D space defined by $d_{N,N}$ and the coordination number between N and surface atoms $C_{N,Fe|Co}$, with the trajectory shown in (b,c) highlighted. Above the top axis, we report the distribution of points where DFT calculations were performed, projected along $d_{N,N}$.

Given the complexity of the reaction pathways and the potential existence of multiple channels, we approached the collection of reactive configurations through a two-step process. In the first, the reactive pathways are discovered, while in the second, the description of these pathways is improved until high accuracy is achieved.

Reactive pathways discovery via uncertainty-aware flooding simulations.

The initial step in harvesting reactive configurations is the discovery of reaction pathways and transition state structures. While for gas phase reactions a simple interpolation between reactants and products can provide a reasonable guess, for catalyzed reactions this is not possible, especially at high temperature. Indeed, the active site is not known beforehand or may change due to dynamics[23] or, again, there may be multiple reaction pathways that need to be sampled[11]. To address this challenge, we used OPES to perform a set of "flooding-like"[51] simulations together with uncertainty-aware MD (see Methods). OPES-flooding introduces an external potential to fill the reactant basin and then let the reactive event occur as spontaneously as possible along the low free-energy pathways. This allowed us to sample reactive processes with minimal knowledge of the reaction mechanism. The only requirement for these methods is the definition of

collective variables (CVs) that can distinguish between reactant and product states. We note that, in the case where the products of the reaction to be studied are not known, generic CVs could be used to discover the possible products[52]. In addition, the integration with uncertainty-aware MD brings two significant advantages: it allows for an efficient selection of reactive configurations based on the uncertainty of the local environments, and it updates the potential energy surface model in an incremental manner, correcting wrong extrapolations to nonphysical regions of phase space.

To illustrate our workflow and demonstrate its effectiveness in sampling reactive pathways, we start with the case of N_2 dissociation, a process that we have extensively examined on Fe surfaces[23,36]. Nitrogen molecules can adsorb onto the (110) surface in two main configurations, either parallel (α) or perpendicular (γ) to the surface (see Fig. 2a). *A priori*, we do not know from which adsorption site the molecule may break. To find out, we performed a set of iterative OPES-flooding simulations, using the same CVs of Ref.[23], namely the N-N distance $d_{N,N}$ and the coordination of N with surface atoms $C_{N,Fe|Co}$. As it can be seen in Fig. 2b, the bias potential promoted the exploration of new regions of the potential energy surface. Whenever the uncertainty of the local environments (Fig. 2c) exceeded a predefined threshold (i.e., it was structurally different from the environments in the training set), a new DFT calculation was performed, and
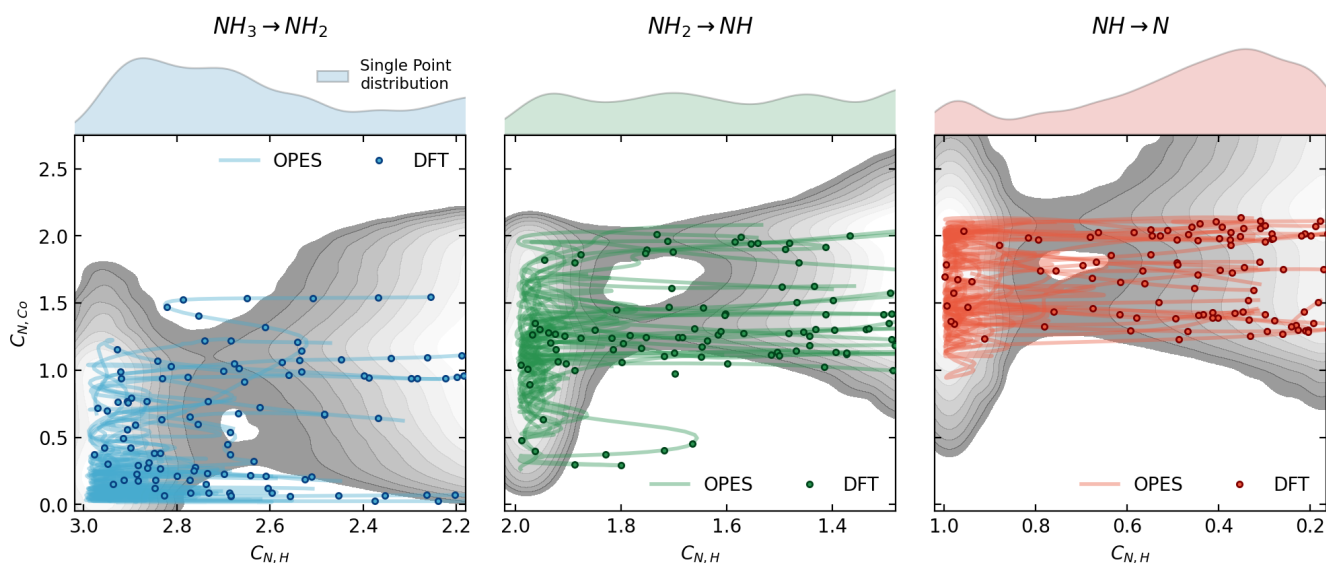
Figure 3. **Reaction path discovery for the hydrogenation steps.** The trajectories are visualized in the space defined by $C_{N,H}$ and $C_{N,Co}$. To facilitate the understanding of the sampled paths, we have shown in the background the free energy surface obtained from the final simulations, highlighting the minimum free energy paths. The dots represent the structures recalculated with DFT, whose distributions along $C_{N,H}$ are projected above the upper axis.

the GP model was updated. The reactive pathway of N_2 dissociation can be effectively visualized in the plane defined by the two CVs (thick orange curve in Fig. 2d), illustrating the progression from the gas phase through the adsorption sites γ and α and finally to the transition state region. Whenever the system reached the product state (vertical dashed line), the simulation was stopped and a new one started. In this way, an ensemble of reactive pathways was iteratively collected (Fig. 2d), ensuring continuous refinement of the potential energy surface and inclusion of a diverse set of reactive environments. Indeed, each pathway exhibited geometric diversity covering a wide range of coordination with surface atoms while originating from the same horizontal adsorption site α.

The need for a flexible and adaptive approach is even more important for modeling the three dehydrogenation steps from ammonia to atomic nitrogen. To harvest such reactive structures, we followed the same protocol, using uncertainty-aware flooding simulations starting from the GP potentials optimized for each reactant. In previous work on the Fe surfaces[11], despite being three steps of the same chemical process (dehydrogenation), the different interactions of the adsorbates with surface atoms had necessitated the use of different CVs to converge the free energy calculations. Since we are in an exploratory phase here, we wanted to use instead a single generic CV (*i.e.*, the planar distance between N and H) for all reaction steps in order to demonstrate the ability of our approach to find the accessible pathways with minimal knowledge. The resulting trajectories are visualized in the plane defined by nitrogen-hydrogen and nitrogen-cobalt coordination (Fig. 3), whose distribution revealed a broad spectrum indicative of the existence of at least two distinct reaction

channels. To put our results into context, we superimposed the reactive pathways on the free energy surface derived later from our study (Fig. S5). The remarkable alignment between the sampled and minimum free energy pathways confirmed the effectiveness of our methodology in accurately sampling the crucial reactive configurations without prior assumptions on the adsorption sites or the transition states. Although each dehydrogenation step occurred from different adsorption sites characterized by increasing coordination with the surface, thanks to the flooding scheme, we were able to find reactive pathways using a generic CV that only distinguished reactants from products. Furthermore, an analysis of the evolution of the sampled paths as a function of iterations demonstrated the importance of an incremental approach: the transition pathways initially pass through high free-energy regions and then converge to low free-energy regions as new data are added (Fig. S2). In total, about 500 reactive configurations were collected in this exploratory phase, demonstrating that our approach enables systematic and efficient sampling of reaction pathways under *operando* conditions, even when multiple ones are available.

Uniform accuracy along reactive paths through GNNs and GP-based active learning

Having collected a first dataset of reactive structures, we moved on to the second phase of our approach, aiming to converge the accuracy along the reactive paths. For this purpose, we used equivariant graph neural networks (GNNs)[12] to represent the potential energy surface. These architectures are more flexible than GPs[53] and have
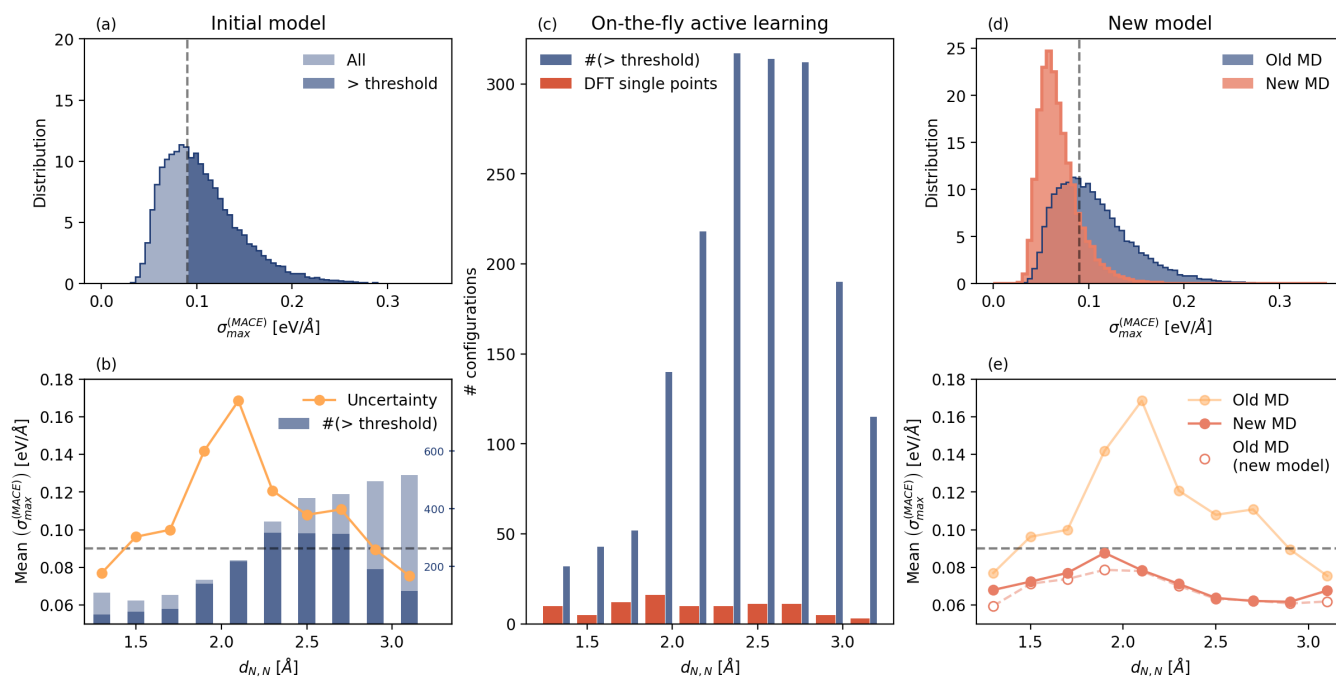
Figure 4. **Convergence stage: uniformly accurate potentials along the reactive path**. We illustrate the GP-based active learning of the MACE structures for the N recombination and dissociation process. a) Distribution of the maximum uncertainty on the forces $\sigma_{max}^{(MACE)}$ (calculated from a committee model) on the simulation performed with MACE before active learning. The grey dotted line at 90 meV/Å represents the chosen threshold for the query-by-committee selection. b) Left axis: average uncertainty $\sigma_{max}^{(MACE)}$ as a function of the collective variable $d_{N,N}$ (orange line). On the right axis: number of configurations sampled (light blue bars) and the subset selected with the query-by-committee criterion (dark blue bars). c) Distribution of the DFT single point performed with our on-the-fly GP-based selection scheme (large red bars), compared with standard query-by-committee selection (narrow blue bars). d) Uncertainty distribution of a new MD simulation performed with the MACE model after the active learning procedure (red histogram), contrasted with the previous distribution from panel a (blue histogram). e) Average uncertainty along the reaction coordinate $d_{N,N}$ for the configurations extracted from the MD simulation performed with the old MACE model and evaluated with the old (yellow solid curve) and new (red dashed line) model. The uncertainty calculated on the new trajectory generated with the new model is represented with a red solid line.

demonstrated remarkable data efficiency, enabling robust simulations as early as hundreds/thousands of training points[54]. In particular, we used MACE[55], which integrates the descriptors used in our previous approach (*i.e.*, ACE) in a message-passing scheme. Resorting to neural networks allowed us to consolidate all previously collected data into a single potential, overcoming the limitations imposed by having separate GPs for each intermediate. In addition, this allowed us to conduct long ($\sim$ns) enhanced sampling simulations instead of the many short ($\sim$ps) simulations of the exploratory phase. This way, we sampled hundreds of different reactive events, providing a more thorough sampling of reaction pathways under dynamic conditions.

To enrich the training dataset further and improve the accuracy of our model, we turned our attention to selecting additional structures from the new MD trajectories. As before, we illustrate our method while focusing on the nitrogen recombination and dissociation process. Our first step is to examine the uncertainty of the MACE model, evaluated using the maximum standard deviation

of a committee of models (see Fig. 4a). The distribution of the maximum uncertainty per configuration peaks around 80 meV Å^{-1}. We note that this value is high due to the large magnitudes of the N forces involved in the reaction: 90 percent of these forces have a relative uncertainty of less than 20 percent, see Fig. S4. Using the query-by-committee criterion[17], we can select a subset of structures whose uncertainty is above a given threshold. This value is typically chosen to be slightly above the uncertainty peak to target configurations which are not already well described[32,35](in this case it was chosen to be equal to 90 meV Å^{-1}). To assess the quality of the potential in describing the reactive event, it was revealing to analyze the average uncertainty as a function of the collective variable describing the progress of the reaction (Fig. 4b). Indeed, this quantity showed a strongly nonuniform behavior, with significantly higher uncertainty in the region between the reactant and product states (*i.e.*, $1.5 < d_{N,N} < 2.3$). Furthermore, by comparing the mean uncertainty with the distribution of the selected configurations (Fig. 4b) we realized that the query-by-committee
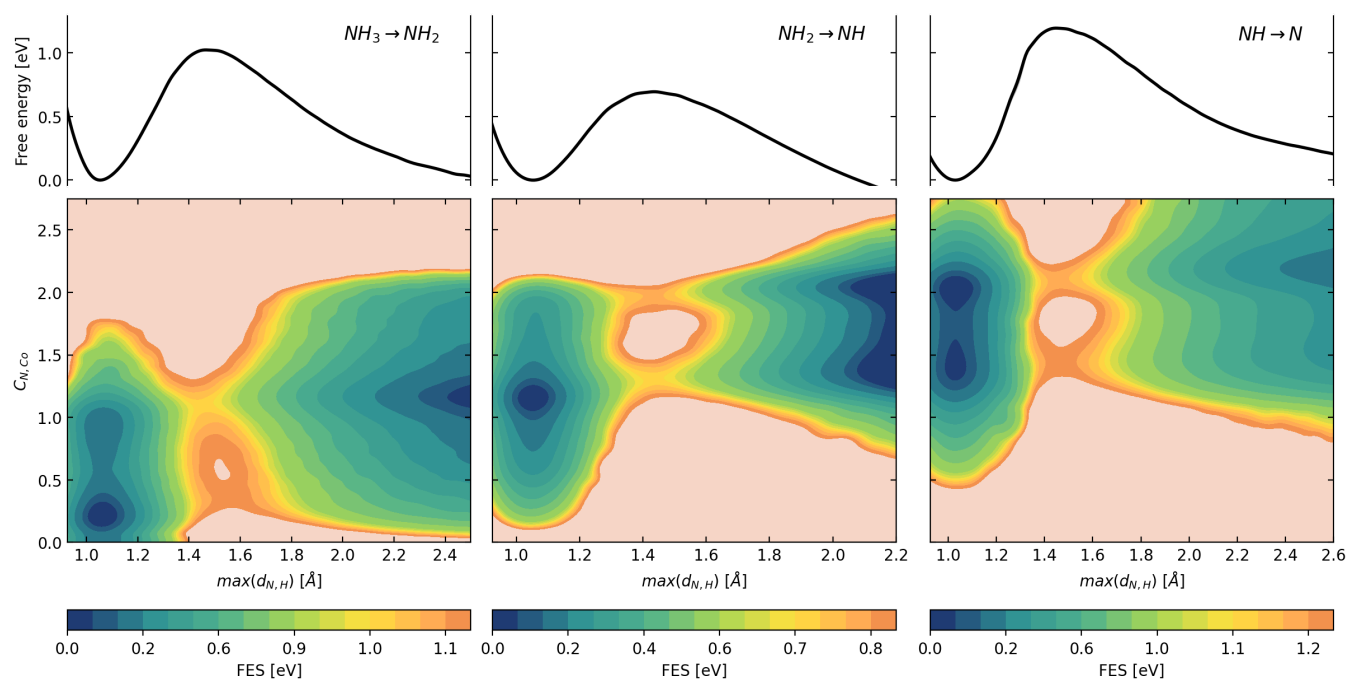
Figure 5. **Free energy profile of dehydrogenation steps** projected along the maximum distance between N and H atoms $max(d_{N,H})$ (top), and on the 2D plane defined by $max(d_{N,H})$ and $C_{N,Co}$ (bottom). The uncertainties on the 1D free energy profiles (shadow region, $\tilde{0}.01$ eV) are calculated with a weighted block average[43].

criterion would lead to an imbalance between the region of the space which is most sampled ($2.3 < d_{N,N} < 2.7$), but whose configurations are only slightly above the threshold, and the one that really needs to be improved. In previous works, it was necessary to manually select multiple configurations with high uncertainty[35] or from a specific region in CV space[32] to accurately model reactive events.

Our objective is to avoid these issues and systematically achieve uniform accuracy along the entire reaction path with a minimal number of DFT calculations. To achieve this goal, we employed Gaussian processes once again, this time as a tool to screen configurations generated by MACE. We estimated the local uncertainty given by the GP on the structures collected by MACE (and selected via query-by-committee criterion) and performed a DFT calculation only when this exceeded a chosen threshold. By updating on-the-fly the uncertainty estimate after each single-point calculation, we selected a set of nonredundant configurations. This is made possible by the correlation between the standard deviation of the MACE force predictions and the local GP uncertainty (Fig. S3). This approach allowed us to perform only $\sim$ 100 DFT calculations, which correspond to about 5% of those selected by the query-by-committee criterion (Fig. 4c). Moreover, this percentage varies significantly as a function of CV, ranging from 2-3% in the most sampled region to 12-25% in the TS region, demonstrating how our algorithm can automatically target the relevant reactive configurations. Adding these selected configurations to the MACE model effectively lowered the uncertainty of the entire previous

MD simulation under the threshold (Fig. 4e). However, a more rigorous test involved running a new MD simulation with the updated potential (Fig. 4e). Remarkably, after one active learning cycle, the average uncertainty of this new simulation was uniformly below the threshold (Fig. 4e), indicating that we have achieved the desired robustness and quality for describing the reaction progress. Otherwise, we would have performed another cycle with the same protocol.

We then applied this approach to the dehydrogenation steps, achieving uniform accuracy along the whole reactive pathways, see Fig. S5. In total, 450 DFT calculations were selected from the reactive region. Furthermore, we added a selection of structures of the products using the query-by-committee criterion, since, unlike the nitrogen recombination case, they had not been sampled in the preliminary simulations.

In summary, we constructed a uniformly accurate potential along all reaction paths while requiring a minimum number of DFT calculations. The root mean square error (RMSE) of the final model on the validation dataset was 0.4 meV/atom for energies and 17 meV/Å for forces (corresponding to a relative error of less than 3 percent). Furthermore, to confirm the accuracy and reliability of our potential in the production simulations, we built a test dataset by performing additional single-point DFT calculations extracted from the dehydrogenation simulations, where we obtained similar RMSE values of 0.2 meV/atom for energies and 14 meV/Å for forces. Fig. S6 reports also the accuracy of force predictions divided

per chemical species. Regarding the training set size, if we focus on nitrogen dissociation/recombination, our approach produced a robust potential with less than 1k configurations (958). Considering the entire catalytic cycle, which consists of 6 different intermediates, a total of 5k DFT calculations were used (see Table S1 for the detailed composition). Compared to the number of configurations used in previous studies[11,23] on Fe surfaces (30k for N_2 dissociation[23] and 110k for ammonia decomposition[11]), we obtained a more than 20-fold improvement in data efficiency. This was possible not only because of data-efficient architectures but, more importantly, due to an efficient protocol for sampling reactive configurations and selecting the most relevant ones.

Unraveling the free energy landscape and dehydrogenation mechanism

Thanks to the potential we have constructed, we can study the mechanism of ammonia decomposition on FeCo alloys in detail. As a representative example, we focus on the dehydrogenation steps of ammonia to atomic nitrogen; showing how the same set of simulations used to construct the potential also allowed us to reconstruct the free energy profiles and unravel the reaction mechanism.

We first performed a set of 25 ns-long OPES simulations, which allowed us to calculate the free energy profiles of the three dehydrogenation steps (as depicted in Fig. 5). The resulting free energy barriers exhibited varying heights, with $NH_2 \rightarrow NH+H$ being the lowest (0.7 eV), followed by $NH_3 \rightarrow NH_2+H$ (1 eV), and finally $NH \rightarrow N+H$ (1.2 eV). The free energies are accurately reconstructed, with a sampling error smaller than 20 meV. Analyzing the 2D free energy surface as a function of the maximum distance between N and H atoms, $max(d_{N,H})$, and the coordination number between N and Co atoms, $C_{N,Co}$, revealed further insights on the reaction mechanism. Multiple pathways are clearly present in all three reaction steps, as already observed in the exploration phase. Notably, these paths exhibit very similar barrier heights but distinct geometric configurations, underscoring the necessity of considering all possible routes. In the case of NH_3 dehydrogenation, we observed two predominant paths characterized by the nitrogen-cobalt coordinations being around 0.5 or 1, with a preference for the latter. Similarly, for NH_2, the reaction occurred at either coordination 1 or 2, albeit with a slight favor towards the former. Finally, for NH, there was substantial parity between the two paths at coordination values 1 and 2. If we look instead from the perspective of hydrogen, its total coordination with the surface in the TS region is always around 1, but the reaction occurs preferentially when the reactive hydrogen is in contact with cobalt atoms (Fig. S7).

To gain a deeper understanding of the reaction paths, we conducted a new set of flooding simulations, yielding an ensemble of 100 reactive trajectories per step (Fig. 6). From the distributions reported on the right of the
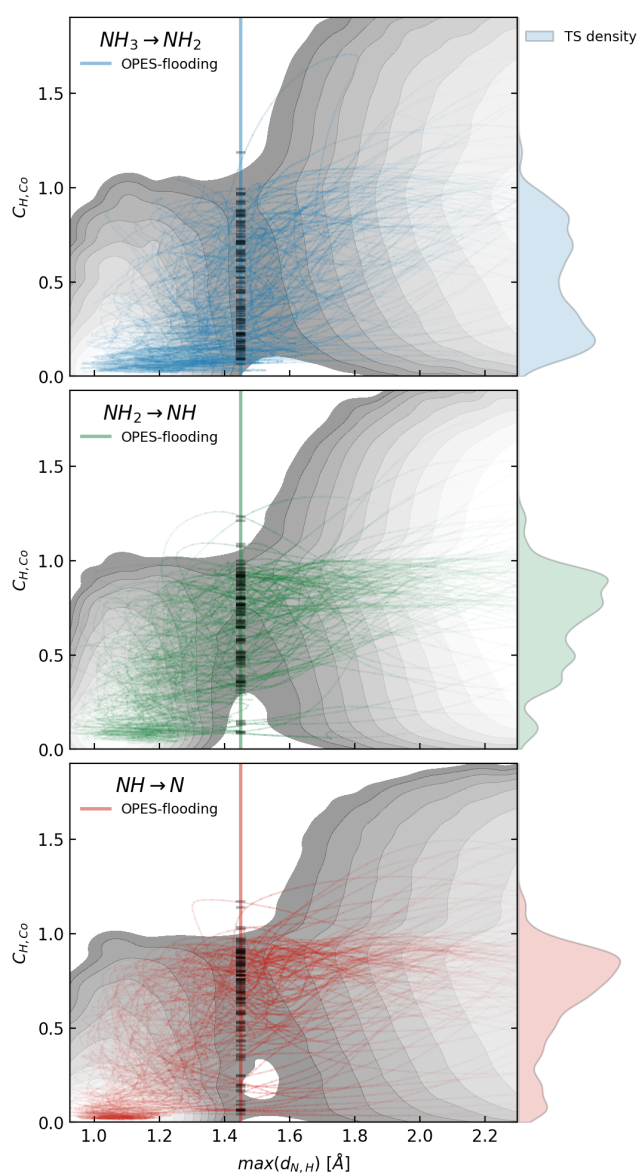
Figure 6. **Reactive path analysis for the dehydrogenation steps.** An ensemble of 100 trajectories is calculated with the OPES-flooding scheme and visualized in the 2D plane defined by $max(d_{N,H})$ and $C_{H,Co}$. Only the 40 fs around the reactive event (determined by $max(d_{N,H})$ crossing 1.45 Å, denoted by a vertical line) are reported for each trajectory. For convenience, we report in the background the free energy surface projected on the same coordinates obtained from the OPES simulations (Fig. S7). On the right axis, we projected the distribution of the coordination number $C_{H,Co}$ when crossing the line at $max(d_{N,H}) = 1.45$ Å(black dashes).

figure, we can learn a qualitatively different behavior between NH_3 and NH_2/NH. Notably, NH_3 trajectories exhibited a broadened distribution in the coordination between H and Co, $C_{H,Co}$, reacting uniformly in contact with either iron ($C_{H,Co} = 0$) and cobalt ($C_{H,Co} = 1$). Conversely, NH_2/NH trajectories displayed a greater flux

8

over higher coordination values, indicating a propensity for dehydrogenation when hydrogen is cobalt-coordinated. See also Fig. S8 for a detailed analysis of TS configurations in terms of N and H coordination. This comprehensive analysis underscores the multiplicity of reaction pathways facilitated by high-temperature dynamic conditions. It also demonstrates the effectiveness of our approach, not only in parsimoniously constructing the potential but also in investigating the process under operative conditions with the same simulations, thus elucidating the reaction mechanism and reconstructing the free energy profile of the different catalytic steps.

CONCLUSIONS

The development of methods that require few QM calculations is essential to enable widespread use of ML-based potentials. In this manuscript, we introduced a data-efficient scheme for constructing reactive potentials. Applied to the study of ammonia decomposition on iron-cobalt alloy catalysts, our protocol effectively captured complex reaction mechanisms composed of multiple channels without prior knowledge of the adsorption sites or the transition states. Remarkably, the number of needed training points (approximately one thousand DFT calculations for each catalytic step) is at least one order of magnitude less than previous work[11,23]. These advances can facilitate the simulation of more realistic systems, both in complexity and underlying level of theory. Our approach achieved this result by relying on both data-efficient architectures and a well-designed protocol for sampling and selecting relevant configurations. Because we do not know in advance which configurations are relevant, especially at high temperatures, we divided the protocol into two phases: first exploring reactive pathways and then obtaining uniform accuracy along them.

The exploratory phase is carried out using flooding simulations that push the system out of the reactants, requiring only collective variables that distinguish reactants and products. To avoid extrapolating into non-physical regions, we performed uncertainty-aware MD based on GPs, which are optimized on-the-fly whenever a new local environment is found. This provided an initial set of reactive structures, most of which lie on the minimum free energy pathways. In the convergence step, we used graph neural networks to extensively sample the reactions with reversible enhanced sampling simulations. Since the query-by-committee criterion was found to under-sample the transition state region, we proposed a novel active learning strategy based on the local environment uncertainty provided by GPs. This way, we extracted a small set of nonredundant configurations from those sampled by the GNN. The result is a uniformly accurate potential along the entire reaction path, achieved with a parsimonious number of DFT calculations.

It is worth noting that these features make our approach general, applicable not only to catalytic processes but also to the broader field of chemical reactions and materials. In addition, ML-based methods could be used to design collective variables[56], thereby reducing the need to identify process-specific CVs. Similarly, path sampling methods could be integrated into the active learning scheme to investigate the mechanism in a more unbiased way[33]. For all these reasons, we believe that our scheme can be effectively used to construct reactive potentials for a wide variety of processes occurring under dynamic conditions, gaining insights into their workings at the atomistic level.

METHODS

Machine learning frameworks

Here we briefly describe the two main architectures that we used in the different stages of our approach to represent the potential energy surface: Gaussian Processes (GPs) and graph neural networks (GNNs).

Gaussian Processes. GPs are kernel-based methods that can be used for regression and probabilistic classification tasks. Starting with the Gaussian Approximation Potential (GAP) method[57], GPs have been successfully used for representing the potential energy surface of extended systems. Here we used the sparse GP implementation of FLARE (Fast Learning of Atomistic Rare Events)[49,58], where the energy is modeled as a sum over atomic contributions that depend on atom-centered local environments: $E = \sum_i \epsilon_i$. These environments are characterized by a set of descriptors based on the atomic cluster expansion (ACE)[50]. What distinguishes GPs from other ML methods is their ability to provide confidence intervals on the predictions. Because of the energy decomposition, we can also measure the variance associated with each atomic contribution ϵ_i, obtaining a measure of the uncertainty of local environments, i.e., how different they are with respect to the training ones. In particular, per each configuration we monitor the maximum local uncertainty defined as:

$$\sigma_{max}^{(GP)} = \max_i \tilde{\sigma}_{\epsilon_i}.$$

where $\tilde{\sigma}_{\epsilon_i}$ is the variance on the local environment rescaled by the kernel hyperparameter to obtain a dimensionless quantity that ranges between 0 and 1 (see Ref.[49] for the details). The uncertainty estimate can be used to guide the active learning process by selecting the new configurations to be labeled with DFT single point calculations. In particular, this can be done on-the-fly by running "uncertainty-aware" MD simulations, in which the GP model is used to perform the next MD step only if the uncertainty is low; otherwise, a DFT calculation is run and the GP is updated[58]. The same property can be used to screen an existing dataset of configurations, using the uncertainty estimate to monitor the presence of new local environments against those in the training set. Again,

the new configuration is computed at the DFT level and added to the GP model. On the other hand, GPs scale unfavorably with the number of points in the training dataset, making them feasible only for small datasets ($\lesssim$ thousands of atomic configurations).

Graph Neural Networks. Compared with kernel-based methods, the parameterization of the potential energy surface by neural networks allows for greater expressiveness[53] and eliminates the scaling with the number of training data points. In the special case of GNNs, the description of local environments (the "message") is propagated across neighbors, resulting in a larger effective receptive field than with kernel methods or standard neural networks. In particular, we used MACE[55], one of the state-of-the-art architectures, excelling in both in-domain and out-domain predictions and demonstrating remarkable data efficiency[59]. MACE accomplishes this through a higher-order equivariant message passing scheme related to ACE, which enables accurate modeling of complex atomic interactions while maintaining efficiency. On the other hand, GNNs do not have a direct estimate of the uncertainty; this can be approximated by the standard deviation of the predictions of a committee of models trained on different partitions of the training dataset. In particular, we focus on the standard deviation of the components α of the forces acting on the i-th atom over a committee of N different models: $\sigma_{i,\alpha} = \sqrt{\frac{\sum_k (f_{i,\alpha} - \bar{f}_{i,\alpha})^2}{N}}$ and monitor the maximum value per each atomic configuration:

$$\sigma_{max}^{(MACE)} = \max_{i,\alpha} \sigma_{i,\alpha}.$$

Enhanced sampling simulations

Even with machine learning potentials, many processes, such as chemical reactions, occur on time scales too long to be simulated with standard MD. For this reason, several enhanced sampling methods[60] have been developed. In particular, a widely used family is based on the identification of selected collective variables (CVs), which describe the most important degrees of freedom involved in the process. Introducing an external potential as a function of these CVs enhances their fluctuations and accelerates the sampling of the transition. Notable examples are Umbrella Sampling[61] and Metadynamics[24].

On-the-fly Probability Enhanced Sampling (OPES)[43] is a recent evolution of Metadynamics, which converges faster and requires fewer hyperparameters. Given a collective variable s, OPES first reconstructs its probability distribution at equilibrium $P(s)$, using a Kernel Density Estimate (KDE). Then, it introduces a potential to guide the system to sample a given target distribution $p^{tg}(s)$:

$$V(s) = -\frac{1}{\beta} \log \frac{p^{tg}(s)}{P(s)}$$

A typical choice of the target distribution is the well-tempered one: $p^{tg}(s) \propto P(s)^{\frac{1}{\gamma}}$, which lowers the free energy barriers by a factor γ, and correspond to the same distribution sampled in Well-Tempered Metadynamics[62]. Since the target is not known a priori, the probability and thus the bias are constructed iteratively during the simulation. Furthermore, a useful feature is the ability to limit the amount of bias deposited. In case one does not know the height of the barrier to be overcome, an exploratory variant is also available[63], which can be used to quickly explore the free energy landscape. When the OPES bias is converged to a quasi-static value, the free energy along the biased CV or any other variables can be recovered from a simple Umbrella Sampling-like reweighting scheme[43].

OPES flooding. OPES-flooding is a variant of the OPES method, in which the potential is applied only in the reactant basin. In OPES this can be ensured with the barrier hyperparameter that sets the maximum amount of deposited bias. In addition, this can be enforced by restricting the application of the potential within a certain region of the CV space that bounds the reactants[51]. In this way, free energy paths that are sampled are more likely to be close to the lowest free energy ones. Once the system has overcome the free energy barrier and reached the product basin, the simulation is terminated. By performing a set of flooding simulations we can also obtain statistics on the reactive pathways and characterize them.

Construction of the reactive ML potential

Here we describe the protocol used for the collection of the configurations for training the reactive potential, focusing on the active learning policies used in the different stages to select the relevant samples.

Preliminary reactants configurations. We ran uncertainty-aware MD through the on-the-fly code implemented in FLARE[64], using LAMMPS[65] as MD engine. For each intermediate (NH_3, NH_2, NH, N/N_2, H), we trained a different GP model. These models were used in parallel to perform a series of simulations starting with the adsorbate in different initial conditions to study the interaction of the intermediates with the surface. The criterion for selecting the configurations is based on the maximum uncertainty over the local environments predicted by FLARE. Whenever this exceeds a given threshold, a DFT calculation is performed and a (sparse) set of local environments is added to the training dataset. In the initial phase, a small threshold of 0.02 was used to avoid extrapolating into unphysical regions. We ran about 10 iterations for each reactant until the simulations proceeded for about 1 ps without DFT calculations. Furthermore, to achieve a more extensive sampling, we performed OPES simulations to sample the different adsorption sites, by enhancing the coordination of nitrogen with the surface, and the diffusion on the surface, by biasing the x,y coordinates of the center of mass; see below for the definition of the CVs.

Pathways discovery with GPs. For the reactive pathway discovery step, we performed a set of OPES flooding-like

simulations with uncertainty-aware MD, biasing collective variables capable of distinguishing reactants from products (e.g., N-N distance for N2 dissociation; see below). For each reaction step (N$_2 \to$2N, NH$_3 \to$NH$_2$+H, NH$_2 \to$NH+H, NH$\to$N+H), we performed 30 iterations, stopping the simulation each time the molecule reacted and starting again to sample a new reactive event. As a threshold for the maximum local uncertainty, we used a value of 0.1 for the first 20 iterations, then lowered it to 0.075 for the remaining ones in order to continue labeling a few structures from each path.

Uniform accuracy with GNNs. For the second step, we trained a single model with MACE[66] using all the configurations collected earlier. With this model we ran long MD simulations (3 ns) for each reaction step, sampling tens of reactive events within each. We then calculated the uncertainty of the predictions $\sigma_{max}^{(MACE)}$ through a committee of 4 models, which we used to select a subset of configurations whose uncertainty was above a given threshold ("query-by-committee" selection). The threshold was chosen based on the peak of the maximum uncertainty distribution, equal to 90 meV Å^{-1} for $N_2 \to 2N$ and 50 meV Å^{-1} for NH$_x \to$NH$_{x-1}$+H. We also calculated the uncertainty as a function of the CV by dividing its interval into 10 bins and taking the mean of $\sigma_{max}^{(MACE)}$ within each. To pick the configurations (from those selected with the query-by-committee approach) on which to actually perform a DFT calculation, we evaluated the maximum uncertainty on the local environments $\sigma_{max}^{(GP)}$ from the GP models built in the previous step. We then performed QM calculations for those configurations whose GP uncertainty exceeded a certain threshold (0.1), thus selecting only structurally different environments. By proceeding incrementally and adding new points to the GP model before evaluating the uncertainty on the next structure, we were able to filter redundant configurations. In addition, to target especially the worst-described configurations, we fed the GP model the structures divided per bin of the CV, ordered by decreasing mean $\sigma_{max}^{(MACE)}$. For the dehydrogenation reactions, in addition to the reactive configurations we also performed calculations of the products with the query-by-committee criterion and the threshold specified above, since they were not sampled in the preliminary dataset.

Computational details

DFT calculations. All DFT calculations used to build the reference database were performed using the PWscf code of Quantum ESPRESSO[67–69]. Exchange-correlation effects were treated using the generalized-gradient approximation with the Perdew-Burke-Ernzerhof (PBE) functional[70]. We employed ultrasoft pseudopotentials selected from the SSSP PBE Precision v1.3.0 library[71] with 1, 5, 16, and 17 valence electrons for H, N, Fe, and Co, respectively. The wave function and the charge density cutoff were set at 90 and 1080 Ry. The Marzari-Vanderbilt

cold smearing technique[72] with a Gaussian spreading of 0.04 Ry was used to treat the collinear spin-polarised electronic occupations. Initial spin polarizations were set to 0.6 for Fe|Co. To ensure the consistency of the dataset, spurious high-energy non-ferromagnetic configurations were removed from the dataset. A vacuum layer of at least 12 Å was included in the z-direction of all the slab models to prevent self-interaction effects. The Brillouin zone was sampled using a Monkhorst-Pack k-point grid[73] with a maximum k-spacing of 0.25 Å^{-1}, and the SCF convergence threshold was set to 10^{-6} Ry.

FLARE. The local environments are described through the ACE B2 descriptors, using a basis expansion with $N_{max} = 8$, and $l_{max} = 3$. We used a cutoff equal to 5.5 Å for Fe and Co and 4 Å for the interaction with the other species. The kernel used to compare the environments was a normalized dot product, squared.

MACE. All the GNN models in this work were constructed using MACE version 0.3.0[66]. The models employed a cutoff radius of 6 Å for the atomic environments with 256 channels and L=0, finding a balance between accuracy and computational efficiency. The dataset was split into training/validation subsets with a ratio of 95:5, and the model was optimized with AMSGrad using a learning rate of 0.01 and a batch size of 4. The performance was evaluated on energy and forces with a weighted root mean square error (RMSE) loss function. The weights of the energy and forces were initially set to 1 and 100, respectively. In the last ̃25% epochs, the weight of energy has been increased by a factor of 10 three times. The optimization was interrupted via an early stopping criterion with a patience of 200 epochs.

Molecular dynamics. MD simulations were performed using the Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS) software[65], supplemented by FLARE[74], MACE v0.3.0[66], and PLUMED v2.9[75]. The canonical sampling through velocity rescaling thermostat[76] with a coupling constant of 50 fs was employed to control the temperature in all the simulations. During the construction of the preliminary reactants potentials, we simulated FeCo slabs cut along the 110 surface containing 60 (3x4x5) and 144 (4x6x6) atoms. A vacuum region of at least 12 Å was added in all configurations to avoid self-interaction effects. The slab was simulated in both bulk-terminated (by fixing the two bottom layers) and double-interface (by fixing a central layer) conditions. In the latter, adsorbate molecules were placed on both sides. The surface dynamics was simulated at 700K and then 800K, while the interaction with the different reaction intermediates (NH$_3$, NH$_2$, NH, N/N$_2$, H) was simulated at 700 K. At this stage, atomic interactions were described by FLARE, and the timestep was set to 1 fs. In the reaction discovery stage, as well as during the GNN active learning and the final simulations, we simulated 120 atoms (4x6x5) bulk-terminated (2-fixed layers) FeCo(110) slabs at T = 700 K. For the final simulations, we repeated the simulations with a larger system (800 atoms) to rule out possible finite-size effects. Periodic boundary conditions

were applied in the x and y directions, while a reflecting wall was implemented over the surface in the z-direction.

Enhanced sampling. Here, we first describe the different CVs used and, later, the details of the OPES simulations employed in the different stages. In addition to the distance between different atoms, we focused on the coordination numbers to measure both the interaction between the adsorbate and the substrate atoms and to distinguish the different NH_x species. The coordination is defined in a continuous and differentiable way as:

$$C_{X,Y} = \sum_{i \in X} \sum_{j \in Y} s(r_{ij}), \quad s(r_{ij}) = \frac{1 - (\frac{r_{ij}-d_0}{r_0})^n}{1 - (\frac{r_{ij}-d_0}{r_0})^m}$$

where r_{ij} is the distance between the i-th atoms of species X and the j-th atoms of species Y, and r_0, d_0, n, and m, are tunable parameters. For the coordination between N and Fe, Co, and Fe|Co, we set the parameters $d_0 = 1$ Å and $r_0 = 1.5$ Å $n = 6$, $m = 12$, while for the coordination between H and Fe, Co, and Fe|Co, we used $d_0 = 0.8$ Å $r_0 = 1.0$ Å. Finally, for the coordination between N and H, we used $d_0 = 0.7$ Å $r_0 = 0.8$ Å $n = 5$, $m = 7$.

In the preliminary step, we used OPES to enhance a) the exploration of the different adsorption states by using as CV the coordination between N and Fe/Co atoms and b) the diffusion of the adsorbed molecule by using the x and y components of its center of mass.

In the exploratory stage, we iteratively performed OPES-flooding simulations to harvest an ensemble of reactive pathways. For N_2 dissociation, we used the distance between the two N atoms $d_{N,N}$ and the coordination $C_{N,Fe|Co}$ as CVs, while for the hydrogenation steps, we used the inverse of the planar distance between N and one H atom. The probability estimate was updated every 50 steps, and the the barrier was chosen to be 50 kJ^{-1}, in order to apply the bias only in the reactant region. Simulations were interrupted after the occurrence of the reaction, namely when $d_{N,N}$ exceeded 2.5 Å or when the coordination $C_{N,H}$ was decreased by 1 with respect to the equilibrium value.

During the active learning stage with MACE and the final production simulations, we used OPES with the standard well-tempered target distribution. To study nitrogen recombination, we use the distance $d_{N,N}$ as CVs. The barrier parameter was set to 200 kJ mol^{-1}, and the bias was updated every 200 steps. For the first dehydrogenation step ($NH_3 \rightarrow NH_2+H$), we biased the simulation along three CVs: C_{N-H}, $C_{N,Fe|Co}$, and the angular variable Θ defined in Ref.[11]. For the second reaction ($NH_2 \rightarrow NH+H$), we used $C_{N,H}$ and $C_{N,Fe|Co}$, while for the last step ($NH \rightarrow N+H$), we employed the distance $d_{N,H}$ and $C_{N,Fe|Co}$. The barrier parameters were set equal to 80, 60, and 100 kJ mol^{-1}, respectively, and the bias was updated every 200 steps. Harmonic restraints were applied in all reactions to facilitate reversible sampling. For the $2N \rightarrow N_2$ process, the restraint was applied if $d_{N,N} \leq 1.2$ or $d_{N,N} \geq 3.2$ Å, while for the dehydro-genation steps if $d_{N,H} \leq 2.5$Å, with an elastic constant of 5000 kJ mol^{-1}Å^{-2}.

Finally, to study in detail the reactive pathways, we perform a set of 100 OPES flooding simulations for each dehydrogenation step. The CV used for all the steps was the coordination $C_{N,H}$. To avoid biasing the TS region, we exclude from bias deposition the regions where $C_{N,H}$ is lower than 2.82, 1.85, and 0.9 for the three hydrogenation steps, using an OPES barrier parameter equal to 70, 30, and 80 kJ mol^{-1}, respectively. All simulations were stopped once the reaction occurred, terminating the simulation when the maximum NH distance exceeded 1.9 Å.

AUTHOR CONTRIBUTIONS

L.B. designed the project; L.B and S.P. performed the calculations, discussed the results, and wrote the manuscript.

DATA AVAILABILITY

Input files of the simulations will be released in the PLUMED-NEST repository.

CODE AVAILABILITY

LAMMPS (www.lammps.org), Quantum ESPRESSO (www.quantum-espresso.org), PLUMED (www.plumed.org), FLARE (github.com/mir-group/flare) and MACE (github.com/ACEsuit/mace) are open-source codes.

REFERENCES

[1] R. Schlögl, "Heterogeneous Catalysis," Angewandte Chemie International Edition **54**, 3465–3520 (2015).

[2] X. Shi, X. Lin, R. Luo, S. Wu, L. Li, Z.-J. Zhao, and J. Gong, "Dynamics of heterogeneous catalytic processes at operando conditions," JACS Au **1**, 2100–2120 (2021).

[3] M. S. Spencer, "Stable and metastable metal surfaces in heterogeneous catalysis," Nature 1986 323:6090 **323**, 685–687 (1986).

[4] B. Eren, D. Zherebetskyy, L. L. Patera, C. H. Wu, H. Bluhm, C. Africh, L. W. Wang, G. A. Somorjai, and M. Salmeron, "Activation of Cu(111) surface by decomposition into nanoclusters driven by CO adsorption," Science **351**, 475–478 (2016).

[5] L. Grajciar, C. J. Heard, A. A. Bondarenko, M. V. Polynski, J. Meeprasert, E. A. Pidko, and P. Nachtigall, "Towards operando computational modeling in heterogeneous catalysis," Chemical Society Reviews **47**, 8307–8348 (2018).

[6] G. Piccini, M. S. Lee, S. F. Yuk, D. Zhang, G. Collinge, L. Kollias, M. T. Nguyen, V. A. Glezakou, and R. Rousseau, "Ab initio molecular dynamics with enhanced sampling in heterogeneous catalysis," Catalysis Science & Technology **12**, 12–37 (2022).

[7] V. Van Speybroeck, M. Bocus, P. Cnudde, and L. Vanduyfhuys, "Operando modeling of zeolite-catalyzed reactions using first-principles molecular dynamics simulations," ACS catalysis **13**, 11455–11493 (2023).

[8] O. T. Unke, S. Chmiela, H. E. Sauceda, M. Gastegger, I. Poltavsky, K. T. Schütt, A. Tkatchenko, and K. R. Müller, "Machine Learning Force Fields," Chemical Reviews **121**, 10142–10186 (2021).

[9] G. Sun and P. Sautet, "Metastable Structures in Cluster Catalysis from First-Principles: Structural Ensemble in Reaction Conditions and Metastability Triggered Reactivity," Journal of the American Chemical Society **140**, 2812–2820 (2018).

[10] Z. Zhang, B. Zandkarimi, and A. N. Alexandrova, "Ensembles of Metastable States Govern Heterogeneous Catalysis on Dynamic Interfaces," Accounts of Chemical Research **53**, 447–458 (2020).

[11] S. Perego, L. Bonati, S. Tripathi, M. Parrinello, and A. Simulations, "How dynamics changes ammonia cracking on iron surfaces," (2024), 10.26434/CHEMRXIV-2024-0S8C9.

[12] I. Batatia, S. Batzner, D. P. Kovács, A. Musaelian, G. N. C. Simm, R. Drautz, C. Ortner, B. Kozinsky, and G. Csányi, "The Design Space of E(3)-Equivariant Atom-Centered Interatomic Potentials," (2022).

[13] J. I. Falk, L. Bonati, A. Simulations, P. Novelli, M. Parrinello, and M. Pontil, "Transfer learning for atomistic simulations using GNNs and kernel mean embeddings," Advances in Neural Information Processing Systems **36**, 29783–29797 (2023).

[14] L. Chanussot, A. Das, S. Goyal, T. Lavril, M. Shuaibi, M. Riviere, K. Tran, J. Heras-Domingo, C. Ho, W. Hu, A. Palizhati, A. Sriram, B. Wood, J. Yoon, D. Parikh, C. L. Zitnick, and Z. Ulissi, "Open Catalyst 2020 (OC20) Dataset and Community Challenges," ACS Catalysis **11**, 6059–6072 (2021).

[15] I. Batatia, P. Benner, Y. Chiang, A. M. Elena, D. P. Kovács, J. Riebesell, X. R. Advincula, M. Asta, M. Avaylon, W. J. Baldwin, F. Berger, N. Bernstein, A. Bhowmik, S. M. Blau, V. Cărare, J. P. Darby, S. De, D. Pia, V. L. Deringer, R. Elijošius, Z. El-Machachi, F. Falcioni, E. Fako, A. C. Ferrari, A. Genreith-Schriever, J. George, R. E. A. Goodall, C. P. Grey, P. Grigorev, S. Han, W. Handley, H. H. Heenen, K. Hermansson, C. Holm, S. Hofmann, J. Jaafar, K. S. Jakob, H. Jung, V. Kapil, A. D. Kaplan, N. Karimitari, J. R. Kermode, N. Kroupa, J. Kullgren, M. C. Kuner, D. Kuryla, G. Liepuoniute, J. T. Margraf, I.-B. Magdău, A. Michaelides, J. H. Moore, A. A. Naik, S. P. Niblett, S. W. Norwood, N. O'neill, C. Ortner, K. A. Persson, K. Reuter, A. S. Rosen, L. L. Schaaf, C. Schran, B. X. Shi, E. Sivonxay, T. K. Stenczel, V. Svahn, C. Sutton, T. D. Swinburne, J. Tilly, C. Van Der Oord, S. Vargas, E. Varga-Umbrich, T. Vegge, M. Vondrák, Y. Wang, W. C. Witt, F. Zills, and G. Csányi, "A foundation model for atomistic materials chemistry," (2023).

[16] J. S. Smith, B. Nebgen, N. Lubbers, O. Isayev, and A. E. Roitberg, "Less is more: Sampling chemical space with active learning," The Journal of chemical physics **148** (2018).

[17] C. Schran, K. Brezina, and O. Marsalek, "Committee neural network potentials control generalization errors and enable active learning," The Journal of Chemical Physics **153** (2020).

[18] S. J. Ang, W. Wang, D. Schwalbe-Koda, S. Axelrod, and R. Gómez-Bombarelli, "Active learning accelerates ab initio molecular dynamics on reactive energy surfaces," Chem **7**, 738–751 (2021).

[19] L. L. Schaaf, E. Fako, S. De, A. Schäfer, and G. Csányi, "Accurate energy barriers for catalytic reaction pathways: an automatic training protocol for machine learning force fields," npj Computational Materials 2023 9:1 **9**, 1–10 (2023).

[20] H. Jung, L. Sauerland, S. Stocker, K. Reuter, and J. T. Margraf, "Machine-learning driven global optimization of surface adsorbate geometries," npj Computational Materials 2023 9:1 **9**, 1–8 (2023).

[21] Y. G. Wang, D. Mei, V. A. Glezakou, J. Li, and R. Rousseau, "Dynamic formation of single-atom catalytic active sites on ceria-supported gold nanoparticles," Nature Communications 2015 6:1 **6**, 1–8 (2015).

[22] W. L. Li, C. N. Lininger, K. Chen, V. Vaissier Welborn, E. Rossomme, A. T. Bell, M. Head-Gordon, and T. Head-Gordon, "Critical Role of Thermal Fluctuations for CO Binding on Electrocatalytic Metal Surfaces," JACS Au **1**, 1708–1718 (2021).

[23] L. Bonati, D. Polino, C. Pizzolitto, P. Biasi, R. Eckert, S. Reitmeier, R. Schlögl, and M. Parrinello, "The role of dynamics in heterogeneous catalysis: Surface diffusivity and N2 decomposition on Fe(111)," Proceedings of the National Academy of Sciences of the United States of America **120**, e2313023120 (2023).

[24] A. Laio and M. Parrinello, "Escaping free-energy minima," Proceedings of the national academy of sciences **99**, 12562–12566 (2002).

[25] D. Yoo, J. Jung, W. Jeong, and S. Han, "Metadynamics sampling in atomic environment space for collecting training data for machine learning potentials," npj Computational Materials **7**, 131 (2021).

[26] J. Xu, X.-M. Cao, and P. Hu, "Accelerating metadynamics-based free-energy calculations with adaptive machine learning potentials," Journal of chemical theory and computation **17**, 4465–4476 (2021).

[27] O. Valsson, P. Tiwary, and M. Parrinello, "Enhancing Important Fluctuations: Rare Events and Metadynamics from a Conceptual Viewpoint," Annual Review of Physical Chemistry **67**, 159–184 (2016).

[28] L. Bonati and M. Parrinello, "Silicon Liquid Structure and Crystal Nucleation from Ab Initio Deep Metadynamics," Physical Review Letters **121** (2018), 10.1103/PhysRevLett.121.265701.

[29] H. Niu, L. Bonati, P. M. Piaggi, and M. Parrinello, "Ab initio phase diagram and nucleation of gallium," Nature Communications **11**, 1–9 (2020).

[30] O. Abou El Kheir, L. Bonati, M. Parrinello, and M. Bernasconi, "Unraveling the crystallization kinetics of the ge2sb2te5 phase change compound with a machine-learned interatomic potential," npj Computational Materials **10**, 33 (2024).

[31] M. Galib and D. T. Limmer, "Reactive uptake of n2o5 by atmospheric aerosol is dominated by interfacial processes," Science **371**, 921–925 (2021).

[32] M. Yang, L. Bonati, D. Polino, and M. Parrinello, "Using metadynamics to build neural network potentials for reactive events: the case of urea decomposition in water," Catalysis Today **387**, 143–149 (2022).

[33] J. Zhang, O. Zhang, L. Bonati, and T. Hou, "Combining transition path sampling with data-driven collective variables through a reactivity-biased shooting algorithm," Journal of Chemical Theory and Computation (2024).

[34] X. Guan, J. P. Heindel, T. Ko, C. Yang, and T. Head-Gordon, "Using machine learning to go beyond potential energy surface benchmarking for chemical reactivity," Nature Computational Science 2023 3:11 **3**, 965–974 (2023).

[35] M. Yang, U. Raucci, and M. Parrinello, "Reactant-induced dynamics of lithium imide surfaces during the ammonia decomposition process," Nature Catalysis 2023 6:9 **6**, 829–836 (2023).

[36] S. Tripathi, L. Bonati, S. Perego, and M. Parrinello, "How Poisoning Is Avoided in a Step of Relevance to the Haber–Bosch Catalysis," ACS Catalysis , 4944–4950 (2024).

[37] F. Mambretti, U. Raucci, M. Yang, and M. Parrinello, "How Does Structural Disorder Impact Heterogeneous Catalysts? The Case of Ammonia Decomposition on Non-stoichiometric Lithium Imide," ACS Catalysis **16**, 1252–1256 (2024).

[38] M. Kulichenko, K. Barros, N. Lubbers, Y. W. Li, R. Messerly, S. Tretiak, J. S. Smith, and B. Nebgen, "Uncertainty-driven dynamics for active learning of interatomic potentials," Nature Computational Science **3**, 230–239 (2023).

[39] C. van der Oord, M. Sachs, D. P. Kovács, C. Ortner, and

G. Csányi, "Hyperactive learning for data-driven interatomic potentials," npj Computational Materials 2023 9:1 **9**, 1–14 (2023).

[40] A. R. Tan, J. C. B. Dietschreit, and R. Gomez-Bombarelli, "Enhanced sampling of robust molecular datasets with uncertainty-based collective variables," (2024).

[41] V. Zaverkin, D. Holzmüller, H. Christiansen, F. Errica, F. Alesiani, M. Takamoto, M. Niepert, and J. Kästner, "Uncertainty-biased molecular dynamics for learning uniformly accurate interatomic potentials," npj Computational Materials **10**, 83 (2024).

[42] S. Vandenhaute, M. Cools-Ceuppens, S. DeKeyser, T. Verstraelen, and V. Van Speybroeck, "Machine learning potentials for metal-organic frameworks using an incremental learning approach," npj Computational Materials 2023 9:1 **9**, 1–8 (2023).

[43] M. Invernizzi and M. Parrinello, "Rethinking Metadynamics: From Bias Potentials to Probability Distributions," Journal of Physical Chemistry Letters **11**, 2731–2736 (2020).

[44] C. E. Rasmussen, "Gaussian processes in machine learning," in *Summer school on machine learning* (Springer, 2003) pp. 63–71.

[45] I. Lucentini, X. Garcia, X. Vendrell, and J. Llorca, "Review of the Decomposition of Ammonia to Generate Hydrogen," (2021).

[46] T. E. Bell and L. Torrente-Murciano, "H2 Production via Ammonia Decomposition Using Non-Noble Metal Catalysts: A Review," Topics in Catalysis 2016 59:15 **59**, 1438–1457 (2016).

[47] Y. Wang, J. Qian, Z. Fang, M. R. Kunz, G. Yablonsky, A. Fortunelli, W. A. Goddard, and R. R. Fushimi, "Understanding Reaction Networks through Controlled Approach to Equilibrium Experiments Using Transient Methods," Journal of the American Chemical Society **143**, 10998–11006 (2021).

[48] S. Chen, J. Jelic, D. Rein, S. Najafishirtari, F.-P. Schmidt, F. Girgsdies, L. Kang, A. Wandzilak, A. Rabe, D. E. Doronkin, *et al.*, "Highly loaded bimetallic iron-cobalt catalysts for hydrogen release from ammonia," Nature Communications **15**, 871 (2024).

[49] J. Vandermause, Y. Xie, J. S. Lim, C. J. Owen, and B. Kozinsky, "Active learning of reactive bayesian force fields applied to heterogeneous catalysis dynamics of h/pt," Nature Communications **13**, 5183 (2022).

[50] R. Drautz, "Atomic cluster expansion for accurate and transferable interatomic potentials," Physical Review B **99**, 014104 (2019).

[51] D. Ray, N. Ansari, V. Rizzi, M. Invernizzi, and M. Parrinello, "Rare Event Kinetics from Adaptive Bias Enhanced Sampling," Journal of Chemical Theory and Computation **18**, 6500–6509 (2022).

[52] U. Raucci, V. Rizzi, and M. Parrinello, "Discover, Sample, and Refine: Exploring Chemistry with Enhanced Sampling Techniques," Journal of Physical Chemistry Letters **13**, 1424–1430 (2022).

[53] C. J. Owen, S. B. Torrisi, Y. Xie, S. Batzner, K. Bystrom, J. Coulter, A. Musaelian, L. Sun, and B. Kozinsky, "Complexity of many-body interactions in transition metals via machine-learned force fields from the tm23 data set," npj Computational Materials **10**, 92 (2024).

[54] X. Fu, Z. Wu, W. Wang, T. Xie, S. Keten, R. Gomez-Bombarelli, and T. Jaakkola, "Forces are not enough: Benchmark and critical evaluation for machine learning force fields with molecular simulations," arXiv preprint arXiv:2210.07237 (2022).

[55] I. Batatia, D. P. Kovacs, G. Simm, C. Ortner, and G. Csányi, "Mace: Higher order equivariant message passing neural networks for fast and accurate force fields," Advances in Neural Information Processing Systems **35**, 11423–11436 (2022).

[56] L. Bonati, E. Trizio, A. Rizzi, and M. Parrinello, "A unified framework for machine learning collective variables for enhanced sampling simulations: mlcolvar," The Journal of Chemical Physics **159** (2023).

[57] A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, "Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons," Physical review letters **104**, 136403 (2010).

[58] J. Vandermause, S. B. Torrisi, S. Batzner, Y. Xie, L. Sun, A. M. Kolpak, and B. Kozinsky, "On-the-fly active learning of interpretable bayesian force fields for atomistic rare events," npj Computational Materials **6**, 20 (2020).

[59] D. P. Kovacs, I. Batatia, E. S. Arany, and G. Csanyi, "Evaluation of the MACE Force Field Architecture: from Medicinal Chemistry to Materials Science," The Journal of Chemical Physics **159**, 44118 (2023).

[60] J. Hénin, T. Lelièvre, M. R. Shirts, O. Valsson, and L. Delemotte, "Enhanced Sampling Methods for Molecular Dynamics Simulations [Article v1.0]," Living Journal of Computational Molecular Science **4**, 1583–1583 (2022).

[61] G. M. Torrie and J. P. Valleau, "Nonphysical sampling distributions in monte carlo free-energy estimation: Umbrella sampling," Journal of computational physics **23**, 187–199 (1977).

[62] A. Barducci, G. Bussi, and M. Parrinello, "Well-tempered metadynamics: a smoothly converging and tunable free-energy method," Physical review letters **100**, 020603 (2008).

[63] M. Invernizzi and M. Parrinello, "Exploration vs Convergence Speed in Adaptive-Bias Enhanced Sampling," Journal of Chemical Theory and Computation **18**, 3988–3996 (2022).

[64] J. Vandermause, S. B. Torrisi, S. Batzner, Y. Xie, L. Sun, A. M. Kolpak, and B. Kozinsky, "On-the-fly active learning of interpretable Bayesian force fields for atomistic rare events," npj Computational Materials 2020 6:1 **6**, 1–11 (2020).

[65] A. P. Thompson, H. M. Aktulga, R. Berger, D. S. Bolintineanu, W. M. Brown, P. S. Crozier, P. J. in 't Veld, A. Kohlmeyer, S. G. Moore, T. D. Nguyen, R. Shan, M. J. Stevens, J. Tranchida, C. Trott, and S. J. Plimpton, "LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales," Computer Physics Communications **271**, 108171 (2022).

[66] I. Batatia, D. P. Kovács, G. N. C. Simm, C. Ortner, and G. Csányi, "MACE: Higher Order Equivariant Message Passing Neural Networks for Fast and Accurate Force Fields," Advances in Neural Information Processing Systems **35**, 11423–11436 (2022).

[67] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, A. Dal Corso, S. De Gironcoli, S. Fabris, G. Fratesi, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos, N. Marzari, F. Mauri, R. Mazzarello, S. Paolini, A. Pasquarello, L. Paulatto, C. Sbraccia, S. Scandolo, G. Sclauzero, A. P. Seitsonen, A. Smogunov, P. Umari, and R. M. Wentzcovitch, "QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials," Journal of Physics: Condensed Matter **21**, 395502 (2009).

[68] P. Giannozzi, O. Andreussi, T. Brumme, O. Bunau, M. Buongiorno Nardelli, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, M. Cococcioni, N. Colonna, I. Carnimeo, A. Dal Corso, S. De Gironcoli, P. Delugas, R. A. Distasio, A. Ferretti, A. Floris, G. Fratesi, G. Fugallo, R. Gebauer, U. Gerstmann, F. Giustino, T. Gorni, J. Jia, M. Kawamura, H. Y. Ko, A. Kokalj, E. Kücükbenli, M. Lazzeri, M. Marsili, N. Marzari, F. Mauri, N. L. Nguyen, H. V. Nguyen, A. Otero-De-La-Roza, L. Paulatto, S. Poncé, D. Rocca, R. Sabatini, B. Santra, M. Schlipf, A. P. Seitsonen, A. Smogunov, I. Timrov, T. Thonhauser, P. Umari, N. Vast, X. Wu, and S. Baroni, "Advanced capabilities for materials modelling with Quantum ESPRESSO," Journal of Physics: Condensed Matter **29**, 465901 (2017).

[69] P. Giannozzi, O. Baseggio, P. Bonfà, D. Brunato, R. Car, I. Carnimeo, C. Cavazzoni, S. De Gironcoli, P. Delugas, F. Ferrari Ruffino, A. Ferretti, N. Marzari, I. Timrov, A. Urru, and S. Baroni, "Quantum ESPRESSO toward the exascale," Journal of Chemical Physics **152** (2020), 10.1063/5.0005082/1058748.

[70] J. P. Perdew, K. Burke, and M. Ernzerhof, "Generalized Gradient Approximation Made Simple," Physical Review Letters **77**, 3865 (1996).

[71] G. Prandini, A. Marrazzo, I. E. Castelli, N. Mounet, and N. Marzari, "Precision and efficiency in solid-state pseudopotential calculations," npj Computational Materials 2018 4:1 **4**, 1–13 (2018).

[72] N. Marzari, D. Vanderbilt, A. De Vita, and M. C. Payne, "Thermal Contraction and Disordering of the Al(110) Surface," Physical Review Letters **82**, 3296 (1999).

[73] H. J. Monkhorst and J. D. Pack, "Special points for Brillouin-zone integrations," Physical Review B **13**, 5188 (1976).

[74] Y. Xie, J. Vandermause, L. Sun, A. Cepellotti, and B. Kozinsky, "Bayesian force fields from active learning for simulation of inter-dimensional transformation of stanene," npj Computational Materials 2021 7:1 **7**, 1–10 (2021).

[75] G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, and G. Bussi, "PLUMED 2: New feathers for an old bird," Computer Physics Communications **185**, 604–613 (2014).

[76] G. Bussi, D. Donadio, and M. Parrinello, "Canonical sampling through velocity rescaling," Journal of Chemical Physics **126**, 14101 (2007).