Document Version
Final published version

Licence
CC BY

Citation (APA)
Wei, Z., Shuang, F., & Dey, P. (2026). From O adsorption to Fe oxide growth: Benchmarking reactive force fields and universal machine learning interatomic potentials against DFT for BCC Fe surface oxidation. *Surface and Coatings Technology*, *521*, Article 133092. https://doi.org/10.1016/j.surfcoat.2025.133092

Surface & Coatings Technology 521 (2026) 133092

Contents lists available at ScienceDirect

Surface & Coatings Technology

journal homepage: www.elsevier.com/locate/sct

From O adsorption to Fe oxide growth: Benchmarking reactive force fields and universal machine learning interatomic potentials against DFT for BCC Fe surface oxidation[☆]

Zixiong Wei, Fei Shuang, Poulumi Dey [*]

Department of Materials Science and Engineering, Faculty of Mechanical Engineering, Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands

ARTICLE INFO

Keywords:
Iron oxidation
Oxygen adsorption
Density functional theory
Reactive force field
Universal machine learning interatomic potential

ABSTRACT

Iron oxidation is a complex process involving critical atomistic events, such as atomic adsorption, diffusion, and surface reconstruction, understanding of which is significant for both surface science and coating technology. Atomistic simulation serves as an useful tool to investigate the processes, where description of interatomic interactions is required. However, selecting appropriate force field or interatomic potential is not only difficult, but also essential for getting accurate result. In this work, we present a detailed benchmark of reactive force fields (ReaxFFs) and universal machine learning interatomic potentials (uMLIPs) against density functional theory (DFT) calculations of oxygen adsorption on various α-iron surfaces, which is the first yet crucial step towards oxidation. The comparisons show the coverage-dependent performance and improvable accuracy of both ReaxFFs and uMLIPs at reproducing DFT results, with ReaxFFs outperforming uMLIPs. Subsequently, iron oxidation is simulated using ReaxFF and uMLIP. The results reveal the strong capability of ReaxFF and poor stability of uMLIP for describing reactive process, i.e., the formation of iron oxide. This may be attributed to the suitable functional form of ReaxFF for the description of bond changes. The insights presented here not only provide an example of benchmarking force field or interatomic potential for system of interest, but also highlight the applicability of ReaxFF and scopes of improvement of uMLIP.

1. Introduction

With the rapid development of automobile industry, the demands for Advanced High Strength Steel (AHSS) are rapidly increasing due to its unique combination of high strength to weight ratio and good ductility [1]. To protect the AHSS from corrosion and oxidation during its service lifespan, hot-dip galvanizing technique is widely adopted to coat the AHSS surface with a zinc (Zn) layer [2]. In order to enhance the Zn layer adhesion on the steel surface, researchers have been working on a variety of methods for several years, such as pre-metallic deposition before annealing [3,4], controlling of the dew point during annealing [5,6], and post-pickling treatment after annealing [7]. Despite these efforts, various defects, such as bare spots, dents, and scratches, can still be observed to be present in these coatings [8], which originate from the poor wettability. The defects are considered detrimental, as the quality of the surface will be degraded leading to inferior mechanical properties of the coating. In addition, they can oppose issues such as tool pollution at the customer when processing the steel. The poor wettability results from the added alloying elements, such as manganese (Mn) and chromium (Cr), which have the tendency to segregate at the steel surface during annealing and undergo selective oxidation [9]. Provided that the selective oxidation of alloying elements gets suppressed, the adhesion of Zn layer will be enhanced resulting in a better quality of the coating. Subsequently, the produced steel will be more robust against corrosion for longer time than when the selective oxidation of the alloying elements cannot be suppressed. Notably, it has been found that iron (Fe) oxidation could suppress the selective oxidation of Mn on the surface of the steel containing Mn and Si [10]. This can be achieved because after Fe pre-oxidation on the steel surface, the thickness of the solute Mn depletion zone in the subsurface of the steel becomes much larger than the outer diffusion distance of Mn from matrix to surface of the steel during annealing. Thus, understanding the Fe oxidation mechanism is crucial for achieving suppression of alloying element oxidation with high precision.

Generally, oxidation happens very fast, normally within a few seconds; and the oxidation layer is quite thin, usually a few hundred

---

[☆] This article is part of a Special issue entitled: 'Women in SCT' published in Surface & Coatings Technology.

[*] Corresponding author.
E-mail address: P.Dey@tudelft.nl (P. Dey).

https://doi.org/10.1016/j.surfcoat.2025.133092
Received 31 October 2025; Received in revised form 10 December 2025; Accepted 17 December 2025

nanometers [11]. It is difficult for existing experiments, such as low-energy electron diffraction and scanning electron microscopy (SEM), to capture this dynamical process in such a short time span with high space resolution, making it difficult to investigate the mechanism. Additionally, the rapid initial reactions involves a variety of Fe oxides [12], such as wüstite (FeO), magnetite (Fe_3O_4), and hematite (Fe_2O_3), making the investigation of Fe oxidation even more complex [13]. Oxidation stems from a series of atomistic processes, ranging from atomic oxygen (O) adsorption and diffusion to chemical bond changing (e.g., formation of new bonds of O with Fe) and surface reconstruction [14]. It is worth to mention that, adsorption is the first step for O to bind with Fe, subsequently leading to the formation of oxides [15]. Though significant progress has been made, such as in situ environmental SEM and transmission electron microscopy, to study the reactivity of O on the surface [16–19], many adsorption properties, such as O coverage dependence and surface electronic structure, are still unclear [20]. Notably, first-principles method, especially density functional theory (DFT), is extremely insightful for studying adsorption behaviors and energetics because of the quantum accuracy of this atomistic technique. Up until now, several studies have employed DFT to investigate O adsorption on Fe surfaces [21–25]. However, these studies are mainly focused on Fe(100) and Fe(110) surfaces, within which DFT is either executed using different softwares such as Vienna Ab-initio Simulation Package (VASP) [26,27] and CASTEP [28] or performed at varied accuracies using distinct exchange–correlation functionals such as Perdew–Burke–Ernzerhof (PBE) [29] and PW91 [30] (see Table 4 in Ref. [31]). Thus, thorough DFT calculations of O adsorption on various Fe surfaces using the same settings are currently lacking, which is of essential significance to obtain an in-depth insight into the initial step of Fe oxidation.

Although O adsorption could be well studied by DFT calculations, conventional DFT gives adsorption energies at 0 K and is limited to a few hundreds of atoms. Hence, it cannot be used to investigate oxidation, which is a dynamical process significantly influenced by the temperature and the O partial pressure. Remarkably, this can be better studied by molecular dynamics (MD) simulation, which is a method that uses force field or interatomic potential to describe the interactions between different atoms. It is suitable for investigating many dynamic processes at the atomistic level, such as surface tribological behavior [32,33], nanoindentation [34,35], and atomic deposition [36,37], which usually contain hundreds of thousands of atoms. Additionally, it should be noted that oxidation constantly involves bond changing. It is worth mentioning that reactive force field (ReaxFF) is effective in describing the bond breaking and forming process, due to the explicit usage of bond order in its functional form [38,39], and hence has been successfully applied in many reactive systems, for instance alumina [34] and diamond-like carbon film [40]. Therefore, it is quite appropriate for the description of the interactions between Fe and O atoms during oxidation. However, there are several ReaxFFs for the Fe–O system [41–45], rendering the choice of the suitable one a challenge. Recently, Thijs et al. benchmarked several ReaxFFs on liquid Fe oxides [33]. Nevertheless, the focus was not on the oxidation case, leaving the problem of finding one suitable ReaxFF for oxidation unsolved, which requires the benchmark study of ReaxFFs for studying oxidation on Fe surface.

Traditionally, force fields or interatomic potentials are derived from empirical methods, such as Lennard-Jones [46], the embedded-atom method [47] and ReaxFF [38,39]. While being computationally efficient, they often lack the necessary applicability and accuracy compared to DFT [48]. Notably, machine learning interatomic potentials (MLIPs) have emerged as a promising counterpart trying to overcome the challenges of high computational costs in DFT and the relatively low accuracy in MD [49]. However, long development period and comprehensive dataset requirement hinder the applicability and transferability of MLIPs, as they are unable to extrapolate to new elements or structures that are not present in the specific training dataset [50]. Recently, universal MLIPs (uMLIPs) have attracted lots of attention due to their universality and versatility [51]. Notably, uMLIPs can not only be applied to many elements on the periodic table directly, but also be utilized to perform geometry optimization or predict properties such as phonon related properties [52] and bulk properties [53]. Due to the increasing number and complexity of uMLIPs, there are a variety of benchmark studies of them, focusing on various materials and properties. For example, Focassio et al. assessed publicly available uMLIPs (MACE [54], CHGNet [55], and M3GNet [56]) for calculating the surface energies of 1497 surfaces comprising of 73 chemical elements [57]. Recently, Mehdizadeh and Schindler presented the benchmark of 19 uMLIPs for cleavage energy prediction of 36,718 surfaces of unary, binary, and ternary compounds [58]. Shuang et al. assessed the performance of 26 uMLIPs in modeling hydrogen (H)-alloy interactions and general defects in metals and random alloys, revealing the exceptional accuracy of EquiformerV2 (eqV2) [59,60] models in predicting energies and forces [61]. However, the capability and transferability of uMLIPs for Fe–O system remain elusive and their comparison with ReaxFF is not uncovered as well.

In this paper, we systematically assess the performance of relevant ReaxFFs and state-of-the-art uMLIPs for Fe–O system. Section 2 details the selection of ReaxFFs and uMLIPs, the construction of atomistic models, and the simulation settings for both O adsorption and Fe surface oxidation. Section 3 first presents the DFT results for O adsorption on various α-Fe surfaces at different coverages; then shows the adsorption benchmark results of the chosen ReaxFFs and uMLIPs; subsequently reveals the performance of ReaxFF and uMLIPs on Fe oxidation to further investigate their effectiveness and stability on dynamical simulation. Section 4 and Section 5 close the paper by discussing the underlying reason for the distinct performance of ReaxFF and uMLIP and drawing the conclusions, respectively.

2. Methods

2.1. Simulation models

All simulation models were constructed using Python Materials Genomics (pymatgen) [62] and Atomic Simulation Environment (ASE) [63] Python libraries. The body-centered cubic (BCC) Fe unit cell was obtained from the Materials Project [64], upon which four slab models for Fe surfaces with different Miller indices were constructed, namely Fe(100), Fe(110), Fe(111), and Fe(211). The slab models consisted of six Fe layers and the bottom three layers were set to be fixed. The thickness of the vacuum layer was set to be 16 Å. To represent cases for low, medium and high O coverage, $\frac{1}{9}$, $\frac{1}{4}$, and $\frac{1}{2}$ monolayer (ML) were considered, which is defined as a ratio of the number of adsorbate atoms (O) to the number of substrate atoms (Fe) in the surface layer [65]. The number of Fe atoms is the same for different surfaces but dependent on the coverage, which are 12, 24, and 54 for 1/2 ML, 1/4 ML, and 1/9 ML, respectively. The O adsorption sites were found using the AdsorbateSiteFinder tool of pymatgen [62]. As shown in Fig. 1(a)–(d), various adsorption sites as marked by green circles and labeled with abbreviations were exhibited on four Fe surfaces at $\frac{1}{9}$ ML coverage. There are mainly three types: on top (OT), bridge (B), and hollow (H). For the bridge site, other variations exist, for instance LB stands for long bridge and SB for short bridge. Notably, on Fe(111) surface, bridge site could also be found between atoms positioned in different layers: sB represents shallow bridge between first and second layers, DB stands for deep bridge between second and third layers, and SP is a saddle point between first and third layers. In addition, there are also variations of hollow site: TH is three fold hollow, while fcc and hcp are another two kinds of adsorption sites depending on different depths from the surface.

The adsorption energy E_{ads} of O on the Fe surface was calculated as:

$$E_{\text{ads}} = E_{\text{sys}} - E_{\text{Fe}} - \frac{1}{2}E_{O_2}, \tag{1}$$

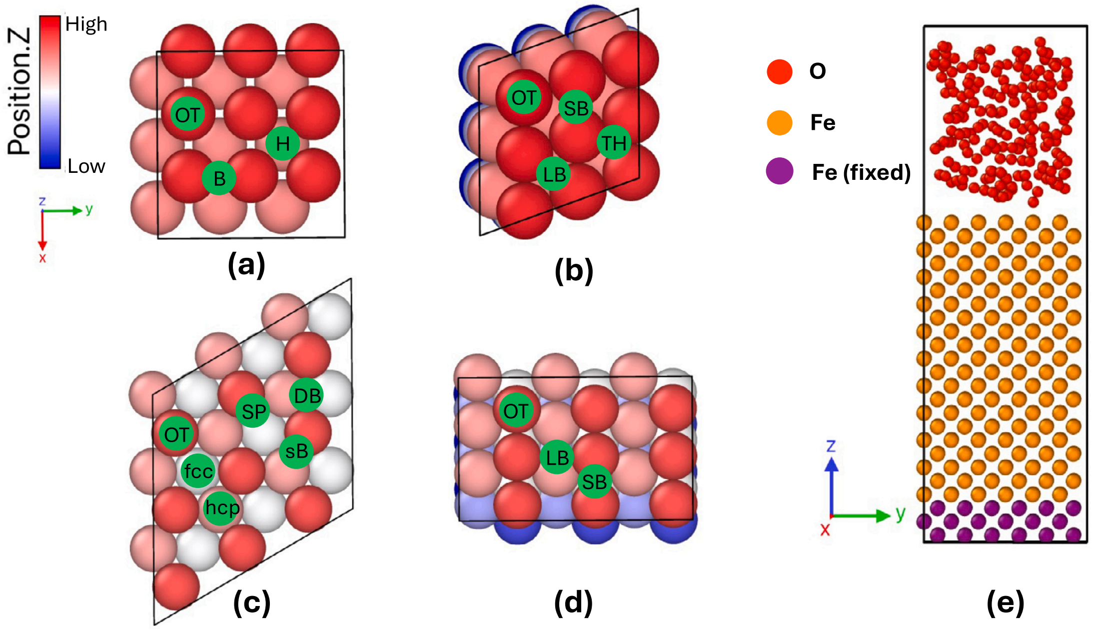


Fig. 1. Simulation models for O adsorption and Fe surface oxidation. Various O adsorption sites are shown on **(a)** Fe(100) surface; **(b)** Fe(110) surface; **(c)** Fe(111) surface; **(d)** Fe(211) surface. **(e)** Initial configuration of oxidation model of Fe(100) surface with O_2.

where E_{sys} is the total energy of the system (i.e., O atom adsorbed on Fe slab), E_{Fe} is the total energy of the relaxed Fe slab without O, and $\frac{1}{2}E_{O_2}$ is equal to half of the O_2 molecule energy. It should be noted that, the calculated E_{ads} is a negative value according to the above definition (Eq. (1)), meaning that the adsorption is energetically favorable. Note that Eq. (1) strictly applies to isolated O-atom adsorption. In this work, different coverages are achieved by varying the numbers of surface layer Fe atoms.

2.2. Selection of ReaxFFs and uMLIPs

Three relevent ReaxFFs were chosen, including those developed by Aryanpour et al. [41] (ReaxFF-Aryanpour), Huang et al. [45] (ReaxFF-Huang), and Liu et al. [44] (ReaxFF-Liu). ReaxFF-Aryanpour was trained on common Fe oxides and oxyhydroxides. Based on ReaxFF-Aryanpour and water ReaxFF [66], ReaxFF-Huang was developed by optimizing the Fe–O, Fe–H, and Fe–O–H parameters. Combining parameters from nickel (Ni)-O-H [67,68], Fe–Ni [69], and Fe–Cr–O–H [42] systems, ReaxFF-Liu additionally developed Ni–Cr parameters to have full description of Ni–Cr–Fe–H–O system.

Based on the Matbench Discovery framework [70], two novel uMLIP architectures (i.e., different uMLIP network model) were firstly adopted, namely eqV2 [59,60] and equivariant Smooth Energy Network (eSEN) [71]. For eqV2 architecture, four models were chosen, of which each has different training dataset. Specifically, OMat24 stands for Open Materials 2024 large-scale open dataset [60]. OAM represents the combination of OMat24, MPtrj (i.e., the dataset of Materials Project DFT relaxation trajectories [55]), and sAlex (i.e., subsampled Alexandria dataset [72]). OC20 and OC22 stand for Open Catalyst 2020 [73] and 2022 Dataset [74], respectively. For eSEN architecture, two models were selected based on different training dataset as well.

Notably, based on the eSEN architecture, Meta FAIR recently presented a family of Universal Models for Atoms (UMA) in order to push the frontier of speed, accuracy, and generalization [75]. In addition to Omat24 and OC20 datasets, the training dataset for UMA also consists of OMol25 (Open Molecules 2025 [76]), OMC25 (Open Molecular Crystals 2025 [77]), and ODAC25 (OpenDAC 2025 [78]) datasets. Since UMA incorporates extra information and is trained on "the largest training runs to date", it is beneficial to add those models for comparison as well.

Additionally, to perform large-scale oxidation simulation, two uMLIPs were chosen, namely NEP89 [79] and GRACE-FS (OMAT) [80], which are based on neuroevolution potential (NEP) [81] and Graph Atomic Cluster Expansion (GRACE) [82], respectively. In total, four state-of-the-art architectures, encompassing a total of 12 models, were chosen in this study, as detailed in Table 1.

2.3. Simulation settings for O adsorption and Fe surface oxidation

In order to obtain high-accuracy adsorption energies, collinear spin-polarized DFT calculations were performed using VASP [26,27]. All DFT calculations are at $T = 0$ K [83,84]. The interactions between the valence electrons and the ionic core were described using the projected augmented wave (PAW) method [85]. The generalized gradient approximation (GGA) in the PBE form was adopted for exchange–correlation energy [29]. The Brillouin zone was sampled by the Monkhorst–Pack methodology [86]. The second order Methfessel–Paxton method was used to treat fractional occupancies with the Fermi surface smearing width of 0.2 eV [87]. The k-points was generated using VASPKIT [88], with a consistent density of $2\pi\times0.03\,\text{Å}^{-1}$. A cutoff energy of 520 eV was chosen for plane-wave basis set. The convergence criteria of the energy for the electronic self-consistency loop and force for the ionic relaxation loop were set to 10^{-6} eV and 0.01 eV/Å, respectively. All ReaxFF MD calculations were performed using Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS) [89,90]. The charge equilibration (QEq) method for handling the electrostatic interactions is implemented in LAMMPS according to Aktulga et al. [91]. The QEq parameters were adopted from ReaxFF-Aryanpour. Conjugate gradient (CG) algorithm was applied for the energy minimization. The stopping tolerance for force was set to be 4×10^{-4} eV/Å. For uMLIPs, all structural relaxations were conducted by the FIRE optimizer [92] using the ASE calculator interfaced with all uMLIPs. Note that the uMLIPs used in our work do not explicitly include a dynamic charge equilibration during MD simulations. The convergence criterion $f_{\max}$ was set to 0.1 eV/Å, which means that the force on all individual atoms should be less than $f_{\max}$. This criterion was adopted by some other studies as well [55,93].

For the Fe surface oxidation simulation, Fe(100), Fe(110), and Fe(111) surfaces were selected. Taking Fe(100) surface as an example as shown in Fig. 1(e), a 6×6×24 supercell was created for each surface,

Table 1
The uMLIPs used in this work.

uMLIP	Training set	Number of parameters
eqV2-31M-omat	OMat24	31M
eqV2-31M-OAM	OMat24 + MPtrj + sAlex	31M
eqV2-31M-S2EF-OC20	OC20	31M
eqV2-S2EFS-OC22	OC22	121M
eSEN-30M-omat	OMat24	30M
eSEN-30M-OAM	OMat24 + MPtrj + sAlex	30M
uma-s-1p1-oc20	OMat24 + OC20 + OMol25 + OMC25 + ODAC25	6.6M
uma-s-1p1-omat	OMat24 + OC20 + OMol25 + OMC25 + ODAC25	6.6M
uma-m-1p1-oc20	OMat24 + OC20 + OMol25 + OMC25 + ODAC25	50M
uma-m-1p1-omat	OMat24 + OC20 + OMol25 + OMC25 + ODAC25	50M
NEP89	OMat24 + MPtrj + SPICE + ANI-1xnr + SSE-ABACUS + SSE-VASP + Protein + UNEP-v1 + CH + CHONPS + Water	N/A
GRACE-FS-OMAT	OMat24	N/A

amounting to 864 Fe atoms. Then, 100 O_2 molecules were randomly distributed on top of each Fe surface. During the oxidation simulation, periodic boundary conditions were adopted in both X and Y directions, while fixed boundary condition was applied along the Z direction, with both the top and bottom simulation box set to be reflective walls. As highlighted in Fig. 1(e), three bottom layers of the Fe slab were fixed to mimic the surface condition. Before the oxidation, the O molecules were relaxed at 300 K in the canonical ensemble (NVT) for 100 ps so as for the initial configuration to be reasonable and randomly distributed over each Fe surface. In the oxidation process, the timestep was set to 0.25 fs and the temperature of the whole system was regulated at a constant 973 K by Nosé–Hoover heat bath method [94,95] for 1 ns. All oxidation simulations were performed using LAMMPS and Open Visualization Tool (OVITO) was utilized for all the visualizations [96].

3. Results

3.1. DFT study of adsorption

To probe the behavior of O adsorption on Fe surface, systematic DFT calculations using consistent set of parameters were performed. Fig. 2 shows the DFT obtained adsorption energy results for various adsorption sites on different Fe surfaces for the three coverages, whose detailed values are displayed in Table 2. It can be seen that the adsorption energies follow a trend: top > bridge > hollow for the Fe(100) and Fe(110) surfaces. This means that the hollow site is the most stable one, followed by the bridge site, with top site being the least energetically favored. Additionally, LB site is found to be more stable than the SB site on Fe(110) surface. When it comes to the difference between coverages, it can be observed that lower coverage is preferred over high coverage, while the difference between $\frac{1}{9}$ ML and $\frac{1}{4}$ ML is observed to be negligible.

Notably, on the Fe(111) surface, there are more adsorption sites and the energetics is slightly different from the other surfaces. Specifically, in terms of site stability, hollow sites (i.e., fcc and hcp) are not the most favored ones. Instead, two bridge sites, namely sB and DB, are more stable than the rest. It is also worth to mention that for sB and DB sites the difference between all coverages are shown to be negligible compared to that of other sites. Despite small energy differences, $\frac{1}{4}$ ML is slightly favored than $\frac{1}{9}$ ML, which is contrary to Fe(100) and Fe(110) surfaces. Moreover, the energy difference between $\frac{1}{2}$ ML and $\frac{1}{4}$ ML is quite large for fcc and SP sites. On Fe(211) surface, surprisingly, no hollow site is found to exist stably. SB and LB sites are quite close in energy values and unlike other surfaces all three adsorption sites are not sensitive to different coverages.

Table 2
DFT adsorption energies of O on various Fe surfaces, for different adsorption sites and coverages. All energies are in eV.

Surface	Site	1/9 ML	1/4 ML	1/2 ML
Fe(100)	top	−1.89	−1.91	−1.62
	bridge	−2.77	−2.74	−2.44
	hollow	−3.35	−3.33	−3.14
Fe(110)	top	−1.77	−1.79	−1.57
	short bridge	−2.90	−2.83	−2.51
	long bridge	−3.33	−3.30	−2.86
	3 fold hollow	−3.35	−3.31	−3.01
Fe(111)	top	−1.85	−1.85	−1.75
	hcp	−2.58	−2.59	−2.40
	fcc	−2.11	−2.16	−1.71
	shallow bridge	−2.74	−2.82	−2.79
	saddle point	−2.39	−2.51	−1.81
	deep bridge	−2.95	−3.02	−3.03
Fe(211)	top	−1.95	−1.95	−1.92
	short bridge	−2.98	−2.99	−2.98
	long bridge	−2.95	−2.99	−2.98

3.2. MD study of adsorption

3.2.1. ReaxFF adsorption results

After the DFT results had been obtained, simulation of O adsorption on four Fe surfaces using the three chosen ReaxFFs was conducted and then the corresponding adsorption energy was compared to the DFT results. The accuracy of ReaxFF and uMLIP adsorption energies is evaluated with two complementary error metrics. First, the mean absolute error (MAE), defined as:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|E_{\mathrm{ads}}^{\mathrm{DFT},i} - E_{\mathrm{ads}}^{\mathrm{FF},i}\right|, \tag{2}$$

which measures the absolute deviation in energy. Second, the mean absolute percentage error (MAPE), defined as:

$$\mathrm{MAPE} = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{E_{\mathrm{ads}}^{\mathrm{DFT},i} - E_{\mathrm{ads}}^{\mathrm{FF},i}}{E_{\mathrm{ads}}^{\mathrm{DFT},i}}\right|, \tag{3}$$

which measures the error relative to the magnitude of the DFT reference energy. Fig. 3 shows the parity plot of the comparison between ReaxFF and DFT, where the metrics (MAE and MAPE) for different coverages are shown as well. The circle, square, and triangle shapes represent ReaxFF-Aryanpour, ReaxFF-Huang, and ReaxFF-Liu, respectively. In addition, the coefficient of determination (R^2) values for the parity plot of adsorption energy between DFT and ReaxFF are calculated (see Supplementary Table S1). It can be seen that all ReaxFFs perform poorly at reproducing the DFT results, as indicated by the negative R^2. Interestingly, ReaxFF-Aryanpour and ReaxFF-Huang have similar distribution of data points. Despite that ReaxFF-Huang

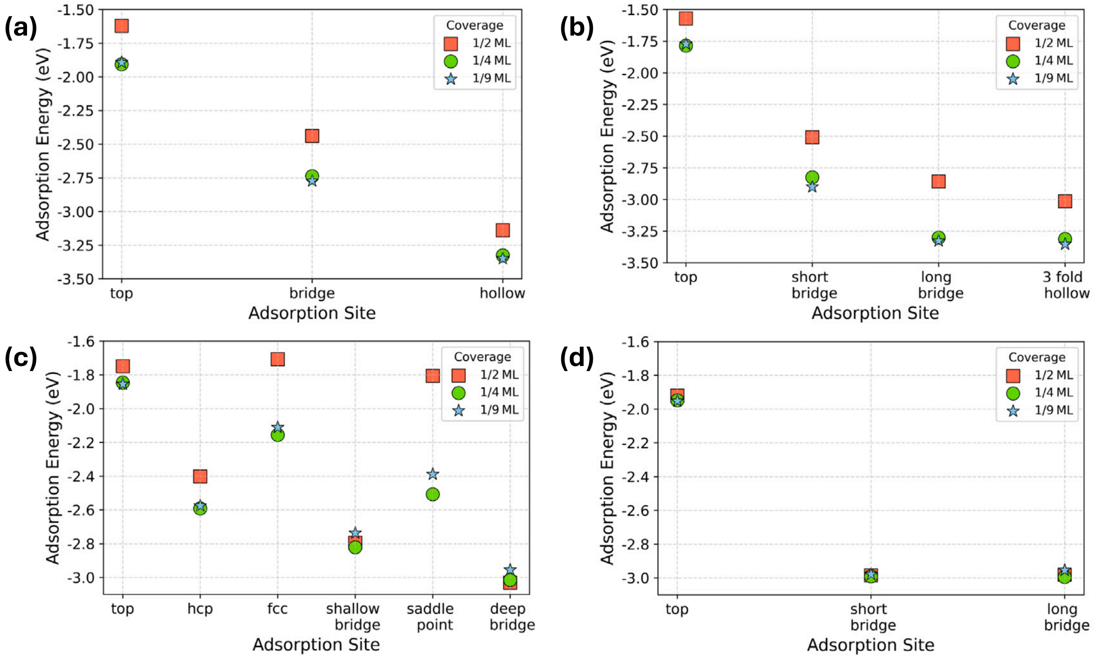


Fig. 2. DFT adsorption energy for O adsorption at various sites on Fe surfaces for three different coverages. (a) Fe(100) surface; **(b)** Fe(110) surface; **(c)** Fe(111) surface; **(d)** Fe(211) surface.

adopted the initial parameters for Fe–Fe from ReaxFF-Aryanpour, the Fe–Fe parameters were re-trained in ReaxFF-Huang. Thus, the observed similarity cannot simply be attributed to inheritance from the ReaxFF-Aryanpour parameterization.

Despite the poor performance of the ReaxFFs, it is worthwhile to investigate and compare the different capabilities of the three ReaxFFs to study O adsorption on Fe surfaces. As shown in Fig. 3(a) and (b), both ReaxFF-Aryanpour and ReaxFF-Huang perform the best in the case of the $\frac{1}{4}$ ML, followed by $\frac{1}{9}$ ML and then $\frac{1}{2}$ ML. ReaxFF-Huang explicitly includes O adsorption configurations in the parameterization of Fe–O parameters. However, ReaxFF-Aryanpour has overall better accuracy compared to ReaxFF-Huang. As indicated in Fig. 3(c), ReaxFF-Liu performs the worst of all the three ReaxFFs. Its Fe–O parameters originate from the ReaxFF developed by Shin et al. [42], whose Fe–O parameters only come from the configurations of the sulfur-substituted Fe_2O_3 and FeO systems. This explains the poor performance of ReaxFF-Liu and also indicates that if not trained explicitly on relevant structures which may appear in the actual simulation, ReaxFF would have reduced performance. In terms of the difference between coverages, ReaxFF-Liu performs the best in the case of $\frac{1}{9}$ ML and the worst in $\frac{1}{4}$ ML. But in the case of $\frac{1}{2}$ ML, it even predicts positive adsorption energy.

3.2.2. *uMLIP adsorption results*

Ten uMLIPs were utilized for the simulation of O adsorption on Fe surface, which are categorized into two sets: (1) eSEN-OAM, eSEN-OMAT, OC20, OC22, eqV2-OAM, and eqV2-OMAT; (2) uma-s-OC20, uma-s-OMAT, uma-m-OC20, and uma-m-OMAT. The first set is focused on models based on the eqV2 and eSEN architectures, which were trained on the OAM, OMAT or OC datasets. The reason for this division is that the models in this set can be conveniently compared between the same or different architectures (e.g., eSEN vs. eqV2) or datasets (e.g., OAM vs. OMAT). Except for OC22 model, all models in the first set have similar number of trainable model parameters. The second set is focused on UMA models, which are divided according to the model sizes (i.e., total number of trainable parameters). Interestingly, various tasks can be chosen when using UMA models. Thus, for comparison, the OC20 and OMAT tasks were chosen in both small and medium UMA models. To quantify the stability of uMLIPs conducting geometry optimization, the failure rate is defined, which specifically refers to the ratio of failure jobs to all calculations for O adsorption on Fe surface. A geometry optimization is considered to fail if the de-adsorption of O, the O adsorption site changing, and unreasonable configuration such as structure distortion were observed in the converged calculations. Fig. 4 shows the failure rate of the two uMLIP sets. It can be seen from Fig. 4(a) that, in the first set of uMLIPs, eSEN-OAM model has the highest failure rate, followed by the eSEN-OMAT model, with both eSEN-OAM and eSEN-OMAT models having higher failure rates than their respective eqV2 counterparts, namely eqV2-OAM and eqV2-OMAT. Notably, though OC20 and OC22 were explicitly trained on a variety of surfaces and adsorbates, their performance is not always the best in all cases. Specifically, despite that OC20 model has the best success rate in $\frac{1}{2}$ ML and $\frac{1}{4}$ ML cases, it loses its superior performance to eqV2-OAM model for the case of $\frac{1}{9}$ ML. OC22 model not only has higher overall failure rate compared to eqV2-OAM model, but also performs worse than OC20 model regardless of the coverage. In addition to the comparison between different uMLIPs, it is also worth to notice that all uMLIPs have coverage-dependent failure rate. Particularly, eSEN-OAM and eSEN-OMAT models have totally opposite failure rate trend. For eqV2-OAM and eqV2-OMAT models, the failure rate follows a trend: $\frac{1}{2}$ ML > $\frac{1}{4}$ ML > $\frac{1}{9}$ ML, with eqV2-OAM model performing slightly better overall. While for OC20 and OC22 models, the failure rate trend follows the order of $\frac{1}{2}$ ML > $\frac{1}{9}$ ML > $\frac{1}{4}$ ML.

Fig. 4(b) shows the failure rate of the second set of uMLIPs. It can be clearly seen that, regardless of the model size, OC20 task has better performance than the OMAT task. It can also be seen that larger model size has higher success rate for the same task. Interestingly, both uma-m-OMAT and uma-s-OMAT models have the same failure rate trends in different coverages as those appeared for the eqV2-OAM and eqV2-OMAT models. Similarly, the pattern between various coverages of OC20/OC22 models also holds for uma-m-OC20/uma-s-OC20 models. Notably, uma-s-OC20 model successfully finished all the calculations for the $\frac{1}{4}$ ML case.

To compare the accuracy of the uMLIPs, direct comparisons of adsorption energy between the two sets of uMLIPs and DFT are shown in Fig. 5 and Fig. 7, respectively. Fig. 5 shows the adsorption energy from all the successful calculations by the first set of uMLIPs, where the

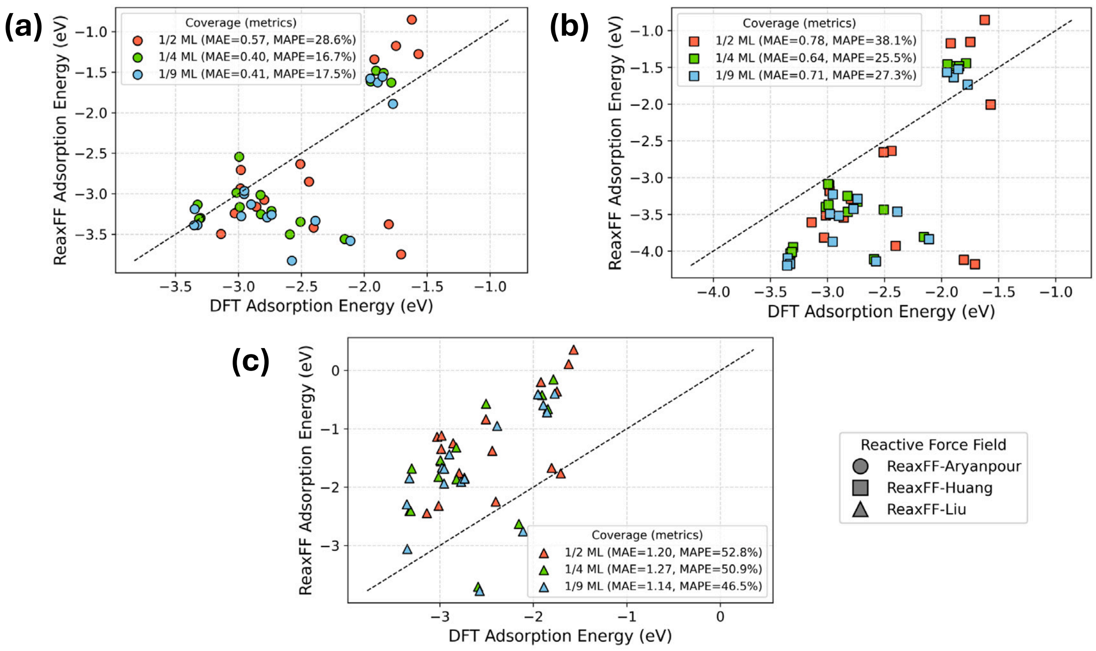


Fig. 3. Comparison of the adsorption energy between ReaxFF and DFT for different coverages. (a) ReaxFF-Aryanpour; **(b)** ReaxFF-Huang; **(c)** ReaxFF-Liu. Note that in each coverage the data points are for all the four Fe surfaces.

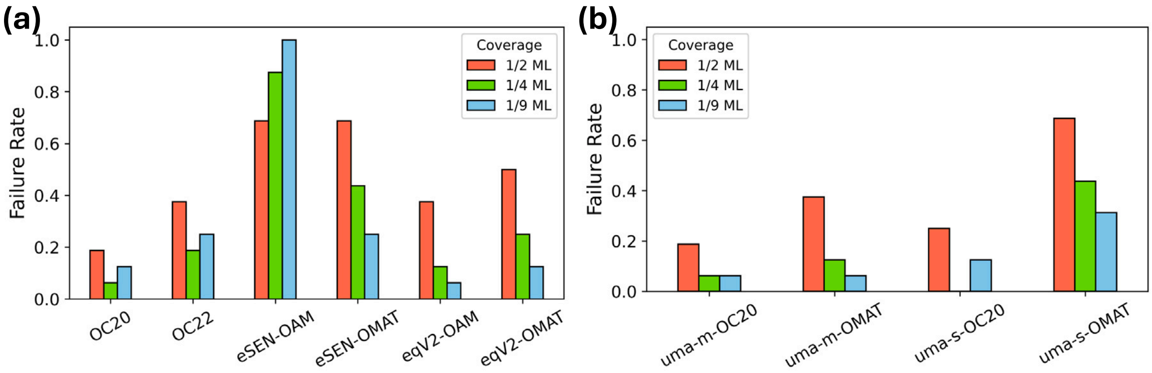
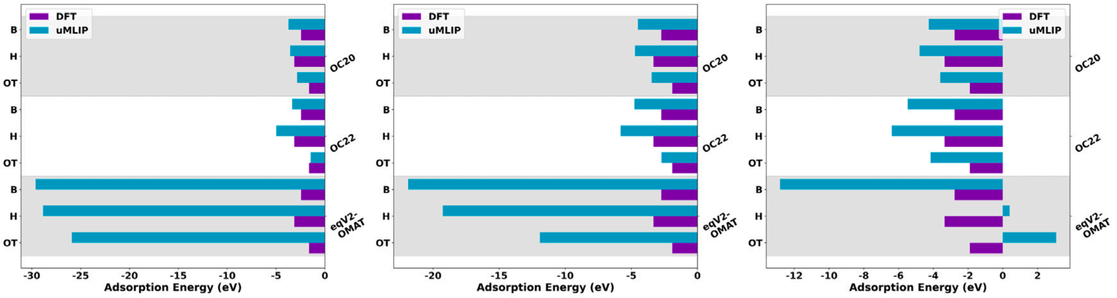


Fig. 4. Failure rate of the geometry optimization by different uMLIPs. (a) First set of uMLIPs; **(b)** Second set of uMLIPs.

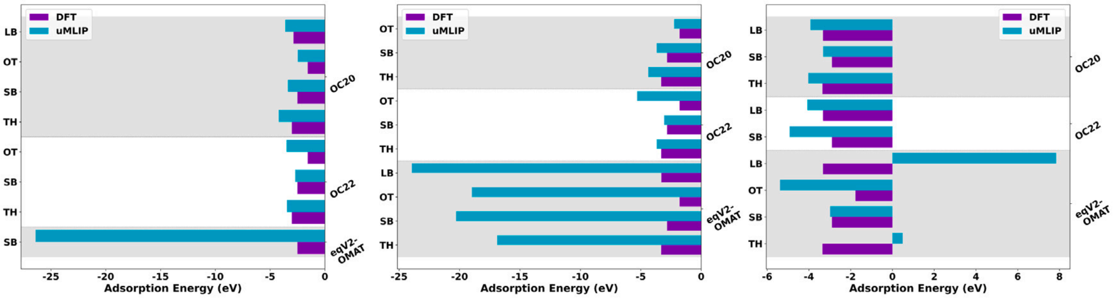
combination of various surfaces and coverages is labeled together with different adsorption sites. Since the failure rates of both eSEN models are high and the adsorption energy calculated by eqV2-OAM model is positive, only results from eqV2-OMAT, OC20, and OC22 models are displayed (see Supplementary Fig. S1 for the other three models). It can be seen that eqV2-OMAT model also gives positive adsorption energies for $\frac{1}{4}$ ML for Fe(111) surface and $\frac{1}{9}$ ML case for all the four surfaces. In addition, the adsorption energies obtained by it are quite extreme compared to those by OC20 and OC22 models.

It should be noted that the performance of the first set of uMLIPs exhibits coverage-dependence. As shown in Fig. 5(a)–(c), on Fe(100) surface, both OC20 and OC22 models align well with DFT results. eqV2-OMAT model shows contradictory results between H and B sites compared to DFT, predicting B site to be more stable. On Fe(110) surface, OC20 still maintains a good match with DFT results. However, as can be seen in Fig. 5(d)–(f), both OC22 and eqV2-OMAT models exhibit conflicting results. Specifically, OC22 predicts OT site to be more stable than TH site for the $\frac{1}{2}$ ML and $\frac{1}{4}$ ML cases and SB site to be more stable than LB site for the $\frac{1}{9}$ ML case. eqV2-OMAT also makes the mistake between OT and TH site for $\frac{1}{4}$ ML and $\frac{1}{9}$ ML cases. On Fe(111) surface, eqV2-OMAT model not only makes the wrong prediction about OT site for all coverages, but also gives positive adsorption energy of other sites for the $\frac{1}{4}$ ML and $\frac{1}{9}$ ML cases. Then, despite the accurate predictions for $\frac{1}{2}$ ML shown in Fig. 5(g) and high success rates for $\frac{1}{4}$ ML and $\frac{1}{9}$ ML shown in Fig. 5(h)–(i), OC20 and OC22 models have more tendency to make mistakes for the lower coverages. For example, as shown in Fig. 5(h)–(i), both models have wrong predictions of the fcc site stability. On Fe(211) surface as shown in Fig. 5(j)–(l), while eqV2-OMAT model has right prediction for $\frac{1}{2}$ ML, it fails for other ML cases. Both OC20 and OC22 models have successfully predicted the energetics for all coverages. Overall, it can be deducted that describing O adsorption on Fe(111) surface is a difficult task for uMLIPs.

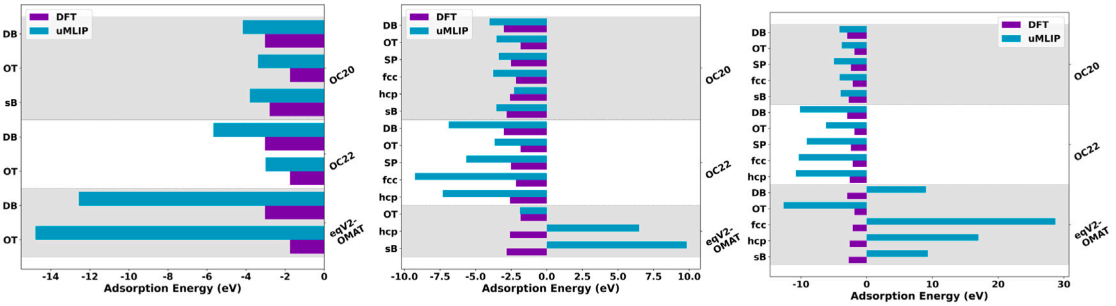
Fig. 6 shows the parity plot of the adsorption energy between DFT and the first set of uMLIPs. The results of eqV2-OMAT, OC20, and OC22 models are displayed, along with the performance metrics for different coverages. Compared to the performance of ReaxFFs, it can be clearly seen that the uMLIPs are worse. Judging by the metrics, it can be seen that OC20 model performs the best, followed by OC22 model and then eqV2-OMAT model. Notably, the performance of these models is also coverage dependent. Specifically, OC20 and OC22 models follow the trend: $\frac{1}{2}$ ML > $\frac{1}{4}$ ML > $\frac{1}{9}$ ML, while eqV2-OMAT model is in a totally opposite order with many wrong predictions. Moreover, the difference between various coverages is insignificant for OC22 model, which is contrary to the other two models.



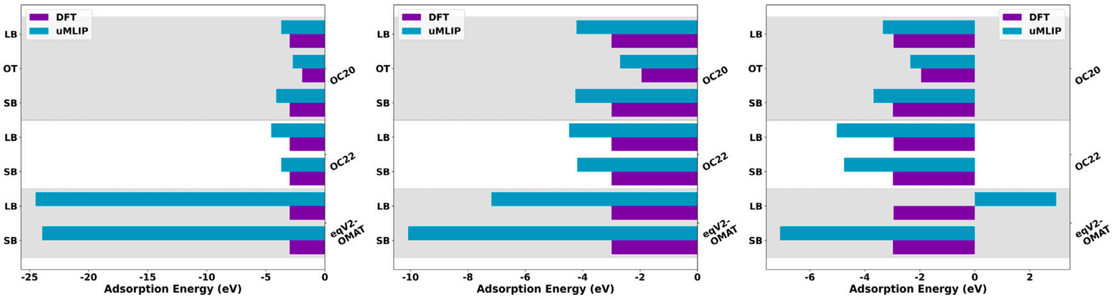
(a) Fe(100)-1/2ML **(b) Fe(100)-1/4ML** **(c) Fe(100)-1/9ML**



(d) Fe(110)-1/2ML **(e) Fe(110)-1/4ML** **(f) Fe(110)-1/9ML**



(g) Fe(111)-1/2ML **(h) Fe(111)-1/4ML** **(i) Fe(111)-1/9ML**



(j) Fe(211)-1/2ML **(k) Fe(211)-1/4ML** **(l) Fe(211)-1/9ML**

Fig. 5. Comparison of the adsorption energy between the first set of uMLIPs and DFT for different Fe surfaces. (a-c) Fe(100) surface for different coverages; **(d-f)** Fe(110) surface for different coverages; **(g-i)** Fe(111) surface for different coverages; **(j-l)** Fe(211) surface for different coverages.

The comparison of adsorption energy between the second set of uMLIPs and DFT is shown in Fig. 7. In general, it can be seen that, similar to eqV2-OMAT model in the first uMLIP set, uma models using the OMAT task also gives positive adsorption energy. However, the main difference lies in the trend of the coverage-dependent accuracy. Specifically, in contrary to the eqV2-OMAT model, both uma-s-OMAT and uma-m-OMAT models have better agreement with DFT with reducing coverage. As shown in Fig. 7(a)–(c), on Fe(100) surface, uma-s-OMAT model gives positive adsorption energy for the $\frac{1}{2}$ ML and $\frac{1}{4}$ ML cases. Despite the negative adsorption energy given for the $\frac{1}{9}$ ML case, the stability prediction is wrong between H and B sites as shown in Fig. 7(c), which is the same mistake as that of eqV2-OMAT model as shown in Fig. 5(a)–(c). For uma-m-OMAT model, it still gives positive energy for the $\frac{1}{2}$ ML, however, it yields negative energy for the $\frac{1}{4}$ ML and $\frac{1}{9}$ ML cases. As shown in Fig. 7(b), the same error of predicting B site to be more stable than H site for the $\frac{1}{4}$ ML is observed as that of uma-s-OMAT model for the $\frac{1}{9}$ ML. It distinguishes the stability of various sites for the $\frac{1}{9}$ ML case. Both uma-s-OC20 and uma-m-OC20 models succeeded for all the calculations and yielded negative adsorption energies. However, in contrary to DFT, they failed to distinguish between H and B sites across all coverages, predicting B site to be more stable.

As shown in Fig. 7(d)–(e), on Fe(110) surface, wrong predictions are made by uma-s-OMAT and uma-m-OMAT models for both $\frac{1}{2}$ ML and $\frac{1}{4}$ ML cases. But for the case of $\frac{1}{9}$ ML, both models show reasonable

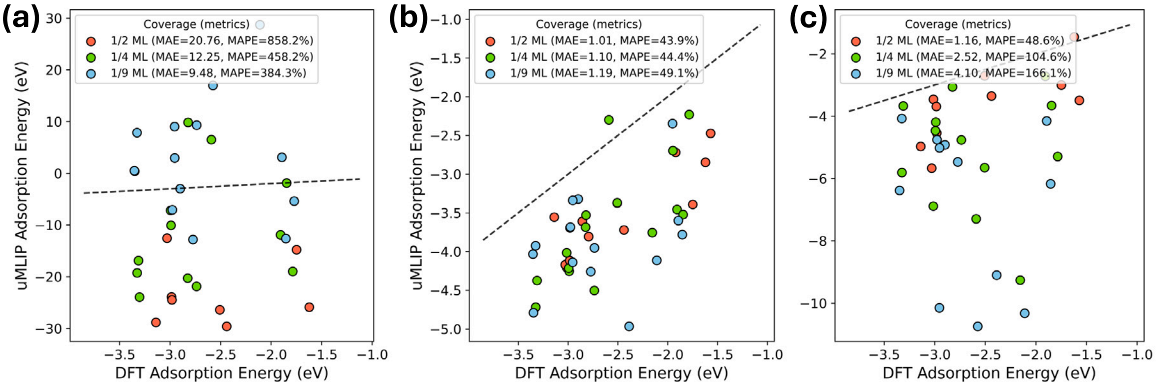


Fig. 6. Comparison of the adsorption energy between three uMLIPs of the first set and DFT for different coverages. (a) eqV2-OMAT model; **(b)** OC20 model; **(c)** OC22 model.

results as shown in Fig. 7(f). Notably, both uma-s-OC20 and uma-m-OC20 models show quite consistent result compared to DFT. As shown in Fig. 7(g)–(i), on Fe(111) surface, a notable distinction in success rate can be observed between high coverage (i.e., $\frac{1}{2}$ ML) and low coverages (i.e., $\frac{1}{4}$ and $\frac{1}{9}$ ML), indicating the poor performance of uMLIPs in this case. For the case of $\frac{1}{4}$ and $\frac{1}{9}$ ML as shown in Fig. 7(h)–(i), uma-s-OMAT model shows bad performance and uma-m-OMAT model performs relatively well except for the prediction of the fcc site. In addition, it is worthwhile to mention that both uma-s-OC20 and uma-m-OC20 models make errors in the prediction of hcp and fcc sites, underestimating and overestimating the stability, respectively. As shown in Fig. 7(j)–(l), on Fe(211) surface, uma-s-OMAT model still shows poor performance, while the performance of uma-m-OMAT model is better. In contrary, both uma-s-OC20 and uma-m-OC20 models exhibits agreeable results with respect to the DFT calculations.

Fig. 8 shows the parity plot of the adsorption energy between DFT and the second set of uMLIPs. It can be observed that totally different trend of coverage-dependent performance exists in uma models on different tasks. For uma-s-OMAT and uma-m-OMAT models, the performance is the best for the case of $\frac{1}{9}$ ML, followed by $\frac{1}{4}$ ML, with $\frac{1}{2}$ ML being the worst. It is also worthwhile to mention that larger model has better performance. While for the other two models, namely uma-s-OC20 and uma-m-OC20, the trend is totally opposite. Both models perform the best for $\frac{1}{2}$ ML and the worst for $\frac{1}{9}$ ML, with uma-s-OC20 model performing better.

3.3. MD study of surface oxidation

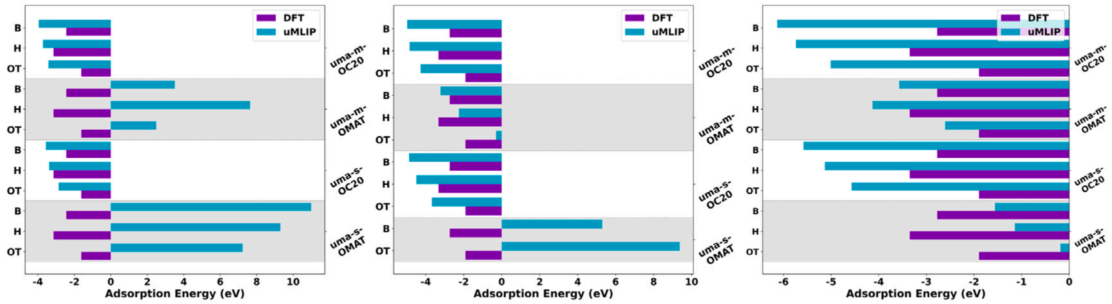
3.3.1. ReaxFF oxidation results

To further investigate the effectiveness of both ReaxFF and uMLIP on the dynamical oxidation process, MD simulation of Fe surface oxidation was conducted, as exemplified in Fig. 1(e). It is known that the initial Fe oxidation is a rapid process, where different oxides can form, such as FeO, Fe_3O_4, and Fe_2O_3 [12]. Normally, thick oxide layer has compositions of Fe:FeO:Fe_3O_4:Fe_2O_3, depending on the distance of the layer from the Fe/oxide interface [97]. The FeO forms first, which is stable above 843.15 K [98], and subsequently transforms to Fe_3O_4 or Fe_2O_3 [99]. Additionally, only FeO may be obtained in the much thinner Fe–Oxides produced during pre-oxidation in the galvanizing lines. Thus, in this study, the temperature of 973 K is adopted in all the MD simulations, which is slightly above the formation temperature of FeO. It is shown in Section 3.2.1 that ReaxFF-Aryanpour has overall better performance than ReaxFF-Huang, which should be chosen for Fe surface oxidation simulation. However, there are several studies available that have performed oxidation simulation using ReaxFF-Aryanpour [100,101], where three surfaces were chosen for comparison, namely Fe(100), Fe(110), and Fe(111). In order to enrich the comparison while keeping consistency with previous studies, ReaxFF-Huang is adopted in this study for the Fe–O atomic interactions to investigate the surface oxidation of Fe(100), Fe(110), and Fe(111). It is worthwhile mentioning that, similar to ReaxFF-Aryanpour, ReaxFF-Huang also includes the lattice constants and heats of formation of various Fe minerals in its parameterization.

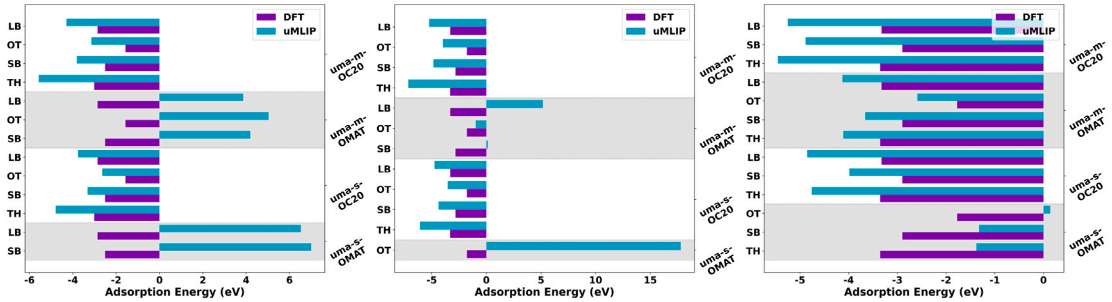
Fig. 9 shows the final configurations of different Fe surfaces after 1 ns oxidation simulation. It can be clearly seen that different surfaces have varied oxidation degree judging by the number of consumed O atoms, which follows the trend: Fe(111) > Fe(110) > Fe(100). This observation aligns well with the results of oxidation at 900 K in the study of Jeon et al. [100], despite that different ReaxFFs were used. In addition, the thicknesses of oxidation layer on three surfaces are measured, which are found to be 16 Å, 20 Å, and 15 Å for Fe(100), Fe(110), and Fe(111), respectively. Notably, it can be seen that oxidation layers on Fe surfaces are quite ordered and exhibit crystalline structures. To probe the structure of oxidation layer, radial distribution function (RDF) $g(r)$ is obtained, which measures the probability of finding a particle at distance r given that there is a particle at position $r = 0$. For comparison, FeO structure is constructed and then relaxed at 973 K for 1 ns. Fig. 10 shows the comparison of the RDF between the final configuration of the relaxed FeO structure and various oxidation layers extracted from Fe surface oxidation. The RDFs of the oxidation layers show similar shapes in terms of the peak positions of different pairs, indicating the similar structures of the oxidation layers. Then, compared to the RDF of the FeO structure, it can be seen that the peaks of both Fe–O and Fe–Fe pairs match well between the FeO structure and oxidation layers. The only difference lies in the peaks of O–O pair. This may be attributed to the insufficient oxidation simulation time. Overall, ReaxFF-Huang is found to be effective for simulating Fe surface oxidation and could provide reasonable structural information. Additionally, it is worthwhile to mention that comparing classical MD oxidation results with ab initio MD (AIMD) would be ideal to validate the force field predictions. However, conducting AIMD for oxidation at finite temperature is computationally prohibitive given the system size and timescales needed for more realistic oxidation. Thus, future work could be done by focusing on much smaller model systems.

3.3.2. uMLIP oxidation results

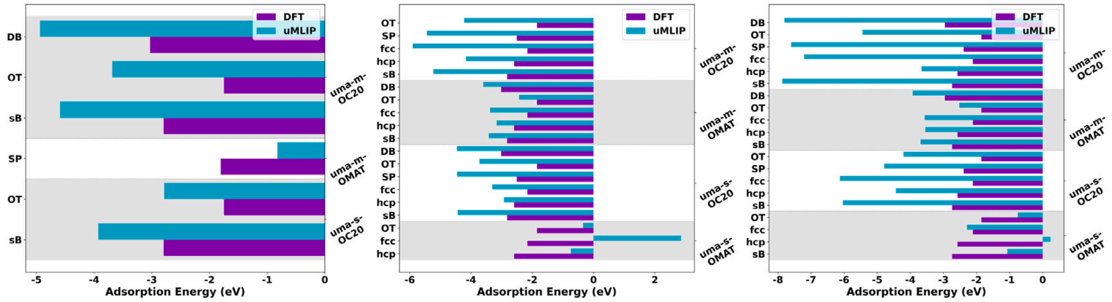
The efficiency of various uMLIPs has been reported in the study of Shuang et al. [61]. It is shown that, despite the increased accuracy compared to other uMLIPs, the eqV2 models are computationally intensive, almost one order of magnitude slower than the Orb model [102]. In addition, in the study of Wood et al. [75], the simulation speeds of UMA and eSEN-OAM models are also compared to the Orb model. The speed is tested on a periodic system of 1000 atoms and measured by steps per second, which is 3 and 1.7 for UMA and eSEN-OAM models, respectively. Still, the two models are one order of magnitude slower than the Orb model (i.e., 30 steps per second for 1000 atoms). Instead,



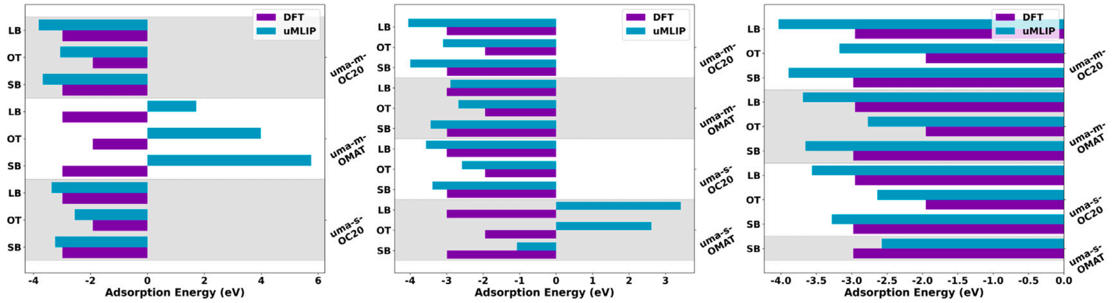
(a) Fe(100)-1/2ML (b) Fe(100)-1/4ML (c) Fe(100)-1/9ML



(d) Fe(110)-1/2ML (e) Fe(110)-1/4ML (f) Fe(110)-1/9ML



(g) Fe(111)-1/2ML (h) Fe(111)-1/4ML (i) Fe(111)-1/9ML



(j) Fe(211)-1/2ML (k) Fe(211)-1/4ML (l) Fe(211)-1/9ML

Fig. 7. Comparison of the adsorption energy between the second set of uMLIPs and DFT for different Fe surfaces. (a-c) Fe(100) surface for different coverages; **(d-f)** Fe(110) surface for different coverages; **(g-i)** Fe(111) surface for different coverages; **(j-l)** Fe(211) surface for different coverages.

the NEP89 [79] and GRACE-FS [80] foundation models are quite fast and have interfaces with LAMMPS, which is very convenient to perform the large-scale simulation. As shown in Table 3, we tested the computational speeds of various uMLIPs for NVE MD simulation of a 115-atom BCC Fe system using ASE on the AMD 9654 CPU. Thus, they are selected for the simulation of Fe surface oxidation. The same simulation models and conditions are adopted and applied as those in ReaxFF MD simulations. During the structural relaxations before oxidation, both uMLIP models have reasonable descriptions for O molecules and Fe surfaces (see Supplementary Fig. S2(a)–(b) and Fig. S3(a)–(b)), which indicates the stability of uMLIPs. However, under the same oxidation condition as that of ReaxFF MD simulations, no oxidation of the fully relaxed system is observed (see Supplementary Fig. S2(c) and Fig. S3(c)). This may be attributed to the absence of related structures in the training dataset. In order to investigate the effectiveness of uMLIPs representing Fe–Oxide, the FeO structure is again considered. It is first geometrically optimized by the DFT calculation. Subsequently, the relaxation of the FeO structure using NEP89 and GRACE-FS models is conducted at 973 K for 1 ns. Then, the RDFs of the final configurations are compared, as shown in Fig. 11. It can be seen that the RDFs from both uMLIP models not only exhibit similar shapes, but also match well with that from DFT calculation. This means that both uMLIP models can provide reasonable description of the FeO structure. Notably, the RDF shapes from both uMLIP models are smoother than the one obtained

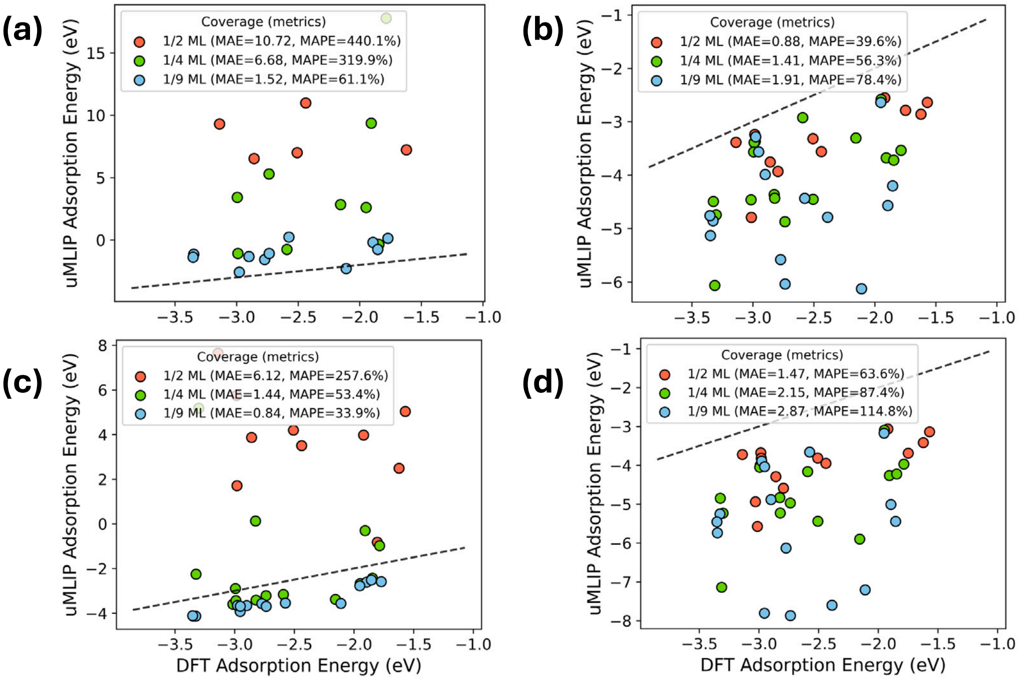


Fig. 8. Comparison of the adsorption energy between the second set of uMLIPs and DFT for different coverages. (a) uma-s-OMAT model; **(b)** uma-s-OC20 model; **(c)** uma-m-OMAT model; **(d)** uma-m-OC20 model.

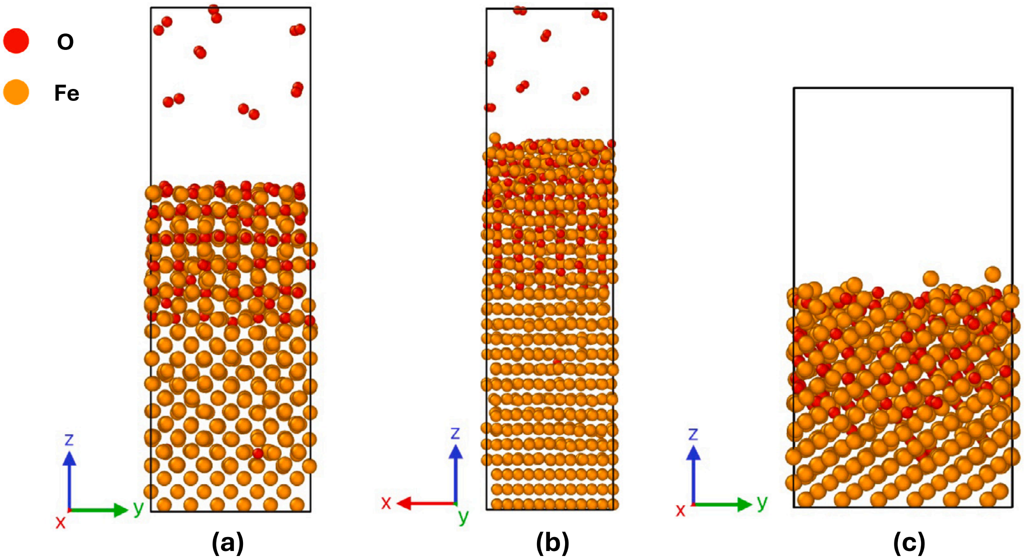


Fig. 9. Final configurations for different Fe surfaces after 1 ns ReaxFF-MD simulation of Fe surface oxidation. (a) Fe(100) surface; **(b)** Fe(110) surface; **(c)** Fe(111) surface.

by DFT, which is due to the smooth potential energy surface generated by the foundation models. In addition, the computational speeds for the relaxation of the wüstite structure are calculated using LAMMPS, which are 2.79, 1.01, and 0.57 katom-step/s/core for NEP89, GRACE-FS, and ReaxFF, respectively. Overall, despite the failure in describing the reaction between O molecules and Fe atoms of both uMLIP models, reasonable structural information could be obtained by them, showing their potential capability to describe the oxidation product.

4. Discussion

In this work, relevant ReaxFFs regarding the Fe–O system and advanced uMLIPs are benchmarked against DFT for the case of O adsorption and Fe surface oxidation. It can be seen from the comparison between ReaxFF and uMLIP that for the Fe–O system, ReaxFF outperforms uMLIP in terms of either the accuracy of reproducing DFT or the stability of describing the oxidation. This indicates that ReaxFF could be the first choice for the reactive systems. However, provided that several relevant ones exist, careful benchmarks need to be done on essential properties in order to make the final choice, for example benchmarking the adsorption energy for the case of oxidation. The rational behind this is that, if not explicitly trained on relevant structures which may appear in the actual simulation, ReaxFF can have poor performance, as exemplified in the comparison between ReaxFF-Aryanpour/ReaxFF-Huang and ReaxFF-Liu in terms of the adsorption energy with regard to DFT.

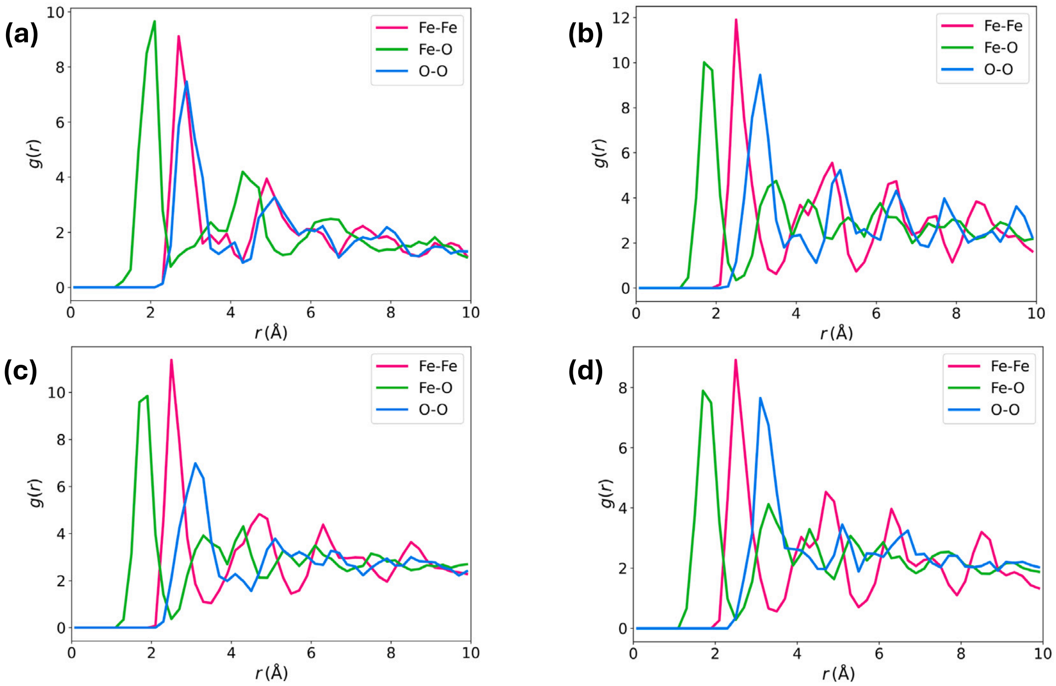


Fig. 10. **Radial distribution function of the wüstite structure and oxidation layers for different Fe surfaces obtained from ReaxFF-MD simulation.** **(a)** the wüstite structure; **(b)** the oxidation layer on Fe(100) surface; **(c)** the oxidation layer on Fe(110) surface; **(d)** the oxidation layer on Fe(111) surface.

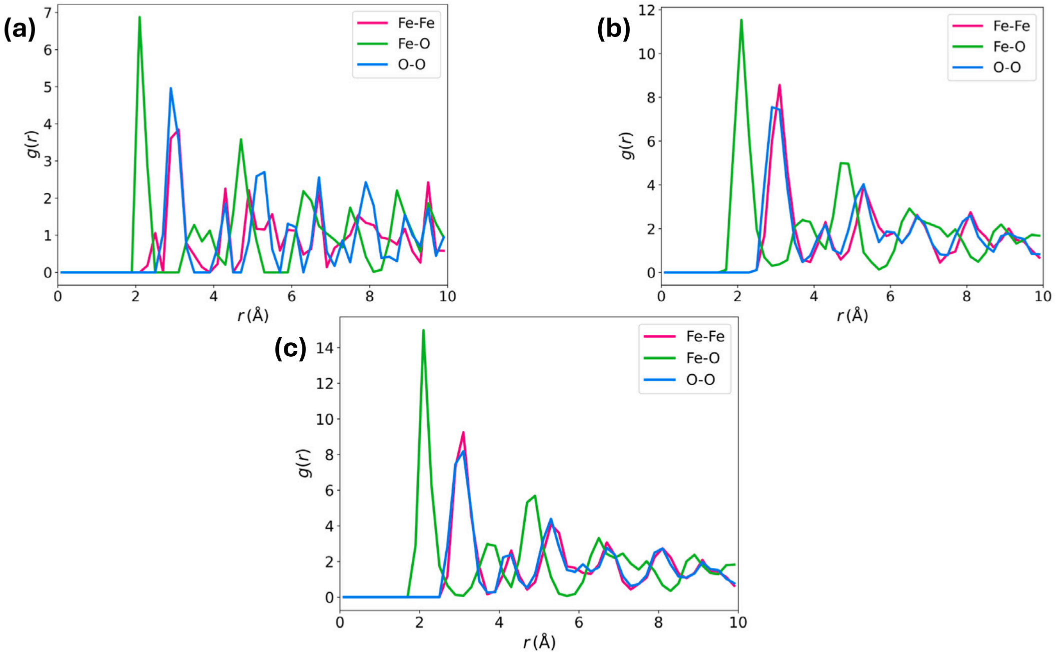


Fig. 11. **Radial distribution function of the wüstite structure obtained by** **(a)** DFT; **(b)** NEP89; **(c)** GRACE-FS.

Notably, the better performance of ReaxFF over uMLIP in describing O adsorption and Fe surface oxidation for the Fe–O system may be attributed to a few factors. First of all, ReaxFF has a functional form that explicitly include the bond order [38], which is appropriate for the description of the reactions [39]. Then, it is known that the Fe–O system itself is a difficult one to study, owing to the challenge to accurately describe the electronic and magnetic properties, such as the electronic band gap and ground state magnetism of various Fe oxides [103]. Next, despite that uMLIPs are trained on millions of structures, they may sacrifice certain degree of accuracy (e.g., using lower convergence parameter) and consistency (e.g., mixed usage of PBE and PBE+U) in order to get the converged DFT results for the structures as diverse as possible. Additionally, unlike in ReaxFF where charges are dynamically redistributed every time step, uMLIP runs are performed under a fixed-charge (or effectively neutral) assumption. The absence of explicit electrostatics/charge transfer likely impairs the ability of uMLIP to describe reactive adsorption and oxidation processes.

Nevertheless, uMLIP also has its unique advantages in both direct use and further development. In terms of its direct use, as shown in the study of Shuang et al. [61], uMLIP demonstrates the remarkable potential for the static property predictions, particularly configurational

Table 3
Computational speed of various uMLIPs on the 115-atom BCC Fe system. All speeds are in atom-step/s/core.

uMLIP	Computational speed
eqV2-OAM	3.59
eqV2-OMAT	3.56
eSEN-OAM	1.18
eSEN-OMAT	1.13
OC20	3.85
OC22	1.31
uma-s-OC20	4.27
uma-s-OMAT	4.36
uma-m-OC20	0.68
uma-m-OMAT	0.68
NEP89	238.89

energy and atomic forces. Hence, uMLIP could be a rough estimate for the energy and force of the structures for the Fe–O system. In addition, combined with the findings from the FeO RDF comparison between uMLIP and DFT, it can be deducted that uMLIP could also be adopted as pre-relaxation calculator if certain structures are difficult to directly get the converged results in DFT. Another advantage of the uMLIP that cannot be ignored is its universality, which is the convenient applicability of including any elements of interest, for instance including Mn or Si for the study of the steel oxidation. It terms of the further development, notably, with the development of several techniques, such as the distillation [104] and fine-tuning [105], both the speed and accuracy of the uMLIP could be significantly enhanced, leading to the improved capability in the simulations such as geometry optimization. For example, in order to boost the applicability of uMLIP in the Fe surface oxidation, future work could be focused on fine-tuning the uMLIP using the structures sampled from the ReaxFF MD simulation of the Fe surface oxidation.

Last but not least, developing the system-specific MLIP (sMLIP) could also be a practical option for the Fe–O system, which is extremely helpful for the correct description of the physical property, for instance the formation mechanisms of interstitial Fe atoms and typical defect clusters in FeO [106]. Notably, in order to improve the sMLIP robustness, uncertainty quantification (UQ) needs to be carefully considered, which assesses the reliability of the energies and forces predicted by the sMLIP [107]. UQ is significantly essential for the dataset generation in the sMLIP development cycle, as it helps detect the out-of-distribution configurations that needs to be included [107]. Moreover, given that UQ in the uMLIPs is quantified, the advantage of the uMLIPs could be better utilized, significantly speeding up the sMLIP development cycle. For example, in a recent study of Liu et al. [108], an uncertainty-aware model distillation framework was proposed using the heterogeneous uMLIP ensemble and then further applied to produce the sMLIPs. It has been verified that with less (e.g., less than 4%) or no DFT data, the produced sMLIPs achieve comparable accuracy to teacher models with significantly reduced computational cost.

5. Conclusion

Systematic DFT calculations using consistent settings were performed to elucidate the O adsorption mechanism on various Fe surfaces for different coverages. On Fe(100) and Fe(110) surfaces, the adsorption energies of various sites follow a trend: top > bridge > hollow. While on Fe(111) and Fe(211) surfaces, the bridge sites are more stable with respective to the top and hollow sites. Then, the comparison of the adsorption energy between ReaxFFs and DFT reveals the good performance of ReaxFFs in reproducing the DFT results. In addition, ReaxFFs exhibits varied performance with regard to coverage. This is also observed in both sets of uMLIPs: (1) eSEN-OAM, eSEN-OMAT, OC20, OC22, eqV2-OAM, and eqV2-OMAT; (2) uma-s-OC20, uma-s-OMAT, uma-m-OC20, and uma-m-OMAT. Next, the comparison of the stability for uMLIPs performing geometry optimization highlights OC20, OC22, eqV2-OAM in the first set and uma-m-OC20, uma-m-OMAT in the second set. Direct comparison of the adsorption energy in the first set of uMLIPs further shows the strength of OC20 and OC22 models, with OC20 model outperforming OC22 model, while revealing numerous wrong predictions of eqV2-OAM model with regard to DFT. For the second set of uMLIPs, despite that uma-s-OC20 and uma-m-OC20 models outperforming uma-s-OMAT and uma-m-OMAT models, they make obvious mistakes, predicting contradictory results about B and H sites on Fe(100) surface and hcp and fcc sites on Fe(111) surface with respect to DFT. Moreover, it is worthwhile to mention that all uMLIPs have poor performances for the Fe(111) surface. Finally, surface oxidation simulation has been conducted for Fe(100), Fe(110), and Fe(111) surfaces using ReaxFF-Huang, NEP89, and GRACE-FS. ReaxFF-Huang exhibits exceptional stability and is quite effective for providing rational structural information. However, both NEP89 and GRACE-FS fail to capture the reactive process, indicating their limited capability in modeling Fe–O system oxidation. Nevertheless, the comparison of the RDFs for the FeO structure obtained from these models and DFT reveals their potential to reasonably describe the final oxidation product.

CRediT authorship contribution statement

Zixiong Wei: Writing – review & editing, Writing – original draft, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Fei Shuang:** Writing – review & editing, Data curation, Conceptualization. **Poulumi Dey:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was carried out under project number T23003 in the framework of the Research Program of the Materials innovation institute (M2i) (www.m2i.nl) supported by the Dutch government. This work was supported by the Nederlandse Organisatie voor Wetenschappelijk Onderzoek, Netherlands (NWO; The Netherlands Organization for Scientific Research), Domain Science, for access to supercomputing facilities. We also acknowledge the use of the DelftBlue supercomputer provided by the Delft High Performance Computing Center (DHPC; https://www.tudelft.nl/dhpc). We thank Dr. Wanda Melfo Prada and Dr. Marga Zuijderwijk from Tata Steel Nederland Technology B.V. for fruitful discussions and insightful suggestions.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.surfcoat.2025.133092.

Data availability

Data will be made available on request.

References

[1] H. Liu, F. Li, W. Shi, S. Swaminathan, Y. He, M. Rohwerder, L. Li, Challenges in hot-dip galvanizing of high strength dual phase steel: Surface selective oxidation and mechanical property degradation, Surf. Coat. Technol. 206 (16) (2012) 3428–3436.

[2] A.R. Marder, The metallurgy of zinc-coated steel, Prog. Mater. Sci. 45 (3) (2000) 191–271.

[3] A. Chakraborty, R. Ghosh, M. Sudan, A. Mondal, Improvement in hot dip galvanized coating microstructure and properties by pre-metallic deposition on steel surface: A comprehensive review, Surf. Coat. Technol. 449 (2022) 128972.

[4] M.G. Walunj, L.C. Pathak, S. Parida, V. Raja, Hot dip galvanization of DP 980 steel sheet using iron based sol pre-layer and its corrosion performance, Surf. Coat. Technol. 504 (2025) 132042.

[5] M. Pourmajidian, J.R. McDermid, Selective oxidation of a 0.1C-6Mn-2Si third generation advanced high-strength steel during dew-point controlled annealing, Met. Mater. Trans. A 49 (5) (2018) 1795–1808.

[6] M.G. Walunj, G.K. Mandal, R.K. Ranjan, R. Pais, S.K. Mishra, T. Venugopalan, L.C. Pathak, Role of dew points and Fe pre-coats on the galvanizing and galvannealing of dual phase steel, Surf. Coat. Technol. 422 (2021) 127573.

[7] X. Jin, G. Hu, L. Wang, H. Wang, Characterization of surface layers of oxidation-reduction treated Si and Mn added advanced high strength steel, Surf. Coat. Technol. 382 (2020) 125172.

[8] X. Jin, J. Hong, Z. Ma, W. Mao, New insights into the formation mechanism of a bare-spot defect in hot-dip galvanized coating, Surf. Coat. Technol. (2025) 132576.

[9] S. Samanta, A.K. Halder, Y. Deo, S. Guha, M. Dutta, Effect of Mn and Cr on the selective oxidation, surface segregation and hot-dip Zn coatability, Surf. Coat. Technol. 377 (2019) 124908.

[10] Y. Fushiwaki, Y. Nagataki, H. Nagano, W. Tanimoto, Y. Sugimoto, Influence of Fe oxidation on selective oxidation behavior of Si and Mn added in high strength sheet steel, ISIJ Int. 54 (3) (2014) 664–670.

[11] M. Blumenau, C.O. Gusek, M. Norden, R. Schönenberg, Industrial use of preoxidation during continuous hot-dip coating of (high) alloyed steels, in: Proc., AISTech 2012 Conference, Atlanta, USA, 2012, pp. 7–10.

[12] A. Atkinson, Transport processes during the growth of oxide films at elevated temperature, Rev. Modern Phys. 57 (2) (1985) 437.

[13] R. Chen, W. Yeun, Review of the high-temperature oxidation of iron and carbon steels in air or oxygen, Oxid. Met. 59 (5) (2003) 433–468.

[14] X. Zhang, P. Zheng, Y. Ma, Y. Jiang, H. Li, Atomic-scale understanding of oxidation mechanisms of materials by computational approaches: A review, Mater. Des. 217 (2022) 110605.

[15] K.R. Lawless, The oxidation of metals, Rep. Progr. Phys. 37 (2) (1974) 231.

[16] Q. Zhu, Z. Pan, Z. Zhao, G. Cao, L. Luo, C. Ni, H. Wei, Z. Zhang, F. Sansoz, J. Wang, Defect-driven selective metal oxidation at atomic scale, Nat. Commun. 12 (1) (2021) 558.

[17] G. Zhou, K.A. Unocic, C. Wang, Z. Shan, S.J. Haigh, J.C. Yang, Revealing atomic-to-nanoscale oxidation mechanisms of metallic materials, MRS Bull. 48 (8) (2023) 852–863.

[18] X. Chen, W. Shan, D. Wu, S.B. Patel, N. Cai, C. Li, S. Ye, Z. Liu, S. Hwang, D.N. Zakharov, et al., Atomistic mechanisms of water vapor–induced surface passivation, Sci. Adv. 9 (44) (2023) eadh5565.

[19] W. Tu, Z. Zeng, Y. Zhao, Z. Jia, C. Meng, W. Liu, X. Sun, J. Huang, In situ multiscale study of iron oxidation at high temperatures, Nano Lett. 25 (20) (2025) 8089–8095.

[20] H. Xing, P. Hu, S. Li, Y. Zuo, J. Han, X. Hua, K. Wang, F. Yang, P. Feng, T. Chang, Adsorption and diffusion of oxygen on metal surfaces studied by first-principle study: A review, J. Mater. Sci. Technol. 62 (2021) 180–194.

[21] P. Błoński, A. Kiejna, J. Hafner, Theoretical study of oxygen adsorption at the Fe(110) and (100) surfaces, Surf. Sci. 590 (1) (2005) 88–100.

[22] P. Błoński, A. Kiejna, J. Hafner, Oxygen adsorption on the clean and O-precovered Fe(110) and (100) surfaces, J. Phys.: Condens. Matter. 19 (9) (2007) 096011.

[23] S. Chen, M.-L. Giorgi, J.-B. Guillot, G. Geneste, Oxidation and diffusion processes at the Mn-doped Fe(001) and Fe(110) surfaces from first-principles, Appl. Surf. Sci. 258 (22) (2012) 8613–8618.

[24] K. Freindl, T. Ossowski, M. Zajac, N. Spiridis, D. Wilgocka-Slezak, E. Madej, T. Giela, A. Kiejna, J. Korecki, Oxygen adsorption on the Fe(110) surface: The old system–new structures, J. Phys. Chem. C 120 (7) (2016) 3807–3813.

[25] J.J. White, J. Liu, J.J. Hinsch, Y. Wang, Theoretical understanding of the properties of stepped iron surfaces with van der waals interaction corrections, Phys. Chem. Chem. Phys. 23 (4) (2021) 2649–2657.

[26] G. Kresse, J. Furthmüller, Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set, Phys. Rev. B 54 (16) (1996) 11169.

[27] G. Kresse, J. Furthmüller, Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set, Comput. Mater. Sci. 6 (1) (1996) 15–50.

[28] M.C. Payne, M.P. Teter, D.C. Allan, T. Arias, a.J. Joannopoulos, Iterative minimization techniques for ab initio total-energy calculations: Molecular dynamics and conjugate gradients, Rev. Modern Phys. 64 (4) (1992) 1045.

[29] J.P. Perdew, K. Burke, M. Ernzerhof, Generalized gradient approximation made simple, Phys. Rev. Lett. 77 (18) (1996) 3865.

[30] J.P. Perdew, J.A. Chevary, S.H. Vosko, K.A. Jackson, M.R. Pederson, D.J. Singh, C. Fiolhais, Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation, Phys. Rev. B 46 (11) (1992) 6671.

[31] H. Balasooriya, C. Li, F. Wang, Understanding steel corrosion: Surface chemistry and defects explored through DFT modelling—A review, Processes 13 (7) (2025) 1971.

[32] R.-t. Tong, Z.-f. Quan, B. Han, G. Liu, Coarse-grained molecular dynamics simulation on friction behaviors of textured ag-coating under vacuum and microgravity environments, Surf. Coat. Technol. 359 (2019) 265–271.

[33] J. Zhang, G. Deng, W. Li, S. Li, Y. Yan, X. Liu, T. Ma, L. Gao, Grain size and scratching depth dependent tribological characteristics of CrCoNi medium-entropy alloy coatings: A molecular dynamics simulation study, Surf. Coat. Technol. 468 (2023) 129772.

[34] H.-T. Luu, S. Raumel, F. Dencker, M. Wurz, N. Merkert, Nanoindentation in alumina coated Al: Molecular dynamics simulations and experiments, Surf. Coat. Technol. 437 (2022) 128342.

[35] G. Zhao, Z. Liu, J. Li, H. Li, L. Ma, et al., Investigating the influence of graphene coating thickness on Al$_{0.3}$CoCrFeNi high-entropy alloy using molecular dynamics simulation, Surf. Coat. Technol. 493 (2024) 131255.

[36] X. Li, S. Xu, P. Ke, A. Wang, Thickness dependence of properties and structure of ultrathin tetrahedral amorphous carbon films: A molecular dynamics simulation, Surf. Coat. Technol. 258 (2014) 938–942.

[37] Y.-N. Chen, T.-B. Ma, P.-Z. Zhu, D.-C. Yue, Y.-Z. Hu, Z. Chen, H. Wang, Growth mechanism of hydrogenated amorphous carbon films: Molecular dynamics simulations, Surf. Coat. Technol. 258 (2014) 901–907.

[38] A.C. Van Duin, S. Dasgupta, F. Lorant, W.A. Goddard, Reaxff: A reactive force field for hydrocarbons, J. Phys. Chem. A 105 (41) (2001) 9396–9409.

[39] T.P. Senftle, S. Hong, M.M. Islam, S.B. Kylasa, Y. Zheng, Y.K. Shin, C. Junkermeier, R. Engel-Herbert, M.J. Janik, H.M. Aktulga, et al., The ReaxFF reactive force-field: Development, applications and future directions, Npj Comput. Mater. 2 (1) (2016) 1–14.

[40] Y. Liu, H. Zhang, Y. Luo, L. Wang, C. Xiao, Probing the low friction mechanism of hydrogen-free DLC film in oxygen and nitrogen environments by first-principles calculations and molecular dynamics simulation, Surf. Coat. Technol. 455 (2023) 129219.

[41] M. Aryanpour, A.C. Van Duin, J.D. Kubicki, Development of a reactive force field for iron- oxyhydroxide systems, J. Phys. Chem. A 114 (21) (2010) 6298–6307.

[42] Y.K. Shin, H. Kwak, A.V. Vasenkov, D. Sengupta, A.C. Van Duin, Development of a reaxff reactive force field for Fe/Cr/O/S and application to oxidation of butane over a pyrite-covered Cr$_2$O$_3$ catalyst, ACS Catal. 5 (12) (2015) 7226–7236.

[43] L. Ai, H. Huang, Y. Zhou, M. Chen, Y. Lü, The oxidation of Fe/Ni alloy surface with supercritical water: A ReaxFF molecular dynamics simulation, Appl. Surf. Sci. 553 (2021) 149519.

[44] X. Liu, S.-Y. Kim, S.H. Lee, B. Lee, Atomistic investigation on initiation of stress corrosion cracking of polycrystalline Ni$_6$0Cr$_3$0Fe$_1$0 alloys under high-temperature water by reactive molecular dynamics simulation, Comput. Mater. Sci. 187 (2021) 110087.

[45] Y. Huang, C. Hu, Z. Xiao, N. Gao, Q. Wang, Z. Liu, W. Hu, H. Deng, Atomic insight into iron corrosion exposed to supercritical water environment with an improved Fe-H$_2$O reactive force field, Appl. Surf. Sci. 580 (2022) 152300.

[46] J.E. Jones, On the determination of molecular fields.—I. From the variation of the viscosity of a gas with temperature, Proc. R. Soc. Lond. Ser. A Contain. Pap. A Math. Phys. Character 106 (738) (1924) 441–462.

[47] M.S. Daw, S.M. Foiles, M.I. Baskes, The embedded-atom method: A review of theory and applications, Mater. Sci. Rep. 9 (7–8) (1993) 251–310.

[48] P. Friederich, F. Häse, J. Proppe, A. Aspuru-Guzik, Machine-learned potentials for next-generation matter simulations, Nat. Mater. 20 (6) (2021) 750–761.

[49] G. Wang, C. Wang, X. Zhang, Z. Li, J. Zhou, Z. Sun, Machine learning interatomic potential: Bridge the gap between small-scale models and realistic device-scale simulations, Iscience 27 (5) (2024).

[50] Y.-W. Zhang, V. Sorkin, Z.H. Aitken, A. Politano, J. Behler, A.P. Thompson, T.W. Ko, S.P. Ong, O. Chalykh, D. Korogod, et al., Roadmap for the development of machine learning-based interatomic potentials, Modelling Simul. Mater. Sci. Eng. 33 (2) (2025) 023301.

[51] R. Jacobs, D. Morgan, S. Attarian, J. Meng, C. Shen, Z. Wu, C.Y. Xie, J.H. Yang, N. Artrith, B. Blaiszik, et al., A practical guide to machine learning interatomic potentials–status and future, Curr. Opin. Solid State Mater. Sci. 35 (2025) 101214.

[52] A. Loew, D. Sun, H.-C. Wang, S. Botti, M.A. Marques, Universal machine learning interatomic potentials are ready for phonons, Npj Comput. Mater. 11 (1) (2025) 178.

[53] H. Kraß, J. Huang, S.M. Moosavi, MOFSimBench: Evaluating universal machine learning interatomic potentials in metal–organic framework molecular modeling, 2025, arXiv preprint arXiv:2507.11806.

[54] I. Batatia, D.P. Kovacs, G. Simm, C. Ortner, G. Csányi, MACE: Higher order equivariant message passing neural networks for fast and accurate force fields, Adv. Neural Inf. Process. Syst. 35 (2022) 11423–11436.

[55] B. Deng, P. Zhong, K. Jun, J. Riebesell, K. Han, C.J. Bartel, G. Ceder, CHGNet as a pretrained universal neural network potential for charge-informed atomistic modelling, Nat. Mach. Intell. 5 (9) (2023) 1031–1041.

[56] C. Chen, S.P. Ong, A universal graph deep learning interatomic potential for the periodic table, Nat. Comput. Sci. 2 (11) (2022) 718–728.

[57] B. Focassio, L.P. M. Freitas, G.R. Schleder, Performance assessment of universal machine learning interatomic potentials: Challenges and directions for materials' surfaces, ACS Appl. Mater. Interfaces 17 (9) (2024) 13111–13121.

[58] A. Mehdizadeh, P. Schindler, Surface stability modeling with universal machine learning interatomic potentials: A comprehensive cleavage energy benchmarking study, 2025, arXiv preprint arXiv:2508.21663.

[59] Y.-L. Liao, B. Wood, A. Das, T. Smidt, EquiformerV2: Improved equivariant transformer for scaling to higher-degree representations, 2023, arXiv preprint arXiv:2306.12059.

[60] L. Barroso-Luque, M. Shuaibi, X. Fu, B.M. Wood, M. Dzamba, M. Gao, A. Rizvi, C.L. Zitnick, Z.W. Ulissi, Open materials 2024 (OMat24) inorganic materials dataset and models, 2024, arXiv preprint arXiv:2410.12771.

[61] F. Shuang, Z. Wei, K. Liu, W. Gao, P. Dey, Universal machine learning interatomic potentials poised to supplant DFT in modeling general defects in metals and random alloys, Mach. Learn.: Sci. Technol. (2025).

[62] S.P. Ong, W.D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V.L. Chevrier, K.A. Persson, G. Ceder, Python materials genomics (pymatgen): A robust, open-source python library for materials analysis, Comput. Mater. Sci. 68 (2013) 314–319.

[63] A.H. Larsen, J.J. Mortensen, J. Blomqvist, I.E. Castelli, R. Christensen, M. Dułak, J. Friis, M.N. Groves, B. Hammer, C. Hargus, et al., The atomic simulation environment—a python library for working with atoms, J. Phys.: Condens. Matter. 29 (27) (2017) 273002.

[64] A. Jain, S.P. Ong, G. Hautier, W. Chen, W.D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, et al., Commentary: The materials project: A materials genome approach to accelerating materials innovation, APL Mater. 1 (1) (2013).

[65] T. Ossowski, A. Kiejna, Oxygen adsorption on Fe(110) surface revisited, Surf. Sci. 637 (2015) 35–41.

[66] W. Zhang, A.C. Van Duin, Second-generation ReaxFF water force field: Improvements in the description of water density and OH-anion diffusion, J. Phys. Chem. B 121 (24) (2017) 6021–6032.

[67] O. Verners, A.C. Van Duin, Comparative molecular dynamics study of fcc-Ni nanoplate stress corrosion in water, Surf. Sci. 633 (2015) 94–101.

[68] C. Zou, Y.K. Shin, A.C. Van Duin, H. Fang, Z.-K. Liu, Molecular dynamics simulations of the effects of vacancies on nickel self-diffusion, oxygen diffusion and oxidation initiation in nickel, using the ReaxFF reactive force field, Acta Mater. 83 (2015) 102–112.

[69] Y.K. Shin, H. Kwak, C. Zou, A.V. Vasenkov, A.C. Van Duin, Development and validation of a reaxff reactive force field for Fe/Al/Ni alloys: Molecular dynamics study of elastic constants, diffusion, and segregation, J. Phys. Chem. A 116 (49) (2012) 12163–12174.

[70] J. Riebesell, R.E. Goodall, P. Benner, Y. Chiang, B. Deng, G. Ceder, M. Asta, A.A. Lee, A. Jain, K.A. Persson, A framework to evaluate machine learning crystal stability predictions, Nat. Mach. Intell. 7 (6) (2025) 836–847.

[71] X. Fu, B.M. Wood, L. Barroso-Luque, D.S. Levine, M. Gao, M. Dzamba, C.L. Zitnick, Learning smooth and expressive interatomic potentials for physical property prediction, 2025, arXiv preprint arXiv:2502.12147.

[72] J. Schmidt, T.F. Cerqueira, A.H. Romero, A. Loew, F. Jäger, H.-C. Wang, S. Botti, M.A. Marques, Improving machine-learning models in materials science through large datasets, Mater. Today Phys. 48 (2024) 101560.

[73] L. Chanussot, A. Das, S. Goyal, T. Lavril, M. Shuaibi, M. Riviere, K. Tran, J. Heras-Domingo, C. Ho, W. Hu, et al., Open catalyst 2020 (OC20) dataset and community challenges, Acs Catal. 11 (10) (2021) 6059–6072.

[74] R. Tran, J. Lan, M. Shuaibi, B.M. Wood, S. Goyal, A. Das, J. Heras-Domingo, A. Kolluru, A. Rizvi, N. Shoghi, et al., The open catalyst 2022 (OC22) dataset and challenges for oxide electrocatalysts, ACS Catal. 13 (5) (2023) 3066–3084.

[75] B.M. Wood, M. Dzamba, X. Fu, M. Gao, M. Shuaibi, L. Barroso-Luque, K. Abdelmaqsoud, V. Gharakhanyan, J.R. Kitchin, D.S. Levine, et al., UMA: A family of universal models for atoms, 2025, arXiv preprint arXiv:2506.23971.

[76] D.S. Levine, M. Shuaibi, E.W.C. Spotte-Smith, M.G. Taylor, M.R. Hasyim, K. Michel, I. Batatia, G. Csányi, M. Dzamba, P. Eastman, et al., The open molecules 2025 (OMmol25) dataset, evaluations, and models, 2025, arXiv preprint arXiv:2505.08762.

[77] V. Gharakhanyan, L. Barroso-Luque, Y. Yang, M. Shuaibi, K. Michel, D.S. Levine, M. Dzamba, X. Fu, M. Gao, X. Liu, H. Ni, K. Noori, B.M. Wood, M. Uyttendaele, A. Boromand, C.L. Zitnick, N. Marom, Z.W. Ulissi, A. Sriram, Open molecular crystals 2025 (OMC25) dataset and models, 2025, arXiv preprint arXiv:2508.02651.

[78] A. Sriram, L.M. Brabson, X. Yu, S. Choi, K. Abdelmaqsoud, E. Moubarak, P. de Haan, S. Löwe, J. Brehmer, J.R. Kitchin, M. Welling, C.L. Zitnick, Z. Ulissi, A.J. Medford, D.S. Sholl, The open DAC 2025 dataset for sorbent discovery in direct air capture, 2025, arXiv preprint arXiv:2508.03162.

[79] T. Liang, K. Xu, E. Lindgren, Z. Chen, R. Zhao, J. Liu, E. Berger, B. Tang, B. Zhang, Y. Wang, et al., NEP89: Universal neuroevolution potential for inorganic and organic materials across 89 elements, 2025, arXiv preprint arXiv:2504.21286.

[80] Y. Lysogorskiy, A. Bochkarev, R. Drautz, Graph atomic cluster expansion for foundational machine learning interatomic potentials, 2025, arXiv preprint arXiv:2508.17936.

[81] Z. Fan, Z. Zeng, C. Zhang, Y. Wang, K. Song, H. Dong, Y. Chen, T. Ala-Nissila, Neuroevolution machine learning potentials: Combining high accuracy and low cost in atomistic simulations and application to heat transport, Phys. Rev. B 104 (10) (2021) 104309.

[82] A. Bochkarev, Y. Lysogorskiy, R. Drautz, Graph atomic cluster expansion for semilocal interactions beyond equivariant message passing, Phys. Rev. X 14 (2) (2024) 021036.

[83] P. Hohenberg, W. Kohn, Inhomogeneous electron gas, Phys. Rev. 136 (3B) (1964) B864.

[84] W. Kohn, L.J. Sham, Self-consistent equations including exchange and correlation effects, Phys. Rev. 140 (4A) (1965) A1133.

[85] G. Kresse, D. Joubert, From ultrasoft pseudopotentials to the projector augmented-wave method, Phys. Rev. B 59 (3) (1999) 1758.

[86] H.J. Monkhorst, J.D. Pack, Special points for Brillouin-zone integrations, Phys. Rev. B 13 (12) (1976) 5188.

[87] M. Methfessel, A.T. Paxton, High-precision sampling for Brillouin-zone integration in metals, Phys. Rev. B 40 (6) (1989) 3616.

[88] V. Wang, N. Xu, J.-C. Liu, G. Tang, W.-T. Geng, VASPKIT: A user-friendly interface facilitating high-throughput computing and analysis using VASP code, Comput. Phys. Comm. 267 (2021) 108033.

[89] S. Plimpton, Fast parallel algorithms for short-range molecular dynamics, J. Comput. Phys. 117 (1) (1995) 1–19.

[90] A.P. Thompson, H.M. Aktulga, R. Berger, D.S. Bolintineanu, W.M. Brown, P.S. Crozier, P.J. In't Veld, A. Kohlmeyer, S.G. Moore, T.D. Nguyen, et al., LAMMPS-a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales, Comput. Phys. Comm. 271 (2022) 108171.

[91] H.M. Aktulga, J.C. Fogarty, S.A. Pandit, A.Y. Grama, Parallel reactive molecular dynamics: Numerical methods and algorithmic techniques, Parallel Comput. 38 (4–5) (2012) 245–259.

[92] E. Bitzek, P. Koskinen, F. Gähler, M. Moseler, P. Gumbsch, Structural relaxation made simple, Phys. Rev. Lett. 97 (17) (2006) 170201.

[93] H. Yu, M. Giantomassi, G. Materzanini, J. Wang, G.-M. Rignanese, Systematic assessment of various universal machine-learning interatomic potentials, Mater. Genome Eng. Adv. 2 (3) (2024) e58.

[94] S. Nosé, A unified formulation of the constant temperature molecular dynamics methods, J. Chem. Phys. 81 (1) (1984) 511–519.

[95] W.G. Hoover, Canonical dynamics: Equilibrium phase-space distributions, Phys. Rev. A 31 (3) (1985) 1695.

[96] A. Stukowski, Visualization and analysis of atomistic simulation data with OVITO–the open visualization tool, Modelling Simul. Mater. Sci. Eng. 18 (1) (2009) 015012.

[97] C.M. Wang, D.R. Baer, L.E. Thomas, J.E. Amonette, J. Antony, Y. Qiang, G. Duscher, Void formation during early stages of passivation: Initial oxidation of iron nanoparticles at room temperature, J. Appl. Phys. 98 (9) (2005).

[98] A. Pignocco, G. Pellissier, Low-energy electron diffraction studies of oxygen adsorption and oxide formation on a (001) iron surface, J. Electrochem. Soc. 112 (12) (1965) 1188.

[99] G.W. Simmons, D.J. Dwyer, A LEED-AES study of the initial stages of oxidation of Fe(001), Surf. Sci. 48 (2) (1975) 373–392.

[100] B. Jeon, Q. Van Overmeere, A.C. Van Duin, S. Ramanathan, Nanoscale oxidation and complex oxide growth on single crystal iron surfaces and external electric field effects, Phys. Chem. Chem. Phys. 15 (6) (2013) 1821–1830.

[101] R. Subbaraman, S.A. Deshmukh, S.K. Sankaranarayanan, Atomistic insights into early stage oxidation and nanoscale oxide growth on Fe(100), Fe(111) and Fe(110) surfaces, J. Phys. Chem. C 117 (10) (2013) 5195–5207.

[102] M. Neumann, J. Gin, B. Rhodes, S. Bennett, Z. Li, H. Choubisa, A. Hussey, J. Godwin, Orb: A fast, scalable neural network potential, 2024, arXiv preprint arXiv:2410.22570.

[103] Y. Meng, X.-W. Liu, C.-F. Huo, W.-P. Guo, D.-B. Cao, Q. Peng, A. Dearden, X. Gonze, Y. Yang, J. Wang, et al., When density functional approximations meet iron oxides, J. Chem. Theory Comput. 12 (10) (2016) 5132–5144.

[104] J.L. Gardner, D.F. Toit, C.B. Mahmoud, Z.F. Beaulieu, V. Juraskova, L.-B. Paşca, L.A. Rosset, F. Duarte, F. Martelli, C.J. Pickard, et al., Distillation of atomistic foundation models across architectures and chemical domains, 2025, arXiv preprint arXiv:2506.10956.

[105] J. Kim, J. Lee, S. Oh, Y. Park, S. Hwang, S. Han, S. Kang, Y. Kang, An efficient forgetting-aware fine-tuning framework for pretrained universal machine-learning interatomic potentials, 2025, arXiv preprint arXiv:2506.15223.

[106] Z. Liang, K. Li, J. Zhang, A.N. Conejo, Insights into defect cluster formation in non-stoichiometric wüstite ($Fe_{1-x}O$) at elevated temperatures: Accurate force field from deep learning, Npj Comput. Mater. 11 (1) (2025) 32.

[107] F. Shuang, Z. Wei, K. Liu, W. Gao, P. Dey, Model accuracy and data heterogeneity shape uncertainty quantification in machine learning interatomic potentials, 2025, arXiv preprint arXiv:2508.03405.

[108] K. Liu, Z. Wei, W. Gao, P. Dey, M.H. Sluiter, F. Shuang, Heterogeneous ensemble enables a universal uncertainty metric for atomistic foundation models, 2025, arXiv preprint arXiv:2507.21297.