

Structure Discovery in Atomic Force Microscopy Imaging of Ice

Fabio Priante, Niko Oinonen, Ye Tian, Dong Guan, Chen Xu, Shuning Cai, Peter Liljeroth,* Ying Jiang,* and Adam S. Foster*



Cite This: <https://doi.org/10.1021/acsnano.3c10958>



Read Online

ACCESS |

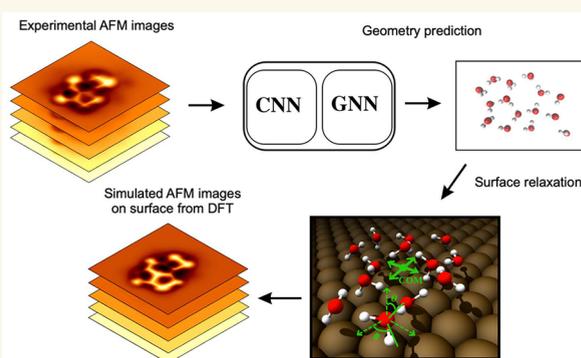
Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: The interaction of water with surfaces is crucially important in a wide range of natural and technological settings. In particular, at low temperatures, unveiling the atomistic structure of adsorbed water clusters would provide valuable data for understanding the ice nucleation process. Using high-resolution atomic force microscopy (AFM) and scanning tunneling microscopy, several studies have demonstrated the presence of water pentamers, hexamers, and heptamers (and of their combinations) on a variety of metallic surfaces, as well as the initial stages of 2D ice growth on an insulating surface. However, in all of these cases, the observed structures were completely flat, providing a relatively straightforward path to interpretation. Here, we present high-resolution AFM measurements of several water clusters on Au(111) and Cu(111), whose understanding presents significant challenges due to both their highly 3D configuration and their large size. For each of them, we use a combination of machine learning, atomistic modeling with neural network potentials, and statistical sampling to propose an underlying atomic structure, finally comparing its AFM simulated images to the experimental ones. These results provide insights into the early phases of ice formation, which is a ubiquitous phenomenon ranging from biology to astrophysics.

KEYWORDS: atomic force microscopy, ice nanoclusters, machine learning, tip functionalization, neural network potentials



INTRODUCTION

Water–solid interfaces feature prominently in a wide spectrum of scientific and technological problems, encompassing materials science, chemistry, biology, and geology. A prerequisite for their understanding is knowing how the water molecules will be structurally organized on the solid surface. The complex interplay between water–water and substrate–water interactions gives rise to a highly diverse range of possible structures, forming one-dimensional,^{1–3} two-dimensional,^{4–6} and three-dimensional⁷ configurations.

Among this rich variety, ice nanoclusters^{7–10} are of particular importance, as they enable sampling of the vast space of metastable configurations explored by water molecules during the heterogeneous ice nucleation process. Atomically resolved images of these nanoclusters can be obtained from atomic force microscopy (AFM) or scanning tunneling microscopy (STM) experiments, under ultrahigh vacuum and low temperature conditions and using tip-functionalization.¹¹ However, beyond clusters of only a few molecules, the resulting images are often difficult to analyze due to the tendency of the nanoclusters to arrange in buckled, bilayered structures once enough water molecules have been deposited.

The difficulty is further increased by the flexibility of their hydrogen bond framework, which, at close distances, can be significantly perturbed by an approaching microscope tip.

Due to these challenges, current investigations often focus on planar, monolayered nanoclusters, for which structural interpretation is relatively straightforward. To approach more general and three-dimensional cases, a promising route could be the application of recent advancements in machine learning (ML) for AFM image analysis. These techniques have already allowed extraction of physical descriptors,¹² electrostatic potential maps,¹³ ball-and-stick molecular representations,¹⁴ and molecular graphs¹⁵ from AFM image inputs. However, these models were trained on organic molecules in a vacuum, which present significantly different chemical features

Received: November 6, 2023

Revised: January 30, 2024

Accepted: January 31, 2024

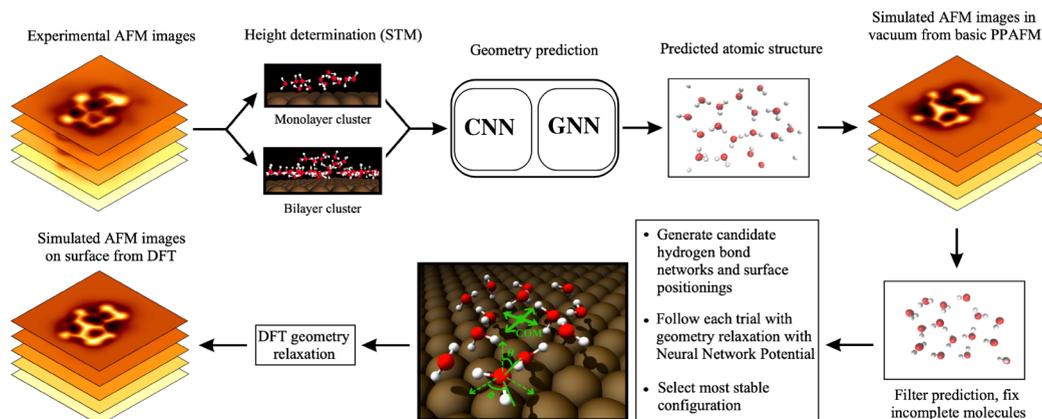


Figure 1. Workflow for structure discovery in AFM images of ice nanoclusters. If scanning tunneling microscopy (STM) is carried out in conjunction with the AFM, the monolayer or bilayer character of the nanocluster can be initially determined from the measured height difference to the substrate (see Figure S2). Then, an atomic structure prediction is inferred from the experimental AFM image stack. Incomplete molecules are fixed, while possibly unrealistic ones are eliminated. Next, multiple hydrogen bond networks and surface positions are rapidly evaluated by carrying out neural network potential relaxations. The most stable resulting structure is further optimized with DFT, from which simulated AFM images can be obtained and compared to the initial experiment.

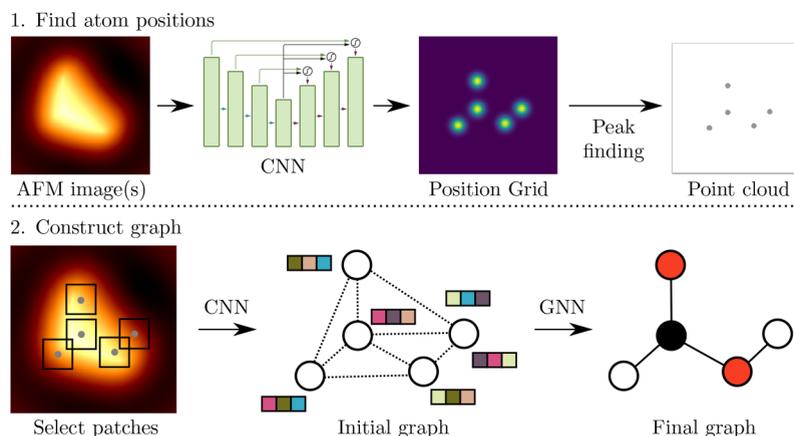


Figure 2. Schematic of the geometry prediction process. The prediction process is split into two parts. Following our previous work,¹⁵ the first part uses a CNN to transform the input AFM image(s) into a grid where the atomic positions are marked by Gaussian peaks which are then identified by a peak-finding algorithm to arrive at a list of coordinates for the atoms. The second part constructs a graph out of the found positions, which unlike previous work is done here in one shot. The process starts on the left by overlaying the found atom positions (gray dots) onto the AFM image and selecting rectangular patches around these positions. A CNN then turns the image patches into the initial embedding vectors (colored squares) for every node of the graph, and edges (dashed lines) are added to the graph based on the proximity of the nodes. Finally, a GNN processes the information in the node vectors and does a classification of the nodes into atom types and a binary classification on each edge whether it corresponds to a chemical bond.

compared to water molecules adsorbed on metal surfaces. Until architectures become available for robustly extrapolating across chemical space, it will remain necessary to generate additional data for applying automated AFM structure discovery to new domains. Indeed, by training on a custom water data set, the two-dimensional configuration of waters in a $\text{Na}^+\cdot 4\text{H}_2\text{O}$ hydrate was successfully predicted.¹⁶

Furthermore, to fully reconcile an experimental observation with its predicted ice nanocluster structure, the underlying substrate must be taken into account. This usually implies carrying out a geometry relaxation of the hypothesized structure onto the surface using quantum mechanical methods such as density functional theory (DFT). For large and bilayered clusters, this is inherently difficult, as very little information can be obtained with AFM about the organization of the lower ice layer. However, even for smaller monolayer cases, the irregular arrangement of water molecules means that

even small variations in the initial hydrogen bond network and adsorption configuration on the substrate can generate drastically different final geometries. This holds true also for weakly reactive, hydrophobic metal surfaces, which are generally favored in experiments, as they do not cause further complications such as hydrogen dissociation.

In this work, we tackle these challenges by developing a simulation workflow for structure discovery in high-resolution AFM imaging of large, buckled, mono- and bilayered ice nanoclusters on Au(111) and Cu(111) surfaces. We utilized the workflow on eight experimental AFM images, obtaining excellent agreement with the simulated AFM from the discovered atomic structures. Furthermore, we demonstrate robustness of the predictions upon their relaxation on surfaces, fully closing the loop between experiment and interpretation.

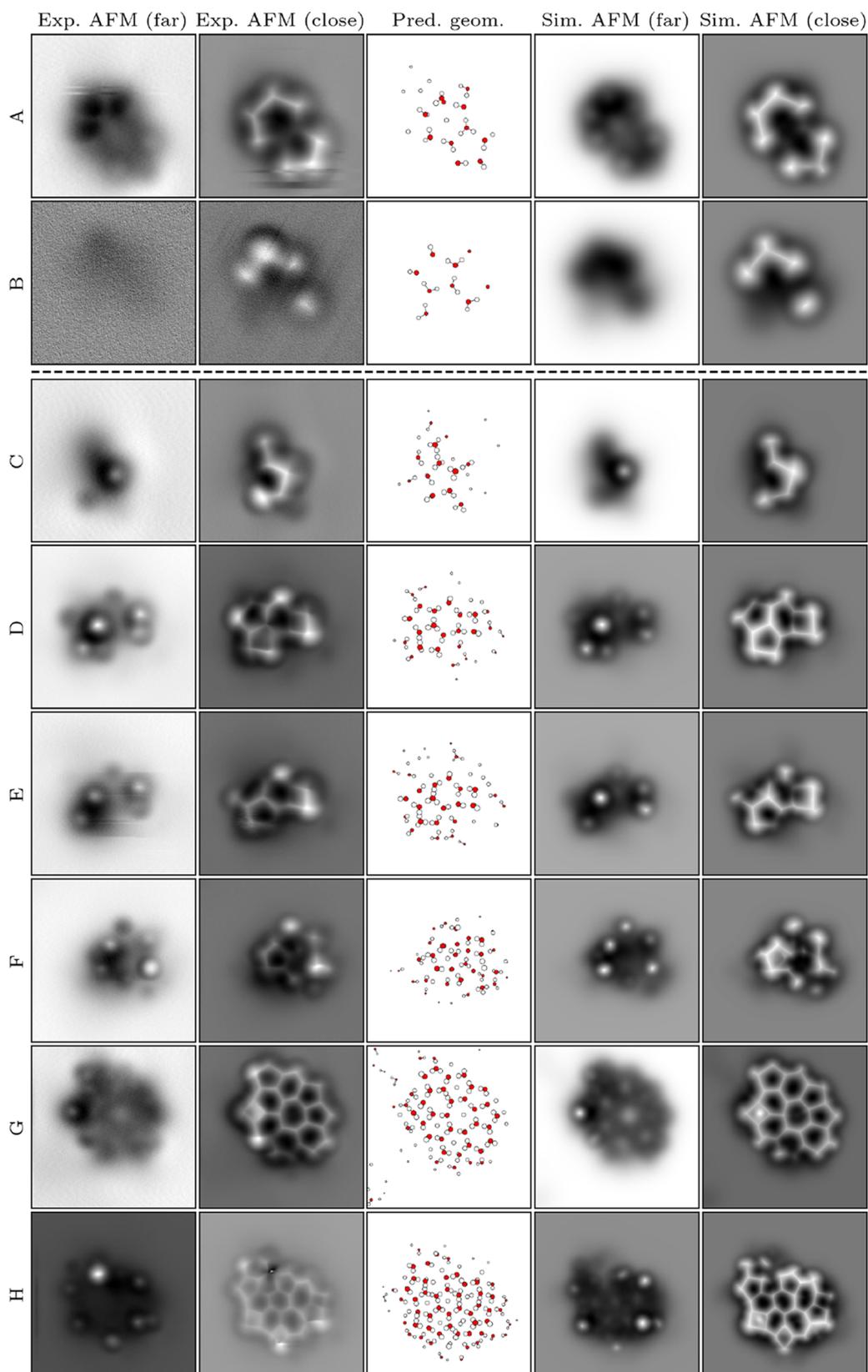


Figure 3. Geometry predictions for experimental AFM images and corresponding simulated AFM images from the PPM (in vacuum). Experiment B was conducted on Cu(111), while others were on Au(111). Each row corresponds to one experiment. Experiments A and B are monolayer, and experiments C–H are bilayer. On the left are the farthest and closest distance experimental AFM images, in the middle the predicted geometries, and on the right the simulations based on the predicted geometries. The tip–sample distances in the simulations are manually chosen in each case to visually match the experimental images. The sizes of the atoms in the geometry indicate the relative depths of the atoms.

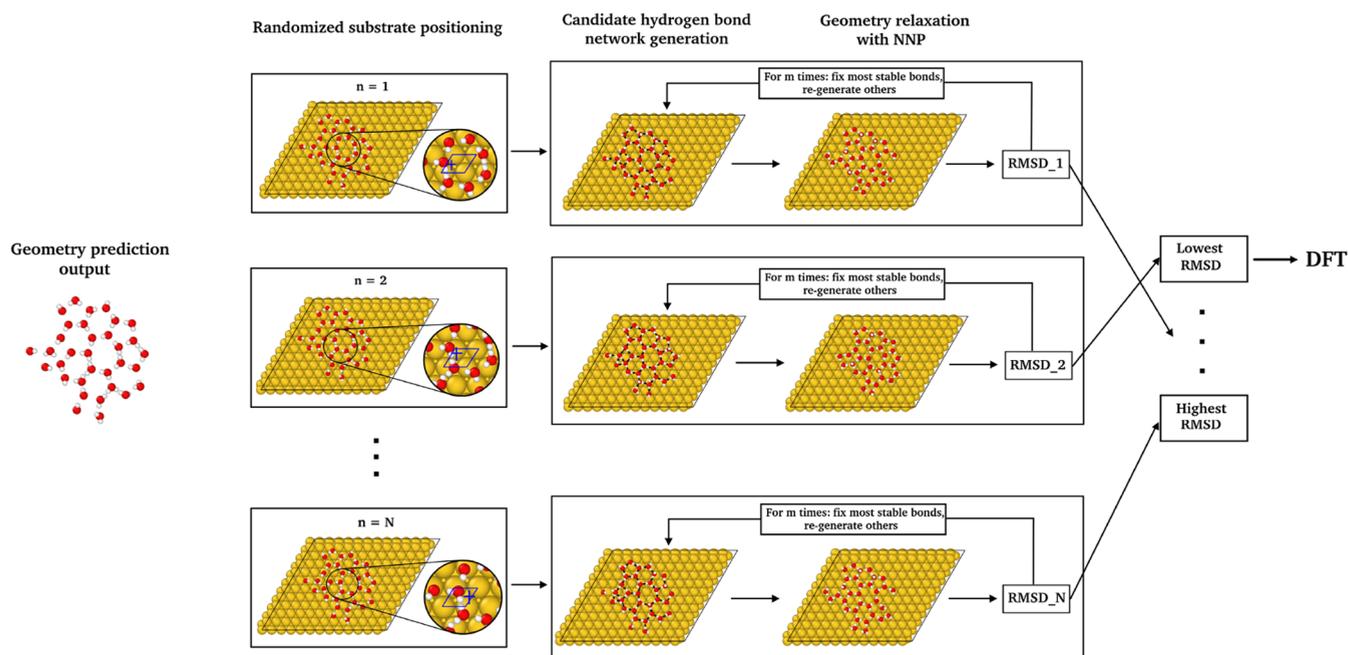


Figure 4. Estimation procedure to connect an ML geometry prediction to a DFT relaxation on a metal substrate. The geometry center of mass (shown as a blue cross in the snapshots) is positioned in a random position of the substrate unit cell, and the cluster is randomly rotated around z . In particular, the initial height of the cluster center of mass is taken randomly from an interval around the central peaks of the [Figure S10](#) (left column). For each of the starting positions, an iterative bond-creation/NNP-relaxation loop is then carried out. A tentative hydrogen bond network is constructed with the algorithm shown in [Figure S8](#). The cluster is then relaxed using the NNP. The RMSD of the relaxed oxygens relative to their initial position is then computed. At this point, more refinements iterations can be carried out, with the oxygens that did move less than a certain threshold being fixed and the rest of the network being constructed again. The sample RMSDs are then sorted and the sample with the lowest one is relaxed with DFT. In the case of bilayer nanoclusters, the only change is the addition of an hexagonal ice layer between the substrate and the prediction, whose orientation is also randomly sampled.

RESULTS

The structure discovery workflow involves a sequence of steps summarized in [Figure 1](#). We apply it on seven experimental samples on a Au(111) surface and on one sample on Cu(111). The full experimental AFM image sets are shown in [Figure S1 in the Supporting Information](#). In the following sections, we outline the individual components of the workflow in detail.

Geometry Prediction. In the initial phase of the workflow, an AFM image stack is fed into our ML geometry prediction model, schematized in [Figure 2](#). Building upon our previous infrastructures,^{13,15} the model comprises an Attention U-Net convolutional neural network (CNN) for predicting atomic positions, and a Graph Neural Network (GNN) for identifying their corresponding atomic species.

Three different ML models were trained, covering each type of system we encountered in experiments. Particularly, we considered separately monolayer and bilayer nanoclusters on Au(111) and monolayers on Cu(111). The models were trained on distinct data sets of AFM images simulated with the probe particle model (PPM).¹⁷ The images were obtained using the hartree potentials of randomly generated ice nanoclusters, previously relaxed with neural network potentials (NNP), as shown in [Figure S3](#). The NNPs themselves were also trained separately, one for each metal, using a diverse range of water structures, exemplified in [Figure S4](#).

The results are shown in [Figure 3](#), where we compare the experimental AFM images of the clusters, their predicted geometries, and the AFM simulations from the predictions. The experimental data here have been preprocessed in two

ways before being input to the machine learning model. First, the model has been trained on a fixed pixel resolution (0.125 Å in both x - and y -directions) and z -height step (0.1 Å), so the experimental images are rescaled by linear interpolation to match the predefined resolution. [Figure S1 in the Supporting Information](#) shows which of the z -slices were constructed by interpolation. Second, each z -slice in the image stack is normalized individually by subtracting the mean and dividing by the sample standard deviation of the pixel values.

Of these experiments, two (A and B) are monolayer and the others bilayer, which is indicated by different height profiles in STM line scans, shown in [Figure S2 in the Supporting Information](#). Sample B is on Cu(111), and the others are on Au(111). For each sample, we include images at both high and low tip heights, emphasizing the markedly three-dimensional character of these structures. We observe an excellent agreement between predictions and observations, with just a single very high atom in sample H that perhaps confused the model and was not accounted for.

We found the performance of the geometry predictions to improve upon increasing the input AFM image stack size, n_z , saturating at around $n_z = 10$. We carry out an in-depth analysis of this trend in section “Effect of the AFM image stack size” and [Figures S5–S7 in the Supporting Information](#). In the same section, we further show how, if using only $n_z = 1$, the images closer to the sample provide more information, as expected.

Prediction Pruning. In the least bright regions of the images, which are the hardest even for human experts to interpret, all our three models tend to generally predict the presence rather than the absence of low-lying atoms. This is

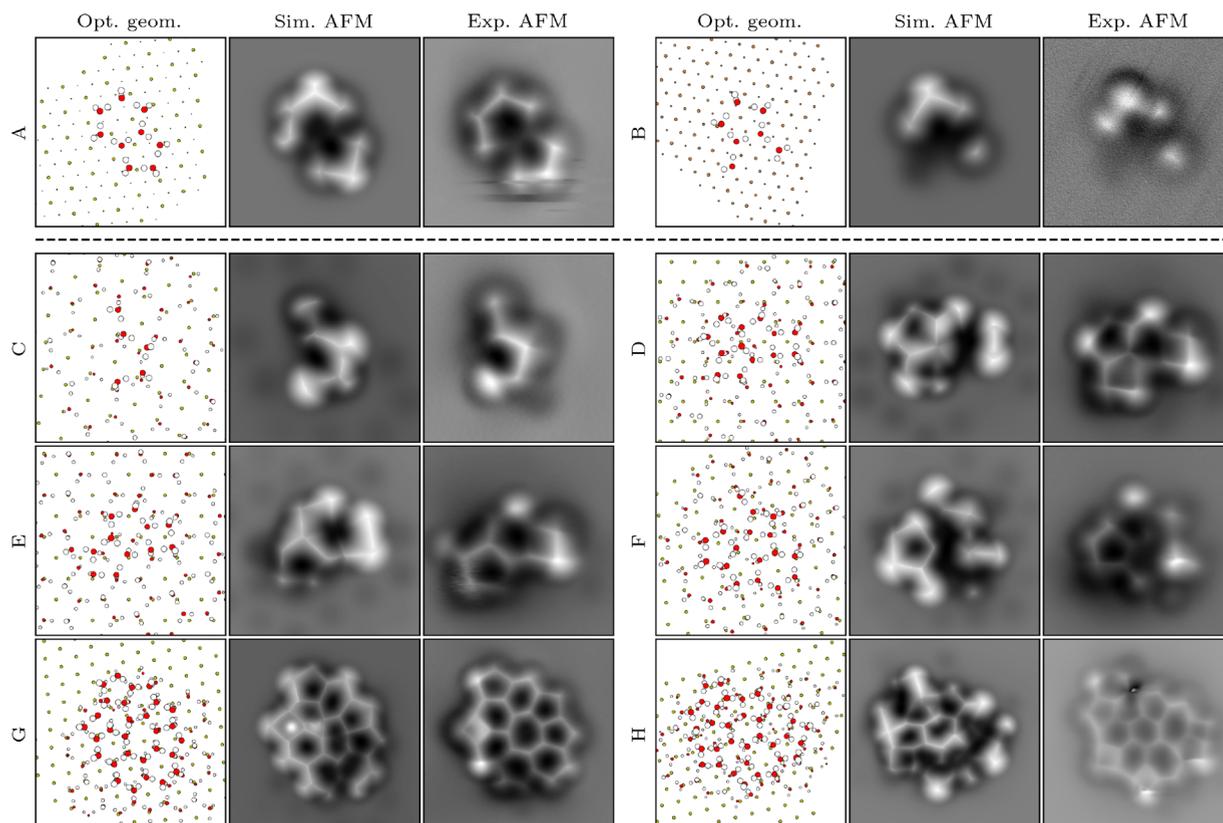


Figure 5. On-surface geometries optimized with an NNP, corresponding simulated AFM images, and comparison to experiment. Rows A–H correspond to the experiments with the same labels as in Figure 3. On each row, on the left is the final optimized geometry, in the middle is the simulated AFM image, and on the right is the experimental image.

likely due to the ML model trying to reproduce as accurately as possible the experimental image features, which are generally sampled from a distribution different from the simulated ones the ML was trained on. This is a well-known problem when adapting models trained on synthetic data onto real-world scenarios, and the development of strategies to overcome it is still an active area of research.^{18–21}

When moving on to the next step of the workflow, the addition of the substrate and the geometry relaxation, the presence of possibly spurious molecules can become problematic. In general, we filter out molecules from the predictions that are low-lying and either isolated or placed inside closed loops, as we have observed that they usually increase in height and form brighter, inconsistent features post-relaxation. For example, this is clearly visible in the predicted geometry of experiment G in Figure 3—the predicted water fragments at the left corners or inside the biggest loops are eventually removed after relaxing on the surface. An informed decision can be made after several trials by observing how the molecule rearranges during the relaxations and by visually evaluating its resulting features in the simulated images. At the moment, the pruning procedure is carried out manually, and it depends on human intuition and experience. Strategies to automate this step are currently being explored.

Lastly, before bringing the surface into the picture, we also adjust eventual incomplete water molecules. In fact, the ML model is not forced to predict complete water molecules but only atomic structures that generate a matching simulated image with experiment. Missing hydrogens are added so that the internal angles of their corresponding molecules are

correct. Beyond this condition, the exact hydrogen positioning is not important at this stage since the molecules will subsequently be rotated in searching for stable hydrogen bond frameworks.

Surface Relaxation. In Figure 4 we illustrate the process for searching for a stable combination of substrate positioning and hydrogen bond arrangement for a given monolayer cluster prediction. We follow the same procedure for bilayers, but also place below the predicted molecules an ice hexagonal layer as in the standard bilayer ice model,²² from which the outer edges are removed to arrive at an isolated, nonperiodic ice nanocluster. In the search, we consider the oxygens in the filtered ML prediction as the ground truth, and we then explore a large array of translations and rotations of the nanocluster center of mass on the surface. For each starting position, a different hydrogen bond network is then generated according to the algorithm in Figure S8, also exemplified in Figure S9 in the Supporting Information. More specifically, hydrogen bonds are created starting from a randomly chosen water molecule, proceeding radially toward the others. A probability is assigned to each molecule to form one or two bonds with its neighbors according to its oxygen height above the surface, following simple statistical considerations from the ML data sets, as shown in Figure 10 and Figure 11 in the Supporting Information. Briefly, we found that in monolayers, low-lying molecules tend to be bonded with two neighbors, arranging in a horizontal position on the surface. Molecules at intermediate heights are instead more frequently positioned vertically (i.e., with one OH bond toward the surface), thus bonding with only one neighbor, while molecules with even

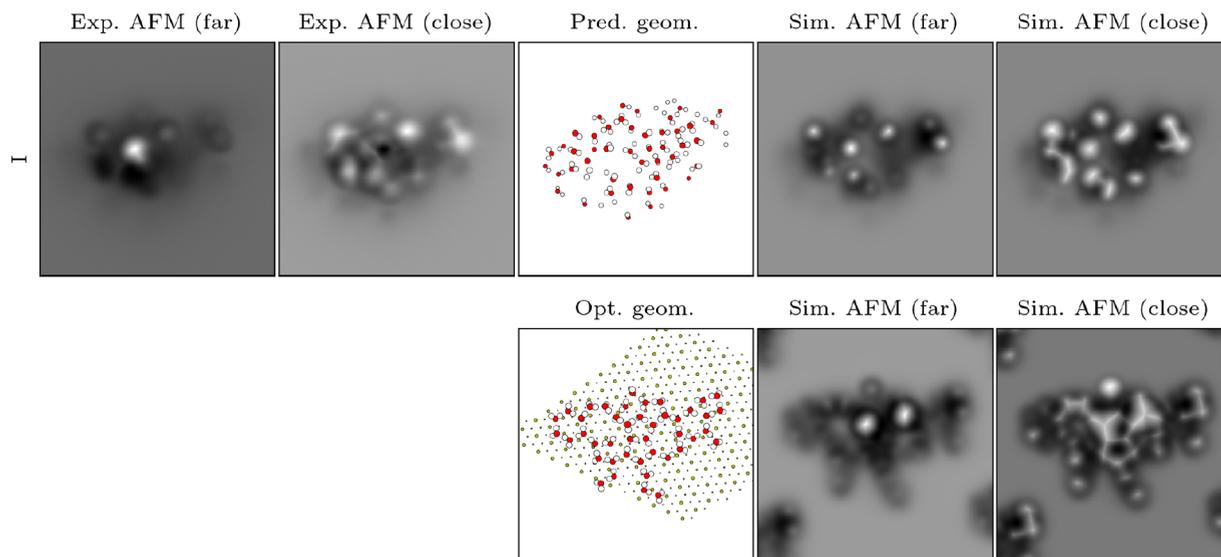


Figure 6. Additional experiment with a prediction and an optimized geometry, and the corresponding simulations in vacuum (top) and on surface (bottom).

higher oxygen z coordinate are instead more diversely oriented. For bilayers, this latter situation applies to all of the predicted molecules, which are effectively decoupled from the metal. In the bottom layer, waters are allowed to rotate upward, if standing below a molecule that was positioned horizontally. Once the position and the bonds of each configuration have been set, we do geometry optimization using the previously described NNPs. The NNP-relaxed structures are then ranked based on how much the oxygens were displaced compared to their ML prediction ground truth. The highest ranked structure is finally relaxed using DFT, and its Hartree potential is used for simulating the AFM images to be compared with experiment.

In Figure 5, we compare the surface-relaxed geometries, their AFM simulations, and the experimental images. We do an extended comparison over three different tip–sample heights in Figure S12 in the Supporting Information. Additionally, we compare experiments, vacuum simulations, and on-surface simulations together in Figure S13 in the Supporting Information. Generally, we find the agreement between simulations and experiments to be still very good after the addition of the underlying bottom ice layer and the metal surface. In addition, we only observe minimal ionic movements during the DFT runs, confirming the accuracy of our NNPs. The agreement is particularly well-preserved for samples A, B, and C, whose small size allowed the positioning and bonding search procedure to thoroughly explore most viable combinations. For the bigger clusters, we notice instead slight readjustments of the molecules, especially in their z coordinate. Sample H in particular, the largest among the clusters, tended to adopt a more three-dimensional structure upon relaxation, diverging from the flat profile predicted by the ML model. This is to be expected, and it results from the higher complexity of these bigger structures, where even small errors in the oxygen positions or slight differences in the choice of hydrogen bonds could have global effects on the cluster geometry.

DISCUSSION

To the best of our knowledge, only cluster B was previously reported in literature, as part of the development of

electrostatic discovery in AFM.¹³ In light of a recent work²³ that showed the presence of a common 15-mer cluster on Cu(111) and Pt(111), we also observe that experiment A resembles the initial state of the 15-mer formation, potentially signaling its presence on Au(111) as well.

We notice that structure predictions from samples A, E, F, and G possess a similar final orientation of the underlying lattice, despite not having imposed such a rotational constraint on the clusters during the search. Certainly, in such amorphous clusters, multiple bonding configurations and center of mass rotations could still result in visually similar and stable geometries. Nevertheless, as the experiments were effectively conducted on an individual clean Au(111) surface, it is a promising result to retrieve the correct orientation. Sample B's final surface orientation was also found to be matching the experimental one. This was the only sample measured on Cu(111), but STM images of its surface had been taken and they allowed us to confirm the matching.

We also want to highlight an additional experiment, shown in Figure 6, for which the geometry prediction failed to extract the complete molecular structure, especially in the center of the cluster. To better understand what caused the ML models to struggle, we attempted to manually build a tentative geometry, which is also shown in the bottom row of the same figure. In all probability, the presence of very high and low molecules next to each other in a concentrated region of the cluster constituted the confounding factor for the ML predictions. The situation might have further been worsened by the lowest images being affected by tip–sample interactions, which may have altered the cluster geometry. In future improvements, we aim to address both problems. First, we will ensure incorporation of similar configurations in our data set, as a postprocessing step on some of the randomly generated clusters. In general, as we analyze more experimental images and thus uncover more problematic patterns, we will correspondingly expand our data set generation to include them. Second, tip-induced relaxation effects in the ice clusters could be directly included in the creation of the data sets, providing the ML with educated guesses about structural changes of the clusters depending on the image height.

CONCLUSIONS

We have demonstrated the effectiveness of our workflow for structure discovery in three-dimensional ice nanoclusters and validated it by comparison to experimental measurements. In contrast to previous efforts in approaching similar problems, we now automate the structure prediction, the hydrogen bond arrangement, and the substrate relaxation, leveraging recent advancements in machine learning which allow us to iterate through tentative structures much more efficiently compared to traditional DFT. Our work enables the partnering of high-resolution AFM with accurate simulations of a wide variety of as yet unexplored complex ice systems, which have generally been inaccessible due to the sheer difficulty in their interpretation. This will broaden and accelerate our understanding of heterogeneous ice nucleation and, more generally, of water–surface interactions, a fundamental concept in numerous scientific and technological domains.

METHODS

AFM Experiments. The experiments were performed on two separate noncontact AFM/STM systems (Createc, Germany) at 5 K using qPlus sensors equipped with a tungsten (W) tip (spring constant $k_0 \approx 1800$ N/m, resonances frequencies of $f_0 \approx 29.1$ kHz, and quality factor $Q \approx 10^5$). All of the AFM frequency shift (Δf) images were obtained with the CO-terminated tips in constant-height mode. The CO-tip was obtained by positioning the tip over a CO molecule on the Au(111) surface at a set point of 100 mV and 10 pA, followed by increasing the current to 400 pA. The CO functionalization on Cu(111) was achieved by positioning the tip over a CO molecule at the set-point of 8 mV and 100 pA, followed by ramping the sample bias from zero to 2.6 V with feedback off. The oscillation amplitude of experimental AFM imaging was 50 pm (cluster B on Cu(111)) or 100 pm (all experiments on Au(111)). The Au(111) and Cu(111) single crystal were purchased from MaTeck. The Au(111) surface was cleaned by repeated Ar⁺ ion sputtering at 1 keV and annealing at about 700 K for multiple cycles. The Cu(111) surface was prepared by repeated Ne⁺ ion sputtering at 750 eV followed by annealing at about 820 K. Ultrapure H₂O (Sigma-Aldrich, deuterium-depleted, 1 ppm) was used and further purified under vacuum by 3–5 freeze-and-pump cycles to remove remaining gas impurities. The water molecules were deposited in situ onto the surface held at 5 K through a dosing tube, followed by annealing at 77 K for 10 min. Note that the fact that we have only monolayer ice clusters on Cu (111), and both monolayer and bilayer ice clusters on Au (111) is a function of the availability of high quality experimental images at a variety of heights. We cannot make any inference on the probability of different ice structures on these limited statistics.

AFM Simulations. The AFM simulations are performed using the probe particle model (PPM)¹⁷ code using a Lennard-Jones interaction and an electrostatic interaction calculated from the Hartree potentials of the samples obtained from the DFT calculations detailed below. The only exception to this is the simulations on the predicted geometries (Figure 3), where only the Lennard-Jones model is used, since the prediction only contains the geometry and atom types without any electrostatics information. The default built-in Lennard-Jones parameters in PPM are used.

In order to augment the data set and make the trained machine learning models more robust, the simulation is performed multiple times for the same sample while varying some of the simulation parameters. Following the example of the QUAM-AFM data set,¹⁴ we vary the lateral spring constant k_{xy} and the oscillation amplitude A . In addition, we vary the closest tip–sample distance d_{ts} and the lateral equilibrium position of the probe particle t_{xy} . In the QUAM-AFM data set, the parameters take all combinations of predetermined values, in total 24 different combinations for the two parameters. In our case where we have 4 parameters to vary, the number of different combinations would grow very large, so we instead choose to do a

fixed number of 10 simulations for each sample, randomly picking the parameters in set ranges from a uniform distribution. The set ranges are 0.2–0.5 N/m for k_{xy} , 0.4–2.0 Å for A , t_{xy} is in a disk of radius 0.3 Å, and $\Delta d_{ts} = 0.5$ Å. The average of d_{ts} is chosen by eye such that sharp features like ones seen in real AFM images at close approach are seen in the simulated images. The simulations are performed at 15 tip–sample distances with a 0.1 Å step, producing 3D stacks of AFM images.

Geometry Prediction Model. The atomic geometry prediction model follows closely our previous work on reconstructing molecule graphs from AFM images¹⁵ with some modifications to make the model more general. The basic structure of the model stays the same: there is a convolutional neural network (CNN) that first predicts the positions of the atoms from the AFM image stack, and a graph neural network (GNN) that uses the predicted positions along with the AFM images to construct a labeled molecule graph, as illustrated in Figure 2. The biggest differences to the original model are that the CNN and GNN networks are now completely separate without any shared layers between them, the CNN is modified to allow arbitrary size inputs in the z -dimension, and the GNN is simplified to label and connect the molecule graph in one shot for the whole graph instead of iteratively for each atom. These modifications are explained in more detail in the following.

Atom Position Prediction. The atom positions are predicted using an Attention U-Net CNN^{24,25} modified to accept variable size inputs while producing fixed-size outputs in z . In the U-Net architecture, the input image is first passed through an encoder that has a series of CNN blocks interleaved with pooling layers that gradually downsample the feature maps to a smaller size and then a decoder that gradually upsamples the feature maps back to the original size. Additionally, there are skip-connections between the corresponding stages of the encoder and decoder that allow information to propagate in the network at multiple scales and allow more efficient back-propagation of gradients. The Attention U-Net²⁴ adds to the skip connections *attention gates* that produce a map of coefficients in $[0, 1]$ for every pixel in a given feature map which is multiplied by that map of coefficients, therefore forcing the model to highlight the relevant regions in the feature map. The attention mechanism is useful by itself for improving the model performance, but they can also be used for gaining insight into what regions the model focuses on for making the prediction,²⁶ although here they do not make use of this aspect of the model.

We construct here a variant of the attention-gate layer that modifies the size of the feature map in z into a fixed size. Suppose we have a feature map X in the middle of the network with size K in the z -dimension, then we produce a new feature map X' with size K' by applying the operation

$$X'_k = \sum_{k=1}^K \sigma(f_k(X))_k \odot X_k \quad \forall k' \in \{1 \dots K'\} \quad (1)$$

where f_k is a convolution block with unique weights for each z -layer k' in the output and with padding to retain the z -size throughout, $\sigma: \mathbb{R} \rightarrow [0, 1]$ is an activation function, and \odot denotes element-wise multiplication. Here we choose to use 3 layers and zero-padding in the convolution block, and we use the sigmoid activation function $\sigma(z) = 1/(1 + \exp(-z))$. We add this layer to the output of the encoder and all of the skip connections so that the decoder can work with fixed z -size feature maps at all scales. The value for K' is a hyperparameter for which we choose here the values 3 for the encoder output and 3, 5, and 10 for the skip-connection outputs from the smallest to largest scale. The encoder does not use any pooling in the z -dimension so that the input AFM image stack can even have just a single z -layer.

Graph Construction. The original graph construction model¹⁵ works in an iterative way, taking one of the predicted atom positions at a time and adding a corresponding node to the graph with associated edges corresponding to chemical bonds. The node labeling process is informed by a channel of information coming from a CNN that is shared with the U-Net. The structure of the model creates

some problems for the predictions. First, the iterative nature of the model makes the predictions dependent on the order in which the atoms are added to the graph. Second, the shared weights between the two networks make the training process more difficult because the two prediction tasks need to be balanced at the same time. Additionally, the way that the GNN uses absolute coordinates of the atoms makes the predictions dependent on the chosen coordinate system and the exact size of the AFM images, and especially makes the predictions close to the edge of the image less reliable.¹⁵

Hence, we modify the graph construction network here in a way that addresses all of these issues (see the schematic at the bottom of Figure 2). Instead of processing the whole stack of AFM images as a whole, we instead choose small patches from the AFM images for each of the predicted atoms based on the proximity of the atom coordinates to the coordinates of the AFM image pixels. The patches are processed with a CNN to produce fixed-size feature vectors for each of the atoms, and these feature vectors are then used as initial hidden vectors for the nodes in a GNN. The GNN mixes the information between the nodes for multiple rounds along edge connections based proximity between the atoms, and finally the nodes are classified by a multilayer perceptron (MLP). This process simplifies the model by getting rid of the iteration for the nodes and makes it one-shot instead, with the AFM features gathered locally in a way that makes the predictions independent of the lateral size of the AFM image. Additionally, only relative coordinates are used within the network, so that the choice of the origin of the coordinate system is arbitrary.

Starting with the AFM images, let $\{r^q\}_{q=1}^N$ be the set of coordinates for the atoms produced by the U-Net model, and let $\{R_{ij}\}$ be the set of xy -coordinates for the voxels of the AFM image stack. For each of the atom coordinates r_q , we gather from the stack of AFM images a square patch of voxels whose xy -coordinates are within a cutoff distance d_a : $\|R_{ij} - r_q\|_{\infty} \leq d_a$. We use the value $d_a = 1.125 \text{ \AA}$ for the cutoff, which with a pixel resolution of 0.125 \AA creates patches of size 19×19 . In the case in which the atom is close to the edge of the image, the image is padded with zeros so that the patches have a constant size. This produces a set of smaller AFM images, which are processed by a CNN to produce a fixed size feature vector for each atom. The CNN has three ResNet²⁷ blocks with 2×2 pooling after each block, an attention gate layer similar to eq 1 to reduce the 3D feature map down to just a single voxel, and a one final fully connected layer that transforms the feature vector size to the one used inside the GNN. The three ResNet blocks have 12, 24, and 48 channels, respectively, and all have 2 layers.

The final feature vectors from the preceding CNN are used as the initial hidden vectors h_v^0 for each node labeled with $v \in \{1 \dots N\}$. The hidden vectors are updated for n_t iterations by a message-passing GNN:²⁸

$$m_{vu}^t = f_m(h_v^{t-1}, h_u^{t-1}, r^u - r^v) \quad \forall u \in \mathcal{N}(v), \quad \forall v \in \{1 \dots N\} \quad (2)$$

$$h_v^t = f_h\left(h_v^{t-1}, \sum_{u \in \mathcal{N}(v)} m_{vu}^t\right) \quad \forall v \in \{1 \dots N\} \quad (3)$$

where $t \in \{1 \dots n_t\}$, f_m is an MLP, f_h is a gated recurrent unit (GRU),²⁹ and $\mathcal{N}(v)$ denotes the set of neighbors for the node v . Note that the message function in eq 2 uses the relative coordinates between the nodes, which makes the update iteration translationally invariant. The set of neighbors of a node is decided based on proximity to other nodes: given a cutoff distance d_e , the set of neighbors for a node v is $\mathcal{N}(v) = \{u \mid \|r^u - r^v\|_2 \leq d_e\}$. Here we use the value $d_e = 3 \text{ \AA}$ for the cutoff, which is enough to capture all possible bonding distances between the atoms, including hydrogen bonding. Additionally, we choose $n_t = 5$, $|h_v^t| = |m_{vu}^t| = 40$, and f_m has two hidden layers of size 196.

The final classification of the node types is done by another MLP f_c :

$$c_v = f_c(h_v^{n_t}) \quad \forall v \in \{1 \dots N\} \quad (4)$$

The final layer of f_c is followed by a softmax activation so that y_v is a probability distribution over the node classes. The loss for the classification task is the cross entropy loss, $L(c_v, c'_v) = -\sum_{i=1}^C c'_{vi} \log c_{vi}$ where C is the number of classes and c'_v is a one-hot vector for the ground-truth class of node v . We choose to use one hidden layer of size 196 in f_c .

In addition to labeling the nodes, the edge connections between the nodes, corresponding to the bonds between the atoms, can be constructed. The bonds are not used here in practice for the subsequent simulations, but the method is described here for generality. The basic idea is to take the neighbor connections $\mathcal{N}(v)$ between the nodes and do a binary classification for each one on whether it corresponds to an edge in the final graph or not. To this end, in addition to maintaining a hidden vector for each node, we also maintain a hidden vector g_{uv}^t for each (unordered) pair of neighboring nodes $(u, v) \in E = \{(u, v) \mid u \in \mathcal{N}(v)\}$. The hidden vector is initialized to the average of the node hidden vectors, $g_{uv}^0 = (h_u^0 + h_v^0)/2$, and then updated on each iteration of the GNN as

$$g_{uv}^t = f_g(g_{uv}^{t-1}, (m_{uv}^t + m_{vu}^t)/2) \quad \forall (u, v) \in E \quad (5)$$

where $t \in \{1 \dots n_t\}$, and f_g is a GRU. The final classification of the edge connections is done as

$$e_{uv} = f_e(g_{uv}^{n_t}) \quad \forall (u, v) \in E \quad (6)$$

where f_e is an MLP with sigmoid activation in the final layer. The loss for the prediction is the binary cross-entropy loss. Like with the node classifier, we use one hidden layer of size 196 in f_e . The activation function for all of the layers in the model is the ReLU function.

Model Training. During training, the AFM images are preprocessed in several ways following previous work,¹³ including normalization, random noise, cutouts, pixel shifts, random background gradients, and random rotations and reflections. Here we also randomize the z -size of the AFM image stack between 1 and 15 slices and randomize the starting slice between 1 and 5. The simulations additionally have 10 different random parameter sets for each sample, as described above. The samples are divided into several shards, and for each training epoch, one of the parameter sets is chosen at random for each shard.

The graph construction model is trained separately from the position prediction model. In order to account for the fact that the predicted positions have some uncertainty in them, we add Gaussian noise ($\sigma = 0.08 \text{ \AA}$) to the input node positions during training of the graph construction model. The model parameters are optimized with the Adam optimizer,³⁰ using the default momentum parameters. The models are trained for 1000–1500 epochs until the loss does not improve significantly anymore. The final model parameters are chosen from the epoch with the lowest validation loss.

It may be of interest that we also tried first training the models on a larger data set with more elements¹³ and then fine-tuning on the water-only data set. However, we found in practice that these models did no better or worse than ones trained from the beginning on the water data set, as measured by the training and validation losses. Finally, while the assumption of intact water molecules is not directly enforced in the ML model, only neutral molecules were considered in the training for simplicity.

DFT Calculations. DFT calculations were conducted with the Vienna Ab Initio Simulation Package (VASP),^{31,32} modeling core electrons with projector augmented wave (PAW) potentials and expanding valence electrons with plane-waves with $E_{\text{cutoff}} = 500 \text{ eV}$. The nonlocal van der Waals-density functional optB86b-vdW-DF^{33–35} was utilized, as it has been shown to accurately describe the adsorption of water molecules on metal substrates.^{4,5,9,10,36}

Depending on the size of the water cluster in consideration, a 9×9 , 11×11 , or 13×13 Cu/Au(111) supercell was selected, using 3 atomic layers and a vacuum separation of 20 \AA along the slab perpendicular direction. Convergence tests of the k -grid showed that

Γ point calculations were already accurate at the meV level for the considered supercells.

Note that the growth of small monolayer clusters and ice bilayers on Au(111) is only weakly affected by the commensurability with the substrate,⁶ so we did not consider the influence of the herringbone reconstruction.

Neural Network Potentials. NNPs were employed via NequIP,³⁷ which allows building of an E(3)-equivariant NNP from a reference data set of ab initio calculations. By leveraging euclidean neural networks from e3nn³⁸ and by utilizing both scalar and higher-order tensor atomic features, the NequIP architecture reaches state of the art accuracy and data efficiency.

Initially, two NequIP models were trained for Au(111) and Cu(111) on 700 structures exemplified in Figure S4, which provided a diverse variety of both high and low energy structures to avoid overfitting and making a robust interatomic potential. These preliminary NNPs were then used to geometry relax a series of randomized water clusters to build the ML data set, as shown in Figure S3.

The relaxations were performed with ASE³⁹ BFGS minimizer, using a force tolerance of 3 meV/Å. After carrying out single-point DFT runs on the relaxed clusters, the NNPs were finally retrained on both their initial data set and on a portion of these new structures, which were more representative of our target distribution compared to the initial data set. For both the Au(111) and the Cu(111), we utilized 1850 data points, a rotation order of $l = 2$, a batch size of 2, a learning rate of 0.0075 and a PerSpeciesL1Loss for the force loss term. For the Au(111) model, we obtained validation energy and force Mean Absolute Errors (MAE) of respectively 0.414 meV/atom and 0.00558 meV/Å (of which H_f_mae = 0.0147, O_f_mae = 0.0184, and Au_f_mae = 0.00362). For the Cu(111) model, the final MAEs were instead 0.859 meV/atom for the validation energy and 0.00950 meV/Å for the validation force (of which H_f_mae = 0.0160, O_f_mae = 0.0215, and Cu_f_mae = 0.00789).

ASSOCIATED CONTENT

Data Availability Statement

All relevant data supporting the findings of this study are available in the Supporting Information. The source code and training data for the machine learning models is available under <https://github.com/SINGROUP/ml-spm>.

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acsnano.3c10958>.

Data sets, AFM images, profiles, simulations (ZIP)

Full sets of experimental AFM images, STM line profiles for each experiment, illustration of data set construction for ML and NNP training, investigation of the effect of AFM image stack size on ML loss, generation of hydrogen bond networks, vertical distribution of water molecules in the ML data set, extended comparisons of experimental data vs ML predictions and on-surface simulations (PDF)

AUTHOR INFORMATION

Corresponding Authors

Peter Liljeroth – Department of Applied Physics, Aalto University, Helsinki FI-00076, Finland; orcid.org/0000-0003-1253-8097; Email: peter.liljeroth@aalto.fi

Ying Jiang – International Center for Quantum Materials, Peking University, Beijing 100871, China; Collaborative Innovation Center of Quantum Matter, Beijing 100871, China; CAS Center for Excellence in Topological Quantum Computation, University of Chinese Academy of Sciences, Beijing 100190, China; Interdisciplinary Institute of Light-

Element Quantum Materials and Research Center for Light-Element Advanced Materials, Peking University, Beijing 100871, China; orcid.org/0000-0002-6887-5503; Email: yjiang@pku.edu.cn

Adam S. Foster – WPI Nano Life Science Institute (WPI-Nano LSI), Kanazawa University, Kanazawa 920-1192, Japan; Department of Applied Physics, Aalto University, Helsinki FI-00076, Finland; orcid.org/0000-0001-5371-5905; Email: adam.foster@aalto.fi

Authors

Fabio Priante – Department of Applied Physics, Aalto University, Helsinki FI-00076, Finland; orcid.org/0000-0001-7052-8570

Niko Oinonen – Department of Applied Physics, Aalto University, Helsinki FI-00076, Finland

Ye Tian – International Center for Quantum Materials, Peking University, Beijing 100871, China

Dong Guan – International Center for Quantum Materials, Peking University, Beijing 100871, China

Chen Xu – Department of Applied Physics, Aalto University, Helsinki FI-00076, Finland; orcid.org/0000-0003-2730-1025

Shuning Cai – Department of Applied Physics, Aalto University, Helsinki FI-00076, Finland

Complete contact information is available at:

<https://pubs.acs.org/doi/10.1021/acsnano.3c10958>

Author Contributions

F.P. and N.O. created the data, performed the machine learning and atomistic simulations, and wrote the first version of the manuscript. Y.T., D.G., C.X., and S.C. performed the experiments. All authors reviewed and commented on the manuscript. A.S.F. supervised the project.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors thank Darina Andriychenko for her contribution to data collection, and Fedor Urtev for his assistance in the ML analysis. This work was supported by World Premier International Research Center Initiative (WPI), MEXT, Japan, and by the Academy of Finland (Projects 347319, 347611, and 346824). The authors acknowledge the computational resources provided by the Aalto Science-IT Project and CSC, Helsinki.

REFERENCES

- (1) Morgenstern, M.; Michely, T.; Comsa, G. Anisotropy in the Adsorption of H₂O at Low Coordination Sites on Pt(111). *Phys. Rev. Lett.* **1996**, *77*, 703–706.
- (2) Yamada, T.; Tamamori, S.; Okuyama, H.; Aruga, T. Anisotropic Water Chain Growth on Cu(110) Observed with Scanning Tunneling Microscopy. *Phys. Rev. Lett.* **2006**, *96*, 036105.
- (3) Carrasco, J.; Michaelides, A.; Forster, M.; Haq, S.; Raval, R.; Hodgson, A. A one-dimensional ice structure built from pentagons. *Nat. Mater.* **2009**, *8*, 427–431.
- (4) Lew, W.; Crowe, M. C.; Campbell, C. T.; Carrasco, J.; Michaelides, A. The Energy of Hydroxyl Coadsorbed with Water on Pt(111). *J. Phys. Chem. C* **2011**, *115*, 23008–23012.
- (5) Forster, M.; Raval, R.; Hodgson, A.; Carrasco, J.; Michaelides, A. c(2×2) water-Hydroxyl Layer on Cu(110): A Wetting Layer Stabilized by Bjerrum Defects. *Phys. Rev. Lett.* **2011**, *106*, 046103.

- (6) Ma, R.; Cao, D.; Zhu, C.; Tian, Y.; Peng, J.; Guo, J.; Chen, J.; Li, X.-Z.; Francisco, J. S.; Zeng, X. C.; Xu, L.-M.; Wang, E.-G.; Jiang, Y. Atomic imaging of the edge structure and growth of a two-dimensional hexagonal ice. *Nature* **2020**, *577*, 60–63.
- (7) Maier, S.; Lechner, B. A. J.; Somorjai, G. A.; Salmeron, M. Growth and Structure of the First Layers of Ice on Ru(0001) and Pt(111). *J. Am. Chem. Soc.* **2016**, *138*, 3145–3151.
- (8) Michaelides, A.; Morgenstern, K. Ice nanoclusters at hydrophobic metal surfaces. *Nat. Mater.* **2007**, *6*, 597–601.
- (9) Liriano, M. L.; Gattinoni, C.; Lewis, E. A.; Murphy, C. J.; Sykes, E. C. H.; Michaelides, A. Water–Ice Analogues of Polycyclic Aromatic Hydrocarbons: Water Nanoclusters on Cu(111). *J. Am. Chem. Soc.* **2017**, *139*, 6403–6410.
- (10) Dong, A.; Yan, L.; Sun, L.; Yan, S.; Shan, X.; Guo, Y.; Meng, S.; Lu, X. Identifying Few-Molecule Water Clusters with High Precision on Au(111) Surface. *ACS Nano* **2018**, *12*, 6452–6457.
- (11) Gross, L.; Mohn, F.; Moll, N.; Liljeroth, P.; Meyer, G. The Chemical Structure of a Molecule Resolved by Atomic Force Microscopy. *Science* **2009**, *325*, 1110–1114.
- (12) Alldritt, B.; Hapala, P.; Oinonen, N.; Urtev, F.; Krejci, O.; Canova, F. F.; Kannala, J.; Schulz, F.; Liljeroth, P.; Foster, A. S. Automated structure discovery in atomic force microscopy. *Sci. Adv.* **2020**, *6*, No. eaay6913.
- (13) Oinonen, N.; Xu, C.; Alldritt, B.; Canova, F. F.; Urtev, F.; Cai, S.; Krejci, O.; Kannala, J.; Liljeroth, P.; Foster, A. S. Electrostatic Discovery Atomic Force Microscopy. *ACS Nano* **2022**, *16*, 89–97.
- (14) Carracedo-Cosme, J.; Romero-Muñiz, C.; Pou, P.; Pérez, R. QUAM-AFM: A Free Database for Molecular Identification by Atomic Force Microscopy. *J. Chem. Inf. Model.* **2022**, *62*, 1214–1223.
- (15) Oinonen, N.; Kurki, L.; Ilin, A.; Foster, A. S. Molecule graph reconstruction from atomic force microscope images with machine learning. *MRS Bull.* **2022**, *47*, 895–905.
- (16) Tang, B.; Song, Y.; Qin, M.; Tian, Y.; Wu, Z. W.; Jiang, Y.; Cao, D.; Xu, L. Machine learning-aided atomic structure identification of interfacial ionic hydrates from AFM images. *Nat. Sci. Rev.* **2023**, *10*, nwac282.
- (17) Hapala, P.; Kichin, G.; Wagner, C.; Tautz, F. S.; Temirov, R.; Jelínek, P. Mechanism of high-resolution STM/AFM imaging with functionalized tips. *Phys. Rev. B* **2014**, *90*, 085421.
- (18) Tobin, J.; Fong, R.; Ray, A.; Schneider, J.; Zaremba, W.; Abbeel, P. Domain Randomization for Transferring Deep Neural Networks from Simulation to the Real World. *arXiv* **2017**, 1703.06907 (<https://arxiv.org/abs/1703.06907>, accessed Dec 21, 2023).
- (19) Taori, R.; Dave, A.; Shankar, V.; Carlini, N.; Recht, B.; Schmidt, L. Measuring Robustness to Natural Distribution Shifts in Image Classification. *arXiv* **2020**, 2007.00644 (<https://arxiv.org/abs/2007.00644>, accessed Dec 21, 2023).
- (20) Hendrycks, D.; Basart, S.; Mu, N.; Kadavath, S.; Wang, F.; Dorundo, E.; Desai, R.; Zhu, T.; Parajuli, S.; Guo, M.; Song, D.; Steinhardt, J.; Gilmer, J. The Many Faces of Robustness: A Critical Analysis of Out-of-Distribution Generalization. *arXiv* **2020**, 2006.16241 (<https://arxiv.org/abs/2006.16241>, accessed Dec 21, 2023).
- (21) Wang, J.; Lan, C.; Liu, C.; Ouyang, Y.; Qin, T.; Lu, W.; Chen, Y.; Zeng, W.; Yu, P. Generalizing to Unseen Domains: A Survey on Domain Generalization. *IEEE Trans. Knowl. Data Eng.* **2022**, *35*, 8052–8072.
- (22) Carrasco, J.; Hodgson, A.; Michaelides, A. A molecular perspective of water at metal interfaces. *Nat. Mater.* **2012**, *11*, 667–674.
- (23) Chen, P.; Xu, Q.; Ding, Z.; Chen, Q.; Xu, J.; Cheng, Z.; Qiu, X.; Yuan, B.; Meng, S.; Yao, N. Identification of a common ice nucleus on hydrophilic and hydrophobic close-packed metal surfaces. *Nat. Commun.* **2023**, *14*, 5813.
- (24) Schlemper, J.; Oktay, O.; Schaap, M.; Heinrich, M.; Kainz, B.; Glocker, B.; Rueckert, D. Attention gated networks: Learning to leverage salient regions in medical images. *Medical Image Analysis* **2019**, *53*, 197–207.
- (25) Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv* **2015**, 1505.04597 (<http://arxiv.org/abs/1505.04597>, accessed Dec 21, 2023).
- (26) Ranawat, Y. S.; Jaques, Y. M.; Foster, A. S. Predicting hydration layers on surfaces using deep learning. *Nanoscale Adv.* **2021**, *3*, 3447–3453.
- (27) He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE, 2016; pp 770–778.
- (28) Gilmer, J.; Schoenholz, S. S.; Riley, P. F.; Vinyals, O.; Dahl, G. E. Neural Message Passing for Quantum Chemistry. *Proceedings of the 34th International Conference on Machine Learning*; PMLR, 2017; pp 1263–1272.
- (29) Cho, K.; van Merriënboer, B.; Bahdanau, D.; Bengio, Y. On the Properties of Neural Machine Translation: Encoder-Decoder Approaches. *arXiv* **2014**, 1409.1259.
- (30) Kingma, D. P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**, 1412.6980.
- (31) Kresse, G.; Furthmüller, J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mater. Sci.* **1996**, *6*, 15–50.
- (32) Kresse, G.; Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **1996**, *54*, 11169–11186.
- (33) Klimeš, J.; Bowler, D. R.; Michaelides, A. Van der Waals density functionals applied to solids. *Phys. Rev. B* **2011**, *83*, 195131.
- (34) Román-Pérez, G.; Soler, J. M. Efficient Implementation of a van der Waals Density Functional: Application to Double-Wall Carbon Nanotubes. *Phys. Rev. Lett.* **2009**, *103*, 096102.
- (35) Dion, M.; Rydberg, H.; Schröder, E.; Langreth, D. C.; Lundqvist, B. I. Van der Waals Density Functional for General Geometries. *Phys. Rev. Lett.* **2004**, *92*, 246401.
- (36) Guo, J.; Meng, X.; Chen, J.; Peng, J.; Sheng, J.; Li, X.-Z.; Xu, L.; Shi, J.-R.; Wang, E.; Jiang, Y. Real-space imaging of interfacial water with submolecular resolution. *Nat. Mater.* **2014**, *13*, 184–189.
- (37) 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.
- (38) Geiger, M.; Smidt, T. e3nn: Euclidean Neural Networks. *arXiv* **2022**, 2207.09453 (<https://arxiv.org/abs/2207.09453>, accessed Dec 21, 2023).
- (39) Larsen, A. H.; et al. The atomic simulation environment—a Python library for working with atoms. *J. Phys.: Condens. Matter* **2017**, *29*, 273002.