ARTICLE TYPE

Cite this: DOI: 00.0000/xxxxxxxxxx

Nature of molybdenum carbide surfaces for catalytic hydrogen dissociation using machine-learned potentials: an ensemble-averaged perspective†

Woodrow N. Wilson,[a] John Michael Lane,[a] Chinmoy Saha,[a] Sony Severin,[a] Vivek S. Bharadwaj,*[b] and Neeraj Rai*[a]

Received Date
Accepted Date

DOI: 00.0000/xxxxxxxxxx

Molybdenum carbides with an electronic structure similar to noble metals have gained attention as a promising low-cost catalyst for biomass valorization and the hydrogen evolution reaction. However, our fundamental understanding of the catalyst surface and how different phases of these catalysts behave at varying reaction conditions is limited to ground state density functional theory calculations as ab initio molecular dynamics (AIMD) is computationally prohibitive at relevant length and time scales. In this work, we train a multi-atomic cluster expansion (MACE) machine-learned interatomic potentials (MLIP) to study hydrogen dissociation and dynamics over Mo, δ-MoC, α-Mo_2C, and β-Mo_2C surfaces at varying temperatures and hydrogen partial pressures. Our simulations identify unique and different molecular and atomic hydrogen adsorption sites on different surfaces that do not depend on the temperature. At low hydrogen pressures, the surface coverage is monolayer, which transitions to two-layer adsorption at higher pressures. We find that atomic hydrogen diffusion and recombinations are preferred over molybdenum atom hollow sites, while the diffusion over carbon-terminated facets was negligible, signifying particularly strong C–H interactions. In contrast, molecular hydrogen adsorption occurs mostly atop Mo or the bridging sites. At a comparable hydrogen loading, β-Mo_2C (001) is the most active surface for hydrogen dissociation reaction. This work provides insights into the dynamic nature of the hydrogen dissociation chemistry and the diversity of hydrogen adsorption sites on molybdenum carbides.

1 Introduction

A significant challenge facing biomass valorization is lowering the oxygen content of bio-oil. A high oxygen content in bio-oil is detrimental as it increases viscosity due to enhanced hydrogen bonding and reduces its lifetime due to oxidative reactions which can occur at ambient conditions[1–3]. Noble metal catalysts have been shown to cleave the oxygen-carbon bonds in biomass thereby reducing its oxygen content, as well as enable hydrogenation reactions to convert reactive ketones and aldehydes to less-reactive alcohols or other compounds.[4–8]. However, the scarcity of noble metal catalysts makes this an economically infeasible option for biomass valorization. A similar problem exists for the hydrogen evolution reaction (HER), where platinum is regarded as one of the best catalysts for the reaction[9–11]. However, due to high cost and lack of availability, finding alternative catalysts with performance similar to platinum and other noble metal catalysts is an active research area in the catalysis community.

A possible alternative for noble metal catalysts is transition metal carbides (TMCs). TMCs consist of interstitial carbon atoms in the lattice of pure metals like tungsten or molybdenum[12–14]. The resulting material has metallic and covalent bonding character. By adjusting the doping of carbon atoms and the phase of the material, the electronic structure of the metal can be tuned to be like that of a noble metal catalyst and have similar cat-

[a] *Dave C. Swalm School of Chemical Engineering and Center for Advanced Vehicular Systems, Mississippi State University, Mississippi State, MS, 39762, USA. Fax: +1-662-325-2482; Tel: +1-662-325-0790; E-mail: neerajrai@che.msstate.edu*

[b] *Renewable Resources and Enabling Sciences Center, National Renewable Energy Laboratory, Golden, CO, 80401, USA.*

† Supplementary Information available: Supplementary material includes crystal symmetry and lattice details of the molybdenum carbides we studied, optimized lattice parameters, source code for adding a ghost atom grid to a training set, an example Python script for running MD with the ghost atom grid, numerical values for activation energy barriers for diffusion and the dissociation-recombination reaction, active learning post processing of Mo (110), α-Mo_2C (001), and β-Mo_2C (101) at the temperatures studied, adsorption heatmaps for Mo (110), α-Mo_2C (001), and β-Mo_2C (101) at the temperatures studied, *MSD* and the corresponging Arrhenius analysis for Mo (110), α-Mo_2C (001), and β-Mo_2C (101), and the Arrhenius analysis of the dissociation-recombination reaction for Mo (110), α-Mo_2C (001), and β-Mo_2C (101). Movies of the diffusion simulations. See DOI: 00.0000/00000000.

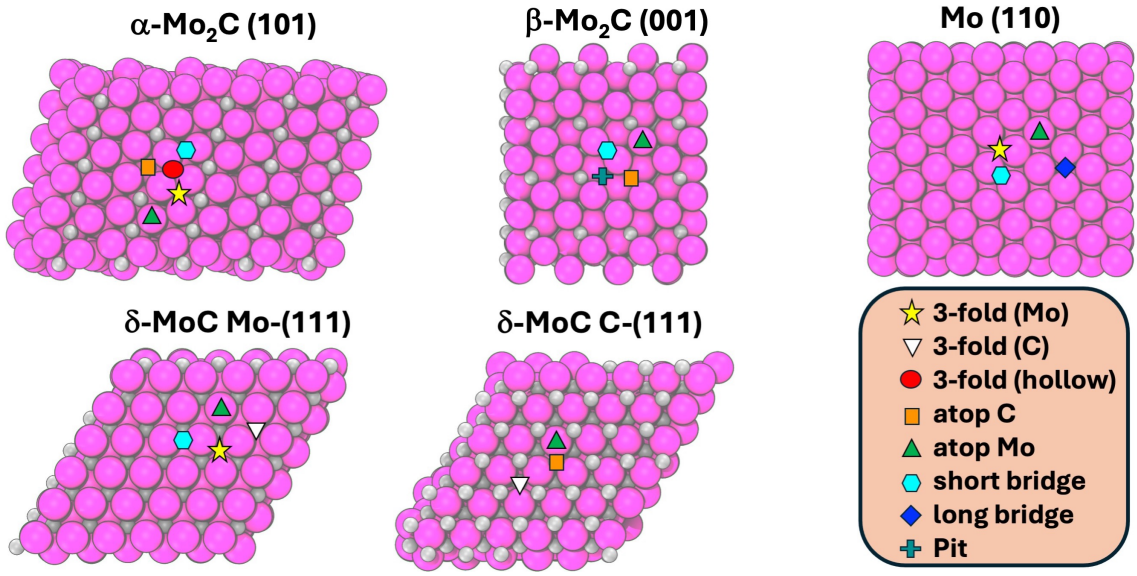

Fig. 1 Active sites on Mo (110), α-Mo_2C (101), β-Mo_2C (001), δ-MoC Mo-(111), and δ-MoC C-(111).

alytic performance as shown originally by Levi and Boudart in their pioneering work with tungsten carbide[15]. Because of this, transition metal carbides have gained attention as a catalyst for biomass valorization[16–21] and deoxygenation[22–24], CO_2 conversion[25–28], the HER[29–32], and other reactions[33–42]. While these materials show promise from an electronic structure perspective, specific challenges must be resolved for wider adoption. TMCs are prone to deactivation via coking, hydrogen and oxygen can either facilitate reactions or poison the catalyst, and depending on the synthesis conditions, they can form various phases with different stoichiometry. Therefore, a fundamental understanding of the physical structure of the surface and phases of these catalysts at synthesis and reaction conditions is necessary to design processes that maximize their lifetime and performance.

Ab initio molecular dynamics (AIMD) has had significant success in modeling reactive systems in catalysis including zeolites[43–47], transition metals[8,48], and homogeneous catalysis[49–51]. However, modeling reactive systems at length and time scales relevant to catalysis can become computationally prohibitive. It is especially true for transition metal catalysis where, due to the number of electrons and the numerical difficulties of solving the Kohn-Sham equations with d-block elements, models are limited to the simplest systems that either neglect environmental effects such as coverage or solvation entirely or only include those effects with short AIMD simulations[8,46,47] or a QM/MM based approach.[52–55]. However, even with these methods and approximations, such calculations still require significant computational resources to sample with MD and obtain a statistically significant trajectory. Recently, machine-learned interatomic potentials (MLIPs) trained on density functional theory data from both single-point calculations and AIMD trajectories have shown success in modeling catalysis with molecular dynamics[56–61], enhanced sampling[62,63], and global optimizations[64].

In this work, we simulate the hydrogen dissociation reaction over molybdenum and molybdenum carbide catalysts using reactive molecular dynamics (MD) simulations with MLIPs developed with MACE[65,66], an equivariant message passing neural network. Hydrogen dissociation is integral to biomass valorization reactions and the HER, with molybdenum carbides gaining significant attention as promising catalysts. We study systems with differing initial hydrogen molecules in the vapor phase and at a range of temperatures to capture both hydrogen partial pressure and temperature effects. We also simulate different catalyst surfaces made of pure Mo, δ-MoC, α-Mo_2C, and β-Mo_2C to capture phase and composition effects. The message-passing formalism and local models are inherently challenged when describing atomic environments for systems with low densities where atoms are outside the cutoff distance to determine neighbors. To circumvent this, we utilize a fixed grid of ghost atoms to facilitate message passing to atoms outside the cutoff radius to determine atomic neighbors but within the effective cutoff radius of the MACE model. Evaluation of the models is done by post-processing of MD trajectories with active learning, and in cases where the error in energy uncertainty becomes larger than 43 meV (1 kcal/mol and the threshold for chemical accuracy) or if the MD was unstable, we ran additional short AIMD simulations from those frames and updated our existing model.

2 Methods

2.1 System details

Crystal structures for Mo, δ-MoC, α-Mo_2C, and β-Mo_2C were obtained from the Materials Project database[67]. Two naming conventions are often used when describing the different phases of molybdenum carbides: an early definition by Christensen[68] or the notation defined by the Joint Committee on Power Diffraction Standards (JCPDS) data files[69]. In this work, we followed the JCPDS convention where β-Mo_2C is the orthorhombic phase of Mo_2C. More details about the crystal symmetries and lattice parameters of Mo, δ-MoC, α-Mo_2C, and β-Mo_2C are presented in Table 1 to eliminate any confusion for what phase of molybdenum

Table 1 Crystal symmetry, facets, and supercell dimensions of molybdenum and molybdenum carbides and the initial number of H_2 molecules in the vapor phase simulated with AIMD

Material	Materials Project ID	Crystal System	Space Group	Facet	Supercell	Layers	H_2 Loadings
Mo	129	Cubic	Im3m	(110)	5 x 6	6	25, 29, 61, 85
δ-MoC	2746	Cubic	Fm3m	(111)	4 x 4	6	23, 29, 83, 124
α-Mo_2C	1221498	Trigonal	P3m1	(101)	6 x 6	3	25, 30, 90
β-Mo_2C	1552	Ortho.	Pbcn	(001)	4 x 4	6	29, 34, 75, 107

carbides we are studying. To orient the readers unfamiliar with the surface topology of molybdenum carbides, we illustrate in Fig. 1 different active sites on Mo (110), α-Mo_2C (101), β-Mo_2C (001), δ-MoC Mo-(111), and δ-MoC C-(111). We will refer to these different active sites throughout the article.

After obtaining the lattice parameters and coordinates for each crystal from the Materials Project[67], we performed a cell optimization enforcing crystal symmetry with CP2K using the level of DFT theory described in the next section. Optimized lattice vectors are presented in Table S1. After cell optimization, we created surfaces with specific Miller indices and dimensions shown in Table 1. We selected these facets based on the dominant peaks in the XRD patterns available at the Materials Project[67] and what other scientists have used in the literature[18,19,29–32,39,40,42,70–73].

Each surface had a vacuum layer of 30 Å to give sufficient space to add H_2 molecules (see Fig. 2). We created systems with different initial hydrogen loadings in the vapor phase to run ab initio and large-scale reactive MD with the machine-learned interatomic potential to capture the effect of different hydrogen partial pressures. In this work, we aimed to investigate how the adsorption, diffusion, and dissociation reaction varies across these phases of molybdenum carbide compared to pure molybdenum at different hydrogen loadings and temperatures. Unfortunately, choosing a priori hydrogen loadings, which will lead to identical pressures across all four materials considered in this work, is difficult. Thus, we approximate the vapor pressure of H_2 for each catalyst and hydrogen loading during the production run of our simulations. Pressure is calculated using the ideal gas law considering the number of H_2 molecules 10 Å away from the surface and is reported in Table S3 of the SI†. We note that the (111) facet of δ-MoC leads to two different surface terminations: a molybdenum terminated facet (Mo-(111)) where hollow sites are present with a subsurface molybdenum atom and three molybdenum atoms coordinate the site and a carbon terminated facet (C-(111)) where carbon atoms occupy the hollow sites and protrude from the catalyst surface. Since these two surface terminations can occur on the same (111) facet, we investigate their differences in this work. We acknowledge that previous work by Li and Reuter[74] and separately by Wang et al.[72] investigated the surface stability and reconstruction of molybdenum carbides with ab initio thermodynamics, and the most thermodynamically favorable surface appears to have a mix of Mo and C surface terminations for the (111) facet. However, in this work, we chose to study surfaces with molybdenum-only or carbon-only terminations.

2.2 Ab initio molecular dynamics

We performed cell optimizations and AIMD simulations with CP2K[75,76] (v2022.1) to model hydrogen dissociation/recombination over molybdenum and molybdenum carbide catalysts. QUICKSTEP[77], the density functional theory module of CP2K, implements the Gaussian-plane wave method, which transforms the density matrix obtained from a Gaussian basis set to a plane-wave representation via fast Fourier transform to solve the long-range electrostatic term in the Kohn-Sham Hamiltonian. It enables tractable AIMD simulations of large systems with moderate computational resources. We use PBE with Grimme's D3 empirical dispersion correction (PBE+D3) for the exchange-correlation functional. The double ζ molecular optimized basis set[78] (DZVP-MOLOPT) with an auxiliary plane wave kinetic energy cutoff of 600 Ry and a relative cutoff of 60 Ry was used to model the valence electrons of C and H. In contrast, the short-range variant with less diffuse functions (DZVP-MOLOPT-SR) was used for Mo. Core electrons were modeled with the dual space pseudopotentials of Goedecker, Teter, and Hutter[79] optimized with PBE (GTH-PBE). The orbital transform (OT) algorithm[80] was used to minimize each SCF cycle to a convergence threshold of 10^{-5} a.u. While OT has poor convergence for metallic systems, after trial and error, we found that turning off the always-stable-predictor-corrector[81] (ASPC) algorithm of CP2K to predict the next wavefunction in favor of simply starting the SCF cycle with the wavefunction from the last MD step improved convergence speed significantly even after the MD had progressed. An example input file we used to perform AIMD with CP2K is provided in the SI† in Code Listing S1.

Born-Oppenheimer MD was performed in the canonical (*NVT*) ensemble with a time step of 0.5 fs and a Nose-Hoover chains thermostat[82] at 1000 K. Running the AIMD simulations at such a high temperature increases the chance of observing rare events like hydrogen surface diffusion and dissociation/recombination. We ran ab initio MD for at least 4,000 steps (2 ps) for each hydrogen loading and material studied.

2.3 Machine learned interatomic potential training and model evaluation

For each catalyst surface, we trained one machine-learned interatomic potential using the AIMD trajectory for the specific catalyst facet with the MACE equivariant message passing neural network architecture[65,66]. MACE, or multi-atomic cluster expansion, embeds atomic environments with invariant and equivariant features and constructs high body order messages passed from atom to atom in the spirit of the atomic cluster expansion[83]. MACE has been previously shown to be able to make accurate force predic-

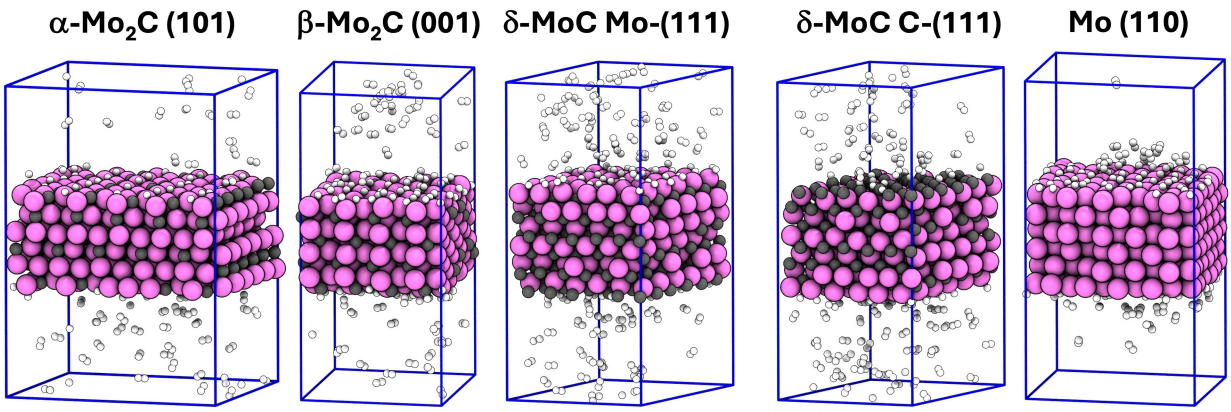

Fig. 2 Visualization of systems considered in the present work. The snapshot of each catalyst surface with the highest hydrogen loading. The magenta, white, and grey spheres represent molybdenum, hydrogen, and carbon, respectively.

tions for a wide range of materials, molecules, and systems in vacuum and the condensed phase[60,84]. MACE models were trained with 2 message passing layers (N_{layers}), a maximum body correlation (ν) of 2, a maximum spherical harmonics degree (l_{max}) of 1, a neighbor distance cutoff (r_{cut}) of 5 Å, and 128 features (N_{feat}). This combination of hyperparameters includes up to 7 body order interactions. To evaluate hyperparameter sensitivity, we also trained models with (ν, l_{max}, N_{feat}) equal to (3, 0, 128), (3, 0, 256), and (3, 1, 128) keeping the number of layers and distance cutoff fixed at 2 and 5Å, respectively. An example shell script for the MACE models we trained is provided in the SI† in Code Listing S2.

One unique feature of message-passing neural networks compared to local models is the receptive field or effective cutoff radius. The effective cutoff is the number of layers of the network multiplied by the distance cutoff used to determine if atoms are neighbors:

$$r_{eff} = N_{layers} * r_{cut} \quad (1)$$

While atoms are only considered neighbors and directly pass messages to each other if they are within a distance $r \leq r_{cut}$, messages can be passed to atoms up to a distance r_{eff} from each other by proxy due to having multiple layers of message passing. This is an advantage for message-passing neural networks in that atoms can interact with each other up to distances that are comparable with traditional classical potentials. However, this relies on at least one other atom being within a distance r_{cut} to pass messages up to a distance r_{eff}, and it can be detrimental for less dense systems where often no atoms might be present within the r_{cut} distance. In these cases, as far as the model is concerned, the atom is isolated in a vacuum. This will become problematic for the treatment of gas-phase heterogeneous catalysis, where long-range effects are an important consideration, as atoms and molecules can be influenced by a catalyst surface several nanometers away from the surface. A similar phenomenon was observed by Kovács et al. when modeling a bucky-ball catcher system, a 2-layer 3 Å cutoff predicted the system dissociating into the bucky-ball and the catcher, whereas a single layer 6 Å cutoff network predicted a bound state[84]. Both networks had the same effective cutoff radius of 6 Å, but only the latter neural network could predict the bound state due to the bucky-ball being 3.5 Å away from the catcher.

To circumvent this issue, we utilize a ghost atom grid whose only function is to facilitate message passing from the surface to the molecules in the gas phase or molecules within the gas phase. These ghost atoms are fixed in space and do not affect the dynamics of the system other than serving as extra nodes for messages to pass. An example code for generating an equidistant ghost atom grid and adding it to a database of XYZ coordinates is provided in the SI† in Code Listing S3. In short, the user sets a target resolution specifying the spacing between atoms on the grid. Then, for each frame of the database, points are evenly spaced from [0,1) in scaled coordinates along each lattice vector at the grid resolution specified by the user. After generating these points, a mesh grid is created, and after converting from scaled coordinates to cartesian coordinates, these points serve as the locations for the ghost atoms. We acknowledge that it is possible to design algorithms and assign grid points in a more elegant manner where points are not included in dense regions, like the inside of the catalyst, where message-passing facilitation is not necessary, and points are only included in less dense regions. While this would improve time-to-solution by increasing training speed and MD efficiency, in this work, we only explore using an equidistant grid with a target grid resolution of 5 Å as it is the r_{cut} distance we use in our MACE model. We concede that grid optimization strategies could be a topic for a future study. A graphic illustrating the need for the ghost atom grid and what the grid looks like for δ-MoC with 124 H_2 molecules is shown in Fig. S1 of the SI†. We caution readers about this approach, as singularities can occur when calculating atomic environments where an atom has the same coordinates as a ghost atom grid point. However, the probability of this occurring in an MD simulation representing coordinates with double-precision floating point numbers is negligible, and we did not experience this in our simulations.

The accuracy of the MLIP is determined using the root mean squared errors (RMSE) for energy (E) and relative root mean squared error (RRMSE) for the force (F). RMSE is calculated

as the square root of the average value of the residuals between the actual ground truth value and the predicted value from the machine learning model.

$$E \text{ RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(E_i - \hat{E}_i)^2} \quad (2)$$

where N is the total number of observations, E_i is the DFT energy of training frame i, and $\hat{E}_i$ is the energy predicted by the model. In contrast, RRMSE is the RMSE normalized by the square root of the mean value of target observations squared:

$$F \text{ RRMSE} = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(F_i - \hat{F}_i)^2}}{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{F}_i)^2}} * 100\% \quad (3)$$

RRMSE has an advantage over RMSE as it normalizes the error with the average force acting on the atoms.

2.4 Large scale reactive molecular dynamics

Reactive molecular dynamics with the MACE MLIPs was done with the MD library in the Atomic Simulation Environment[85] (ASE) at 450, 525, 600, 750, 825, and 900 K. A Langevin thermostat[86,87] with a 0.5 fs time step and friction coefficient of 0.01 fs^{-1} was used to integrate Newton's equations of motion for 1,000,000 steps (500 ps). MD was done with the same systems that we had trained our MLIPs on described in Table 1, except for each surface, we also ran additional simulations with a single adsorbed hydrogen atom to study surface diffusion at the limit of low coverage, and another simulation consisting of 53 H_2 molecules in the vapor phase for δ-MoC to test force-field transferability. Snapshots of the last MD frame of each surface with the highest hydrogen loading studied are depicted in Fig. 2.

In each of these MD simulations, we included the ghost atom grid used in the training and kept it fixed throughout the simulation. An example Python script for utilizing the ghost atom grid during an MD simulation is provided in the SI[†] in Code Listing S4. We did encounter instances when some trajectories experienced run-away temperatures resulting in unphysical conformations, likely caused by insufficient training of 'high-energy' or 'repulsive' configurations that were absent in the training data. During post-processing with active learning, these frames were identified as high-uncertainty frames and were used to initiate additional short AIMD simulations. These calculations were then used to supplement the training dataset, and the MD simulations were restarted from the beginning. The final datasets we used to train our MACE models are provided along with the SI[†].

All analyses in the following subsections consider only the last 200 ps of the trajectory, as at this point, the hydrogen dissociation-recombination reaction had reached a steady state for all systems.

Table 2 MACE root mean squared error for energy, E (meV/atom), and relative root mean squared error for forces, F (%) and MD efficiency reported in wall time/ps (s) for δ-MoC with and without the ghost atom grid on a node utilizing 1 NVIDIA H100 GPU

(ν, l_{max}, N_{feat})	Grid	E RMSE	F RRMSE	Walltime/ps
(2, 1, 128)	Yes	1.8	8.1	85.8
(2, 1, 128)	No	3.2	9.2	79.7
(3, 0, 128)	Yes	1.6	8.9	76.3
(3, 0, 128)	No	3.1	10.5	54.8
(3, 0, 256)	Yes	1.8	7.5	82.4
(3, 0, 256)	No	3.0	8.7	64.8
(3, 1, 128)	Yes	2.0	7.6	93.2
(3, 1, 128)	No	3.3	8.6	76.6

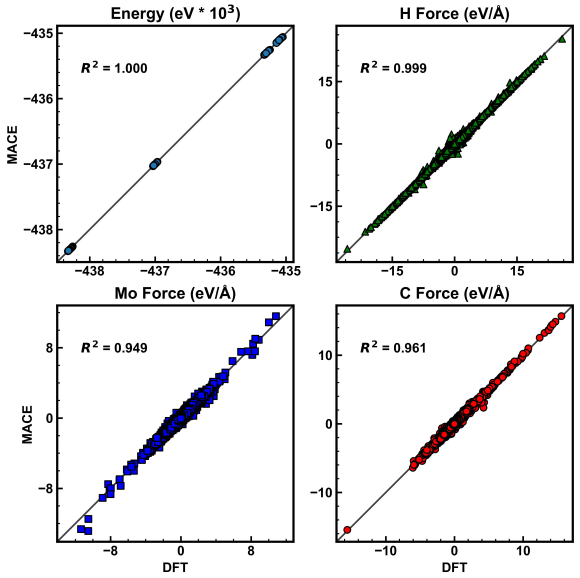

Fig. 3 The energy and force parity plots for a MACE model trained on test data with a ghost-atom grid for δ-MoC.

3 Results and discussion

3.1 MLIP efficiency, accuracy, and hyperparameter sensitivity

The RMSE for energy (E) and the RRMSE for the force (F) predictions for the δ-MoC system and MD efficiency reported as compute time (wall time) for 1 ps on a node with a single NVIDIA H100 GPU for a system containing 124 H_2 molecules over δ-MoC is shown in Table 2.

We also show how including the ghost atom grid (Grid=Yes/No) affects the root mean squared errors and MD efficiency. For each MACE model trained with and without the ghost atom grid, we obtain energy percent errors within the threshold of chemical accuracy (below 43 meV = 1 kcal/mol), and F RRMSE are comparable to relative mean absolute errors for force that Owen et al. reported for bulk transition metals with a vacancy defect[88] where forces were obtained using the NequIP framework[59,89], another equivariant message passing neural network architecture. With the ghost atom grid, energy predictions are improved between 30 - 55 % depending on which set of hyperparameters are used, whereas the RRMSE on forces improves by approximately 1 percent. We also observe that the

ghost atom grid increases the wall time/ps and worsens MD efficiency by 8-39 % depending on the hyperparameter set used. Despite the ghost atom grid reducing MD efficiency with only a small gain in force accuracy, we still feel it is necessary to include it for these systems because of the significant improvement to energy predictions and to help avoid situations for sparse hydrogen loadings wherein atoms just outside the 5 Å cutoff radius would be treated as isolated atoms in a vacuum. In this work, we chose to run molecular dynamics with $(N_{layers}, r_{cut}, \nu, l_{max}, N_{feat})$ equal to (2, 5 Å, 2, 1, 128) with the ghost atom grid as it balances both accurate force and energy predictions with reasonable MD efficiency.

Energy and force parity plots on the test data for δ-MoC are shown in Fig. 3, and the plots for α-Mo_2C, β-Mo_2C, and Mo are given in the SI† in Fig. S2, S3, and S4, respectively. The neural networks model, with the ghost atom grid approach, demonstrates remarkable energy and force parity when compared to DFT. Unsurprisingly, Mo is the most difficult atom type for force evaluation, given that it is a d-block transition metal, whereas H is the easiest. Regardless, we are still able to reliably predict forces on Mo with a R^2 value of 0.949 over a range of hydrogen loadings.

As a final test to validate our force fields, we trained two additional MACE models with different random number seeds. We performed active learning by evaluating all three MLIPs on each MD trajectory, predicting the energy. In active learning, an ensemble of neural networks is used to assess which frames in a dataset are unlike others on which the models have been trained by calculating variance in the energy prediction of each network. When the energy predictions between the models diverge and the variance is high, the frame has high uncertainty and needs to be included in the training set. Plots of the energy variance for 53 H_2 molecules over δ-MoC at each temperature studied are shown in Fig. 4, and plots for α-Mo_2C, β-Mo_2C, δ-MoC, and Mo at each hydrogen loading studied are provided in the SI† in Figs. S5-S19. Over the simulation trajectory, the variance for each system and process condition simulated stays below the threshold of chemical accuracy. We also note that the 53 H_2 over δ-MoC system was not included in the AIMD training data. It demonstrates the transferability of the MACE MLIPs we have trained and implies that we can use these force fields to study other hydrogen loadings we did not include in the training data.

3.2 Effect of surface type on reaction and diffusion energetics

The simulation setup allows us to directly observe (and count) the dissociation and association events occurring on the catalytic surface as a function of time to calculate reaction rates[56]. For the equilibrium reaction:

$$H_2 \cdot S + S \rightleftharpoons 2H \cdot S \tag{4}$$

where S denotes an active site, and the dot denotes hydrogen adsorbed to a particular site, the forward reaction is hydrogen dissociation, and the reverse reaction is recombination. Counting reactive events is facilitated by keeping track of the coordination

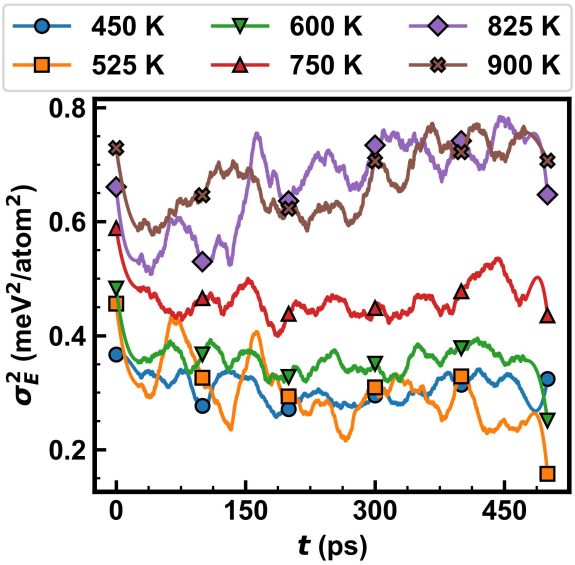

Fig. 4 The evolution of variance in energy prediction as a function of simulation time using post-process active learning on a MD trajectory of 53 H_2 molecules over δ-MoC.

number of hydrogen atoms at every time step. If the coordination number is equal to 1 for a given hydrogen atom, then that hydrogen atom is bound to one other hydrogen and is part of a hydrogen molecule. In contrast, a coordination number of 0 indicates that hydrogen is atomic. We used a distance cutoff criterion of 0.992 Å to determine if two hydrogen atoms were bonded to one another. This cutoff distance ensured that molecular hydrogen bond vibrations in the gas phase were not counted as reactive events. Between two MD frames, if the difference in coordination number for a given hydrogen atom is equal to 0, then no reaction occurred. Otherwise, if the difference equals +1 or -1, a recombination or dissociation reaction occurred, respectively. After counting the number of reactions that occur as a function of simulation time, the reaction rate can be determined from the slope of the line. From equation 4, the forward reaction rate law is:

$$-r_{H_2 \cdot S} = k_f C_{H_2 \cdot S} C_S \tag{5}$$

whereas for the reverse:

$$-r_{H \cdot S_H} = k_r C_{H \cdot S}^2 \tag{6}$$

where k_f and k_r are the rate constants, and C denotes the surface concentration of either an adsorbed species or a vacant adsorption site. Once the reaction rate and all surface concentrations are known for a range of temperatures, using the well-known Arrhenius relationship:

$$k = A \exp(-E_a / k_B T) \tag{7}$$

the apparent activation energy (E_a) and Arrhenius pre-exponential factor (A) for the reaction can be computed via linear regression. To estimate $C_{H \cdot S}$ and $C_{H_2 \cdot S}$, we average the number of adsorbed species during the production run of the simulations after equilibrium is reached. The number of vacant sites can be

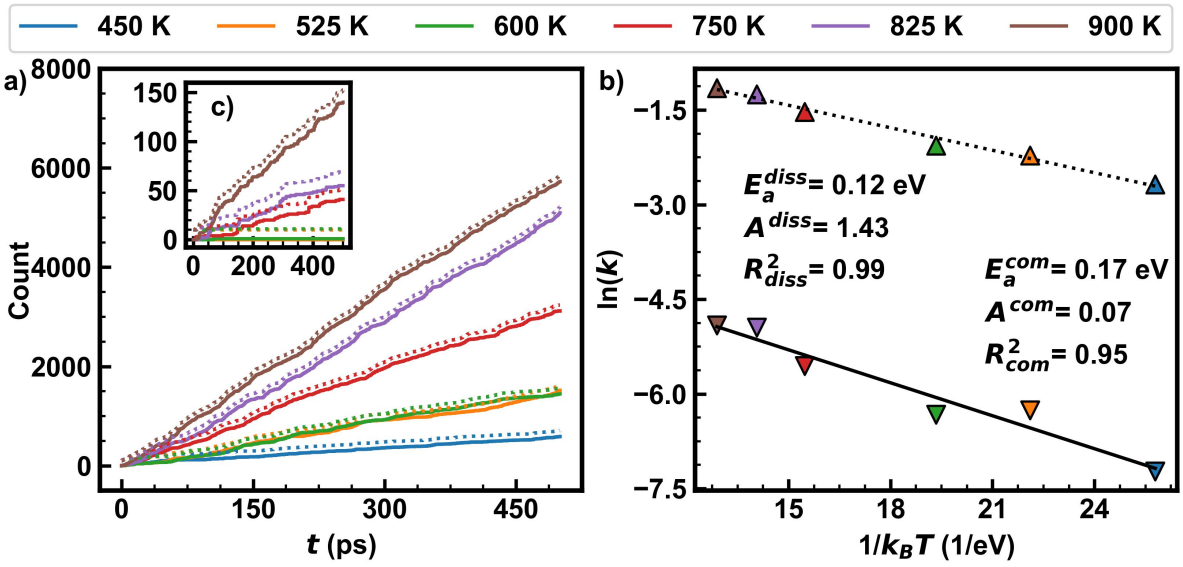

Fig. 5 The number of hydrogen recombination (solid lines and triangles pointing up) and dissociation (dotted lines and triangles pointing down) over the course of the MD simulation (a) and the corresponding Arrhenius analysis (b) for 53 H_2 molecules over δ-MoC Mo-(111). The inset (c) refers to the number of reactions that take place over the C-(111) surface. For readability, the number of dissociation reactions has been shifted by 100 and 10 for the Mo-(111) (a) and C-(111) (c) facets, respectively.

calculated from the site balance equation:

$$C_S = C_{Total} - C_{H_2 \cdot S} - C_{H \cdot S} \tag{8}$$

where C_{Total} is the total number of active sites. The total number of active sites depends on the catalyst surface. For our calculations, we count per periodic surface the number of atop Mo and 3-fold Mo sites for α-Mo_2C (101), atop Mo and short bridge sites for β-Mo_2C (001), atop Mo and 3-fold (Mo) sites for δ-MoC Mo-(111), atop C and 3-fold (C) for δ-MoC C-(111), and atop Mo and 3-fold (Mo) sites for Mo (110) to estimate the value of C_{Total}. The reason for choosing these sites as opposed to the others shown in Figure 1 is discussed in more detail in Section 3.3.

A similar procedure can also be used to calculate diffusion barriers. The diffusion coefficient of hydrogen can be calculated by first calculating the mean squared displacement *MSD*:

$$MSD(\tau) = \langle |\mathbf{r}(\tau) - \mathbf{r}(0)|^2 \rangle. \tag{9}$$

where $\mathbf{r}(\tau)$ is the coordinates of the hydrogen atom at a correlation depth τ and $\mathbf{r}(0)$ is the initial coordinates of the hydrogen atom. Then, the diffusion coefficient can be calculated via the Einstien relationship:

$$D = \frac{1}{2d} \lim_{\tau \to \infty} \frac{d}{d\tau} MSD(\tau) \tag{10}$$

where d is the dimensionality of the system (here $d = 3$ as the hydrogen atom is allowed to move in the x, y, and z directions). By calculating the diffusion coefficient at different temperatures, we can then apply an Arrhenius relationship to calculate the activation energy barrier of diffusion:

$$\ln D = \ln D_0 - E_a / k_B T \tag{11}$$

Diffusion calculations were done at the low coverage limit, where only one hydrogen atom was adsorbed onto the surface simulated. We acknowledge that this approach neglects the environmental effects of molecular hydrogen in the vapor phase and coverage effects of other hydrogen atoms and molecules adsorbed on the surface, but this approach isolates diffusion rare events from reactions and enables studying hydrogen transport across the surface without the possibility of it reacting with another hydrogen atom and moving into the vapor phase.

A plot of the cumulative number of dissociation and recombination reactions over the course of the simulation for δ-MoC, Mo-(111) and C-(111) with 53 H_2 molecules and the corresponding Arrhenius analyses are given in Fig. 5. The *MSD* and their corresponding Arrhenius analysis for other systems considered in this work are presented in Figs. S20-S24. For δ-MoC C-(111), we observe frequent dissociation and recombination reactions at higher temperatures, whereas the reaction rate at the lower temperatures is negligible. Furthermore, at higher temperatures, it is also clear, based on the number of reactions that occur, that the Mo-(111) facet is much more active than the C-(111) facet for δ-MoC. We observe this same behavior overall hydrogen loadings studied for δ-MoC Mo-(111) and C-(111).

In Fig. 6, we summarize the activation energy barriers for diffusion and the dissociation and recombination reaction reaction for all surfaces and loadings studied. Numerical values for the reaction rates as well as the activation energy barriers and pre-exponential factors are given in the SI in Tables S3-S7. Unfortunately, we do not have enough reactive events to estimate the activation energy barrier reliably for some simulations with smaller hydrogen loadings at lower temperatures. In these scenarios, the surface diffusion of atomic hydrogen becomes the rate-controlling step. The δ-MoC C-(111) system presents the most extreme case,

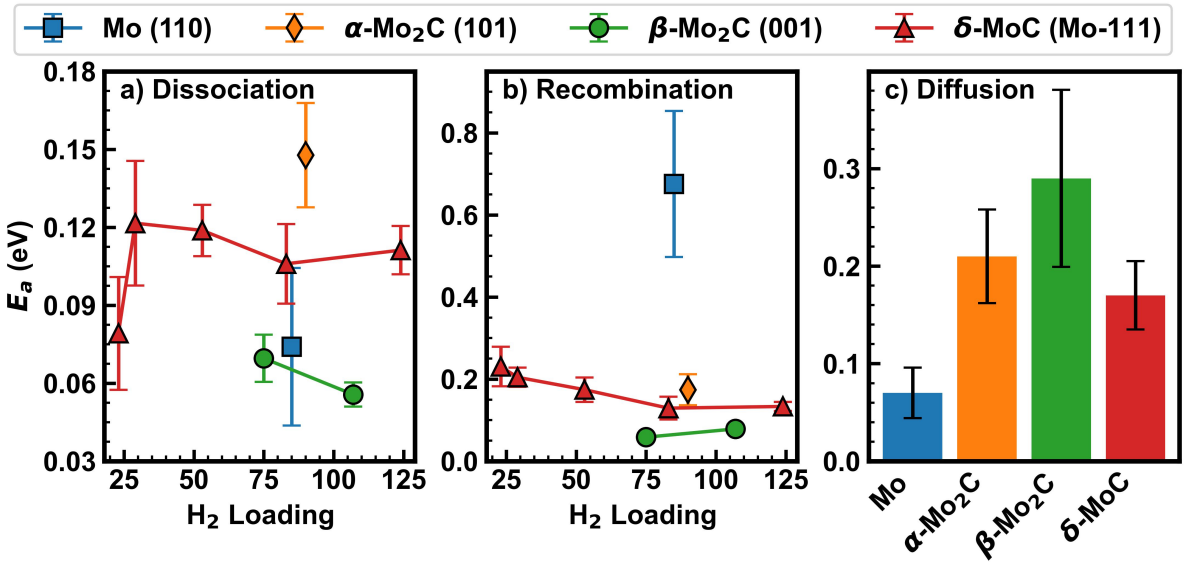

Fig. 6 The activation energy barrier for hydrogen dissociation (a) and recombination (b) is shown for cases with sufficient reactive events in the 500 ps trajectory to evaluate statistical averages and the activation energy barrier for diffusion (c) for each system studied. Error bars show 1 standard deviation.

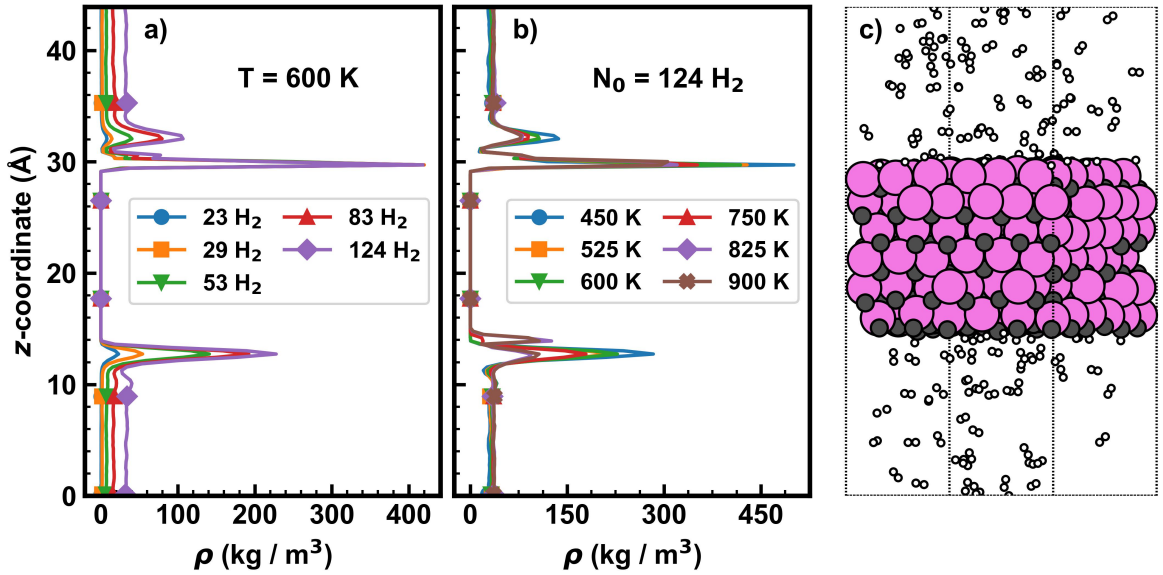

Fig. 7 Density profile of hydrogen over δ-MoC as a function of H_2 loading at 600 K (a), as a function of temperature with 124 H_2 molecules (b), and a simulation snapshot of the last frame of 124 H_2 over δ-MoC at 900 K (c).

where regardless of H_2 loading, the activation energy barrier cannot be reliably estimated. Also, the diffusion barrier of δ-MoC C-(111) cannot be computed since the *MSD*s have a slope of 0 (see Fig. S23 in the SI[†]) as the interaction between under coordinated carbon atoms on the surface is particularly strong leading to non-diffusive behavior of the hydrogen on this facet.

As hydrogen loading is increased for δ-MoC, the activation energy barrier for recombination decreases, whereas for dissociation, the barrier is largely unaffected (except at the lowest two loadings). However, it can be argued that this decrease for δ-MoC we observe due to an increase in the number of hydrogen molecules is not significant as these data points lie within a 95% confidence interval (1.96 times the standard deviation). Also, including carbon atoms in the molybdenum crystal lattice dramatically reduces the activation energy barrier of recombination from 0.68 eV to 0.05-0.23 eV depending on the H_2 loading and phase of molybdenum carbide. Interestingly, the diffusion and recombination barriers appear to have an inverse relationship. Mo has the smallest diffusion barrier but the largest combination reaction energy barrier, whereas β-Mo_2C has the smallest reaction energy barrier but the largest diffusion barrier. This is further elaborated in the next section.

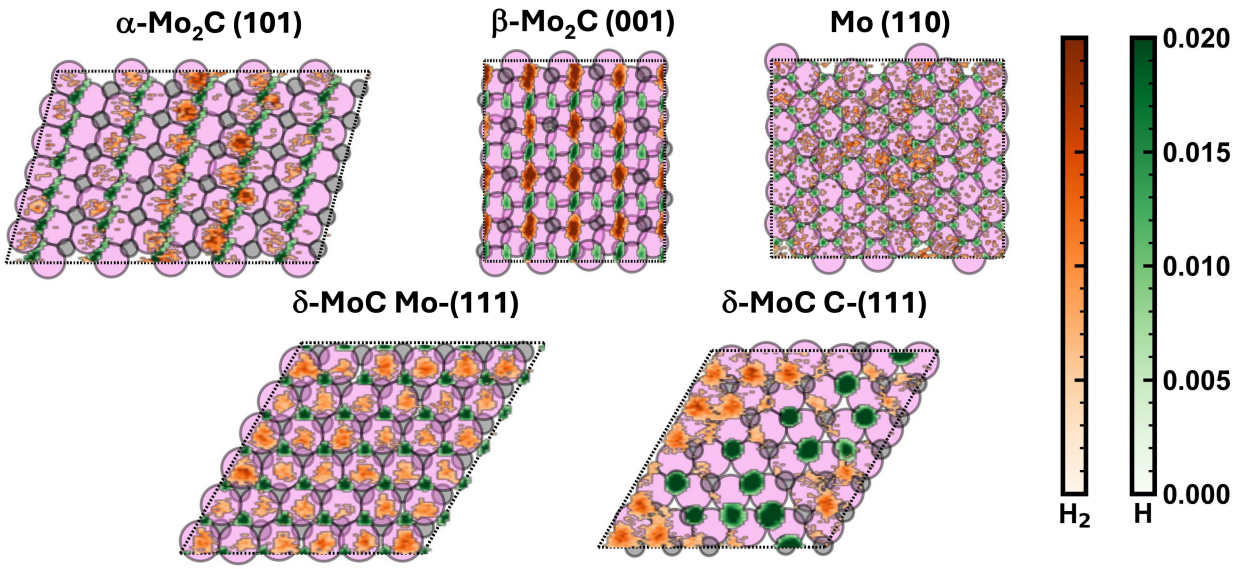

Fig. 8 The adsorption probabilities for molecular hydrogen (depicted in orange) and atomic hydrogen (depicted in green) mapped onto the δ-MoC Mo-(111) surfaces with 124 hydrogen molecules.

3.3 Preferential adsorption sites and diffusion paths on different surfaces

Carbon atoms in Mo_xC_y catalysts lead to heterogeneous surface sites with a different affinity towards molecular or atomic hydrogen. In this Section, we examine the nature of surface coverage (monolayer vs. multilayer adsorption), preferential adsorption sites, and surface diffusion characteristics for atomic hydrogen. Density profiles as a function of the z-coordinate normal to the catalyst surface reveal that hydrogen adsorption increases as the H_2 loading increases and a second adsorption layer is formed. Fig. 7 shows the density profile of hydrogen over δ-MoC as a function of loading and temperature. Figs. S25-S27 show the same data for the other catalyst surfaces studied here. A common behavior we see among all catalyst surfaces is the formation of a 2nd monolayer above the catalyst surface. As expected, the density of this layer increases with increasing hydrogen loading. However, we do not see an increase in the density in the 1st layer at a given temperature, indicating that we have 100 % coverage on the catalyst surface. With increased temperatures, we see a sharp decrease in the density of the 1st adsorbed layer. We also see a decrease in the height of the peak corresponding to the 2nd layer, albeit not as significant as the decrease in the first adsorbed layer. It is expected that the temperature increase will lead to increased diffusivity and activity, leading to a decrease in the structure in the first adsorbed layer. Interestingly, except for the C-(111) facet of δ-MoC, all catalyst surfaces studied demonstrated this behavior. The C-(111) facet of δ-MoC shows no pronounced 2nd layer. At the highest temperature, the peak splits, and another peak closer to the interface appears. It results from molecular hydrogen from the first monolayer dissociating on the surface of the surface-exposed carbon atoms. This implies that dissociation and adsorption of atomic hydrogen on this facet only occur at higher temperatures and high H_2 pressures.

From the hydrogen-hydrogen coordination numbers used to calculate the hydrogen dissociation-recombination reaction, we can also compute where atomic and molecular hydrogen prefers to adsorb on the catalyst surface by building a histogram of the x and y coordinates for atoms adsorbed on the catalyst surface. A 2 Å cutoff was used from the average z-coordinate of surface Mo atoms to define whether or not hydrogen was adsorbed on the surface. The choice of 2 Å was made by observing the width of the density profiles in Figs. 7 and S26-28.

Fig. 8 shows atomic and molecular hydrogen adsorption heat maps for the highest hydrogen loading studied for each catalyst at 900 K. Heatmaps at the highest hydrogen loading studied for the catalysts as a function of temperature are provided in the SI[†] in Figs. S28-S32. When available, atomic hydrogen prefers to bind to 3-fold Mo sites regardless of temperature and the catalyst surface. An exception is α-Mo_2C (101), where hydrogen prefers the 3-fold hollow site. However, even on this surface, we observe hydrogen at the 3-fold Mo site at higher temperatures. In the case of β-Mo_2C (001), where there are no 3-fold sites, atomic hydrogen prefers to adsorb on the bridge sites. For the δ-MoC C-(111) facet, when atomic hydrogen is present on the surface, it binds exclusively atop C atoms, while molecular hydrogen is observed at the 3-fold site. Our simulation results agree with prior computational work based on single-point DFT calculations. For example, Matanovic et al. report that for δ-MoC (111), the threefold Mo site is the most stable for atomic H adsorption[32]. Furthermore, similar to our δ-MoC C-(111) facet, uncoordinated C atoms on the surface are the preferred site for the H atom adsorption.[37,73]

For the molecular hydrogen, it is most likely to be found atop molybdenum atoms for Mo (110), α-Mo_2C (101), and δ-MoC Mo-(111). We do not observe any molecular hydrogen atop Mo for β-Mo_2C (001) and δ-MoC C-(111). A unique feature of β-Mo_2C (001) compared to every other surface studied here is that there are several pits on the surface between surface carbon atoms and rows of Mo atoms connected by bridge sites, and it's at these pits

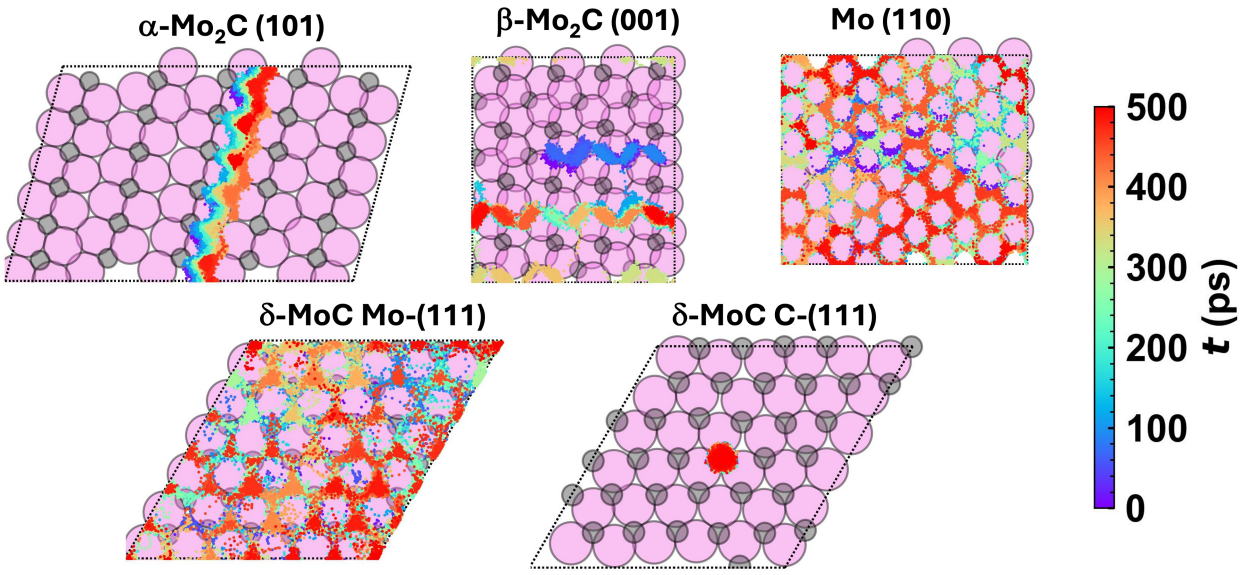

Fig. 9 The trajectories of surface adsorbed hydrogen at the limit of low coverage for each surface studied at 900 K. Blue depicts initial positions, while red depicts final positions of adsorbed hydrogen.

we see molecular hydrogen instead of atop Mo. It indicates that this is the only site where the dissociation-recombination reaction occurs on β-Mo_2C (001).

To help explain why we observe these interesting phenomena for adsorption and the dissociation-recombination reactions, we visualize the trajectory of the hydrogen atom from the diffusion simulations on each surface at 900 K in Fig. 9. In addition to this figure, we also provide animations of these graphs along with the SI[†].

For δ-MoC Mo-(111) and Mo (110), the hydrogen atom is observed to freely move across the surface. It also agrees with the Arrhenius analysis from Fig. 6, where these surfaces had the lowest activation energy. Interestingly, for δ-MoC C-(111), we do not see the hydrogen atom diffuse across the surface at the temperatures studied. Instead, we see it adsorb to and simply vibrate on the carbon atom on which it was initially placed. Hydrogen also has an interesting diffusion trajectory on α-Mo_2C (101) and β-Mo_2C (001) moving along active sites connected by Mo atoms and does not diffuse across surface-exposed carbon atoms. From these simulations and considering that Mo (110) has an order of magnitude lower activation energy barrier of diffusion (0.07 eV) compared to the molybdenum carbides studied, we conclude that surface carbon atoms inhibit diffusion with molybdenum carbides but can significantly lower the energy barrier for the dissociation-recombination reaction.

Conclusions

In this work, we have trained an equivariant message-passing neural network model to learn interatomic potentials with the MACE architecture for the hydrogen dissociation-recombination reaction over pure molybdenum and molybdenum carbides. To facilitate message passing for less dense systems, we utilize a ghost-atom grid approach, which improved the accuracy of energy and force predictions for the machine-learning interatomic potential. The kinetics and energetics of the hydrogen dissociation-recombination reaction, the diffusion of atomic hydrogen, and the adsorption of atomic and molecular hydrogen were studied with reactive molecular dynamics at time scales statistically relevant to catalysis. We found that atomic hydrogen prefers to diffuse along and adsorb at sites coordinated by multiple Mo atoms. Molecular hydrogen, on the other hand, prefers to adsorb atop Mo atoms or in the pit sites of β-Mo_2C (001). We also observe that surface carbon atoms impede diffusion and raise the diffusion barrier for hydrogen. In the case of δ-MoC on the C-(111) facet where carbon atoms protrude from the surface, the reaction only occurs at high temperatures, and hydrogen does not diffuse from one site to the next. Furthermore, carbon atoms limit diffusion pathways and force the reaction to occur on select sites of the molybdenum carbide surfaces. In contrast, for pure molybdenum, hydrogen can diffuse easily across the surface. Among the facets considered in our work, at a comparable hydrogen loading, β-Mo_2C (001) is the most active surface for hydrogen dissociation reaction. We also observe that surfaces with carbon termination are not an active site for this reaction, and their exposure could have a deleterious effect on the chemical kinetics. This work provides unique insights into the hydrogen reaction dynamics on the surface for several phases of molybdenum carbide compared to pure molybdenum.

Author contributions

Woodrow N. Wilson: Conceptualization (equal); Data curation (equal); Formal analysis (lead); Investigation (lead); Methodology (lead); Software (lead); Validation (lead); Visualization (lead); Writing – original draft (lead); Writing – review & editing (equal). **John Michael Lane**: Formal analysis (supporting); Investigation (supporting); Visualization (supporting); Writing – review & editing (supporting). **Chinmoy Saha**: Formal analysis (supporting); Investigation (supporting); Visualization (support-

ing); Writing – review & editing (supporting). **Sony Severin**: Formal analysis (supporting); Investigation (supporting); Visualization (supporting); Writing – review & editing (supporting). **Vivek S. Bharadwaj**: Formal analysis (supporting); Investigation (supporting); Visualization (supporting); Resources (equal); Writing – review & editing (equal) **Neeraj Rai**: Conceptualization (equal); Funding acquisition (lead); Project administration (lead); Resources (equal); Project Supervision (lead); Writing – review & editing (equal).

Conflicts of interest

There are no conflicts to declare.

Data availability

Data for this article, including starting coordinates and example input files are available at GitHub at https://github.com/MSELRai/H2-MoC-MLIP. The AIMD datasets used to train the MACE models are available at 10.5281/zenodo.14206465. Additional data supporting this article have been included as part of the Supplementary Information.

Acknowledgements

This work was supported by the National Science Foundation through the CAREER award program: CBET-1752036. This work was authored in part by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. VSB would like to acknowledge support from NREL's Laboratory Directed Research and Development (LDRD) Program. WNW would like to acknowledge support from the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists, Office of Science Graduate Student Research (SCGSR) program that enabled the collaboration between Mississippi State University and National Renewable Energy Laboratory. The SCGSR program is administered by the Oak Ridge Institute for Science and Education for the DOE under contract number DE-SC0014664. The views expressed in the article do not necessarily represent the views of the DOE or the U.S. Government. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes. This research used computational resources provided by the High Performance Computing Collaboratory at Mississippi State University, NREL's Computational Sciences' resources (Eagle and Kestrel) supported by the DOE Office of EERE under contract no. DE-AC36-08GO28308, and resources of the National Energy Research Scientific Computing Center (NERSC), a Department of Energy Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231 using NERSC award BES-ERCAP0029138. The Institute of Computational Molecular Science Education (i-CoMSE) provided key insights and provided foundational concepts to work at the 5th i-CoMSE Workshop: Machine Learning for Molecular Science hosted at the University of Minnesota. We would like to acknowledge the VMD software for rendering simulation snapshots[90]. WNW would like to thank Evan Komp (National Renewable Energy Laboratory) for helpful discussions regarding message passing neural networks and suggesting the use of ghost atoms to facilitate message passing.

Notes and references

1 P. Mortensen, J.-D. Grunwaldt, P. Jensen, K. Knudsen and A. Jensen, *Applied Catalysis A: General*, 2011, **407**, 1–19.
2 H. Wang, J. Male and Y. Wang, *ACS Catalysis*, 2013, **3**, 1047–1070.
3 K. Yang, K. Wu and H. Zhang, *Energy*, 2022, **254**, 124320.
4 J. J. Sims, C. A. Ould Hamou, R. Réocreux, C. Michel and J. B. Giorgi, *The Journal of Physical Chemistry C*, 2018, **122**, 20279–20288.
5 S. Sitthisa and D. E. Resasco, *Catalysis Letters*, 2011, **141**, 784–791.
6 R. Ormerod, C. J. Baddeley, C. Hardacre and R. M. Lambert, *Surface Science*, 1996, **360**, 1–9.
7 J. Justicia, J. A. Baeza, L. Calvo, F. Heras and M. A. Gilarranz, *Chemical Engineering Journal*, 2023, **477**, 146860.
8 Y. Yoon, R. Rousseau, R. S. Weber, D. Mei and J. A. Lercher, *Journal of the American Chemical Society*, 2014, **136**, 10287–10298.
9 I. T. McCrum and M. T. M. Koper, *Nature Energy*, 2020, **5**, 891–899.
10 B. K. Devendra, B. M. Praveen, V. S. Tripathi, D. H. Nagaraju and K. O. Nayana, *Iranian Journal of Science and Technology, Transactions A: Science*, 2021, **45**, 1993–2000.
11 N. Cheng, S. Stambula, D. Wang, M. N. Banis, J. Liu, A. Riese, B. Xiao, R. Li, T.-K. Sham, L.-M. Liu, G. A. Botton and X. Sun, *Nature Communications*, 2016, **7**, 13638.
12 J. Pang, J. Sun, M. Zheng, H. Li, Y. Wang and T. Zhang, *Applied Catalysis B: Environmental*, 2019, **254**, 510–522.
13 C. Chan-Thaw and A. Villa, *Applied Sciences*, 2018, **8**, 259.
14 H. Prats, J. J. Piñero, F. Viñes, S. T. Bromley, R. Sayós and F. Illas, *Chemical Communications*, 2019, **55**, 12797–12800.
15 R. B. Levy and M. Boudart, *Science*, 1973, **181**, 547–549.
16 K. Xiong, W. Lee, A. Bhan and J. G. Chen, *ChemSusChem*, 2014, **7**, 2146–2149.
17 A. A. Smirnov, I. N. Shilov, A. A. Saraev, O. A. Bulavchenko, S. G. Zavarukhin and V. A. Yakovlev, *WSEAS Transactions on Environmnet and Development*, 2019, **15**, 2224–3496.
18 Q. Yang, K. Sun, Y. Xu, Z. Ding and R. Hou, *Applied Catalysis A: General*, 2022, **630**, 118455.
19 Y. Shi, Y. Yang, Y.-W. Li and H. Jiao, *ACS Catalysis*, 2016, **6**, 6790–6803.
20 B. Yan, X. Lin, Z. Chen, Q. Cai and S. Zhang, *Bioresource Technology*, 2021, **321**, 124503.
21 J. Engelhardt, P. Lyu, P. Nachtigall, F. Schüth and Á. M. Garcá, *ChemCatChem*, 2017, **9**, 1985–1991.
22 B. Martínez, F. Viñes, P. H. McBreen and F. Illas, *Journal of Catalysis*, 2021, **393**, 381–389.
23 C. Fortin, A. Adnot and P. H. McBreen, *The Canadian Journal*

of Chemical Engineering, 2018, **96**, 2138–2143.
24 N. Dubuc and P. H. McBreen, *Topics in Catalysis*, 2015, **58**, 232–239.
25 J. A. Rodriguez, C. Jimenez-Orozco, E. Flórez, F. Viñes and F. Illas, *The Journal of Physical Chemistry C*, 2023, **127**, 16764–16780.
26 S. Posada-Pérez, P. J. Ramírez, J. Evans, F. Viñes, P. Liu, F. Illas and J. A. Rodriguez, *Journal of the American Chemical Society*, 2016, **138**, 8269–8278.
27 M. Figueras, R. A. Gutiérrez, F. Viñes, P. J. Ramírez, J. A. Rodriguez and F. Illas, *ACS Catalysis*, 2021, **11**, 9679–9687.
28 C. Jimenez-Orozco, M. Figueras, E. Flórez, F. Viñes, J. A. Rodriguez and F. Illas, *Physical Chemistry Chemical Physics*, 2022, **24**, 16556–16565.
29 G.-Q. Yu, B.-Y. Huang, X. Chen, D. Wang, F. Zheng and X.-B. Li, *The Journal of Physical Chemistry C*, 2019, **123**, 21878–21887.
30 J. Jia, T. Xiong, L. Zhao, F. Wang, H. Liu, R. Hu, J. Zhou, W. Zhou and S. Chen, *ACS Nano*, 2017, **11**, 12509–12518.
31 X. Huang, J. Wang, H. Bing Tao, H. Tian, Z. Zhang and H. Xu, *Journal of Catalysis*, 2020, **389**, 461–467.
32 I. Matanovic and F. H. Garzon, *Physical Chemistry Chemical Physics*, 2018, **20**, 14679–14687.
33 L. Sun, J. Xu, X. Liu, B. Qiao, L. Li, Y. Ren, Q. Wan, J. Lin, S. Lin, X. Wang, H. Guo and T. Zhang, *ACS Catalysis*, 2021, **11**, 5942–5950.
34 H. Tominaga and M. Nagai, *The Journal of Physical Chemistry B*, 2005, **109**, 20415–20423.
35 X. Zhang, M. E. Reece, C. B. Cockreham, H. Sun, B. Wang, H. Xu, J. Sun, X. Guo, H. Su, Y. Wang and D. Wu, *Industrial & Engineering Chemistry Research*, 2021, **60**, 13991–14003.
36 J. Wang, M. Castonguay, J. Deng and P. H. McBreen, *Surface Science*, 1997.
37 Q. Luo, T. Wang, G. Walther, M. Beller and H. Jiao, *Journal of Power Sources*, 2014, **246**, 548–555.
38 M. Nagai, H. Tominaga and S. Omi, *Langmuir*, 2000, **16**, 10215–10220.
39 L. Wang, B. Wang, M. Fan, L. Ling and R. Zhang, *Fuel*, 2023, **336**, 127131.
40 S. Yao, B. Yan, Z. Jiang, Z. Liu, Q. Wu, J. H. Lee and J. G. Chen, *ACS Catalysis*, 2018, **8**, 5374–5381.
41 B. Fang, M. Yang, C. Zhang, J. Li, C. Li, J. Ni, X. Wang, J. Lin, B. Lin and L. Jiang, *Chemical Engineering Science*, 2022, **259**, 117834.
42 A. Vojvodic, *Catalysis Letters*, 2012, **142**, 728–735.
43 V. Van Speybroeck, M. Bocus, P. Cnudde and L. Vanduyfhuys, *ACS Catalysis*, 2023, **13**, 11455–11493.
44 F. Haase, J. Sauer and J. Hutter, *Chemical Physics Letters*, 1997, **266**, 397–402.
45 J. Randrianandraina, M. Badawi, B. Cardey, M. Grivet, J.-E. Groetz, C. Ramseyer, F. T. Anzola, C. Chambelland and D. Ducret, *Physical Chemistry Chemical Physics*, 2021, **23**, 19032–19042.
46 V. Jain, W. N. Wilson and N. Rai, *The Journal of Chemical Physics*, 2019, **151**, 114708.
47 W. N. Wilson, J. Whittington and N. Rai, *The Journal of Chemical Physics*, 2024, **160**, 224703.
48 N. Humphrey, S. Bac and S. Mallikarjun Sharada, *The Journal of Physical Chemistry C*, 2020, **124**, 24187–24195.
49 V. Sinha, N. Govindarajan, B. De Bruin and E. J. Meijer, *ACS Catalysis*, 2018, **8**, 6908–6913.
50 F. H. Hodel and S. Luber, *ACS Catalysis*, 2016, **6**, 1505–1517.
51 N. Govindarajan, H. Beks and E. J. Meijer, *ACS Catalysis*, 2020, **10**, 14775–14781.
52 R. A. Garcia Carcamo, X. Zhang, A. Estejab, J. Zhou, B. J. Hare, C. Sievers, S. Sarupria and R. B. Getman, *iScience*, 2023, **26**, 105980.
53 C. J. Bodenschatz, T. Xie, X. Zhang and R. B. Getman, *Physical Chemistry Chemical Physics*, 2019, **21**, 9895–9904.
54 A. Estejab, R. A. García Cárcamo and R. B. Getman, *Physical Chemistry Chemical Physics*, 2022, **24**, 4251–4261.
55 X. Zhang, R. S. DeFever, S. Sarupria and R. B. Getman, *Journal of Chemical Information and Modeling*, 2019, **59**, 2190–2198.
56 J. Vandermause, Y. Xie, J. S. Lim, C. J. Owen and B. Kozinsky, *Nature Communications*, 2022, **13**, 5183.
57 C. Zhou, H. T. Ngan, J. S. Lim, Z. Darbari, A. Lewandowski, D. J. Stacchiola, B. Kozinsky, P. Sautet and J. A. Boscoboinik, *Journal of the American Chemical Society*, 2022, **144**, 15132–15142.
58 A. Johansson, Y. Xie, C. J. Owen, J. S. Lim, L. Sun, J. Vandermause and B. Kozinsky, *Micron-scale heterogeneous catalysis with Bayesian force fields from first principles and active learning*, 2022, `http://arxiv.org/abs/2204.12573`, arXiv:2204.12573 [physics].
59 S. Batzner, A. Musaelian, L. Sun, M. Geiger, J. P. Mailoa, M. Kornbluth, N. Molinari, T. E. Smidt and B. Kozinsky, *Nature Communications*, 2022, **13**, 2453.
60 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, F. Della 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, J. Jaafar, S. Hofmann, 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, E. Varga-Umbrich, T. Vegge, M. Vondrák, Y. Wang, W. C. Witt, F. Zills and G. Csányi, *A foundation model for atomistic materials chemistry*, 2024, `http://arxiv.org/abs/2401.00096`, arXiv:2401.00096 [cond-mat, physics:physics].
61 W. G. Stark, C. van der Oord, I. Batatia, Y. Zhang, B. Jiang, G. Csányi and R. J. Maurer, *Benchmarking of machine learn-*

ing interatomic potentials for reactive hydrogen dynamics at metal surfaces, 2024, http://arxiv.org/abs/2403.15334, arXiv:2403.15334 [physics].

62 S. Vandenhaute, M. Cools-Ceuppens, S. DeKeyser, T. Verstraelen and V. Van Speybroeck, *npj Computational Materials*, 2023, **9**, 19.

63 J. Xu, X.-M. Cao and P. Hu, *Journal of Chemical Theory and Computation*, 2021, **17**, 4465–4476.

64 J. S. Jestilä, N. Wu, F. Priante and A. S. Foster, *Journal of Chemical Theory and Computation*, 2024, acs.jctc.3c01292.

65 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, http://arxiv.org/abs/2205.06643, arXiv:2205.06643 [cond-mat, physics:physics, stat].

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*, 2023, http://arxiv.org/abs/2206.07697, arXiv:2206.07697 [cond-mat, physics:physics, stat].

67 A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder and K. A. Persson, *APL Materials*, 2013, **1**, 011002.

68 A. N. Christensen, *Acta Chemica Scandinavica*, 1977, **31**, 509–511.

69 *Powder Diffraction File; JCPDS International Center for Diffraction Data: Pennsylvania*, 2004.

70 J. Wang, F. Coppari, R. F. Smith, J. H. Eggert, A. E. Lazicki, D. E. Fratanduono, J. R. Rygg, T. R. Boehly, G. W. Collins and T. S. Duffy, *Physical Review B*, 2015, **92**, 174114.

71 J. S. Lee, K. H. Lee and J. Y. Lee, *The Journal of Physical Chemistry*, 1992, **96**, 362–366.

72 T. Wang, X. Liu, S. Wang, C. Huo, Y.-W. Li, J. Wang and H. Jiao, *The Journal of Physical Chemistry C*, 2011, **115**, 22360–22368.

73 T. Wang, Y.-W. Li, J. Wang, M. Beller and H. Jiao, *The Journal of Physical Chemistry C*, 2014, **118**, 8079–8089.

74 H. Li and K. Reuter, *ACS Catalysis*, 2022, **12**, 10506–10513.

75 T. D. Kühne, M. Iannuzzi, M. Del Ben, V. V. Rybkin, P. Seewald, F. Stein, T. Laino, R. Z. Khaliullin, O. Schütt, F. Schiffmann, D. Golze, J. Wilhelm, S. Chulkov, M. H. Bani-Hashemian, V. Weber, U. Borštnik, M. Taillefumier, A. S. Jakobovits, A. Lazzaro, H. Pabst, T. Müller, R. Schade, M. Guidon, S. Andermatt, N. Holmberg, G. K. Schenter, A. Hehn, A. Bussy, F. Belleflamme, G. Tabacchi, A. Glöß, M. Lass, I. Bethune, C. J. Mundy, C. Plessl, M. Watkins, J. VandeVondele, M. Krack and J. Hutter, *The Journal of Chemical Physics*, 2020, **152**, 194103.

76 J. Hutter, M. Iannuzzi, F. Schiffmann and J. VandeVondele, *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2014, **4**, 15–25.

77 J. VandeVondele, M. Krack, F. Mohamed, M. Parrinello, T. Chassaing and J. Hutter, *Computer Physics Communications*, 2005, **167**, 103–128.

78 J. VandeVondele and J. Hutter, *The Journal of Chemical Physics*, 2007, **127**, 114105.

79 S. Goedecker, M. Teter and J. Hutter, *Physical Review B*, 1996, **54**, 1703–1710.

80 J. VandeVondele and J. Hutter, *The Journal of Chemical Physics*, 2003, **118**, 4365–4369.

81 J. Kolafa, *Journal of Computational Chemistry*, 2004, **25**, 335–342.

82 G. J. Martyna, M. L. Klein and M. Tuckerman, *The Journal of Chemical Physics*, 1992, **97**, 2635–2643.

83 R. Drautz, *Physical Review B*, 2019, **99**, 014104.

84 D. P. Kovács, I. Batatia, E. S. Arany and G. Csányi, *The Journal of Chemical Physics*, 2023, **159**, 044118.

85 A. Hjorth Larsen, J. Jørgen Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dułak, J. Friis, M. N. Groves, B. Hammer, C. Hargus, E. D. Hermes, P. C. Jennings, P. Bjerre Jensen, J. Kermode, J. R. Kitchin, E. Leonhard Kolsbjerg, J. Kubal, K. Kaasbjerg, S. Lysgaard, J. Bergmann Maronsson, T. Maxson, T. Olsen, L. Pastewka, A. Peterson, C. Rostgaard, J. Schiøtz, O. Schütt, M. Strange, K. S. Thygesen, T. Vegge, L. Vilhelmsen, M. Walter, Z. Zeng and K. W. Jacobsen, *Journal of Physics: Condensed Matter*, 2017, **29**, 273002.

86 W. G. Hoover, A. J. C. Ladd and B. Moran, *Physical Review Letters*, 1982, **48**, 1818–1820.

87 D. J. Evans, *The Journal of Chemical Physics*, 1983, **78**, 3297–3302.

88 C. J. Owen, S. B. Torrisi, Y. Xie, S. Batzner, K. Bystrom, J. Coulter, A. Musaelian, L. Sun and B. Kozinsky, *npj Computational Materials*, 2024, **10**, 92.

89 M. Geiger, T. Smidt, A. M, B. K. Miller, W. Boomsma, B. Dice, K. Lapchevskyi, M. Weiler, M. Tyszkiewicz, S. Batzner, D. Madisetti, M. Uhrin, J. Frellsen, N. Jung, S. Sanborn, M. Wen, J. Rackers, M. Rød and M. Bailey, *Euclidean neural networks: e3nn*, 2022, https://doi.org/10.5281/zenodo.6459381.

90 W. Humphrey, A. Dalke and K. Schulten, *Journal of Molecular Graphics*, 1996, **14**, 33–38.