THE JOURNAL OF PHYSICAL CHEMISTRY C
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCC
Article

Multidimensional Neural Network Interatomic Potentials for CO on NaCl(100)

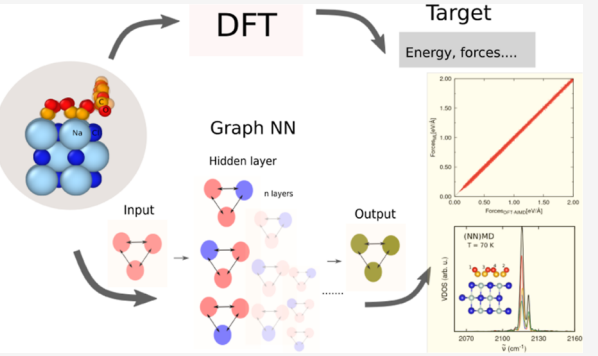
Published as part of The Journal of Physical Chemistry C special issue "Alec Wodtke Festschrift".

Shreya Sinha, Bruno Mladineo, Ivor Lončarić, and Peter Saalfrank*

Cite This: J. Phys. Chem. C 2024, 128, 21117−21131
Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The advent of machine learning (ML) models has unlocked new possibilities in the realm of interatomic potentials. We use an equivariant graph Neural Network (NN) to construct interatomic potentials for a versatile system, CO on a NaCl(100) surface, and mediate efficient large-scale atomistic simulations with ab initio molecular dynamics accuracy. We report two NN potentials, one trained on equilibrium configurations at finite temperatures (T = 30, 300 K), and the other additionally trained upon nonequilibrium trajectories of pre-excited CO adsorbates. We demonstrate first applications of the ML potentials for (i) adsorption energies and barriers for reactions, (ii) potential energy landscapes for submonolayer and monolayer coverages, (iii) vibrational spectra at finite temperatures, and (iv) vibrational relaxation dynamics. Further possible applications are discussed.



1. INTRODUCTION

Atomic scale simulations equip researchers with an effective tool for gaining fundamental insights into microscopic mechanisms of processes occurring in materials;[1] their accuracy and reliability often determined by the quality of the interatomic potentials. Access to time scales in the order of nanoseconds in molecular dynamics simulations, for instance, is usually provided by classical interatomic potentials. With ab initio potentials, however, this becomes computationally expensive and time scales in the order of picoseconds are mostly affordable. Additionally, fitting ab initio energy points to a smooth analytical function is often nontrivial and requires a flexible, accurate, reliable, and fast interpolation scheme, not introducing additional errors on top of the level of theory utilized to obtain the energy points. Several interpolation methods have been discussed in the literature,[2−5] with specific advantages and disadvantages. With increasing dimensionality, finding appropriate analytical forms for the interatomic potential becomes cumbersome or even impossible, however. This is where machine learning (ML) comes into play.

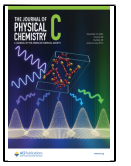
The fundamental concept of ML potentials is to bypass traditional physics-based potential construction and instead predict the potential energy of the system through numerical interpolation, based on known reference data, such as energies and forces obtained from quantum mechanical calculations. Neural networks (NN) are one of the most frequently used ML approaches to construct interatomic potentials and mediate efficient large-scale atomistic simulations.[6] NN potentials utilizing invariant representations have been successfully constructed for polyatomic molecules interacting with frozen surfaces, such as H_2O/Ni(111),[7] H_2/Ru(0001),[8] and CH_3OH/Ni(111).[9] In addition, high-dimensional NN potentials that incorporate surface atoms have also been developed, with notable examples including HCl/Au(111),[10] and N_2/Ru(0001),[11] and more recently for systems like CO/Au(111),[12] CO/Pd(111),[13] and CO/Ru(0001).[14] Further, reviews on the development of atomistic NN potentials for gas-surface systems can be found in refs 15 and 16.

The use of so-called *equivariant* NNs is becoming increasingly popular as enforcing invariance across all intermediate layers of the NN can lead to the loss of relative pose information of local features. A *G-equivariant* NN is a linear map, Φ: $f(X_1) \rightarrow f'(X_2)$, with the property[17]

$$\Phi[U_g f(x)] = U'_g \Phi[f(x)], \qquad \forall\, g \in G, \forall\, f(x)$$

where the group, G has actions on two homogeneous spaces X_1 and X_2, and U_g as well as U'_g are G-function transforms on the two spaces.

Received: August 27, 2024
Revised: November 5, 2024
Accepted: November 11, 2024
Published: November 26, 2024



https://doi.org/10.1021/acs.jpcc.4c05765
J. Phys. Chem. C 2024, 128, 21117−21131

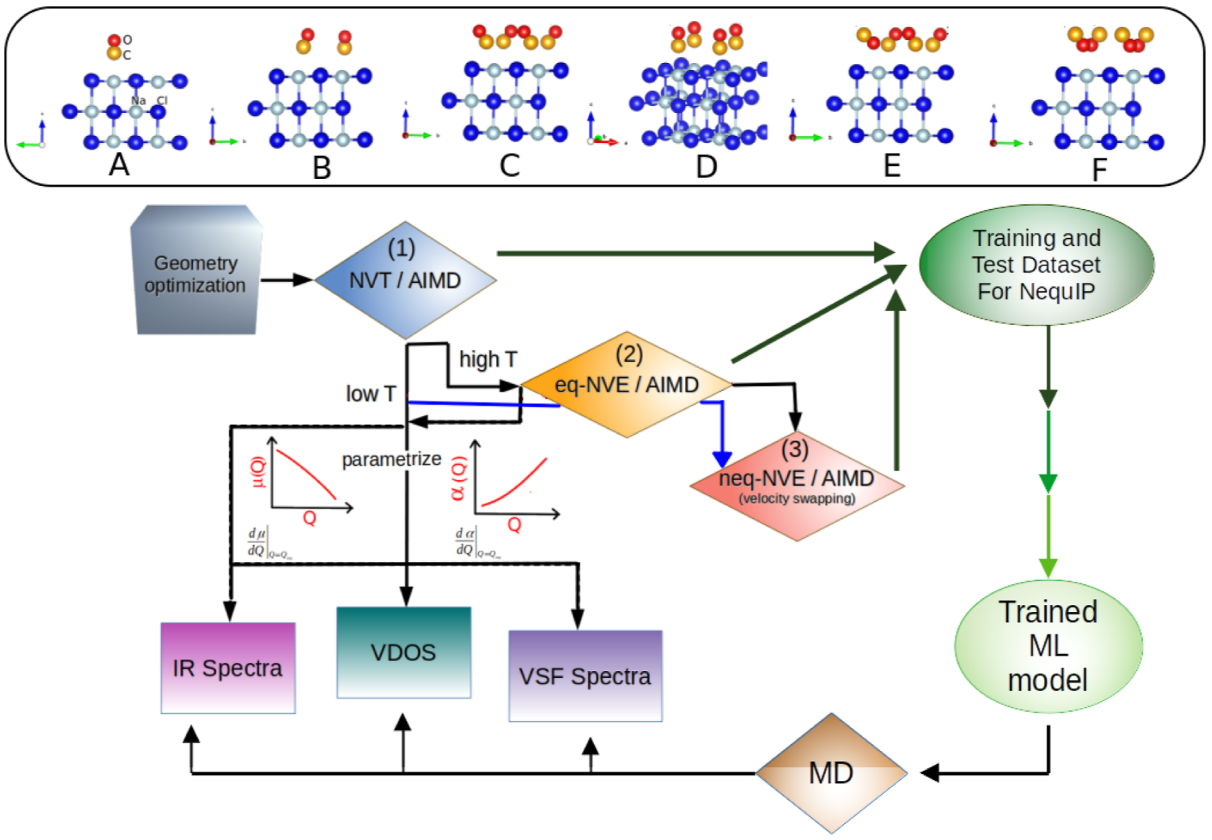


Figure 1. Upper panel: optimized geometries of some initial configurations (A−F) representing different coverages and phases, which are propagated in the AIMD trajectories. Lower panel: a schematic diagram of the generation of the ML training and test data set from the AIMD simulations. Three phases of the AIMD simulations are run with periodic PBE + D2 using the VASP package, namely, the equilibration phase in the *NVT* ensemble (1), followed by the production phase in the *NVT* ensemble (also 1), and then a short NVE phase (eq-NVE) to remove thermostat effects (2). The figure also illustrates how molecular dynamics (MD) using the final trained ML model can be used to compute vibrational spectra [IR, vibrational density of states (VDOS), and vibrational sum frequency (VSF)], which also usually requires a parametrization of molecular dipole moments (μ) and polarizabilities (α)—see text and Section SD for details. To make certain other special applications possible (vibrational energy relaxation, see below), also nonequilibrium AIMD [neq-NVE, (3)] calculations carried out in the *NVE* ensemble were used to train the network, in which excess energy was initially pumped into particular modes of CO molecules, by a "velocity swapping" algorithm (see below). The blue arrows indicate eq-NVE runs conducted for low-temperature equilibrated configurations only when utilized for nonequilibrium vibrational relaxation dynamics studies. In our case, low T corresponds to T = 30 K, and high T corresponds to T = 300 K.

One example of an [$G = E(3)$] equivariant NN is "NequIP" (neural equivariant interatomic potentials), recently developed by Batzner et al.[18,19] NequIP employs $E(3)$ equivariant convolutions for interaction of geometric tensors, resulting in a significant reduction in the size of training data sets required to achieve high-fidelity potentials.

This reduction allows us to employ high-quality reference data for training and to perform molecular dynamics simulations on the trained potentials over extended time scales. NequIP has been shown to be effective in training energies and interatomic forces of different molecular or periodic systems, sometimes reducing the amount of necessary input data for training by 3 orders of magnitude.[18] As only one example, we mention applications of NequIP to study phase transformation in amorphous lithium phosphate and kinetic transport (Li-ion diffusion), by using reference data sets obtained from ab initio molecular dynamics (AIMD) simulations.[18]

Here, we wish to utilize this approach to CO adsorbed on NaCl(100), with data points, for ML potential construction, obtained from DFT-based, few-picosecond AIMD calculations. This will allow us to study MD on ML potentials over long time scales (>100 ps) and for complex systems, originally not in the training set. CO adsorbed on NaCl(100) is a prototypical model system for studying structural and dynamical aspects of weak surface−molecule interaction. A plethora of interesting effects have been demonstrated for this system, such as vibrational energy pooling,[20,21] orientational order disorder phenomena,[22] very long vibrational lifetimes of the C−O stretch mode (~4 ms),[20] and more recently orientational isomerization after infrared (IR) vibrational excitation, wherein the CO molecule adsorbs with the O-end pointing toward the surface ("O-bound"), rather than the canonical C-end ("C-bound").[23] Several aspects related to the isomerization process, such as structure characterization, density functional theory (DFT)-based low-dimensional potential energy surfaces (PES) at low coverages,[24] vibrational spectra and picosecond scale AIMD dynamics (to be published), as well as kinetic isomerization rates[25,26] for various isotopes of the system form a part of recent investigation in our group. However, computationally expensive AIMD simulations have still restricted our dynamics studies to the picosecond time scales and because good PES are the key to simulating correct dynamics in most systems and also often the main bottleneck in studying higher dimensional dynamics for complex systems, the construction of ML

https://doi.org/10.1021/acs.jpcc.4c05765
J. Phys. Chem. C 2024, 128, 21117−21131

potentials is highly desirable. This will be achieved in this paper, and the new potentials will be utilized for selected applications.

This work is organized as follows. First, in Section 2, we give a description about data sets and models used in this work. Results and discussions follow in Section 3; in Section 3.1, we validate the constructed NN with a few examples. In Section 3.2, various first applications of the ML potentials will be demonstrated: (i) the determination of barriers for isomerization of CO/NaCl(100) from "C-bound" to "O-bound" for various coverages and phases (Section 3.2.1). (ii) The construction and discussion of multidimensional PES for submonolayer and monolayer configurations of CO on NaCl(100) (Section 3.2.2). (iii) The use of MD via ML potentials to obtain IR spectra of CO/NaCl—here long propagation times allow to resolve the Davydov splitting of CO vibrational stretch peaks of a few cm^{-1} for various "C-bound" and "O-bound" configurations, found in previous low-temperature experiments (Section 3.2.3). (iv) The use of long-time nonequilibrium MD (>100 ps) to follow the relaxation dynamics of vibrational pre-excited CO molecules on NaCl (Section 3.2.4). Finally, Section 4 concludes this work and emphasizes further possible applications of the new ML potentials.

2. METHODS AND MODELS

2.1. NN Architecture. In the NequIP approach, $E(3)$ (3D space) equivariant convolutions for interactions of geometric tensors are employed, yielding a higher information density and a good representation of the atomic environment. There is a mapping between the atomic positions $\underline{r}_i$ of the chemical species to the total potential energy, E_{pot}, and the forces acting on the atoms, $\underline{F}_i$. With M being the number of atoms, and i the atom index

$$E_{pot} = \sum_{i=1}^{M} E_{i,atomic} \tag{1}$$

and

$$\underline{F}_i = -\underline{\nabla}_i E_{pot} \tag{2}$$

One special feature of this NN is that it is energy conserving, in comparison to tensor field networks,[27] making it possible to reproduce structural and kinetic properties from molecular dynamics simulations with high fidelity.

Even though the network prediction (E_{pot}) is invariant under translations, reflections, and rotations, the convolutions that operate on the internal features associated with every atom (scalars, vectors, and higher order tensors) are equivariant functions instead of invariant ones. For instance, the feature vectors contain a direct sum of the irreducible representations of the O(3) symmetry group; for rotational equivariance, they are constrained to be products of learnable radial functions and spherical harmonics, the learnable weights being embedded into the radial function

$$S_m^{(l)}(\underline{r}_{ij}) = R(r_{ij}) \cdot Y_m^{(l)}(\hat{\underline{r}}_{ij}) \tag{3}$$

Here, $\underline{r}_{ij}$ is the relative interatomic distance vector from an atom, i to atom j, $\underline{r}_{ij} = \underline{r}_i - \underline{r}_j$, r_{ij}, and $\hat{\underline{r}}_{ij}$ are the associated interatomic distance and unit vector, respectively. $S_m^{(l)}(\underline{r}_{ij})$ is the corresponding convolution filter. $l = 0, 1, 2, \ldots$ corresponds to the rotation order and m is the representation index which takes up values $\in [-l, l]$. More details on the NN are described in ref 18. There, in particular the evaluation and precise form of the radial function $R(r_{ij})$ is explained, which involves a multilayer construction with n_{layer} layers and a cutoff radius, r_{cutoff}, both of which are (hyper) parameters of the NN (see below).

In the current work, we have used the NequIP software package,[19] fed with data points obtained from AIMD calculations for CO/NaCl(100). Details used for the NequIP input can be found in the Supporting Information, Section SA1.

2.2. Data Set Creation. We ran AIMD simulations for various starting geometries of CO/NaCl(100), as shown in the upper box of Figure 1A–F. For this, periodic supercell calculations using periodic DFT, employing the PBE exchange correlation functional[28] and a projector wave-augmented basis,[29] as implemented in the Vienna Ab initio simulation package (VASP),[30−32] were added. Grimme's D2 correction[33] was applied to account for van der Waals interactions. For all calculations, we used a $2 \times 2 \times 1$ Γ-centered k-point mesh and a kinetic energy cutoff of 700 eV. The specific surface models, some of them analogous to ref 24, are all based on $2\sqrt{2} \times 2\sqrt{2} \times 3$ supercells, with dimensions $8.017 \times 8.017 \times 30$ Å. For $T = 300$ K simulations, the supercell dimensions are $8.017 \times 8.017 \times 50$ Å instead.

These starting configurations arose from geometry optimizations, for one CO "C-bound" molecule per cell (A, coverage 1/4), and four CO "C-bound" molecules per cell (D, coverage 1) in a parallel/upright (P/U) configuration. These P/U forms are known to be the most stable at either low coverages, or at high coverages for temperatures above 35 K (see refs 24 and 34 and references therein). Also four other (one submonolayer and others monolayer) models with two CO "C-bound" molecules per cell (B, coverage 1/2) and four CO molecules per cell were adopted as starting configurations, namely, the tilted/antiparallel (T/A) "all C-bound" form (C), the most stable monolayer configuration at temperatures below 35 K,[24,34,35] as well as metastable T/A forms with one (E) or all four CO molecules (F) inverted to the "O-bound" orientations. Further details on these models and their relative stabilities will be published elsewhere (Sinha, Saalfrank, to be published).

Three phases of AIMD simulations were conducted with these starting structures, namely, an equilibration phase in the NVT ensemble using a Nosé−Hoover thermostat,[36] a production phase also in the NVT ensemble, and then a short NVE phase to remove thermostat effects. The AIMD approach as well as data set acquisition for both training and validation of the NN potentials is shown schematically in Figure 1, lower panel. Note that the AIMD simulations were performed with an aim to study vibrational spectra (IR spectra, VDOS, and VSF spectra) of orientational CO adsorbates on the NaCl surface (indicated in Figure 1) and are upcycled data for building the interatomic potentials.

We construct an initial data set from a uniform distribution of randomly selected snapshots from AIMD trajectories starting from the different equilibrium structures given in the upper panel of Figure 1. The trajectories cover configurations equilibrated at two temperatures, $T = 30$ and 300 K. We note that some $T = 300$ K trajectories encompass configurations with one or more desorbed CO(s). This so-called "Generation 0" data set, representing the above data uniformly, comprises

https://doi.org/10.1021/acs.jpcc.4c05765
J. Phys. Chem. C 2024, 128, 21117−21131

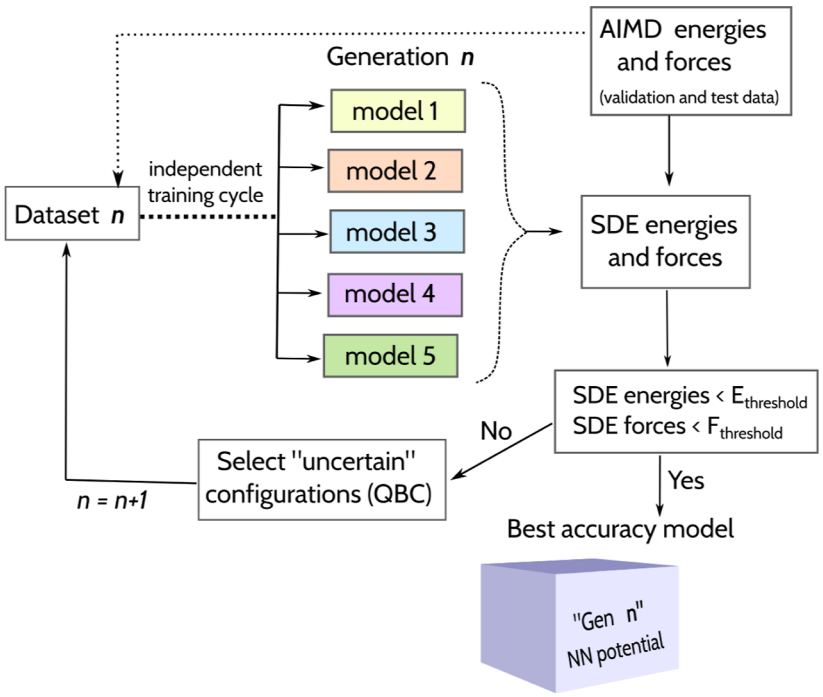


Figure 2. Training algorithm of the NN potential. For a description, see text.

nearly ∼1500 structures and makes up only 0.25% of the whole data set. Further details on data set creation and AIMD simulations can be found in Section SA2.

2.3. Hyperparameter Optimization. To optimize the NN architecture, the initial step is to optimize the so-called hyperparameters. Hyperparameters are parameters that control the NN and the learning process. In our system, we scan and select two parameters; the number of layers, n_{layer}, and the cutoff distance up to which interactions between atoms are considered, r_{cutoff}. We find that for $n_{\text{layer}} = 4$ and $r_{\text{cutoff}} = 5$ Å, the mean absolute error (MAE) on forces in the training and validation data set and the MAE on energies are quite reasonable (<0.002 eV for the energies and <0.005 eV/Å for the forces), and, therefore, we proceed with these values. Note that the MAEs were evaluated with the *Weights & Biases* package.[37]

2.4. Error Analysis Tools and Query by Committee Data Sets. Note that the (i) MAE as a first error analysis tool is given as

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N} |(y_i - \overline{y}_i)| \tag{4}$$

where y_i is the true value, $\overline{y}_i$ is the NN predicted value, and N is the number of data points.

Other quantities used to assess the quality of our NN include:

(ii) The root-mean-square deviation (RMSD)

$$\text{RMSD} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (y_i - \overline{y}_i)^2} \tag{5}$$

(iii) Standard deviation error (SDE), described below.

Mainly, RMSD and SDE will be utilized in the forthcoming sections to assess the predictive quality of our NN.

After the hyperparameter optimization step, we construct five models using the Generation 0 data set. The only variations between these models are the data set split and initial random weight seed numbers specified in the input files, which then randomly partition the data set into training (90%) and test (10%) subsets. Following the initial training cycle, we validate the NN potentials generated from our five models using subsets of the entire data set. We then evaluate the SDE in energies and forces per atom across all five models in the following way. The energies are given as

$$E_{\text{avg}} = \frac{\sum_i^{N_m} E_i}{N_m} \tag{6}$$

where E_i is the energy from the model i for a single configuration of CO/NaCl(100) and N_m is the number of models ($N_m = 5$ in our case). The SDE for energies is

$$\sigma_E = \frac{1}{M}\sqrt{\sum_i (E_{\text{avg}} - E_i)^2 / N_m} \tag{7}$$

Further, σ_E for each configuration is normalized by the number of atoms M present in each unit cell of the corresponding configuration because we have different coverages included in the total data set.

For forces, we have

$$F^a_{\text{avg}} = \frac{\sum_i^{N_m} F^a_i}{N_m} \tag{8}$$

where F^a_i is the force component along the direction a, where $a = x, y, z$ from the model i for a single configuration of CO/NaCl(100). This is computed for every atom in the simulation cell. The SDE for forces is

$$\sigma^a_F = \sqrt{\sum_i (F^a_{\text{avg}} - F^a_i)^2 / N_m},$$
$$\sigma^{\text{tot}}_F = \max\left(\sqrt{\sigma^{x2}_F + \sigma^{y2}_F + \sigma^{z2}_F}\right) \tag{9}$$

https://doi.org/10.1021/acs.jpcc.4c05765
J. Phys. Chem. C 2024, 128, 21117−21131

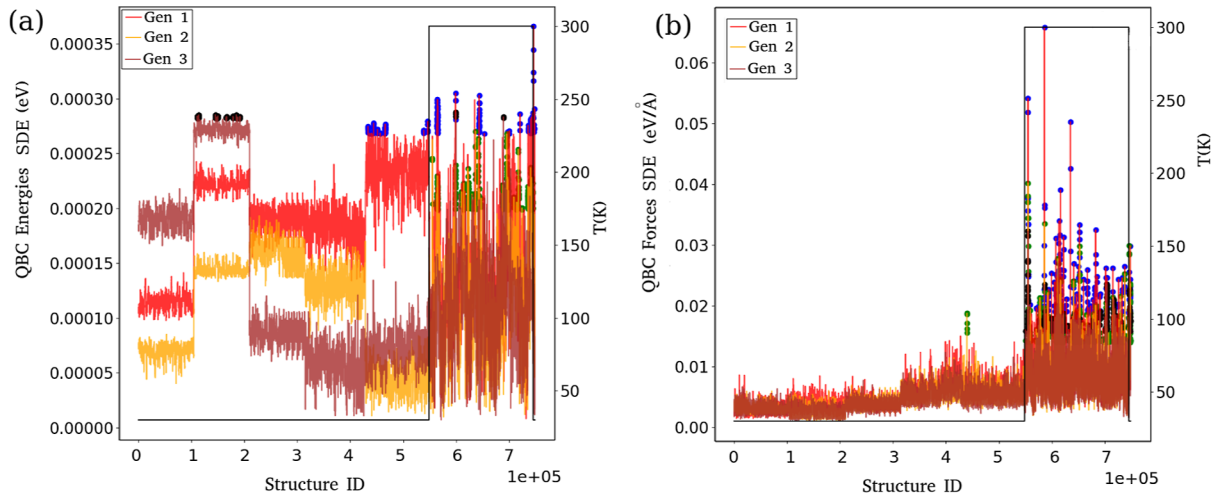


Figure 3. (a) SDE in energies, σ_E (eq 7) and (b) maximal forces, σ_F^{tot} (eq 9) in consecutive ML trainings, namely, Generation 1, 2, and 3 for the CO/NaCl(100) system, obtained with the QBC algorithm and N_m = 5 models. The different (∼0.8 million) structures, labeled by a structure ID, arise from AIMD simulations of various initial structure models A−F and different temperatures (T = 30 K and T = 300 K, right scale and black curve). Bullets (black: Gen 1, green: Gen 2, blue: Gen 3) indicate structures of maximal disagreement across the five models which are added to the training data set used for developing a Generation (n + 1) NN from a Generation n NN.

The notion "max" for σ_F^{tot} means that out of the total forces for each atom in the simulation cell, the one with the maximum value is taken as a conservative measure for force errors below.

In practice, every 5th configuration (out of ∼0.8 million), for instance, is chosen from the whole data set to obtain the SDE in energies and forces. Different NN models are trained over independent cycles, c_n, and are subsequently termed Generation n NNs. A subset of test configurations showcasing a maximal disagreement among the Generation n trained NNs are then added to the training data set used for developing a Generation (n + 1) NN. This is the so-called query by the committee (QBC) method,[38] a well-known acquisition strategy for active learning. A schematic overview of the training approach of the NN is displayed in Figure 2.

The QBC approach is adopted for a few cycles until the SDE in energy falls below 0.4 meV and that of forces per atom falls below 70 meV/Å. In some instances, we observe that after one round of training the SDE of certain configurations increases compared to the last round of training. This is due to the addition of new high uncertainty configurations to update our current data set. When incorporating new data, the distribution of configurations, representing different coverages and temperature scenarios within the data set may shift, leading to changes in the emphasis placed on different configurations during training. As a result, configurations that were previously well represented or favored may receive less emphasis in the updated data set, leading to higher SDE values for those configurations in subsequent rounds of training. Note again that each round of NN training is independent of the previous round. Here, in general, configurations near room temperatures, T = 300 K, of CO/NaCl(100) are found to be among the most disagreeable ones in the whole validation data set, followed by the low coverage configurations because the latter are somewhat under-represented proportionally in the Generation n data set. This trend improves with higher iterations of the training data set.

The (normalized) SDE in energies, σ_E (eq 7), and maximal forces, σ_F^{tot} (eq 9), of configurations in the full validation data set comprising of nearly ∼0.8 million configurations, in three consecutive generation cycles are shown in Figure 3.

2.5. Inclusion of Nonequilibrium Trajectories into the NN Model. We further aim for our ML potential to be able to describe vibrational relaxation and simulate vibrational energy pooling dynamics. We therefore augment our data set with reference configurations (∼2100) extracted from nonequilibrium NVE AIMD simulations (see Figure 1). These additional configurations result from nonequilibrium dynamics, following bond excitations; they feature initial pre-excitation of a CO stretch mode and a frustrated rotational mode, simulated independent of each other, by vibrational quanta in the respective normal mode (see below and Section SE). As a result, the entire data set (available for validation and training) increases to a size of ∼1.3 × 10^6.

Following training with the Generation 4 data set, a modest increase is found in the SDE energies (σ_E) by 5 meV; however, the SDE in forces (σ_F^{tot}) now yield a much larger error of 1 eV/Å (not shown). A few more training cycles are iterated in the framework of the QBC method described earlier, until Generation 6, where we are able to achieve σ_E and σ_F^{tot} lower than 0.6 and 150 meV/Å, respectively, as shown in Figure S3 of the Section SB. Note that the data set (∼8000 configurations) utilized for training the Gen 6 NN is still quite small, i.e., ∼0.8% of the total data set available, showcasing the efficiency of NequIP.

3. RESULTS AND DISCUSSION

3.1. Validating the NN Potentials. Deconstructing an ML potential is as important as constructing one, because trained potentials may display unexpected behavior in specific scenarios, making it crucial to evaluate the fundamental quantities they produce. At this juncture, we select the most accurate model(s) from the trained Generation 3 and Generation 6 NN potentials based on the (normalized) validation MAE on energies from the training cycles; model 3 and model 2 were chosen for Gen 3 and Gen 6 NNs, respectively, and will be henceforth used to test their accuracy in predicting properties of interest.

https://doi.org/10.1021/acs.jpcc.4c05765
J. Phys. Chem. C 2024, 128, 21117−21131

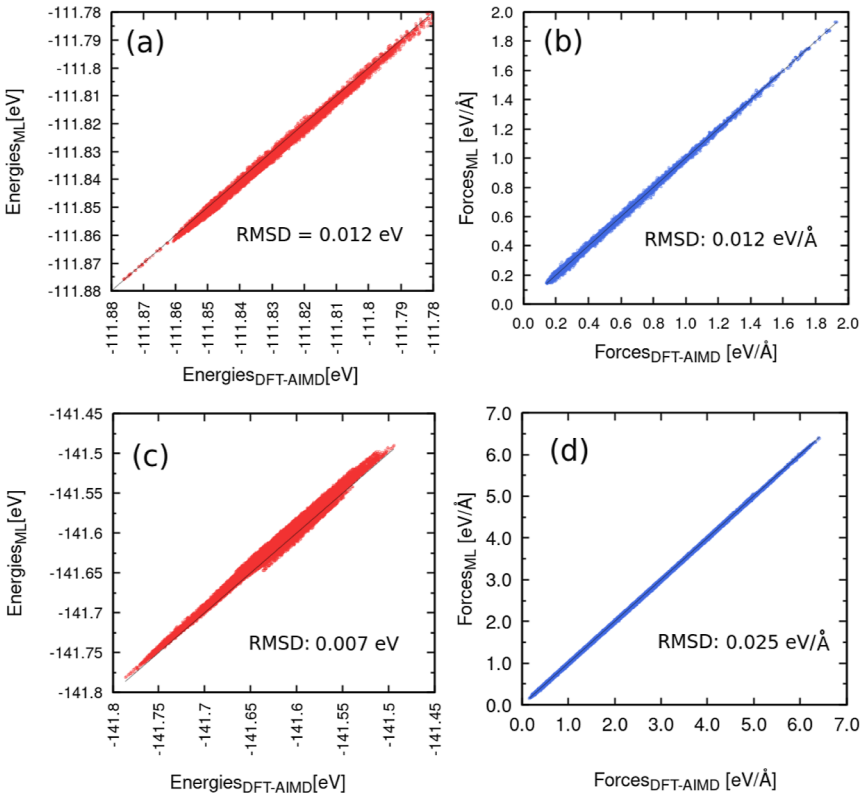


Figure 4. (a) Raw AIMD energies vs energies obtained from the ML potential (Generation 3, model 3) and (b) raw AIMD max (forces per atom) vs max (forces per atom) obtained from the ML potential (Generation 3, model 3) for a subset of the full validation data set. (c) Raw AIMD energies vs energies obtained from the ML potential (Generation 6, model 2) and (d) raw AIMD max(forces per atom) vs max(forces per atom) obtained from the ML potential (Generation 6, model 2) for a subset of the data set comprising of nonequilibrium trajectories. RMSD values w.r.t. to all configurations considered are also given.

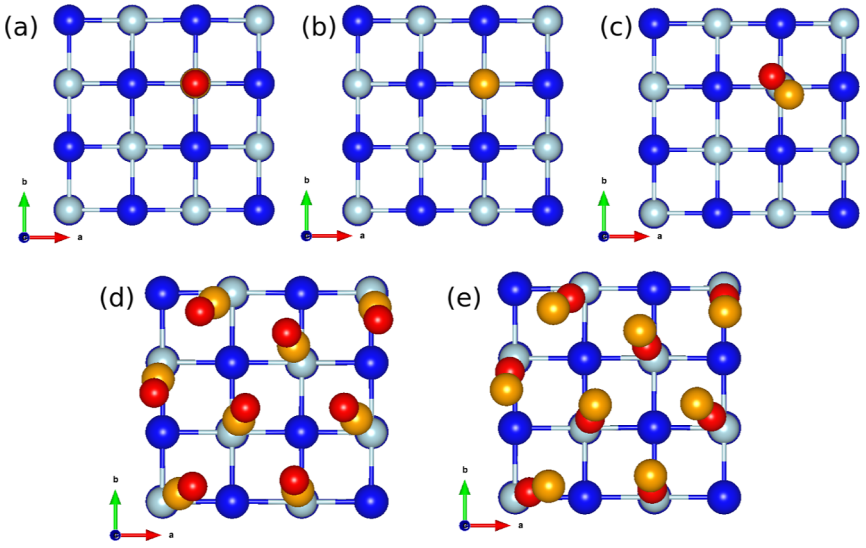


Figure 5. Optimized PBE + D2 geometries of CO/NaCl(100) configurations, modeled by a 4 × 4 × 3 cell, for (a) 1/8 ML "C-bound", (b) "O-bound", and (c) transition state configurations. (d) PBE + D2 optimized "all C-bound" T/S phase (see text) and (e) the corresponding "all O-bound" phase. All structures are utilized to compute classical isomerization barriers and energetic differences via the constructed NN potentials.

Specifically, we evaluate energies and forces for configurations of CO/NaCl(100) picked at random, equilibrated over finite temperatures, and compared them to AIMD energies and forces. In Figure 4a,b, we show a comparison plot for a subset of test configurations, where raw AIMD potential energies (for the whole simulation cell) and max(forces/atom) are plotted against NN (Gen 3) energies and max(forces/atom). We find that the energies and max(forces per atom) are distributed along the diagonal of the comparison plot, showcasing the predictive ability of our NN potential with respect to configuration energies and forces. More specifically, RMSD values, now obtained for all configurations considered in the figures (a) and (b), are 12 meV (for potential energies in the order of −100 eV) for

https://doi.org/10.1021/acs.jpcc.4c05765
J. Phys. Chem. C 2024, 128, 21117−21131

energies and 12 meV/Å for forces per atom (which extend up to 2 eV/Å). Errors on a subset of nonequilibrium configurations, computed via the Gen 6 NN (c,d) are also fairly small.

We also tested other measures for the trained ML potentials, notably, energy conservation and fluctuations in kinetic and potential energies along trajectories. In Section SB, we show one example for a 0.5 monolayer CO/NaCl(100) propagated for 50 ps in the *NVE* ensemble. Potential and kinetic energy fluctuations over this time interval are shown there in Figure S1. We see moderate usual fluctuations in the two quantities over time, with the total energy being conserved, as expected. We also report IR spectra, CO bond lengths, and angular probability distributions for the adsorbates during the time scale of the trajectory, as described in Section SA.2 and Figure S2, which further validates the ability of the NN model in predicting finite-temperature spectra and molecular motion with AIMD level accuracy.

3.2. Selected Applications. After validating our NN models, we proceeded to compute some static and dynamical properties of the CO/NaCl(100) system. These include (i) adsorption energies and barriers for reactions, (ii) potential energy landscapes for submonolayer and monolayer coverages, (iii) vibrational spectra at finite temperatures, and (iv) vibrational relaxation dynamics. Note again that we use two trained interatomic potentials, namely, Gen 3 and Gen 6, the properties derived from which will be compared occasionally against each other and with AIMD data.

3.2.1. Adsorption Energies and Isomerization Barriers. Herein, we compute adsorption energies of a single "O-bound"/"C-bound" CO on a NaCl(100) surface modeled by a $4 \times 4 \times 3$ unit cell, using both Gen 3 and Gen 6 NN potentials as well as periodic DFT (PBE + D2, using the same computational settings as before). Note that the supercell utilized here is larger compared to the ones used to train the NN potentials (see above, and Figure 1), namely, $a = b = 11.338$ Å. The PBE + D2 optimized geometries for various models considered can be found in Figure 5. Here, (a) corresponds to a low-coverage situation (coverage 1/8) with "C-bound", (b) the same with "O-bound", and (c) the transition state for an isomerization reaction between the two. For another low-temperature (meta)stable phase of monolayer CO on NaCl(100), slightly higher in energy than T/A and so far identified only theoretically by Boese and Saalfrank,[39] the "tilted/spiral" (T/S) phase, the "all C-bound" configuration is shown in panel (c), and its "all O-bound" counterpart in (e).

Using the DFT geometries as in (a)−(c), for the low-coverage case, we find that Gen 3 and Gen 6 potentials predict an isomerization barrier of 136 (Gen 3) and 138 meV (Gen 6) w.r.t. to the "C-bound" minimum and 74 meV (for both Gen 3 and Gen 6) w.r.t. the "O-bound" minimum, respectively. This is very close to the PBE + D2 barriers, which are 136 meV with respect to the "C-bound" minimum and 74 meV with respect to the "O-bound" minimum. The energy difference between the two isomers is 62 meV from Gen 3 NN and 64 meV from Gen 6 NN, compared to 64 meV from DFT + D2. Note that this energy difference was found to be 60.7 meV (without zero-point vibrational corrections for a low coverage situation using multireference-embedded cluster model) in a recent theoretical work.[40]

For the monolayer "tilted/spiral" (T/S) phase, the "all O-bound" configuration is 68 meV (Gen 3) and 70 meV (Gen 6) higher in energy per CO molecule, compared to the "all C-bound" configuration, yielding an "O-bound" → "C-bound" flipping energies of 84 and 87 meV per CO from the Gen 3 and Gen 6 NNs, respectively. PBE + D2 predicts an energy difference of 71 meV with flipping of CO requiring 89 meV per CO. Note that experimentally, energetic differences between the two isomers are not known so far to the best of our knowledge, but an activation barrier of ∼55 meV was found in a (CO overlayer) buried monolayer CO/NaCl(100) for the tunneling dominated O-bound to C-bound isomerization.[25]

The reliability of energy (and force) prediction in structures or coverages like those in Figure 5 not included in the data set used for constructing the NN, later useful for conducting dynamics studies in "arbitrary" phases and cell sizes of the system, is therefore illustrated by these examples. Additionally, it should be noted that no geometry optimizations were performed with the NN potential; instead, the same DFT-optimized geometries were used to calculate the energetics with the NN.

3.2.2. Potential Energy Surfaces. Next, we construct potential energy contours (reduced-dimensional) for low and monolayer coverages of CO/NaCl(100) along different internal degrees of freedom (DoFs) by utilizing the NN potentials. A schematic representation of the internal DoFs r, θ, and ϕ for an isolated CO molecule on the surface are shown in Figure 6a. In addition, the molecule is characterized by its center of mass (COM) coordinates, X, Y, and Z. Further, for a "CO-dimer like" configuration angles are indicated in Figure 6b, as well as the intermolecular distance, d.

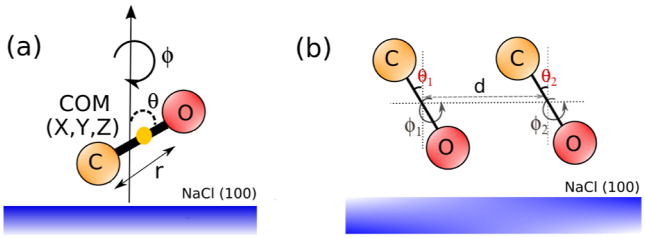


Figure 6. (a) Definition of DoF for an "adsorbate-on-surface" system (here CO on NaCl), where $X, Y, Z \in (-\infty, +\infty)$, $r \in [0, \infty)$, $\theta \in [0, \pi]$, $\phi \in [0, 2\pi]$, and (b) selected coordinates for "CO dimers" adsorbed on the surface; d is the distance between the COM of the two CO adsorbates. θ_1 and θ_2 are the respective tilt angles, and ϕ_1 and ϕ_2 are the respective azimuthal angles.

3.2.2.1. Submonolayer Configurations. First, we present 2D contour plots along different pairs of internal DoFs for a 0.25 ML coverage CO/NaCl(100) modeled by a $2\sqrt{2} \times 2\sqrt{2} \times 3$ unit cell (cf. Figure 7d,e). All PESs are computed via Gen 6 NN unless otherwise specified. So far, we reported in ref 24, two- and three-dimensional PESs along θ, Z, and r, computed via periodic PBE + D2. Other recent studies reporting PES for this system include six-dimensional PESs by Chen et al.[41] for 0.25 and 1 ML CO/NaCl(100) based on first-principles and feed-forward NN, as well as Evangelista and co-workers[40] who presented reduced PESs of CO interacting with NaCl(100) obtained from the second-order multireference perturbation theory.

Figure 7a−c shows 2D contour plots along (a) (θ, r) at $Z_{eq} = 3.33$ Å (eq = equilibrium), (b) (θ, ϕ) at $r_{eq} = 1.14$ Å, $Z_{eq} = 3.33$ Å, and (c) (Z, r) at $\theta_{O\text{-bound}} = 180°$. Other DoFs not mentioned are fixed at their optimized (equilibrium) values or values specified in the captions.

https://doi.org/10.1021/acs.jpcc.4c05765
J. Phys. Chem. C 2024, 128, 21117−21131

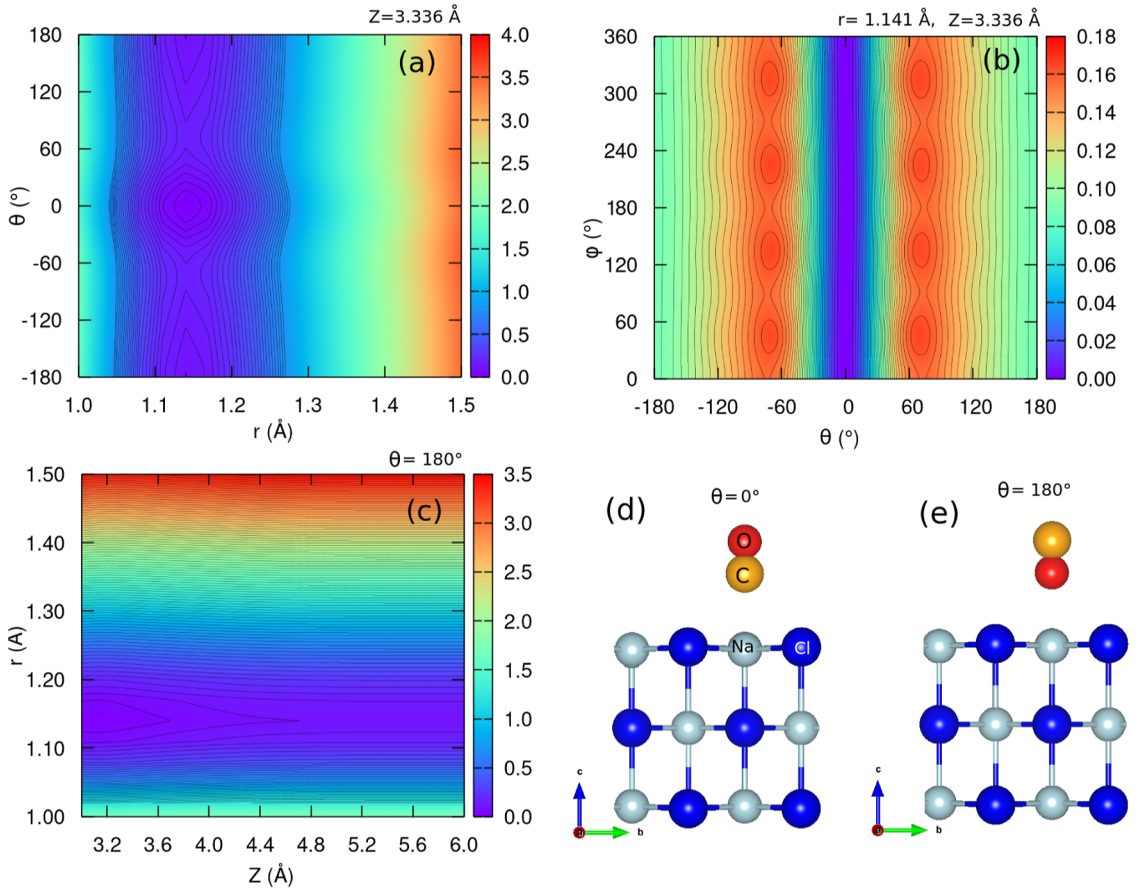


Figure 7. Two-dimensional (a) (r, θ), (b) (θ, ϕ), and (c) (Z, r) contours of the potential energy surface for an optimized PBE + D2 geometry of 0.25 ML CO/NaCl(100), modeled by a $2\sqrt{2} \times 2\sqrt{2} \times 3$ unit cell. Missing (four) coordinates in (a–c) were fixed at the values shown close to the panels, or at their equilibrium values for "C-bound" if not printed. The contour spacing is 0.03 eV for (r, θ), 0.005 eV for (θ, ϕ), and 0.04 eV for (Z, r) plots, respectively. Optimized PBE + D2 geometries of 0.25 ML CO/NaCl(100) are shown in (d) C-bound and (e) O-bound configurations. The definition of DoFs can be found in Figure 6. The Gen 6 NN (model 2) is utilized to compute the potential profiles.

In the (r, θ) 2D contour plot (a), we find the minimum located around $\theta = 0°$ corresponding to the C-bound configuration at $r = r_{eq} = 1.14$ Å. With increasing r, corresponding to vibrational excitation of the C−O bond, the energetic difference between the C-bound ($\theta \sim 0°$) and O-bound minimum ($\theta \sim 180°$) decreases (cf. Figure S4); this is concordant with previous findings.[24,41] We find that a reversal in the energetic ordering occurs beyond the outer turning point of the vibrational quantum number, $\nu_r = 20$ for CO, in our Gen 6 NN PES, which is in agreement with DFT (PBE + D2). This is, however, in contrast to ref 41, where the reversal is found to occur at a higher ν_r threshold ($\nu_r = 30$), albeit with different surface models and a feed-forward NN PES. We also note ν_r threshold values of 16 and 22, obtained from a multireference quantum embedding model[40] and quasi-classical trajectories using a 13-CO cluster model,[42] respectively. Such a prediction seems to be rather dependent upon the level of theory underlying the training data rather than on the quality of the NN itself. Note that a higher r, indicating a "stretched" CO, corresponds to a higher vibration in the CO stretch mode, ν_r. As shown in Figure S4, we see that with increasing CO bond length, both the height and width of the "C-bound → O-bound" isomerization barrier decrease, making it more likely for the C-bound adsorbates to undergo flipping at higher vibrational excitations of the CO stretch mode.

For the (θ, ϕ) contours (Figure 7b), we find overall a very low azimuthal barrier for the 0.25 ML CO/NaCl(100) system, supporting azimuthal disorder observed during the MD simulations at low temperatures ($T = 30$ K). 1D cuts of the potential energy as a function of ϕ at different tilt angles, $\theta = 30, 60, 90, 150°$ are displayed in Section SC and Figure S4b. The energy barriers along θ are found to have a maximum magnitude of 0.01 eV, for $\theta = 60°$, with only 6 and 3 meV for $\theta = 30$ and 90°, respectively. The energy profile becomes nearly flat when $\theta = 150°$.

For the (Z, r) contour plots of an "O-bound" CO on NaCl (Figure 7c), findings are concordant with previous literature.[24] In Section SC and Figure S4c, one can again find 1D cuts along Z for increasing CO bond lengths, representing vibrationally excited configurations. With increasing r, the potential well shifts to larger Z values, and the desorption barrier becomes smaller; decreasing from ~726 cm^{-1} at $r = 1.14$ Å to only 354 cm^{-1} at $r = 1.53$ Å.

We also present 2D contour plots (Z, θ), at larger CO-distances, $r = 1.596$ Å in Figure S5 in Section SC. There we note non-negligible differences between the contours generated by the two different NNs, namely, Gen 3 and Gen 6. It can be discerned, from previous literature,[41] that the Gen3 NN gives somewhat poor results for extrapolated (low) Z values for a stretched CO because the training data did not comprise of configurations with very low Z values. In fact, Gen 6 performs

https://doi.org/10.1021/acs.jpcc.4c05765
J. Phys. Chem. C 2024, 128, 21117−21131

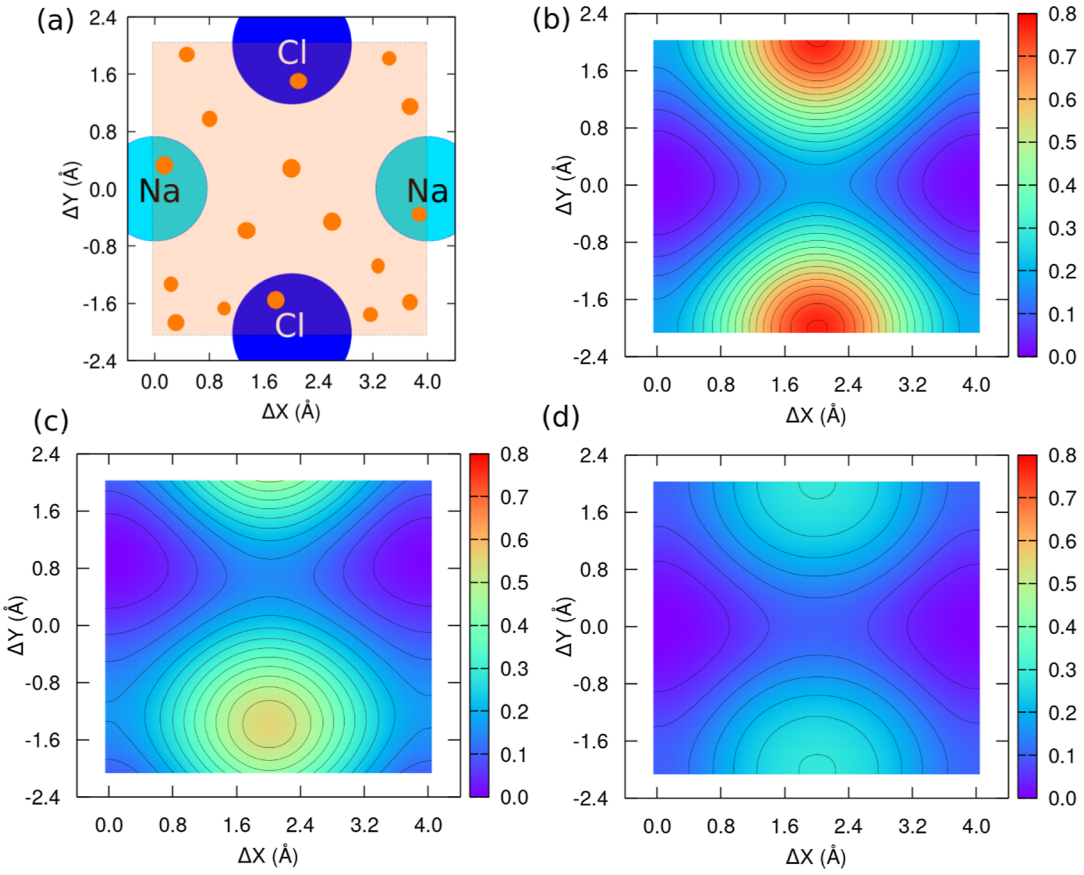


Figure 8. (a) Schematic representation of the sampled adsorption sites on the NaCl(100) surface (in orange squares the dots represent instances of (ΔX, ΔY) scanned for the adsorbed CO molecule). Contour potential plots for 0.25 ML CO/NaCl(100), modeled by a $2\sqrt{2} \times 2\sqrt{2} \times 3$ unit cell, scanned along translational COM coordinates, ΔX and ΔY, for (b) upright "C-bound" CO ($\theta = 0°$), (c) tilted "C-bound" CO ($\theta = 30°$), and (d) upright "O-bound" CO ($\theta = 180°$), at fixed CO bond length (~1.142 Å), $Z = 3.336$ Å being the distance of the monomers from the NaCl surface, and $\phi = 90°$. The energy contour spacing is 0.04 eV. The Gen 6 NN (model 2) is utilized to compute the potential profiles.

somewhat better than Gen 3 NN, due to the nonequilibrium configurations for highly excited CO molecules included in the training data, but also fails to predict (highly repulsive) energies at $Z \leq 2.7$ Å. It is important to highlight such pitfalls of the network to know and utilize it in a correct manner, only in the relevant configuration space. Additional configurations with small Z can be added to further train the NN, however, we note that all properties discussed below are computed with the Gen 6 NNs, with comparisons sometimes to Gen 3 NN.

Next, Figure 8b–d shows 2D potential contour plots along ΔX and ΔY (displacement of the COM coordinates) for an upright "C-bound" CO, a tilted ($\theta = 30°$) "C-bound" CO, and an "O-bound" CO, scanning over a 4 Å region over the NaCl surface, for coverage 1/4, as shown in Figure 7a. We find that the minima region over the Na atoms extends over 0.8 Å along both the ΔX and ΔY directions for a "C-bound" molecule, with adsorption sites over Cl being highly repulsive, as expected. The diffusion barrier along ΔX from one Na atom to another via a "hollow" site near $\Delta Y = 0$ is 0.18 eV. The potential becomes less and less repulsive between the Cl–Cl atoms, as expected, when one replaces an upright "C-bound" CO with a tilted "C-bound" CO or even an "O-bound" CO. For the latter case, the diffusion barrier is in fact lower, explaining extended diffusion relative to the "C-bound" CO in our AIMD simulations. Similar plots are obtained with both Gen 3 and Gen 6 NN potentials. Also note that the minima region along ΔY is shifted by 0.8 Å in the "tilted C-bound CO" case compared to the upright "C-bound" and "O-bound" CO case. Further detailed investigations can now be performed with the availability of NN potentials.

3.2.2.2. Monolayer Configurations. For the monolayer configurations, we study potential energy contours of two phases, the tilted/antiparallel (T/A) phase and the slightly less stable[39] tilted spiral (T/S) phase.

In the first case, we scan along the tilt angles, θ_1 and θ_2, of two of the CO adsorbates diagonal to each other, with a nearest neighbor distance of ~4 Å; the orientation of the other two neighboring molecules, in the $2\sqrt{2} \times 2\sqrt{2} \times 3$ unit cell, being either all "C-bound" or all "O-bound". The CO bond lengths, CO-surface distances, COM coordinates, and azimuthal angles are kept fixed at the respective PBE + D2 equilibrium values ($r \sim 1.14$ Å, $\phi_1 = 90°$; $\phi_2 = 270°$). Also fixed are the positions of the two other CO molecules in the unit cell, at their equilibrium geometries.

Figure 9b,d displays 2D contours of the potential along (θ_1, θ_2) for monolayer T/A CO/NaCl(100) phases, one with "O-bound" neighbors (b), and one with "C-bound" neighbors (d). For both PESs, a mirror plane symmetry occurs at the $\theta_1 = \theta_2$ axis. In case of "O-bound" neighbors (Figure 9a,b), we find, when looking at 1D cuts of the potential along, for instance, θ_2, two minima corresponding to $\theta_1 = -24°$ and $\theta_1 = -142°$, with maxima regions lying around $\theta_1 = -110°$ and $\theta_1 = 40°$, and respective classical activation barriers ~0.16 and 0.18 eV, w.r.t.

https://doi.org/10.1021/acs.jpcc.4c05765
J. Phys. Chem. C 2024, 128, 21117–21131

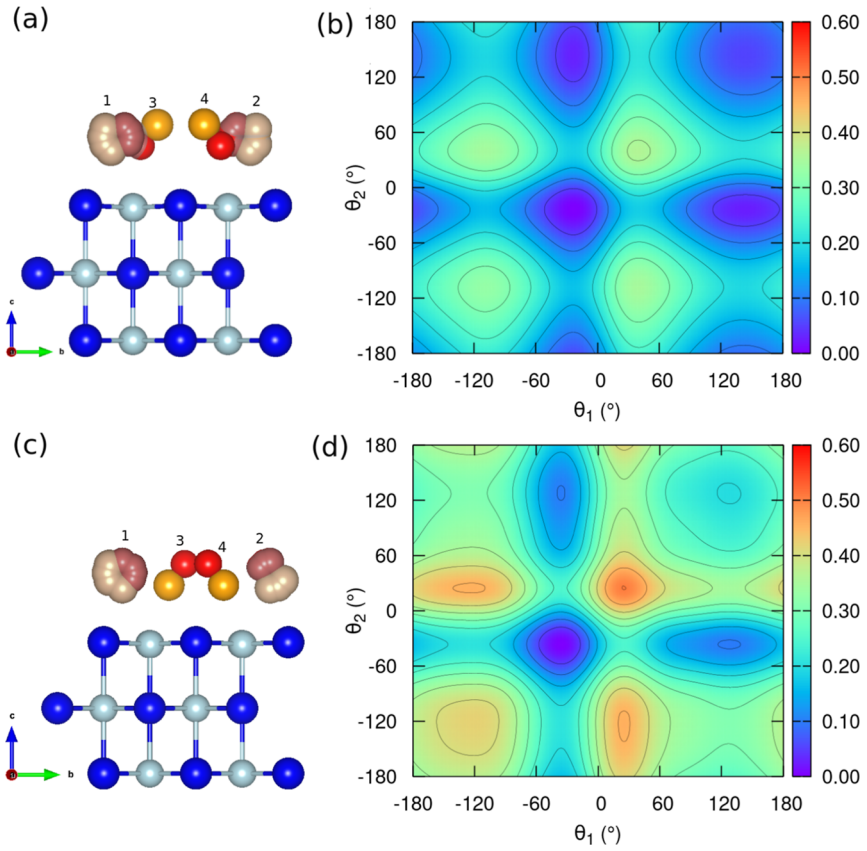


Figure 9. T/A monolayer phases of CO/NaCl(100). (a,c) Configuration for which the potential along (θ_1 and θ_2) is evaluated, with either fixed "O-bound" neighbors (a) or "C-bound" neighbors (c). For the rotating two CO molecules (indicated by blurry structures), the CO bond lengths (∼1.142 Å), distance from the surface, their (X, Y) COM coordinates, and positions of neighboring COs are fixed, all at equilibrium PBE + D2 geometries. (b,d) Two-dimensional contours of the potential energy surface with respect to θ_1 and θ_2; the energy contour spacing is 0.05 eV. The Gen 6 NN (model 2) is utilized to compute the potential profiles.

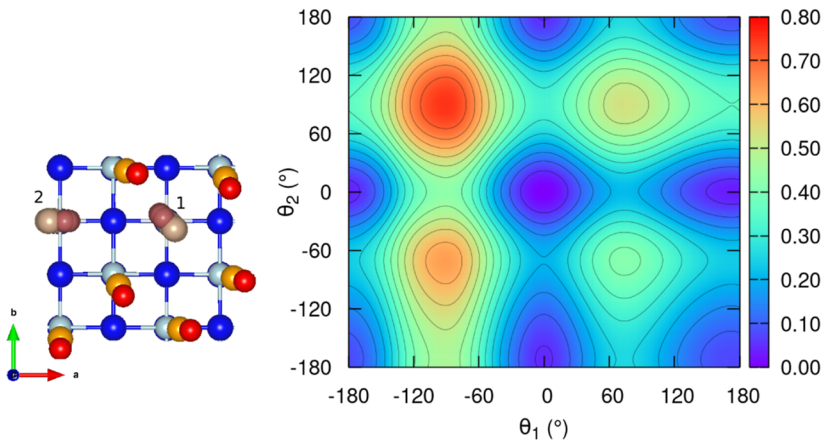


Figure 10. Two-dimensional contours of the potential energy surface with respect to θ_1 and θ_2 for a 1 ML T/S phase CO/NaCl(100) (right); the energy contour spacing is 0.05 eV. The corresponding configuration for which the potential is evaluated is given in the left panel. The two CO molecules 1 and 2 (highlighted, blurred structures) are scanned along θ_1 and θ_2, while keeping their CO bond lengths (∼1.142 Å), distance from the surface, their (X, Y) COM coordinates, and positions of neighboring COs fixed, all at equilibrium PBE + D2 geometries. The Gen 6 NN (model 2) is utilized to compute the potential profiles.

the minimum geometry at $\theta_1 = -24$ and $\theta_2 = -24°$(cf. Section SC and Figure S6a).

In the other case, for the monolayer configuration, where the neighboring, fixed molecules are all "C-bound" (Figure 9c,d), we again find two minima regions occurring around the same θ_1 region as above, with maxima regions yielding relatively higher classical activation barriers of 0.21 and 0.25 eV, respectively. These values are w.r.t. the global minimum in the respective 2D contour plot in Figure 9d. Another point of difference is the relative energy of the two minima configurations; while the two minima in the former case differ only by 0.02 eV, in the latter case, this difference increases to 0.1 eV. Inspection of the above potential energy profiles (albeit unrelaxed in the other coordinates), therefore, hints at the CO

https://doi.org/10.1021/acs.jpcc.4c05765
J. Phys. Chem. C 2024, 128, 21117−21131

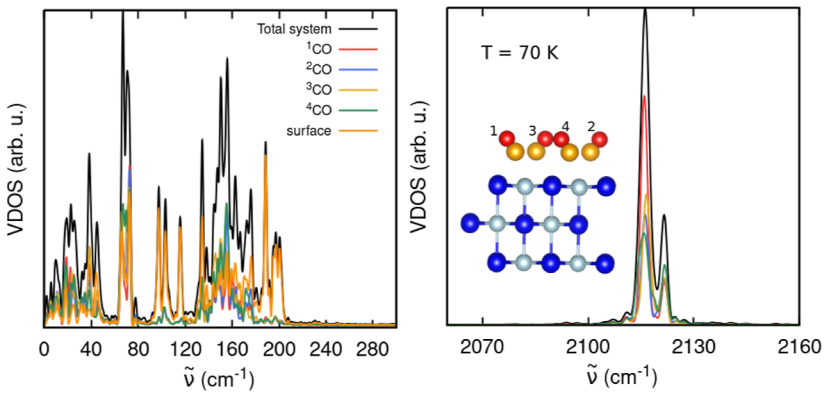


Figure 11. VDOS for one-monolayer CO/NaCl modeled by a $2\sqrt{2} \times 2\sqrt{2} \times 3$ unit cell, computed by averaging over 2 trajectories of length 50 ps each; t_{sample} = 16 ps with a spectral resolution 2.1 cm^{-1} via NVE-MD and the NN Gen 6 (model 2) potential, at T = 70 K.

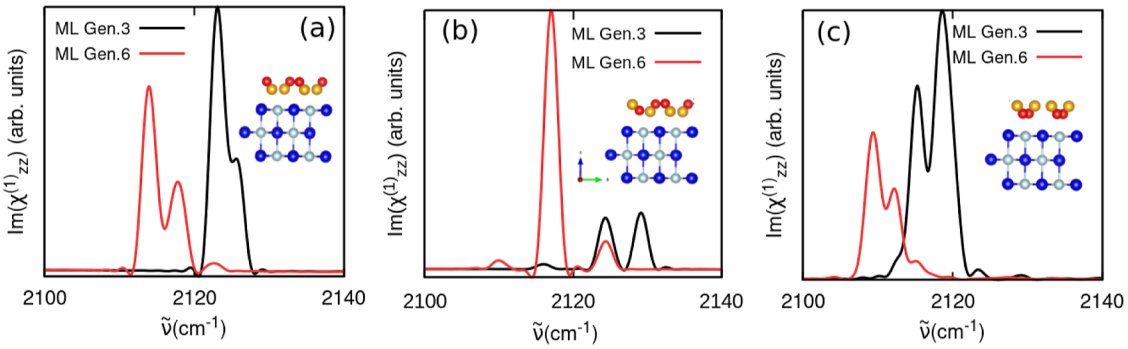


Figure 12. (a) Equilibrium z-polarized IR spectra in the CO stretch region at T = 5 K, computed with two ML potentials (Gen 3 and Gen 6) for (a) monolayer "all C-bound" configuration, (b) monolayer "one O-bound" configuration, and (c) monolayer "all O-bound" configurations. We use t_{sample} = 16 ps, giving a spectral resolution of 2.1 cm^{-1} (cf. Section SD.2). One trajectory is propagated in each case for a total of 200 ps.

orientational isomerization being more energetically favored in the presence of already flipped ("O-bound") CO adsorbates, relative to unflipped ("C-bound") ones. Note that the same minima geometries are obtained from the NN potential energy scans, as reported in Figures 7−9 compared to DFT.

We also constructed PES $V(\theta_1, \theta_2)$ via the Gen 6 NN for CO dimers adsorbed on the NaCl(100) surface, for the T/S monolayer phase, starting from a 4 × 4 × 3 unit cell "all C-bound" (see Figure 5d). Again, two CO molecules, denoted as 1 and 2 in Figure 10, left panel, were scanned along their θ angles while keeping all other coordinates (and neighboring molecules) fixed. In this case, the symmetry of the PES (Figure 10, right panel) is lifted, which is otherwise the case for the adsorption of the CO "dimer" molecules without any neighboring adsorbates (see Figure S7). We find roughly three minima regions around θ_1 = −180, 0 and 180°, each for θ_2 = −180, 0, and 180°.

Without going into further details, we thus demonstrate here how the NN PES is useful for identifying and computing crucial adsorption and isomerization energetics across different phases of the same system outside the training set.

3.2.3. Vibrational Spectra and Davydov Splittings. Next, we attempt to capture the Davydov split "in-phase" and "out-of-phase" vibrational CO-stretch modes, observed experimentally for the monolayer CO/NaCl(100) at 7 K,[43] with "C-bound" and "O-bound" CO adsorbates. We recall also earlier experimental observations of two CO stretch peaks around 2149 and 2154 cm^{-1}, respectively, with a Davydov splitting of ∼5 cm^{-1} in the T/A monolayer ("all C-bound") spectra.[34]

The vibrational response of the configurations C, E, and F in Figure 1 (upper panel) are computed from velocity−velocity autocorrelation functions (VVAFs) extracted from MD simulations using both Gen 3 and Gen 6 NN. As shown in Section SD.1 and eq 1, where also details of this well-established procedure are given (see ref 44 and references therein), we compute vibrational density of state curves VDOS(ω) as Fourier-transformed VVAFs. The VDOS curves give a vibrational frequency distribution but no reliable representation of an actual measured spectrum. Therefore, we compute first-order susceptibilities $\chi^{(1)}_{ab}(\omega)$ according to eq 3 in Section SD.1, also from VVAFs and dipole derivates of the vibration of interest (in our case, the dipole derivative μ'_{CO} of a C−O molecule on NaCl(100)). The imaginary part of $\chi^{(1)}_{ab}$ is a measure for the intensity found in an IR spectrum, with a, b = x, y, or z. In particular, we consider z-polarized IR spectra (where z is the direction perpendicular the surface), i.e., Im($\chi^{(1)}_{zz}$) in what follows. In practice, the VVAF is evaluated over a finite time interval, t_{sample}, which limits the resolution (the width) of vibrational peaks in VDOS or IR spectra (cf. Section SD.2). This is where a distinct advantage of ML potentials comes into play: once generated long propagation times and hence spectral resolutions can be afforded, even for complex systems. In particular, Davydov splitting in CO/NaCl can be resolved in this way.

We first discuss VDOS curves computed for a monolayer "all C-bound" configuration (C), modeled by a $2\sqrt{2} \times 2\sqrt{2} \times 3$ cell, at a temperature not included in the NN model training, T = 70 K. VVAFs were computed by averaging over two

https://doi.org/10.1021/acs.jpcc.4c05765
J. Phys. Chem. C 2024, 128, 21117−21131

trajectories of length 50 ps each via NVE-MD and the NN Gen 6 (model 2) potential. For details of the calculation, see Section SD.2.

In Figure 11, right panel, one can see two peaks in the CO stretch region around 2120 cm^{-1}. The red-shifted peak (~2115.3 cm^{-1}), in addition to the fundamental CO stretch transition (at ~2121.5 cm^{-1}) speculatively either stems from the vibrationally hot CO molecules at T = 70 K or is a result of Davydov splitting, experimentally observed for this system at lower temperatures.[34,43] The asymmetric doublet peaks occur also for the individual CO adsorbates, computed by molecule-resolved VDOS analysis and indicted as ^{1}CO−^{4}CO and by different colors in the figure. Note that further evaluation of VDOS based on internal DoFs (not shown) instead of Cartesian components rules out any rovibrational origin of the doublet peaks in the total VDOS.

In Figure 11, left panel, the highly structured VDOS in the low-frequency regime up to 300 cm^{-1} is also shown. As can be seen, while the spectrum is dominated by the motion of surface atoms (phonons), the low-frequency modes of the CO molecules (notably frustrated translations) also make a contribution in some parts of the spectrum.

We now discuss results for the IR spectrum in the CO stretch region, measured by $\mathrm{Im}(\chi^{(1)}_{zz})$, using the Gen 3 and Gen 6 NN models, at a very low temperature, T = 5 K. We first discuss the results from the Gen 3 NN. Note that the dipole derivative, μ'_{CO} is simply taken as −1 e in this case. Figure 12 reports the corresponding spectral profiles with the configurations under investigation, models C (a), E (b), and F (c). For an "all C-bound" configuration (model C) at 5 K, two CO stretch peaks are found centered at 2123 and 2126 cm^{-1}, while for the "all O-bound" configuration (model F), the two peaks are centered around 2115 and 2118 cm^{-1}. In the case of one "O-bound" configuration (model E), the "O-bound" CO stretch is found to peak at 2110 cm^{-1}, while the "C-bound" peak maxima from the other adsorbates are found at 2117 and 2124 cm^{-1}. These correspond to an excitonic splitting of ~3, 3, and 7 cm^{-1} for the above "C-bound" ("O-bound") molecules, respectively, from Gen 3 NN, and can be compared to 4.5, 3, and 8 cm^{-1} computed via Gen 6 NN and 5, 1, and 7 cm^{-1} computed via normal-mode analysis (NMA). Additionally, an overall red shift of ~9, 5, and 6.7 cm^{-1} is introduced in the peaks obtained from Gen 6 NN simulated z-polarized IR spectra compared to that of Gen 3 NN simulated spectra. Table S1 in Supporting Information, Section SD.3 lists the harmonic as well as IR frequencies generated from DFT and both NN models. Setting aside minor deviations, the NN potentials, therefore, achieve a very good agreement in predicting IR spectra of the system at extrapolated temperatures, such as 5 K, in this case.

In passing, we note that we have also calculated VSF spectra for CO/NaCl(100) using the new NN PESs (cf. Figure 1), which will be published elsewhere (Sinha, Saalfrank, to be published).

3.2.4. Nonequilibrium Relaxation Dynamics. In order to showcase the applicability of the NN potential in the context of vibrational dynamics, we highlight here a case study of vibrational relaxation of pre-excited CO adsorbates on the NaCl(100) surface using the Gen 6 NN potential. Specifically, we study vibrational relaxation in the monolayer T/A "all C-bound" (model C), for all (four) CO molecules pre-excited in their CO stretch by one vibrational quantum each (~4 × 0.26 eV), amounting to a total of 1.04 eV excess energy pumped into the system. Details on the pre-excitation approach can be found in Supporting Information, Section SE.

A single trajectory of 150 ps starting from the pre-excited CO monolayer is run, and a subsequent redistribution of vibrational energies onto the different subsystems is followed, via instantaneous kinetic energy analysis. In Figure 13,

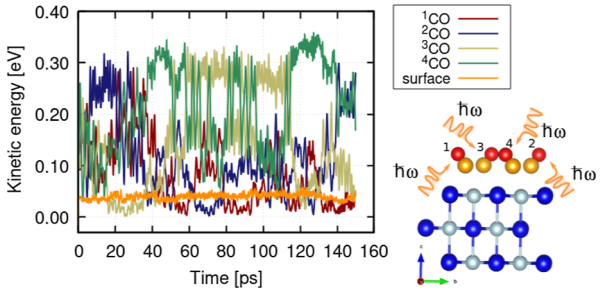


Figure 13. Averaged trendlines of instantaneous kinetic energies decomposed into subsystems, for a nonequilibrium trajectory of 150 ps, after pre-excitation of all CO adsorbates by one vibrational quantum in the CO stretch mode.

averaged trendlines[a] of instantaneous kinetic energies of the total system ($E_{kin}(t) = \frac{1}{2}\sum_i^{N_a} m_i v_i^2(t)$, where N_a = total number of movable atoms in the model system), decomposed into adsorbate subsystems and the NaCl surface are plotted. Note that the top layer of the surface was allowed to move, keeping the remaining two layers frozen during the simulations.

One can discern an energy redistribution occurring among the adsorbates within the 150 ps time scale with nearly negligible energy transfer to the surface, which is expected owing to the long vibrational lifetime[20] of CO adsorbates (approaching the so-called Sommerfeld limit[21]) on the NaCl(100) surface.

A deeper analysis of vibrational relaxation after CO stretch excitation will follow elsewhere (Sinha, Saalfrank, to be published). We also studied cases where a single "O-bound" CO molecule is highly excited in its bending mode, which led to relaxation, roaming, and flipping of the pre-excited molecule to "C-bound" orientation. Details of this will also be published in a forthcoming publication. Note that the roaming trajectories in our simulations are reminiscent of those reported in a previous study by Nandi et al.[42]

4. SUMMARY AND CONCLUSIONS

In this work, we constructed multidimensional machine-learned potentials of PBE + D2 accuracy trained on AIMD simulation data using an $E(3)$ equivariant NequIP NN. NequIP is a data-efficient ML approach to construct multidimensional PES for complex systems, allowing for long-time MD and, potentially, the use of more accurate electronic structure methods beyond DFT, which was the workhorse in this work.

The NN potentials demonstrated very good predictive power and accuracy in adsorption energetics and barriers, multidimensional potentials, vibrational spectra, and dynamics across different adsorbate orientations and coverages for CO on NaCl(100). They showcased generalization to different phases as well as temperatures in the extrapolated region not included in the training procedure.

https://doi.org/10.1021/acs.jpcc.4c05765
J. Phys. Chem. C 2024, 128, 21117−21131

We validated and tested two NN models based on configuration energies and forces across different coverages and phases of the system at different temperatures, called Gen 3 and Gen 6. While the former made use of equilibrium AIMD simulations (NVT and NVE) for CO/NaCl(100), the latter was trained on nonequilibrium configurations encompassing desorption and highly excited adsorbates in addition. In general, Gen 6 can be considered more accurate than Gen 3.

Applications pertaining to potential energy profiles, identifying isomerization barriers ("C-bound" ↔ "O-bound") across different phases of CO/NaCl(100), IR spectra, and nonequilibrium relaxation dynamics were showcased. Of special mention is the Davydov splitting reproduced theoretically in the T = 5 K IR spectra and insights into real-time relaxation dynamics of excited CO molecules at NaCl(100). Details of vibrational spectra and relaxation dynamics will be exploited in forthcoming publications (Sinha and Saalfrank, to be published). Other possible applications of the new potentials include:

- Investigating longer time scale (ns) processes like vibrational energy pooling in CO/NaCl, i.e., CO(ν_r = 1) + CO(ν_r = 1) → CO(ν_r = 2) + CO(ν_r = 0) and higher excitations, also possibly spread over larger distances, which can now be modeled via much larger supercells. Its role in the recently observed isotope-dependent isomerization of the more pragmatic "C-bound" CO to "O-bound" CO on the NaCl(100) system can also be explored via long time scale relaxation dynamics, leading to identification of active vibrational energy-transfer pathways. Note a previous extensive study[45] on the pooling dynamics in gas-phase CO molecules, using a NN fitted full-dimensional PES.
- Looking at the order−disorder (phase) transition occurring in this system, e.g., from the monolayer (2 × 1) T/A phase to a p(1 × 1) P/U phase.
- Extending kinetic isomerization rate models for "O-bound" → "C-bound" reactions[25] based on "system-bath" models to higher dimensions, i.e., including more system DoFs and higher order system-bath couplings by increasing the dimensionality of the potential energy surface.

Some of this work is in progress in our laboratory.

■ ASSOCIATED CONTENT

Data Availability Statement

The neural network (Gen 3, Gen 6) models are available in Zenodo under 10.5281/zenodo.13371065. They can be used in association with the NequIP code, and called upon as ASE calculator objects[46] or in integration with the LAMMPS software.[47] Other AIMD data and data related to NN-MD simulations can be obtained upon reasonable request from the author.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpcc.4c05765.

Details on the NN architecture and learning parameters; details of AIMD simulations and training data set generation; details regarding validation of the NN potentials; figures of potential contours constructed with the NN; details on computation of vibrational spectra (VDOS and IR); comments on spectral resolution of the corresponding spectra with comparison of vibrational frequencies obtained from NMA and NN(MD) simulations; and details on nonequilibrium NN(MD) simulations (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Peter Saalfrank − *Theoretical Chemistry, Institute of Chemistry, University of Potsdam, 14476 Potsdam, Germany;* orcid.org/0000-0002-5988-5945; Phone: +49 (0)331−977−5232; Email: peter.saalfrank@uni-potsdam.de

Authors

Shreya Sinha − *Theoretical Chemistry, Institute of Chemistry, University of Potsdam, 14476 Potsdam, Germany;* orcid.org/0000-0003-2726-1537

Bruno Mladineo − *Institut Ruđer Bošković, 10000 Zagreb, Croatia*

Ivor Lončarić − *Institut Ruđer Bošković, 10000 Zagreb, Croatia;* orcid.org/0000-0002-5554-4641

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpcc.4c05765

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

S.S., B.M., I.L., and P.S. thank DAAD and the Croatian Ministry of Science, Education, and Youth for the scientific exchange funding which helped initiate this project. Fruitful discussions with Giacomo Melani (Pisa) on spectra calculations are also acknowledged. On the German side, this work was funded through Deutsche Forschungsgemeinschaft (DFG), through project Sa 548/18-1. I.L. and B.M. acknowledge support by Croatian Science Foundation under the project UIP-2020-02-5675. We also thank the North-German Supercomputing Alliance (HLRN) for providing HPC resources. S.S. also thanks the Max-Planck Graduate school "IMPRS-EPPC" (Berlin) for support. We would like to specially thank Prof. Dr. Alec M. Wodtke for his pioneering and wonderful contributions to surface science and particularly his work on the pragmatic CO/NaCl(100) system which motivated us to pursue theoretical investigations on this system. PS and SS are grateful to him for many fruitful discussions and successful collaborations.

■ ADDITIONAL NOTE

[a]A moving average trendline averages a specific number of data points over equal data interval periods (here period = 0.25 fs), and these averages are plotted as points on the trendline; this smoothens out data fluctuations and offers better visualization.

■ REFERENCES

(1) Mishin, Y. Machine-learning interatomic potentials for materials science. *Acta Mater.* **2021**, *214*, 116980.

(2) Forni, A.; Wiesenekker, G.; Baerends, E. J.; Tantardini, G. F. A dynamical study of the chemisorption of molecular hydrogen on the Cu(111) surface. *J. Phys.: Condens. Matter* **1995**, *7*, 7195−7207.

(3) Forni, A.; Tantardini, G. F. A simulation study of the chemisorption dynamics of molecular hydrogen on the Ni(111) surface. *Surf. Sci.* **1996**, *352*−*354*, 142−147.

(4) Busnengo, H. F.; Salin, A.; Dong, W. Representation of the 6D potential energy surface for a diatomic molecule near a solid surface. *J. Chem. Phys.* **2000**, *112*, 7641−7651.
(5) Kresse, G. Dissociation and sticking of H_2 on the Ni(111), (100), and (110) substrate. *Phys. Rev. B* **2000**, *62*, 8295−8305.
(6) Zaverkin, V.; Holzmüller, D.; Steinwart, I.; Kästner, J. Fast and Sample-Efficient Interatomic Neural Network Potentials for Molecules and Materials Based on Gaussian Moments. *J. Chem. Theory Comput.* **2021**, *17*, 6658−6670.
(7) Jiang, B.; Guo, H. Dynamics of Water Dissociative Chemisorption on Ni(111): Effects of Impact Sites and Incident Angles. *Phys. Rev. Lett.* **2015**, *114*, 166101.
(8) Lindner, S.; Lončarić, I.; Vrček, L.; Alducin, M.; Juaristi, J. I.; Saalfrank, P. Femtosecond Laser-Induced Desorption of Hydrogen Molecules from Ru(0001): A Systematic Study Based on Machine-Learned Potentials. *J. Phys. Chem. C* **2023**, *127*, 14756−14764.
(9) Chen, J.; Zhou, X.; Zhang, Y.; Jiang, B. Vibrational control of selective bond cleavage in dissociative chemisorption of methanol on Cu(111). *Nat. Commun.* **2018**, *9*, 4039.
(10) Kolb, B.; Luo, X.; Zhou, X.; Jiang, B.; Guo, H. High-Dimensional Atomistic Neural Network Potentials for Molecule−Surface Interactions: HCl Scattering from Au(111). *J. Phys. Chem. Lett.* **2017**, *8*, 666−672.
(11) Shakouri, K.; Behler, J.; Meyer, J.; Kroes, G.-J. Accurate Neural Network Description of Surface Phonons in Reactive Gas−Surface Dynamics: N2 + Ru(0001). *J. Phys. Chem. Lett.* **2017**, *8*, 2131−2136.
(12) Meng, G.; Gardner, J.; Hertl, N.; Dou, W.; Maurer, R. J.; Jiang, B. First-Principles Nonadiabatic Dynamics of Molecules at Metal Surfaces with Vibrationally Coupled Electron Transfer. *Phys. Rev. Lett.* **2024**, *133*, 036203.
(13) Serrano Jiménez, A.; Sánchez Muzas, A. P.; Zhang, Y.; Ovčar, J.; Jiang, B.; Lončarić, I.; Juaristi, J. I.; Alducin, M. Photoinduced Desorption Dynamics of CO from Pd(111): A Neural Network Approach. *J. Chem. Theory Comput.* **2021**, *17*, 4648−4659.
(14) Žugec, I.; Tetenoire, A.; Muzas, A. S.; Zhang, Y.; Jiang, B.; Alducin, M.; Juaristi, J. I. Understanding the Photoinduced Desorption and Oxidation of CO on Ru(0001) Using a Neural Network Potential Energy Surface. *JACS Au* **2024**, *4*, 1997−2004.
(15) Zhang, Y.; Hu, C.; Jiang, B. Embedded Atom Neural Network Potentials: Efficient and Accurate Machine Learning with a Physically Inspired Representation. *J. Phys. Chem. Lett.* **2019**, *10*, 4962−4967.
(16) Fu, B.; Zhang, D. H. Accurate fundamental invariant-neural network representation of ab initio potential energy surfaces. *Nat. Sci. Rev.* **2023**, *10*, nwad321.
(17) Kondor, R.; Trivedi, S. On the Generalization of Equivariance and Convolution in Neural Networks to the Action of Compact Groups. *Proceedings of Machine Learning Research*, 2018; Vol. *80*; pp 2747−2755.
(18) Batzner, S.; Musaelian, A.; Sun, L.; Geiger, M.; Mailoa, J. P.; Kornbluth, M.; Molinari, N.; Smidt, T. E.; Kozinsky, B. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nat. Commun.* **2022**, *13*, 2453.
(19) Batzner, S.; Musaelian, A.; Sun, L.; Geiger, M.; Mailoa, J. P.; Kornbluth, M.; Molinari, N.; Smidt, T. E.; Kozinsky, B. *The NeQuIP package*, 2022. An open source package which can be obtained from. https://github.com/mir-group/nequip.
(20) Chang, H. C.; Ewing, G. E. Infrared fluorescence from a monolayer of CO on NaCl(100). *Phys. Rev. Lett.* **1990**, *65*, 2125−2128.
(21) Chen, L.; Lau, J. A.; Schwarzer, D.; Meyer, J.; Verma, V. B.; Wodtke, A. M. The Sommerfeld ground-wave limit for a molecule adsorbed at a surface. *Science* **2019**, *363*, 158−161.
(22) Heidberg, J.; Kampshoff, E.; Kühnemuth, R.; Schönekäs, O. Surface diffusion in the 2D-phase transition of CO_2 adsorbed on NaCl(100). *J. Electron Spectrosc. Relat. Phenom.* **1993**, *64*−*65*, 803−812.
(23) Lau, J. A.; Choudhury, A.; Li, C.; Schwarzer, D.; Verma, V. B.; Wodtke, A. M. Observation of an isomerizing double-well quantum system in the condensed phase. *Science* **2020**, *367*, 175−178.
(24) Sinha, S.; Saalfrank, P. Inverted" CO molecules on NaCl(100): a quantum mechanical study. *Phys. Chem. Chem. Phys.* **2021**, *23*, 7860−7874.
(25) Choudhury, A.; DeVine, J. A.; Sinha, S.; Lau, J. A.; Kandratsenka, A.; Schwarzer, D.; Saalfrank, P.; Wodtke, A. M. Condensed-phase isomerization through tunnelling gateways. *Nature* **2022**, *612*, 691−695.
(26) Choudhury, A.; Sinha, S.; Harlander, D.; DeVine, J.; Kandratsenka, A.; Saalfrank, P.; Schwarzer, D.; Wodtke, A. M. Manipulating tunnelling gateways in condensed phase isomerization. *Nat. Sci.* **2023**, *3*, No. e20230006.
(27) Thomas, N.; Smidt, T.; Kearnes, S.; Yang, L.; Li, L.; Kohlhoff, K.; Riley, P. Tensor field networks: Rotation- and translation-equivariant neural networks for 3D point clouds. *arXiv* **2018**, arXiv:1802.08219.
(28) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868.
(29) Blöchl, P. E. Projector Augmented-Wave Method. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1994**, *50*, 17953−17979.
(30) Kresse, G.; Hafner, J. Ab Initio Molecular-Dynamics Simulation of the Liquid-Metal−Amorphous-Semiconductor Transition in Germanium. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1994**, *49*, 14251−14269.
(31) Kresse, G.; Hafner, J. Ab Initio Molecular Dynamics for Liquid Metals. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1993**, *47*, 558−561.
(32) Kresse, G.; Furthmüller, J. Efficient Iterative Schemes for Ab Initio Total-Energy Calculations using a Plane-Wave Basis Set. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1996**, *54*, 11169−11186.
(33) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.
(34) Heidberg, J.; Kampshoff, E.; Suhren, M. Correlation field, structure, and phase transition in the monolayer CO adsorbed on NaCl(100) as revealed from polarization Fourier-transform infrared spectroscopy. *J. Chem. Phys.* **1991**, *95*, 9408−9411.
(35) Vogt, J.; Vogt, B. The structure of carbon monoxide adsorbed on the NaCl(100) surface—A combined LEED and DFT-D/vdW-DF study. *J. Chem. Phys.* **2014**, *141*, 214708.
(36) Nosé, S. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* **1984**, *81*, 511−519.
(37) Biewald, L. Experiment Tracking with Weights and Biases, 2020.https://www.wandb.com/,Softwareavailablefromwandb.com (accessed September 2, 2022).
(38) Seung, H. S.; Opper, M.; Sompolinsky, H. Query by committee. *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, 1992, pp 287−294.
(39) Boese, A. D.; Saalfrank, P. CO Molecules on a NaCl(100) Surface: Structures, Energetics, and Vibrational Davydov Splittings at Various Coverages. *J. Phys. Chem. C* **2016**, *120*, 12637−12653.
(40) He, N.; Huang, M.; Evangelista, F. A. CO Inversion on a NaCl(100) Surface: A Multireference Quantum Embedding Study. *J. Phys. Chem. A* **2023**, *127*, 1975−1987.
(41) Chen, J.; Hariharan, S.; Meyer, J.; Guo, H. Potential Energy Landscape of CO Adsorbates on NaCl(100) and Implications in Isomerization of Vibrationally Excited CO. *J. Phys. Chem. C* **2020**, *124*, 19146−19156.
(42) Nandi, A.; Zhang, P.; Chen, J.; Guo, H.; Bowman, J. M. Quasiclassical simulations based on cluster models reveal vibration-facilitated roaming in the isomerization of CO adsorbed on NaCl. *Nat. Chem.* **2021**, *13*, 249−254.
(43) Lau, J. A.; Schönemann, A. M.; Schwarzer, D.; Wodtke, A. M. The coverage dependence of the infrared absorption of CO adsorbed to NaCl(100). *J. Chem. Phys.* **2020**, *153*, 154703.
(44) Melani, G.; Nagata, Y.; Campen, R. K.; Saalfrank, P. Vibrational spectra of dissociatively adsorbed D_2O on Al-terminated α-Al_2O_3(0001) surfaces from ab initio molecular dynamics. *J. Chem. Phys.* **2019**, *150*, 244701.

https://doi.org/10.1021/acs.jpcc.4c05765
J. Phys. Chem. C 2024, 128, 21117−21131

(45) Chen, J.; Li, J.; Bowman, J. M.; Guo, H. Energy transfer between vibrationally excited carbon monoxide based on a highly accurate six-dimensional potential energy surface. *J. Chem. Phys.* **2020**, *153*, 054310.
(46) Hjorth Larsen, A.; et al. The atomic simulation environment—a Python library for working with atoms. *J. Phys.: Condens. Matter* **2017**, *29*, 273002.
(47) Thompson, A. P.; Aktulga, H. M.; Berger, R.; Bolintineanu, D. S.; Brown, W. M.; Crozier, P. S.; in't Veld, P. J.; Kohlmeyer, A.; Moore, S. G.; Nguyen, T. D.; et al. LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Comput. Phys. Commun.* **2022**, *271*, 108171.

https://doi.org/10.1021/acs.jpcc.4c05765
J. Phys. Chem. C 2024, 128, 21117−21131