

Peer Review Information

Journal: Nature Computational Science

Manuscript Title: Machine-learning-accelerated simulations to enable automatic surface reconstruction

Corresponding author name(s): Professor Rafael Gomez-Bombarelli

Editorial Notes:

Reviewer Comments & Decisions:

Decision Letter, initial version:

Date: 5th July 23 15:32:52

Last Sent: 5th July 23 15:32:52

Triggered By: Kaitlin McCardle

From: kaitlin.mccardle@us.nature.com

To: rafagb@mit.edu

BCC: kaitlin.mccardle@us.nature.com

Subject: Decision on Nature Computational Science manuscript NATCOMPUTSCI-23-0568

Message: ** Please ensure you delete the link to your author homepage in this e-mail if you wish to forward it to your co-authors. **

Dear Professor Gomez-Bombarelli,

Your manuscript "Machine-learning-accelerated simulations enable heuristic-free surface reconstruction" has now been seen by 3 referees, whose comments are appended below. You will see that while they find your work of interest, they have raised points that need to be addressed before we can make a decision on publication.

The referees' reports seem to be quite clear. Naturally, we will need you to address all of the points raised.

While we ask you to address all of the points raised, the following points need to be substantially worked on:

- Please provide quantitative comparisons of your approach to existing methods in

order to demonstrate the level of improvement over the state-of-the-art.

- Please provide experiments on additional benchmark systems, such as Si(111).
- Please discuss, in the Discussion section, the limitations of your work
- Please provide additional methodological and experimental details, as noted by the referees.

You will also need to make some editorial changes so that it complies with our Guide to Authors at <https://www.nature.com/natcomputsci/for-authors>.

In particular, I would like to highlight the following points of our style:

Figure legends must provide a brief description of the figure and the symbols used, including definitions of any error bars employed in the figures.

Please use the following link to submit your revised manuscript and a point-by-point response to the referees' comments (which should be in a separate document to any cover letter):

[REDACTED]

** This url links to your confidential homepage and associated information about manuscripts you may have submitted or be reviewing for us. If you wish to forward this e-mail to co-authors, please delete this link to your homepage first. **

To aid in the review process, we would appreciate it if you could also provide a copy of your manuscript files that indicates your revisions by making use of Track Changes or similar mark-up tools. Please also ensure that all correspondence is marked with your Nature Computational Science reference number in the subject line.

In addition, please make sure to upload a Word Document or LaTeX version of your text, to assist us in the editorial stage.

To improve transparency in authorship, we request that all authors identified as 'corresponding author' on published papers create and link their Open Researcher and Contributor Identifier (ORCID) with their account on the Manuscript Tracking System (MTS), prior to acceptance. ORCID helps the scientific community achieve unambiguous attribution of all scholarly contributions. You can create and link your ORCID from the home page of the MTS by clicking on 'Modify my Springer Nature account'. For more information please visit www.springernature.com/orcid.

We hope to receive your revised paper within three weeks. If you cannot send it within this time, please let us know.

We look forward to hearing from you soon.

Best regards,
Kaitie

--

Kaitlin McCardle, PhD

Associate Editor
Nature Computational Science

Reviewers comments:

Reviewer #1 (Remarks to the Author):

The manuscript by Xiaochen Du et al. presents a computational framework termed "Heuristic-Free Surface Reconstruction (HFSurfRecon)" for predicting surface phase diagrams of complex, multi-component materials. A new algorithm, the Virtual Surface Site Relaxation-Monte Carlo (VSSR-MC) sampling algorithm, is presented to efficiently sample across surface configurations and compositions in the semi-grand canonical ensemble. Energy evaluations are first carried out with a classical force field for the GaN surface. Then a neural network force field (NFF) is trained as part of the computational framework for the more challenging SrTiO₃(001) surface. The training of the NFF is carried out using uncertainty-driven active learning to simultaneously improve the accuracy of the NFF and the statistics of the sampling of surface reconstructions.

I believe this work is of quite high novelty and will be of high interest to the community. In particular, I think the approach of combining discrete sampling using fixed adsorption sites with NFF continuous relaxation is very powerful. However, while the use of fixed, virtual adsorption sites makes the sampling efficient, it is also a main drawback of the method, since it would not be able to tackle more complex off-lattice surface reconstructions. These limitations are, however, clearly outlined in the discussion.

The development of the NFF is state-of-the-art, with the inclusion of both active learning and adversarial attack to iteratively improve the NFF.

My recommendation would be to consider this manuscript for publication, although I am a little unsure about the practical usefulness of this method, given that off-lattice surface reconstructions are out of the reach, and given that such reconstructions have frequently been reported in the literature.

Addressing the below points could help to clarify this point and other minor points related to the methodology itself:

- 1) It would be useful to provide some concrete examples of other surfaces where this method can (i.e. known reconstructions follows virtual site geometries) or cannot be applied to make it easier to judge the practical usefulness of this method.
- 2) It seems that the method still requires the user to specify the cell size, which determines the periodicity of surface reconstructions that can be found. How were the cell sizes chosen in this work? And is it correctly understood that this choice still needs guidance from experimental work? If yes, this point should also be discussed in connection with the other limitations of this method.
- 3) What is meant by "heuristic-free" in the naming of the method?
- 4) Why is the predicted force standard deviation and not the predicted energy standard deviation used in the active learning? And is an ensemble size of 3 large enough to reliably estimate the standard deviation?
- 5) I did not understand the point "no additional runs varying μO were necessary". Couldn't there in practice exist surface reconstructions that would not follow the O:Sr and O:Ti stoichiometries from the bulk? Why were these not considered here?
- 6) DFT with the PBE functional is known to exhibit a very high error for the O2

molecule in the gas phase. Were any corrections applied to remedy this? If not, could this error influence the presented results?

7) The sentence "The Perdew-Burke-Ernzerhof (PBE) functional within the generalized-gradient approximation (GGA) were utilized [53] along with the Hubbard U correction [54] for spin-polarized calculations" does not make sense. The Hubbard U correction is a correction for the self-interaction error in DFT, and it must be stated exactly which U value was used for a given element, since this is an empirical parameter. Please comment on the value used, and whether it is appropriate to use a fixed value for all chemical environments of the given element.

8) Reading through the README on the Github page, I could not find any actual installation instructions? Only System requirements. I may have overlooked this, in that case, please clarify.

Reviewer #1 (Remarks on code availability):

Reading through the README on the Github page, I could not find any actual installation instructions? Only System requirements. I may have overlooked this, in that case, please clarify.

I have not reviewed the code itself.

Reviewer #2 (Remarks to the Author):

The manuscript presents a methodological framework for efficiently exploring complex surface reconstructions of multi-component materials under challenging environments. Accurate determination of surface structures is crucial for understanding material properties and performance in applications such as catalysis and energy storage. However, accurately predicting surface energies and effectively sampling surface compositions and configurations poses significant challenges. Machine-learning force fields (MLFFs) offer a potential solution to these challenges due to their lower computational cost compared to density-functional theory (DFT) calculations and their high accuracy. However, the authors claimed that the development of MLFFs with an efficient sampling scheme that can explore compositional and configurational spaces in a heuristic-free manner is currently lacking. To overcome these limitations, they propose a workflow called Heuristic-Free Surface Reconstruction (HFSurfRecon), which combines MLFFs with active learning (AL) and the Virtual Surface Site Relaxation-Monte Carlo (VSSR-MC) sampling algorithm. The framework is validated through case studies on the GaN(0001) surface using a classical force field and the SrTiO₃(001) surface using a neural network force field. The results demonstrate the effectiveness of the AL strategy in conjunction with VSSR-MC sampling, as it acquires new data points from DFT calculations only in relevant regions of the surface phase space. The generated surface phase diagram aligns well with existing literature and reveals previously unknown low-energy surface terminations. The manuscript is highly relevant to the community, and the study is well-executed. However, some statements made in the manuscript raise questions and require further clarification or justification.

Major comments

1. The authors employed the VSSR-MC sampling algorithm to explore compositional

and configurational search spaces and to sample complex, stable surface structures. I understand that semi-grand-canonical sampling is only the addition and removal of particles. The authors later (in section 4.1) mention that they also do canonical sampling, but it is unclear what the trial moves are there. Swaps? Simultaneous addition and removal?

2. Closely related to comment 1, while the algorithm focuses on high-likelihood virtual sites related to the symmetry of the pristine slab, there is no mention of atomic displacements or rearrangements of coordinates of any kind. In my opinion, this completely fails to explore the configurational space of surface reconstructions, and it is possible that non-trivial surface reconstructions deviating from slab symmetry can be easily missed. And these reconstructions can still exhibit periodicity and should not be considered amorphous (e.g., Si(111) DAS reconstructions). This could be okay, as long as the method essentially samples site occupations, but this is unclear in the text. It would be beneficial for the authors to clarify whether they have considered these possibilities and the potential impact on the sampling process. Or else, the sentence "Our algorithm can be easily applied to any surface of interest and paves the way to study unknown multi-component materials under challenging environments such as in aqueous electrochemistry" is a little bit of an overstatement.

3. The authors performed latent space clustering based on the geometries obtained from VSSR-MC sampling. They embedded the geometries using the first three principal components, capturing over 90% of the variances. However, the details of how the samples were clustered into different clusters are not clearly explained. The authors should provide further clarification regarding the clustering methodology they employed, as there are various clustering algorithms available that could have been used.

4. The authors use two model surfaces of GaN(0001) and SrTiO₃(001) to validate the framework. However, that Si(111) is not mentioned bothers me in particular because it exhibits one of the most complex surface reconstructions known to date, and cheap, accurate potentials are available for it (either Stillinger-Weber or GAP). I think it would have been a much more convincing benchmark than GaN(0001) that then would really justify publication in a prestigious journal like npj Comp. Mat.

5. The authors mentioned that a total of 6500 structures for SrTiO₃(001) were generated over six iterations, with fewer than 5,000 structures used for training the final NFF. It thus appears that approximately 80% of the sampled structures required DFT calculations. This raises some ambiguity since the authors selected surfaces based on the highest force standard deviation rather than high energy. Also, it would be interesting to know if there are structurally similar or nearly identical samples in the training set. If so, wouldn't sparsifying the new samples effectively reduce the total number of required DFT calculations while maintaining the accuracy of the NFF (also potentially speed up the NFF potential)? Considering the good extrapolability of the NN-based force field, such a large amount of training data may not be necessary to obtain an accurate MLFF for a single surface orientation. As it stands right now, the approach by the authors is not data efficient at all – and thus not really compatible with other already existing approaches in the literature.

6. Although the samples from the VSSR-MC algorithm are obtained by varying the chemical potential of Sr from -12 to 0 eV, it is unclear how the stable surfaces for

each chemical potential were selected for further DFT calculations and refinement of the training data. The authors seem to use NFF energy (after screening with force standard deviation) in the selection process. It would be worth considering using surface free energy (by varying the chemical potential) instead of NFF energy to identify the relevant structures for further DFT calculations. This could naturally expand the compositional search space to surface structures stable under more extreme conditions. Currently, Fig 5(c-e) shows a limited number of structures found at $\Gamma \leq -5$ or ≥ 5 with the current settings.

7. Fig 4(b) demonstrates the correlation between force standard deviation and force mean absolute error. However, the details of the figure are not explained in the manuscript. The meaning of the colormap terms "estimated density" and "binned average" is unclear. The authors should provide a more detailed explanation of these terms or simplify the figure to enhance clarity.

8. The active learning workflow for the SrTiO₃ model system achieved decent accuracy of the NFF at the 6th generation employed by the authors. However, it would be helpful to know if the number of generations would differ when applying this workflow to other systems with varying levels of complexity. Is the total number of generations fixed, or are there other criteria for determining when to stop the generations, such as energy or force thresholds?

9. The manuscript does not clearly describe how the authors perturbed the given structure in the adversarial attack process. Since there are no specifically defined collective variables in these systems, further elucidation is needed regarding how the perturbation was implemented in their system. Additional details regarding the methodology of the adversarial attack would be beneficial.

Minor comments

1. On page 2, the authors mentioned that "Machine learning (ML) force fields are much faster than DFT calculations while preserving accuracy [21, 22], but they have not been connected to an efficient sampling scheme to explore the phase space of multi-component surfaces in a heuristic-free fashion." seems a bit overstated. There are many works that connect ML and heuristic-free sampling. The authors should provide more detailed comparisons of their methodologies to others in the literature. At present it seems that they do not have a good overview of the field or try to oversell the novelty of their approach.

2. At least, in fairness to the very active community of surface structure determinations, referring to new MLFF fitting + sampling frameworks (such as that in Ref. 14 and 23) seems to be cutting it a bit short. A few others for a more balanced view might include:

<https://doi.org/10.1103/PhysRevLett.124.086102>

<https://doi.org/10.1103/PhysRevLett.125.206101>

<https://doi.org/10.1063/5.0071249>

<https://doi.org/10.1063/5.0121748>

<https://doi.org/10.1103/PhysRevB.105.245404>

3. The global optimization literature is immense, and the authors do not even mention e.g. basin hopping, genetic algorithms, AIRSS, etc. Admittedly, there are not

too many GO works for surfaces and surface reconstructions, but this is precisely because of the particular complexity of the problem, which makes it expensive at the ab initio level. This is a good motivation for this work but not clearly mentioned. Nonetheless, there are some GO works, so giving a reflection of this literature is appropriate. E.g.:

TiO₂ reconstructions <https://doi.org/10.1103/PhysRevLett.113.266101> (GA/USPEX)

Molecular adlayers: <https://doi.org/10.1021/acs.nanolett.7b01637>

<https://doi.org/10.1002/advs.202000992> and others from the group

<https://doi.org/10.1021/acs.nanolett.5b03388>

SrTiO₃: <https://doi.org/10.1103/PhysRevB.90.035424> (AIRSS)

Si(111): <https://doi.org/10.1021/acs.jpca.2c00647> (basin hopping vs GA)

4. The authors do perform relaxation after every MC move, so effectively this is a form of basin hopping and it would be nice of them to cite the original paper: <https://doi.org/10.1021/jp970984n>

5. In section 4.4, a kcal/mol unit has been used, but using a consistent eV unit for energy might be preferable.

6. In Fig 5(b), what are the ticks at the right y-axis (denoted as red and orange ticks with values of 1 and 15) besides scales of oxygen partial pressure? Please add a proper explanation in the caption or main text.

7. In Fig 5(c-e), what is the specific value of oxygen chemical potential? Please add the text.

8. The authors mentioned that "The goal of using adversarial attack for the final AL generation was to make the force field more robust for low-energy structures." Could this be justified?

9. In Fig C2, the test data seems to include very large force components close to 100 eV/Å, and the NFF model surprisingly predicts well of these large forces. Does your final training data also include these high-force configurations? If so, were they only selected and added during the initial generations of NFF when its prediction of energies and forces was poor?

Reviewer #3 (Remarks to the Author):

Du et al. present a computational technique to obtain atomistic models for ideal, i.e., defect-free surfaces under varying external conditions, which is an important problem in material science with important implications e.g. in heterogeneous catalysis. They present what they call a "heuristic-free surface reconstruction framework" hinging on the "Virtual Surface Site Relaxation-Monte Carlo (VSSR-MC)" sampling algorithm to automate the exploration of configurational space and its concomitant potential energy surface.

The presented applications of this methodology to the GaN(0001) and SrTiO₃(001) surfaces seem to have been carried out very carefully. Existing atomistic models available in the literature are being reproduced for both surface and suggest additional 2x2-reconstructions for SrTiO₃(001), one of which is energetically

favorable and thus might also be experimentally relevant. Figure 3a) does provide a glimpse of how the virtual adsorption sites have been chosen for GaN(0001). However, I miss details on how this can be done (heuristic-free!) in general. Since this is an essential and distinguishing feature of the suggested VSSR-MC algorithm, a discussion how such choices can be easily validated - if e.g. specified by parameters that influence the virtual site density. Reconstructions with larger supercells cannot be discovered as the presented approach requires a (surface) super cell to be specified a priori. Consequently, it is not able to explore arbitrarily large NxM-reconstructions (e.g. the famous Au(111) heringbone reconstruction), which might also be based on surface unit cells that are rotated with respect to the primitive surface unit cell. The discussion (in Section 3) mentions this systematic limitation in the context of amorphous overlayers, but does not discuss the aforementioned implications. Furthermore, the present formulation of the VSSR-MC algorithm neglects vibrational entropy in the underlying surface Gibbs free energy (see Section 4.5 and Appendix B). These can play a crucial role to stabilize particular surface structures (and reconstructions) over others in particular when varying external conditions. Consequently, this limitation should be discussed as well.

Altogether, I do not recommend publication of the manuscript in Nature Computational Science in its present form. Considering other computational approaches presented in the literature before (see e.g. Refs. 1 and 2 - neither of which is cited), I fail to see the "groundbreaking" aspects of the work presented here.

[1] S. A. Meldgaard et al., Structure prediction of surface reconstructions by deep reinforcement learning, *J. Phys.: Condens. Matter* 32 404005 (2020).

[2] Yu Han et al., Prediction of surface reconstructions using MAGUS, *J. Chem. Phys.* 158, 174109 (2023).

minor issues:

- Section 2.1, 2nd paragraph:

"Runs are conducted at different external conditions such as pressure, ..., and applied electrical potential."

This is somewhat confusing. If I understand correctly, those two external conditions are not varied in the present work, so the authors should modify/extend this sentence accordingly.

- Section 4.5 and Appendix B:

Appendix B repeats a lot of the contents of Section 4.5 and does not add much over it. So the authors should reconsider whether the former is really needed.

Reviewer #3 (Remarks on code availability):

I have glimpsed at the Python source code and had a good impression regarding the coding style. However, due to lack of time I did not manage to run the code for any of the provided examples.

Author Rebuttal to Initial comments

NATCOMPUTSCI-23-0568: Machine-learning-accelerated simulations enable heuristic-free surface reconstruction

Reviewer #1 (Remarks to the Author):

The manuscript by Xiaochen Du et al. presents a computational framework termed "Heuristic-Free Surface Reconstruction (HFSurfRecon)" for predicting surface phase diagrams of complex, multi-component materials. A new algorithm, the Virtual Surface Site Relaxation-Monte Carlo (VSSR-MC) sampling algorithm, is presented to efficiently sample across surface configurations and compositions in the semi-grand canonical ensemble. Energy evaluations are first carried out with a classical force field for the GaN surface. Then a neural network force field (NFF) is trained as part of the computational framework for the more challenging SrTiO₃(001) surface. The training of the NFF is carried out using uncertainty-driven active learning to simultaneously improve the accuracy of the NFF and the statistics of the sampling of surface reconstructions.

I believe this work is of quite high novelty and will be of high interest to the community. In particular, I think the approach of combining discrete sampling using fixed adsorption sites with NFF continuous relaxation is very powerful. However, while the use of fixed, virtual adsorption sites makes the sampling efficient, it is also a main drawback of the method, since it would not be able to tackle more complex off-lattice surface reconstructions. These limitations are, however, clearly outlined in the discussion.

The development of the NFF is state-of-the-art, with the inclusion of both active learning and adversarial attack to iteratively improve the NFF.

My recommendation would be to consider this manuscript for publication, although I am a little unsure about the practical usefulness of this method, given that off-lattice surface reconstructions are out of the reach, and given that such reconstructions have frequently been reported in the literature.

We thank the Reviewer for the positive feedback. We want to highlight two reasons our method is not fixed to on-lattice surface reconstructions or reconstructions that exactly follow virtual site geometries. First, our virtual sites contain all possible adsorption sites generated algorithmically, much more than the number of on-lattice sites. As an example, our 2x2 SrTiO₃(001) surface has 16 lattice sites, but ~100 virtual sites. Second, by performing relaxations after discrete sampling, we can access the rest of the configurational space. Our GaN(0001) surface demonstrates one such reconstruction that does not obey the bulk lattice structure (the reconstructed layer has a rotated lattice with respect to the pristine surface). We have updated Section 2.2.1 to highlight this. For SrTiO₃(001) surfaces, the algorithm enumerated through many off-lattice structures before determining reconstructions that follow the bulk lattice symmetries are the most stable. We do anticipate that amorphous structures might be harder but not impossible and have edited Section 2.1 and the Discussion section to better reflect our point of view. We have also created an additional figure, Fig. C4 to better convey the strengths and limitations of our approach.

Addressing the below points could help to clarify this point and other minor points related to the methodology itself:

1) It would be useful to provide some concrete examples of other surfaces where this method can (i.e. known reconstructions follows virtual site geometries) or cannot be applied to make it easier to judge the practical usefulness of this method.

We have added the Si(111) reconstructions as proposed by Reviewer 2 to our manuscript in Section 2.2.2 and created Fig. C5 to illustrate the results. These reconstructions neither follow bulk lattice geometries nor virtual site geometries. In a separate work, we are studying the much less understood amorphous reconstruction of SrTiO₃(001) in electrochemistry conditions (additional pH and electrical potential) using our HFSurfRecon algorithm.

2) It seems that the method still requires the user to specify the cell size, which determines the periodicity of surface reconstructions that can be found. How were the cell sizes chosen in this work? And is it correctly understood that this choice still needs guidance from experimental work? If yes, this point should also be discussed in connection with the other limitations of this method.

Yes, the cell size will have to be picked in advance. We chose a small unit cell size commensurate with other cited computational studies to demonstrate our method. But our method is not limited to the cell sizes that we have chosen. Since our future goal is to apply our work to new multicomponent systems, we envision larger unit cell sizes and experimental collaborations to inform our work and verify our results. We have updated Section 3 to better reflect this point.

3) What is meant by "heuristic-free" in the naming of the method?

It means no manual construction of trial surfaces (compositions or configurations) by a human is required. The surface phase diagram is automatically generated based on the sampling results.

4) Why is the predicted force standard deviation and not the predicted energy standard deviation used in the active learning? And is an ensemble size of 3 large enough to reliably estimate the standard deviation?

In active learning, we improve the NFF by obtaining ground-truth samples (using DFT) in regions of high error. Since we do not know the error beforehand (before performing DFT calculations), we use prediction standard deviation as a proxy for prediction error. In the present (see Fig. 4(b)) and previous works by our research group [R1] and others [R2], we have found that the force standard deviation as predicted through an ensemble neural network has a stronger correlation with force error than the energy standard deviation with energy error. In our work, an ensemble size of 3 was sufficient to give a reliable estimate of force standard deviation that correlates well with force error at each generation, which we show in Fig. C8. We have edited Sections 4.3 and 4.4 to reflect this information.

References:

[R1] Schwalbe-Koda, D., Tan, A. R. & Gómez-Bombarelli, R. Differentiable sampling of molecular geometries with uncertainty-based adversarial attacks. *Nat Commun* 12, 5104 (2021).

[R2] Gasteiger, J., Giri, S., Margraf, J. T. & Günnemann, S. Fast and Uncertainty-Aware Directional Message Passing for Non-Equilibrium Molecules. *arXiv.org* <https://arxiv.org/abs/2011.14115v3> (2020).

NATCOMPUTSCI-23-0568: Machine-learning-accelerated simulations enable heuristic-free surface reconstruction

5) I did not understand the point "no additional runs varying μO were necessary". Couldn't there in practice exist surface reconstructions that would not follow the O:Sr and O:Ti stoichiometries from the bulk? Why were these not considered here?

Exactly, our code is designed to handle reconstructions that do not follow bulk stoichiometries. In fact, unlike previous works, our SGC VSSR-MC code does not require the specification of surface stoichiometries or specific stoichiometric ratios to the bulk. During VSSR-MC runs, our code visited many stoichiometries (a total of 683 unique ones) and many do not follow the bulk stoichiometries, e.g., $\text{O}_3\text{7Sr}_{17}\text{Ti}_{18}$ and $\text{O}_{39}\text{Sr}_{15}\text{Ti}_{13}$, and assigned high surface energies to these structures. As a result, the stable surface structures that we find in the phase diagram (Fig. 5(b)) follow the bulk stoichiometries. By "no additional runs varying μO were necessary," we meant that we ran VSSR-MC only at $\mu\text{O} = 0$. This was possible due to the way we are modeling our slab (the pristine surface was stoichiometric to the bulk). A decrease in μO has the same effect on surface free energies as a decrease in μSr by the same amount (Equation 7). We have included these clarifications in Section 2.3.2.

6) DFT with the PBE functional is known to exhibit a very high error for the O_2 molecule in the gas phase. Were any corrections applied to remedy this? If not, could this error influence the presented results?

Thank you for pointing this out. We did not correct for oxygen overbinding in the gas phase. Using the method by Wang et al. [R3] of computing the difference between experimental and DFT formation energies of binary metal oxides, we find each O atom was calculated by DFT to be more stable by around 0.5 eV in the gas phase. This means oxygen should be more likely to adsorb onto surfaces than in our simulations. The results that might be affected would be the surface phase diagram Fig. 5(b) and the surface free energy plots at different μSr , Fig. 5(c-e). Recalculating our surface free energy plots, we find that the DL TiO_2 (Fig. 5(c)) and SL TiO_2 (Fig. 5(d)) terminations are still the lowest energy structures at $\mu\text{Sr} = -10$ eV and $\mu\text{Sr} = -7$ eV respectively as predicted by our NFF and confirmed with DFT. For the SL SrO structure, it became the 12th lowest energy structure at $\mu\text{Sr} = -4$ eV as predicted by our NFF, but according to DFT it still has the lowest surface free energy. Altogether, we should still see three stable phases on the phase diagram, although the cross-over points changed. In Equation 7 we demonstrate that the increase in E_{O_2} is analogous to an increase in μO and thus our new phase diagram would be shifted downwards. Nevertheless, the trend of Fig. 5(b) should stay the same and our conclusions do not change. We have added to Section 4.5 to clarify our procedure.

Reference:

[R3] Wang, L., Maxisch, T. & Ceder, G. Oxidation energies of transition metal oxides within the GGA+U framework. Phys. Rev. B 73, 195107 (2006).

7) The sentence "The Perdew-Burke-Ernzerhof (PBE) functional within the generalized-gradient approximation (GGA) were utilized [53] along with the Hubbard U correction [54] for spin-polarized calculations" does not make sense. The Hubbard U correction is a correction for the self-interaction error in DFT, and it must be stated exactly which U value was used for a given element, since this is an empirical parameter. Please comment on the value used, and whether it is appropriate to use a fixed value for all chemical environments of the given element.

Thank you for spotting this error. We did not use the Hubbard U correction in SrTiO_3 and have revised Section 4.7. We followed the Materials Project guidelines for our calculations. Values of the U parameter for Ti have yet to be calibrated on the Materials Project (<https://docs.materialsproject.org/methodology/materials-methodology/calculation-details/qqa+u-calculations/hubbard-u-values>).

8) Reading through the README on the Github page, I could not find any actual installation instructions? Only System requirements. I may have overlooked this, in that case, please clarify.

We thank the Reviewer for letting us know. The README was written as a procedure to be followed in order. We assumed the reader to be familiar with GitHub and will download the repository to a local directory before proceeding with the instructions. We have updated our Overview and Setup section in the README to make this clearer. Our GitHub repository is available here: <https://github.com/learningmatter-mit/surface-sampling>.

Reviewer #1 (Remarks on code availability):

Reading through the README on the Github page, I could not find any actual installation instructions? Only System requirements. I may have overlooked this, in that case, please clarify.
I have not reviewed the code itself.

NATCOMPUTSCI-23-0568: Machine-learning-accelerated simulations enable heuristic-free surface reconstruction

Reviewer #2 (Remarks to the Author):

The manuscript presents a methodological framework for efficiently exploring complex surface reconstructions of multi-component materials under challenging environments. Accurate determination of surface structures is crucial for understanding material properties and performance in applications such as catalysis and energy storage. However, accurately predicting surface energies and effectively sampling surface compositions and configurations poses significant challenges. Machine-learning force fields (MLFFs) offer a potential solution to these challenges due to their lower computational cost compared to density-functional theory (DFT) calculations and their high accuracy. However, the authors claimed that the development of MLFFs with an efficient sampling scheme that can explore compositional and configurational spaces in a heuristic-free manner is currently lacking. To overcome these limitations, they propose a workflow called Heuristic-Free Surface Reconstruction (HFSurfRecon), which combines MLFFs with active learning (AL) and the Virtual Surface Site Relaxation-Monte Carlo (VSSR-MC) sampling algorithm. The framework is validated through case studies on the GaN(0001) surface using a classical force field and the SrTiO₃(001) surface using a neural network force field. The results demonstrate the effectiveness of the AL strategy in conjunction with VSSR-MC sampling, as it acquires new data points from DFT calculations only in relevant regions of the surface phase space. The generated surface phase diagram aligns well with existing literature and reveals previously unknown low-energy surface terminations. The manuscript is highly relevant to the community, and the study is well-executed. However, some statements made in the manuscript raise questions and require further clarification or justification.

We thank the Reviewer for their detailed evaluation of our manuscript.

Major comments

1. The authors employed the VSSR-MC sampling algorithm to explore compositional and configurational search spaces and to sample complex, stable surface structures. I understand that semi-grand-canonical sampling is only the addition and removal of particles. The authors later (in section 4.1) mention that they also do canonical sampling, but it is unclear what the trial moves are there. Swaps? Simultaneous addition and removal?

The canonical sampling for GaN(0001) was done with swaps between adsorption sites, with the number of Ga adsorbates pre-defined. We have updated Sections 2.2.1 and 4.1.

2. Closely related to comment 1, while the algorithm focuses on high-likelihood virtual sites related to the symmetry of the pristine slab, there is no mention of atomic displacements or rearrangements of coordinates of any kind. In my opinion, this completely fails to explore the configurational space of surface reconstructions, and it is possible that non-trivial surface reconstructions deviating from slab symmetry can be easily missed. And these reconstructions can still exhibit periodicity and should not be considered amorphous (e.g., Si(111) DAS reconstructions). This could be okay, as long as the method essentially samples site occupations, but this is unclear in the text. It would be beneficial for the authors to clarify whether they have considered these possibilities and the potential impact on the sampling process. Or else, the sentence "Our algorithm can be easily applied to any surface of interest and paves the way to study unknown multi-component materials under challenging environments such as in aqueous electrochemistry" is a little bit of an overstatement.

The Reviewer is correct in that the coordinates of the virtual adsorption sites, once defined, are fixed. The discrete sampling step (atom addition, identity change, or removal) is always done at fixed sites. However, as the Reviewer noted in Minor comments 4, we perform relaxation of the surface and adsorbed atoms after each MC move, which allows us to explore the configurational space. The energy and structure saved (if the MC move is accepted) correspond to the relaxed structure. In the GaN(0001) case, we did see periodic surface reconstructions deviating from slab symmetry (the reconstructed layer has a rotated lattice with respect to the pristine surface). We have updated Section 2.2.1. For SrTiO₃(001), the most stable surfaces were shown to be closely aligned with the bulk after many off-lattice structures with lower symmetry were enumerated. We have also added a clarifying figure, Fig. C4, to show the strengths and limitations of our approach.

3. The authors performed latent space clustering based on the geometries obtained from VSSR-MC sampling. They embedded the geometries using the first three principal components, capturing over 90% of the variances. However, the details of how the samples were clustered into different clusters are not clearly explained. The authors should provide further clarification regarding the clustering methodology they employed, as there are various clustering algorithms available that could have been used.

The clustering of our embedded geometries is achieved through hierarchical agglomerative clustering on the first three principal components. This is a bottom-up clustering method where initially, each data point forms its own cluster. Then, pairs of clusters are successively merged based on their closeness, defined by a Euclidean distance metric. We used Ward's method in our linkage computation, which minimizes the total within-cluster variance. The number of clusters can be tuned. To save on compute time, a first-pass clustering was run for every 1000 samples to yield around 100-200 structures at each μ Sr. Structures across all μ Sr were pooled together for a second clustering step to select around 800 structures per AL generation. We have edited Section 4.4 to make our clustering method clearer. We also have an example in our GitHub repository 'tutorials/latent_space_clustering.ipynb'.

4. The authors use two model surfaces of GaN(0001) and SrTiO₃(001) to validate the framework. However, that Si(111) is not mentioned bothers me in particular because it exhibits one of the most complex surface reconstructions known to date, and cheap, accurate potentials are available for it (either Stillinger-Weber or GAP). I think it would have been a much more convincing benchmark than GaN(0001) that then would really justify publication in a prestigious journal like npj Comp. Mat.

We thank the Reviewer for this suggestion. We originally wanted our tool to focus on sampling the free-energy landscapes across compositions and configurations of multi-component materials instead of optimizing the configuration

NATCOMPUTSCI-23-0568: Machine-learning-accelerated simulations enable heuristic-free surface reconstruction

of a particular surface. We have since performed experiments on the Si(111) 3x3, 5x5, and 7x7 surfaces that we summarize in Section 2.2.2 with the Stephenson, Radny & Smith (SRS) modified Stillinger-Weber potential [R4] and have added Fig. C5 to describe the results.

For the 3x3 surface, using the same sampling approach as GaN(0001) (pymatgen virtual site generation, canonical sampling at fixed number of adsorbates, and single annealing step) was sufficient to recover the dimer-adatom-stacking-fault (DAS) reconstruction in Fig. C5(a).

For the larger 5x5 and 7x7 surfaces, we use an additional measure to improve convergence: exponentially decaying distance bias in trial move selection to favor swaps between closer virtual sites. For the 5x5 surface in Fig. C5(b), we sampled DAS-like structures within the thermal energy (25.7 meV/atom) accessible at room temperature in a single MC chain. Including these low-energy structures as seeds in subsequent VSSR-MC runs allowed us to recover the 5x5 DAS target. Additionally, VSSR-MC also sampled the pristine surface, which is about 2 meV/atom of the DAS structure according to the SRS potential.

For the 7x7 surface in Fig. C5(c), we saw various pristine-like structures with four adatoms (due to stoichiometry differences) during our VSSR-MC runs. These surfaces are also similar in energy compared with the DAS target (3.5 meV/atom higher in energy). Using the same strategy as in the 5x5 surface, we similarly obtained low energy, DAS-like structures, which upon further VSSR-MC runs resulted in the 7x7 DAS target structure.

In summary, VSSR-MC was able to thoroughly sample the phase space of Si surface reconstructions and not only recover the DAS structures, but also discover unexpected low-energy structures using the SRS potential. We further describe our methodology in Sections 4.1, 4.2 & 4.8 and update the Abstract, Section 1, and Section 3.

Compared with previous method development studies that used Si(111) as a benchmark, the work by Han et al. [R5] demonstrated only the 3x3 reconstruction. In the GOFEE method by Bisbo and Hammer [R6], they used the full 6mm symmetry to achieve the Si(111) 7x7 reconstruction while our method does not require enforcing cell symmetry *a priori*.

In a work that focused on the comparison of genetic algorithms (GA) and basin hopping (BH) for Si(111) reconstructions, Bauer et al. [R7] utilized custom trial moves for covalent structures that might not translate well to metals and metal oxides. They also started their runs from a perturbed version of the reconstructed surface, unlike the random initial structures we did. Finally, it appears both GA and BH did not discover the pristine and pristine-like reconstructions for 5x5 and 7x7 respectively, which are well within the thermally accessible window. This suggests their methods might not be comprehensive in their exploration of the phase space under non-zero temperature conditions and are better suited to pure energy minimization, rather than free energy minimization, which is the intended objective of VSSR-MC.

References:

- [R4] Stephenson, P. C. L., Radny, M. W. & Smith, P. V. A modified Stillinger-Weber potential for modelling silicon surfaces. *Surface Science* 366, 177–184 (1996).
- [R5] Han, Y. et al. Prediction of surface reconstructions using MAGUS. *The Journal of Chemical Physics* 158, 174109 (2023).
- [R6] Bisbo, M. K. & Hammer, B. Efficient Global Structure Optimization with a Machine-Learned Surrogate Model. *Phys. Rev. Lett.* 124, 086102 (2020).
- [R7] 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).

5. The authors mentioned that a total of 6500 structures for SrTiO₃(001) were generated over six iterations, with fewer than 5,000 structures used for training the final NFF. It thus appears that approximately 80% of the sampled structures required DFT calculations. This raises some ambiguity since the authors selected surfaces based on the highest force standard deviation rather than high energy.

Also, it would be interesting to know if there are structurally similar or nearly identical samples in the training set. If so, wouldn't sparsifying the new samples effectively reduce the total number of required DFT calculations while maintaining the accuracy of the NFF (also potentially speed up the NFF potential)?

Considering the good extrapolability of the NN-based force field, such a large amount of training data may not be necessary to obtain an accurate MLFF for a single surface orientation. As it stands right now, the approach by the authors is not data efficient at all – and thus not really compatible with other already existing approaches in the literature.

We want to clarify that the structures selected from latent space clustering for DFT evaluation constitute a subset of the total number of structures sampled from VSSR-MC. The 6500 structures mentioned here were all evaluated using DFT but fewer than 5000 were used for training because initially we reserved a set of structures at each iteration for validation and testing. In our manuscript, we only showed the MC-sampled test set (which is obtained by running VSSR-MC with the 6th generation NFF model) as we believe it is more representative of our use case. We have updated Section 2.3.1 to clarify our AL data acquisition method and the size of the training dataset.

As in our comment to Reviewer 1, we selected surfaces for DFT evaluation using force variance because it is a better estimate of the NFF uncertainty compared with energy variance.

At each AL generation, we have measures to avoid structure duplication. For our first generation structures and our adversarial attack structures, we used a symmetry equivalence checker ('ase.util.structure_comparator.SymmetryEquivalenceCheck') to identify duplicate structures. In our latent space clustering method, we are already sparsifying our samples by avoiding the selection of similar structures. Without clustering, there would be many more than 6500 structures. While we do not expect duplicates between AL generations because of the large configurational space and each generation using NFFs with different parameters, we agree that it might be worthwhile to consider the similarity of new structures generated at each AL iteration with the existing dataset to reduce the number of DFT calculations.

For a neural network force field, the training speed scales favorably as O(N) in the number of structures (datapoints) and evaluation speed stays constant with the number of structures. Given the relatively small number of

NATCOMPUTSCI-23-0568: Machine-learning-accelerated simulations enable heuristic-free surface reconstruction

training structures, we are not too concerned about slower training speeds. General-purpose universal machine learning potentials are trained on millions of structures.

We politely disagree with the Reviewer that NN-based force fields have good extrapolability and that our approach is not data efficient. Machine learning extrapolation has been a challenge [R1, R8] and NN-based interatomic potentials are notoriously brittle outside of the training distributions (they are *overfit*, as opposed to classical potentials which are typically less accurate but more robust, since they are relatively *underfit*). In other works that we compare to in Table C1 (appended in this Response as Table R1), they require hundreds or thousands of structures to train a machine learning force field for few distinct compositions (usually 10s) and in some instances only one composition. Our NFF model considers as many compositions sampled during our VSSR-MC algorithm runs (more than 600 distinct compositions in our case) and requires fewer than 5000 single-point DFT calculations to achieve a good performance.

Reference:

[R8] Unke, O.T., Chmiela, S., Sauceda, H.E., Gastegger, M., Poltavsky, I., Schütt, K.T., Tkatchenko, A., Müller, K.-R.: Machine Learning Force Fields. *Chemical Reviews* 121(16), 10142–10186 (2021). <https://doi.org/10.1021/acs.chemrev.0c01111>.

6. Although the samples from the VSSR-MC algorithm are obtained by varying the chemical potential of Sr from -12 to 0 eV, it is unclear how the stable surfaces for each chemical potential were selected for further DFT calculations and refinement of the training data. The authors seem to use NFF energy (after screening with force standard deviation) in the selection process. It would be worth considering using surface free energy (by varying the chemical potential) instead of NFF energy to identify the relevant structures for further DFT calculations. This could naturally expand the compositional search space to surface structures stable under more extreme conditions. Currently, Fig 5(c-e) shows a limited number of structures found at $\Gamma \leq -5$ or ≥ 5 with the current settings.

We thank the Reviewer for their comment. The surfaces used for further DFT calculations were obtained either from latent space clustering or from adversarial attack. In the latent space clustering method, structures sampled from VSSR-MC were pooled together after an initial clustering step regardless of the Sr chemical potential for the second clustering step. We have updated Section 4.4 to clarify this procedure. There was no energy consideration in the clustering selection process. In adversarial attack, there were no changes in composition, and thus the NFF energy is monotonic with respect to the surface free energy across configurational changes.

VSSR-MC in the semi-grand canonical ensemble samples according to the grand potential, which is equivalent to the surface free energy after approximations (Equation B10). Fig. 5(c-e) show surface free energy as a function of the difference in the number of Sr and Ti atoms. The surface free energies for structures with $\Gamma \leq -5$ or ≥ 5 are high at all chemical potentials, which is why we naturally sample fewer structures in this range using our semi-grand canonical MC acceptance criterion.

7. Fig 4(b) demonstrates the correlation between force standard deviation and force mean absolute error. However, the details of the figure are not explained in the manuscript. The meaning of the colormap terms "estimated density" and "binned average" is unclear. The authors should provide a more detailed explanation of these terms or simplify the figure to enhance clarity.

We appreciate the Reviewer's observation. We have revised Fig. 4 and its associated description in the manuscript. To elucidate, the term 'estimated density' refers to the estimated probability density at any given point on the plot. This density essentially quantifies the frequency or occurrence of data points in a specific region of the graph. The term 'binned average' is derived by partitioning the force SD represented on the horizontal axis, into uniformly sized bins. Similarly, the force MAE values, indicated on the vertical axis, are also divided into corresponding bins. For each bin, we compute the average force MAE, which is subsequently plotted against the median force SD of the corresponding bin. The 'binned average' thus visualizes the average force MAE for respective force SD ranges. The best-fit line in the graph has been computed using these 'binned average' values to better illustrate the correlation between force SD and force MAE.

8. The active learning workflow for the SrTiO₃ model system achieved decent accuracy of the NFF at the 6th generation employed by the authors. However, it would be helpful to know if the number of generations would differ when applying this workflow to other systems with varying levels of complexity. Is the total number of generations fixed, or are there other criteria for determining when to stop the generations, such as energy or force thresholds?

Thank you for the insightful query. As observed in our study, the NFF accuracy improvements began to plateau between generations 4 and 5 (refer to Fig. 4(d)), and the convergence of VSSR-MC sampled structures from generations 3&4 to 5 is evident (as depicted in Fig. 4(c)). Based on these findings, we halted further VSSR-MC sampling after the 5th generation and carried out a final round of adversarial attack.

We agree that a more precise or automated stopping criterion could be beneficial. An approach could involve reducing the number of new structures per generation to approximately 100-200. This would increase the number of active learning generations, allowing for more granular observation of improvements. The trade-off is the added computational cost associated with training the neural networks.

Using force or energy thresholds as a stopping criterion is an excellent suggestion. However, considering the diversity of systems we work with, setting a universal cutoff point is challenging. We anticipate that as we continue to explore more surfaces, we will be able to establish more robust guidelines regarding the number of generations, structures per generation, and suitable force and energy error thresholds.

We have updated Section 2.3.1 to reflect our point of view.

9. The manuscript does not clearly describe how the authors perturbed the given structure in the adversarial attack process. Since there are no specifically defined collective variables in these systems, further elucidation is needed regarding how the perturbation was implemented in their system. Additional details regarding the methodology of the adversarial attack would be beneficial.

NATCOMPUTSCI-23-0568: Machine-learning-accelerated simulations enable heuristic-free surface reconstruction

We used the published code and many of the default settings in the referenced adversarial attack paper by Schwalbe-Koda et al. [R1], which is described in Section 4.4. We have since rewritten the relevant paragraphs and added more details regarding the adversarial attack methodology and the perturbation method. We clarified the purpose and functionality of the adversarial loss function, \mathcal{L}_{adv} , underscoring that it aims to optimize atomic displacements, denoted as δ to maximize the predicted Boltzmann probabilities and the variance of predicted forces. Additionally, we have provided more detail on how the perturbation was conducted, specifying that each atom's nuclear coordinates are perturbed independently in an effort to maximize the adversarial loss function.

Minor comments

1. On page 2, the authors mentioned that "Machine learning (ML) force fields are much faster than DFT calculations while preserving accuracy [21, 22], but they have not been connected to an efficient sampling scheme to explore the phase space of multi-component surfaces in a heuristic-free fashion." seems a bit overstated. There are many works that connect ML and heuristic-free sampling. The authors should provide more detailed comparisons of their methodologies to others in the literature. At present it seems that they do not have a good overview of the field or try to oversell the novelty of their approach.

It is true that there are several works employing ML and sampling to study surfaces. We apologize for any oversight in citing relevant literature. However, our definition of heuristic-free sampling is that a human is not involved in either the compositional or configurational selection of guess structures. Many of the previous works have the configurational aspect but not the compositional aspect. We have edited Section 1 to include this literature and added a comparison in Table C1 (Table R1 in this Response) with all other cited works. Additionally, we emphasize our active learning algorithm automatically sampled over 600 different compositions of Sr, Ti, and O to achieve an NFF with force MAE of 100 meV/Å and energy MAE of 5.18 meV/atom by training on an average of only 8 structures per composition. This is better than the work of Xu et al. that relied on manual construction of a PbO_2 dataset with 95 distinct compositions at over 500 structures per composition to attain a force MAE of 230 meV/Å and energy MAE of 6 meV/atom. Even though Wanzenböck et al. obtained a superior force MAE of 77 meV/Å and energy MAE of 0.836 meV/atom for their NFF, they fitted on a single composition of SrTiO_3 requiring 3000 single point DFT calculations, 3 orders of magnitude higher than what we required per composition. Finally, we highlight that throughout our work, we only performed a total of 6500 single point DFT calculations compared with 33,000 for Wanzenböck et al. and 250,265 for Xu et al.

2. At least, in fairness to the very active community of surface structure determinations, referring to new MLFF fitting + sampling frameworks (such as that in Ref. 14 and 23) seems to be cutting it a bit short. A few others for a more balanced view might include:

<https://doi.org/10.1103/PhysRevLett.124.086102>

<https://doi.org/10.1103/PhysRevLett.125.206101>

<https://doi.org/10.1063/5.0071249>

<https://doi.org/10.1063/5.0121748>

<https://doi.org/10.1103/PhysRevB.105.245404>

We thank the Reviewer for pointing out these works. Some of these works we have seen, and some are new to us. Again, we apologize for not citing these works earlier. We have included these citations in Section 1. We also appended an additional table in Table C1 (appended in this Response as Table R1) to show the improvement of our HFSurfRecon method over these works and other previously cited works.

NATCOMPUTSCI-23-0568: Machine-learning-accelerated simulations enable heuristic-free surface reconstruction

Reference	Algorithm	Samples across compositions?	Trains a surrogate energy model?	Best surrogate model performance	Elements for most complicated run	Distinct compositions for each system	DFT calculations per composition
Bisbo and Hammer [R9, R9]	GOPEE (EA + GPR)	No	Yes, GPR	Not specified	3 (C and O on V)	~12	>200 single point
Merte et al. [R10]	GOPEE	No	Yes, GPR	Not specified	3 (O on Pt ₃ Sn)	16	>200 single point
Timmermann et al. [R11, R12]	MD, SA + GAP	No	Yes, GAP	E: 25 meV/atom	2 (IrO ₂ , RuO ₂)	20-30	>5 relaxations, ~10 single point
Rønne et al. [R13]	AGOX (BH + GAP)	No	Yes, GAP	E: 103.1 meV/atom	2 (Ag ₂ O ₂)	12	>16 single point
Han et al. [R5]	MAGUS (EA)	Manually specified	No, DFT/pretrained FF	-	2 (SnO ₂ , SiC)	~10	>10 relaxations
Meldgaard et al. [R14]	ASLA (RL + DFT)	No	No, DFT	-	2 (TiO ₂ , SnO ₂)	1	1000 single point
Wang et al. [R15]	USPEX (EA + DFT)	Automatic	No, DFT	-	2 (TiO ₂)	Not specified	1800 total relaxations
Schusterich and Pickard [R16]	AIRSS (RSS + DFT)	Manually specified	No, DFT	-	3 (SrTiO ₃)	5	-
Bauer et al. [R7]	BH and EA	No	No, classical FF	-	1 (Si)	1	-
Wexler et al. [R17]	GCMC BH	Automatic	No, DFT	-	2 (Ag ₂ O ₂)	Not specified	>6000 total relaxations
Hess and Yildiz [R18]	MCMC + SA	No	No, electrostatic forces and DFT	-	4 (La _{0.75} Sr _{0.25} MnO ₃)	1	450 total relaxations
Wanzenböck et al. [R19]	CMA-ES + NFF	No	Yes, NFF	F: 77.01 meV/Å E: 0.836 meV/atom	3 (SrTiO ₃)	1 (training)	3000 single point
Xu et al. [R20]	GCMC + NFF	Automatic	Yes, NFF	F: 230 meV/Å E: 6 meV/atom	2 (PtO ₂)	95 (training)	562 single point
Du et al. (Present work)	HFSurfRecon (VSSR-MC + NFF)	Automatic	Yes, NFF	F: 100 meV/Å E: 9.18 meV/atom	3 (SrTiO ₃)	665	<8 single point

Table R1. Comparison of HFSurfRecon with existing computational methods for surface reconstruction. HFSurfRecon automatically samples across many surface compositions and configurations while training an accurate NFF for low-cost energy prediction. Under "Algorithm", we have evolutionary algorithms (EA), Gaussian process regression (GPR), molecular dynamics (MD), simulated annealing (SA), Gaussian approximation potential (GAP), basin hopping (BH), reinforcement learning (RL), random structure search (RSS), grand canonical Monte Carlo (GCMC), Markov-chain Monte Carlo (MCMC), and covariance matrix adaptation evolution strategy (CMA-ES). Under "Best surrogate model performance", "F" refers to Force MAE while "E" refers to Energy MAE.

References:

- [R9] Bisbo, M. K. & Hammer, B. Global optimization of atomic structure enhanced by machine learning. *Phys. Rev. B* 105, 245404 (2022).
- [R10] Merte, L. R. et al. Structure of an Ultrathin Oxide on Pt₃Sn(111) Solved by Machine Learning Enhanced Global Optimization. *Angewandte Chemie International Edition* 61, e202204244 (2022).
- [R11] Timmermann, J. et al. IrO₂ Surface Complexions Identified through Machine Learning and Surface Investigations. *Phys. Rev. Lett.* 125, 206101 (2020).
- [R12] Timmermann, J. et al. Data-efficient iterative training of Gaussian approximation potentials: Application to surface structure determination of rutile IrO₂ and RuO₂. *The Journal of Chemical Physics* 155, 244107 (2021).
- [R13] Rønne, N. et al. Atomistic structure search using local surrogate model. *The Journal of Chemical Physics* 157, 174115 (2022).
- [R14] Meldgaard, S. A., Mortensen, H. L., Jørgensen, M. S. & Hammer, B. Structure prediction of surface reconstructions by deep reinforcement learning. *J. Phys.: Condens. Matter* 32, 404005 (2020).
- [R15] Wang, Q., Oganov, A. R., Zhu, Q. & Zhou, X.-F. New Reconstructions of the (110) Surface of Rutile TiO₂ Predicted by an Evolutionary Method. *Phys. Rev. Lett.* 113, 266101 (2014).
- [R16] Schusterich, G. & Pickard, C. J. Predicting interface structures: From SrTiO₃ to graphene. *Phys. Rev. B* 90, 035424 (2014).
- [R17] 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).
- [R18] Hess, F. & Yildiz, B. Polar or not polar? The interplay between reconstruction, Sr enrichment, and reduction at the La_{0.75}Sr_{0.25}MnO₃ (001) surface. *Phys. Rev. Mater.* 4, 015801 (2020).
- [R19] Wanzenböck, R. et al. Neural-network-backed evolutionary search for SrTiO₃(110) surface reconstructions. *Digital Discovery* 1, 703–710 (2022).
- [R20] 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).

3. The global optimization literature is immense, and the authors do not even mention e.g. basin hopping, genetic algorithms, AIRSS, etc. Admittedly, there are not too many GO works for surfaces and surface reconstructions, but this is precisely because of the particular complexity of the problem, which makes it expensive at the ab initio level. This is a good motivation for this work but not clearly mentioned. Nonetheless, there are some GO works, so giving a reflection of this literature is appropriate. E.g.:

TiO₂ reconstructions <https://doi.org/10.1103/PhysRevLett.113.266101> (GA/USPEX)

Molecular adlayers: <https://doi.org/10.1021/acs.nanolett.7b01637>

<https://doi.org/10.1002/adv.202000992> and others from the group

<https://doi.org/10.1021/acs.nanolett.5b03388>

SrTiO₃: <https://doi.org/10.1103/PhysRevB.90.035424> (AIRSS)

Si(111): <https://doi.org/10.1021/acs.jpca.2c00647> (basin hopping vs GA)

NATCOMPUTSCI-23-0568: Machine-learning-accelerated simulations enable heuristic-free surface reconstruction

We thank the Reviewer again for providing these references for global optimization. We agree that these references are important – and that the field of global optimization is immense. We do highlight that our algorithm is less of a global optimization for energy, and rather it seeks to sample *free energy*, which necessarily mandates acquiring statistics. We have added references to these works in Section 1 and for the works directly related to surface reconstruction (highlighted above), we have added a comparison in Table C1 (Table R1 in this Response).

4. The authors do perform relaxation after every MC move, so effectively this is a form of basin hopping and it would be nice of them to cite the original paper: <https://doi.org/10.1021/p970984n>

We thank the Reviewer for the reference and have cited the basin hopping paper in Section 1.

5. In section 4.4, a kcal/mol unit has been used, but using a consistent eV unit for energy might be preferable.

We agree and have updated the units in Section 4.4.

6. In Fig 5(b), what are the ticks at the right y-axis (denoted as red and orange ticks with values of 1 and 15) besides scales of oxygen partial pressure? Please add a proper explanation in the caption or main text.

The smaller scales (located on the phase diagram and denoted in orange and red) represent an abbreviated range, only marked at the endpoints of 1 and 15. Correspondingly, the larger scales, stretching in equal intervals from 1 to 15, depict the negative logarithm (-log) of oxygen partial pressure values at distinct temperatures: orange corresponds to -log pO₂ at 900 K, while red is for -log pO₂ at 1250 K. These larger scales provide a detailed division for better resolution of the variation in oxygen partial pressure across these two temperatures. We have summarized these details into the caption in Fig. 5(b).

7. In Fig 5(c-e), what is the specific value of oxygen chemical potential? Please add the text.

$\mu_{\text{O}} = 0$ eV for all subfigures in Fig. 5(c-e). We have updated the figures.

8. The authors mentioned that “The goal of using adversarial attack for the final AL generation was to make the force field more robust for low-energy structures.” Could this be justified?

The purpose of employing adversarial attacks, as presented in the reference paper [R1], is to train an NFF to efficiently encompass the entire phase space of molecular dynamics (MD) simulations. This strategy aims to mitigate the risk of dramatic failures that may occur when encountering infrequent events not captured in the training data. Our use of adversarial attacks shares this objective—we strive to enhance the resilience of our NFF.

However, it is worth noting that our work differs in that we do not perform MD simulations. Given that our VSSR-MC method can automatically reject high energy states, we can focus our attention on bolstering the NFF’s capability to effectively relax lower-energy structures. We have imposed an energy limit on our seed structures at 1000 kcal/mol (43.4 eV) above the lowest NFF energy structure. This allows us to focus solely on structures that lie roughly within the bottom quartile of the NFF energy spectrum. In Fig. 4(c), we demonstrate that the energies of the new structures we generate do not stray excessively high, due to the thermodynamic-likelihood-weighted adversarial loss function.

9. In Fig C2, the test data seems to include very large force components close to 100 eV/Å, and the NFF model surprisingly predicts well of these large forces. Does your final training data also include these high-force configurations? If so, were they only selected and added during the initial generations of NFF when its prediction of energies and forces was poor?

Yes, our training data include these high-force compositions and configurations ranging in magnitude up to 150 eV/Å. The vast majority of high-force structures were added in the first AL generation (random structures) and the two adversarial attack AL generations (generations 2 and 6). Since adversarial attack uses all available data at each generation, generated structures include those with high force components. The three VSSR-MC AL generations produced structures with low force values (mostly 50 eV/Å or less). These forces are not closer to 0 because we did not fully relax structures in the NFF VSSR-MC runs. We have added details on the number of maximum steps for our relaxation runs into Section 4.1. The force distribution over AL generations is summarized in Fig. C6.

NATCOMPUTSCI-23-0568: Machine-learning-accelerated simulations enable heuristic-free surface reconstruction

Reviewer #3 (Remarks to the Author):

Du et al. present a computational technique to obtain atomistic models for ideal, i.e., defect-free surfaces under varying external conditions, which is an important problem in material science with important implications e.g. in heterogeneous catalysis. They present what they call a "heuristic-free surface reconstruction framework" hinging on the "Virtual Surface Site Relaxation-Monte Carlo (VSSR-MC)" sampling algorithm to automate the exploration of configurational space and its concomitant potential energy surface.

We thank the Reviewer for reviewing our manuscript. We would like to emphasize that our VSSR-MC algorithm automatically samples not only the configurational space, but more importantly, the compositional space that many other works do not.

The presented applications of this methodology to the GaN(0001) and SrTiO₃(001) surfaces seem to have been carried out very carefully. Existing atomistic models available in the literature are being reproduced for both surface and suggest additional 2x2-reconstructions for SrTiO₃(001), one of which is energetically favorable and thus might also be experimentally relevant.

Figure 3a) does provide a glimpse of how the virtual adsorption sites have been chosen for GaN(0001).

However, I miss details on how this can be done (heuristic-free!) in general. Since this is an essential and distinguishing feature of the suggested VSSR-MC algorithm, a discussion how such choices can be easily validated - if e.g. specified by parameters that influence the virtual site density.

Our virtual adsorption sites are defined algorithmically using existing software packages: pymatgen and CatKit (see Section 4.2). Our VSSR-MC algorithm is not sensitive to the exact placement of virtual sites, i.e., distance from pristine surface or the site density. Both pymatgen

(`pymatgen.analysis.adsorption.AdsorbateSiteFinder.find_adsorption_sites`) and CatKit (`catkit.gen.adsorption.get_adsorption_sites`) produce well-covered, visually dense adsorption sites with minimal changes from default settings. For GaN(0001), we used both pymatgen and CatKit to generate adsorption sites that worked equally well. The only change from default parameters is to set the symmetry reduce option for both methods to False to obtain more virtual sites. 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 SrTiO₃(001), we only used pymatgen virtual sites.

For SrTiO₃(001), we chose to place the adsorption sites at 1.55 Å (cf. 1.96 Å lattice parameter in the unit cell) from the pristine surface. This does not change the density of sites or site patterns in any way. We also emphasize that this distance is arbitrary; testing different adsorption site distances up to 1.70 Å in early runs showed adsorbates consistently relaxed to a similar distance away from the pristine surface. Similarly, for SrTiO₃(001), we disabled symmetry reduction of sites.

We have updated Section 4.2 to elaborate on the algorithmic generation of virtual sites. We have also included an additional figure to show both the pymatgen and CatKit virtual sites for GaN(0001) as well as our pymatgen virtual sites for SrTiO₃(001) in Fig. C7.

Reconstructions with larger supercells cannot be discovered as the presented approach requires a (surface) super cell to be specified a priori. Consequently, it is not able to explore arbitrarily large NxM-reconstructions (e.g. the famous Au(111) herringbone reconstruction), which might also be based on surface unit cells that are rotated with respect to the primitive surface unit cell. The discussion (in Section 3) mentions this systematic limitation in the context of amorphous overlayers, but does not discuss the aforementioned implications.

Yes, the cell size will have to be specified a priori, as in other works. We also mentioned in Section 3 that "... larger ones [unit cells] could show more complex reconstruction patterns." However, our VSSR-MC algorithm, as demonstrated in the GaN(0001) contracted Ga overlayer reconstruction, can already recover rotated unit cells with respect to the surface unit cell. We have updated Section 2.2.1 to highlight this fact. We have also created an additional figure, Fig. C4 to better convey the strengths and limitations of our approach. Our algorithm can be extended to larger supercells, which might require transfer learning of the interatomic potential trained on smaller unit cells as in Wanzenböck et al. [R18]. We have added a line in Section 3 to reflect the modifications anticipated to apply our work to larger unit cells of multicomponent materials.

Furthermore, the present formulation of the VSSR-MC algorithm neglects vibrational entropy in the underlying surface Gibbs free energy (see Section 4.5 and Appendix B). These can play a crucial role to stabilize particular surface structures (and reconstructions) over others in particular when varying external conditions. Consequently, this limitation should be discussed as well.

We agree with the Reviewer there could be instances where vibrational effects play an important role in determining surface stability, especially at higher temperatures. We have supplemented Section 3 (Discussion).

Altogether, I do not recommend publication of the manuscript in Nature Computational Science in its present form. Considering other computational approaches presented in the literature before (see e.g. Refs. 1 and 2 - neither of which is cited), I fail to see the "groundbreaking" aspects of the work presented here.

[1] S. A. Meldgaard et al., Structure prediction of surface reconstructions by deep reinforcement learning, *J. Phys.: Condens. Matter* 32 404005 (2020).

[2] Yu Han et al., Prediction of surface reconstructions using MAGUS, *J. Chem. Phys.* 158, 174109 (2023).

We thank the Reviewer for pointing out these works. We note that Ref. 2 was only published shortly before our manuscript submission. We have included these works in the references. Our HFSurfRecon framework presented in the submitted manuscript offers several key advantages over the methodologies in Meldgaard et al. and Han et al., making it a superior choice for predicting complex material surfaces. The main strengths of HFSurfRecon are its ability to:

NATCOMPUTSCI-23-0568: Machine-learning-accelerated simulations enable heuristic-free surface reconstruction

1. **Sample across arbitrarily many compositions:** The HFSurfRecon framework enables efficient and accurate exploration of complex, stable surface structures across both compositions and configurations. This advantage is particularly important compared to Meldgaard et al., which requires a fixed composition, and Han et al., which manually specifies limited compositions for the investigated surfaces.
2. **Utilize a simple, heuristic-free sampling strategy:** Our method does not require manual specification of grid sizes, as seen in Meldgaard et al. and does not require manual guesses of the number of adsorbed species, unlike in both references. Our approach to sampling and optimization is more straightforward in contrast to the complicated trial moves and heuristics involved in Han et al. This makes our proposed method more general and less dependent on problem-specific expertise.
3. **Learn accurate ML force fields while sampling:** In Meldgaard et al., the authors themselves note, "... finding the optimal structure is still a time-consuming process due to costly computational chemistry methods such as density functional theory (DFT) and the immense configurational and compositional space." The uncertainty-driven active learning (AL) strategy in HFSurfRecon works in tandem with VSSR-MC sampling to acquire new DFT data points only in 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 (683 in our SrTiO₃ example). This approach is more data-efficient than the RL-based approach in Meldgaard et al., which requires 1,000-10,000 DFT single-point calculations for a single composition, and the genetic-algorithm-based method in Han et al., which neither develops a tandem training-sampling method nor uses an existing active learning framework.
4. **Produce various reconstructions of a 3-component material and beyond:** HFSurfRecon is demonstrated to be effective in predicting multiple phases of the complex SrTiO₃(001) and revealing new low-energy surface terminations. This accomplishment contrasts with Han et al., which is limited to a maximum of two-component structures, SiC and SnO₂, and optimizes for only a single phase in each material. This stands in similar contrast to Meldgaard et al., which focuses on optimizing for only one structure in each of TiO₂(001) and SnO₂(110).

In conclusion, the submitted manuscript offers a more versatile and efficient method for predicting material surfaces and interfaces, overcoming the limitations of Meldgaard et al. and Han et al. We believe HFSurfRecon is a significant advancement that will be a powerful tool for researchers to study a wide range of material surfaces and interfaces under various conditions.

We show a table summarizing the distinguishing features between our method and the two references.

Reference	Algorithm	Samples across compositions?	Trains a surrogate energy model?	Seel surrogate model performance	Elements for most complicated run	Distinct compositions for each system	DFT calculations per composition
Meldgaard et al.	ASLA (RL + DFT)	No	No, DFT	-	2 (TiO ₂ , SnO ₂)	1	1000 single point
Han et al.	MAGUS (EA)	Manually specified	No, DFT/pretrained FF	-	2 (SnO ₂ , SiC)	<10	>10 relaxations
Du et al. (Present work)	HFSurfRecon (VSSR-MC + NFF)	Automatic	Yes, NFF	F: 100 meV/Å E: 6.18 meV/atom	3 (SrTiO ₃)	683	*8 single point

We would also like to refer the Reviewer to our comments to Reviewer 2 and our new Table C1 (Table R1 in this Response) that further illustrate the improvement of our method over a more comprehensive set of previous works.

minor issues:

- Section 2.1, 2nd paragraph:

"Runs are conducted at different external conditions such as pressure, ..., and applied electrical potential."

This is somewhat confusing. If I understand correctly, those two external conditions are not varied in the present work, so the authors should modify/extend this sentence accordingly.

Thank you, we have made the changes.

- Section 4.5 and Appendix B:

Appendix B repeats a lot of the contents of Section 4.5 and does not add much over it. So the authors should reconsider whether the former is really needed.

We acknowledge that Section 4.5 overlaps quite a bit with Appendix B and have removed details from Section 4.5.

Reviewer #3 (Remarks on code availability):

I have glimpsed at the Python source code and had a good impression regarding the coding style. However, due to lack of time I did not manage to run the code for any of the provided examples.

We thank the Reviewer for their confidence in our coding style.

Decision Letter, first revision:

Date: 1st November 23 09:13:37
Last Sent: 1st November 23 09:13:37
Triggered By: Kaitlin McCardle
From: kaitlin.mccardle@us.nature.com
To: rafagb@mit.edu
CC: computationalscience@nature.com
BCC: kaitlin.mccardle@us.nature.com
Subject: AIP Decision on Manuscript NATCOMPUTSCI-23-0568A
Message: Our ref: NATCOMPUTSCI-23-0568A

1st November 2023

Dear Dr. Gomez-Bombarelli,

Thank you for submitting your revised manuscript "Machine-learning-accelerated simulations enable heuristic-free surface reconstruction" (NATCOMPUTSCI-23-0568A). It has now been seen by two of the original referees and their comments are below. The reviewers find that the paper has improved in revision, and therefore we'll be happy in principle to publish it in Nature Computational Science, pending minor revisions to satisfy the referees' final requests and to comply with our editorial and formatting guidelines.

We are now performing detailed checks on your paper and will send you a checklist detailing our editorial and formatting requirements in about a week. Please do not upload the final materials and make any revisions until you receive this additional information from us.

TRANSPARENT PEER REVIEW

Nature Computational Science offers a transparent peer review option for original research manuscripts. We encourage increased transparency in peer review by publishing the reviewer comments, author rebuttal letters and editorial decision letters if the authors agree. Such peer review material is made available as a supplementary peer review file. **Please remember to choose, using the manuscript system, whether or not you want to participate in transparent peer review.**

Please note: we allow redactions to authors' rebuttal and reviewer comments in the interest of confidentiality. If you are concerned about the release of confidential data, please let us know specifically what information you would like to have removed. Please note that we cannot incorporate redactions for any other reasons. Reviewer names will be published in the peer review files if the reviewer signed the comments to authors, or if reviewers explicitly agree to release their name. For more information, please refer to our <a href="https://www.nature.com/documents/nr-

transparent-peer-review.pdf" target="new">FAQ page.

Thank you again for your interest in Nature Computational Science. Please do not hesitate to contact me if you have any questions.

Sincerely,

Kaitlin McCardle, PhD
Senior Editor
Nature Computational Science

ORCID

IMPORTANT: Non-corresponding authors do not have to link their ORCIDs but are encouraged to do so. Please note that it will not be possible to add/modify ORCIDs at proof. Thus, please let your co-authors know that if they wish to have their ORCID added to the paper they must follow the procedure described in the following link prior to acceptance: <https://www.springernature.com/gp/researchers/orcid/orcid-for-nature-research>

Reviewer #1 (Remarks to the Author):

I have reviewed the author's answers to the points raised by both myself and reviewer 2, and I think they have addressed them all satisfactory. I believe the manuscript is now suitable for publication.

Reviewer #1 (Remarks on code availability):

I have only skimmed through the README file. I appreciate the updates made to the Overview and Setup section and have no further comments.

Reviewer #3 (Remarks to the Author):

In my opinion, the authors have addressed many of the reviewers' points clearly and adequately.

However, their response to my question about the VSSR-MC algorithm is very technical, i.e., mainly pointing to routines from the pymatgen and catkit Python packages which the authors employ for creating virtual sites in their code. This does not elucidate how this is / can be done in general, e.g., for stepped metal surfaces where bonding of adatoms is not directional" and thus not favors sharp, distinct sites like in all the examples (only semiconductor and insulator surfaces!) for which the approach is demonstrated here.

Furthermore, I am still very skeptic about calling the approach "heuristic free". The dictionary definition of "heuristic" in the context of computing is "proceeding to a solution by trial and error or by rules that are only loosely defined." For that reason, any computational approach relying on a Monte Carlo technique can hardly be called "heuristic free".

Since these are crucial points of the current study, which have also been brought up by the other reviewers, I expect them to be (better) addressed - so that a broad readership obtains a realistic impression about the approach described here. Therefore, I recommend further revision of the manuscript before I can recommend publication.

minor issues:

- The bibliography still contains a lot of verbatim LaTeX elements and thus requires a type-setting overhaul.

Reviewer #3 (Remarks on code availability):

I have glimpsed at the Python source code during the first iteration and had a good impression regarding the coding style. It is my understanding that the authors have not modified the code since. Consequently, I have not look at the code during this (second) iteration..

Author Rebuttal, first revision:

Reviewer #1:

Remarks to the Author:

I have reviewed the author's answers to the points raised by both myself and reviewer 2, and I think they have addressed them all satisfactory.

I believe the manuscript is now suitable for publication.

[We thank the Reviewer for their positive feedback.](#)

Reviewer #3:

Remarks to the Author:

In my opinion, the authors have addressed many of the reviewers' points clearly and adequately.

[We thank the Reviewer for their positive feedback.](#)

However, their response to my question about the VSSR-MC algorithm is very technical, i.e., mainly pointing to routines from the pymatgen and catkit Python packages which the authors employ for creating virtual sites in their code. This does not elucidate how this is / can be done in general, e.g., for stepped metal surfaces where bonding of adatoms is not directional" and thus not favors sharp, distinct sites like in all the examples (only semiconductor and insulator surfaces!) for which the approach is demonstrated here.

VSSR-MC, with its discrete sampling followed by continuous relaxation strategy, is indeed designed with the flexibility to handle various surface geometries, including stepped metal surfaces. Generally speaking, virtual sites should entirely cover the surface and may overlap. In the case of stepped metal surfaces, the edges and terraces should be covered. Even if the bonding of adatoms is not directional, complete coverage of the surface ensures that the relevant reconstructions can be sampled. More virtual site layers can be added if necessary. The virtual sites can be constructed manually or retrieved from higher-fidelity simulations. We mention pymatgen and CatKit because they

are two well-used tools within the materials science community, and they can automatically generate desired virtual sites with negligible computational cost.

Furthermore, I am still very skeptic about calling the approach "heuristic free". The dictionary definition of "heuristic" in the context of computing is "proceeding to a solution by trial and error or by rules that are only loosely defined." For that reason, any computational approach relying on a Monte Carlo technique can hardly be called "heuristic free".

We have changed all instances of "heuristic-free" to "automatic." "HFSurfRecon" has been changed to "AutoSurfRecon." We believe "automatic" will better describe the behavior of not having manual constructions by humans.

Since these are crucial points of the current study, which have also been brought up by the other reviewers, I expect them to be (better) addressed - so that a broad readership obtains a realistic impression about the approach described here. Therefore, I recommend further revision of the manuscript before I can recommend publication.

We hope the Reviewer will find these changes satisfactory.

minor issues:

- The bibliography still contains a lot of verbatim LaTeX elements and thus requires a type-setting overhaul.

We have fixed issues related to the bibliography section.

Final Decision Letter:

Date: 13th November 23 16:55:27
Last Sent: 13th November 23 16:55:27
Triggered By: Kaitlin McCardle
From: computacionalscience@nature.com
To: rafagb@mit.edu
CC: computacionalscience@nature.com
Subject: Decision on Nature Computational Science submission NATCOMPUTSCI-23-0568B
Message: 13th November 2023

Dear Dr. Gomez-Bombarelli,

I am delighted to tell you that your manuscript NATCOMPUTSCI-23-0568B has been accepted for publication in Nature Computational Science.

As discussed, due to the exceptional nature of your work, we will publish your paper on

an accelerated schedule. **Please carefully review the details below and contact us immediately at computationalscience@nature.com if you have any travel plans or other conflicts that may make you unable to respond to us for the next 5-7 days.**

In approximately 2 business days you will receive a link to choose the appropriate publishing options for your paper and complete the appropriate grant of rights necessary to publish your work. As it is vital that this process not be delayed, we strongly encourage you to [whitelist](https://www.simpleminds.com/how-to-check-your-spam-filter-and-whitelist-emails/) the email address do-not-reply@springernature.com to ensure that this message is received.

You will receive a link to your electronic proof via email with a request to make any necessary corrections as soon as possible. You will find that we have made minor changes to enhance the clarity of the text and to ensure that your paper conforms to the journal's style so we ask that you review these proofs carefully to ensure that we have not inadvertently introduced errors or altered the sense of your text in any way.

Please return your proof within 24 hours of receiving it. If you have any questions about your proofs or anticipate any delays please contact rjsproduction@springernature.com immediately.

Once a publication date is set for your paper, the Springer Nature press office will be in touch with the full embargo details. We request that you do not send out your own publicity or contact any journalists until you hear from us that the paper has a confirmed publication date.

If you would like to inform your Public Relations or Press Office about your paper, we suggest that you do so immediately to allow them as much time as possible to prepare an appropriate press release and organize publicity if they choose to do so. Please include your manuscript tracking number NATCOMPUTSCI-23-0568B and the name of the journal, which they will need if they contact our press office.

Please note that Nature Computational Science is a Transformative Journal (TJ). Authors may publish their research with us through the traditional subscription access route or make their paper immediately open access through payment of an article-processing charge (APC). Authors will not be required to make a final decision about access to their article until it has been accepted. [Find out more about Transformative Journals](https://www.springernature.com/gp/open-research/transformative-journals/)

Authors may need to take specific actions to achieve [compliance](https://www.springernature.com/gp/open-research/funding/policy-compliance-faqs) with funder and institutional open access mandates. If your research is supported by a funder that requires immediate open access (e.g. according to [Plan S principles](https://www.springernature.com/gp/open-research/plan-s-compliance)) then you should select the gold OA route, and we will direct you to the compliant route where possible. For authors selecting the subscription publication route, the journal's standard licensing terms will need to be accepted, including [self-archiving policies](https://www.springernature.com/gp/open-research/policies/journal-policies). Those licensing terms will supersede any other terms that the author or any third party may assert apply to

any version of the manuscript.

If you have any questions about our publishing options, costs, Open Access requirements, or our legal forms, please contact ASJournals@springernature.com.

If you have not already done so, we strongly recommend that you upload the step-by-step protocols used in this manuscript to the Protocol Exchange. Protocol Exchange is an open online resource that allows researchers to share their detailed experimental know-how. All uploaded protocols are made freely available, assigned DOIs for ease of citation and fully searchable through nature.com. Protocols can be linked to any publications in which they are used and will be linked to from your article. You can also establish a dedicated page to collect all your lab Protocols. By uploading your Protocols to Protocol Exchange, you are enabling researchers to more readily reproduce or adapt the methodology you use, as well as increasing the visibility of your protocols and papers. Upload your Protocols at www.nature.com/protocolexchange/. Further information can be found at www.nature.com/protocolexchange/about.

You can now use a single sign-on for all your accounts, view the status of all your manuscript submissions and reviews, access usage statistics for your published articles and download a record of your refereeing activity for the Nature journals.

An online order form for reprints of your paper is available at <https://www.nature.com/reprints/author-reprints.html>. All co-authors, authors' institutions and authors' funding agencies can order reprints using the form appropriate to their geographical region.

Sincerely,
Kaitie

--

Kaitlin McCardle, PhD
Senior Editor
Nature Computational Science

P.S. Click [here](#) if you would like to recommend Nature Computational Science to your librarian - this will link directly to the Recommend page.

<http://www.nature.com/subscriptions/recommend.html#forms>

** Visit the Springer Nature Editorial and Publishing website at <https://group.springernature.com/gp/group/careers/editorial> for more information about our career opportunities. If you have any questions please click [here](mailto:editorial.publishing.jobs@springernature.com). **