
Equivariant Modelling for Catalysis on 2D MXenes

Pavlo Melnyk^{*1} Anmar Karmush^{*1} Ania Beatriz Rodríguez-Barrera^{2,3}
Mårten Wadenbäck¹ Michael Felsberg¹ Johanna Rosen^{2,3} Jonas Björk^{*†2,3}

¹Computer Vision Laboratory, Department of Electrical Engineering (ISY) & AI4X,

²Materials Design Division, Department of Physics, Chemistry and Biology (IFM), and

³Wallenberg Initiative Materials Science for Sustainability (WISE);

Linköping University (LiU), 58183 Linköping, Sweden.

^{*} equal contribution [†] corresponding author jonas.bjork@liu.se

Abstract

Merging advanced computations with machine learning, we aim to accelerate the exploration of catalytic behaviour in novel materials. We focus on two-dimensional (2D) Ti_2CT_y MXenes, whose versatile surface chemistry makes them particularly compelling candidates for catalysis. However, resolving their composition and structure under realistic conditions requires going beyond the systems typically studied with density functional theory (DFT), as the computational cost of such calculations limits accessible system sizes and timescales, calling instead for more efficient approaches. To address this challenge, we generate a comprehensive dataset of 50,000 DFT calculations for training and 10,000 for testing, encompassing both Ti_2CT_y MXene configurations and molecular systems, along with an augmented dataset where systems are artificially repeated to investigate how well models generalise to larger systems. Employing advances in geometric deep learning, we train and validate an equivariant (i.e. symmetry-aware) model (EquiformerV2) that accurately predicts atomic forces and formation energies — quantities that DFT must repeatedly compute for structural and catalytic investigations — for these 2D materials. This combined DFT–ML framework achieves computational acceleration of the order $\sim 10^3$ – 10^4 (on a CPU) while maintaining DFT-level accuracy ($\sim \pm 45$ meV/Å for forces and $\sim \pm 6$ meV for per-atom energies), paving the way for more efficient investigations of MXene catalytic behaviour. Moreover, we confirm that the total energy prediction error of the model grows linearly with the number of atoms in an input system, while the force error remains the same, which, along with the equivariant model design, is a necessity for a robust model. The dataset and the trained models with the code are available at <https://github.com/CataLiUst>.

1 Introduction

A central challenge in computational catalysis is balancing accuracy and efficiency. Density functional theory (DFT) provides a reliable description of surface chemistry, delivering energies and atomic forces that can be used to model molecule-surface interactions and reaction pathways. However, the computational cost of obtaining such data restricts simulations to relatively small systems and short timescales — far from resembling realistic operating conditions of catalysts. This limitation is particularly problematic because catalysts are inherently dynamic: their structure and reactivity evolve under reaction conditions, and idealised models may risk missing the relevant chemistry.

Machine learning (ML) interatomic potentials (MLIP) [1], often within the framework of geometric deep learning (GDL) [2, 3], offer a promising way forward by approximating Density functional theory (DFT)-calculated energies and forces with near-DFT accuracy while operating orders of magnitude faster [4–7]. This facilitates the study of catalytic systems under more realistic conditions, helping to bridge the gap between idealised DFT models and experimental environments. Such simulations can, in turn, identify structural motifs and environments that are truly representative, guiding more focused and accurate mechanistic studies at the DFT level. While the existing benchmarks for related catalysis problems [9, 10] involve metal oxides and metal-intermetallics, another group of materials

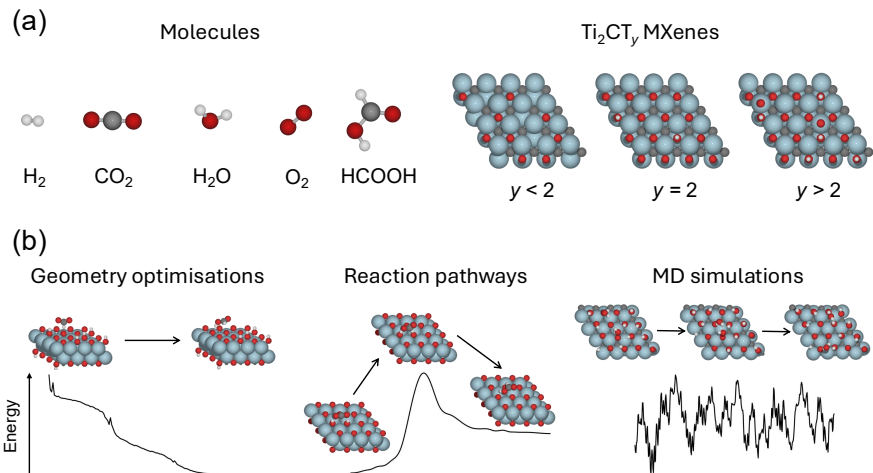


Figure 1: Types of (a) systems and (b) DFT calculations included in the dataset. It comprises DFT calculations of five different molecules adsorbed on Ti₂CT_y MXenes with varying surface termination configurations (denoted T_y). The calculations include geometry optimisations, reaction pathways, and high-temperature molecular dynamics simulations, providing a diverse set of configurations for training. Atomic simulation environment (ASE) [8] was used for structural visualisation.

for which the above considerations are particularly relevant is that of two-dimensional transition metal carbides (MXenes). The surfaces of these materials are far from well-defined, as sluggish adsorption processes during synthesis lead to a variety of stable surface functionalisations with distinct chemical reactivities. At the same time, experimental studies have demonstrated promising catalytic performance for MXenes, and DFT investigations have shown that the catalytic activity is highly sensitive to the precise surface chemical environment. For example, for MXenes with mixed O and OH terminations, the dehydrogenation of alkanes to olefins has a reactivity that decreases linearly with increasing number of OH terminations [11, 12]. At the same time, MXenes with high OH coverage have been predicted to have a high reactivity to CO₂ reduction [13]. Thus, even for MXenes in their idealised, well-defined form, the catalytic activity can be systematically tuned by modifying the surface terminations.

However, despite their great promise in catalysis, very little is known about the actual state of MXenes under operating conditions. Capturing how their surface terminations and reactivity evolve in reactive environments requires simulations that are both accurate and computationally feasible — beyond the scope of conventional DFT approaches. Although there have been ML approaches applied to catalysis on MXenes — for example, to predict MXene-supported single-atom catalysts for oxygen reduction and evolution reactions [14] as well as ammonia synthesis [15] — most work has focused on static screening of catalytic activity. The application of ML models to capture the dynamic surface chemistry and evolving reactivity of MXenes under operating conditions has largely been overlooked. In this context, MLIP trained on high-quality DFT data provide a powerful route to explore the dynamic surface chemistry of MXenes and to establish a more realistic understanding of their catalytic behaviour under working conditions.

In this work, we build on recent advances in geometric deep learning — in particular, equivariant (i.e. symmetry-aware) modelling for interatomic potentials — to investigate its utility for catalysis on 2D MXenes. Specifically, we summarise our contributions as follows:

- (i) We generate a dataset for catalysis on Ti₂CT_y MXenes, CATALIUST Ti₂C-MXENE, comprising 50,000 DFT calculations for training and 10,000 for testing, including MXene configurations with and without molecular adsorbates, along with an augmented dataset, in which systems are artificially repeated to investigate how well our modelling generalises to larger systems.
- (ii) We train an equivariant state-of-the-art model, EquiformerV2 [16], on the proposed dataset, and evaluate its ability to generalise to larger atomic systems.
- (iii) We integrate the trained model into a computational workflow, which achieves 1,113–8,848-fold computational acceleration (on a CPU) compared to the DFT method while maintaining desired accuracy and robustness, enabling more efficient studies of MXene catalytic behaviour.

2 Background

2.1 Density Functional Theory (DFT)

DFT is a first-principles method: given only the atomic species and coordinates, it computes the total energy, forces on atoms, and other properties derived from the electron density. At its core, DFT replaces the many-electron wave function with the electron density—a simpler quantity that, in principle, contains all ground-state information about the system [17, 18].

In practice, however, DFT is approximate at several levels. The primary challenge lies in describing the non-classical part of the electron-electron interaction, namely the exchange and correlation effects. These contributions are encompassed in the so-called exchange–correlation (XC) functional. Since the exact form of this functional is unknown, a range of approximations exist, each balancing accuracy and computational cost. The predictive power of DFT therefore depends critically on how well the chosen XC functional represents the relevant electronic interactions.

Accurately describing non-local dispersion forces, or van der Waals (vdW) interactions, is a particular challenge, as generalised gradient approximation (GGA) functionals cannot capture them. To address this, we employ the vdW density functional (vdW-DF) [19] in the form introduced by Hamada [20] (rev-vdW-DF2). This functional accurately describes structural parameters of MXenes [21], molecular adsorption energies and distances [22], lattice parameters and interlayer binding energies in graphite [20], and other weakly bonded layered systems [23].

2.2 Equivariant Modelling with Geometric Deep Learning

Incorporating correct inductive biases in learning systems facilitates the learning process and enables tractable learning as the space of transformations acting on the input grows [3]. In general, *symmetries*, both as (i) the transformations of an object (or its properties) that leave it unchanged (*invariance*) and (ii) the transformations that change it in a predictable manner (*equivariance*), are powerful inductive biases [24, 25]. Note that (i) is a special case of (ii). More formally, GDL [3] uses the language of group and representation theory and provides a framework for constructing neural network (NN) architectures that adhere to the symmetries in the data the NNs are to model.

One such example in deep learning is convolutional neural networks (CNNs) [26–28], which have built-in translation equivariance, enabling the identification of objects in images regardless of their shifted positions within the images. This is achieved thanks to the inherent properties of cross-correlation being the central operation of CNNs. Thus, a CNN classifying images containing a particular object should not change the class attribute if the object in the image is shifted. Another example is models equivariant under rotations (encompassed by the group $SO(n)$, where n is the dimensionality of the space) [29–33]. For instance, for a classification model taking in an input 3D point cloud representing the surface of an object, rotating the point cloud should not affect the class assignment produced by the model (invariance under rotations a.k.a. $SO(3)$ -invariance) [34, 35]. At the same time, a model predicting forces acting on the object should output correspondingly rotated force vectors if the input point cloud is rotated ($SO(3)$ -equivariance). In the following section, we show how the GDL framework is used in our work for catalysis on 2D MXenes.

3 Method

We aim to enable accelerated computational investigations of catalytic processes on 2D MXenes. To attain this, we use first principles DFT calculations to produce a dataset of MXene systems, which we later use to train a model to predict atomic forces and formation energy for each of the systems.

3.1 CATALIUST Ti2C-MXENE: DFT dataset of molecules interacting with MXene surfaces

While prominent ML catalysis efforts, such as the Open Catalyst Project [9, 10], focus on screening across a wide range of materials to identify promising candidates, here we adopt a complementary approach. Instead of targeting broad screening, we focus on a single material — the Ti_2CT_y MXene — chosen for its catalytic relevance and rich surface chemistry. This allows us to construct a highly detailed and physically consistent dataset that captures the complexity of a single catalyst under realistic conditions. Our dataset, called CATALIUST Ti2C-MXENE, spans a large range of surface terminations, from non-terminated to over-terminated configurations, including mixed O and OH coverage. Furthermore, it includes configurations far from equilibrium obtained from high-temperature molecular dynamics and reaction pathway calculations, enabling the model to capture chemical environments encountered under operating conditions. The dataset covers molecules

relevant to CO₂ reduction to formic acid, as well as other key catalytic processes such as the hydrogen evolution reaction (HER) and the oxygen evolution reaction (OER).

The diversity and depth of configurations are made possible through systematic DFT calculations, which form the foundation of the dataset. Each data point corresponds to a static DFT calculation that provides the total energy of the system and the forces acting on its atoms. For the energy, we use the **formation** energy, defined as

$$\Delta_f E = E_{\text{DFT}}(\text{system}) - \sum_i E_{\text{ref}}(\text{atom}_i), \quad (1)$$

where $E_{\text{DFT}}(\text{system})$ is the total DFT energy and $E_{\text{ref}}(\text{atom}_i)$ is the reference energy of atom i in the system, which contains the positions, i.e. coordinates, pos and atomic species (numbers) of the atoms. Reference energies are taken as the energies of the respective elemental phases — graphitic C, hcp Ti, O₂ gas, and H₂ gas — normalised per atom.

Note that $\Delta_f E$ is the formation energy of the complete unit cell representing a system, and can be viewed as its total potential energy with a shifted reference relative to the raw DFT energy. In contrast to reporting adsorption energies for specific adsorbates, formation energies provide a consistent physical reference independent of the number or identity of adsorbates, enabling comparisons across different structures and chemical environments, while still allowing derivation of meaningful energy differences — such as adsorption energies, as well as reaction and activation energies — all relevant in catalysis. The data, outlined in Figure 1, comprise DFT calculations for:

- Individual molecules (H₂, CO₂, H₂O, O₂, HCOOH).
- Ti₂CT_y MXenes with various surface termination configurations (denoted T_y). Both the coverage of surface terminations (value of y) and the types of terminations (O and/or OH) are considered to capture the diversity of MXene surface chemistry.
- Molecules adsorbed on the different MXene surfaces.

The data originate from several calculation types:

1. **Geometry optimisations**

Configurations correspond largely to local minima, where atomic forces are generally small.

2. **Reaction pathway calculations**

These probe reaction mechanisms by optimising along a reaction coordinate (using the nudged elastic band method). Forces are not necessarily small, as these configurations represent specific regions of the potential energy surface connecting local minima.

3. **High temperature molecular dynamics (MD) simulations (>1000 K)**

These explore configurations far from local minima and reaction coordinates. This diverse dataset spans a broader portion of configurational space, compelling the model to learn system behaviour far from ideal conditions — a crucial capability for accurately describing catalysts under realistic operating conditions.

Notably, our dataset contains only four atom types, but the complexity of the system is amplified by the diversity of their chemical environments. In contrast to systems with a larger number of distinct elements, where surface and molecular species are inherently distinguished, here each element appears in multiple, chemically distinct bonding motifs. For example, a carbon atom in a MXene — coordinated to six titanium atoms — is fundamentally different from a carbon atom in a molecule, and carbon atoms in different molecules also differ significantly in bonding. Likewise, oxygen behaves very differently when bound as a surface termination to titanium atoms compared to when participating in various molecular configurations. This rich variety in the chemical surrounding of different atom types greatly increases the challenge for any model.

To augment the training and test datasets, we generated replicated systems by repeating each original system along the three unit cell vectors defining the periodic boundary conditions. Repetition factors n_1, n_2, n_3 were randomly chosen with equal probability, with $n_1, n_2 \in \{1, 2, 3\}$ along the in-plane lattice vectors and $n_3 \in \{1, 2\}$ along the out-of-plane vector. The total energy of each augmented system was scaled by $n_1 \times n_2 \times n_3$, while atomic forces for repeated atoms were copied directly from the original system. This augmented dataset serves as a benchmark for testing the ability of the model to generalise with an increase in the size of input molecular systems, without adding new chemical environments. This constitutes separate training and test datasets with the same number of systems as in the original dataset, denoted *with repetitions*.

Dataset		Number of atoms				Number of systems
Repetitions	Split	min	median	max	mean $\pm$ std	
w/o rep.	Train	2	98.0	156	86.7 $\pm$ 38.2	50,000
	Test	2	96.0	158	86.4 $\pm$ 36.1	10,000
w/ rep. ($\approx 6\times$)	Train	2	396.0	2808	519.1 $\pm$ 466.6	50,000
	Test	2	396.0	2844	521.6 $\pm$ 465.4	10,000

Table 1: Number of atoms in the molecular systems in datasets *without* and *with repetitions* (which contains, on average, $\approx 6\times$ larger systems).

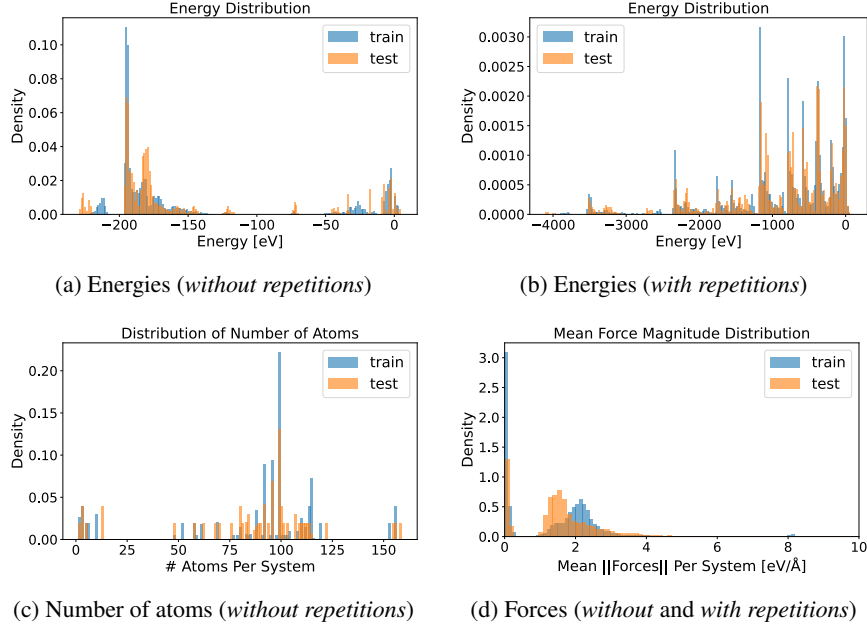


Figure 2: Distributions for energies, number of atoms, and mean force magnitudes across training and test sets. By definition, the L2-norm of atomic forces will be the same for the two datasets.

Details of DFT calculations Periodic density functional theory calculations are performed with the VASP code [36], using the project-augmented wave method [37, 38], and with a plane wave basis expanded to a kinetic energy cutoff of 400 eV. The MXenes are represented in p (4×4) unit cell together with a Γ -point only k -point sampling.

Dataset statistics Each system is stored as a row in an HDF5 file containing atomic numbers, number of atoms, atomic positions, forces, formation energy, and unit cell. Table 1 summarises the atomic composition of the datasets *without* and *with repetitions* (which contains, on average, $\approx 6\times$ larger systems). Some systems contain up to 2,844 atoms, making the prediction task particularly challenging. Figure 2 shows the distributions of formation energies, average L2-norms of atomic forces, and the number of atoms per system for the training and test sets. As seen in both Figures 2a and 2b, the energy distributions differ between the training and test sets, whereas Figure 2d reveals that many training structures are close to equilibrium (with low mean force magnitudes). Similar to the energies, certain force ranges are absent in the test data, further emphasising the difficulty of generalisation to unseen configurations. A comparison of the proposed dataset with OC20 [9] is presented in Appendix A.

3.2 Combined DFT and ML approach

To unlock the possibility of simulating MXenes under realistic conditions, we approximate DFT by modelling forces $\mathbf{f} \in \mathbb{R}^{N \times 3}$ and energies $E \in \mathbb{R}$ of MXene systems, with the atomic positions represented as the point set $\text{system}_{\text{pos}} = \{\text{pos}_i\}_{i=1}^N \in \mathbb{R}^{N \times 3}$ with $\text{pos}_i \in \mathbb{R}^3$ being the (x, y, z) -coordinates of atom i , and N being the number of atoms in the system. Note that despite the two-dimensional periodic structure of MXenes, the atomic positions are still defined in three-dimensional Euclidean space $\mathbb{R}^3$.

Model (size) & Training set	Training time (GPU hours)	Validation		Test	
		Energy MAE (meV) ↓	Force MAE (meV/Å) ↓	Energy MAE (meV) ↓	Force MAE (meV/Å) ↓
<i>Original</i> (30.8M) w/o rep.	253.6	49.0 ± 1.2 (47.7)	5.8 ± 0.1 (5.6)	472.9 ± 9.0 (464.6)	43.5 ± 0.4 (43.0)
<i>Original PT</i> (30.8M) w/o rep.	253.7	84.0 ± 15.3 (63.8)	6.5 ± 0.5 (5.9)	467.9 ± 15.2 (452.0)	33.9 ± 0.3 (33.6)
<i>Small</i> (4.8M) w/o rep.	141.3	238.6 ± 43.2 (195.4)	11.0 ± 0.2 (10.8)	512.0 ± 35.8 (477.0)	44.9 ± 0.7 (44.3)
<i>Small</i> (4.8M) w/ rep. ($\approx 6\times$)	457.6	729.2 ± 8.5 (720.8)	11.1 ± 0.5 (10.5)	2922.3 ± 29.3 (2881.2)	44.5 ± 0.5 (44.0)

Table 2: The performance of EquiformerV2 on the proposed dataset. *W/ rep.* refers to the augmentation of the original proposed dataset (*w/o rep.*) performed by artificially repeating the systems (see details in the text). The final model is selected based on the validation performance. Mean and std over 3 runs are presented with the best result in parentheses. PT denotes models that were pre-trained on the OC20 dataset. Validation and test sets are *w/ rep.* for *small* trained on *w/ rep.*

When the coordinate system of $\text{system}_{\text{pos}}$ is rotated, the forces $\mathbf{f}$, acting on the atoms, rotate accordingly, exhibiting the rotation equivariance of the system. That is, for a rotation-equivariant function $M_{\mathbf{f}} : \mathbb{R}^{N \times 3} \rightarrow \mathbb{R}^{N \times 3}$ and any rotation $R \in \text{SO}(3)$ represented as a matrix in $\mathbb{R}^{3 \times 3}$,

$$M_{\mathbf{f}}(\text{system}_{\text{pos}} R) = M_{\mathbf{f}}(\text{system}_{\text{pos}}) R. \quad (2)$$

Simultaneously, this does not affect the energy of system, i.e. the energy is rotation-invariant in this context. That is, for a rotation-invariant function $M_E : \mathbb{R}^{N \times 3} \rightarrow \mathbb{R}$,

$$M_E(\text{system}_{\text{pos}} R) = M_E(\text{system}_{\text{pos}}). \quad (3)$$

There is a plethora of work demonstrating that incorporating rotation equivariance as an inductive bias into the model operating on point sets is useful [30, 39–44]. One such model is EquiformerV2, which integrates these $\text{SO}(3)$ -symmetries *by design*, simultaneously approximating forces and energy in accordance with Eqs. (2) and (3), respectively, and exhibits state-of-the-art performance in the related applications [16], which is why we consider it as a primary method for our experiments.

We argue that, beyond being robust to different coordinate system orientations, a model approximating atomic forces and energies must also generalise to systems containing a larger number of atoms. We investigate this property in the case of EquiformerV2 in the experiments presented in the following section.

4 Experiments and Results

We conduct experiments using two variants of EquiformerV2: the 31M-parameter model from the original paper [16] (referred to as *original*), as well as its version pretrained on the OC20 dataset (*original PT*), and a smaller 4.8M-parameter model (referred to as *small*). Among others, the main changes between *original* and *small* were reducing the number of layers from 8 to 5 and the number of attention heads from 8 to 4. For the dataset with artificial repetitions, we report results only with the *small* model, since the *original* model could not be trained due to GPU memory constraints.

4.1 Training setup

We first split the training set into training and validation sets with the 80/20 ratio, aiming at having a similar distribution of system sizes (the validation indices will be provided in the dataset). We then utilise the official PyTorch [45] implementation of EquiformerV2¹ and train all the models end-to-end for 100 epochs, with batch size set to 1. The learning rate is initially set to $4 \cdot 10^{-4}$ with a cosine annealing decay down to $4 \cdot 10^{-6}$ at the last epoch, with the AdamW optimiser and the mean absolute error (MAE) objective

$$\mathcal{L}_{\text{total}} := \frac{1}{N} \mathcal{L}_E + \lambda_{\mathbf{f}} \mathcal{L}_{\mathbf{f}} := \frac{1}{N} |E^{\text{gt}} - E^{\text{pred}}| + \lambda_{\mathbf{f}} \frac{1}{3N} \sum_{i=1}^N \sum_{j=1}^3 |\mathbf{f}_{ij}^{\text{gt}} - \mathbf{f}_{ij}^{\text{pred}}|, \quad (4)$$

where *gt* and *pred* are respectively ground-truth and predicted quantities, N is the number of atoms in the system, and $\lambda_{\mathbf{f}} = 25$ is the weight we experimentally find to work well. We train all the models on NVIDIA A100 (40GB), and for the *small* model trained on the dataset *with repetitions* (with the maximum number of atoms per system being as great as 2808 in the training set), we had to use A100 fat (with twice as much memory) to make sure the input can fit into the memory.

¹https://github.com/atomicarchitects/equiformer_v2

Test set	Model	Inference time (CPU)		Inference time (GPU)	
		DFT	MLIP	DFT	MLIP
w/o rep.	<i>Original</i> (30.8M)	890 s	0.80 s	–	0.07 s
	<i>Small</i> (4.8M)		0.29 s	–	0.04 s
w/ rep. ($\approx 6\times$)	<i>Small</i> (4.8M)	10,264 s	1.16 s	–	0.06 s

Table 3: Comparison of CPU (Intel Xeon Gold 6130 CPU @ 2.10GHz) and GPU (A100 fat) inference times between traditional DFT calculations and MLIP-based predictions using EquiformerV2. The inference time is calculated for a median-sized sample in *w/o rep.* (96 atoms) and *w/ rep.* (396 atoms) test sets, respectively.

Training set	Model	Test w/o rep., MAE			Test w/ rep. ($\approx 6\times$), MAE		
		Energy (meV) ↓	Energy/atom (meV) ↓	Force (meV/Å) ↓	Energy (meV) ↓	Energy/atom (meV) ↓	Force (meV/Å) ↓
w/o rep.	<i>Small</i> (4.8M)	512.0 ± 35.8 (477.0)	5.9 ± 0.4 (5.5)	44.9 ± 0.7 (44.3)	3063.6 ± 215.0 (2853.3)	5.9 ± 0.4 (5.5)	44.7 ± 0.9 (43.9)
w/ rep. ($\approx 6\times$)	<i>Small</i> (4.8M)	488.3 ± 3.4 (483.6)	5.6 ± 0.0 (5.6)	44.9 ± 0.5 (44.3)	2922.3 ± 29.3 (2881.2)	5.6 ± 0.1 (5.5)	44.5 ± 0.5 (44.0)

Table 4: Cross-evaluation of EquiformerV2 *small* on test data *without* and *with* artificial repetitions. Reported are the mean and standard deviation over three runs, with the best result in parentheses.

4.2 Results

Table 2 summarises the results on the two datasets where energy MAE and force MAE refer to $\mathcal{L}_E$ and $\mathcal{L}_f$, respectively, in Eq. (4) additionally averaged over all the systems in the corresponding dataset. The evaluation on the test set is conducted using the models with the best validation performance.

We observe that the relative performance of the models differs on the validation and test sets. For example, *original PT* exhibits lower performance on validation compared to *original*, but better on the test set. Notably, all the models performed worse on the test set in comparison to the validation set. We attribute this to the challenge posed by the different distributions of systems in the training/validation and test sets (see Figure 2).

Generalisation to larger systems To assess the ability of the model to generalise to larger systems, we perform cross-evaluation using the *small* model. Specifically, we train the model on one version of the dataset (*with* or *without* artificial repetitions) and evaluate it on the other. Due to the GPU memory constraints, we are not able to perform this analysis on the *original* model. The results are summarised in Table 4, where the energy MAE is also further averaged by the number of atoms (energy/atom). Comparing the performance of the *small* model trained on the dataset *without repetitions* on the test sets *without* and *with repetitions*, we observe that, albeit increased, the energy error of EquiformerV2 scales linearly with the number of atoms (see the energy/atom MAE column), while the force error remains virtually the same. At the same time, the energy error of the *small* model trained on the data *with repetitions*, on average, improves over *small w/o rep.* by 4.63% on the test data *without repetitions*, and by 4.61% on the test data *with repetitions*, while the energy/atom MAE in both cases reduces by 5.08%.

Overall, this indicates that the model trained on the data *without repetitions* does not require training with augmentation (artificially repeated systems) to maintain the per-atom energy prediction accuracy. Besides, training with system size augmentation (*with repetitions*) yields only modest improvements in energy/atom and total energy prediction accuracy on average, although it considerably reduces variance across runs, which comes at the cost of prolonged training (see Table 2).

Speed comparison We display the comparison between the inference time of the trained models and DFT in Table 3. We note that even on a CPU, both the *original* and *small* models run significantly faster than DFT: 1113 and 3069 times, respectively, for a median-sized sample in the original dataset, and 8848 times for a median-sized sample in the larger-system dataset (*small* model).

Qualitative results In our computational workflow, DFT calculations are replaced by the trained EquiformerV2 model for the analysis of 2D MXenes. As shown in Figure 3, the model accurately reproduces DFT geometry optimisation trajectories. The agreement is particularly strong for HCOOH and CO₂ adsorption, while minor systematic energy deviations are observed for H₂O adsorption and the pristine MXene, with force predictions remaining consistent with DFT throughout. Additional results and analysis are presented in Appendix B.

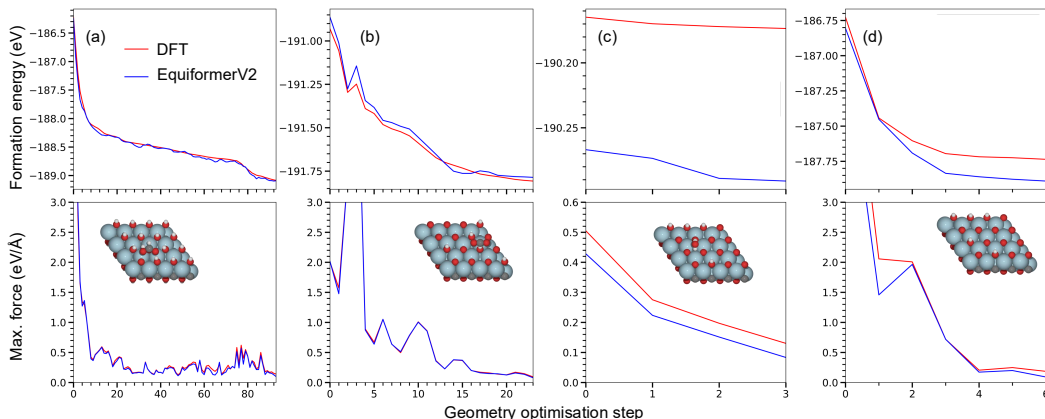


Figure 3: Formation energies and maximum atomic forces along the trained EquiformerV2 (*original*, trained on data *w/o rep.*) geometry optimisation trajectories for four test systems (shown as insets) – (a) HCOOH, (b) CO₂ and (c) H₂O adsorbed on Ti₂CT_y MXenes with different termination configurations, and (d) Ti₂CT_y MXene distorted from its equilibrium geometry – benchmarked against static DFT calculations performed at each step. The comparison illustrates the model’s accuracy in reproducing DFT energetics and forces during relaxation. ASE [8] was used for structural visualisation.

5 Conclusions

In this work, we have generated a dataset for catalysis on Ti₂CT_y MXenes, called CATALIUST Ti₂C-MXENE, comprising a total of 60,000 DFT calculations — 50,000 for training and 10,000 for testing — as well as its version with artificially repeated systems. Using this dataset, we have performed equivariant modelling for interatomic potentials (i.e. predicting atomic forces and formation energies of molecular systems) using EquiformerV2, and accelerated our computational workflow for catalysis on 2D MXenes 1113 – 8848-fold (on a CPU), while maintaining the desired quality that the DFT method provides. Furthermore, we have shown that the energy error of the trained model scales linearly with the size of the input systems, while the force error remains the same, which is a necessity for a robust model approximating DFT.

Limitations In this work, we have focused exclusively on the EquiformerV2 model due to its state-of-the-art performance on the OC20 dataset. Future studies could explore other variants or extensions of EquiformerV2, such as integrating DeNS [46], which might come with additional computational requirements. Besides, there exist competing methods on other benchmarks, such as GotenNet [44] for molecular systems; however, adopting it for our application might not be a trivial task. Additionally, the hyperparameter search we have conducted is non-exhaustive due to computational limitations. In generating the DFT training data, we balanced numerical accuracy and computational cost by adopting a moderate k -point sampling density, which is particularly important for total energies. Although a denser grid would shift the absolute formation energies, the correction is expected to be smaller than the current model’s energy loss. If higher energy accuracy is required, the dataset should be recalculated with tighter k -point convergence. Another limitation concerns the molecular systems included. The chosen molecules represent prototypical catalytic reactions, providing chemically meaningful examples for MXene interactions. Future work could extend the molecular scope to additional reactants and intermediates relevant to other catalytic processes.

Broader impact This work accelerates the analysis of catalysis on 2D MXenes while maintaining robustness through the trained model. Faster evaluations of MXene catalysis open opportunities for identifying more efficient catalysts, supporting efforts to reduce the carbon footprint of the chemical industry. In addition, the generated dataset provides a resource for the community to develop and benchmark models for catalysis on Ti₂CT_y MXenes and related materials.

Acknowledgments

This work was supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP) and the Wallenberg Initiative Materials Science for Sustainability (WISE), by the Swedish Research Council through a grant for the project Uncertainty-Aware Transformers for Regression Tasks in Computer Vision (2022-04266), and the strategic research environment ELLIIT. The computations were enabled by resources provided by the National Academic Infrastructure for Supercomputing in Sweden (NAISS) partially funded by the Swedish Research Council through grant agreement no. 2022-06725, and by the Berzelius resource provided by the Knut and Alice Wallenberg Foundation at the National Supercomputer Centre.

References

- (1) Wang, G.; Wang, C.; Zhang, X.; Li, Z.; Zhou, J.; Sun, Z. Machine learning interatomic potential: Bridge the gap between small-scale models and realistic device-scale simulations. *Iscience* **2024**, *27*.
- (2) Bronstein, M. M.; Bruna, J.; LeCun, Y.; Szlam, A.; Vandergheynst, P. Geometric deep learning: going beyond Euclidean data. *IEEE Signal Processing Magazine* **2017**, *34*, 18–42.
- (3) Bronstein, M. M.; Bruna, J.; Cohen, T.; Veličković, P. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. *arXiv preprint arXiv:2104.13478* **2021**.
- (4) Deringer, V. L.; Caro, M. A.; Csányi, G. Machine learning interatomic potentials as emerging tools for materials science. *Adv. Mater.* **2019**, *31*, 1902765.
- (5) Chen, C.; Ong, S. P. A universal graph deep learning interatomic potential for the periodic table. *Nat. Comput. Sci.* **2022**, *2*, 718–728.
- (6) Yang, H.; Hu, C.; Zhou, Y.; Liu, X.; Shi, Y.; Li, J.; Li, G.; Chen, Z.; Chen, S.; Zeni, C., et al. MatterSim: A deep learning atomistic model across elements, temperatures and pressures. *arXiv preprint arXiv:2405.04967* **2024**.
- (7) Batzner, S.; Musaelian, A.; Sun, L.; Geiger, M.; Mailoa, J. P.; Kornbluth, M.; Molinari, N.; Smidt, T. E.; Kozinsky, B. E (3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nat. Commun.* **2022**, *13*, 2453.
- (8) Larsen, A. H. et al. The atomic simulation environment—a Python library for working with atoms. *J. Phys. Condens. Matter* **2017**, *29*, 273002.
- (9) Chanussot, L. et al. Open catalyst 2020 (OC20) dataset and community challenges. *ACS Catal.* **2021**, *11*, 6059–6072.
- (10) Tran, R. et al. The Open Catalyst 2022 (OC22) Dataset and Challenges for Oxide Electrocatalysts. *ACS Catal.* **2023**, *13*, 3066–3084.
- (11) Niu, K.; Chi, L.; Rosen, J.; Björk, J. C–H activation of light alkanes on MXenes predicted by hydrogen affinity. *Phys. Chem. Chem. Phys.* **2020**, *22*, 18622–18630.
- (12) Niu, K.; Chi, L.; Rosen, J.; Björk, J. Structure-activity correlation of Ti₂CT₂ MXenes for C–H activation. *J. Phys. Condens. Matter* **2021**, *33*, 235201.
- (13) Parui, A.; Srivastava, P.; Singh, A. K. Selective reduction of CO₂ on Ti₂C(OH)₂ MXene through spontaneous crossing of transition states. *ACS Appl. Mater. Interfaces* **2022**, *14*, 40913–40920.
- (14) Guo, H.; Lee, S. G. Machine learning-guided discovery of thermodynamically stable single-atom catalysts on functionalized MXenes for enhanced oxygen reduction and evolution reactions. *J. Mater. Chem. A* **2025**, *13*, 22730–22744.
- (15) Lin, G.; Guo, T.; Lin, W.; Fan, H.; Guo, L.; Zhang, Z.; Li, B.; Wang, J.; Ji, H.; Song, W.; Fu, J. Machine learning accelerated screening advanced single-atom anchored MXenes electrocatalyst for nitrogen fixation. *ACS Catal.* **2025**, *15*, 13534–13548.
- (16) Liao, Y.-L.; Wood, B.; Das, A.; Smidt, T. EquiformerV2: Improved Equivariant Transformer for Scaling to Higher-Degree Representations. *The Twelfth International Conference on Learning Representations* **2024**.
- (17) Hohenberg, P.; Kohn, W. Inhomogeneous electron gas. *Phys. Rev.* **1964**, *136*, B864–B871.
- (18) Kohn, W.; Sham, L. J. Self-consistent equations including exchange and correlation effects. *Phys. Rev.* **1965**, *140*, A1133–A1138.
- (19) Dion, M.; Rydberg, H.; Schröder, E.; Langreth, D. C.; Lundqvist, B. I. Van der Waals density functional for general geometries. *Phys. Rev. Lett.* **2004**, *92*, 246401.
- (20) Hamada, I. van der Waals density functional made accurate. *Phys. Rev. B* **2014**, *89*, 121103.
- (21) Niu, K.; Björk, J.; Rosen, J. First-principles exploration of Sc- and Y-based MXenes with halogen terminations. *npj 2D Mater. Appl.* **2025**, *9*, 69.
- (22) Chen, L.; Rosen, J.; Björk, J. A density functional benchmark for dehydrogenation and dehalogenation reactions on coinage metal surfaces. *ChemPhysChem* **2025**, *26*, e202400865.
- (23) Tran, F.; Kalantari, L.; Traoré, B.; Rocquefelte, X.; Blaha, P. Nonlocal van der Waals functionals for solids: Choosing an appropriate one. *Phys. Rev. Materials* **2019**, *3*, 063602.
- (24) Van Gool, L.; Moons, T.; Pauwels, E.; Oosterlinck, A. Vision and Lie’s approach to invariance. *Image and vision computing* **1995**, *13*, 259–277.
- (25) Weiler, M.; Forré, P.; Verlinde, E.; Welling, M. Equivariant and coordinate independent convolutional networks. *A Gauge Field Theory of Neural Networks* **2023**, 110.
- (26) Fukushima, K. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological cybernetics* **1980**, *36*, 193–202.
- (27) Zhang, W.; Tanida, J.; Itoh, K.; Ichioka, Y. Shift-invariant pattern recognition neural network and its optical architecture. *Proceedings of annual conference of the Japan Society of Applied Physics* **1988**, 564.

- (28) LeCun, Y.; Boser, B.; Denker, J. S.; Henderson, D.; Howard, R. E.; Hubbard, W.; Jackel, L. D. Backpropagation applied to handwritten zip code recognition. *Neural computation* **1989**, *1*, 541–551.
- (29) Weiler, M.; Geiger, M.; Welling, M.; Boomsma, W.; Cohen, T. S. 3D steerable CNNs: Learning rotationally equivariant features in volumetric data. *Advances in Neural Information Processing Systems* **2018**, 10381–10392.
- (30) Anderson, B.; Hy, T. S.; Kondor, R. Cormorant: Covariant molecular neural networks. *Advances in Neural Information Processing Systems* **2019**, 14510–14519.
- (31) Bökman, G.; Kahl, F.; Flinth, A. Zz-net: A universal rotation equivariant architecture for 2d point clouds. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* **2022**, 10976–10985.
- (32) Cohen, T. S.; Geiger, M.; Köhler, J.; Welling, M. Spherical CNNs. *arXiv preprint arXiv:1801.10130* **2018**.
- (33) Esteves, C.; Allen-Blanchette, C.; Makadia, A.; Daniilidis, K. Learning SO(3) Equivariant Representations With Spherical CNNs. *CoRR* **2017**.
- (34) Fuchs, F.; Worrall, D.; Fischer, V.; Welling, M. SE(3)-Transformers: 3D Roto-Translation Equivariant Attention Networks. *Advances in Neural Information Processing Systems* **2020**, 33, ed. by Larochelle, H.; Ranzato, M.; Hadsell, R.; Balcan, M. F.; Lin, H., 1970–1981.
- (35) Melnyk, P.; Robinson, A.; Felsberg, M.; Wadenbäck, M. TetraSphere: A Neural Descriptor for O(3)-Invariant Point Cloud Analysis. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* **2024**, 5620–5630.
- (36) Kresse, G.; Furthmüller, J. Efficient iterative schemes for *ab initio* total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **1996**, *54*, 11169–11186.
- (37) Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **1994**, *50*, 17953–17979.
- (38) Kresse, G.; Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **1999**, *59*, 1758–1775.
- (39) Liao, Y.-L.; Smidt, T. Equiformer: Equivariant Graph Attention Transformer for 3D Atomistic Graphs. *The Eleventh International Conference on Learning Representations* **2023**.
- (40) Ruhe, D.; Brandstetter, J.; Forré, P. Clifford Group Equivariant Neural Networks. *Thirty-seventh Conference on Neural Information Processing Systems* **2023**.
- (41) Melnyk, P.; Felsberg, M.; Wadenbäck, M.; Robinson, A.; Le, C. On Learning Deep O(n)-Equivariant Hyperspheres. *Proceedings of the 41st International Conference on Machine Learning* **2024**, 235, 35324–35339.
- (42) Satorras, V. G.; Hoogeboom, E.; Welling, M. E(n) equivariant graph neural networks. *International conference on machine learning* **2021**, 9323–9332.
- (43) Passaro, S.; Zitnick, C. L. Reducing SO(3) Convolutions to SO(2) for Efficient Equivariant GNNs. *Proceedings of the 40th International Conference on Machine Learning* **2023**, 202, ed. by Krause, A.; Brunskill, E.; Cho, K.; Engelhardt, B.; Sabato, S.; Scarlett, J., 27420–27438.
- (44) Aykent, S.; Xia, T. Gotennet: Rethinking efficient 3d equivariant graph neural networks. *The Thirteenth International Conference on Learning Representations* **2025**.
- (45) Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L., et al. PyTorch: An imperative style, high-performance deep learning library. *Advances in Neural Information Processing Systems* **2019**, 8024–8035.
- (46) Liao, Y.-L.; Smidt, T.; Shuaibi, M.; Das, A. Generalizing Denoising to Non-Equilibrium Structures Improves Equivariant Force Fields. *Transactions on Machine Learning Research* **2024**.

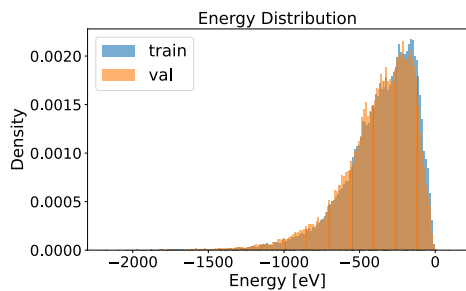
A Comparison with the OC20 dataset

In Figure 4, we compare the training and validation splits of OC20² with the corresponding splits of the proposed CATALIUST Ti2C-MXENE (*without repetitions*). This comparison is limited to the training and validation subsets, as the OC20 test set is private, and the labels are not publicly available. Additional details about OC20 can be found in [9].

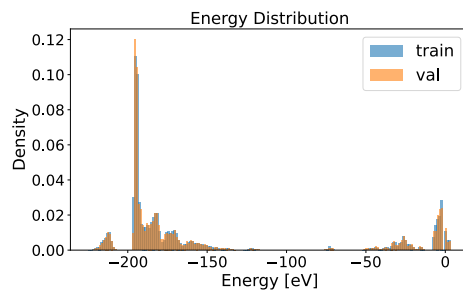
A key distinction between the CATALIUST Ti2C-MXENE and OC20 datasets lies in the chemical nature of the systems they contain. OC20 primarily focuses on molecules adsorbed on extended periodic catalyst slabs representing bulk surfaces, whereas CATALIUST Ti2C-MXENE contains 2D Ti₂C MXenes interacting with molecular species. The surface chemistry of MXenes depends strongly on the configuration of surface terminations, which our dataset explicitly samples, enabling the description of MXene chemistry under diverse conditions. Furthermore, elements such as oxygen and carbon appear both as part of the solid MXene and in molecular adsorbates, introducing an additional level of complexity. This diversity requires the model to distinguish between chemically identical atoms in different environments, which is less pronounced in OC20.

The most informative distinction between the datasets is in the force distributions. Unlike the energy histograms, which are strongly influenced by system size and composition, force magnitudes reflect how far configurations are from equilibrium. CATALIUST Ti2C-MXENE exhibits substantially larger forces compared to OC20, indicating that the dataset samples geometries further from local minima. This diversity is advantageous for model training, as it allows learning of far-from-equilibrium chemistry, in contrast to OC20 where most adsorbates lie near relaxed surface sites. Regarding system size, OC20 is heterogeneous, with configurations ranging widely in atom count, while our dataset is more homogeneous, with most systems around ≈ 100 atoms. Thus, the train/val mismatch is more noticeable in OC20, with our dataset presenting a more consistent split. In the absence of the OC20 test set, it is difficult to draw additional conclusions about the relative dataset difficulty.

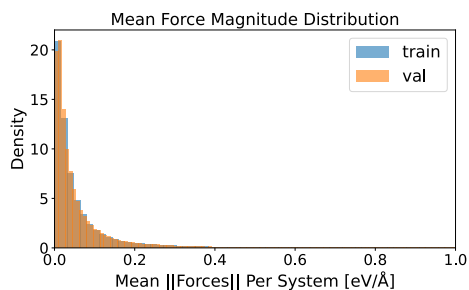
²The OC20 dataset can be downloaded at <https://github.com/facebookresearch/fairchem/blob/main/docs/catalysts/datasets/oc20.md>. The histograms were computed from the S2EF table, using all of the validation data, and the 20M training set.



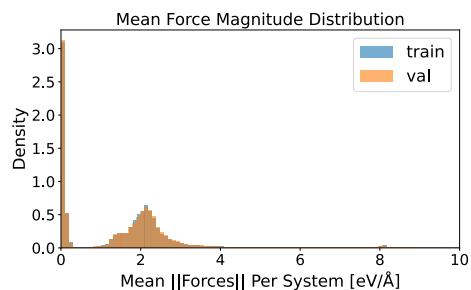
(a) Energies (OC20)



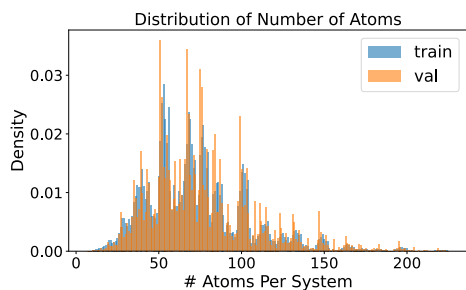
(b) Energies (Ours)



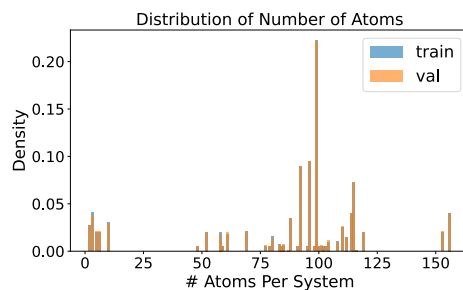
(c) Forces (OC20)



(d) Forces (Ours)



(e) Number of atoms (OC20)



(f) Number of atoms (Ours)

Figure 4: Distributions of energies, mean force magnitudes, and number of atoms for the OC20 training and validation subsets in comparison to our proposed dataset *without repetitions*.

B Additional Qualitative Results

In Figure 5, we present qualitative results alike those in Figure 3, but for the rest of the models trained on the data *without repetitions*. Similar to the results showcased for the *original* model in Figure 3, the *small* model (see Figure 5, bottom) can also quite accurately reproduce DFT geometry optimisation trajectories for HCOOH and CO₂ adsorption. However, for H₂O adsorption and the pristine MXene, the force predictions of the *small* model are notably poor—significantly worse than the overall force error observed for the *small* model evaluated on the test set—indicating that these errors are system-specific rather than general model limitations. Systematic energy deviations are also observed in these cases. In either case, the *original* model can be employed for consistently accurate force prediction.

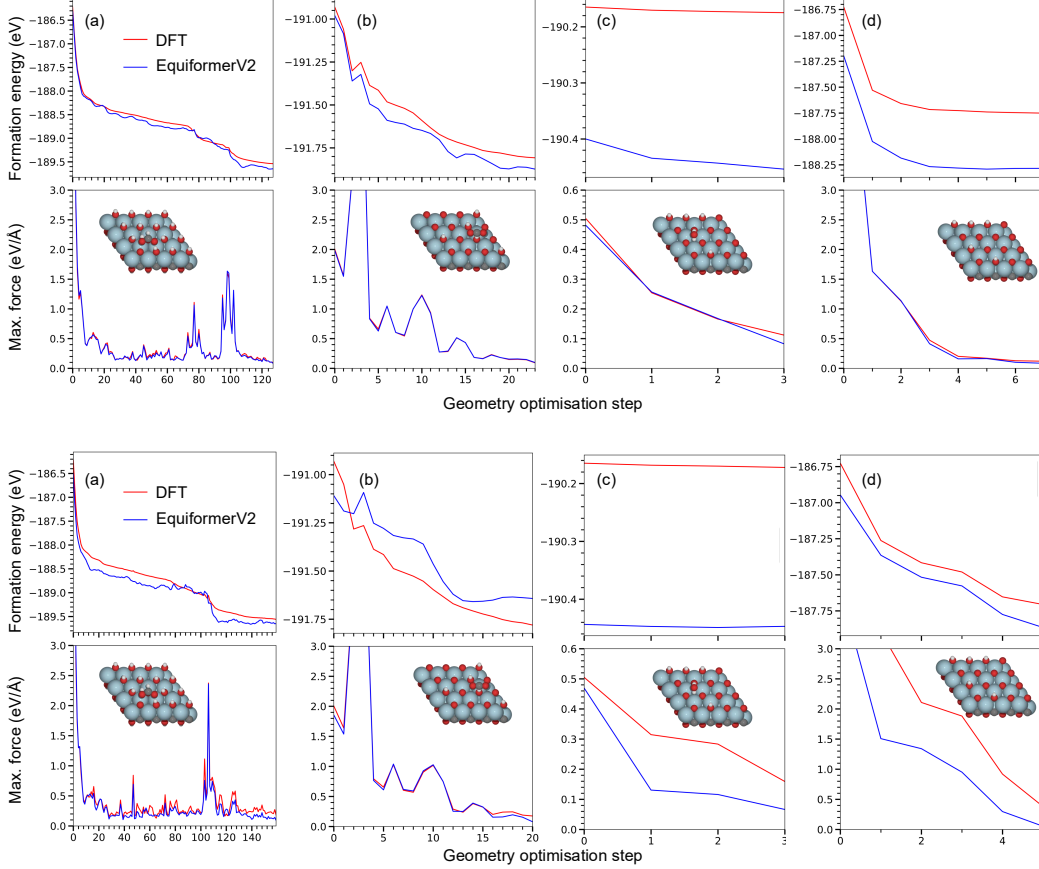


Figure 5: Formation energies and maximum atomic forces along the trained EquiformerV2 (top: *original PT*, bottom: *small*, both trained on data w/o rep.) geometry optimisation trajectories for four test systems (shown as insets) — (a) HCOOH, (b) CO₂ and (c) H₂O adsorbed on Ti₂CT_y MXenes with different termination configurations, and (d) Ti₂CT_y MXene distorted from its equilibrium geometry — benchmarked against static DFT calculations performed at each step. ASE [8] was used for structural visualisation.