

Electronic Supplementary Information to: A first-principles machine-learning force field for heterogeneous ice nucleation on microcline feldspar

Pablo M. Piaggi,^a Annabella Selloni,^a Athanassios Z. Panagiotopoulos,^b Roberto Car^{a,c}, and Pablo G. Debenedetti^b

Contents

1	Electronic structure calculations	1
2	Machine-learning model definition and training	1
3	Molecular dynamics simulations	2
4	Enhanced sampling simulations	2
5	Initial training data	3
6	Active learning	4
7	Generalization errors	4
8	Calculation of surface energies	6
9	Formation of ice cluster at the microcline/water interface	6

1 Electronic structure calculations

Plane-wave DFT calculations were performed using the QUANTUM ESPRESSO suite for electronic structure calculations v6.4.1^{1,2} The SCAN exchange and correlation functional^{3,4} was evaluated with the LIBXC 4.3.4 library⁵ We employed norm-conserving, scalar-relativistic pseudopotentials⁶ for K, Al, Si, O, H parametrized using the PBE⁷ functional with 9, 11, 4, 6, and 1 valence electrons, respectively. Kinetic energy cutoffs of 110 and 440 Ry were used for the wave functions and the charge density. k -point convergence was evaluated for each configuration with energy convergence threshold 1 meV/atom. The convergence absolute error for the self-consistent procedure was set to 10^{-6} Ry. For energy minimization with respect to the box or atomic coordinates, the criterion for ending the minimization was a change in energy smaller than 10^{-4} Ry in subsequent iterations. For surface configurations, we employed a dipole correction based on a saw-like potential in the direction perpendicular to the interface⁸. The mixing factor

for self consistency was set to 0.5. All other parameters were set to their default values in QUANTUM ESPRESSO.

2 Machine-learning model definition and training

Machine-learning interatomic potentials were constructed using the Deep Potential methodology developed by Zhang *et al.*⁹ as implemented in the DEEPM-D-KIT v2.0.0.b2¹⁰. Below, we describe the smooth version of this framework¹¹, which is based on deep neural networks and descriptors learned on the fly during the training process.

The total energy E of a configuration of N atoms with atomic coordinates \mathbf{R} is written as a sum over per-atom energies E_i , i.e.,

$$E(\mathbf{R}) = \sum_{i=1}^N E_i = \sum_{i=1}^N E^{\alpha_i}(\mathbf{R}_i) \quad (1)$$

where \mathbf{R}_i are the relative atomic coordinates of N_i neighbors in an environment with cutoff r_c around atom i , α_i is the atom type of atom i , and E^{α_i} is an energy function for atoms of the chemical species α_i . In order to preserve the natural symmetries of the problem, i.e., rotation and permutation of atoms of the same type, we define a vector of descriptors \mathbf{D}_i for atom i . Then, the energy of a configuration can be written as,

$$E(\mathbf{R}) = \sum_{i=1}^N E^{\alpha_i}(\mathbf{D}_i). \quad (2)$$

The starting point for the definition of the descriptors \mathbf{D}_i is a continuous and differentiable switching function,

$$s(r) = \begin{cases} \frac{1}{r}, & r < r_s \\ \frac{1}{r} \{u^3(-6u^2 + 15u - 10) + 1\}, & r_s \leq r < r_c \\ 0, & r \geq r_c \end{cases} \quad (3)$$

where $u = (r - r_s)/(r_c - r_s)$, and r_s and r_c are smooth and hard cutoffs, respectively. Next, we construct a matrix $\tilde{\mathbf{R}}_i \in \mathbb{R}^{N_i \times 4}$ of generalized coordinates with rows,

$$(\tilde{\mathbf{R}}_i)_j = \left[s(r_{ij}), \frac{s(r_{ij})x_{ij}}{r_{ij}}, \frac{s(r_{ij})y_{ij}}{r_{ij}}, \frac{s(r_{ij})z_{ij}}{r_{ij}} \right] \quad (4)$$

where (x_{ij}, y_{ij}, z_{ij}) is the distance vector from atom j to atom i , and r_{ij} is the norm of such distance. Furthermore, we define an embedding matrix $\mathbf{G}^i \in \mathbb{R}^{N_i \times M_1}$ with row j

^a Department of Chemistry, Princeton University, Princeton, NJ 08544, USA

^b Department of Chemical and Biological Engineering, Princeton University, Princeton, NJ 08544, USA

^c Department of Physics, Princeton University, Princeton, NJ 08544, USA

given by,

$$(\mathbf{G}^i)_{j,*} = g^{\alpha_i, \alpha_j}(s(r_{ij})) \quad (5)$$

where g^{α_i, α_j} is a function that maps a scalar into M_1 outputs, and is different for each pair of chemical species α_i and α_j . We also define a secondary embedding matrix $\mathbf{G}^i \in \mathbb{R}^{N_i \times M_2}$ with the first $M_2 < M_1$ columns of \mathbf{G}^i . With these ingredients, we now write the descriptor matrix $\mathbf{D}_i \in \mathbb{R}^{M_1 \times M_2}$ as,

$$\mathbf{D}_i = (\mathbf{G}^i)^T \tilde{\mathbf{R}}_i (\tilde{\mathbf{R}}_i)^T \mathbf{G}^i \quad (6)$$

which is subsequently flattened into a vector of $M_1 \times M_2$ elements and is used as input in Eq. (2). In our simulations, we used a model for five chemical species $\boldsymbol{\alpha} = (\text{K}, \text{Al}, \text{Si}, \text{O}, \text{H})$, E^{α_i} in Eq. (2) was represented by a neural network with three layers and 120 neurons per layer, and g^{α_i, α_j} in Eq. (5) was represented by a three-layer neural network with sizes 25, 50 and 100, respectively. Other parameters of our model are $M_1 = 100$, $M_2 = 16$, $r_s = 3 \text{ \AA}$, and $r_c = 6 \text{ \AA}$.

The parameters in the neural networks E^{α_i} and g^{α_i, α_j} described above are determined through the minimization of the following loss function,

$$\mathcal{L} = \frac{1}{N_{\mathcal{B}}} \left(\sum_{l \in \mathcal{B}} \frac{w_{\epsilon}}{N_l} |E_l - E(\mathbf{R}^l)|^2 + \frac{w_f}{3N_l} \|\mathbf{F}_l - \mathbf{F}(\mathbf{R}^l)\|^2 \right) \quad (7)$$

where \mathcal{B} is a mini-batch (i.e., a subset of the training set) with $N_{\mathcal{B}}$ atomic configurations, w_{ϵ} and w_f are weights, E_l and \mathbf{F}_l are reference energies and forces, $E(\mathbf{R}^l)$ and $\mathbf{F}(\mathbf{R}^l) = -\nabla_{\mathbf{R}} E(\mathbf{R}^l)$ are the energy and force predictions of our model described in Eq. (2) for configuration l in the minibatch, and \mathbf{R}^l and N_l are the atomic coordinates and the number of atoms in configuration l . The parameters in the training procedure were as follows. We trained the models using the Adam optimizer with learning rate $\alpha(i) = \alpha_0 \lambda^{i/\tau}$ where $\alpha_0 = 0.002$ is the initial learning rate, $\lambda = 0.97$, $\tau = 5 \times 10^3$, and i is the step number. The batch size $N_{\mathcal{B}}$ was set to one and we trained for a total number of steps equal to 2×10^6 . w_{ϵ} and w_f were varied according to $w_{\epsilon}(i) = w_{\epsilon}^1 + (w_{\epsilon}^0 - w_{\epsilon}^1) \lambda^{i/\tau}$ and $w_f(i) = w_f^1 + (w_f^0 - w_f^1) \lambda^{i/\tau}$, with $w_{\epsilon}^0 = 0.02$, $w_{\epsilon}^1 = 1$, $w_f^0 = 1000$, and $w_f^1 = 1$. This scheme gives a higher weight to the force term in Eq. (7) at the beginning of the training process, and by the end of it both the energy and force term have equal weights. After the training process, the model was subsequently compressed as described in ref. 12 in order to improve the computational performance.

An example of the training curves resulting from the protocol described above is shown in Fig. 1.

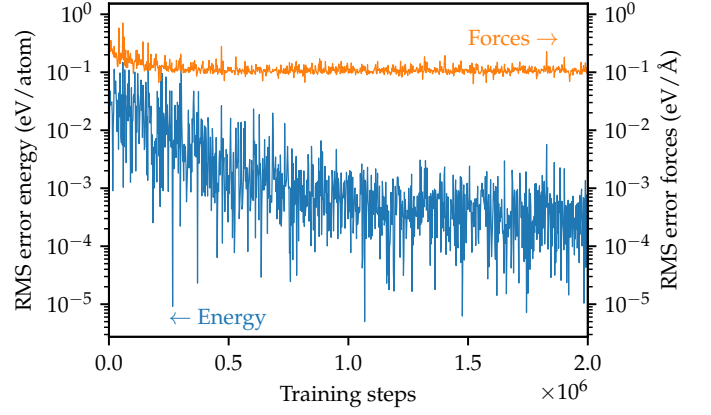


Fig. 1 Training curve for the machine learning model of the potential energy surface of liquid water, ice, microcline feldspar, feldspar/vacuum interfaces, and feldspar/water interfaces. The root mean square (RMS) error in energy and forces is shown as a function of the training steps.

3 Molecular dynamics simulations

We performed molecular dynamics simulations using LAMMPS^{13,14}. The time step for the integration of the equations of motion was 0.5 fs. We used the standard atomic weights for the masses of all elements (K, Al, Si, O) except H, for which we used a mass of 2 grams/mol, in order to improve the stability of the integration of the equations of motion. We controlled the temperature using the stochastic velocity-rescaling thermostat¹⁵ with a relaxation time of 0.1 ps.

For constant pressure simulations we employed a Parrinello-Rahman barostat with relaxation time 1 ps. For the simulation of bulk liquid water we used an isotropic barostat (box vector lengths coupled, and angles fixed) and for bulk ice I_h we used a fully anisotropic barostat (box vector lengths and angles can vary independently). In both cases we employed systems with 288 water molecules.

4 Enhanced sampling simulations

Enhanced sampling simulations were carried out using the enhanced sampling plugin PLUMED^{16,17}. We employed the global Steinhardt parameter Q_6 as collective variable¹⁸⁻²⁰, which we define below. We first consider the local Steinhardt parameter for oxygen atom i ,

$$q_{6m}(i) = \frac{\sum_j \sigma(r_{ij}) Y_{6m}(\mathbf{r}_{ij})}{\sum_j \sigma(r_{ij})} \quad (8)$$

where the sums run over all neighboring oxygen atoms j closer than a distance r_2 (defined below), \mathbf{r}_{ij} is the distance vector between i and j with norm $r_{ij} = |\mathbf{r}_{ij}|$, Y_{6m} are spherical harmonic with $l = 6$, and $-6 < m < 6$ is an integer.

The switching function $\sigma(r_{ij})$ makes $q_{6m}(i)$ continuous and differentiable and is defined as,

$$\sigma(r_{ij}) = \begin{cases} 1 & \text{if } y < 0 \\ (y-1)^2(1+2y) & \text{if } 0 < y < 1 \\ 0 & \text{if } y > 1 \end{cases}, \quad (9)$$

with $y = (r_{ij} - r_1)/(r_2 - r_1)$, $r_1 = 0.3$ nm, and $r_2 = 0.35$ nm. The values of r_1 and r_2 were chosen based on the position of the first peak of the radial distribution function, such that $\sigma(r_{ij}) \approx 1$ for first neighbors and $\sigma(r_{ij}) \approx 0$ for more distant neighbors.

Using $q_{6m}(i)$ described above, we define the global Steinhardt order parameter as,

$$Q_6 = \sqrt{\frac{4\pi}{2l+1} \sum_{m=-l}^l \left| \frac{1}{N} \sum_{i=1}^N q_{6m}(i) \right|^2} \quad (10)$$

where $l = 6$, and N is the total number of oxygen atoms for which $q_{6m}(i)$ was calculated. This quantity was used to drive the formation of ice in the microcline/water interfacial systems used during the active learning procedure.

In large simulation boxes, biasing Q_6 leads to the formation of ice throughout the simulation box, instead of localized in a region of space, as expected in nucleation phenomena. For this reason, we used a variant of Q_6 , denoted Q_6^{sph} , which localizes the effect of the collective variable in a spherical region centered at a position \mathbf{r}^0 . We define this collective variable in the following way,

$$Q_6^{\text{sph}} = \sqrt{\frac{4\pi}{2l+1} \sum_{m=-l}^l \left| \frac{\sum_{i=1}^N \gamma(r_{i0}) q_{6m}(i)}{\sum_{i=1}^N \gamma(r_{i0})} \right|^2} \quad (11)$$

where $r_{i0} = |\mathbf{r}_i - \mathbf{r}^0|$ is the distance from the i -th oxygen atom to a reference position \mathbf{r}^0 (center of the sphere), which we define as the position of a K atom at the microcline/water surface, and other symbols have the same meaning as above. The switching function $\gamma(r_{i0})$ was defined as,

$$\gamma(r_{i0}) = 1 - \tanh(r_{i0}/r^*) \quad (12)$$

with $r^* = 1.5$ nm, Q_6^{sph} was used to drive the formation of ice the large microcline/water interfaces described in the main part of manuscript.

5 Initial training data

The starting point for training the machine-learning model consisted in configurations of bulk microcline feldspar, bulk water and ice Ih, and feldspar surfaces in contact with liquid water. The procedure to obtain such configuration is described below.

Microcline feldspar The initial configuration for microcline feldspar was obtained from the American Mineralogist Crystal Structure Database and corresponds to the experimental structure determined in ref. 21. Experiments show that Al and Si can exhibit disorder in the four non-equivalent sites $T_1(0)$, $T_1(m)$, $T_2(0)$, and $T_2(m)$. The probability of finding Si in each of these sites is in the ranges 0.23-0.52, 0.64-0.87, 0.92-0.98, and 0.92-0.98, respectively²¹⁻²³. We made the simplification of assigning Al to $T_1(0)$ sites, and Si to $T_1(m)$, $T_2(0)$, and $T_2(m)$. Thus, our microcline configurations do not show Al/Si disorder, and the stoichiometry is exactly KAlSi_3O_8 .

We first minimized the energy of the microcline structure with respect to the cell dimensions and the atomic coordinates using DFT and the SCAN functional. The configuration contained 52 atoms, corresponding to 4 formula units. The lattice constants thus obtained are reported and discussed in the Results section below. Starting from the relaxed configuration, we perturbed randomly the atomic coordinates and box dimensions to generate structures that sample the potential energy surface of bulk feldspar. Each atomic coordinate is displaced by $\delta R = \delta R_{\text{max}}(2x-1)$ where δR_{max} is the maximum displacement and $x \in (0, 1)$ is a random number, and each box dimension L is scaled with $L = L_0[1 + \epsilon_{\text{max}}^L(2x-1)]$ where L_0 is the initial box dimension and ϵ_{max}^L is the maximum fractional change in box dimensions. The number of perturbed configurations and the parameters used are summarized in Table 1.

Bulk water, ice Ih, and ice Ic In order to obtain initial configurations for bulk water, ice Ih, and ice Ic we used trajectories from our previous work²⁴⁻²⁶, which were driven by a previously trained DeepMD model²⁷. Information about these simulations and the methodology for extracting frames from the trajectories is given in Table 2. In total, around 1000 configurations of liquid water, ice Ih, and ice Ic were used. Note that in two of the simulations liquid water and ice were sampled reversibly²⁴. Thus, the training data includes interfacial configurations and ice configurations with diverse proton arrangements (proton disorder).

Table 1 Parameters for the generation of microcline feldspar configurations with randomly perturbed atomic coordinates and box dimensions. δR_{max} is the maximum displacement of atomic coordinates and ϵ_{max}^L is the percent change in box dimensions.

# configurations	δR_{max} (Å)	ϵ_{max}^L (%)
100	0.01	1
100	0.05	2
100	0.1	3
10	0.2	5

Table 2 Configurations of bulk water, ice I_h , and ice I_c used as starting point for training the model. Each row corresponds to a different simulation trajectory. We report the number of configurations extracted from a given trajectory, the phases that were explored, a description of the origin of the simulation trajectory, the time interval between successive frames Δt , the temperatures or range of temperatures of the simulations T , the pressure P , and the number of water molecules in the simulation box N_{H_2O} .

# configurations	Phases	Description	Δt (ns)	T (K)	P (bar)	N_{H_2O}
300	water, ice I_h	Biased simulation ²⁴	0.25	300-350	1	96
300	water, ice I_c	Biased simulation ²⁴	0.25	300-350	1	64
40	water	Standard simulation ²⁵	1	260,270,280,300	1	288
200	LDL, HDL	Standard simulation ²⁶	2.5	235	3200	192
122	water	Biased simulation ²⁴	0.25	260-350	1	64

der). Although most simulations were performed at constant pressure 1 bar, we also employed a simulation at 3200 bar which is below the liquid-liquid critical point²⁶. We used these simulation trajectory to provide the model with examples of the atomic environments in the low-density liquid (LDL) and high-density liquid (HDL)²⁸.

Feldspar surfaces We constructed feldspar surfaces with the (100), (010), and (001) crystallographic planes exposed to vacuum. We identified several terminations for each surface and we saturated all undercoordinated oxygen atoms with protons. For each of the studied terminations, we minimized the energy with respect to the atomic coordinates. Then we generated ten perturbed configurations for each of the terminations and the $\delta R_{\max} - \epsilon_{\max}^L$ parameters in Table 1. The surface configurations contained between 58 and 128 atoms.

6 Active learning

With the data described above, we trained an initial ensemble of four models, with different initialization random seeds. The models were subsequently improved with an active learning procedure. This approach is iterative and each iteration k can be described as follows,

1. use the models obtained in iteration $k-1$ to run molecular dynamics simulations. The simulations are described in detail below.
2. extract high-error configurations from the resulting trajectories. The error ϵ_i in the i -th force component is calculated as proposed in ref. 29, i.e. $\epsilon_i^2 = \langle |f_i - \bar{f}_i|^2 \rangle$, where $\bar{f}_i = \langle f_i \rangle$ and the average $\langle \cdot \rangle$ is taken over the ensemble of models. The errors were evaluated every 0.25 ps and high-error configurations were defined by a maximum error in the forces $\epsilon = \max_i \epsilon_i > 0.2$ eV/Å.
3. compute energies and forces of the extracted configurations via DFT calculations (see section on electronic-structure calculations for further details).

4. train a new set of four models which correspond to iteration k . The data for the models in iteration k includes all or a subset of the data obtained in iterations $i = 1, \dots, k$.

This process was repeated until the convergence criterion $\max_i \epsilon_i < 0.2$ eV/Å was met. Around 20 iterations were needed to reach the desired accuracy.

In Fig. 1 we show the training curves for the final machine learning model of the *ab initio* potential energy surface obtained in this work.

Feldspar/water and feldspar/ice interfaces Configurations for the active learning process were generated using molecular dynamics simulations of the water/feldspar interface at constant volume, and constant temperature equal to 300 K (7 K of supercooling). We used the most stable terminations for the (100), (010), and (001) surfaces of feldspar. The configurations contained 544, 900, and 664 atoms, respectively. Since the water to ice transition is a rare event in the time scale accessible to a molecular simulation, we used advanced sampling to observe this phase transformation at feldspar surfaces during our simulations. This was achieved through a harmonic bias potential,

$$V(Q_6(\mathbf{R})) = \frac{k}{2} (Q_6(\mathbf{R}) - Q_6^{\text{ctr}}(t))^2 \quad (13)$$

where $k = 10^5$ kJ/mol, $Q_6(\mathbf{R})$ is the Steinhardt parameter defined above, and $Q_6^{\text{ctr}}(t) = Q_6^0 + (Q_6^f - Q_6^0)t/t_{\text{tot}}$ is the center of the harmonic bias which changes linearly with time t from initial value Q_6^0 to final value Q_6^f over the total simulation time $t_{\text{tot}} = 1$ ns. $Q_6^0 \approx 0$ and $Q_6^f = 0.55$ are values of $Q_6(\mathbf{R})$ compatible with liquid water and ice, respectively.

7 Generalization errors

We tested the reliability of our force field by analyzing its ability to reproduce DFT energies and forces of configurations not provided in the training set. Using our force field we performed simulations of the (100), (010), and (001) microcline/water interface as described in Sec. 6. The width

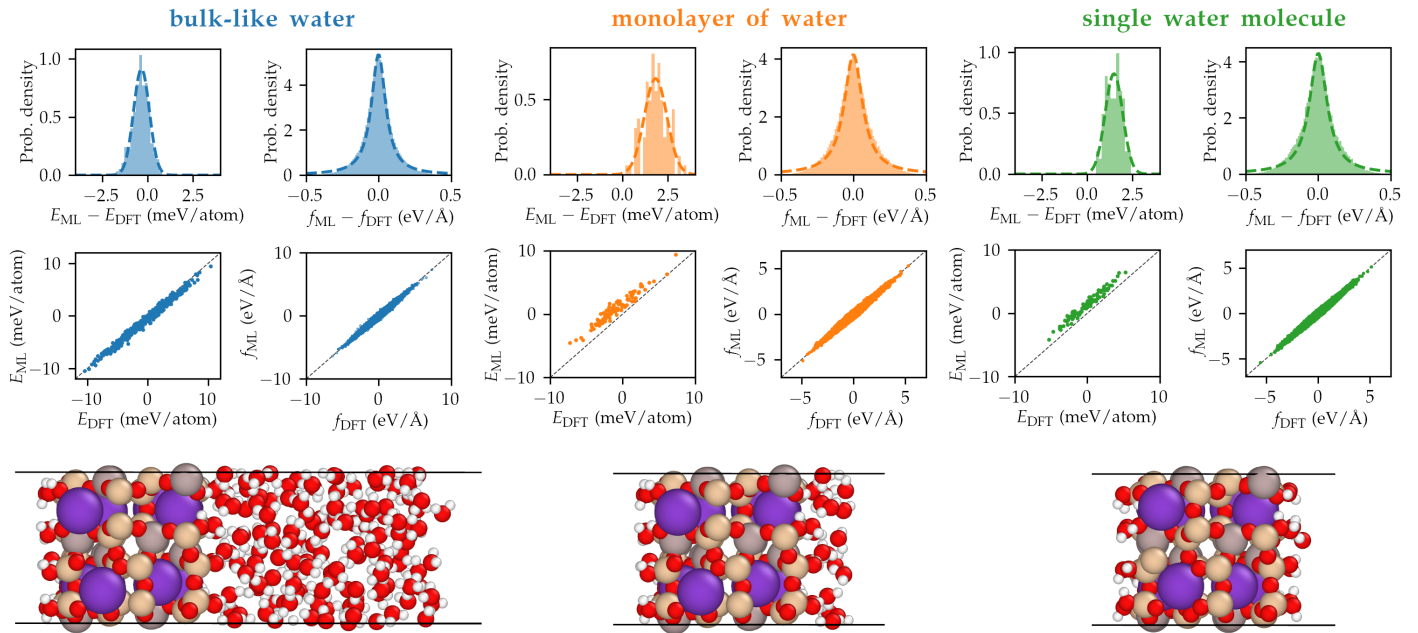


Fig. 2 Analysis of the accuracy of the trained model for bulk-like water (blue), a monolayer of water (orange), and a single water molecule (green) at the (001) surface of microcline (termination b). We show the distribution of errors in the energy $E_{ML} - E_{DFT}$ and the forces $f_{ML} - f_{DFT}$. The errors in energy and forces were fitted to a Gaussian and Lorentzian probability density function, respectively, and the corresponding curves are shown with a dashed line. Parity plots E_{ML} vs. E_{DFT} and f_{ML} vs. f_{DFT} are also shown. Representative atomistic configurations are shown for bulk-like water, a monolayer of water, and a single water molecule at the (001) surface of microcline.

of the water layer above the microcline surfaces is around 15 Å, sufficient to provide bulk-like atomic environments around water molecules. This choice of configurations was motivated by our goal to study heterogeneous ice nucleation in immersion mode. From each 1 ns simulation we extracted configurations every 10 ps for a total of 300 configurations for the three interfaces. The same simulation setup was used during the active-learning-based training process, and thus these configurations contain atomic environments similar to others already seen during the training of the force field. In Fig. 2 we analyze the errors of our force field compared to direct DFT calculations in configurations of bulk-like water at microcline surfaces. The errors were $\epsilon_E^{\text{RMS}} = 0.60$ meV/atom and $\epsilon_f^{\text{RMS}} = 107$ meV/Å, very similar to the errors made on configurations present in the training set.

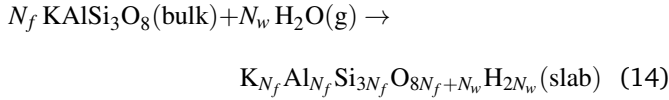
We also tested the ability of our force field to extrapolate to environments somewhat different to the ones present in the training set. For this purpose, we prepared a configuration of the (001) microcline surface (termination b) with a monolayer of water, and another configuration with a single water molecule at the surface (see Fig. 2 for representative atomistic configurations). We then performed 1 ns simulations at 300 K for each of these configurations and extracted configurations every 10 ps. In total, we extracted 100 configurations both for the monolayer of water

and the single water molecule at the surface. The atomic environments at these surfaces were not represented in the training set. The errors in energies and forces for these configurations with respect to DFT results are studied in Fig. 2 using error distributions and parity plots. Both for the monolayer of water and the single water molecule at the surface, the distribution of per-atom energy errors is centered at around 1.7 meV/atom, implying a small systematic deviation of the energy predicted by the model with respect to DFT. The RMS error in energy was $\epsilon_E^{\text{RMS}} = 1.9$ meV/atom for the monolayer of water and $\epsilon_E^{\text{RMS}} = 1.5$ meV/atom for the single water molecule. Both values are higher than the error $\epsilon_E^{\text{RMS}} = 0.60$ meV/atom found for bulk-like water at microcline interfaces. These results suggest that the force field is still valid for this type of configuration, although it loses some accuracy in the energy prediction. On the other hand, the prediction of the forces retains a relatively high-accuracy even for configurations not seen in the training set. Both for the monolayer of water and the single water molecule at the surface, the force error distribution is centered at the origin and the parity plots show excellent correlation between our force field and DFT predictions. The RMS error in forces was $\epsilon_f^{\text{RMS}} = 126$ meV/Å for the monolayer of water and $\epsilon_f^{\text{RMS}} = 123$ meV/Å for the single water molecule. These errors are only marginally higher than the error $\epsilon_f^{\text{RMS}} = 107$ meV/Å of the force field for bulk-

like water at microcline surfaces. These results suggest that the dynamics created by the force field are reliable, also for this type of configurations not present in the training set. An accurate description of single water molecules as well as a monolayer of water at microcline is important to study heterogeneous ice nucleation in deposition mode (from the vapor phase). Although training a force field to study nucleation in deposition mode was not the objective of our work, the results show that the model extrapolates relatively well to such conditions.

8 Calculation of surface energies

We evaluated the zero-temperature stability of different surfaces and terminations as follows. We consider slab configurations with two identical surface terminations which can be described as N_f unit formulas of bulk feldspar plus N_w water molecules,



The potential energy of the slab configurations can be written as,

$$E_{\text{slab}} = N_f E_{\text{feldspar}}^{\text{bulk}} + N_w E_{\text{H}_2\text{O}}^{\text{gas}} + 2A\gamma \quad (15)$$

where $E_{\text{feldspar}}^{\text{bulk}}$ is the potential energy of bulk feldspar, $E_{\text{H}_2\text{O}}^{\text{gas}}$ is the potential energy of an isolated water molecule, A is the surface area, and γ is the surface energy. Eq. (15) is only valid in the limit $N_f \rightarrow \infty$, when the slab is infinitely thick. As proposed by Fiorentini and Methfessel³⁰, we first compute E_{slab} for slabs of different thicknesses, same area A , and same number of water molecules N_w at the surfaces. Then, we calculate $E_{\text{feldspar}}^{\text{bulk}}$ as the slope of a linear fit of the $E_{\text{slab}} - N_w E_{\text{H}_2\text{O}}^{\text{gas}}$ vs N_f data. With the value of $E_{\text{feldspar}}^{\text{bulk}}$ thus calculated, the surface energy can be obtained via the definition,

$$\gamma = \lim_{N_f \rightarrow \infty} \frac{1}{2A} \left(E_{\text{slab}} - N_w E_{\text{H}_2\text{O}}^{\text{gas}} - N_f E_{\text{feldspar}}^{\text{bulk}} \right) \quad (16)$$

Note that according to this equation, the surface energy can be lowered either by reducing E_{slab} or by increasing the number of dissociated water molecules N_w at the surface.

The surface energy vs. slab thickness is shown in Fig. 3.

9 Formation of ice cluster at the microcline/water interface

The formation of ice at microcline/water interfaces was driven by the following bias potential,

$$V(Q_6^{\text{sph}}(\mathbf{R})) = \frac{k}{2} \left(Q_6^{\text{sph}}(\mathbf{R}) - Q_6^{\text{ctr}}(t) \right)^2 \quad (17)$$

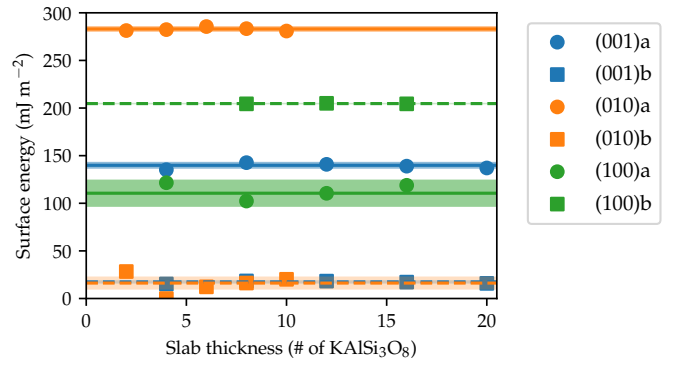


Fig. 3 Zero-temperature surface energy of different fully-hydroxylated terminations of the (100), (010), and (001) surfaces of microcline feldspar vs slab thickness. Fits of a constant function to the data are shown with lines. The width of the semi-transparent regions around each line are one-sigma errors of the fitting procedure.

where $k = 10^6$ kJ/mol, $Q_6^{\text{sph}}(\mathbf{R})$ is the Steinhardt parameter localized within a sphere (see definition above), and $Q_6^{\text{ctr}}(t) = Q_6^0 + (Q_6^f - Q_6^0)t/t_{\text{tot}}$ is the center of the harmonic bias which changes linearly with time t from initial value Q_6^0 to final value Q_6^f over the total simulation time $t_{\text{tot}} = 10$ ns. $Q_6^0 \approx 0$ and $Q_6^f = 0.4$ are values of $Q_6^{\text{sph}}(\mathbf{R})$ compatible with liquid water and ice, respectively.

Notes and references

- 1 P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo *et al.*, *J. Phys.: Condens. Matter*, 2009, **21**, 395502.
- 2 P. Giannozzi, O. Andreussi, T. Brumme, O. Bunau, M. B. Nardelli, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, M. Cococcioni *et al.*, *J. Phys.: Condens. Matter*, 2017, **29**, 465901.
- 3 J. Sun, A. Ruzsinszky and J. P. Perdew, *Phys. Rev. Lett.*, 2015, **115**, 036402.
- 4 J. Sun, R. C. Remsing, Y. Zhang, Z. Sun, A. Ruzsinszky, H. Peng, Z. Yang, A. Paul, U. Waghmare, X. Wu *et al.*, *Nat. Chem.*, 2016, **8**, 831.
- 5 M. A. Marques, M. J. Oliveira and T. Burnus, *Comput. Phys. Commun.*, 2012, **183**, 2272–2281.
- 6 D. Hamann, *Phys. Rev. B*, 2013, **88**, 085117.
- 7 J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865.
- 8 L. Bengtsson, *Physical Review B*, 1999, **59**, 12301.
- 9 L. Zhang, J. Han, H. Wang, R. Car and E. Weinan, *Phys. Rev. Lett.*, 2018, **120**, 143001.
- 10 H. Wang, L. Zhang, J. Han and E. Weinan, *Comput. Phys. Commun.*, 2018, **228**, 178–184.

- 11 L. Zhang, J. Han, H. Wang, W. Saidi, R. Car *et al.*, *Adv. Neural Inform. Process. Syst.*, 2018, **31**, 4436–4446.
- 12 D. Lu, W. Jiang, Y. Chen, L. Zhang, W. Jia, H. Wang and M. Chen, *Journal of chemical theory and computation*, 2022, **18**, 5559–5567.
- 13 S. Plimpton, *J. Comput. Phys.*, 1995, **117**, 1–19.
- 14 A. P. Thompson, H. M. Aktulga, R. Berger, D. S. Bolinteanu, W. M. Brown, P. S. Crozier, P. J. in't Veld, A. Kohlmeyer, S. G. Moore, T. D. Nguyen *et al.*, *Comput. Phys. Commun.*, 2022, **271**, 108171.
- 15 G. Bussi, D. Donadio and M. Parrinello, *J. Chem. Phys.*, 2007, **126**, 014101.
- 16 G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni and G. Bussi, *Comput. Phys. Commun.*, 2014, **185**, 604–613.
- 17 M. Bonomi, G. Bussi, C. Camilloni, G. A. Tribello, P. Banáš, A. Barducci, M. Bernetti, P. G. Bolhuis, S. Bottaro, D. Branduardi *et al.*, *Nat. Methods*, 2019, **16**, 670–673.
- 18 P. Steinhardt, D. Nelson and M. Ronchetti, *Phys. Rev. B*, 1983, **28**, 784–805.
- 19 J. Van Duijneveldt and D. Frenkel, *J. Chem. Phys.*, 1992, **96**, 4655–4668.
- 20 P. Rein ten Wolde, M. J. Ruiz-Montero and D. Frenkel, *J. Chem. Phys.*, 1996, **104**, 9932–9947.
- 21 S. Bailey, *American Mineralogist: Journal of Earth and Planetary Materials*, 1969, **54**, 1540–1545.
- 22 S. Liu and Y. Zhai, *Periodico di Mineralogia*, 2021, **90**,.
- 23 H. Kroll and P. H. Ribbe, *American Mineralogist*, 1987, **72**, 491–506.
- 24 P. M. Piaggi, A. Z. Panagiotopoulos, P. G. Debenedetti and R. Car, *J. Chem. Theory Comput.*, 2021, **17**, 3065–3077.
- 25 P. M. Piaggi, J. Weis, A. Z. Panagiotopoulos, P. G. Debenedetti and R. Car, *Proc. Natl. Acad. Sci. U.S.A.*, 2022, **119**, e2207294119.
- 26 T. E. Gartner III, P. M. Piaggi, R. Car, A. Z. Panagiotopoulos and P. G. Debenedetti, *Phys. Rev. Lett.*, 2022, **129**, 255702.
- 27 L. Zhang, H. Wang, R. Car and W. E, *Phys. Rev. Lett.*, 2021, **126**, 236001.
- 28 P. Gallo, K. Amann-Winkel, C. A. Angell, M. A. Anisimov, F. Caupin, C. Chakravarty, E. Lascaris, T. Loerting, A. Z. Panagiotopoulos, J. Russo *et al.*, *Chem. Rev.*, 2016, **116**, 7463–7500.
- 29 L. Zhang, D.-Y. Lin, H. Wang, R. Car and E. Weinan, *Physical Review Materials*, 2019, **3**, 023804.
- 30 V. Fiorentini and M. Methfessel, *Journal of Physics: Condensed Matter*, 1996, **8**, 6525.