The recent developments in machine learning (ML) technology have introduced new functional materials for applications such as photo-voltaics, thermoelectrics, electrolytes, and heterogeneous catalysts. Often, advances in creating new such materials are lead by experimental discoveries and subsequent characterization of the compounds serve to identify their precise structure and composition at the atomic scale. Once the structure of a material is known, its physico-chemical properties may be rationalized on theoretical grounds providing a basis for further improvements. The process of structure determination is highly non-trivial and is hampered by the rapid growth with problem size of both the configurational space and the local relaxations can be circumvented. Typically, the process of structure determination is subjected to a computationally inexpensive local relaxation of the structural candidates before the single-point energy evaluation with the computationally expensive first-principles single-point calculation is conducted. The Atomistic Structure Learning Algorithm (ASLA) proposes a new structure-stability datapoint to train the neural network.

I. INTRODUCTION

Structure determination is a cornerstone in the development of new functional materials for applications as photo-voltaics, thermoelectrics, electrolytes, and heterogeneous catalysts. Often, advances in creating new such materials are lead by experimental discoveries and subsequent characterization of the compounds serve to identify their precise structure and composition at the atomic scale. Once the structure of a material is known, its physico-chemical properties may be rationalized on theoretical grounds providing a basis for further improvements. The process of structure determination is highly non-trivial and is hampered by the rapid growth with problem size of both the configurational space and the local relaxations can be circumvented. Typically, the process of structure determination is subjected to a computationally inexpensive local relaxation of the structural candidates before the single-point energy evaluation with the computationally expensive first-principles single-point calculation is conducted. The Atomistic Structure Learning Algorithm (ASLA) proposes a new structure-stability datapoint to train the neural network.

In this work, we extend ASLA to include a surrogate energy landscape that is constructed on-the-fly. Whenever ASLA proposes a new structure, the structure is subjected to a computationally inexpensive local relaxation in the model before the computationally expensive single-point energy calculation is conducted. The method is shown to dramatically increase the performance of ASLA when applied to solving for the most stable structure-stability datapoint to train the neural network.

Further we apply this model-enhanced ASLA in a thorough investigation of the c(4 × 8) phase of the Ag(111) surface oxide. ASLA successfully identifies a surface reconstruction which has previously only been guessed on the basis of scanning tunnelling microscopy images.
The paper is outlined as follows. In the first section, we introduce the details of the model. In the second section, we demonstrate the performance increase for a simple benzene system, where many restarts can be made and good statistics obtained. Lastly, in the final section, we employ our algorithm to identify the aforementioned Ag(111) oxide surface structure. Doing so, we employ our algorithm to identify the aforementioned Ag(111) oxide surface structure. The reinforcement learning episode proceeds with the evaluation phase which is a single-point DFT calculation of the total energy, $E_{\text{tot}}$, of structure, $s_T$. This means that ordinarily it is the expected most rewarding action (greedy), which is chosen, but occasionally with some small likelihood (epsilon) a stochastic element is used in choosing the action.

After constructing a candidate from the policy we add in the present version of ASLA a new element, namely a structural relaxation. The changes that a structural candidate undergoes during this are illustrated in Fig. 2c. The structure relaxation seeks to identify the nearest local minimum energy structure in the vicinity of the structure just built. To avoid any significant computational expense of this operation, it is carried out in a surrogate energy landscape described in the next section. After completing the relaxation, the atom positions are adjusted slightly (snapped to grid) to ensure that the relaxed structure, now $s_T$, can be represented on the grid used for training the agent.

The reinforcement learning episode proceeds with the evaluation phase which is a single-point DFT calculation of the total energy, $E_{\text{tot}}$, of structure, $s_T$. This means that ordinarily it is the expected most rewarding action (greedy), which is chosen, but occasionally with some small likelihood (epsilon) a stochastic element is used in choosing the action.

II. ASLA

In the Atomistic Structure Learning Algorithm (ASLA)\textsuperscript{46,50,51}, the global optimization problem of structure search is formulated as a reinforcement learning problem, where an agent learns to build the most stable atomistic structure. The learning appears in episodes, in each of which, the agent itself creates a new prospective atomistic structure whose total energy is provided by an external density functional theory (DFT) program\textsuperscript{52,53}. Each learning episode contains three distinct phases, the building phase, the evaluation phase, and the training phase as illustrated in the flowchart given in Fig. 1.

In the building phase, ASLA commences by constructing a new structural candidate by following the policy as guided by the current state of the agent. It starts from a template structure, $s_0$, (possibly an empty computational cell) and proceeds iteratively to a final state, $s_T$, such that a total of $N_{\text{atom}}$ atoms have been placed, as further detailed in Fig. 2a. In each iteration the intermediate state, $s_t$, is given to the agent equipped with a convolutional neural network (CNN), that responds with its action, $a_t$, taken at state $s_t$ leading to state $s_{t+1}$. We employ a modified epsilon-greedy policy. This means that ordinarily it is the expected most rewarding action (greedy), which is chosen, but occasionally with some small likelihood (epsilon) a stochastic element is used in choosing the action.

FIG. 1. Flowchart showing the three phases of ASLA. The green nodes highlight the contributions from this paper. The blue polygon represents the CNN producing the $Q$-value map, while the surrogate energy model, in which the relaxation is performed, is schematically represented as the blue curve in the structure-energy plot. The true energy curve is shown in black and black dots are training points, while the shaded area marks the model uncertainty. Both the CNN and the surrogate energy model improve in each episode, leaving ASLA better equipped for building low-energy structures in later episodes.
A candidate structure is constructed atom-by-atom in the candidate-from-policy operation by iteratively consulting a CNN (shown in b) and following the modified epsilon-greedy policy described in the text. This results in the sequence of structures, $s_0, \ldots, s_T$. The final structure, $s_T$, is relaxed in the surrogate energy model, resulting in the $\tilde{s}_T$-structure, for which the total energy is evaluated. The $\tilde{s}_T$-structure is decomposed atom-by-atom to obtain the sequence, $\tilde{s}_0, \ldots, \tilde{s}_T$, which is stored in the ASLA memory together with the energy, and used for training of both the CNN and the surrogate energy model. Date construction was preceded by some relaxation, the state-action pairs will most likely deviate from the actual states, $s_t$, visited and actual actions, $a_t$, taken during the build action. Assuming, however, that atoms are placed in the same order, the state-action pairs, $(\tilde{s}_t, \tilde{a}_t)$, that would have led directly to the snapped, relaxed structure, $\tilde{s}_T$, can be determined, cf. Fig. 2d, and stored with $E_{\text{tot}}$ in the memory.

The reinforcement learning episode continues with a training phase, where the CNN and surrogate energy model are updated. The CNN is trained by first extracting a batch of state-action-energy pairs from the most recent, most-favorable and random episodes from the memory, such that the batch comprises $5 \times N_{\text{atom}}$ state-action-energy pairs. The energies are transformed to reward values, $r \in [-1, 1]$, and the batch is expanded by rotated and mirrored versions of the state-action-energy pairs. The weights of the CNN are updated by a backpropagation step with a learning rate of $10^{-3}$ that decreases the total mean squared error between the predicted value and the reward for state-action-energy pairs in the batch.

As a final operation during the learning phase of a reinforcement learning episode, an improved surrogate energy model is prepared for the next episode. This is done in the train-surrogate-model operation, which represents the final new element introduced to ASLA in this work. The surrogate energy model extracts from the memory $(\tilde{s}_T, E_{\text{tot}})$ datapoints and constructs a structure-energy model. The details of the model are given in the next section.

Upon starting the reinforcement learning cycle, the neural network representing the agent can be initialized randomly meaning that ASLA learns completely autonomously from its interaction with the DFT program with no other input than the template, $s_0$, and the stoichiometry of the final structure, $s_T$. It may, however, benefit largely in a transfer learning setting, where network weights are inherited from a prior ASLA run solving a similar task.
pler problem. Initially, before sufficient structure-energy datapoints have been collected and stored in memory, a sufficiently reliable surrogate model cannot be established, and the relax operation is simply skipped.

III. SURROGATE MODEL

For the surrogate model, a Gaussian process (GP) regression model is used. The model is trained on feature representations of the final structures and their total energies. The use of a feature representation means that the GP model incorporates e.g. rotational and translational energy invariances directly, unlike the CNN model used for the agent, which has to learn that via the applied data augmentation. The GP regression model, thus learns the underlying energy landscape more efficiently from the training data compared to what would be expected from using a CNN model. However, contrary to using a CNN for the Q-value map evaluation, using a GP model would be computationally expensive owing to the cost of evaluating the feature representation at every pixel in the map. It can therefore not form the basis for the policy guided building step in ASLA, where a CNN with its inexpensive image representation is preferred. A further reason why a CNN is preferred for the feature map prediction is that Q-values depend on the final structures that can be built from incomplete ones. Thus the Q-value evaluation must predict the consequences of future actions, something which is better done with the deep learning capabilities of the CNN.

A GP is specified by its prior mean, \( \mu(x) \), and covariance function (the kernel), \( k(x, x') \), which expresses the beliefs about the system prior to any observations. The training consists of inferring the posterior distribution, which takes into account the observed training data, \( D = \{X, E\} \). The posterior can be analytically expressed due to the well-behaved marginalization properties of the GP. Namely, the posterior is also a GP with mean function given by

\[
E(x_*) = k^T(X) (K + \sigma_n^2 I)^{-1} [E - \mu(x)] + \mu(x)
\]

where \( K = k(X, X) \) is the kernel matrix, \( k_* = k(X, x_*) \) and \( \sigma_n \) is a regularization hyperparameter that acts to prevent overfitting (we use \( \sigma_n = 3.2 \cdot 10^{-3} \)). The posterior mean is taken as the model prediction of the energy. Computationally, the training is dominated by the inversion of the kernel matrix, which scales with the number of data points as \( O(n^3) \).

To take advantage of the rotational, translational and permutation symmetries of quantum chemistry, structures are represented by feature vectors that exhibit these symmetries. In this work, we choose the fingerprint descriptors of Oganov and Valle.

The kernel is chosen to be a sum of two Gaussians with different length scales, similar to that proposed in,

\[
k(x, x') = \theta_0 ((1 - \beta) e^{-|x - x'|^2 / 2 \lambda_1^2} + \beta e^{-|x - x'|^2 / 2 \lambda_2^2})
\]

where \( x \) are the feature vectors and \( \theta_0 \) is the maximal covariance. Using two length scales, \( \lambda_1 \) and \( \lambda_2 \), improves the model by allowing the kernel to capture trends in a large, sparsely sampled configuration space, while maintaining resolution in smaller, more densely sampled regions. The prior mean is chosen to be a short-ranged repulsive potential, which is generally present between atoms. This naturally prohibits atoms to get too close during structure relaxations with the model, which may cause convergence problems when the total energy is evaluated. The feature vector is analytically differentiable with respect to the Cartesian coordinates, which allows us to obtain the predicted force according to the model. The relaxation is performed by iteratively moving the atoms based on the predicted force.

In training the model, the hyperparameters, \((\theta_0, \lambda_1, \lambda_2)\), are optimized by maximizing the log marginal likelihood. For the training data, only the 500 lowest energy structures from the ASLA memory are used. This ensures that the model resolves the low-energy part of the configuration space, while keeping the computational training time small. No effort is thus made to capture high-energy structures in irrelevant parts of the configuration space.

The model is not meant to reproduce total energies with high accuracy. Rather, the purpose of the model is to guide ASLA into local minima of the configuration space in order to aid the global search. Therefore, we measure the models performance by the improvement in the number of ASLA episodes needed to solve a global optimization problem, instead of an error with respect to a test set.

IV. BENZENE EXAMPLE

Having introduced to ASLA that candidates built from the policy may be relaxed in a surrogate energy landscape, we now turn to probe how it affects the overall performance of ASLA. In order to be able to make convincing statistics we choose to have ASLA solve the problem of building the most stable molecule given six carbon atoms and six hydrogen atoms, i.e. \( \text{C}_6\text{H}_6 \). The total energy expression is provided by a density functional based tight binding (DFTB) energy expression using the DFTB+ implementation since this provides a sufficiently accurate description that the correct chemical bonds are formed, yet at a much reduced computational cost compared to a DFT energy expression. As the solution turns out to be the planar molecule benzene, we address this problem with a 2D space. For this a grid spacing of 0.25 Å is used.

The performance of ASLA is gauged by restarting the structural search a great number of times with no data in the memory and with random initialization of the CNN. By compiling as a function of episode count, the share of search runs that have found the benzene molecule, we arrive at the success curves displayed in Fig. 3. The blue curve shows the performance of standard ASLA without
model relaxations, while the green curve shows the performance when the model relaxations are included. The here-proposed ASLA with model relaxations outperforms the standard ASLA by a wide margin. While standard ASLA needs $\approx 7000$ episodes for 10\% of the restarts to identify benzene, ASLA with model relaxations attains this after less than 1000 episodes.

The green success curve for the model-relaxed ASLA shows, however, one convergence issue. The fact that it levels off at about 60\% success after about 2000 episodes means that the 40\% of the restarts that have not found benzene within the first 2000 episodes stand a small chance of doing so in the remaining 6000 episodes of the runs. To probe the origin of this, we conducted a set of search runs, in which the relaxations of the policy-built candidate structures were done in the full DFTB energy landscape rather than with the on-the-fly learnt surrogate energy landscape. The grey curve in Fig. 3 shows the resulting success curve, which evidences that when relaxing candidates in the DFTB energy landscape, the stagnation almost vanishes. About 80\% of the restarts have found benzene after 2000 episodes, and in the following 6000 episodes about half of the remaining restarts do so.

We consequently attribute the stagnation of the model-relaxed ASLA to errors in the surrogate energy model. Note, however, that the DFTB relaxation scheme uses several order of magnitudes more DFTB energy evaluations per episode, than does the model relaxation scheme, where one episode corresponds to one DFTB single-point energy evaluation. This renders the model relaxation scheme far superior when more refined and computationally expensive DFT or quantum chemistry methods are used as the total energy expression.

Possible solutions to the stagnation issue could be the introduction of more advanced model energy expressions than the presently used one, or the development of schemes to reset the surrogate model upon detection of stagnation. Note that a single restart will not know if it has ceased to find better structures because it has identified the global minimum energy structure or because it uses an insufficiently accurate model for the relaxation and finds a higher lying energy structure. In practice, when searching for an unknown global minimum energy structure, all restarts would therefore need to have their model reset once they consistently produce the same best

![FIG. 3. Success curves based on 200 restarts. The standard ASLA (blue curve) is outperformed by ASLA augmented with model relaxations (green curve), without any additional energy evaluations. The grey curve shows the performance when DFTB relaxations are used in place of the model relaxations, imitating a perfectly trained model.](image)

![FIG. 4. The structures built according to the $Q$-value directed policy (left) and their appearance being relaxed in the model and snapped to the grid (right) for different stages of a search. The model potential quickly learns to locally optimize the structure that ASLA produces. In turn, ASLA learns more quickly to build stable structures (cf. Fig. 3).](image)
Having seen that ASLA improves significantly upon adding the presently proposed model-relaxation, it is instructive to inspect the degree of relaxation as a function of reinforcement episodes. Figure 4 presents one such study for a randomly chosen restart. It shows the policy-built structure and the corresponding model-relaxed structure after 100, 200, 500, and 900 episodes. After 100 and 200 episodes, the agents CNN has not developed sufficiently yet as to consistently build molecules with chemical meaningful coordinations. However, the surrogate energy model has learnt enough to provide for relaxation of the atoms into having more proper interatomic distances. After 500 episodes, the policy-built structure still lacks a bit on C-C coordination and C-H bond lengths and it leaves some H as isolated atoms. All of these deficiencies are remedied by the model-relaxation and by 900 episodes, the agents starts to know how to build benzene in need of only very minor bond adjustments, mainly pertaining to the large grid spacing used. Note, that the CNN benefits from the improved training examples provided by the model relaxations, and learns to build reasonable structures faster than in the standard ASLA scheme.

V. AG(111) OXIDE SURFACE

We now turn to apply ASLA with model relaxation to an outstanding problem in materials science. Specifically, we will be concerned with the oxidation of the Ag(111)-surface, which has been shown to exhibit a rich variety of stoichiometries and phases during growth of surface oxide layers. Schnadt and co-workers have reported scanning tunneling microscopy (STM) topographs of a number of such phases, one of which, the \(c(4 \times 8)\)-phase, is reproduced in Fig. 5. To the best of our knowledge,
no systematic structural search has been carried out for this particular surface oxide phase on Ag(111), yet in the original paper a structural model was put forward\[32\]. By having ASLA perform the structural search given the experimentally determined surface unit cell, but covering a large range of Ag-O stoichiometries for the surface oxide, we confirm below the originally conjectured model in a set of ASLA runs.

In our search for the Ag(111)-c(4 × 8) silver oxide phase, ASLA is used to identify the best possible structure of a mono-layer silver oxide on top of a bulk truncated Ag(111) template. For the surface oxide, it is assumed that all Ag atoms occupy positions whose heights coincide with that of an extra unrelaxed Ag(111) layer, z₀, while all O atoms occupy some of seven different heights evenly distributed around such a Ag(111) layer, i.e. ranging from 1.6 to 3.1 Å above the Ag(111) template. For the in-plane positions of the Ag and O atoms, any values on a 2D grid with grid spacing 0.255 Å are allowed. Figure 6 depicts these discretized positions as they are fed into the neural network. The seven brown layers hold the O atoms, while the three grey layers hold the Ag atoms. As a template of preplaced Ag atoms, the bottom two grey layers are prepopulated with two Ag(111) layers. The CNN, shown schematically as the blue polygon in Fig. 6, outputs Q-values for all oxygen layers and for the upper Ag layer as shown by the bluish raster plots in the figure. The network does not need to output any Q-values for the lower two Ag layers, since these are already fully occupied.

During the model relaxation the silver atoms are constrained to move in the xy-plane, while the oxygen atoms are constrained to move freely between z = 1.6 Å and z = 3.1 Å, which does not represent any further approximation given that the atomic position are snapped to the grid heights within these bounds at the end of the building phase.

\[ \Delta G(T, p) = E_{\text{DFT}} - E_{\text{slab}}^{\text{DFT}} - X\mu_{\text{Ag}} - Y\mu_{\text{O}}(T, p), \]

where \( E_{\text{DFT}} \) is the DFT energy of the full structure, \( E_{\text{slab}}^{\text{DFT}} \) is the DFT energy of the Ag(111)-c(4 × 8) five layer slab without the surface oxide, and where \( \mu_{\text{Ag}} \) and \( \mu_{\text{O}} \) are the chemical potentials of silver and oxygen, respectively. Note, that we neglect the small vibrational and configurational contributions to the Gibbs free energy of the surface structures. The silver chemical potential, \( \mu_{\text{Ag}} \), is taken to be the chemical potential of a silver atom situated in the bulk position, calculated as the difference per Ag atom of 6 and 5 layer thick Ag(111) slabs. The oxygen chemical potential is a function of the di-oxygen chemical potential.
pressure, \( p \), and the temperature, \( T \):

\[
\mu_O(T, p) = \frac{1}{2} \left( \bar{E}_\text{DFT}^{O_2} + \tilde{\mu}_O(T, p^0) + k_B T \ln \left( \frac{p}{p^0} \right) \right)
\]

(4)

where \( E_{\text{DFT}}^{O_2} \) is the DFT total energy of the di-oxygen molecule, \( \tilde{\mu}_O(T, p^0) \) is the translational, rotational, and vibrational contributions to the free energy of an \( O_2 \) gas at a reference pressure, \( p^0 \), and \( k_B \) is Boltzmann's constant. It is seen that the absolute value of \( \mu_O(T, p) \) depends on the specific computational settings through \( E_{\text{DFT}}^{O_2} \). To circumvent that, it is convenient to quote only the \((T, p)\)-dependent part of the chemical potential:

\[
\Delta \mu_O(T, p) = \mu_O(T, p) - \frac{1}{2} E_{\text{DFT}}^{O_2}
\]

(5)

whenever a chemical potential is specified.

With the thermodynamic considerations in place, it is now possible to compare the stability of the most stable structures found by ASLA. Figure 7 does so in the form of a raster plot of the free energy, \( \Delta G \), evaluated at an oxygen chemical potential of \( \Delta \mu_O = -0.5 \text{ eV} \), corresponding approximately to ambient conditions, i.e. a pressure of 1 atm and a temperature of 300 K. The plot shows a clear optimal stability for a surface oxide of \( \text{Ag}_5\text{O}_4 \) stoichiometry, and it brings evidence that sufficient variation in the stoichiometry has been considered to call this the thermodynamic most stable state at this chemical potential for oxygen.

In Fig. 8 a diagram of the free energy as a function of \( \Delta \mu_O \) is shown. The diagram builds on the most stable \( \text{Ag}_X\text{O}_Y \) structure for every considered value of \( Y \), i.e. \( \text{Ag}_4\text{O}_2, \text{Ag}_6\text{O}_3, \text{Ag}_5\text{O}_4, \text{Ag}_6\text{O}_5, \) and \( \text{Ag}_4\text{O}_6 \) surface oxides. The diagram shows that over a wide range of chemical potential for oxygen, the \( \text{Ag}_5\text{O}_4 \) surface oxide remains the most stable.

The structures leading to the free energy diagram in Fig. 8 are shown in Fig. 9. They expose a rich variety of chemical bonding motifs involving the Ag and O atoms within the surface oxide. Inspecting the structures in Fig. 9, the first one appears network type with voids, the next one resembles the first, but with reduced \( \text{Ag}_4 \) islands in the voids, while the remaining three structures appear to have both Ag and O atoms highly dispersed in the surface oxide layer. As the oxygen content increases from left to right in the figure, ASLA eventually identifies the need for including \( O_2 \) motifs when tasked with accommodating a large amount of oxygen within the surface oxide.
Ag$_3$O$_2$, where it is seen that ASLA in independent searches for one O or contain an extra O compared to the optimal.

5. Likewise, Fig. 10d-f shows how Ag$_2$O$_2$ and Ag$_3$O$_2$ stoichiometries find again the optimal Ag$_2$O$_2$ structure only with one Ag atom removed or added. Likewise, Fig. 10a-c, where it is seen that ASLA in independent searches for Ag$_3$O$_2$ and Ag$_3$O$_2$ stoichiometries finds again the optimal Ag$_3$O$_2$ structure and of the sub-optimal structures found, it can be realized that ASLA often finds the same structural skeleton for diverse set of chemically meaningful structures as the stoichiometry is varied. Conversely, collecting series of some of chemical bonding within the surface oxide. It does so without requiring any human input of expected structural motifs or other presumed knowledge of chemical bonding within the surface oxide.

VI. CONCLUSION

We have augmented ASLA with relaxations in a model potential, dramatically increasing the performance. The improvements come with no need for extra electronic-structure energy evaluations, and little overhead due to training the model and the relaxation procedure itself. We applied the improved ASLA to a Ag(111) oxide surface, where a proposed structure was confirmed by a thorough search.

VII. ACKNOWLEDGEMENTS

We acknowledge support from VILLUM FONDEN (Investigator grant, Project No. 16562).
