

Supporting Information: Data efficient machine learning potentials for modeling catalytic reactivity via active learning and enhanced sampling

Simone Perego^{1,2} and Luigi Bonati^{1, a)}

¹⁾Atomistic Simulations, Italian Institute of Technology, 16156 Genova, Italy

²⁾Department of Materials Science, University of Milano-Bicocca, 20126 Milano, Italy

LIST OF TABLES

1	Dataset composition	2
---	-------------------------------	---

LIST OF FIGURES

1	Snapshots of NH_x adsorbates on the FeCo (110) surface	3
2	OPES-flooding iterative discovery of reactive pathways (NH_x)	4
3	Uncertainty along the CV and query-by-committee distribution	5
4	Correlation between GP and GNN uncertainties	5
5	MACE relative uncertainty before/after DEAL	6
6	MACE uncertainty before/after DEAL (NH_x)	7
7	DEAL: selected configurations by the different variants (pre-train GP vs from scratch)	8
8	DEAL: force accuracy along the reactive pathway (pre-train GP vs from scratch)	8
9	MACE forces accuracy of the final stage divided per chemical species	9
10	MACE energy accuracy along the reactive pathway	9
11	MACE forces accuracy along the reactive pathways (NH_x)	10
12	MACE forces accuracy vs uncertainty	11
13	Comparison between DeepMD and MACE	12
14	Free energy surfaces of dehydrogenation steps	13
15	Transition state analysis.	14

^{a)}luigi.bonati@iit.it

I. SUPPLEMENTARY TABLES

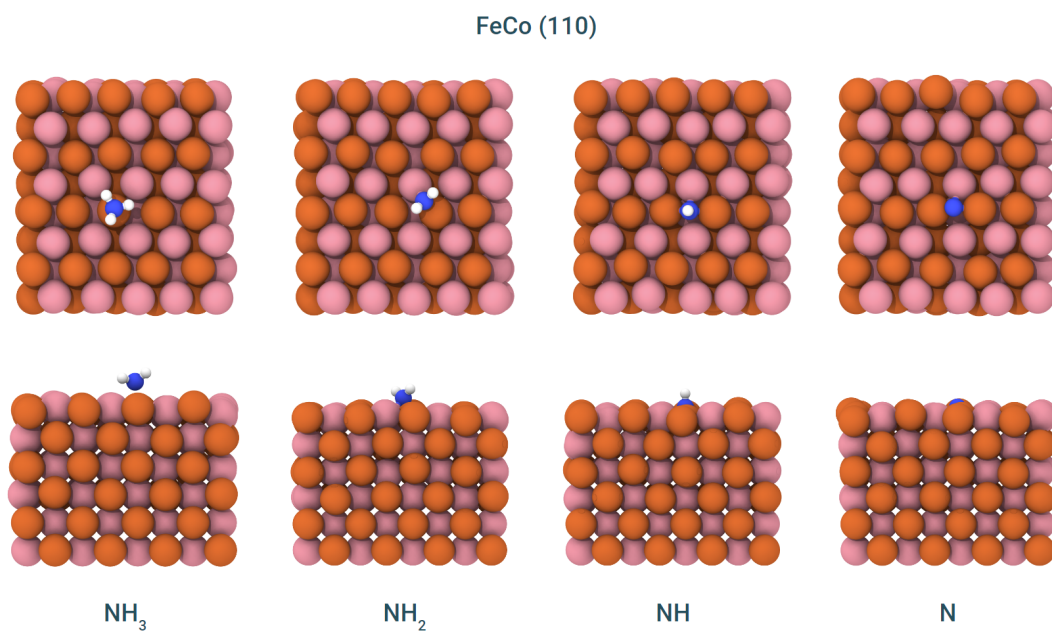
Dataset composition				
System	Stage	MD	System	# configurations
FeCo	Preliminary	AIMD ¹	T=700 K	124
		FLARE	T=800 K	6
				130
N ₂ / FeCo	Preliminary	FLARE	MD	651
		FLARE	OPES (adsorption/diffusion)	91
	Explore Convergence	FLARE	OPES-flooding	86
		MACE	OPES	130
				958
NH _x / FeCo	Preliminary	FLARE	MD	1184
		FLARE	OPES (adsorption/diffusion)	286
	Explore Convergence	FLARE	OPES-flooding	323
		MACE	OPES (reactive)	450
		MACE	OPES (products) ²	1681
				3924
TOTAL				5012

Supplementary Table 1. **Dataset composition.** Summary of the configurations collected during the different stages, divided based on the system, the MD driver, and the type of sampling.

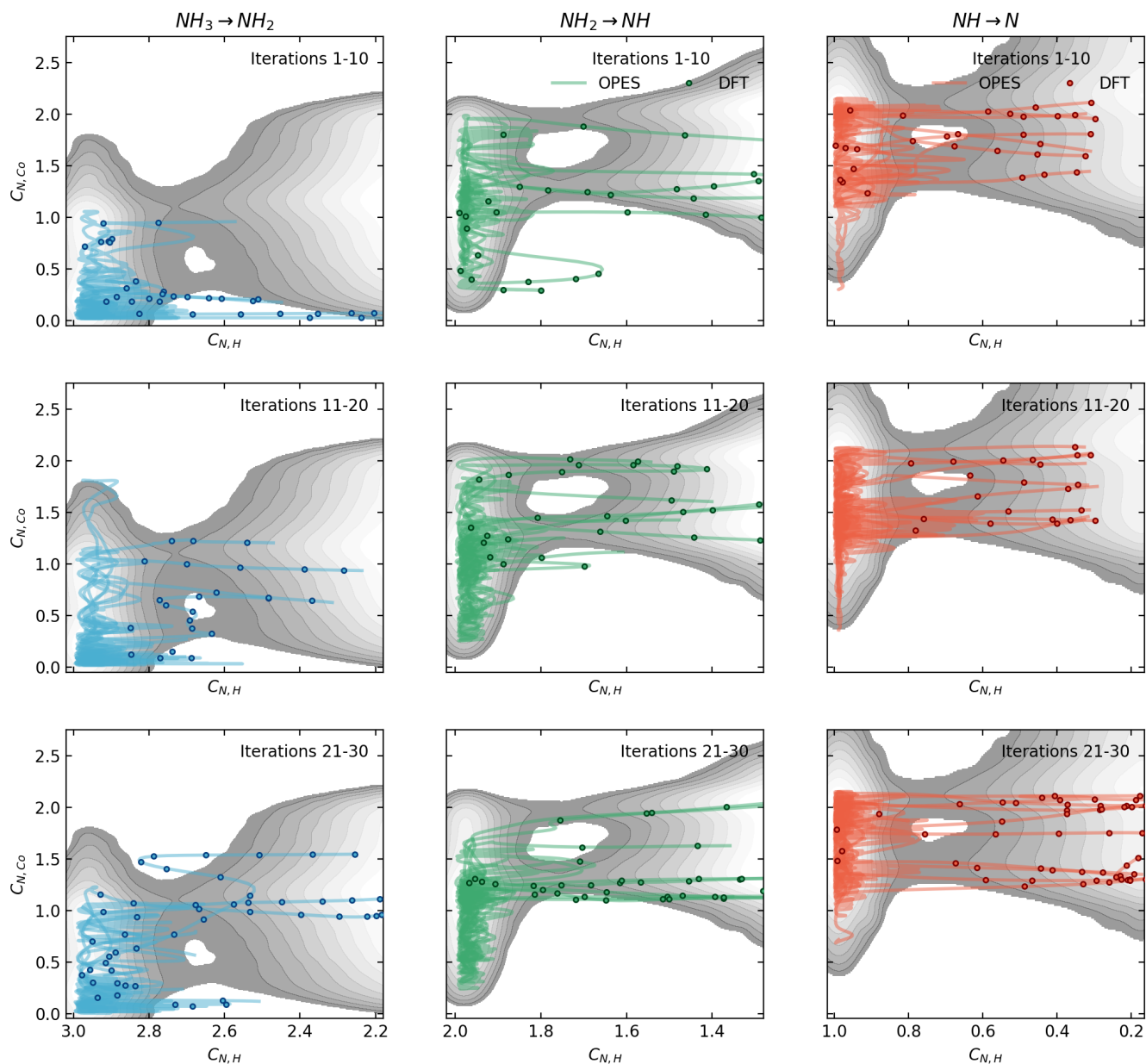
¹: Short (400 fs) AIMD simulation, with configurations extracted every 3 steps. These configurations were used to initialize the GP model.

²: Configurations from the product states extracted from the MACE simulations using the query-by-committee criterion.

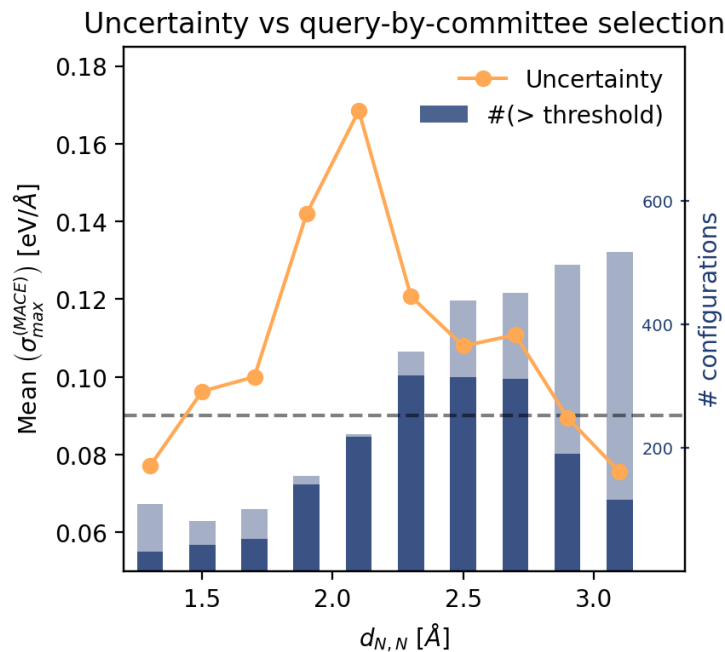
II. SUPPLEMENTARY FIGURES



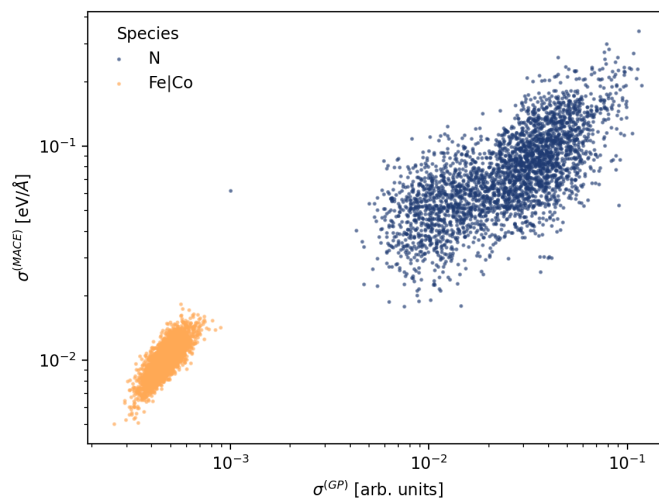
Supplementary Figure 1. **Snapshots of NH_x adsorbates on the FeCo (110) surface, top and left view.**



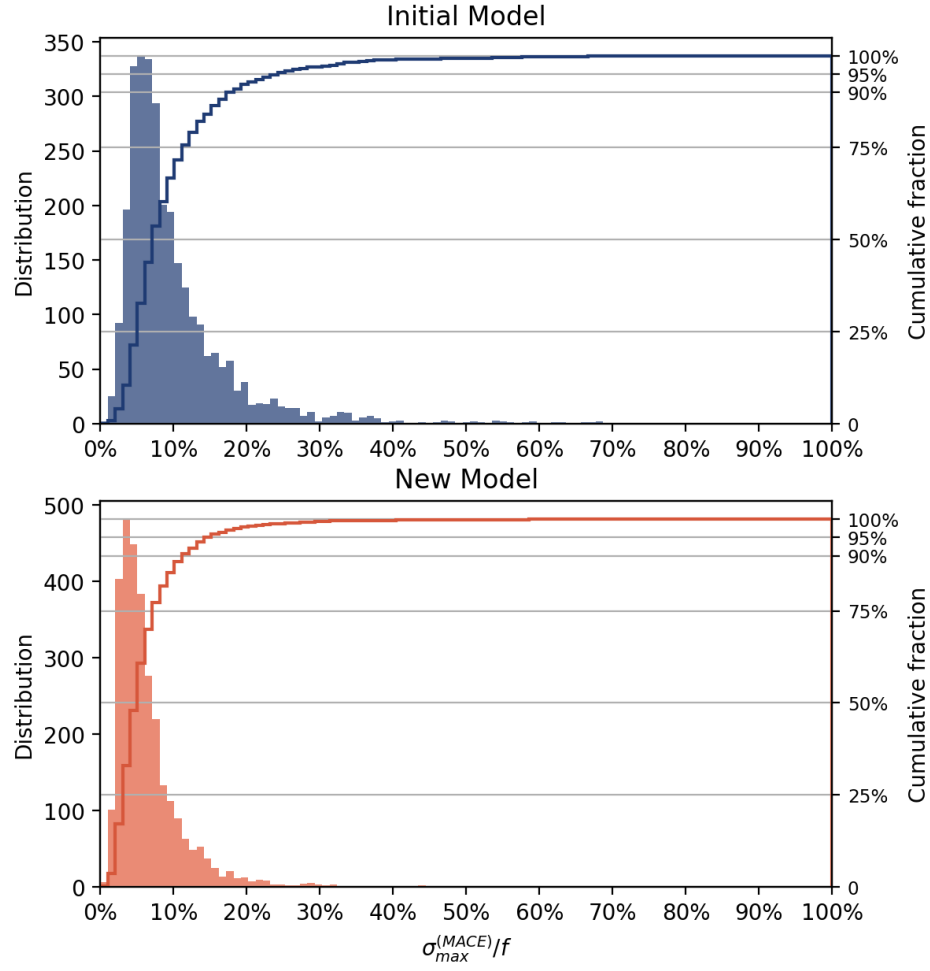
Supplementary Figure 2. **OPES-flooding iterative discovery of reactive pathways** for the dehydrogenation reactions. This is the same plot as in Main Text Fig. 3, divided here into three rows according to the number of iterations. The sampled paths in the first few iterations mostly involve regions of higher free energy. As the flooding iterations proceed and the potential is updated with these new data points, the sampled paths converge to lower free energy regions, emphasizing the importance of an incremental approach. This is particularly evident in the case of the first step of dehydrogenation of NH_3 , which initially occurs at very low coordination numbers between N and Co and then converges to the minimum free energy pathway (that is around 1).



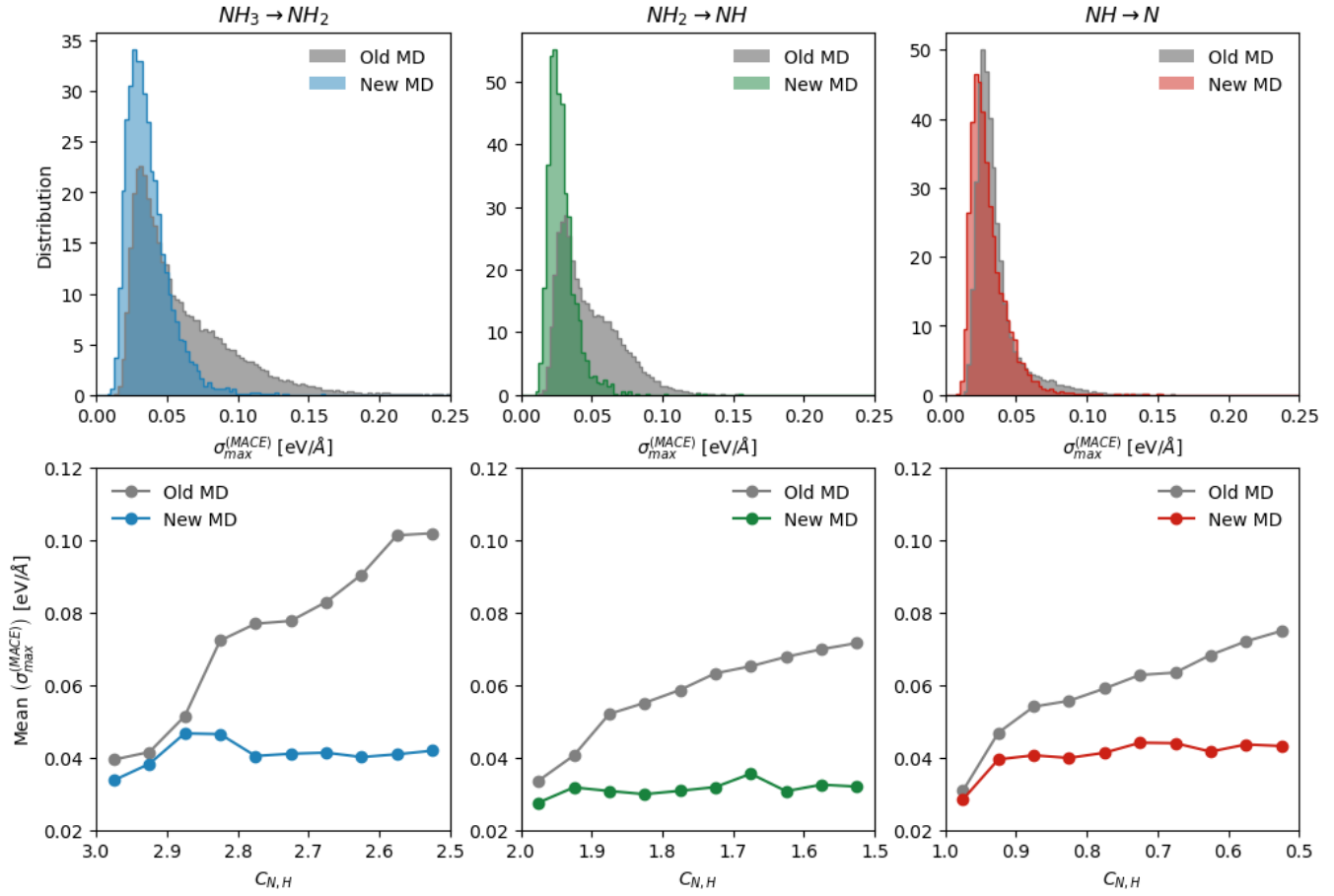
Supplementary Figure 3. **Uncertainty along the CV and query-by-committee distribution** for the $N_2 \rightleftharpoons 2N$ reaction. On the left axis is reported the average uncertainty $\sigma_{\max}^{(MACE)}$ as a function of the collective variable $d_{N,N}$ (orange line). On the right axis there are the number of configurations sampled in the simulation (light blue bars) and the subset selected with the query-by-committee criterion (dark blue bars).



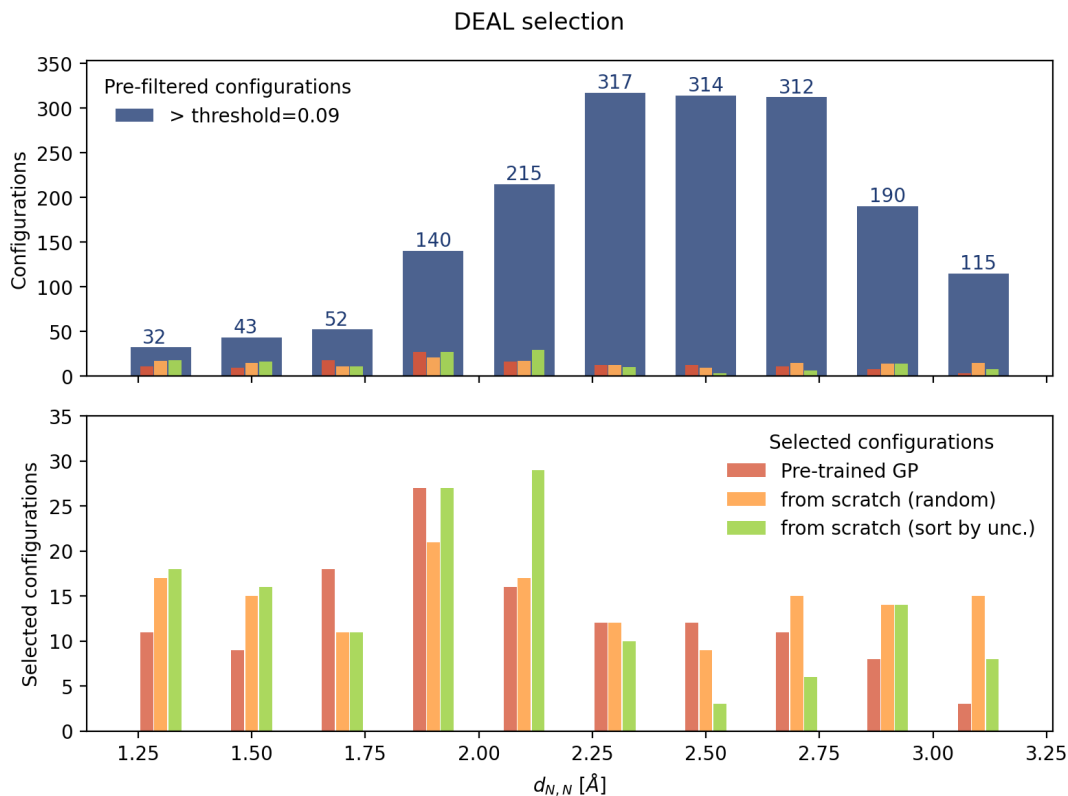
Supplementary Figure 4. **Correlation between GP and GNN uncertainties.** We compared the GP maximum local atomic uncertainty $\sigma_{\max}^{(GP)}$ versus the MACE maximum force standard deviation $\sigma_{\max}^{(MACE)}$, for the $N_2 \rightleftharpoons 2N$ reaction step, for both the adsorbate molecule/atoms (N) and the surface (Fe|Co).



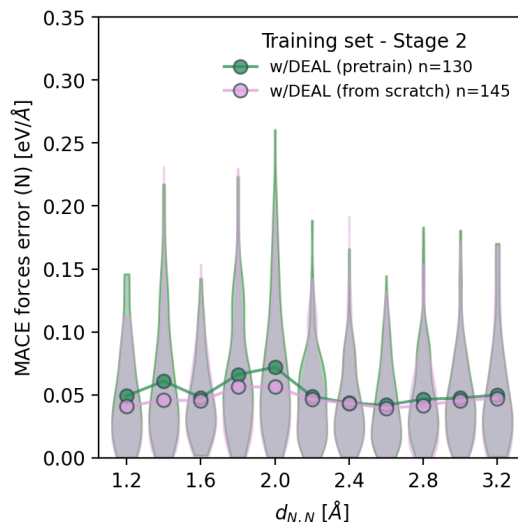
Supplementary Figure 5. **MACE relative uncertainty before/after DEAL.** We plotted the distribution of the relative maximum force uncertainty $\sigma_{max}^{(MACE)}/f$ for the initial MACE model (top) and after (bottom) the DEAL the active learning scheme described in Main Text Fig. 4. On the right axis, the corresponding cumulative fraction is reported.



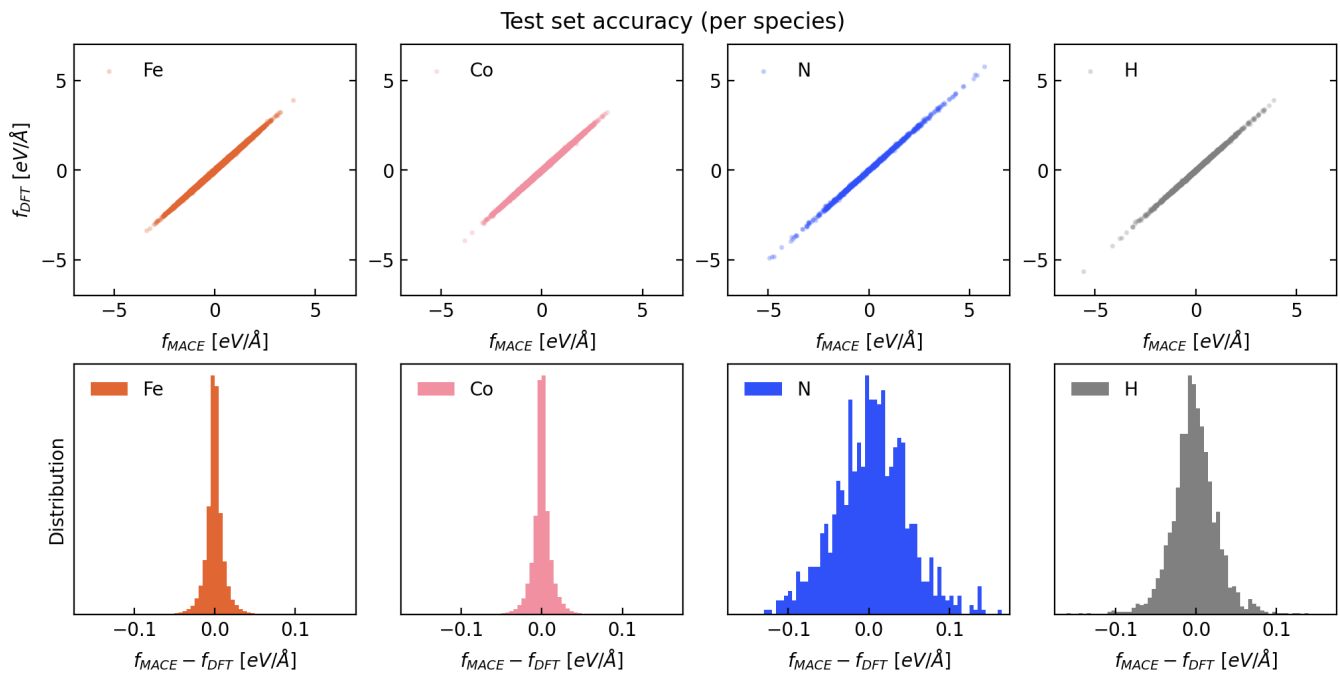
Supplementary Figure 6. **MACE uncertainty before/after DEAL (dehydrogenation reactions).** We present here the results of the dehydrogenation steps. MD simulations are performed with MACE and the OPES bias to reversibly sample the transition ($NH_x \rightleftharpoons NH_{x-1} + H$, $x=1,2,3$). The first row reports the distribution of the maximum uncertainty on the forces $\sigma_{max}^{(MACE)}$ on the MD simulations performed with the MACE models before (grey histograms) and after (colored histograms) the active learning procedure. The second row shows the distribution of the average (maximum) uncertainty along the reaction coordinate $C_{N,H}$ for the configurations extracted from the MD simulations before (grey lines) and after (colored lines) the active learning. After the active learning, the uncertainty was uniformly below the threshold (50 meV/Å). Unlike the case of N_2 recombination, the uncertainty in the product regions (low coordination) was even higher than the TS. This was due to the absence of configurations with short-range interaction between the products in the preliminary dataset.



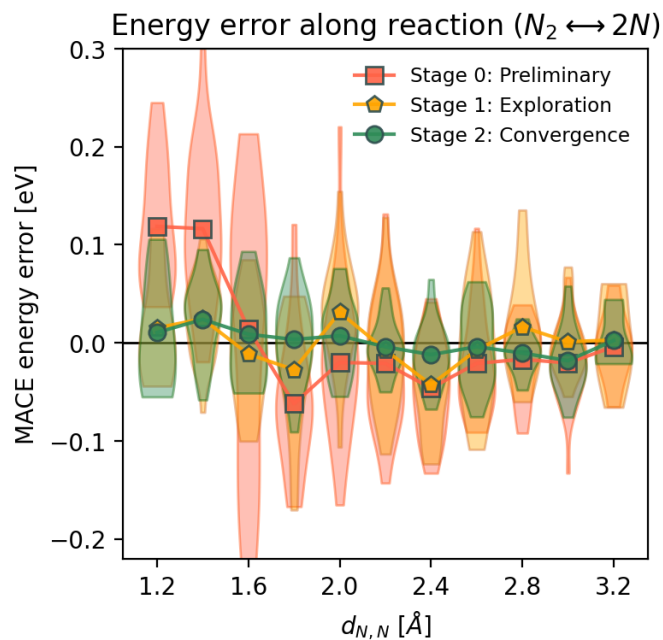
Supplementary Figure 7. **DEAL selected configurations by the different variants (pre-train GP vs from scratch).** Here we compare the distribution of configurations selected by DEAL variants along the CV. (top) Distributions of high uncertainty configurations pre-selected with the query-by-committee criterion. (bottom) Configurations selected by the GP pre-trained on the data collected in steps 0 and 1 and those trained from scratch only on the configurations generated by MD, using MACE energy/forces as labels (see Methods). In particular, since training a model from scratch would result in selecting more configurations from the first part of the dataset, we tested two different schemes: shuffling the trajectory or sorting the configurations according to uncertainty in order to select more configurations with high uncertainty. In both cases, we found remarkable agreement with the selection made with the pre-trained model.



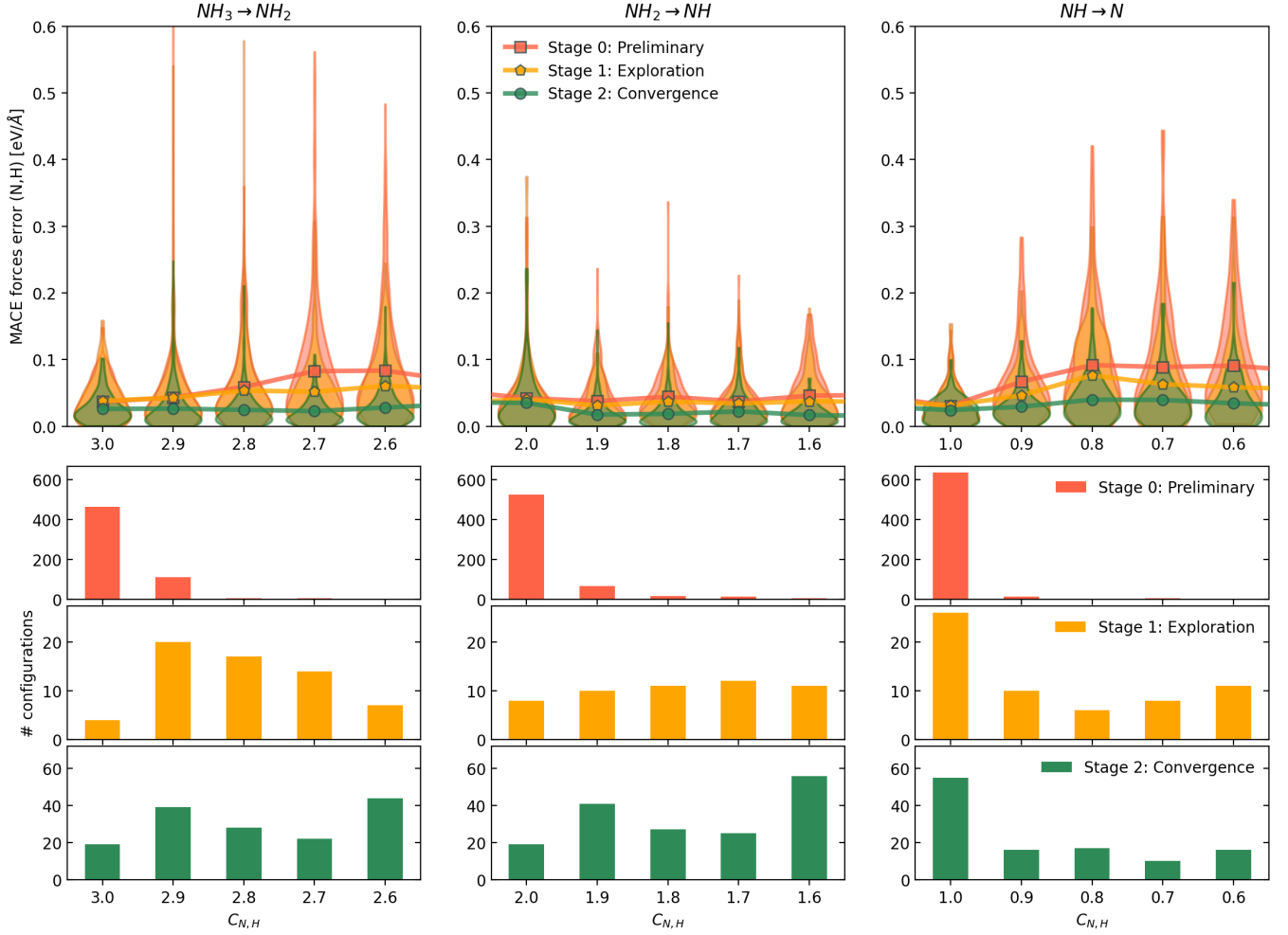
Supplementary Figure 8. **DEAL selection: forces accuracy along the reactive pathway for the two variants (pre-train GP vs. from scratch).** We compared the accuracy of the force predictions of the N atoms (as in Main Text Fig. 5) along the reactive pathway using DEAL-selected configurations with a pre-trained or trained ex novo GP model. In both cases, a uniformly good description is obtained along the entire reactive path.



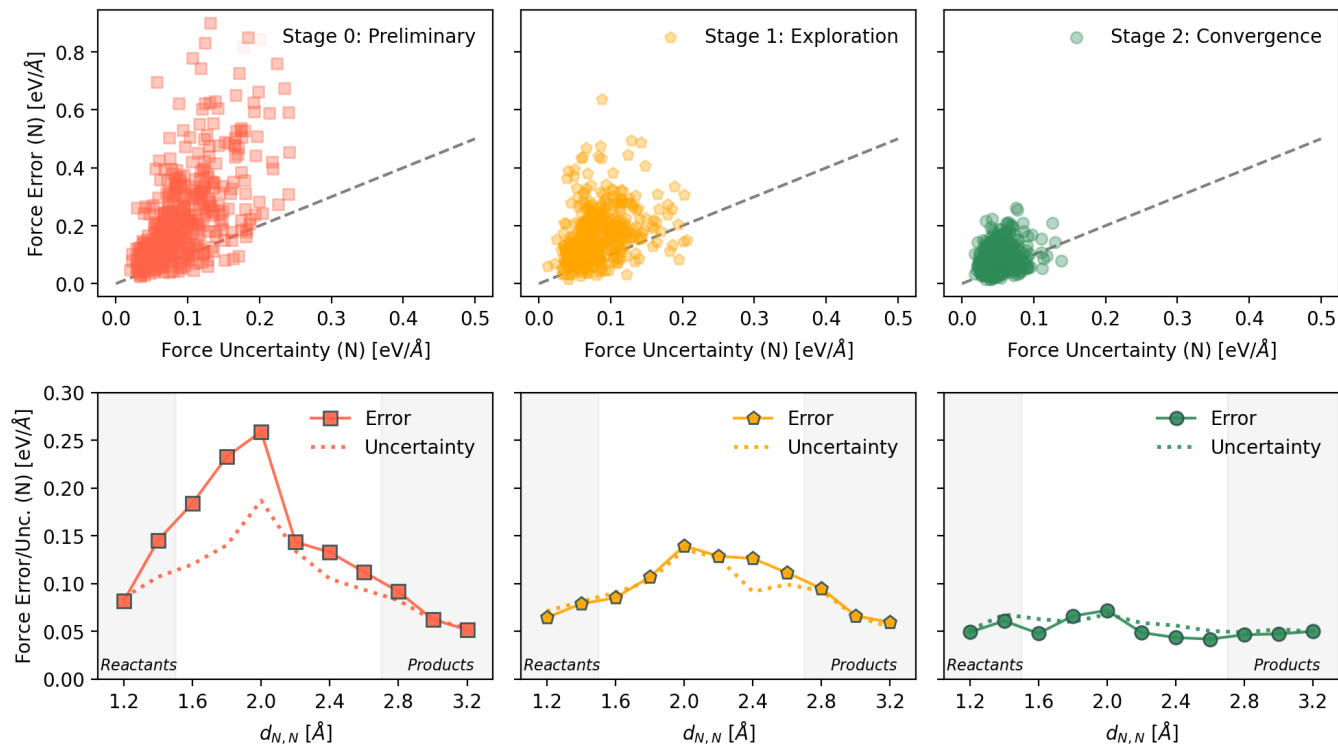
Supplementary Figure 9. **MACE forces accuracy of the final stage divided per chemical species** (top row) MACE forces prediction versus DFT reference, resolved per atomic species. (bottom row) Distribution of the error per each species. The test dataset is composed of a subset of configurations extracted from the final dehydrogenation simulations and recalculated with DFT single point calculations.



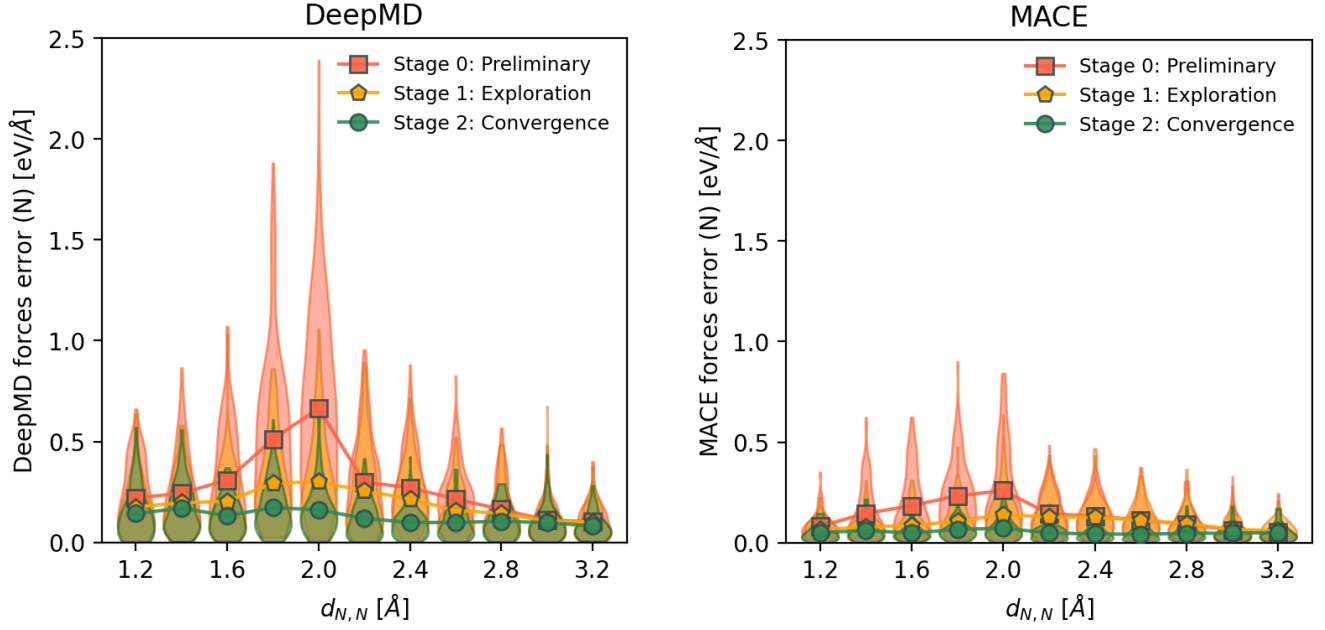
Supplementary Figure 10. **MACE energy accuracy along the reactive pathway** for $N_2 \leftrightarrow 2N$, as a function of the N-N distance. The violin plots represent the distribution of the errors, and the lines denote the average over the points belonging to each bin.



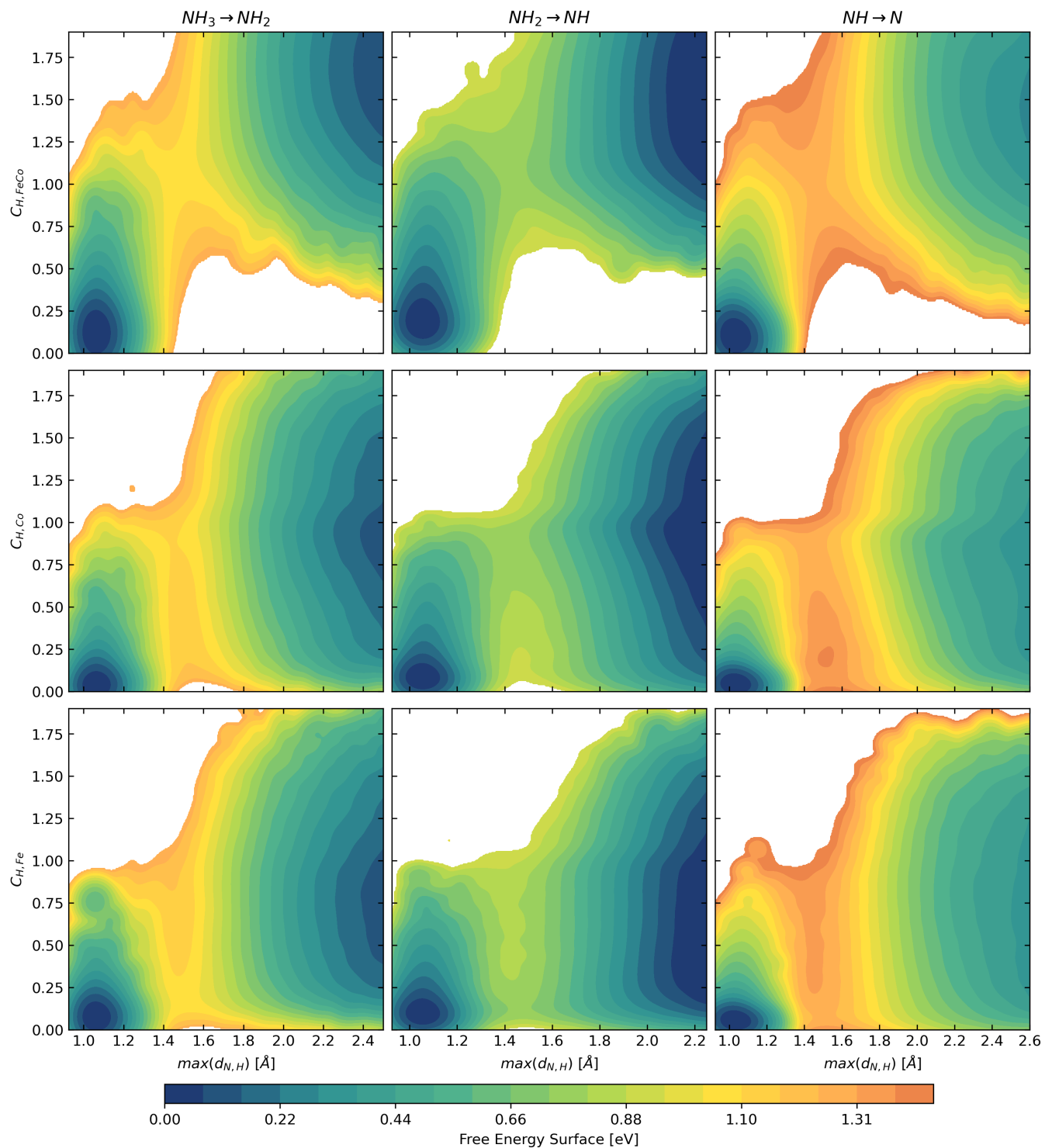
Supplementary Figure 11. **MACE forces accuracy along the reactive pathways of $NH_x \rightleftharpoons NH_{x-1} + H, x=1,2,3$.** (top) This is the same plot as Main Text Fig. 5 but for the dehydrogenation reactions. (top) Force accuracy for the N forces along the CV (coordination N,H), with the violin plot representing the distribution and the solid lines the average along the CV. Each line/violinplot represents the MACE model obtained with the training set at a given stage. (bottom) Number of configurations in the reactive range collected at each stage.

Accuracy and uncertainty along reaction profile ($N_2 \leftrightarrow 2N$)

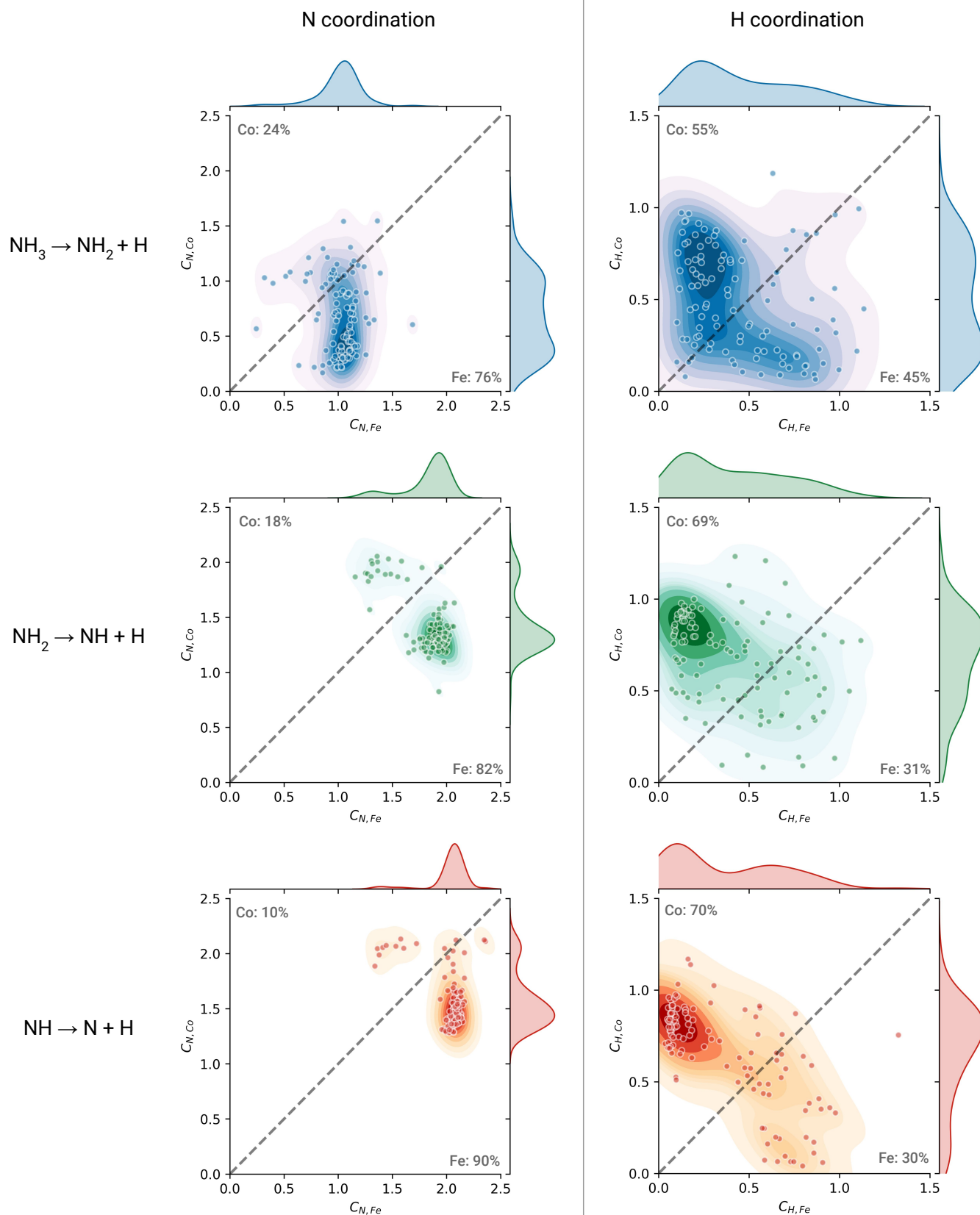
Supplementary Figure 12. **MACE forces accuracy vs uncertainty for MACE GNN models.** Top row: scatter plot between the MACE uncertainty and the MACE error for the predictions of the N forces, calculated on the reactive pathway of the $N_2 \leftrightarrow 2N$ reaction. Bottom row: Average MACE error/uncertainty along the CV (N-N distance), highlighting the reactant and product regions. Each column corresponds to the MACE models trained with the training set collected at the different stages.



Supplementary Figure 13. **Comparison between DeepMD and MACE for the forces accuracy at the different stages** for the $N_2 \rightleftharpoons 2N$ reaction. To evaluate how important it is to use a data-efficient GNN architecture in our scheme, we optimized another set of ML potentials using the DeepMD scheme on the same data collected at different stages. We notice that the errors are systematically larger: after the initial exploratory phase (Stage 1), DeepMD errors still reach more than $1 \text{ eV}/\text{\AA}$, which leads the simulations to be highly unstable. As a result, it's not possible to use a potential with DeepMD with so few reactive configurations. At the same time, this analysis shows us that, at the end of the sampling process (Stage 2), the accuracy of DeepMD is also uniform (although much higher) throughout the entire reactive path, confirming the goodness of the data selection scheme.



Supplementary Figure 14. **Free energy surfaces of dehydrogenation steps** $NH_x \rightleftharpoons NH_{x-1} + H$ ($x=1,2,3$) projected along the maximum distance between N and H atoms $\max(d_{N,H})$ and the coordination with surface atoms ($C_{H,Fe|Co}$, top panels), cobalt ($C_{H,Co}$, central panels), and iron ($C_{H,Fe}$, bottom panels). The free energy surfaces are shown up to $3k_B T$ above the barriers for each reaction. In terms of coordination with the surface, the minimum free energy pathways pass through a value of $C_{H,Fe|Co} \approx 1$ (top panels). However, if we resolve the coordination with respect to Fe and Co, we see a marked tendency for the reactive hydrogen to form bonds with Co rather than Fe, indicated by the minimum free energy path passing through $C_{H,Co} \approx 1$ and $C_{H,Fe} \approx 0$.



Supplementary Figure 15. **Transition state analysis.** Analysis of geometries related to the transition state configurations extracted from the OPES-flooding trajectories (see Main Text Fig. 7) for the three dehydrogenation steps. The left column shows the coordination number between iron (x) and cobalt (y) with N, while the right column shows the one with H. The distribution shows that, at the transition state, N is more coordinated with Fe in most of the reactive events, while H is more closely coordinated with Co. This can be quantified by the percentage of reactive events for which the coordination with N/H is higher with Co (Fe), which is reported in the top left (bottom right) corner of each plot.