



Cite this: DOI: 10.1039/d5cp03121d

## Completing the hierarchy of rotational defects in monolayer MoS<sub>2</sub> through symmetry-aware evolutionary search

Alexander Adel, Ralf Wanzenböck and Georg K. H. Madsen \*

Monolayer molybdenum disulfide (MoS<sub>2</sub>) shows a plethora of defect configurations, which constitutes the basis for tailoring material properties through defect engineering. Detailed characterization of these defects remains challenging due to the complexity of the potential energy surface. We efficiently explore the three-fold rotational defect potential energy surface in monolayer MoS<sub>2</sub> by combining an evolutionary algorithm with a machine-learning force field. To improve the performance of the structure searches, the algorithm hierarchically restricts the exploration process to a lower-dimensional subspace, utilizing the symmetry operators associated with the investigated defects. We demonstrate that these constrained trajectories exhibit lower global uncertainty measures during the evolution, produce final structures with lower energy distributions and converge faster. Our approach results in the discovery of several novel structures with reasonable computational effort, thereby completing the hierarchy of rotational defects in MoS<sub>2</sub>.

Received 14th August 2025,  
Accepted 10th November 2025

DOI: 10.1039/d5cp03121d

rsc.li/pccp

### Introduction

As one of the most prominent members of the transition-metal-chalcogenides family, molybdenum disulfide (MoS<sub>2</sub>) features extensive studies regarding its layered structure, from the preparation of very thin crystals<sup>1–3</sup> to the isolation of individual crystal planes.<sup>4</sup> Monolayer MoS<sub>2</sub> consists of three atomic layers of alternating Mo and S. Notably, this two-dimensional material is a direct band gap semiconductor.<sup>5–7</sup> Monolayer MoS<sub>2</sub> shows a plethora of defect configurations. The size and complexity of these configurations range from small native point defects such as vacancies, antisites, adatoms, and interstitials<sup>8–10</sup> to large structural disturbances such as extended line defects<sup>11–13</sup> and grain boundaries.<sup>14,15</sup> This large variety constitutes the basis for the manipulation of material properties and defect engineering.<sup>16–21</sup>

A specific type of defect in MoS<sub>2</sub> are the topological defects obtained when six S atoms (three in the upper and three in the lower atomic layer) are rotated by 60° around an Mo atom in the central layer.<sup>22</sup> Parallel to how the well-known Stone–Wales defects in graphene<sup>23</sup> can order in extended patterns,<sup>24</sup> the rotational MoS<sub>2</sub> defects, in conjunction with sulfur double vacancies, have been found to produce quite extensive reconstructions, dependent on the number of these transformations occurring in the vicinity of each other.<sup>22</sup> As the defect structures

grow in size, so does the complexity of the energy landscape<sup>25</sup> and finding the stable structures by constructing viable low-energy defect configurations according to domain knowledge becomes untenable.

Evolution strategies are a frequently used approach for rugged energy landscapes.<sup>26–28</sup> To enhance the efficiency of global atomistic structure optimizations, search algorithms can exploit the inherent symmetries of the system under investigation. One approach is to bias the potential energy surface directly,<sup>29</sup> but a conceptually simpler approach is to restrict the search space to configurations that obey predefined symmetry constraints.<sup>30–34</sup> This reduction in the number of degrees of freedom narrows the configurational space and allows the algorithm to identify plausible structures more efficiently.

In general, stochastic approaches like evolution strategies still require a substantial volume of fitness evaluations. To make the calculations feasible, highly parametrized models based on machine learning have been used.<sup>35–37</sup> In this paper, we show that the utilization of machine-learning force fields (MLFF), combined with the systematic reduction of the degrees of freedom using symmetry, makes the efficient treatment of complex defect structures possible. Specifically, we apply our constrained evolutionary algorithm to investigate the aforementioned large-scale rotational defects in monolayer MoS<sub>2</sub>.

The paper is organized as follows. We start with a description of the process of the constrained structure search and the generation of the data set required for the MLFF training. Next, we discuss the characteristics of the constrained trajectories



based on the predicted energies and force uncertainties obtained by the MLFF. This is followed by a presentation of the resulting structures in the rotational defect hierarchy. Finally, the MLFF is also applied to monolayer MoS<sub>2</sub> grain boundaries. The paper ends with a conclusion and an outlook.

## Methods

### DFT calculations

The reference values for the data sets were obtained by performing spin-polarized calculations with the VASP<sup>38,39</sup> implementation of the projector augmented-wave formalism,<sup>40</sup> using the Perdew–Burke–Ernzerhof (PBE) functional.<sup>41</sup> The energy cutoff was set to 258 eV (the default value for sulfur) and the break condition for the electronic self-consistent loop to  $10^{-6}$  eV. Only the  $\Gamma$ -point was used for the defected supercell and the width of Gaussian smearing was set to  $10^{-4}$  eV. All supercells were set up with one monolayer separated from the vertical borders of the supercell by vacuum with a thickness of 22 Å. Convergence tests of the energy cutoff and  $k$ -mesh were performed by comparing the energy difference of the most stable structures found with symmetry constraint  $C_3 + \sigma_h$  and with symmetry constraint  $\sigma_h$ . The energy difference was found to change by less than 0.01 eV compared to a total energy difference of 1.17 eV.

### Structure search

We utilize the covariance matrix adaptation evolution strategy (CMA-ES)<sup>28</sup> as implemented and tailored to atomistic structure searches in CLINAMEN2.<sup>42</sup> The CLINAMEN2 code was extended by a module implementing symmetry operations such as rotations and reflections. Before the random sampling step takes place, all individual structures of the current population are reduced to the asymmetric unit corresponding to the selected symmetry constraint. After sampling, the structures are rebuilt so that the energy evaluation step is performed for the complete structures.

### Symmetry operations

The relevant point group of the three-fold rotational defects is  $D_{3h}$ . We considered four symmetry constraints, none ( $C_1$ ) as a baseline, a horizontal mirror plane ( $\sigma_h$ ), a threefold rotation axis perpendicular to the MoS<sub>2</sub> plane ( $C_3$ ), and the combination of  $C_3$  and  $\sigma_h$ . The two symmetry operators are illustrated in Fig. 1, together with the initial or founder structure used for the smallest defect.

### 3V<sub>S<sub>2</sub></sub> data set generation

The data set for MoS<sub>2</sub> structures with three sulfur double vacancies inside a (5 × 5 × 1) supercell (symbolized by 3V<sub>S<sub>2</sub></sub>) was constructed as follows: for each combination of symmetry operations, one CLINAMEN2 evolution (with Vasp as evaluation back-end) was performed for 100 generations (with  $\sigma^{init} = 0.40$  Å and population sizes between 12 and 18, dependent on the symmetry). Non-physical structures were filtered out by

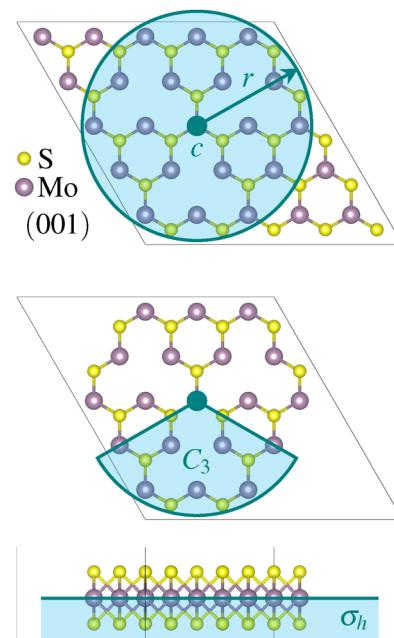


Fig. 1 Overview of the applied symmetry constraints. Top view of a monolayer MoS<sub>2</sub> founder structure with three sulfur double vacancies (3V<sub>S<sub>2</sub></sub>) inside a (5 × 5 × 1) supercell. The atoms modified during the evolution are placed inside a sphere defined by radius  $r$  around the center  $c$ . The atoms outside the sphere are fixed. This setup corresponds to the baseline none ( $C_1$ ). Further reduction of the number of degrees of freedom is achieved by applying the threefold ( $C_3$ ) rotational axis and the horizontal ( $\sigma_h$ ) mirror plane.

discarding all data points that contained at least one force component larger than a specific value (between 100 and 240 meV Å<sup>-1</sup>, dependent on the symmetry) and/or had an energy value larger than 0.0 eV. To decrease the redundancy of the data set, it was further reduced by selecting only five structures per generation (highest energy, lowest energy, and additional three randomly chosen) for further treatment. To clean the data set from structures that do not represent two-dimensional materials, all structures that contained at least one atom with a distance larger than two Mo–S bond lengths from the horizontal monolayer plane were also removed. In the end, the split consisted of 921 structures for the training set (50%), 180 structures for the validation set (10%), 373 structures for the active learning sample set (20%) and 373 structures for the test set (20%), in total 1847 structures.

### Machine-learning force fields

The 3V<sub>S<sub>2</sub></sub> data set (described above) was used for the training of the MACE<sup>43</sup> MLFF. The hidden layers were set to 64 channels and the cutoff radius to 4 Å. In the first phase of training (maximum 1200 epochs), the energy and force weights were set to 1 and 100, respectively. In the second phase (maximum 300 epochs), the weight for the energy was increased to 1000. A second MLFF, based on the NeuralIL<sup>44,45</sup> architecture, was used only for the generation of the 5V<sub>S<sub>2</sub></sub> test set (described below). The training fraction was set to 0.8, the cutoff radius

to 4 Å, the number of basis functions to 5, the energy weight to 0.5 and the number of epochs to 51. The ResNet core widths were chosen as 64, 32 and 16.

### 5V<sub>S<sub>2</sub></sub> test set generation

To generate the test set of MoS<sub>2</sub> structures with five sulfur double vacancies inside a (7 × 7 × 1) supercell (symbolized by 5V<sub>S<sub>2</sub></sub>), the following steps were performed: for each symmetry operation, the 100 structures with the highest uncertainties, see eqn (1), were selected from 10 CLINAMEN2 evolutions. These were executed with  $\sigma^{\text{init}} = 0.40$  Å, different random seeds, and the NeuralIL MLFF as evaluation back-end. The same procedure was repeated with additional 10 evolutions, but now with  $\sigma^{\text{init}} = 0.60$  Å. The removal of high force, high energy, and non-two-dimensional structures (similar to the 3V<sub>S<sub>2</sub></sub> data set) resulted in a test set consisting of 496 structures in total.

### Uncertainties

The force uncertainties were estimated by training two committees (MACE and NEURALIL) consisting of 5 models each. The only difference between the models of a committee is the random seed that determines the initial weights of the neural network. Every model in the committee produces their own predictions for the force components of the atoms. The individual force uncertainties were aggregated over all atoms in a given configuration to the global uncertainties

$$s_i = \frac{1}{3N_{\text{atom}}} \sum_j \sum_k^{xyz} s_{ij}^{(k)}, \quad (1)$$

with  $k$  denoting the spatial direction,  $j$  the atom, and  $i$  the configuration. These uncertainties have been shown to provide reliable estimates of the error, exhibiting a strong correlation

with the corresponding differences between the predicted and reference DFT values.<sup>45–47</sup>

### Active learning

For the active learning step, the structures from the 3V<sub>S<sub>2</sub></sub> active learning sample set were sorted by their uncertainty as defined in eqn (1). The 74 structures in the highest 20% interval were removed from the active learning sample set and instead added to the training and validation sets. The resulting data set represents iteration 1 (see Fig. S2 in the SI).

## Results

### Force field predictions

The data set for the MACE MLFF committee contained structures only from evolutions that started with the smallest founder, consisting of a single layer of MoS<sub>2</sub> with three double vacancies (3V<sub>S<sub>2</sub></sub>). Additionally, a test set including larger founders with five double vacancies (5V<sub>S<sub>2</sub></sub>) was generated. To measure the performance of the MLFF, parity plots for both 3V<sub>S<sub>2</sub></sub> and 5V<sub>S<sub>2</sub></sub> data sets were generated, see Fig. 2. The error values demonstrate the predictive power of the force field for the 3V<sub>S<sub>2</sub></sub> structures and the ability to generalize well to the larger 5V<sub>S<sub>2</sub></sub> structures, which were not part of the training set.

The correlation between the uncertainties and error values is visualized in a correlation plot, Fig. 3, where also linear fits of the data points are shown. The value of the slope  $\alpha$  of the linear fit for the 3V<sub>S<sub>2</sub></sub> structures is smaller than for the 5V<sub>S<sub>2</sub></sub> test set. Since higher  $\alpha$  values are generally related to higher model bias,<sup>47</sup> it seems plausible to receive a higher  $\alpha$  value when the force field has to extrapolate the prediction to larger, previously unseen structures. Nevertheless, the difference is small, and the uncertainties can, therefore, be used as a reliable measure

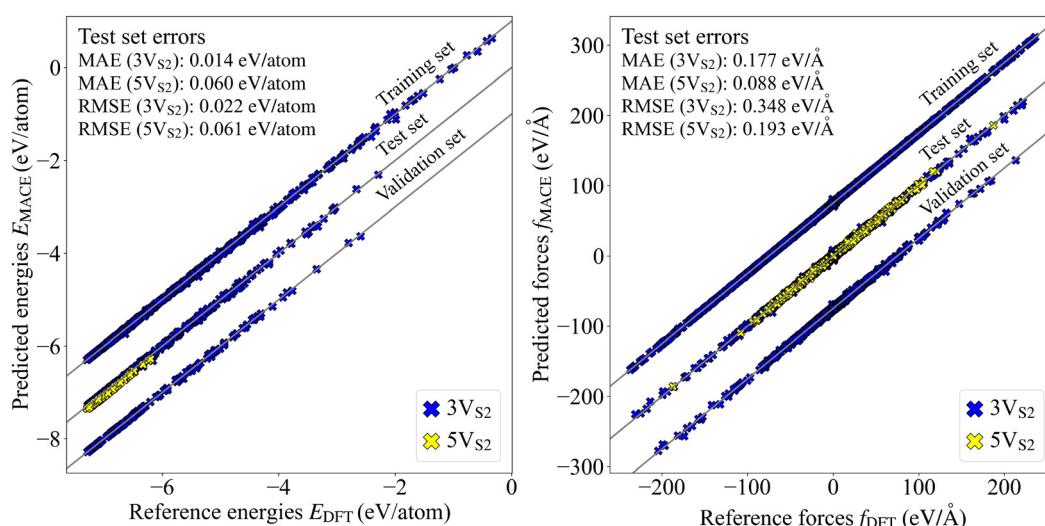


Fig. 2 Parity plots for the predicted and reference energies and forces. The plots compare the predicted energies per atom  $E_{\text{MACE}}$  and forces  $f_{\text{MACE}}$  from the MLFF with reference energies per atom  $E_{\text{DFT}}$  and forces  $f_{\text{DFT}}$  calculated by DFT. Shown are all three  $3V_{S_2}$  data set splits, where the training and validation set lines are shifted up and down relative to the test set line, respectively. The  $5V_{S_2}$  data set is only used as a test set. Also given are the test set errors (MAE and RMSE) for both types of founder structures.

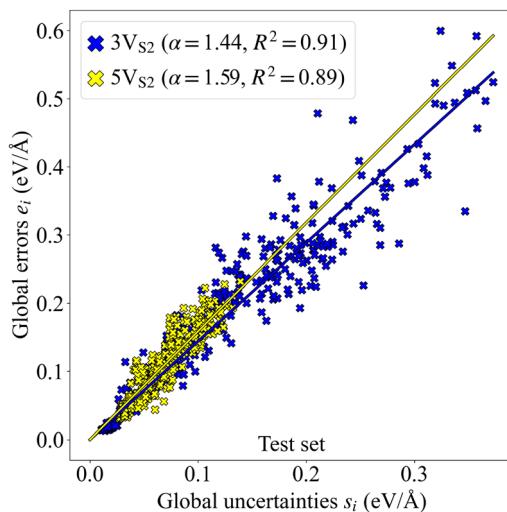


Fig. 3 Alpha correlation plot. The global errors  $e_i$  (the difference between the predicted and the reference forces) are plotted over the global uncertainties  $s_i$  resulting from the force predictions of the MLFF committee. Linear fits for both  $3V_{S_2}$  and  $5V_{S_2}$  data sets are visualized by colored lines. Additional characteristics of the linear fits such as the slope  $\alpha$  and the coefficient of determination  $R^2$  are given in the legend.

for both founder types to describe the performance of the MLFF during the executed evolutions.

The global uncertainties classify all obtained structures into two categories. The first contains configurations which are familiar to the force field (low uncertainty) and the second contains configurations where the force field is unsure if the predictions are accurate (high uncertainty). This classification more or less decides which structures contain the most information that could be useful for the improvement of the MLFF. Keeping this principle in mind, active learning steps can be performed, where the structures with the highest uncertainties are selected and added to the training set for a new force field. In our case, the top 20%  $3V_{S_2}$  structures exhibiting the largest uncertainty were removed from the active learning sample set and instead added to the training and validation sets. Then a new MLFF was trained on these augmented data sets. In the following, the original force field before the active learning step will be called iteration 0, while the new force field after the active learning step will be called iteration 1. Parity and alpha correlation plots were generated to compare the predictive power between these two MLFFs (see Fig. S2 in the SI). The test set errors contained in these plots are summarized in Table 1. The results indicate that the active learning step does not show a large improvement regarding the predictive power. The energy and force errors for both the  $3V_{S_2}$  and  $5V_{S_2}$  structures only changed minutely. We therefore settled on the MLFF trained purely on the  $3V_{S_2}$  structures obtained from the original CMA-ES evolutions.

### Constrained trajectories

Three-fold rotational defects can be ordered into a hierarchy, where the geometric properties of each level, labeled as  $T_n$  with

Table 1 Test set errors for one active learning step. Test set errors for the MLFF before (iteration 0) and after (iteration 1) one active learning step. Given are the MAE and RMSE for the predicted energies per atom  $E_{\text{MACE}}$  and the predicted forces  $f_{\text{MACE}}$  for both  $3V_{S_2}$  and  $5V_{S_2}$  data sets

$E_{\text{MACE}}$ (eV per atom)	$3V_{S_2}$		$5V_{S_2}$	
	MAE	RMSE	MAE	RMSE
Iteration 0	0.014	0.022	0.060	0.061
Iteration 1	0.013	0.021	0.057	0.058
$f_{\text{MACE}}$ (eV Å <sup>-1</sup> )	MAE	RMSE	MAE	RMSE
Iteration 0	0.177	0.348	0.088	0.193
Iteration 1	0.175	0.342	0.087	0.189

$n \in \{1, 2, 3, \dots\}$ , can be described by structural parameters. One example would be the number of octagons and double-pentagons which are part of the defect.<sup>22</sup> The requirement for the formation of these defects is a specific number of sulfur double vacancies in the vicinity of potential bond rotation centers. The smallest defect, called  $T_1$ , requires three double vacancies ( $3V_{S_2}$ ) around one rotation center (see Fig. 4a). The vacancies in the founder were placed symmetrically around the center, since the search algorithm had to start the evolutions with the same symmetry constraints as the desired final structure. This was also the reason why one of the sulfur pairs of the founder of the defect level  $T_2$ , which requires five double vacancies ( $5V_{S_2}$ ), was shifted to the intersection between the rotational axis and the horizontal monolayer plane (see Fig. 4b).

To demonstrate the influence of the applied constraints, we ran 20 separate CLINAMEN2 evolutions with different random

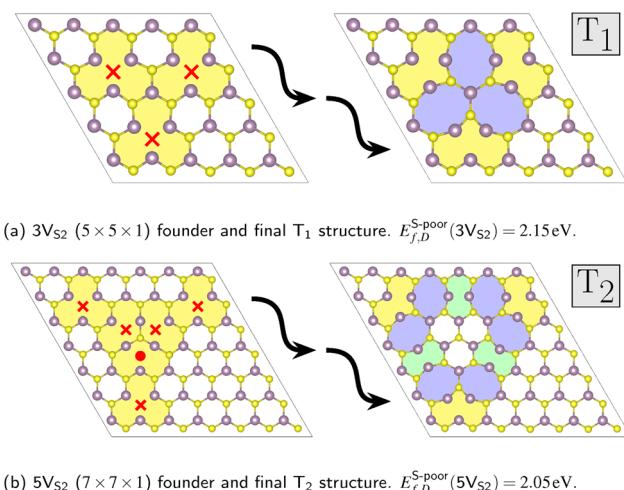
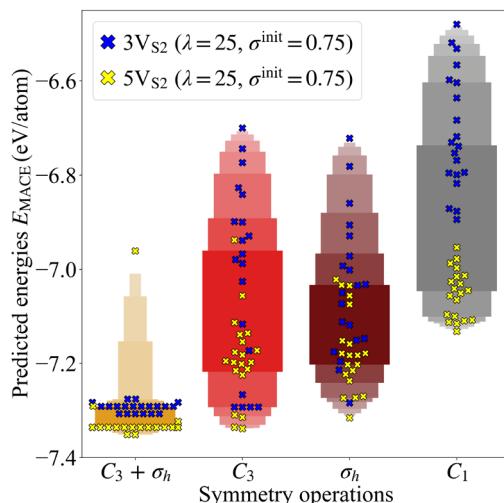


Fig. 4 Founders with double vacancies and final structures with rotational defects. The founders are placed in differently sized super cells ( $D \times D \times 1$ ): (a)  $D = 5$  and (b)  $D = 7$ . Both contain sulfur double vacancies  $XV_{S_2}$  with  $X \in \{3, 5\}$ , respectively. The double vacancies are symbolized by red crosses. Notice that in the case of  $5V_{S_2}$ , one pair of sulfurs is not removed (shown as a red dot), but shifted to the center to preserve the rotational symmetry. The final structures, produced by  $C_3 + \sigma_h$  constrained evolutions, are colored to highlight the octagons (blue), double-pentagons (green), and ten-fold rings (yellow) of the defects. Also given are the defect formation energies  $E_{f,D} = X^{-1}(E_D - E_{\text{bulk}} - n_z E_z)$ , normalized by the number of double vacancies  $X$ .



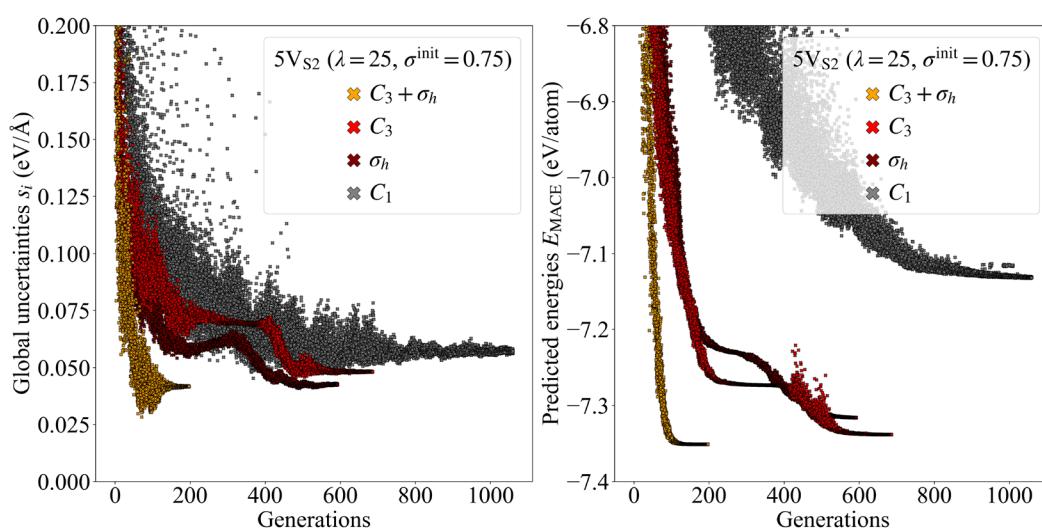


**Fig. 5** Predicted energy per atom distributions for all symmetry operations. Shown are the predicted energies per atom  $E_{\text{MACE}}$  for the final structures from in total 160 evolutions. Both  $3V_{S_2}$  and  $5V_{S_2}$  founders were included. The population size was selected as  $\lambda = 25$  and the initial step size set to  $\sigma^{\text{init}} = 0.75 \text{ \AA}$ . The utilized constraints are ordered from the highest symmetry ( $C_3 + \sigma_h$ ) on the left to the lowest symmetry ( $C_1$ ) on the right. The shading in the background indicates the accumulation area around the mean value of the distributions.

seeds and the MACE MLFF as evaluation back-end for each of the four symmetry operations. This was done for both the smaller  $3V_{S_2}$  and the larger  $5V_{S_2}$  founders, thus yielding 160 CMA-ES evolutions in total. The population size was selected as  $\lambda = 25$ , while the initial step size was chosen as  $\sigma^{\text{init}} = 0.75 \text{ \AA}$ . The advantages of these constrained evolutions can be shown by analyzing the energy per atom distributions of the final structures of the trajectories. For every evolution, the predicted energy per atom of the most stable structure in the population

of the last generation was plotted, see Fig. 5. The distribution of both founder types shows that evolutions with symmetry constraints lead to lower average values of the energies per atom of the final structures. Out of all  $3V_{S_2}$  evolutions (20 runs) with the  $C_3 + \sigma_h$  constraint, the search found the smallest of the rotational defects,  $T_1$  in Fig. 4a, six times. Out of all  $5V_{S_2}$  evolutions run with the same constraint (also 20 runs), the larger  $T_2$  defect was found two times (see Fig. 4b). Both these structures are in excellent agreement with the combined DFT and experimental study in ref. 22, underlining the significant promise of the transferable MLFF. Another important benefit shown by Fig. 5 is the fact that evolutions augmented with symmetry constraints permit both lower population sizes  $\lambda$  and higher initial step sizes  $\sigma^{\text{init}}$ , while still producing stable final structures at the same time. Low population sizes lead to faster execution times, while high initial step sizes increase the possibility of the evolution mean to leave the minimum of the founder, opening up the opportunity to find other stable minima. Additional energy per atom distributions for other initial step and population sizes can be found in Fig. S3 in the SI.

Another way to visualize the usefulness of our symmetry constraints is to plot the calculated uncertainties of all structures as a function of the number of generations, see Fig. 6. The uncertainty trajectories were extracted from the four  $5V_{S_2}$  evolutions that led to the final structures with the lowest energies out of all evolutions. The plot shows that constrained trajectories, in general, exhibit global uncertainties with a lower mean and a smaller standard deviation per generation than completely free trajectories. In addition, on average, the termination criteria are satisfied earlier, which leads to faster convergence. Furthermore, the final structures produced by these high-symmetry evolutions possess lower predicted energy per atom distributions, leading to more stable configurations.



**Fig. 6** Uncertainties and predicted energies per atom plotted over the number of generations. The left figure shows the global uncertainties  $s_i$  of all structures produced by selected evolutions as a function of the generations. For every symmetry operation, one evolution is plotted, coded by color. All evolutions started with the  $5V_{S_2}$  founder. The population size was selected as  $\lambda = 25$  and the initial step size set to  $\sigma^{\text{init}} = 0.75 \text{ \AA}$ . The right figure plots the predicted energies per atom  $E_{\text{MACE}}$  of the same evolutions over the generations.

To substantiate the claim of a symmetry-related advantage, we performed a series of DFT calculations for selected final structures shown in Fig. 5. The deviations in  $\Delta E$  between the MACE model and DFT are provided in Table S1 in the SI and fall within the range expected from the model's MAE and RMSE (Fig. 2). Moreover, the DFT energy differences between the final structures corroborate the stability ordering predicted by the force field.

The dynamic change of the uncertainty values during the evolutions offers insight into different phases of the search process. For example, the trajectory for the evolution with the  $C_3$  constraint (see Fig. 6) experiences an interesting sequence of character shifts. Approaching generation 300, the spread of the uncertainties is already very slim and decreases even further, which normally indicates the end of an evolution. However, around generation 400, the standard deviation increases sharply, suggesting that the CMA-ES found a new path for further exploration. The same characteristics can be seen for the energies. The evolution is following this track until around generation 500, during which both the uncertainties and the energies are decreasing. Finally, around generation 600, the search converges to the stable configuration, while at the same time scaling back once again to small standard deviations for both quantities.

### Structure search

The MLFF was not only able to reproduce the two known members of the hierarchy, Fig. 4, but also capable to find additional rotational defects. These require an amount of double vacancies between the numbers already presented, specifically  $4V_{S_2}$  and  $6V_{S_2}$ , see Fig. 7a and b. These defects contain the same structural 8-5-5-8 ring building blocks as the known structures in Fig. 4. They also exhibit corners consisting of one octagon and one pentagon each

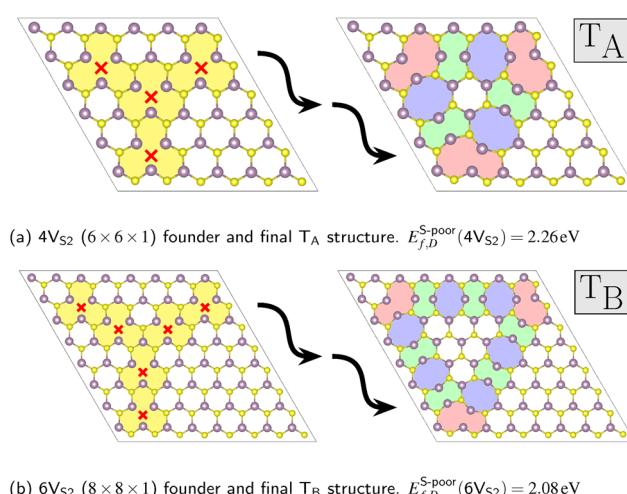


Fig. 7 Founders and final structures with new rotational defects from the structure search. The founders are placed in differently sized super cells ( $D \times D \times 1$ ): (a)  $D = 6$  and (b)  $D = 8$ . Both contain sulfur double vacancies  $XV_{S_2}$  with  $X \in \{4, 6\}$ , respectively. The coloring scheme follows Fig. 4. The corners of the  $T_A$  and  $T_B$  structures colored in red consist of one octagon and one pentagon each.

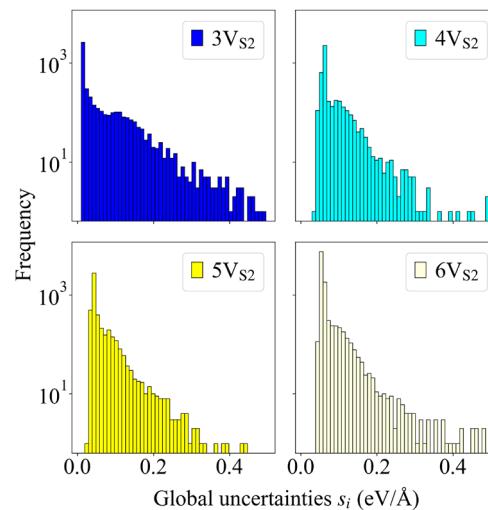


Fig. 8 Histograms of global uncertainties. Shown are the frequency distributions of the global uncertainties  $s_i$  for all four evolutions that generated the final structures illustrated in Fig. 4 and 7, visualized as histograms. Uncertainties larger than  $s_i = 0.5 \text{ eV } \text{\AA}^{-1}$  are not shown, since these values are very low in number and arise only at the very beginning of the evolutions.

(red colored areas in Fig. 7a and b). These  $T_A$  and  $T_B$  defect structures fill the gaps in the defect hierarchy, completing the series ranging from  $3V_{S_2}$  to  $6V_{S_2}$  founders.

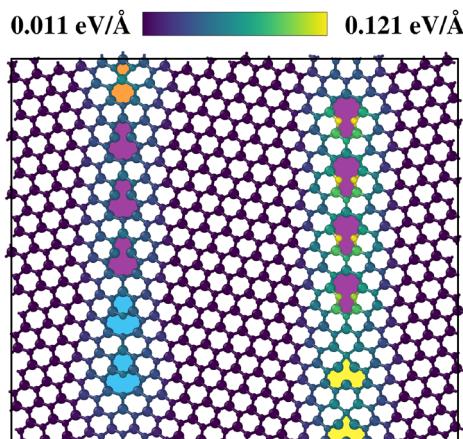
Although the force field was trained on structures that only contained small  $T_1$  defects, it was able to generalize well to predictions of larger  $T_2$  structures and to extrapolate from the existing data to  $T_A$  and  $T_B$  defects. Fig. 8 indicates that the global uncertainties are still a reliable measure of performance when utilized for the new  $4V_{S_2}$  and  $6V_{S_2}$  founders, since the visual comparison shows frequency distributions similar to those started with the  $3V_{S_2}$  and  $5V_{S_2}$  structures.

As a final validation of the search procedure, we initiated new evolutionary runs using the high-symmetry structures as founders, while enforcing lower symmetry constraints. The low-energy threefold-rotational defects shown in Fig. 4 and 7 were used as starting configurations for evolutions constrained by  $C_3$ ,  $\sigma_h$ , and  $C_1$  symmetries. In all cases, the evolutions converged to the same defect structures, thereby reinforcing the conclusion that the rotational defects correspond to genuine local minima.

### Grain boundaries

The MLFF was trained on only 1101 structures from the original DFT CMA-ES trajectories of small  $3V_{S_2}$  defects, but showed the ability to generalize to structures not included in training. To further demonstrate this feature, it was applied to MoS<sub>2</sub> monolayer grain boundaries. These geometries contain similar building blocks as the rotational defects discussed above.<sup>14</sup> Specifically, we used the MACE committee to calculate the local uncertainties

$$s_{ij}^{\text{local}} = \frac{1}{3N_{\text{neigh}}} \sum_n \sum_k^{xyz} s_{in}^{(k)} \quad (2)$$



**Fig. 9** Local uncertainties of grain boundary ring motifs. The color bar represents the local uncertainties  $s_{ij}^{\text{local}}$  for every atom, calculated by the MACE committee for a MoS<sub>2</sub> monolayer grain boundary tilted by 18.5°. The cutoff radius for the uncertainty calculation was chosen as  $r_{\text{cut}} = 4.0$  Å. Motifs such as 4|6 (orange), 5|7 (violet) and 6|8 (cyan) rings are highlighted by coloring. Ten-fold rings also present in the corners of the T<sub>1</sub> and T<sub>2</sub> defects are indicated by yellow.

for every atom in the vicinity of a grain boundary with a tilt angle of 18.5°.  $N_{\text{neigh}}$  denotes here the neighbouring atoms of atom  $j$ , located within a cutoff radius  $r_{\text{cut}}$ . These locally aggregated uncertainties have been shown to correlate with the local errors in a similar way as the globally aggregated uncertainties in Fig. 3.<sup>47</sup>

Fig. 9 illustrates the calculated  $s_{ij}^{\text{local}}$ . Considering first the Mo-rich interface on the left side of the structure, it is seen that motifs already present in the threefold rotational defect structures, such as Mo-rich five- and seven-fold rings (5|7), display local uncertainty values almost as low as the pristine environment. More surprisingly, unknown motifs, such as Mo-rich four- and six-fold (4|6) and six- and eight-fold (6|8) rings, do not increase the resulting values in a significant way. To assess the transferability of the force field to grain boundaries, we computed DFT reference forces for the grain boundary structure shown in Fig. 9. An  $\alpha$ -correlation plot illustrating the relationship between local uncertainties and force errors for this structure is provided in Fig. S6 in the SI. As in previous cases, the local uncertainties are found to be predictive of the corresponding local force errors. This underlines that a CMA-ES run visits a diverse set of structures.<sup>48,49</sup>

The other S-rich interface on the right consists of the same motifs, but here the Mo and S atoms are switched in place (sometimes called opposite polarity). For this reason, two sulfur atoms of the pentagon structure in the S-rich 5|7 rings are placed closer to each other than in the structures known by the force field. This circumstance leads to the increased local uncertainties that are displayed for these specific sulfur atoms.

## Conclusions

We investigated the potential energy landscape of threefold rotational defects in monolayer MoS<sub>2</sub> by coupling a

machine-learning force field with an evolutionary structure search algorithm. To enhance search efficiency, the algorithm systematically constrained the exploration to a reduced-dimensional space defined by the symmetry operations of the respective defects. We found that these symmetry-guided trajectories maintained lower global uncertainty throughout the search process, led to energetically more favorable structures, and achieved faster convergence. This strategy allowed us to identify several previously unreported defect configurations at a moderate computational cost, thus completing the catalog of rotational defects in MoS<sub>2</sub>.

While machine-learning force fields can already efficiently and transferable parametrize potential energy surfaces, the structure search itself still relies on heuristic algorithms that require extensive manual tuning and struggle to escape local minima. Replacing these heuristics with data-driven approaches capable of learning adaptive search strategies from large datasets presents a promising direction for future work.

## Author contributions

Alexander Adel: conceptualization, investigation, methodology, software, validation, visualization, writing – original draft. Ralf Wanzenböck: conceptualization, methodology, software, supervision, writing – review & editing. Georg K. H. Madsen: conceptualization, methodology, project administration, supervision, writing – review & editing.

## Conflicts of interest

There are no conflicts to declare.

## Data availability

Additional figures including parity, correlation, distribution and uncertainty plots have been included as part of the supplementary information (SI). Supplementary information is available. See DOI: <https://doi.org/10.1039/d5cp03121d>.

The trained models and data set splits, together with structure files containing the founders and results presented in this paper, are available on Zenodo at <https://doi.org/10.5281/zenodo.16812610>. The latest version of CLINAMEN2, along with its documentation and example evolution scripts, is available at <https://github.com/Madsen-s-research-group/Clinamen2-public-releases>.

## Acknowledgements

This research was funded in part by the Austrian Science Fund (FWF) through the doctoral college TU-DX <https://doi.org/10.55776/DOC142>, and in part through <https://doi.org/10.55776/F8100>. For open access purposes, the authors have applied a CC BY public copyright license to any author accepted manuscript version arising from this submission.



## References

- 1 R. F. Frindt and A. D. Yoffe, *Proc. R. Soc. A*, 1963, **273**, 69–83.
- 2 R. F. Frindt, *J. Appl. Phys.*, 1966, **37**, 1928–1929.
- 3 P. Joensen, R. F. Frindt and S. R. Morrison, *Mater. Res. Bull.*, 1986, **21**, 457–461.
- 4 K. S. Novoselov, D. Jiang, F. Schedin, T. J. Booth, V. V. Khotkevich, S. V. Morozov and A. K. Geim, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 10451–10453.
- 5 K. F. Mak, C. Lee, J. Hone, J. Shan and T. F. Heinz, *Phys. Rev. Lett.*, 2010, **105**, 136805.
- 6 A. Kuc, N. Zibouche and T. Heine, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2011, **83**, 245213.
- 7 P. Miró, M. Audiffred and T. Heine, *Chem. Soc. Rev.*, 2014, **43**, 6537–6554.
- 8 H.-P. Komsa and A. V. Krasheninnikov, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2015, **91**, 125304.
- 9 J. Hong, Z. Hu, M. Probert, K. Li, D. Lv, X. Yang, L. Gu, N. Mao, Q. Feng, L. Xie, J. Zhang, D. Wu, Z. Zhang, C. Jin, W. Ji, X. Zhang, J. Yuan and Z. Zhang, *Nat. Commun.*, 2015, **6**, 6293.
- 10 J.-Y. Noh, H. Kim and Y.-S. Kim, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2014, **89**, 205417.
- 11 H.-P. Komsa, S. Kurasch, O. Lehtinen, U. Kaiser and A. V. Krasheninnikov, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2013, **88**, 035301.
- 12 A. N. Enyashin, M. Bar-Sadan, L. Houben and G. Seifert, *J. Phys. Chem. C*, 2013, **117**, 10842–10848.
- 13 S. Wang, G.-D. Lee, S. Lee, E. Yoon and J. H. Warner, *ACS Nano*, 2016, **10**, 5419–5430.
- 14 W. Zhou, X. Zou, S. Najmaei, Z. Liu, Y. Shi, J. Kong, J. Lou, P. M. Ajayan, B. I. Yakobson and J.-C. Idrobo, *Nano Lett.*, 2013, **13**, 2615–2622.
- 15 A. M. V. D. Zande, P. Y. Huang, D. A. Chenet, T. C. Berkelbach, Y. You, G.-H. Lee, T. F. Heinz, D. R. Reichman, D. A. Muller and J. C. Hone, *Nat. Mater.*, 2013, **12**, 554–561.
- 16 Z. Lin, B. R. Carvalho, E. Kahn, R. Lv, R. Rao, H. Terrones, M. A. Pimenta and M. Terrones, *2D Mater.*, 2016, **3**, 022002.
- 17 Z. Wu and Z. Ni, *Nanophotonics*, 2017, **6**, 1219–1237.
- 18 Z. Hu, Z. Wu, C. Han, J. He, Z. Ni and W. Chen, *Chem. Soc. Rev.*, 2018, **47**, 3100–3128.
- 19 Q. Liang, Q. Zhang, X. Zhao, M. Liu and A. T. S. Wee, *ACS Nano*, 2021, **15**, 2165–2181.
- 20 S. Zeng, F. Li, C. Tan, L. Yang and Z. Wang, *Front. Phys.*, 2023, **18**, 53604.
- 21 S. Mahendran, J. Carrete, A. Isacsson, G. K. H. Madsen and P. Erhart, *J. Phys. Chem. C*, 2024, **128**, 1709–1716.
- 22 Y.-C. Lin, T. Björkman, H.-P. Komsa, P.-Y. Teng, C.-H. Yeh, F.-S. Huang, K.-H. Lin, J. Jadczak, Y.-S. Huang, P.-W. Chiu, A. V. Krasheninnikov and K. Suenaga, *Nat. Commun.*, 2015, **6**, 6736.
- 23 A. J. Stone and D. J. Wales, *Chem. Phys. Lett.*, 1986, **128**, 501–503.
- 24 A. Cresti, J. Carrete, H. Okuno, T. Wang, G. K. H. Madsen, N. Mingo and P. Pochet, *Carbon*, 2020, **161**, 259–268.
- 25 D. J. Wales, *Annu. Rev. Phys. Chem.*, 2018, **69**, 401–425.
- 26 H.-G. Beyer and H.-P. Schwefel, *Nat. Comput.*, 2002, **1**, 3–52.
- 27 T. Bartz-Beielstein, J. Branke, J. Mehnen and O. Mersmann, *WIREs Data Mining Knowl. Discov.*, 2014, **4**, 178–195.
- 28 N. Hansen, *arXiv*, 2023, preprint, arXiv:1604.00772v2 [cs.LG], DOI: [10.48550/arXiv.1604.00772](https://doi.org/10.48550/arXiv.1604.00772).
- 29 H. Huber, M. Sommer-Jørgensen, M. Gubler and S. Goedecker, *Phys. Rev. Res.*, 2023, **5**, 013189.
- 30 A. O. Lyakhov, A. R. Oganov and M. Valle, *Comput. Phys. Commun.*, 2010, **181**, 1623–1632.
- 31 C. J. Pickard and R. J. Needs, *J. Phys.: Condens. Matter*, 2011, **23**, 053201.
- 32 X. Shao, J. Lv, P. Liu, S. Shao, P. Gao, H. Liu, Y. Wang and Y. Ma, *J. Chem. Phys.*, 2022, **156**, 014105.
- 33 Y. Zhao, E. M. D. Siriwardane, Z. Wu, N. Fu, M. Al-Fahdi, M. Hu and J. Hu, *npj Comput. Mater.*, 2023, **9**, 38.
- 34 F. Brix, M.-P. V. Christiansen and B. Hammer, *J. Chem. Phys.*, 2024, **160**, 174107.
- 35 M. Arrigoni and G. K. H. Madsen, *npj Comput. Mater.*, 2021, **7**, 71.
- 36 I. Mosquera-Lois, S. R. Kavanagh, A. M. Ganose and A. Walsh, *npj Comput. Mater.*, 2024, **10**, 121.
- 37 Z. Yang, X. Liu, X. Zhang, P. Huang, K. S. Novoselov and L. Shen, *npj Comput. Mater.*, 2025, **11**, 229.
- 38 G. Kresse and J. Furthmüller, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1996, **54**, 11169–11186.
- 39 G. Kresse and D. Joubert, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1999, **59**, 1758–1775.
- 40 P. E. Blöchl, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1994, **50**, 17953–17979.
- 41 J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865–3868.
- 42 R. Wanzenböck, F. Buchner, P. Kovács, G. K. H. Madsen and J. Carrete, *Comput. Phys. Commun.*, 2024, **297**, 109065.
- 43 I. Batatia, D. P. Kovács, G. N. C. Simm, C. Ortner and G. Csányi, *arXiv*, 2023, preprint, arXiv:2206.07697v2 [stat.ML], DOI: [10.48550/arXiv.2206.07697](https://doi.org/10.48550/arXiv.2206.07697).
- 44 H. Montes-Campos, J. Carrete, S. Bichelmaier, L. M. Varela and G. K. H. Madsen, *J. Chem. Inf. Model.*, 2022, **62**, 88–101.
- 45 J. Carrete, H. Montes-Campos, R. Wanzenböck, E. Heid and G. K. H. Madsen, *J. Chem. Phys.*, 2023, **158**, 204801.
- 46 D. Schwalbe-Koda, A. R. Tan and R. Gómez-Bombarelli, *Nat. Commun.*, 2021, **12**, 5104.
- 47 E. Heid, J. Schörghuber, R. Wanzenböck and G. K. H. Madsen, *J. Chem. Inf. Model.*, 2024, **64**, 6377–6387.
- 48 R. Wanzenböck, M. Arrigoni, S. Bichelmaier, F. Buchner, J. Carrete and G. K. H. Madsen, *Digital Discovery*, 2022, **1**, 703–710.
- 49 R. Wanzenböck, E. Heid, M. Riva, G. Franceschi, A. M. Imre, J. Carrete, U. Diebold and G. K. H. Madsen, *Digital Discovery*, 2024, **3**, 2137–2145.

