

# Machine-learning-accelerated simulations to enable automatic surface reconstruction

Received: 11 May 2023

Accepted: 13 November 2023

Published online: 7 December 2023

 Check for updates

Xiaochen Du<sup>1,2</sup>, James K. Damewood<sup>2,3</sup>, Jaclyn R. Lunger<sup>3</sup>, Reisel Millan<sup>3</sup>, Bilge Yildiz<sup>3,4,5</sup>, Lin Li<sup>1,6</sup> & Rafael Gómez-Bombarelli<sup>1,3</sup> 

Understanding material surfaces and interfaces is vital in applications such as catalysis or electronics. By combining energies from electronic structure with statistical mechanics, ab initio simulations can, in principle, predict the structure of material surfaces as a function of thermodynamic variables. However, accurate energy simulations are prohibitive when coupled to the vast phase space that must be statistically sampled. Here we present a bi-faceted computational loop to predict surface phase diagrams of multicomponent materials that accelerates both the energy scoring and statistical sampling methods. Fast, scalable and data-efficient machine learning interatomic potentials are trained on high-throughput density-functional-theory calculations through closed-loop active learning. Markov chain Monte Carlo sampling in the semigrand canonical ensemble is enabled by using virtual surface sites. The predicted surfaces for GaN(0001), Si(111) and SrTiO<sub>3</sub>(001) are in agreement with past work and indicate that the proposed strategy can model complex material surfaces and discover previously unreported surface terminations.

Surface structure determines the properties and performance of materials in application areas such as heterogeneous catalysis<sup>1,2</sup>, electrocatalysis<sup>3–5</sup> and electrochemical energy storage<sup>6–8</sup>. Material surfaces are not just pristine cuts of the bulk structure, and even for a single surface facet, equilibrium reconstruction can lead to vastly different terminations and patterns as a function of temperature, external chemical potentials and applied electrical potential<sup>9–13</sup>. Experimental methods for studying surfaces at the atomic level are costly and involved so they cannot cover the wide range of experimental conditions<sup>13,14</sup>.

Simulations have the potential to capture complex surface structures at a wide range of external conditions. To do so, accurate and computationally affordable surface-energy predictions are needed, along with efficient statistical sampling across surface compositions and configurations. On the energy modeling front, while classical interatomic force fields with few fitted parameters are sufficient for simple surfaces such as those of gold and gallium nitride (GaN)<sup>15,16</sup>,

their simple functional forms are often unsuited for multicomponent surfaces.

Accurate computational studies of complex surfaces have relied on energetics derived from expensive density-functional-theory (DFT) simulations of human-input guess surfaces, but this strategy does not easily scale to the diversity of possible structures and cannot gather enough statistics for converging thermodynamic averages. A previous study manually constructed candidate surfaces of BaTiO<sub>3</sub>(001) with various coverage levels comprising vacancies and adsorbates of Ba, Ti and O based on chemical intuition and previous experimental data<sup>17</sup>. Then DFT relaxations were used to produce an energy-based phase diagram that connected external conditions to surface structures. Nevertheless, human intuition does not guarantee sufficient exploration of the phase space to uncover the most thermodynamically stable structures, and cannot capture the role of entropy through single static structures<sup>18–20</sup>.

<sup>1</sup>Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA. <sup>2</sup>Center for Computational Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA. <sup>3</sup>Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA. <sup>4</sup>Department of Nuclear Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA. <sup>5</sup>Microsystems Technology Laboratories, Massachusetts Institute of Technology, Cambridge, MA, USA. <sup>6</sup>Massachusetts Institute of Technology Lincoln Laboratory, Lexington, MA, USA.  e-mail: [rafagb@mit.edu](mailto:rafagb@mit.edu)

Global optimization techniques such as basin hopping<sup>18,21–25</sup>, evolutionary algorithms<sup>25,26</sup>, random structure search<sup>27</sup>, reinforcement learning<sup>28</sup> and simulated annealing<sup>29</sup> offer a principled approach to enhance the generalizability of such computational studies, enabling the discovery of unexpected phases in materials science. Nonetheless, these methods often entail trade-offs among the explored phase space, energy accuracy and computational cost. In addition, the focus is on finding energy minima, not free-energy minima. A previous study achieved a balance between energy accuracy and computational cost, but the exploration of compositions while investigating SrTiO<sub>3</sub> grain boundaries was somewhat limited<sup>27</sup>. Two studies managed extensive sampling across different compositions<sup>18,26</sup>. However, their reliance on computationally expensive DFT increased computational costs despite maintaining high energy accuracy. Meanwhile, in a study of La<sub>0.75</sub>Sr<sub>0.25</sub>MnO<sub>3</sub>(001), costs were reduced by utilizing electrostatic energy calculations derived from Bader charges and limiting the phase space using canonical Monte Carlo (MC)<sup>29</sup>. However, this lower-fidelity approach does not assure the stability of final structures.

Machine learning (ML) force fields are much faster than DFT calculations while preserving accuracy<sup>30,31</sup> but it has been challenging to connect them to an efficient sampling scheme to explore the phase space of multicomponent surfaces in an self-directed fashion. Recent studies have introduced reconstruction pipelines leveraging ML force fields, although many require guess compositions and lack automatic consideration of external conditions in their sampling procedure<sup>14,20,32–36</sup>. Furthermore, although ML-based sampling methods that vary surface compositions have demonstrated potential, they are currently limited to single adsorbate types on monocomponent surfaces and may employ less efficient sampling schemes<sup>37</sup>.

Here we present our Automatic Surface Reconstruction (AutoSurfRecon) framework, which achieves thorough statistical sampling of thermodynamic states and avoids relying on hand-picked trial surfaces while being computationally efficient. AutoSurfRecon utilizes ML force fields and active learning (AL) to enable fast and accurate energetics. In addition, because surface reconstruction typically takes place on sites that emerge from the underlying slab, but not necessarily following slab symmetry, we utilize the computationally efficient Virtual Surface Site Relaxation-Monte Carlo (VSSR-MC) sampling algorithm in the semigrand canonical ensemble, instead of the more intensive grand canonical MC. By populating virtual sites followed by relaxation, VSSR-MC samples across compositional and configurational search spaces to efficiently and accurately explore complex, stable surface structures across a range of external chemical potentials. To validate our sampling strategy based on virtual sites, we recover known reconstructions of the well-studied GaN(0001) and Si(111) surfaces using classical force fields. We then show the effectiveness of the full AutoSurfRecon pipeline on SrTiO<sub>3</sub>(001), a complex perovskite unsuited for classical force fields, using a neural network force field (NFF) energy model. We demonstrate that our uncertainty-driven AL strategy working in tandem with VSSR-MC sampling acquires new DFT data points only at relevant regions of the surface phase space. In fewer than 5,000 DFT single-point calculations, the NFF can be trained to obtain accurate energy predictions for many distinct chemical compositions. Finally, we construct an SrTiO<sub>3</sub>(001) surface phase diagram that compares well with literature results and reveals unexpected low-energy surface terminations. Our algorithm can be easily applied to other surfaces of interest and we anticipate that it can be used to study multicomponent materials under challenging environments such as in aqueous electrochemistry.

## Results

### Development of the end-to-end automatic framework

Figure 1 shows our AutoSurfRecon computational workflow. Starting solely with a clean-cut surface and algorithmically generated virtual adsorption sites as inputs (see ‘Surface slab modeling’ in Methods), a surface phase diagram is produced. In this workflow, VSSR-MC-sampled

structures have the dual purpose of improving the NFF energy model and obtaining surface reconstruction statistics, further improving the efficiency of our pipeline.

The phase diagram is obtained through successive surface reconstruction runs that, beginning with the pristine surface as input, explore the surface phase space using VSSR-MC. Runs are conducted at different elemental chemical potentials.

Every individual VSSR-MC sampling iteration starts with discretely choosing one elemental identity at a selected adsorption site. The absorption sites are created algorithmically from the pristine slab geometry. A key innovation in VSSR-MC is that empty sites are defined as virtual atoms thus avoiding the higher volume space of grand canonical sampling in Cartesian coordinates. As such, adding an atom or removing an existing atom become instances of changing elemental identity and close contacts corresponding to very high repulsive energies can be easily rejected based on hard-sphere cut-offs.

Following discrete sampling, all adsorbates and surface atoms undergo continuous relaxation of atomic positions away from their assigned sites (see ‘Virtual Surface Site Relaxation-Monte Carlo’ in Methods), allowing us to efficiently explore the vast configurational space as lower-energy relaxations of easy-to-sample virtual sites. The energy model used is a classical force field for mono- and bi-component materials or a machine-learned force-field model for complex materials. The structure obtained after discrete sampling and continuous relaxation is evaluated with the acceptance criterion corresponding to the semigrand canonical ensemble<sup>38,39</sup>:

$$P = \min \left\{ 1, \exp \left( -\frac{\Delta E_{\text{slab}} - \Delta \mu}{k_B T} \right) \right\} \quad (1)$$

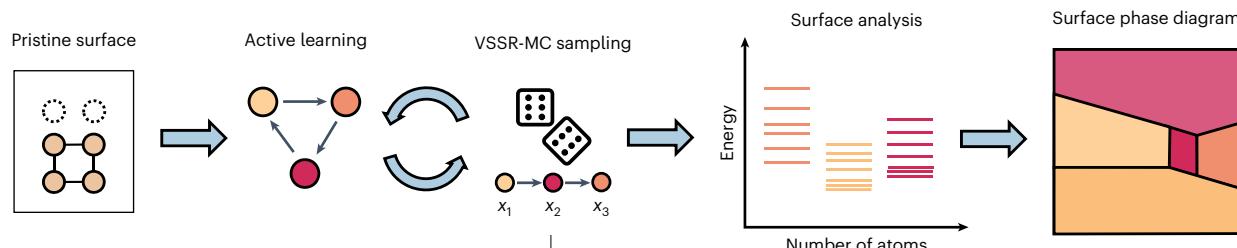
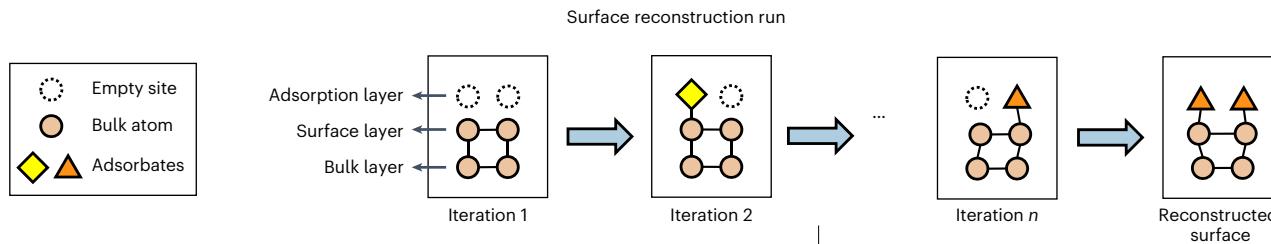
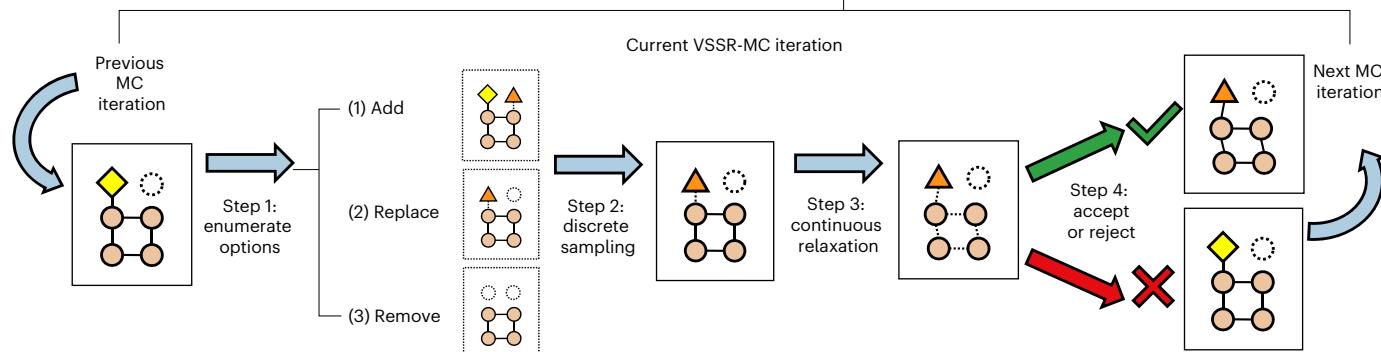
where  $P$  is the probability of acceptance,  $\Delta E_{\text{slab}}$  is the change in slab energy after both discrete sampling and continuous relaxation,  $\Delta \mu$  is the change in chemical potential due to discrete sampling,  $k_B$  is the Boltzmann constant and  $T$  is the sampling temperature.

We also developed an AL strategy (Fig. 2) to efficiently acquire the fewest possible DFT-evaluated structures required for NFF training. At each AL iteration, we train an ensemble of NFF models for uncertainty quantification<sup>40,41</sup> (see ‘Neural network force field’ in Methods). To iteratively select structures for DFT evaluation, we use uncertainty-based adversarial attack<sup>42</sup> or VSSR-MC combined with latent space clustering to seek out structures that maximize NFF prediction error, which in practice is estimated by the predicted force standard deviation (s.d.) (see ‘Active learning’ in Methods). The adversarial attack algorithm displaces atomic positions of existing structures to regions of high s.d. but not high energy<sup>42</sup>, thereby improving the accuracy and stability of the force field<sup>43</sup>. However, adversarial attacks cannot autonomously sample different compositions. VSSR-MC samples both new compositions and configurations across chemical potentials and guides the NFF to learn only relevant subsets of the vast phase space. At the same time, clustering candidate structures reduces the number of training data by selecting only those whose local chemical environments are unique<sup>44</sup>.

### Reconstructions with classical potentials

To demonstrate our VSSR-MC sampling method, we investigated the  $3 \times 3$  reconstruction of the GaN(0001) surface using a Tersoff potential<sup>16</sup> and the  $3 \times 3$ ,  $5 \times 5$  and  $7 \times 7$  reconstructions of the Si(111) surface using a modified Stillinger–Weber potential<sup>45</sup>.

**GaN(0001) reconstruction.** GaN is a well-studied semiconductor and the (0001) surface is described as having a contracted adsorption layer on a rotated lattice with respect to the pristine surface<sup>46</sup>. When running the VSSR-MC algorithm at a fixed number of Ga adsorbates, several reconstructions matching the literature were obtained (Fig. 3). When viewed from the top, our structure in Fig. 3a shows the same rhombus patterns as the literature structure. The adsorbate distance from the

**a****b****c**

**Fig. 1 | AutoSurfRecon framework.** **a**, Beginning with a pristine surface and computer-generated virtual adsorption sites, VSSR-MC sampling is conducted in tandem with AL of an NFF. The MC nature of VSSR-MC is denoted by a pair of dice, and its Markov chain characteristic is illustrated by sequential changes from structure  $x_1$  to  $x_2$  followed by  $x_3$ . Following multiple rounds of surface reconstruction runs, the surface free energy of sampled structures are separated by their composition and analyzed at different external conditions. Finally, the surface phase diagram is constructed, with each color representing the dominant surface across a set of external conditions. **b**, The proposed surface reconstruction run consists of multiple iterations of VSSR-MC. Starting from

the same pristine surface with empty virtual sites, a series of iterations alter both the atomic identities at adsorption sites and the relative atomic positions within the structure. **c**, In each VSSR-MC iteration, there are three steps. Starting from a surface taken from the previous iteration, in step 1, possible options for discrete sampling on a single adsorption site are enumerated. The possible actions are: add, replace and remove. In step 2, the chosen discrete sampling action is performed to produce a candidate structure. This mutation is followed by continuous relaxation in step 3. For step 4, an energy model is required, such as an NFF or a classical force field. Finally, in step 4, the candidate structure is evaluated using an MC acceptance criterion.

pristine surface layer matched the literature value of  $2.42 \text{ \AA}$  (side view, Fig. 3b) and the energy difference stood at a mere  $0.008 \text{ meV per atom}$ . The energy approached the ground state and VSSR-MC acceptance rate neared 0 at around 20,000 to 25,000 iterations in Fig. 3c. Additional reconstructed surfaces can be found in Extended Data Fig. 1.

well as various  $7 \times 7$  pristine-like structures with four adatoms (due to stoichiometry differences) that are  $3.5 \text{ meV per atom}$  higher in energy than the DAS target. By discovering these unexpected low-energy structures alongside DAS structures, VSSR-MC demonstrated the ability to thoroughly sample a large phase space of surface reconstructions.

**Si(111) reconstructions.** Si(111) is known to show complex dimer-adatom stacking fault (DAS) surface reconstructions that vary with supercell size<sup>47,48</sup>. These reconstructions follow neither bulk lattice geometries nor virtual site geometries. We separately investigated the  $3 \times 3$ ,  $5 \times 5$  and  $7 \times 7$  surfaces and recovered the DAS reconstructions in our VSSR-MC runs at a fixed adsorbate number for each supercell (Extended Data Fig. 2). Multiple DAS-like structures that are within the thermally accessible energy range ( $25.7 \text{ meV per atom}$ ) at room temperature were obtained for all supercell sizes—two of such  $3 \times 3$  structures are shown in Extended Data Fig. 2a and one of each for  $5 \times 5$  and  $7 \times 7$  is presented in Extended Data Fig. 2b,c, respectively. Moreover, VSSR-MC also sampled the  $5 \times 5$  pristine surface, which is within  $2.1 \text{ meV per atom}$  of the DAS structure according to the potential, as

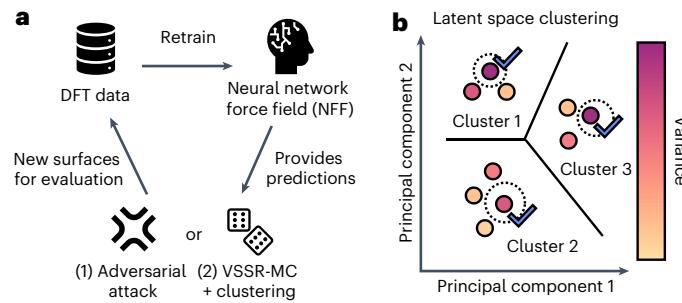
### SrTiO<sub>3</sub>(001) reconstructions with NFF

We validated the performance of our full AutoSurfRecon framework on a challenging surface for which there is no known analytical potential. We chose SrTiO<sub>3</sub>(001) because SrTiO<sub>3</sub> is representative of the complex perovskite oxide family and the (001) surface, in particular, is stable and demonstrates a variety of surface reconstruction patterns under different elemental chemical potentials<sup>9–12</sup>.

**AL for NFF.** Including the initial dataset, AL was run for a total of six iterations (see ‘Active learning’ in Methods for details) on SrTiO<sub>3</sub>(001) slabs of varying compositions (example in Fig. 4a). A total of 6,500 structures were selected for DFT evaluation: the first and last AL generations resulted in about 1,500 structures each while AL generations

2–5 yielded approximately 800 structures each. Across our AL runs, we found a good correlation between force mean absolute error (MAE) and predicted force s.d., demonstrating the validity of our error estimation procedure. Figure 4b shows one such correlation plot derived from the final NFF model and sixth-generation structures. Correlation plots across all generations are provided in Extended Data Fig. 3.

Figure 4c shows a principal component analysis (PCA), two-dimensional projection of all new surface structures generated through our AL process. To create a consistent representation across generations, PCA was applied to neural network embeddings from the last-generation model on the full dataset. In Fig. 4c, we see a distinct pattern of VSSR-MC structures compared with the initial structures, with VSSR-MC structures evolving over generations 3–5 to show streaks. These patterns suggest that with NFF improvement, VSSR-MC was able to more effectively sample structures corresponding to surface-energy minima. We also observe a correlation of the predicted energy with the first two principal axes (see training objective in ‘Neural network force field’ in Methods). Additional information on the distribution of forces obtained at each AL generation is provided in Extended Data Fig. 4.



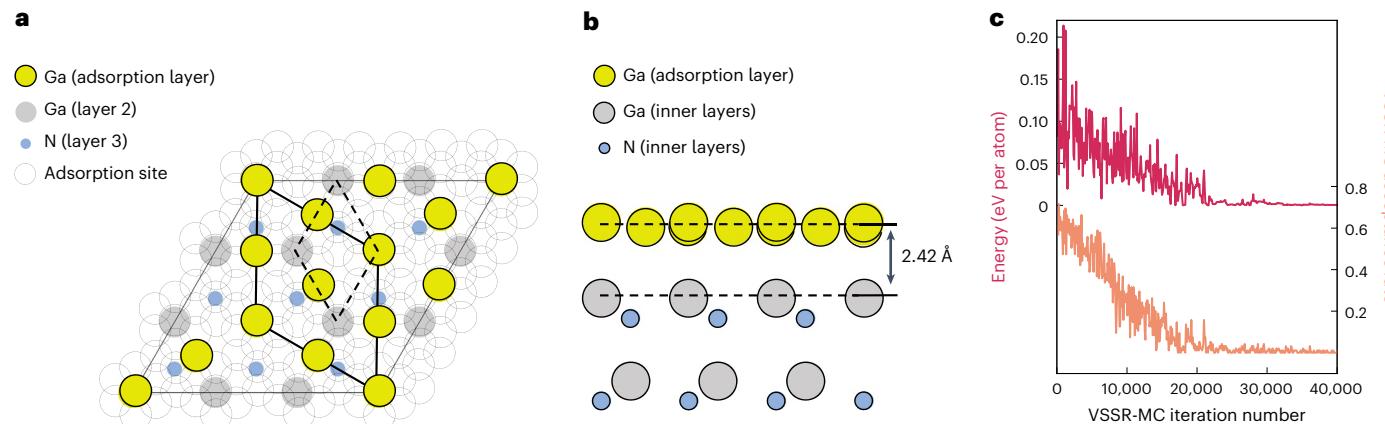
**Fig. 2 | AL procedure for NFF.** **a**, An initial ensemble of NFF models is trained using a common dataset of available DFT data. Using the NFF to provide predictions of forces, energies and their s.d., either adversarial attack or VSSR-MC with latent space clustering is used to generate new surfaces for DFT evaluation. After performing DFT calculations on these new structures, they are added to the data pool and the NFF is retrained. **b**, Latent space clustering procedure for MC-generated structures. Structures are clustered according to the first few principal components of latent space embeddings and the structure with the highest force s.d. in each cluster is selected for DFT evaluation.

Figure 4d shows the performance of the NFF model, measured by energy MAE and force MAE, after each AL iteration over common test data. The force and energy MAE drop substantially within four AL iterations, suggesting the effectiveness of the AL process for sampling the most informative surfaces for ML model training. While the overall trend of both MAEs is down, the increase in energy MAE during generation 2 can be attributed to the distribution mismatch between the starting structures and the MC-sampled structures, suggesting that the first-generation model was overfit to the initial dataset and struggled to generalize to the much broader and realistic phase space explored during the AL cycles. The decrease from generation 3 to 4 was the most dramatic, and in generation 4, an improved NFF allowed VSSR-MC samples to more closely follow the underlying distribution, resulting in the performance improvement.

Overall, after data splitting, the final NFF was fitted on a training set with fewer than 5,000 structures and achieved a force MAE of  $0.10 \text{ eV } \text{\AA}^{-1}$  and an energy MAE of 5.18 meV per atom (Extended Data Fig. 5) across the phase space needed to power production VSSR-MC runs. It is important to note that the number of AL generations is not fixed, but is determined by the observed plateau in NFF accuracy improvements and the convergence of VSSR-MC structures. The goal of our AL methodology is to optimize computational resources while attaining good prediction accuracy for a diverse set of structures. Future research may include the exploration of more precise or automated stopping criteria, such as the implementation of force or energy thresholds.

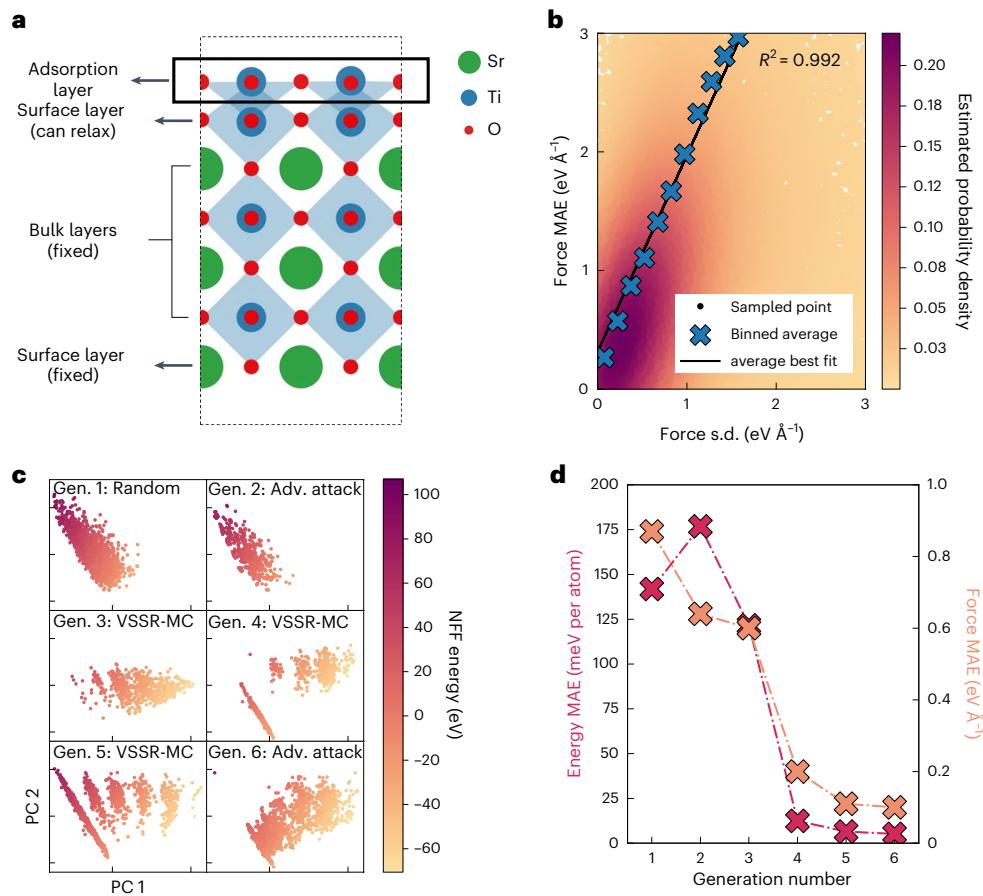
**Data analysis and constructing the phase diagram.** We first present a summary of literature data on  $\text{SrTiO}_3(001)$  reconstructions. The structure of  $\text{SrTiO}_3(001)$  has inspired intensive research and debate. A double-layer  $\text{TiO}_2$ -terminated surface is supported by both experiments and theory<sup>11,49,50</sup>. But single-layer  $\text{TiO}_2$  and single-layer  $\text{SrO}$  terminations have also been reported<sup>9,10,51,52</sup>. All these surfaces can be related to the chemical potential of the constitutive elements in the material bulk, which is assumed to be in equilibrium with the environment. In this case, the Sr chemical potential,  $\mu_{\text{Sr}}$ , is the most relevant. By increasing  $\mu_{\text{Sr}}$ , which can be achieved by evaporating Sr metal into the reaction environment, adding Sr to the surface becomes more favorable, resulting in the depletion of Ti and finally the formation of an  $\text{SrO}$  adlayer<sup>9,12</sup>.

In our surface reconstruction runs, we observed all three terminations starting from double-layer  $\text{TiO}_2$  in low  $\mu_{\text{Sr}}$  to single-layer  $\text{TiO}_2$  (that is, the unreconstructed surface) in intermediate  $\mu_{\text{Sr}}$ , and single-layer



**Fig. 3 | Obtaining the GaN(0001) contracted-monolayer surface reconstruction using VSSR-MC and a classical potential.** **a**, Top view of surface reconstruction. Only the first three layers are shown. Our structure is visually identical to the literature optimized structure from ref. 46 with an energy difference of 0.008 meV per atom. **b**, Side view of the reconstructed surface. The

average distance of the first two layers is  $2.42 \text{ \AA}$ , in agreement with the literature. The dashed lines are a guide for the eye. **c**, Typical VSSR-MC run profile for GaN(0001). Annealing from a high temperature allows convergence to a low-energy structure.



**Fig. 4 | AL for NFF.** **a**, Example SrTiO<sub>3</sub>(001) surface for training NFF. The surface consists of an adsorption layer with variable chemical compositions and configurations, and inner layers with fixed compositions. The blue diamonds represent the octahedral of O atoms surrounding a Ti atom. **b**, Correlation plot of force MAE with force s.d. The binned average is calculated by dividing both the force s.d. and force MAE into equal-size bins. The average force MAE is then plotted against the median force s.d. for each corresponding bin. The average

best-fit line is based on the binned average values and the trend implies force s.d. is a fair surrogate for estimating force MAE. **c**, PCA of surfaces obtained at each AL generation. Starting with generation (gen.) 2, a mix of adversarial (adv.) attack and VSSR-MC with latent space clustering were used. **d**, Retrospective performance of NFF trained using all surfaces available at each AL generation. Performance was measured using a test set consisting of VSSR-MC structures sampled from the sixth-generation NFF model to reflect our use case.

SrO in high  $\mu_{\text{Sr}}$ . A schematic is provided in Fig. 5a to illustrate this relationship. Given our choice of modeling SrTiO<sub>3</sub>(001), the pristine surface mirrored the stoichiometry of the bulk. As such, alterations to  $\mu_{\text{Sr}}$  influenced surface free energies in the same manner as equivalent changes in  $\mu_{\text{O}}$ , the O chemical potential, eliminating the need for additional runs varying  $\mu_{\text{O}}$  (as referenced in equation (7)).

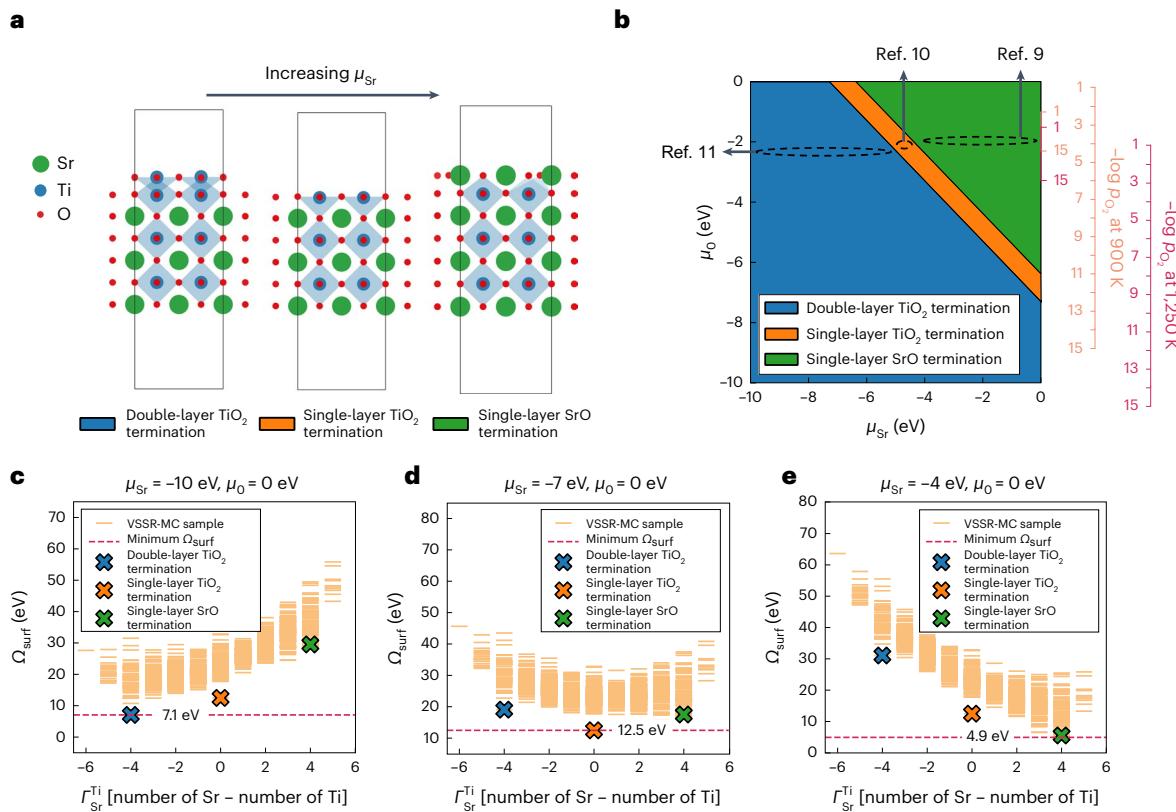
Analyzing our data obtained from VSSR-MC, we constructed a surface phase diagram of SrTiO<sub>3</sub>(001) in Fig. 5b that maps chemical potentials of Sr and O to the most stable surface terminations. Our phase diagram matches the expected trend in  $\mu_{\text{Sr}}$ . It is also similar to the one obtained by ref. 12 with a narrow strip of single-layer TiO<sub>2</sub> phase sandwiched between a double-layer TiO<sub>2</sub> phase at substantial oxygen vacancies (low  $\mu_{\text{O}}$ ) and low  $\mu_{\text{Sr}}$ , and a single-layer SrO phase at considerable Sr concentrations in an oxygen atmosphere (high  $\mu_{\text{O}}$  and  $\mu_{\text{Sr}}$ ). We additionally assign three experimental SrTiO<sub>3</sub>(001) surfaces to our phase diagram, taking into account  $\mu_{\text{Sr}}$  is loosely related to experimental procedures while  $\mu_{\text{O}}$  can be calculated from  $p_{\text{O}_2}$ , the partial pressure of O<sub>2</sub> gas, and the experimental temperature.

To construct the phase diagram, the surface free energy ( $\Omega_{\text{surf}}$ ) was recalculated for each structure using the final NFF model (see ‘Surface stability analysis’ in Methods) at various  $\mu_{\text{Sr}}$  and we plotted  $\Omega_{\text{surf}}$  against the difference in the number of Sr and Ti atoms ( $T_{\text{Sr}}^{\text{Ti}}$ ) for each slab. In the plots for  $\mu_{\text{Sr}} = -10$  eV in Fig. 5c,  $\mu_{\text{Sr}} = -7$  eV in Fig. 5d and  $\mu_{\text{Sr}} = -4$  eV in Fig. 5e, we see that structures near the minimum  $\Omega_{\text{surf}}$  correspond to

the three known terminations. Double-layer TiO<sub>2</sub> (Fig. 5c) and single-layer TiO<sub>2</sub> (Fig. 5d) had the lowest energy at  $\mu_{\text{Sr}} = -10$  eV and  $\mu_{\text{Sr}} = -7$  eV respectively. Single-layer SrO (Fig. 5e) had the second lowest energy at  $\mu_{\text{Sr}} = -4$  eV, and became the lowest after filtering for stoichiometry.

**Comparing double-layer TiO<sub>2</sub> terminations.** We additionally show that VSSR-MC faithfully recreated different reconstruction patterns of the double-layer TiO<sub>2</sub> surface. As reported by refs. 11,49 and others, double-layer TiO<sub>2</sub> does not consist of a single termination; 2 × 2, 2 × 1 and 1 × 1 terminations are possible and we describe these terminations in Fig. 6. The most common 2 × 2 termination in literature is denoted in Fig. 6 as 2 × 2-A. The dominant double-layer TiO<sub>2</sub> terminations could vary based on the exact surface structure and exchange-correlation functional, calculation settings, and empirical dispersion or Hubbard corrections. In this case, VSSR-MC samples contain two out of the three literature terminations: 2 × 2-A and 1 × 1. These two TiO<sub>2</sub> terminations are close to one another in stability, as in ref. 49.

The algorithm also discovered two surface terminations not previously reported, which we denote as 2 × 2-B and 2 × 2-C. The 2 × 2-C termination has a similar energy to the two observed literature terminations but 2 × 2-B is lower in energy. The 2 × 1 reconstruction reported in the literature was not observed during surface reconstruction runs. Energy prediction using NFF and confirmed using DFT demonstrate an energy more than 2 eV above that of the most stable surface (2 × 2-B)



**Fig. 5 | Analyzing dominant  $\text{SrTiO}_3(001)$  surface terminations observed during VSSR-MC runs at different chemical potentials.** **a**, Side view of change in dominant surface termination as  $\mu_{\text{Sr}}$  increases. The blue diamonds represent the octahedral of O atoms surrounding a Ti atom. **b**, Computed phase diagram showing the stable surface terminations at varying  $\mu_{\text{Sr}}$  and  $\mu_0$  along with estimated positions of three experimental  $\text{SrTiO}_3(001)$  surfaces<sup>9–11</sup>. On the right, four vertical axes are illustrated. The smaller axes on the phase diagram, ending at points 1 and 15, provide an abbreviated view. The larger scales extend from

1 to 15 at equal intervals, indicating  $-\log p_{\text{O}_2}$  values at 900 K (orange) and 1,250 K (red), offering a detailed perspective on the oxygen partial pressure across these temperatures. **c–e**, Surface free energy ( $\Omega_{\text{surf}}$ ) plots of sampled structures as a function of the difference in the number of Sr and Ti atoms. Plots shown correspond to  $\Omega_{\text{surf}}$  at various  $\mu_{\text{Sr}}$ :  $\mu_{\text{Sr}} = -10 \text{ eV}$  (**c**),  $\mu_{\text{Sr}} = -7 \text{ eV}$  (**d**) and  $\mu_{\text{Sr}} = -4 \text{ eV}$  (**e**).  $\mu_0 = 0 \text{ eV}$  in all three plots. The minimum energy surfaces are crossed out and correspond to those from literature.

Type	2 × 2-A	2 × 1	1 × 1	2 × 2-B	2 × 2-C
Top layer					
Occurrence					
DFT energy (eV)	1.116	2.422	1.306	0.000	1.442
NFF energy (eV)	0.988 ± 0.050	2.218 ± 0.344	0.948 ± 0.139	0.000 ± 0.151	1.139 ± 0.124

**Fig. 6 | Comparing DFT and NFF energies of double-layer  $\text{TiO}_2$  terminations, all with the same composition.** The NFF predicted energies are close to DFT energies. In addition, the previously unreported 2 × 2-B and 2 × 2-C terminations have roughly equal or lower energies than those of the other three literature-reported surfaces.

and almost 1 eV above that of the next highest energy termination (2 × 2-C). We also observe the consistency between NFF predictions and DFT energies, which again shows the accuracy of NFFs as an energy model.

## Discussion

The presented algorithm overcomes limitations of previous computational methods, as shown in Supplementary Data 1. We anticipate

VSSR-MC will be broadly applicable and aim to extend it to more challenging multicomponent solid/liquid interfaces under electrical potential.

VSSR-MC is advantageous because its trials are limited to high-likelihood virtual sites responsible for reconstructions that relate to the symmetry of the pristine slab. A disadvantage is that it is focused on thin, periodic reconstructions and may struggle to reconstruct amorphized thicker slabs that do not follow virtual site

geometries. A schematic on the strengths and limitations of our sampling approach is given in Extended Data Fig. 6.

In this work, we sampled only one adsorption layer of a complex oxide. Future work would benefit from an improved understanding of complex surface reconstructions as results of stoichiometric changes across multiple adsorption layers, or to vacancies in the bulk.

While our approach is not strictly constrained by the size of the unit cell, for  $\text{SrTiO}_3(001)$  especially, we chose to use a relatively small unit cell whereas larger ones could show more complex reconstruction patterns. Moving to a larger unit cell could require transfer learning from NFFs trained on existing data. Well-trained NFFs typically generalize well to larger symmetry-breaking supercells<sup>53,54</sup>. As with other methods for studying surface reconstruction, the choice of the unit cell should be informed by past experimental and computational studies.

As is common in the field, we approximated the surface free energy directly from DFT energies. However, there could be instances where vibrational contributions to the surface free energy can be important, especially at higher temperatures. The speed of NFFs would in principle allow adding free-energy corrections based on the harmonic approximation at tractable cost.

In addition, there is a wide literature of  $\text{SrTiO}_3(001)$  reconstructions and not all past studies agree with each other. Further experimental studies probing different combinations of chemical potentials will help validate our phase diagram.

Finally, we acknowledge that Markov chain-based sampling is difficult to parallelize and we envision an ML-based sampling method in the future to improve sampling speed.

## Methods

### Virtual Surface Site Relaxation-Monte Carlo

The Metropolis–Hastings Markov chain Monte Carlo algorithm was adapted to implement VSSR-MC. VSSR-MC simulations were performed in the canonical ensemble in addition to the semigrand ensemble. In the canonical ensemble simulation, the number of each adsorbate type is fixed. For each canonical VSSR-MC iteration, a pair of adsorption sites with different adsorbate types (empty virtual sites count as one adsorbate type) are randomly chosen and the adsorbate identities are swapped. In the semigrand ensemble simulation, the chemical potentials of adsorbates are supplied and the total number of adsorbates may vary across an MC run. For each semigrand VSSR-MC iteration, one adsorption site is randomly chosen to change state.

For  $\text{GaN}(0001)$ , the semigrand ensemble method was first used with  $\mu_{\text{Ga}} = 5 \text{ eV}$  (arbitrary positive value) to increase the number of Ga adsorbates to 12 before switching over to the canonical ensemble for annealing with annealing parameter  $\alpha = 0.99$ . The starting sampling temperature was varied between 5,000 K and 12,000 K (temperature at which  $k_{\text{B}}T = 1$ ).

For  $\text{Si}(111)$ , a similar procedure was followed for canonical ensemble sampling with  $\alpha = 0.99$  and starting temperature at 12,000 K. Low-temperature constant-temperature sampling varying between 300 K and 2,500 K was also employed and was more effective at converging the simulations for the  $5 \times 5$  and  $7 \times 7$  supercell reconstructions. The  $3 \times 3$ ,  $5 \times 5$  and  $7 \times 7$  supercells had 16, 50 and 102 Si adsorbates respectively. For the larger  $\text{Si}(111)$   $5 \times 5$  and  $7 \times 7$  supercells, swaps between closer virtual sites were favored using exponentially decaying distance weights:

$$w_{ij} = \frac{\exp(-d_{ij}/d_0)}{\sum_{\text{all } j} \exp(-d_{ij}/d_0)} \quad (2)$$

where  $w_{ij}$  is the normalized weight between the first selected virtual site  $i$  and the second selected virtual site  $j$ ,  $d_{ij}$  is the Euclidean distance between  $i$  and  $j$ , and  $d_0$  is a user-defined distance decay factor that was set to  $2.35 \text{ \AA}$ , the nearest-neighbor distance in Si bulk. For the  $5 \times 5$  and  $7 \times 7$  surfaces, multiple DAS-like structures within the thermal

energy threshold were sampled in a single MC chain. Including these low-energy structures as seeds in subsequent VSSR-MC runs produced the DAS structures.

For  $\text{SrTiO}_3(001)$ , semigrand VSSR-MC was run at various  $\mu_{\text{Sr}}$  and a sampling temperature of at least 1,000 K. (See ‘Active learning’ below for details.) The chemical potential of empty virtual surface sites is set to 0 eV.

Following the discrete sampling step, continuous relaxation was performed using the conjugate gradient method in the Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS)<sup>55</sup> (for the adsorbate atoms in  $\text{GaN}(0001)$  and both surface and adsorbate atoms in  $\text{Si}(111)$ ) or the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm in the Atomic Simulation Environment (ASE)<sup>56</sup> (for both surface and adsorbate atoms in  $\text{SrTiO}_3(001)$ ). For the  $\text{GaN}(0001)$  and  $\text{Si}(111)$   $3 \times 3$  surfaces, a maximum of 500 relaxation steps was allowed after each discrete sampling step. For the  $\text{Si}(111)$   $5 \times 5$  and  $7 \times 7$  surfaces, a maximum of 100 relaxation steps was set to save on computational cost. For  $\text{SrTiO}_3(001)$ , a maximum ranging from 5 to 20 steps in 5-step increments was allowed to reduce computational cost and to capture non-zero forces for AL.

For each iteration in the canonical ensemble, the acceptance probability  $P$  is given by the minimum of unity and the ratio of the Boltzmann weights between the proposed and current state:

$$P = \min \left\{ 1, \exp \left( -\frac{\Delta E_{\text{slab}}}{k_{\text{B}} T} \right) \right\}. \quad (3)$$

### Surface slab modeling

The Python ASE, CatKit<sup>57</sup> and pymatgen<sup>58</sup> libraries were used to create and manipulate surfaces as well as to generate virtual adsorption sites defined from the bulk. Both pymatgen pymatgen.analysis.adsorption.AdsorbateSiteFinder.find\_adsorption\_sites and CatKit catkit.gen.adsorption.get\_adsorption\_sites produce well-covered, visually dense top, bridge and hollow adsorption sites with minimal changes from default settings. Within CatKit, adsorption sites are defined on the same plane as the surface and depend on another method to adsorb atoms at the correct distance from the surface. Converting CatKit coordinates to virtual sites for use in VSSR-MC required additional steps. Thus, for  $\text{Si}(111)$  and  $\text{SrTiO}_3(001)$ , we used only pymatgen virtual sites. The AdsorbateSiteFinder class creates adsorption sites using a Delaunay triangulation-based algorithm. The analogous CatKit method also uses a geometry-based method to generate adsorption sites. VESTA v.3.5.8 was used for visualization and producing figures<sup>59</sup>.

**GaN(0001) slab modeling.** A GaN hexagonal unit cell from the Materials Project<sup>60</sup> (mp-804) was cut in the (0001) plane to form a  $3 \times 3$  supercell with two layers. A vacuum spacing of  $15 \text{ \AA}$  in total was added to the ends of the slab. Both pymatgen- and CatKit-generated virtual sites worked equally well. Setting the symmetry reduce option to ‘False’ to generate more virtual sites was the only change made to the default settings for both methods. A side view of the pymatgen and CatKit virtual sites can be found in Extended Data Fig. 7a,b. For this surface, Ga was the only adsorbate.

**Si(111) slab modeling.** A Si cubic unit cell from the Materials Project (mp-149) was cut in the (111) plane to form  $3 \times 3$ ,  $5 \times 5$  and  $7 \times 7$  supercells with four layers each comprising 9, 25 and 49 atoms, respectively. A vacuum spacing of  $20 \text{ \AA}$  in total was added to the ends of the slab. pymatgen sites with up to two layers were employed in separate MC runs. For a single layer, the adsorption sites were defined at  $3.0 \text{ \AA}$ , the expected distance of adsorbates from the surface. For two layers, the bottom layer ranged from  $2.0 \text{ \AA}$  to  $3.0 \text{ \AA}$  in  $0.5 \text{ \AA}$  increments and the top layer ranged from  $3.5 \text{ \AA}$  to  $5.0 \text{ \AA}$  also in  $0.5 \text{ \AA}$  increments. A single layer was sufficient to sample structures within the thermally accessible window and the pristine(-like)  $5 \times 5$  and  $7 \times 7$  surfaces. Similar to

GaN(0001), symmetry reduction of sites was disabled. A side view of the  $5 \times 5$  supercell pymatgen virtual sites can be found in Extended Data Fig. 7c. For this surface, Si was the only adsorbate.

**SrTiO<sub>3</sub>(001) slab modeling.** An SrTiO<sub>3</sub> cubic unit cell from the Materials Project (mp-5229) was optimized and cut in the (001) plane to create a  $2 \times 2$  supercell with three layers of TiO<sub>2</sub> and SrO. A vacuum spacing of 15 Å in total was set at the ends of the slab. Overlapping adsorption sites ( $\sim 100$ ) were defined at an arbitrary distance of 1.55 Å from the TiO<sub>2</sub> surface (compare 1.96 Å lattice parameter in the unit cell) using the same pymatgen AdsorbateSiteFinder class. Testing different adsorption site distances up to 1.7 Å in early runs showed adsorbates consistently relaxed to around the same distance away from the pristine surface. Symmetry reduction of sites was similarly disabled. A side view of the pymatgen virtual sites can be found in Extended Data Fig. 7d. For this surface, Sr, Ti and O were possible adsorbates.

### Neural network force field

NFFs for SrTiO<sub>3</sub>(001) were trained using an internal implementation of the equivariant polarizable atom interaction neural network (PaiNN) architecture<sup>61</sup>. The NFF energies are not the same as surface free energies (as in ‘Surface stability analysis’ below). A linear interpolation was performed over the lowest DFT energy structure for each composition to derive the energy offset per atom type. The corresponding atomic energy offset was subtracted from each atom in all structures to produce the target NFF energies. The original network hyperparameters were determined with SigOpt<sup>62</sup> to provide good results and thus unmodified. Briefly, the interaction cut-off between atoms was 5.0 Å, feature dimension was 128, number of radial basis functions was 20 and the number of convolutions was 3. The swish activation function was employed<sup>63</sup>. The loss function was a weighted sum of the mean-squared errors of forces and energy in the 100:1 ratio. The Adam optimizer<sup>64</sup> was used with a starting learning rate of 0.001, a learning rate patience of 15 steps and a learning rate decay factor of 0.3. A stopping potential in the form of  $V = (\sigma/r)^2$  was also used, where  $\sigma$  is related to the size of the atoms and set to 1.5 Å, and  $r$  is the distance between atoms. An ensemble of three NFF models was trained for 500 epochs each, sufficient to yield reliable estimates of energy, forces and their respective variances.

### Active learning

To create the starting SrTiO<sub>3</sub>(001) dataset, Sr, Ti and O adsorbates were randomly added to adsorption sites, using a minimum distance of 1.5 Å between atoms to prevent non-physical results.

The adversarial attack method, as implemented by ref. 42, centers around the adversarial loss function,  $\mathcal{L}_{\text{adv}}$ , which maximizes the predicted Boltzmann probabilities,  $p(X_\delta)$ , and the variance of predicted forces,  $\sigma_F^2(X_\delta)$ , via direct nuclear coordinate perturbation:

$$\max_{\delta} \mathcal{L}_{\text{adv}}(X, \delta; \theta) = \max_{\delta} p(X_\delta) \sigma_F^2(X_\delta) \quad (4)$$

In this equation,  $X$  denotes initial atomic positions,  $\delta$  represents atomic displacements from these positions and  $\theta$  signals the use of a neural network to estimate  $\mathcal{L}_{\text{adv}}$ . The predicted Boltzmann probabilities,  $p(X_\delta)$ , are proportional to  $\exp\left(\frac{-E_{\text{slab}}(X_\delta)}{k_B T_{\text{adv}}}\right)$ , where  $E_{\text{slab}}(X_\delta)$  is the energy of the displaced slab,  $k_B$  is the Boltzmann constant and  $T_{\text{adv}}$  is the adversarial attack sampling temperature. Both  $p(X_\delta)$  and  $\sigma_F^2(X_\delta)$  are estimated by our NFF. The atomic coordinates of each atom are perturbed independently to maximize the adversarial loss function. The adversarial attack was performed using default parameters, with  $k_B T_{\text{adv}}$  set at 0.7 eV, learning rate at  $5 \times 10^{-4}$  and run for 100 epochs. For the second-generation structures, starting structures (seeds) were selected randomly, while for the sixth-generation structures, seeds with energies up to 43.4 eV (1,000 kcal mol<sup>-1</sup>) above the lowest-energy structure were chosen.

AL using latent space clustering was done by first running VSSR-MC sampling from  $\mu_{\text{Sr}} = -12$  eV to  $\mu_{\text{Sr}} = 0$  eV in increments of 2 eV. The MC-generated structures were clustered according to the first three principal components (PCs;  $\sim 90\%$  explained variance) of their NFF embeddings and the most uncertain structure for each cluster was selected. The predicted force s.d., rather than energy s.d., was used to estimate uncertainty, as it more accurately estimates the NFF error as seen in this work and previous research<sup>42,65</sup>. Our approach applies hierarchical agglomerative clustering with Ward’s method. This technique initially treats each data point as a separate cluster and then progressively merges clusters, minimizing total within-cluster variance. To save on compute time, a first-pass clustering was run for every 1,000 samples to yield around 100–200 structures at each  $\mu_{\text{Sr}}$ . The results of one such clustering run is shown in Extended Data Fig. 8a. Structures sampled at all  $\mu_{\text{Sr}}$  were pooled together for a second clustering step to select around 800 structures per AL generation.

Including starting structures, AL was run for a total of six generations. The second AL generation used adversarial attack to create a stable NFF for continuous relaxation. The next three generations (3–5) employed VSSR-MC with latent space clustering to train our NFF in the phase space most relevant to surface reconstruction. The goal of using adversarial attack for the final AL generation was to make the force field more robust for low-energy structures.

### Surface stability analysis

The analysis method is outlined below and details are in ‘Surface stability analysis’ in Supplementary Information. SrTiO<sub>3</sub>(001) surfaces were compared using the surface Gibbs free energy  $\Omega_{\text{surf}}^{\text{SrTiO}_3}$  (ref. 12):

$$\Omega_{\text{surf}}^{\text{SrTiO}_3} = G_{\text{slab}} - N_{\text{Sr}}\mu_{\text{Sr}} - N_{\text{Ti}}\mu_{\text{Ti}} - N_{\text{O}}\mu_{\text{O}} \quad (5)$$

where  $G_{\text{slab}}$  refers to the Gibbs free energy of the slab. For each element  $a$ ,  $N_a$  refers to the number of  $a$  atoms in the slab,  $\mu_a$  refers to the chemical potential of  $a$ .

By the following relationship:

$$\mu_{\text{SrTiO}_3} = \mu_{\text{Sr}} + \mu_{\text{Ti}} + 3\mu_{\text{O}} = g_{\text{SrTiO}_3}^{\text{bulk}}$$

where  $\mu_{\text{SrTiO}_3}$  and  $g_{\text{SrTiO}_3}^{\text{bulk}}$  refer to the chemical potential and the Gibbs free energy of the SrTiO<sub>3</sub> unit cell, respectively, we obtain:

$$\Omega_{\text{surf}}^{\text{SrTiO}_3} = G_{\text{slab}} - N_{\text{Ti}}g_{\text{SrTiO}_3}^{\text{bulk}} - \Gamma_{\text{Sr}}^{\text{Ti}}\mu_{\text{Sr}} - \Gamma_{\text{O}}^{\text{Ti}}\mu_{\text{O}} \quad (6)$$

where  $\Gamma_a^{\text{Ti}} = N_a - N_{\text{Ti}} \frac{N_a^{\text{bulk}}}{N_{\text{Ti}}^{\text{bulk}}}$  refers to the excess  $a$  component in the surface with respect to the number of Ti atoms and  $\frac{N_a^{\text{bulk}}}{N_{\text{Ti}}^{\text{bulk}}}$  refers to the bulk stoichiometric ratio of  $a$  to Ti.

Approximating Gibbs free energies by DFT energies<sup>66,67</sup> and redefining chemical potentials by subtracting reference state energies  $E_a$  for each component  $a$  obtained from DFT calculations:

$$\Omega_{\text{surf}}^{\text{SrTiO}_3} = \phi - \Gamma_{\text{Sr}}^{\text{Ti}}\mu_{\text{Sr}} - \Gamma_{\text{O}}^{\text{Ti}}\mu_{\text{O}} \quad (7)$$

where

$$\phi \approx E_{\text{slab}} - N_{\text{Ti}}E_{\text{SrTiO}_3}^{\text{bulk}} - \Gamma_{\text{Sr}}^{\text{Ti}}E_{\text{Sr}}^{\text{bulk}} - \Gamma_{\text{O}}^{\text{Ti}}\frac{E_{\text{O}_2}}{2}$$

$E_{\text{slab}}$  is the slab energy, and  $E_{\text{SrTiO}_3}^{\text{bulk}}$  and  $E_{\text{Sr}}^{\text{bulk}}$  are the DFT energies of the SrTiO<sub>3</sub> and Sr bulk, respectively.  $E_{\text{O}_2}$  is the DFT energy of an isolated O<sub>2</sub> molecule. In this work, we did not correct for oxygen overbinding in the gas phase. Even with such a correction, we expect only a minor change of the cross-over points and a downwards shift in our surface phase diagram.

## Phase diagram creation

$\text{SrTiO}_3$  has three elements but only two degrees of freedom in chemical potential, as in equation (7). The chemical potential of empty virtual sites is fixed. The MC-sampled structures show different  $\Omega_{\text{surf}}^{\text{SrTiO}_3}$  as  $\mu_{\text{Sr}}$  and  $\mu_{\text{o}}$  change. By selecting the most stable structure at each  $\mu_{\text{Sr}}$  and  $\mu_{\text{o}}$ , a two-dimensional phase diagram with  $\mu_{\text{Sr}}$  and  $\mu_{\text{o}}$  axes was obtained.

## DFT calculations

The Vienna ab initio Simulation Package v.6.2.1<sup>68</sup> was employed to relax bulk structures and for single-point DFT calculations of  $\text{SrTiO}_3(001)$  surfaces, using the projector-augmented-wave method to describe core electrons<sup>69</sup>. The Perdew–Burke–Ernzerhof functional<sup>70</sup> within the generalized-gradient approximation was utilized for spin-polarized calculations. Dipole corrections to the total energy were enabled along the z axis for surfaces. The kinetic energy cut-off for plane waves was set to 520 eV. In the self-consistent field cycle, a limit of  $10^{-6}$  eV was adopted as the stopping criterion.

## Workflow management and compute time

An internal library, HTVS (for high-throughput virtual simulations) managed the DFT calculations, NFF training and adversarial attacks. VSSR-MC and latent space clustering were run in separate procedures. DFT calculations took about 0.5–1 hours each on 30 cores of an Intel Xeon Platinum 8260 central processing unit (CPU). NFF training increased with dataset size and approached a maximum of 3 hours per model with an Nvidia Volta V100 32 GB graphics processing unit (GPU). VSSR-MC runs sampled at a rate of about 1 hour per 10,000 iterations using the NFF energy model (~20,000 iterations for each run for  $\text{SrTiO}_3(001)$ ) on an Nvidia GeForce RTX 2080 Ti 11 GB GPU and about 15 min per 10,000 iterations using the classical Tersoff potential (~50,000 iterations for each run for GaN(0001)) on 4 cores of an Intel Core i9-7920X CPU. For the Si(111)  $3 \times 3$  surface using the Stephenson, Radny and Smith (SRS) modified Stillinger–Weber potential implemented in OpenKIM<sup>71</sup>, it took about 5 min per 10,000 iterations (~50,000 iterations for each run) on 4 cores of the same Intel Core i9-7920X CPU. Meanwhile, the Si(111)  $5 \times 5$  and  $7 \times 7$  runs were carried out on the same machine used for DFT calculations (4 cores of an Intel Xeon Platinum 8260 CPU). The Si(111)  $5 \times 5$  surface required just 4 min per 10,000 iterations (~150,000 iterations for each run) and the Si(111)  $7 \times 7$  surface required 15 min per 10,000 iterations (~200,000 iterations for each run). Adversarial attack and latent space clustering were comparatively fast, taking less than 15 min for each generation on an Nvidia Volta V100 32 GB GPU and an Nvidia GeForce RTX 2080 Ti 11 GB GPU, respectively.

## Data availability

The trained models, DFT data and Jupyter notebooks used for data analysis are available on Zenodo at <https://doi.org/10.5281/zenodo.7758174> (ref. 72). Source data are provided with this paper.

## Code availability

The VSSR-MC algorithm reported in this work is available on GitHub at <https://github.com/learningmatter-mit/surface-sampling>. The version of code used in this work is available on Zenodo at <https://doi.org/10.5281/zenodo.10086398> (ref. 73).

## References

- Shi, R., Waterhouse, G. I. & Zhang, T. Recent progress in photocatalytic  $\text{CO}_2$  reduction over perovskite oxides. *Solar RRL* **1**, 1700126 (2017).
- Sumaria, V., Nguyen, L., Tao, F. F. & Sautet, P. Atomic-scale mechanism of platinum catalyst restructuring under a pressure of reactant gas. *J. Am. Chem. Soc.* **145**, 392–401 (2023).
- Fabbri, E. et al. Dynamic surface self-reconstruction is the key of highly active perovskite nano-electrocatalysts for water splitting. *Nat. Mater.* **16**, 925–931 (2017).
- Zhang, Z., Wei, Z., Sautet, P. & Alexandrova, A. N. Hydrogen-induced restructuring of a Cu(100) electrode in electroreduction conditions. *J. Am. Chem. Soc.* **144**, 19284–19293 (2022).
- Sha, Z., Shen, Z., Cali, E., Kilner, J. A. & Skinner, S. J. Understanding surface chemical processes in perovskite oxide electrodes. *J. Mater. Chem.* **11**, 5645–5659 (2023).
- Jung, S.-K. et al. Understanding the degradation mechanisms of  $\text{LiNi}_{0.5}\text{Co}_{0.2}\text{Mn}_{0.3}\text{O}_2$  cathode material in lithium ion batteries. *Adv. Energy Mater.* **4**, 1300787 (2014).
- Han, B. et al. From coating to dopant: how the transition metal composition affects alumina coatings on Ni-rich cathodes. *ACS Appl. Mater. Interfaces* **9**, 41291–41302 (2017).
- Xu, C. et al. Bulk fatigue induced by surface reconstruction in layered Ni-rich cathodes for Li-ion batteries. *Nat. Mater.* **20**, 84–92 (2021).
- Hirata, A., Saiki, K., Koma, A. & Ando, A. Electronic structure of a SrO-terminated  $\text{SrTiO}_3(100)$  surface. *Surf. Sci.* **319**, 267–271 (1994).
- Castell, M. R. Scanning tunneling microscopy of reconstructions on the  $\text{SrTiO}_3(001)$  surface. *Surf. Sci.* **505**, 1–13 (2002).
- Erdman, N. et al. The structure and chemistry of the  $\text{TiO}_2$ -rich surface of  $\text{SrTiO}_3(001)$ . *Nature* **419**, 55–58 (2002).
- Heifets, E., Piskunov, S., Kotomin, E. A., Zhukovskii, Y. F. & Ellis, D. E. Electronic structure and thermodynamic stability of double-layered  $\text{SrTiO}_3(001)$  surfaces: ab initio simulations. *Phys. Rev. B* **75**, 115417 (2007).
- Li, H., Jiao, Y., Davey, K. & Qiao, S.-Z. Data-driven machine learning for understanding surface structures of heterogeneous catalysts. *Angew. Chem. Int. Ed.* **135**, e202216383 (2023).
- Merte, L. R. et al. Structure of an ultrathin oxide on  $\text{Pt}_3\text{Sn}(111)$  solved by machine learning enhanced global optimization. *Angew. Chem. Int. Ed.* **61**, e202204244 (2022).
- Foiles, S. M., Baskes, M. I. & Daw, M. S. Embedded-atom-method functions for the fcc metals Cu, Ag, Au, Ni, Pd, Pt, and their alloys. *Phys. Rev. B* **33**, 7983–7991 (1986).
- Nord, J., Albe, K., Erhart, P. & Nordlund, K. Modelling of compound semiconductors: analytical bond-order potential for gallium, nitrogen and gallium nitride. *J. Phys. Condensed Matter* **15**, 5649 (2003).
- Kolpak, A. M., Li, D., Shao, R., Rappe, A. M. & Bonnell, D. A. Evolution of the structure and thermodynamic stability of the  $\text{BaTiO}_3(001)$  surface. *Phys. Rev. Lett.* **101**, 036102 (2008).
- Wexler, R. B., Qiu, T. & Rappe, A. M. Automatic prediction of surface phase diagrams using ab initio grand canonical Monte Carlo. *J. Phys. Chem. C* **123**, 2321–2328 (2019).
- Zhou, X.-F., Oganov, A. R., Shao, X., Zhu, Q. & Wang, H.-T. Unexpected reconstruction of the  $\alpha$ -boron (111) surface. *Phys. Rev. Lett.* **113**, 176101 (2014).
- Timmermann, J. et al.  $\text{IrO}_2$  surface complexions identified through machine learning and surface investigations. *Phys. Rev. Lett.* **125**, 206101 (2020).
- Wales, D. J. & Doye, J. P. K. Global optimization by basin-hopping and the lowest energy structures of Lennard–Jones clusters containing up to 110 atoms. *J. Phys. Chem. A* **101**, 5111–5116 (1997).
- Panosetti, C., Krautgasser, K., Palagin, D., Reuter, K. & Maurer, R. J. Global materials structure search with chemically motivated coordinates. *Nano Lett.* **15**, 8044–8048 (2015).
- Obersteiner, V., Scherbela, M., Hörmann, L., Wegner, D. & Hofmann, O. T. Structure prediction for surface-induced phases of organic monolayers: overcoming the combinatorial bottleneck. *Nano Lett.* **17**, 4453–4460 (2017).
- Egger, A. T. et al. Charge transfer into organic thin films: a deeper insight through machine-learning-assisted structure search. *Adv. Sci.* **7**, 2000992 (2020).

25. Bauer, M. N., Probert, M. I. J. & Panosetti, C. Systematic comparison of genetic algorithm and basin hopping approaches to the global optimization of Si(111) surface reconstructions. *J. Phys. Chem. A* **126**, 3043–3056 (2022).
26. Wang, Q., Oganov, A. R., Zhu, Q. & Zhou, X.-F. New reconstructions of the (110) surface of rutile TiO<sub>2</sub> predicted by an evolutionary method. *Phys. Rev. Lett.* **113**, 266101 (2014).
27. Schusteritsch, G. & Pickard, C. J. Predicting interface structures: from SrTiO<sub>3</sub> to graphene. *Phys. Rev. B* **90**, 035424 (2014).
28. Meldgaard, S. A., Mortensen, H. L., Jørgensen, M. S. & Hammer, B. Structure prediction of surface reconstructions by deep reinforcement learning. *J. Phys. Condensed Matter* **32**, 404005 (2020).
29. Hess, F. & Yıldız, B. Polar or not polar? The interplay between reconstruction, Sr enrichment, and reduction at the La<sub>0.75</sub>Sr<sub>0.25</sub>MnO<sub>3</sub>(001) surface. *Phys. Rev. Mater.* **4**, 015801 (2020).
30. Unke, O. T. et al. Machine learning force fields. *Chem. Rev.* **121**, 10142–10186 (2021).
31. Axelrod, S. et al. Learning matter: materials design with machine learning and atomistic simulations. *Acc. Mater. Res.* **3**, 343–357 (2022).
32. Bisbo, M. K. & Hammer, B. Efficient global structure optimization with a machine-learned surrogate model. *Phys. Rev. Lett.* **124**, 086102 (2020).
33. Bisbo, M. K. & Hammer, B. Global optimization of atomic structure enhanced by machine learning. *Phys. Rev. B* **105**, 245404 (2022).
34. Timmermann, J. et al. Data-efficient iterative training of Gaussian approximation potentials: application to surface structure determination of rutile IrO<sub>2</sub> and RuO<sub>2</sub>. *J. Chem. Phys.* **155**, 244107 (2021).
35. Rønne, N. et al. Atomistic structure search using local surrogate model. *J. Chem. Phys.* **157**, 174115 (2022).
36. Han, Y. et al. Prediction of surface reconstructions using MAGUS. *J. Chem. Phys.* **158**, 174109 (2023).
37. Xu, J., Xie, W., Han, Y. & Hu, P. Atomistic insights into the oxidation of flat and stepped platinum surfaces using large-scale machine learning potential-based grand-canonical Monte Carlo. *ACS Catal.* **12**, 14812–14824 (2022).
38. Bernardin, F. E. & Rutledge, G. C. Semi-grand canonical Monte Carlo (SGMC) simulations to interpret experimental data on processed polymer melts and glasses. *Macromolecules* **40**, 4691–4702 (2007).
39. Damewood, J., Schwalbe-Koda, D. & Gómez-Bombarelli, R. Sampling lattices in semi-grand canonical ensemble with autoregressive machine learning. *npj Comput. Mater.* **8**, 61 (2022).
40. Carrete, J., Montes-Campos, H., Wanzenböck, R., Heid, E. & Madsen, G. K. H. Deep ensembles vs committees for uncertainty estimation in neural-network force fields: comparison and application to active learning. *J. Chem. Phys.* **158**, 204801 (2023).
41. Tan, A. R., Urata, S., Goldman, S., Dietschreit, J. C. B. & Gómez-Bombarelli, R. Single-model uncertainty quantification in neural network potentials does not consistently outperform model ensembles. Preprint at <https://arxiv.org/abs/2305.01754> (2023).
42. Schwalbe-Koda, D., Tan, A. R. & Gómez-Bombarelli, R. Differentiable sampling of molecular geometries with uncertainty-based adversarial attacks. *Nat. Commun.* **12**, 5104 (2021).
43. Fu, X. et al. Forces are not enough: benchmark and critical evaluation for machine learning force fields with molecular simulations. *Transactions on Machine Learning Research* <https://openreview.net/forum?id=A8pqQipwkt> (2023).
44. Damewood, J. et al. Representations of materials for machine learning. *Annu. Rev. Mater. Res.* **53**, 399–426 (2023).
45. Stephenson, P. C. L., Radny, M. W. & Smith, P. V. A modified Stillinger–Weber potential for modelling silicon surfaces. *Surf. Sci.* **366**, 177–184 (1996).
46. Northrup, J. E., Neugebauer, J., Feenstra, R. M. & Smith, A. R. Structure of GaN(0001): the laterally contracted Ga bilayer model. *Phys. Rev. B* **61**, 9932–9935 (2000).
47. Štich, I., Payne, M. C., King-Smith, R. D., Lin, J.-S. & Clarke, L. J. Ab initio total-energy calculations for extremely large systems: application to the Takayanagi reconstruction of Si(111). *Phys. Rev. Lett.* **68**, 1351–1354 (1992).
48. Smeu, M., Guo, H., Ji, W. & Wolkow, R. A. Electronic properties of Si(111)-7×7 and related reconstructions: density functional theory calculations. *Phys. Rev. B* **85**, 195315 (2012).
49. Herger, R. et al. Surface of strontium titanate. *Phys. Rev. Lett.* **98**, 076102 (2007).
50. Hong, C. et al. Anomalous intense coherent secondary photo-emission from a perovskite oxide. *Nature* **617**, 493–498 (2023).
51. Szot, K. & Speier, W. Surfaces of reduced and oxidized SrTiO<sub>3</sub> from atomic force microscopy. *Phys. Rev. B* **60**, 5909–5926 (1999).
52. Kubo, T. & Nozoye, H. Surface structure of SrTiO<sub>3</sub>(100). *Surf. Sci.* **542**, 177–191 (2003).
53. Winter, G. & Gómez-Bombarelli, R. Simulations with machine learning potentials identify the ion conduction mechanism mediating non-Arrhenius behavior in LGPS. *J. Phys. Energy* **5**, 024004 (2023).
54. Millan, R., Bello-Jurado, E., Moliner, M., Boronat, M. & Gómez-Bombarelli, R. Effect of framework composition and NH<sub>3</sub> on the diffusion of Cu<sup>+</sup> in Cu-CHA catalysts predicted by machine-learning accelerated molecular dynamics. *ACS Cent. Sci.* **9**, 2044–2056 (2023).
55. Thompson, A. P. et al. LAMMPS—a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Comput. Phys. Commun.* **271**, 108171 (2022).
56. Larsen, A. H. et al. The atomic simulation environment—a Python library for working with atoms. *J. Phys. Condensed Matter* **29**, 273002 (2017).
57. Boes, J. R., Mamun, O., Winther, K. & Bligaard, T. Graph theory approach to high-throughput surface adsorption structure generation. *J. Phys. Chem. A* **123**, 2281–2285 (2019).
58. Ong, S. P. et al. Python Materials Genomics (pymatgen): a robust, open-source python library for materials analysis. *Comput. Mater. Sci.* **68**, 314–319 (2013).
59. Momma, K. & Izumi, F. VESTA 3 for three-dimensional visualization of crystal, volumetric and morphology data. *J. Appl. Crystallogr.* **44**, 1272–1276 (2011).
60. Jain, A. et al. The Materials Project: a materials genome approach to accelerating materials innovation. *APL Mater.* **1**, 011002 (2013).
61. Schütt, K., Unke, O. & Gastegger, M. Equivariant message passing for the prediction of tensorial properties and molecular spectra. In *Proc. 38th International Conference on Machine Learning, Proc. Machine Learning Research Vol. 139* (eds Meila, M. & Zhang, T.) 9377–9388 (PMLR, 2021).
62. Martinez-Cantin, R., Tee, K. & McCourt, M. Practical Bayesian optimization in the presence of outliers. In *Proc. Twenty-First International Conference on Artificial Intelligence and Statistics, Proc. Machine Learning Research Vol. 84* (eds Storkey, A. & Perez-Cruz, F.) 1722–1731 (PMLR, 2018).
63. Ramachandran, P., Zoph, B. & Le, Q. V. Searching for activation functions. Preprint at <https://arxiv.org/abs/1710.05941> (2017).
64. Kingma, D. P. & Ba, J. Adam: a method for stochastic optimization. In *Proc. 3rd International Conference on Learning Representations, ICLR 2015* (eds Bengio, Y. & LeCun, Y.) (2015).
65. Gasteiger, J., Giri, S., Margraf, J. T. & Günemann, S. Fast and uncertainty-aware directional message passing for non-equilibrium molecules. *Machine Learning for Molecules Workshop, NeurIPS 2020* [https://ml4molecules.github.io/papers2020/ML4Molecules\\_2020\\_paper\\_35.pdf](https://ml4molecules.github.io/papers2020/ML4Molecules_2020_paper_35.pdf) (2020).

66. Reuter, K. & Scheffler, M. Composition, structure, and stability of RuO<sub>2</sub>(110) as a function of oxygen pressure. *Phys. Rev. B* **65**, 035406 (2001).
67. Heifets, E., Ho, J. & Merinov, B. Density functional simulation of the BaZrO<sub>3</sub>(011) surface structure. *Phys. Rev. B* **75**, 155431 (2007).
68. Kresse, G. & Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169–11186 (1996).
69. Kresse, G. & Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **59**, 1758–1775 (1999).
70. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
71. Tadmor, E. B., Elliott, R. S., Sethna, J. P., Miller, R. E. & Becker, C. A. The potential of atomistic simulations and the knowledgebase of interatomic models. *JOM* **63**, 17 (2011).
72. Du, X. Data for: Machine-learning-accelerated simulations to enable automatic surface reconstruction. Zenodo <https://doi.org/10.5281/zenodo.7758174> (2023).
73. Du, X. learningmatter-mit/surface-sampling. Zenodo <https://doi.org/10.5281/zenodo.10086398> (2023).

## Acknowledgements

We thank G. Winter, J. Peng, N. Frey and M. Liu for helpful discussions. We also appreciate editing by J. Peng and A. Hoffman. X.D. acknowledges support from the National Science Foundation Graduate Research Fellowship under grant no. 2141064. J.K.D. was supported by the Department of Defense through the National Defense Science and Engineering Graduate Fellowship Program. We are grateful for computation time allocated on the MIT SuperCloud cluster, the MIT Engaging cluster and the NERSC Perlmutter cluster. This material is based on work supported by the Under Secretary of Defense for Research and Engineering under Air Force Contract No. FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Under Secretary of Defense for Research and Engineering. Delivered to the US Government with Unlimited Rights, as defined in DFARS Part 252.227-7013 or 7014 (February 2014). Notwithstanding any copyright notice, US Government rights in this work are defined by DFARS 252.227-7013 or DFARS 252.227-7014 as detailed above. Use of this work other than as specifically authorized by the US Government may violate any copyrights that exist in this work.

## Author contributions

X.D. implemented the sampling algorithm, performed surface modeling, ran DFT calculations, trained the neural networks and carried out surface stability analysis. J.K.D. assisted with sampling algorithm implementation and provided guidance with surface modeling. J.R.L. provided guidance with surface modeling and ran DFT calculations. R.M. provided guidance with neural network training and active learning. B.Y. provided guidance with the choice of surfaces and surface stability analysis. L.L. supervised the research and contributed to securing funding. R.G.-B. conceived the project, supervised the research and contributed to securing funding. All authors contributed to results discussion and paper writing.

## Competing interests

The authors declare no competing interests.

## Additional information

**Extended data** is available for this paper at  
<https://doi.org/10.1038/s43588-023-00571-7>.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s43588-023-00571-7>.

**Correspondence and requests for materials** should be addressed to Rafael Gómez-Bombarelli.

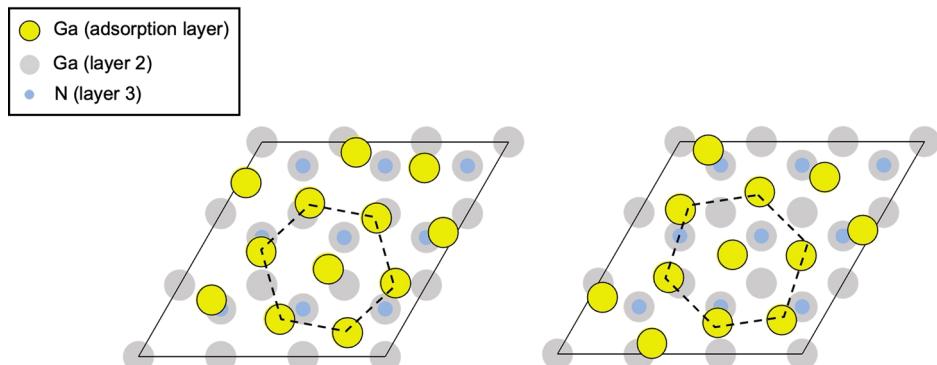
**Peer review information** *Nature Computational Science* thanks Mie Andersen and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: Kaitlin McCardle, in collaboration with the *Nature Computational Science* team. Peer reviewer reports are available.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

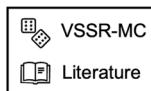
**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2023



**Extended Data Fig. 1 | Top view of additional GaN(0001) MC-sampled structures.** The surface reconstructions are rotated in comparison with the reference structure from ref. 46 but contain the same hexagonal pattern.



(a)	Type (3x3)	DAS	DAS-like 1	DAS-like 2
Top view				
Side view				
Occurrence				
SRS energy [meV/atom]	0.0	12.5	18.4	
DFT energy [meV/atom]	0.0	30.3	0.0*	

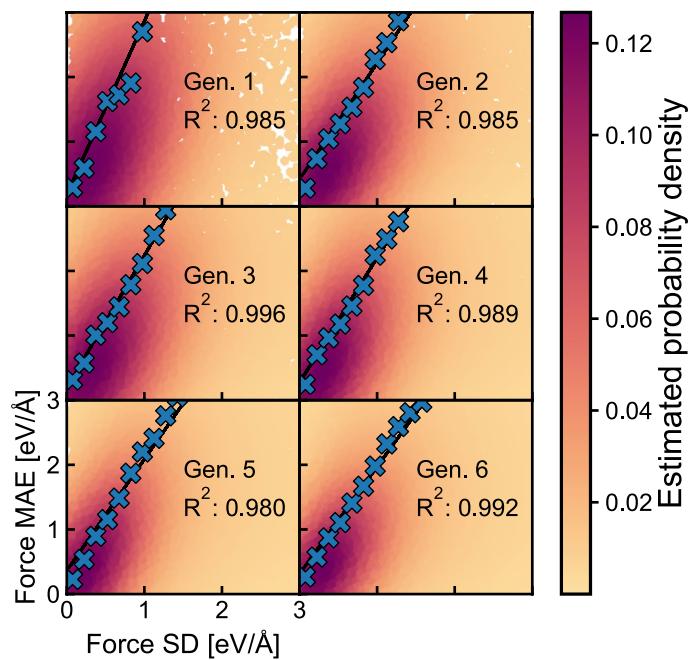
(b)	Type (5x5)	DAS	DAS-like	Pristine
Top view				
Side view				
Occurrence				
SRS energy [meV/atom]	0.0	20.4	2.1	
DFT energy [meV/atom]	0.0	32.6	19.7	

(c)	Type (7x7)	DAS	DAS-like	Pristine with adatoms
Top view				
Side view				
Occurrence				
SRS energy [meV/atom]	0.0	11.9	3.5	
DFT energy [meV/atom]	0.0	30.0	16.7	

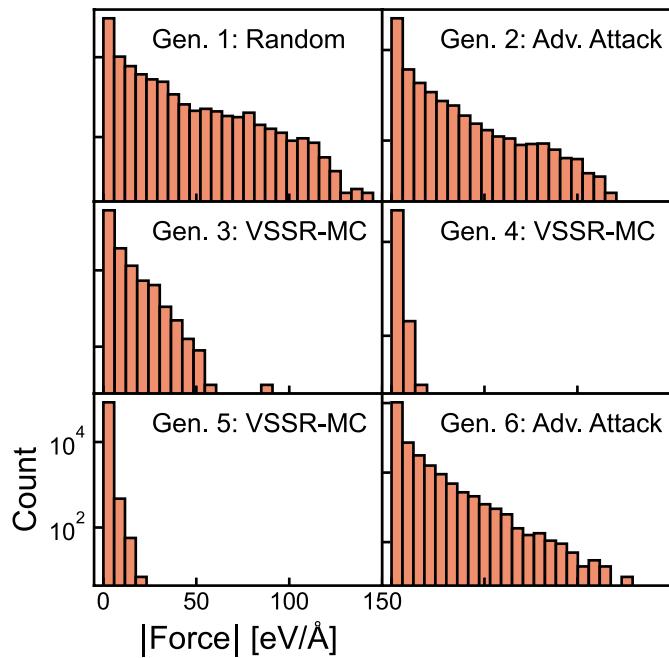
**Extended Data Fig. 2 | Comparing classical potential and DFT energies of Si(111) sampled surface reconstructions.** a–c, Structures shown were obtained from constant-composition (canonical) VSSR-MC sampling using the SRS modified Stillinger-Weber potential<sup>45</sup> with 3x3 (a), 5x5 (b) and 7x7 (c) unit

cells. The SRS energies were obtained from the depicted structures while the DFT energies came from structures further relaxed at the DFT level. \*Further relaxation using DFT resulted in the 3x3 DAS structure.

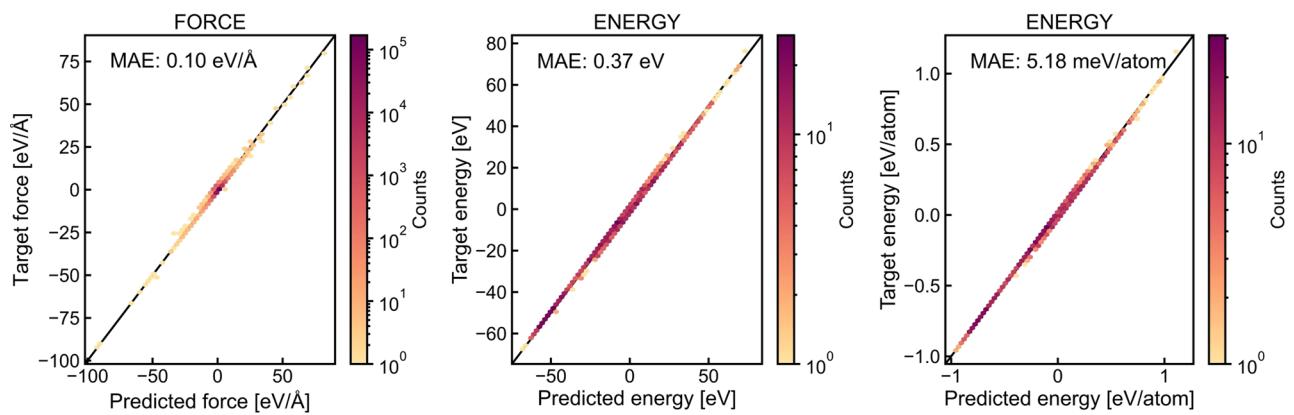
**Extended Data Fig. 3 | Correlation plot of force MAE with force s.d. over AL generations.**

At each AL generation, an ensemble of just three NFF models was able to estimate force s.d. that correlated strongly with force error. Each individual data point represents a sampled structure. Each blue 'X' represents

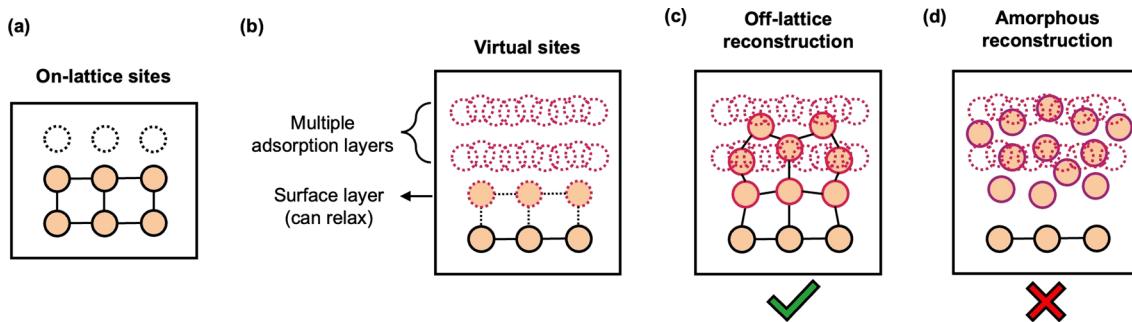
a binned average and a best-fit line is drawn through the binned averages. The binned average is calculated by dividing both the force s.d. and force MAE into equal-sized bins. The average force MAE is then plotted against the median force s.d. for each corresponding bin.



**Extended Data Fig. 4 | Force distribution over AL generations.** The majority of high-force structures were added in AL generations 1, 2 and 6, which correspond either to random structures or structures obtained through adversarial attack. The three VSSR-MC AL generations produced structures with low force values mostly around 50 eV Å<sup>-1</sup> or less.

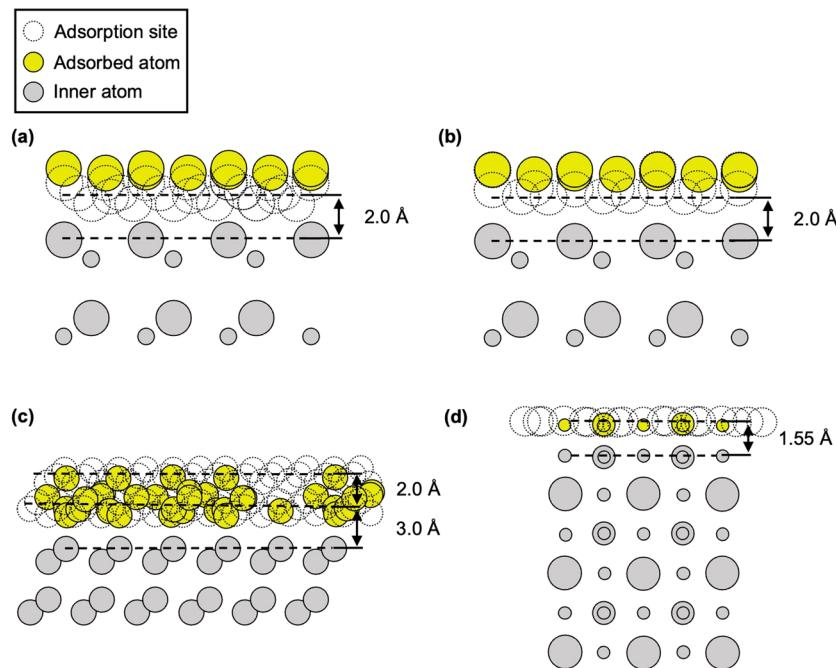


**Extended Data Fig. 5 | Test performance of the best NFF model.** As described in the main paper, the test data is obtained from VSSR-MC runs using the sixth-generation NFF model.



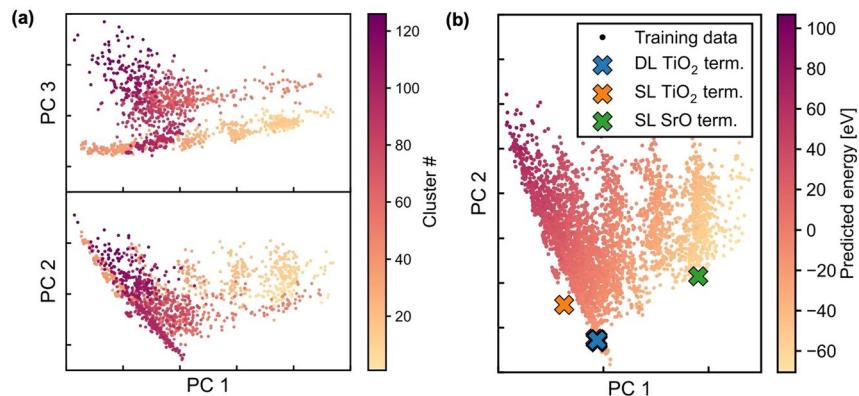
**Extended Data Fig. 6 | Strengths and limitations of VSSR-MC.** **a,b**, Comparison of limited fixed on-lattice sites (**a**) and denser algorithmically-generated virtual surface sites that can overlap (**b**). **c**, Off-lattice reconstructions can be obtained

following VSSR-MC discrete sampling at virtual sites and continuous relaxation of surface atoms and adsorbates. **d**, Amorphous reconstructions with many local minima, however, will likely be difficult for VSSR-MC to sample.



**Extended Data Fig. 7 | Side view of virtual sites for surfaces studied in this work.** a–d, Pymatgen (a) and CatKit (b) virtual sites for GaN(0001) against the contracted Ga monolayer reconstruction, two-layer pymatgen sites for

Si(111) against the 5x5 DAS reconstruction (c), and pymatgen virtual sites for SrTiO<sub>3</sub>(001) against the double-layer TiO<sub>2</sub> reconstruction (d). The dashed lines are a guide for the eye.



**Extended Data Fig. 8 | Visualizations in the latent space.** **a**, Clustering of VSSR-MC structures in the NFF latent space visualized in the first three principal components. In the VSSR-MC with clustering AL method, the surface from

each cluster with the highest force s.d. is selected for DFT evaluation. **b**, PCA of training data and the dominant terminations (term.) in the latent space of the sixth-generation model.