npj | computational materials

Article

Published in partnership with the Shanghai Institute of Ceramics of the Chinese Academy of Sciences

https://doi.org/10.1038/s41524-025-01754-8

Substrate-aware computational design of two-dimensional materials

Arslan Mazitov[1] ✉, **Ivan Kruglov**[1,2], **Alexey V. Yanilkin**[1], **Aleksey V. Arsenin**[1,2], **Valentyn S. Volkov**[2], **Dmitry G. Kvashnin**[3], **Artem R. Oganov**[4] & **Kostya S. Novoselov**[5,6,7]

Two-dimensional (2D) materials attract considerable attention due to their remarkable electronic, mechanical and optical properties. Despite their use in combination with substrates in practical applications, computational studies often neglect the effects of substrate interactions for simplicity. This study presents a novel method for predicting the atomic structure of 2D materials on substrates by combining an evolutionary algorithm, a lattice-matching technique, an automated machine-learning interatomic potentials training protocol, and the ab initio thermodynamics approach. Using the molybdenum-sulfur system on a sapphire substrate as a case study, we reveal several new stable and metastable structures, including previously known 1H-MoS_2 and newly found *Pmma* Mo_3S_2, $P\bar{1}$ Mo_2S, $P2_1m$ Mo_5S_3, and *P4mm* Mo_4S, where the Mo_4S structure is specifically stabilized by interaction with the substrate. Finally, we use the ab initio thermodynamics approach to predict the synthesis conditions of the discovered structures in the parameter space of the commonly used chemical vapor deposition technique.

The discovery of graphene — a single layer of carbon atoms arranged in a honeycomb lattice — in 2004 marked a milestone in the field of 2D materials and sparked a wave of research into this novel class of materials[1]. Since then, the family of 2D materials has expanded to include transition metal dichalcogenides (TMDs), hexagonal boron nitride (h-BN), phosphorene, and other layered materials, each with their own unique properties and potential applications[2–4]. Moreover, this family is only getting larger[5], and many new 2D materials, such as $Ru_2Si_xO_y$, $Cr_2B_2F_2$ and $Ni_2Si_2O_2$ have recently been discovered by large-scale computational studies, while the stability of $Ru_2Si_xO_y$ has been confirmed by further experimental synthesis[6]. At the time of writing, the Computational 2D Materials Database[7,8] already contains more than 16,000 entries.

The unique properties of 2D materials, such as high carrier mobility, exceptional mechanical flexibility, and tunable bandgaps, have spurred interest in exploring their use in a wide range of electronic and optoelectronic devices[2,9–13]. For example, TMDs have emerged as promising candidates for field-effect transistors, photodetectors, and light-emitting diodes due to their large bandgaps and strong light-matter interaction[2]. Similarly, the atomic thickness and excellent mechanical properties of graphene make it an ideal candidate for flexible electronics, transparent conductive electrodes, and sensors[14]. Moreover, a wide range of ways to tune properties of 2D layers using lateral and vertical heterostructures fabrication[15–17], chemical functionalization[10,18], strain[19–21], defect[22,23] and substrate engineering[11], makes them ideal candidates for developing a new class of electronic devices. According to the International Roadmap for Devices and Systems[24], the use of 2D materials and their heterostructures in the fabrication of a new generation of transistors can improve the technological process from ~ 5 nm to 1 nm by 2031. Moreover, many promising applications in the fields of photonics[25], photovoltaics[18,26], valleytronics[27], energetics[28], and catalysis[29] have already been realized in practice.

2D materials provide a versatile platform to study a large range of physical phenomena. However, even more control can be achieved through the manipulation of the interaction with the substrate. Indeed, the fabrication of 2D materials typically relies on physical (PVD) and chemical (CVD) vapor deposition techniques, molecular-beam (MBE) and atomic-layer (ALE) epitaxy, or direct mechanical exfoliation method[13,30], where the material is either directly grown or finally placed on top of a substrate. Substrate engineering not only offers higher-quality devices, as in the case of graphene on hBN[31,32], but also allows to modify the stability, electronic structure, and mechanical properties of the 2D material-substrate system.

---

[1]Moscow Center for Advanced Studies, Kulakova str. 20, Moscow, 123592, Russian Federation. [2]Emerging Technologies Research Center, XPANCEO, Internet City, Emmay Tower, Dubai, United Arab Emirates. [3]Emanuel Institute of Biochemical Physics, Kosigina st. 4, Moscow, 119334, Russian Federation. [4]Materials Discovery Laboratory, Skolkovo Institute of Science and Technology (Skoltech), Bolshoy Boulevard 30, bld. 1, Moscow, 121205, Russian Federation. [5]National Graphene Institute (NGI), University of Manchester, Manchester, M13 9PL, UK. [6]Department of Materials Science and Engineering, National University of Singapore, Singapore, 03-09 EA, Singapore. [7]Institute for Functional Intelligent Materials, National University of Singapore, Singapore, 117544, Singapore.
✉e-mail: arslan.mazitov@phystech.edu

For example, the choice of substrate can induce strain in the 2D material, thereby modulating its electronic properties and band structure[17,23]. The interaction with the substrate can also lead to a significant reduction in graphene's thermal conductivity on α-SiO$_2$ compared to the freestanding graphene due to enhanced phonon-phonon scattering rate[11]. In addition, the substrate can influence charge transfer processes[33], affect device performance, induce semi-metal - metal transition[34], or even lead to the emergence of novel phenomena, such as moiré patterns in heterostructures[35].

In computational 2D materials design, substrate effects, however, are often overlooked for the sake of simplicity. Most of the studies based on high-throughput screening of 2D materials with machine learning and/or ab initio techniques typically consider isolated 2D materials without taking into account the presence of substrates or other surrounding layers[36]. This includes data-driven high-throughput ab initio screening of 2D materials for properties optimization[8,37–40], evolutionary and global optimization methods for 2D materials discovery[40–50], and machine learning based prediction of new 2D materials[51–54]. In the case of crystal structure prediction (CSP), predicting the structure of two-dimensional materials in a wide range of stochiometries and in the presence of the substrate at the ab initio accuracy was desired, yet previously inaccessible for a few reasons. First, the periodic representation of the 2D layer and the substrate requires their unit cells to be as close as possible to each other to avoid the unphysical mistfit stress[55]. Since primitive cells generally do not meet this requirement, a suitable supercell construction is necessary to minimize the lattice mismatch. This inevitably results in structures with a large number of atoms, which do not allow one to use conventional density functional theory (DFT) calculations in a high-throughput manner.

This paper introduces a method for substrate-aware computational design of 2D materials. It is based on a combination of the evolutionary algorithm (EA) USPEX[56–58], a lattice-matching technique, and a machine learning interatomic potentials (MLIP)-based relaxation protocol. This combination is used to predict the possible stable 2D crystals in the presence of a substrate, while the ab initio thermodynamics approach is employed as a post-processing tool to study their possible synthesis conditions. Our method allows one to automatically explore the entire space of 2D atomic configurations and their compositions for a given chemical space, while estimating their relative stability in the presence of the substrate, and linking the calculated stability patterns to a realistic set of parameters controlled by a specific experimental setup. As the training of the MLIP for CSP purposes represents a crucial and non-trivial part of the computational pipeline, we also introduce an automatic workflow for creating a robust interatomic potential suitable for CSP. We demonstrate the reliability of our technique by predicting the stable two-dimensional crystals in the molybdenum-sulfur system on a *c*-cut sapphire substrate. First, we create the interatomic potential for the Mo–S@ Al$_2$O$_3$ system, which is suitable for both predicting the structure of the freestanding two-dimensional molybdenum-sulfur layers and those joined with the substrate. Next, we perform the evolutionary search for stable two-dimensional crystals using the trained MLIP for local relaxation and stability evaluation of the Mo–S structures. Finally, we predict the synthesis conditions of all the stable structures in the parameter space of the commonly used CVD technique.

Results and discussion

Automatic self-consistent training of MLIPs

In order to train the interatomic potential used in our work, we developed the automatic self-consistent training (ASCT) algorithm based on iterative sampling of the new structures from the molecular dynamics (MD) trajectories followed by ab initio calculations of their energies, interatomic forces, and stresses. MLIPs offer several advantages, such as a reasonable accuracy in predicting energies and forces, and a significantly lower computational cost, which allows one to consider large structures with low values of lattice mismatch during the evolutionary search. Given our focus on CSP, the training set of the ML potential typically requires covering a large configuration space, handling the random 2D crystal structures generated within the evolutionary algorithm and performing their structural optimization and/or MD annealing at a constant pressure and temperature. This includes both the freestanding 2D structures, those joined with the substrate, and the freestanding surface slab of the substrate due to the relaxation technique used in our work (see Methods for details).

A schematic representation of ASCT is given in Figure 1. At each ASCT iteration, a set of seed atomic configurations is either generated by the random structure generator of the USPEX code or selected from a user-provided list of structures. These configurations are then used to initialize the parallel MD sampling of the new configurations within the LAMMPS package[59,60] using the ML potential from the previous training iteration. During the MD run, each structure is compared to the existing training set of the ML potential, and is saved if it gets beyond the corresponding region of the configuration space. In order to avoid sampling unphysical configurations while the potential is not fully trained, the sampling procedure is

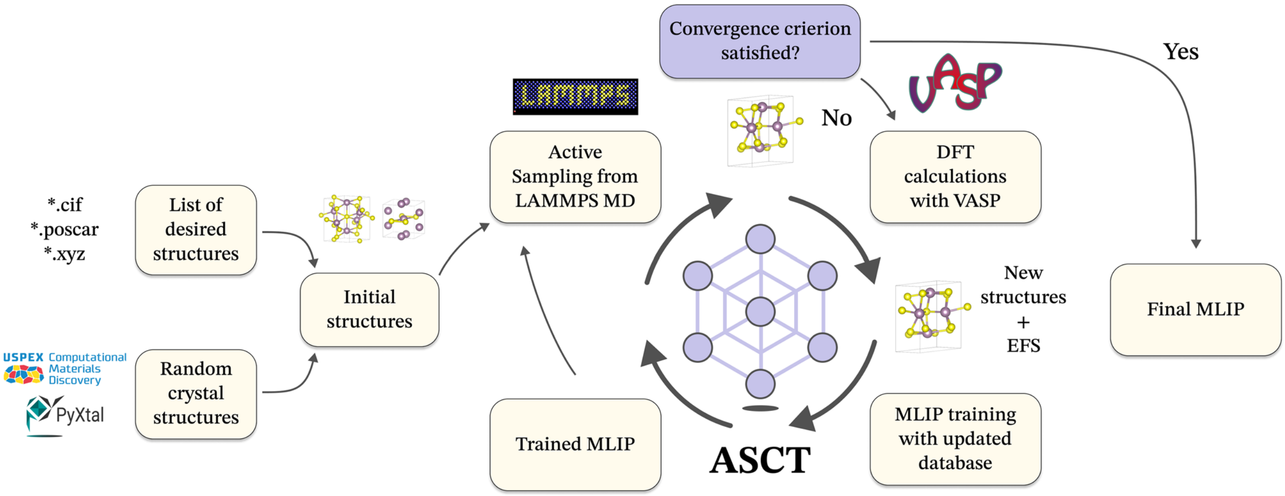

Fig. 1 | Automatic self-consistent training workflow (ASCT). At each iteration of the ASCT, a set of candidate structures is selected from the MD trajectory initialized with either randomly generated crystals or a given list of desired atomic configurations. New structures are sampled using the ML potential from the previous iteration based on the extrapolation criterion. Selected structures are used to update the training set of the potential after calculating their energies, forces and stresses in the DFT. Next, the MLIP is re-trained on the new data, and the training cycle is repeated. Finally, the training stops if no new structures are sampled from the MD runs for a certain number of cycles.

usually interrupted if the structures start breaking a certain extrapolation threshold of the ML potential. At the next step, all the configurations from parallel MD runs are aggregated, cleaned from duplicates, and their energies, interatomic forces and stresses are calculated at the DFT level. Finally, the new data set is merged with the training set from the previous iteration and the ML potential is retrained. The configuration space of a given system is considered to be fully explored (and thus the ASCT is converged) once the new configurations are not sampled during the MD runs for a reasonable number of iterations (usually about 10-15 iterations). ASCT allows for both initializing the training with a pre-trained MLIP using a desired minimalistic training set, or "from scratch" with a bare potential and no user input, thus effectively including all the random structures from the first iteration to the initial training set. Our ASCT workflow is currently interfaced only with the Moment Tensor Potentials (MTP) framework[61], which was chosen as the primary MLIP implementation in this work because of its relatively high accuracy[62,63], convenient code implementation[64], and active learning capabilities[65]. Moreover, it has been successfully used in many recent computational studies for predicting the crystal structure of bulk materials[66], phase behavior of alloys[67–70], and characterizing the phonon, kinetic and mechanical properties of 2D materials[71–73].

We used the ASCT framework to train the MLIP for the 2D Mo–S/Al_2O_3 system, suitable for crystal structure prediction purposes. Since its computational cost is much lower compared to DFT, the potential can predict energies, forces and stresses for large structures, thus allowing one to consider large cells with low lattice mismatch. We provide the technical details of the training in Methods, and only demonstrate the final results in this part of the manuscript. Table 1 shows the accuracy of the final model in predicting energies, interatomic forces and stresses of the structures in the validation set. When interpreting these results, it is important to keep in mind that the training set consists mostly of random, and thus highly diverse and non-equilibrium structures, resulting in a huge range of energies and forces to be fitted. The percentage values of the root mean square error (RMSE) are derived as the ratio of the RMSE in models' predictions to the root mean square deviation of the target values from their mean. This essentially gives an insight into how large the errors are compared to the overall range of values of a corresponding target quantity. In Supplementary Fig. 1, we also provide both training and validation pair plots for energies and forces predictions.

Even though the errors presented in Table 1 are not negligible, the trained MLIP can be used for a high-throughput screening of potentially stable candidates in a given system, with a subsequent refinement of the results within accurate DFT calculations. Nevertheless, as we show in the following sections, the binary phase diagrams of the Mo-S system, predicted with MTP and recalculated with DFT are quite close. This may indicate that the training error accumulates on highly non-equilibrium structures, while the predictions on relaxed structures are more accurate.

Substrate-aware structure prediction of 2D crystals

While predicting the structure of the free-standing 2D crystals is a common practice[41–49], incorporating substrate effects into this type of simulation has been challenging for several reasons. First, the primitive cells of a 2D crystal and a substrate are usually not identical, which leads to the non-physical strain caused by the lattice mismatch. One possible solution is to find a suitable supercell representation matching both the 2D crystal and the substrate. This reduces the effect of mismatch on the structure and properties of the considered material. Second, the resulting supercell size of 10^3–10^4 atoms does not allow for high-throughput DFT relaxations of the crystals that are typically performed during the EA run. Machine learning-based models address this problem by combining the near ab initio accuracy and low computational cost, but require a thorough preparation of the training data for each particular system. We have already discussed the latter in the previous section and focus on a robust and convenient way to perform the CSP below.

The evolutionary search for the new stable structures in a 2D Mo–S/Al_2O_3 system was performed with a modified version of the USPEX code. USPEX is widely used[74] for predicting the crystal structure of bulk materials, two-dimensional materials in a vacuum[45], reconstructions of surfaces[75] and nanoclusters[76], as well as for discovering stable materials with optimal properties[77,78].

The workflow of the developed algorithm is summarized in Fig. 2. In each generation, a set of 2D crystal structures with a constrained thickness is first produced using either USPEX evolutionary operators or a random structure generator[79,80]. These structures then undergo a preliminary relaxation and annealing step first without a substrate with the trained ML potential. Here, we assume that the equilibrium lattice parameters of the 2D crystal are by greater extent defined by its composition, while the effect of the substrate is considered as a next-order correction. As the equilibrium lattice parameters and atomic positions are determined, we join the 2D crystal with the substrate using the lattice-matching algorithm (see Methods for details) and perform a conjugate gradient relaxation of the atomic positions, followed by evaluation of the total energy of the system. The contribution of the freestanding substrate is then subtracted from the total energy by detaching it from the structure and evaluating its energy, while keeping its atoms fixed. Finally, each structure in the set is ranked according to its value of fitness, which is calculated as an energy above hull in a binary Mo-S phase diagram. We note, that the convex hull construction is valid precisely because of the fixed thickness (see more details in ref. 50). The structures in subsequent generations of EA are produced based on the set of best representatives identified up to that point using heredity and transmutation evolutionary operators (details are provided in Methods). Thus, the algorithm works iteratively until the list of the best structures remains unchanged for a reasonable number of generations.

Stable 2D Mo–S crystals on Al_2O_3 substrate

2D TMDs are typically manufactured using CVD techniques, in which tiny flakes of the material are condensed from the gas phase of the precursors onto the surface of the substrate, where they undergo a chemical reaction. Sapphire substrates are often chosen for their durability and high chemical stability[81]. We follow the outlined experimental setup and predict the stable Mo–S crystals on the same *c*-cut sapphire substrate. Our main motivation here is to computationally explore the possible outcomes of the CVD growth upon varying the growth parameters (e.g., partial pressures and temperatures of precursors), which affect the chemical potentials of the elements and can lead to stabilization of different 2D Mo–S phases.

The results of the evolutionary search for stable Mo–S crystals on the Al_2O_3 substrate (see the details of the calculation in Methods) are shown in Fig. 3. In order to study the relative thermodynamic stability of the predicted Mo-S phases, we use a convex hull approach in the binary phase diagram[56,57]. The convex hull, by definition, connects phases that have lower energy than any other phase or any linear combination of phases at the same overall composition. Therefore, by plotting different structures found in the EA search as points in the formation energy-composition space, one can determine the stable structures by identifying the smallest set of points that form a convex hull.

The top left panel in Fig. 3 shows the convex hull for the freestanding 2D crystals, while the top right panel shows the convex hull for the crystals attached to the substrate. The bottom panel contains top and side views of the 2D visualization of each Mo–S structure from the convex hull. The green shading around the convex hull segments represents the error of the MTP potential in predicting the formation energy, which is equal to $\sqrt{3\sigma^2_{\mathrm{MTP}}}$ and $\sqrt{4\sigma^2_{\mathrm{MTP}}}$ for freestanding and substrate-bound structures, respectively (where σ_{MTP} is a RMSE per atom of the MTP potential in predicting energies from Table 1). We chose the error window according to the errors summation rule for the formation energy formula (see Methods). For each structure falling within this error window, we re-evaluated the formation energies in DFT (relaxing both the atomic positions and cell parameters for the freestanding 2D layers, while doing only single-point DFT calculations for those joined with the substrate).

First of all, we note that the 1H-MoS_2 structure with $P\bar{6}m2$ space group and a lattice parameter $a^{\mathrm{MTP}}_{\mathrm{MoS_2}} = 3.15$ Å was successfully found during the

EA search. The lattice constant predicted within the MTP relaxation is very close to the experimental one, which varies from 3.12 Å[82] to 3.22 Å[83,84]. This observation indicates that despite being trained on random structures, our MLIP can successfully identify the equilibrium. In addition, we found three new Mo–S structures, namely *Pmma* Mo_3S_2, low-symmetry $P\bar{1}$ Mo_2S, $P2_1m$ Mo_5S_3, and *P4mm* Mo_4S, which are either close to or located on the convex hull sections and therefore are stable. The lattice parameters of the new structures are listed in the Supplementary Table 1.

Since our study is mainly focused on the overall effect of the substrate on the properties of 2D materials, we next analyzed the way substrate changes the stability of 2D structures. Indeed, the set of stable structures on a convex hull is changed under the effect of the substrate: in both the MTP and DFT cases, a new Mo_4S structure appears on the convex hull after joining with the substrate (see the upper right panel in Fig. 3), while Mo_2S loses its stability in the MTP case. The presence of structures such as Mo_4S on a convex hull indicates that chemical bonding with the substrate saturates a certain fraction of the bonds in the 2D layer, stabilizing structures that would not be stable in a freestanding 2D case (see Supplementary Fig. 7). Moreover, the relative formation energies of the phases alter upon joining with the substrate (see Supplementary Table 2), therefore allowing one to favor a desired phase during experimental fabrication by choosing a suitable substrate and adjusting the synthesis parameters. For example, changes in the slope of the convex hull sections are directly related to the changes in the chemical potentials of molybdenum and sulfur, and thus to the partial concentrations of CVD precursors under which a specific phase of the material can be stabilized. We discuss the synthesis conditions under a given range of chemical potentials in more detail in the following sections and in the SI.

Table 1 | Performance of the trained MTP potential for the Mo–S/Al_2O_3 system in predicting energies, forces and stresses on a validation set

	Energies, meV/at.	Forces, meV/Å	Stresses, kbar
MAE	37.9	307.4	2.11
RMSE	50.6	442.6	3.37
RMSE, %		6.5	0.64
Total number of configurations	3240		
Total number of atoms in all configurations	283059		

The mean absolute error (MAE) and root mean square error (RMSE) are shown along with the ratio of the RMSE to the overall range of forces and stresses in the data set.

Despite the certain inaccuracy of the MLIP, the sets of stable structures predicted by the DFT and MTP overlap significantly. This justifies the proposed pipeline for high-throughput stability studies a preliminary MLIP-based screening of the configuration space within the evolutionary search is performed first, and then followed by a re-evaluation of the formation energies of a subset of stable structures falling within the window of MLIP errors. This approach reduces the overall cost of the study by several orders of magnitude and allows one to achieve the scaling of the system size that is typically inaccessible in a pure DFT approach.

The last limitation we want to discuss here is the use of potential energies to analyze the relative stability of the structures. Realistic experimental setups require Gibbs formation energy calculations and consideration of entropy effects. In this study, for simplicity, we focused on potential energy contributions and analyzed the substrate effect in terms of a single well-understood quantity, leaving more accurate energy calculations to future research. As our recent work has shown (see ref. 85), the most laborious part here is the calculation of the vibrational entropy contribution, since one must use accurate large-scale MD simulations with the pre-trained MLIPs to obtain accurate results. Fortunately, a suitable MLIP for the Mo–S/Al_2O_3 system has already been trained in this study, allowing us to account for the entropy effects in the future.

We also compared our results with the study of the stable freestanding 2D Mo–S structures[86], where the standard DFT approach was used to perform the local relaxations and energy evaluations. We note that most of these reference stable structures (except Mo_5S and Mo_5S_4) were successfully found in our simulations. In Supplementary Fig. 8, we compare our results with those reproduced from ref. 86. All these reference structures appear to be above the convex hull, indicating their thermodynamic instability. There could be several reasons for such a discrepancy. First, there is no robust and trustworthy way to determine whether the evolutionary algorithm has successfully found the global minimum. Compared to ref. 86, we performed the relaxation of the structures at finite temperature. This usually leads to a smoother potential energy landscape and eliminates some of the local minima that exist in the zero temperature DFT relaxation[85] and could cause the evolutionary search to get stuck. Another related reason is a dynamical instability of some of the thermodynamically stable structures found in

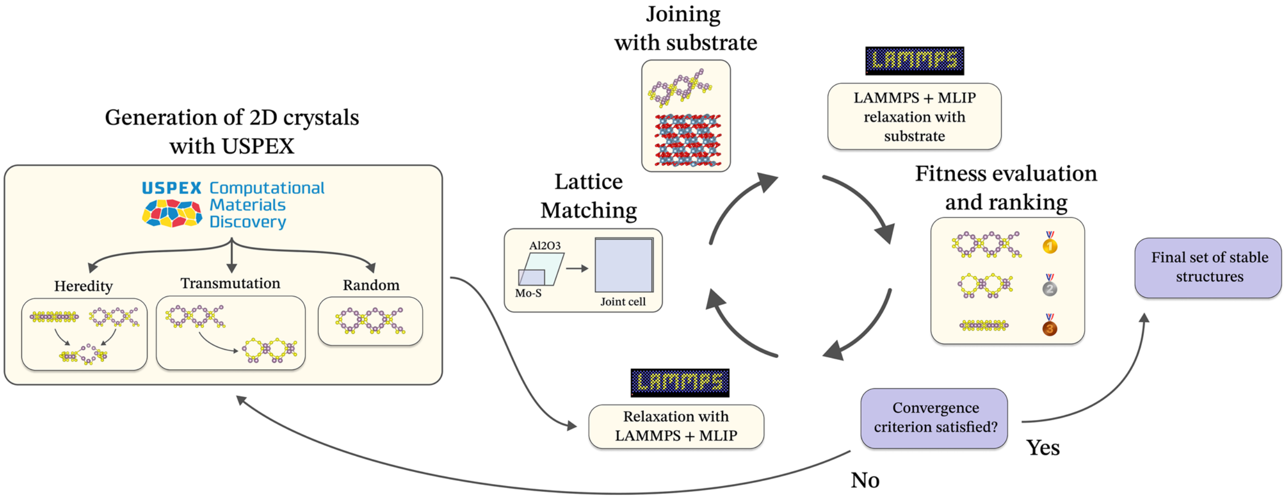

Fig. 2 | Evolutionary algorithm workflow. At each iteration, a set of 2D structures is generated using the evolutionary operators in a given chemical space. After a preliminary relaxation of the atomic positions and lattice parameters, these structures are joined with the substrate using a lattice-matching algorithm that ensures no lattice mismatch. Finally, the structures are ranked by their fitness value, and the cycle is repeats. The calculation stops when the set of best structures remains unchanged for a given number of iterations. A detailed explanation of the algorithm and evolutionary operators used is given in Methods.

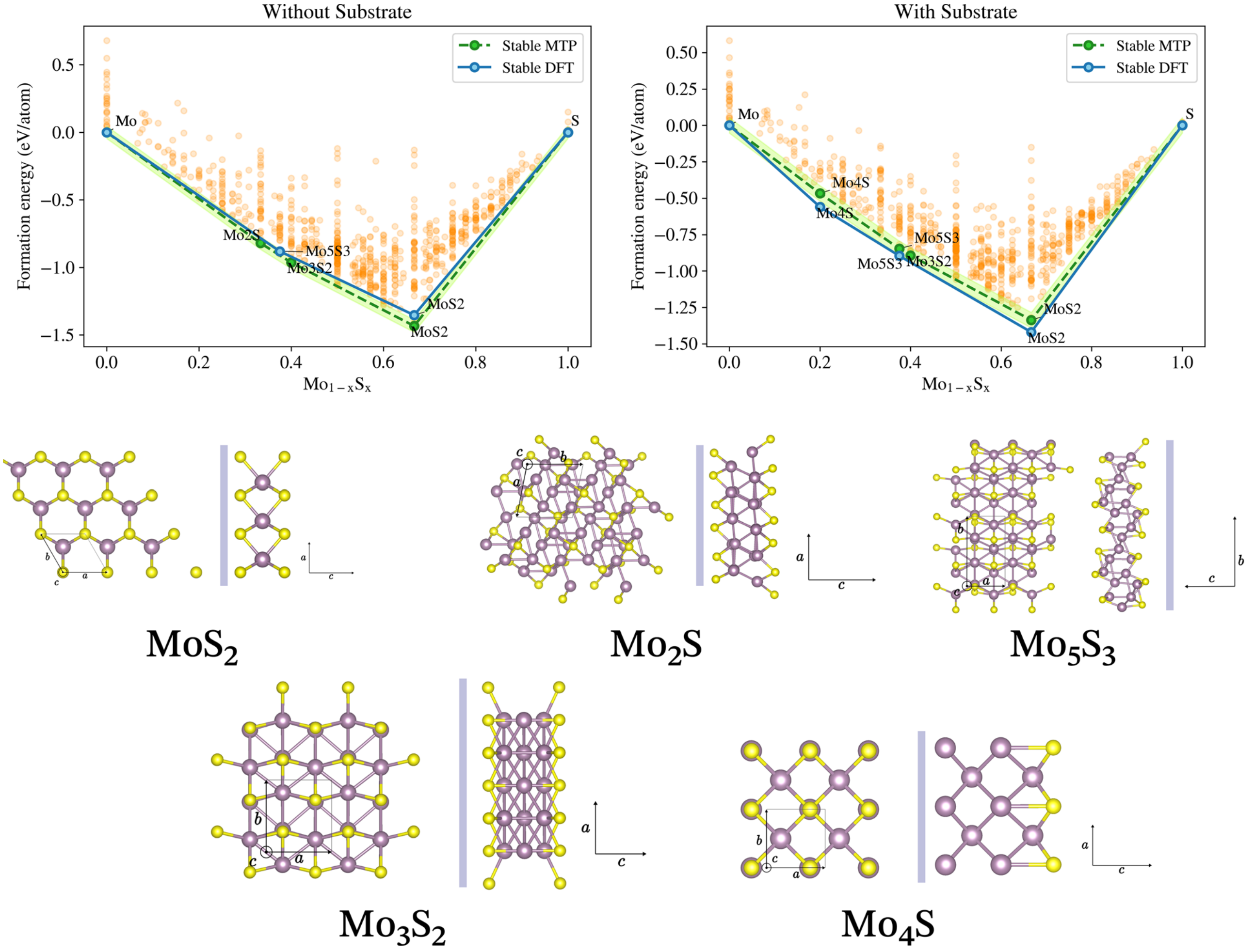

Fig. 3 | Results of the evolutionary search in the Mo–S/Al_2O_3 system. The upper left panel shows the convex hull together with the set of stable (green and blue circles) and metastable (orange circles) freestanding 2D Mo–S structures. The upper right panel shows similar results, but for the structures joined with the substrate. The green and blue convex hulls show the results predicted with MTP and recomputed with DFT, respectively. The shaded green area represents the error of the MTP potential in predicting the formation energies. Stable structures are visualized in the lower panel, where molybdenum atoms are drawn as larger violet spheres and sulfur atoms as smaller yellow spheres. The position of the substrate relative to the 2D layers is shown schematically in a side view of each structure.

ref. 86, such as Mo_5S and Mo_5S_4. Due to the finite temperature relaxation, the structures located at saddle points on the potential energy surface could undergo the structural transformation and thus never be found during the evolutionary search.

Electronic and phonon properties

In addition to stability study, it is important to analyze the electronic and phonon properties of the discovered phases. In Supplementary Fig. 9, we show the DFT-calculated electronic band structures of all five Mo–S structures presented earlier. Most of the stable structures found during the calculationexhibit metallic behavior, except for the semiconducting MoS_2. The latter has a direct DFT band gap of 1.8 eV, which is in perfect agreement with existing DFT results in the literature[87]. The special feature of the MoS_2 monolayer is that its DFT band gap value is nearly identical to that obtained from photoluminescence (PL) and optical absorption experiments[88] (1.85 eV). Although DFT calculations are known to underestimate the band gap values due to the lack of electronic correlations, here they provide better agreement with the experiment compared to the GW approach (2.8 eV)[87]. However, this is mostly explained by a strong exciton binding (~ 1 eV), which compensates the lack of correlations and can be observed in more accurate first-principles methods (e.g. solving the Bethe–Salpeter equation (BSE))[89–91].

Since other stable phases are rich in molybdenum, the electron density is mostly concentrated on the molybdenum atoms, therefore determining the metallic behavior of the corresponding structures. This property can be extremely useful in fabrication of low-dimensional electronic and optical devices. For instance, the Mo_3S_2 and Mo_5S_3 layers can actually coexist with MoS_2 according to the convex hull (as shown in Fig. 3), thus allowing for creation of 2D metal-semiconductor lateral heterostructures, where the metallic phase can act as an electrode[86].

In order to study the dynamic stability of the predicted 2D structures and examine the way they behave in realistic conditions, we performed the calculation of the phonon band structures in a quasi-harmonic approximation using a finite-displacements method for all the new freestanding 2D crystals from Supplementary Table 2 using the trained MLIP and DFT for energy and force evaluation (see Supplementary Fig. 10). Most of the structures in the list have a certain number of imaginary frequencies associated with the out-of-plane vibrational modes of the 2D layer in the DFT results. However, the density of states corresponding to these frequencies is often close to zero, resulting in a small contribution of these modes to the total lattice vibrations. MLIP tends to overestimate the stability of the structures and shows no imaginary modes in the majority of cases. This is most likely due to the extremely large configuration space presented in the training set, the coverage of which inevitably leads to a relatively worse accuracy for structures close to equilibrium. Nevertheless, these results demonstrate that with certain assumptions, the trained MLIP is capable of both performing high-throughput search for stable structures within the structure prediction protocol, and estimating the dynamical stability of the stable candidates.

Effect of substrate on phonon properties

The most interesting part, however, is identifying the effect of the substrate on the dynamical stability of the structures. However, this requires using a different method to calculate the phonon properties, since the finite displacement approach typically requires solving an eigenproblem of a $3N\times 3N$ dynamical matrix, which is not accessible due to the large size of the systems joined with the substrate. Therefore, we calculated the phonon DOS for the MoS_2 structure both with and without the substrate as the Fourier transform of the velocity autocorrelation function (see Methods for details). This method not only allows us to compute the phonon DOS for large systems, but also reveals a full anharmonic picture of the lattice vibrations. Our results are shown in Fig. 4.

Interaction with the substrate leads to the energy transfer between the 2D layer and the substrate itself, resulting in both shifting and broadening of the DOS peaks. This can affect both the specific heat capacity of the layer and its thermal conductivity, enabling the design of desired vibrational patterns by choosing a suitable substrate. We also note that similar effects can sometimes be observed when the lattice vibrations are highly anharmonic. Here, however, the calculations were performed in a fully anharmonic setup to eliminate this effect. A more detailed discussion of the anharmonicity effect in the case of the 2D MoS_2 can be found in the SI.

Prediction of synthesis conditions

Although the evolutionary search yields a set of thermodynamically stable structures, synthesizing these structures experimentally remains quite complicated and, in fact, hardly related to the computational study we have presented so far. In particular, even the synthesis of 2D MoS_2 is represented by at least six different techniques, including CVD growth[81,92], ALD[93–95], and electron-beam deposition (EBD)[96–98]. A comprehensive review of 2D MoS_2 growth methods can be found in ref. 99.

Nevertheless, one of the key properties of the convex hull phase diagrams (Fig. 3) is the direct relationship between the slope of the convex hull sections and the chemical potential value at which the corresponding phases can be stabilized. In order to link the predicted binary phase diagram of the 2D Mo-S system with the experimental synthesis conditions, it is necessary to relate the values of the chemical potentials to the parameters that can affect and control them in each specific experimental setup. In this case, a key quantity that determines the stability of the structure is its Gibbs formation energy G_f:

$$G_f(P,T)=\frac{1}{N}\left(G(P,T)-G_{sub}(P,T)-\sum_i n_i\mu_i(P,T)\right) \tag{1}$$

where $G(P, T)$ is a Gibbs energy of a structure joined with the substrate, calculated at a given pressure P and temperature T, $G_{sub}(P, T)$ is the Gibbs energy of a clean substrate, n_i, $\mu_i(P, T)$ are number of atoms and chemical potential of atomic type i, and $N=\sum_i n_i$ is a total number of atoms in the 2D layer. For each value of μ_i, the most stable structure is then given by a minimum value of the Gibbs formation energy among the set of considered structures. For simplicity, we only consider the potential energy contribution to the Gibbs free energies of the structures and the substrate.

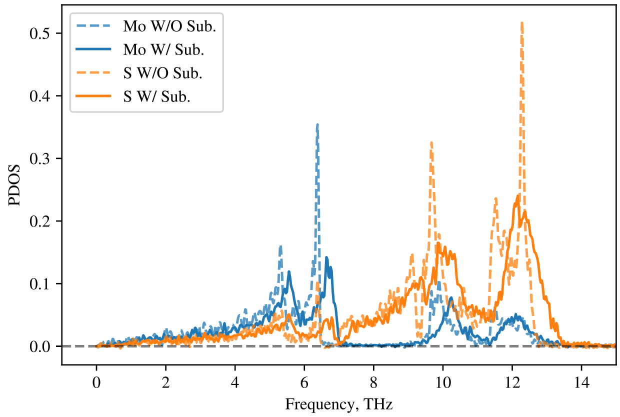

Fig. 4 | Phonon density of states of the 2D MoS_2 layer with (solid lines) and without (dashed lines) substrate. The partial contributions of molybdenum and sulfur atoms are shown in blue and orange, respectively. The effect of the substrate leads to a broadening of the DOS peaks and a shift in their positions.

We used an experimental setup for the CVD synthesis of 2D MoS_2 from ref. 81 to demonstrate how the synthesis conditions can be predicted from the ab initio thermodynamics approach[100]. In this setup, sulfur is vaporized from a sulfur boat and transferred to a furnace in a quartz tube with a sapphire substrate and MoO_3 precursor, leading to the formation of MoS_2 layers (see left panel in Fig. 5). The vaporization rate of sulfur (controlled by the temperature of the sulfur boat), the temperature of the furnace, and the carrier gas flow rate mostly determine the pressure and temperature of the reaction components. Therefore, we linked the values of the chemical potentials of molybdenum and sulfur to these quantities to identify the stability regions of the structures presented on the convex hull (Fig. 3). First, we assumed that the partial pressure of sulfur is determined by a certain fraction of its saturated vapor pressure at the temperature of the sulfur boat T_{SB}, and sulfur is transferred to the reaction chamber without loss. Additionally, even though sulfur vapor can exhibit different molecular configurations (from S_2 to S_8), we nevertheless considered only S_2 molecules, since their concentration is predominant at $T > 1000$ K[101]. Similarly, the partial pressure of MoO_3 molecules is determined by a fraction its saturated vapor pressure at the furnace temperature T_F. Finally, we assumed that in the thermodynamic limit, the reaction between MoO_3 and sulfur leads to the formation of molybdenum and SO_2 gas:

$$S_2(g)+MoO_3(g)\rightarrow Mo(s)+SO_2(g). \tag{2}$$

We used the ideal gas approximation for sulfur vapor, MoO_3 vapor and SO_2 gas, to derive the chemical potentials of molybdenum and sulfur, and linked them to the stability patterns from the EA search. A detailed explanation of the underlying calculations, as well as a theoretical background of the ab initio thermodynamics is presented in the SI.

Our results are shown in the right panel of Fig. 5. For simplicity, we assumed that the flow rate of the carrier gas is constant (the reliability of this assumption is discussed further in the text), and only the temperatures of the furnace T_F and the sulfur boat T_{SB} affect the chemical potentials of molybdenum and sulfur. As expected, MoS_2 remains thermodynamically stable in a wide range of temperatures, including the experimental synthesis conditions ($T_{SB}\sim 800$ K and $T_F\sim 1000$ K)[81]. Both decrease in the temperature of the sulfur boat (and thus decrease in the sulfur vaporization rate) and increase in the furnace temperature (which promotes the MoO_3 evaporation) can lead to stabilization of the new structures with higher molybdenum content.

While interpreting these results, it is important to keep in mind that the predictions are made in a purely thermodynamic limit and do not take into account the kinetics of the process. For this reason, altering the carrier gas flow rate goes beyond the scope of our model. The specific values of the temperatures T_F and T_{SB} are therefore descriptive in their nature, and should serve as a guide to the actual experimental procedure rather than a solid synthesis recipe. Nevertheless, the universality of the proposed approach enables the construction of phase diagrams for nearly any experimental setup, once the relationship between the synthesis parameters and the chemical potential of components is established.

In this study, we have developed and demonstrated a comprehensive method for predicting the atomic structure and stability of two-dimensional materials on arbitrary substrates, as well as their experimental synthesis conditions. We integrated an evolutionary algorithm with a lattice-matching subroutine and machine learning interatomic potentials for relaxation to enable efficient exploration of a wide range of atomic configurations and chemical compositions. The

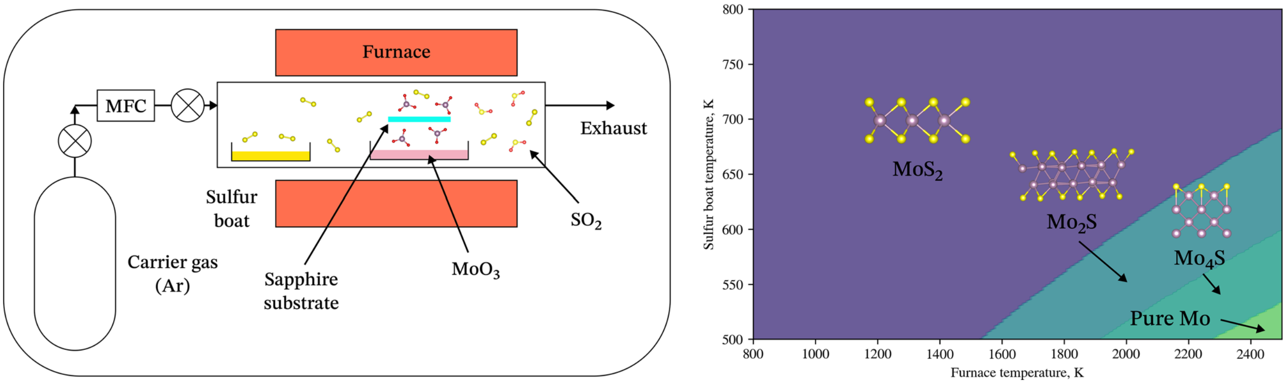

Fig. 5 | Schematic representation of the CVD synthesis of MoS_2 using sulfur and MoO_3 precursors (left) and the phase diagram of the 2D Mo-S system (right). In an argon atmosphere, sulfur is vaporized from a sulfur boat and transferred to a furnace in a quartz tube with a sapphire substrate and MoO_3 precursor. As the precursors react, MoS_2 layers grow on the substrate, and by-products (such as S_x and SO_2 molecules) are removed from the reaction chamber. By controlling the temperature of the furnace (T_F) and the sulfur boat (T_{SB}), it is possible to alter the chemical potentials of sulfur and molybdenum. Phase diagram of the 2D Mo–S system from the evolutionary search is plotted in (T_F, T_{SB}) coordinates. At the experimental conditions of MoS_2 synthesis ($T_{SB} \sim 800$ K and $T_F \sim 1000$ K)[81], MoS_2 is predicted to be stable. By increasing the T_F and decreasing the T_{SB}, it is possible to stabilize other Mo–S phases found in this work.

proposed ASCT algorithm for MLIP training allow for the automatic generation of fast and accurate interatomic potentials for arbitrary simulation scenarios, including crystal structure prediction. Our approach was validated on the Mo–S system on a *c*-cut sapphire (Al_2O_3) substrate, demonstrating its efficiency in identifying stable 2D crystal structures and evaluating their electronic and phonon properties. The evolutionary search revealed several new stable configurations: *Pmma* Mo_3S_2, $P\bar{1}$ Mo_2S, $P2_1m$ Mo_5S_3, and *P4mm* Mo_4S, where the Mo_4S structure is specifically stabilized by interaction with the substrate. These structures extend the known landscape of 2D materials and open new possibilities for practical applications. Electronic band structure calculations showed that all the new structures exhibit metallic behavior. This suggests the potential to create 2D metal-semiconductor lateral heterostructures, where the metallic phases could serve as electrodes in electronic devices. Phonon calculations demonstrated that some of the predicted structures show dynamical instabilities associated with the out-of-plane vibrations of the 2D layer. Interaction with the substrate was found to alter the phonon density of states, potentially allowing modulation of the thermal and mechanical properties of the 2D materials. Based on the ab initio thermodynamics approach, we demonstrated the phase diagram of the predicted 2D Mo–S structures in the parameter space of the CVD experimental setup. Our results highlight the importance of including substrate effects in computational studies to accurately predict the stability and properties of 2D materials. The trained MLIP not only facilitates accurate predictions but also significantly reduces computational costs, making it feasible to study large systems that are otherwise inaccessible within the conventional DFT methods. Overall, our approach opens new horizons in 2D materials discovery, allowing to find new substrate-stabilized phases even in well-known and thoroughly studied systems, obtain the configurations for characterizing the substrate effect on various properties of 2D materials, and predict the possible conditions for their synthesis to design the next generation of electronic and optoelectronic devices. Future work will focus on refining the MLIP training process further and extending the application of this method to other material systems and substrates, aligning the computational approaches better with real experimental setups, and experimental verification of newly discovered 2D materials.

Methods

MLIP training details for 2D Mo–S/Al_2O_3 system

2D Mo–S system. The ML potential for the Mo–S/Al_2O_3 system was trained within the ASCT framework in three stages, where we first prepared the MLIP for 2D Mo-S system. Each iteration of the ASCT was initialized with random 2D structures in the Mo–S system with 2 to 16 atoms in the unit cell, generated using the symmetry-based random structure generator from a PyXtal package[80]. These structures were then used to initialize the sampling in 20 parallel MD simulations. Each MD run began with a replication of the unit cell in the in-plane directions to obtain the structure with approximately 64 atoms in the unit cell. The following steps were then performed for each structure:

1. A conjugate gradient relaxation of the atomic positions with a fixed cell and a convergence criterion of 10^{-10} eV.
2. NPT annealing at T = 10 K and external stress of 1 bar for 10 ps
3. NPT heating from 10 K to 300 K at external stress of 1 bar for 50 ps
4. NPT anneal at T = 300 K and external stress of 1 bar for 50 ps

A time step of 1 fs was used for all simulations, and the external stress was only applied along non-periodic dimensions. The MTP extrapolation grade thresholds for sampling and run interruption were set to 3 and 10, respectively. The cutoff radius for the local atomic environment representation was 5 Å. The complexity of the model in terms of the size of the basis set was fixed by selecting the 24g.mtp initial potential file. Each iteration of training was performed with weights of 10, 0.01, 0.001 for energies, forces, and stresses contributions in the loss function, while the weight scaling for energies and forces was set to 2 and 1, respectively. The ASCT convergence criterion was set to 25 iterations. Although it is technically possible to initialize the ASCT with a pre-trained potential, in this work, we did not use pre-training and instead explored how the ASCT algorithm naturally builds the potential from scratch. This was done to show that a sufficiently accurate potential can be obtained automatically without user involvement. Therefore, during the first few ASCT iterations, all generated random structures entered the training set immediately until sufficient coverage of the configuration space was achieved to begin sampling from MD trajectories. Finally, the ML potential for the 2D Mo–S system was obtained after 41 iterations of the ASCT and the training set contained 727 structures.

Al_2O_3 surface slab. In the second stage, the *c*-cut surface slab of α-Al_2O_3 with 48 Al and 72 O atoms and a thickness of 10 Å was used to initialize the sampling and to train a separate interatomic potential. The ASCT routine and MTP training parameters were almost identical to those used in the case of a Mo–S system, except for the replication of the unit cell and a lower convergence criterion of 15 iterations. Thus, the potential was obtained in 27 iterations of ASCT with a total of 1140 structures in the training set.

2D Mo–S/Al_2O_3. In the final, third stage, the training sets for the Mo–S and Al_2O_3 systems were merged to train the initial version of the desired Mo–S/Al_2O_3 potential. This time, the MD sampling part of ASCT was initialized with the random Mo–S structures containing up to 16 atoms in the unit cell, stacked with a surface plate of Al_2O_3 using a lattice matching algorithm (see details below). To avoid abnormally large structures, we did not apply lattice multiplication in this step and just put the generated random structure into the fixed unit cell of the Al_2O_3 surface slab. Despite the lattice mismatch strain induced by this operation, these structures are still suitable for training the interatomic potential, especially given their "random" nature. The rest of the training was done in the same way as in the previous cases. The convergence criterion of ASCT was 15 iterations, and the training was done in a total of 37 iterations with 3218 structures in the training set.

Details of the ab initio calculations

Calculation of data for MLIP training. Training data on energies, interatomic forces and stresses of the new configurations within the ASCT workflow were obtained at the density functional theory (DFT) level using the VASP package[102,103]. The convergence criteria for an electronic self-consistent cycle was 10^{-6} eV, and the cutoff energy for a plane-wave basis set was 600 eV. We used a Gaussian smearing scheme with a width of 0.05 eV to represent the band occupancy. The first Brillouin zone was represented by a Γ-centered uniform grid with a density of $2\pi \cdot 0.03$ Å. The behavior of the core electrons and their interaction with the valence electrons was described within Projector-Augmented Wave (PAW) pseudopotentials[104] with 12 [4p5s4d], 6 [s2p4], 3 [s2p1], and 6 [s2p4] valence electrons for Mo, S, Al, and O, respectively. The electronic exchange correlation effects were modeled within a generalized gradient approximation (Perdew-Burke-Ernzerhof functional)[105]. A vacuum layer of 20 Å was added to all structures along the direction normal to the surface plane to avoid surface interactions due to periodic boundary conditions. Due to the short-range nature of the trained MLIP and the lack of explicit dispersion corrections in the MTP package, dispersion corrections were not included in the DFT setup. While this approach may limit the overall accuracy of the predictions, it is not possible to properly describe long-range interactions using the short-range MLIP approach. Therefore, using DFT with van der Waals (vdW) functionals as a source of training data would lead to systematically worse prediction errors.

Band structure calculations. The electronic band structure calculations were performed at the DFT level using a *RelaxBandStructureMaker* utility implemented in the *atomate2* package[106] and the VASP backend. All crystal structures (i.e. their positions and cell parameters) were first relaxed until all interatomic forces were less than $2 \cdot 10^{-2}$ eV/Å. Next, the standard SCF calculation was performed to obtain the Kohn-Sham orbitals on a coarse k-point grid with a spacing of 0.3 Å^{-1}. Finally, a band structure calculation was performed for a specific k-path generated by the SeeK-path utility[107] based on crystal symmetry. We used the same set of PAW potentials as for MLIP training and evaluation, but chose a slightly higher cutoff of 680 eV to accurately describe the surface effects of the electrons. We also switched to the PBEsol functional[108] to describe the exchange-correlation effects of the electrons, as this usually results in more accurate electronic band structures. Partial occupancies of the orbitals were set using a Gaussian smearing method with a width of 0.01 eV.

Phonon properties calculations. For the free-standing 2D crystals, phonon band structures and densities of states were calculated in a quasi-harmonic approximation with a *Phonopy* package[109–111]. We used both VASP- and MTP-based relaxation and force constant evaluation to evaluate the accuracy of the trained ML potential. First, the positions and cell parameters of all crystals were relaxed until the interatomic forces were less than $5 \cdot 10^{-4}$ eV/Å. Next, the set of symmetric nonequivalent finite displacements was generated using *Phonopy* utilities, and the resulting interatomic forces were calculated to evaluate the dynamical matrix. This matrix was finally diagonalized to obtain a set of eigenfrequencies for the corresponding set of wavevectors. Compared to previous DFT calculations, we used a slightly denser k-point grid with a spacing of 0.2 Å^{-1}, but a larger electronic smearing of 0.05 eV, as this helped to obtain better converged results.

For the systems connected to the substrate, we used a different approach to calculate phonon density of states based on the Fourier transform of the velocity autocorrelation function (VACF) (Eq. (3)).

$$g(\nu) = 4\int_0^\infty cos(2\pi\nu t)\frac{\langle\overline{\nu(0)\nu(t)}\rangle}{\langle\overline{\nu(0)^2}\rangle}dt \tag{3}$$

VACF was calculated during NVE molecular dynamics simulation (with trained MTP potential) for each studied system. It allows one to calculate the phonon DOS in the full anharmonic picture and is generally more accessible for large systems than a finite displacement method.

We also used the same approach to calculate the phonon DOS for the freestanding 2D layers, in order to identify the effect of the substrate and the potential influence of the anharmonicity.

Details of the evolutionary search

The evolutionary search for the new stable structures in a 2D Mo–S/Al_2O_3 system was performed with a modified version of the USPEX code. During the search, the structures were allowed to have from 4 to 16 atoms of variable Mo–S composition in the unit cell, while their thickness was constrained to 6 Å. Each generation of the evolutionary search consisted of 120 structures, except for the first one, which had 180 structures. The first generation was generated with a symmetry-based random structure generator, while structures in all subsequent generations were generated with a heredity (40 %) and transmutation (30 %) evolutionary operator using the subset of the best structures from the previous generation. The heredity operator has been adapted to 2D crystals from its 3D analog in the USPEX code[56,57]. It first slices two parent structures along a random direction and then alternately combines the slices to create a new structure. The transmutation operator uses only one parent structure and creates a new one by randomly assigning new chemical identities to a group of atoms. The remaining 30 % of the structures in each generation were generated randomly. The total number of generations was limited to 150, while the evolutionary search was considered to converge when the list of best structures remained unchanged for 25 consecutive generations.

Local relaxation of the structures was performed in three steps. In the first step, the generated 2D Mo–S configurations were relaxed and annealed with the pre-trained MTP potential and a LAMMPS package, following the same protocol used to train the MLIP (see section 8). Next, the relaxed 2D structures were joined to the Al_2O_3 substrate using a lattice-matching algorithm (see details below) with an initial gap value of 2.0 Å, a maximum mismatch criterion of 5×10^{-3}, and a maximum value of the resulting substrate area of 1000 Å^2. This joint structure was again relaxed and annealed in the same manner, while all the atoms of the substrate located 3.0 Å below the surface were frozen. In the last step, the resulting configuration of the substrate was used for a single-point calculation of the substrate energy for further calculation of the fitness function of the structures. Finally, as the relaxation is done, the fitness function of the structures was calculated in terms of the energy over the composition convex hull. To do this, we first calculate the formation energy ΔE_f of each structure $\Delta E_f = \frac{1}{N}(E - \sum_i n_i E_i)$, where E is the energy of the structure, n_i, E_i are the number of atoms and energies of each pure component, and $N = \sum_i n_i$ is the total number of atoms in the structure. The energy of the structure is either equal to the total energy from the DFT calculation E_{tot}, or to the difference between the total energy and the clean substrate energy E_{sub}: $E = E_{tot} - E_{sub}$, depending on whether the free-standing 2D layers or those joined with the substrate are considered. Finally, we build the convex hull in the (ΔE_f - composition) space and evaluate the energy over the convex hull to determine the stability of the structure.

Lattice-matching algorithm

To obtain the appropriate supercell representation of the 2D material on top of the substrate, we adapted the Zur-McGill lattice matching algorithm[112] implemented in the PyMatGen package[113]. The algorithm essentially builds a set of supercell matrices for both substrate and 2D layer unit cells by considering arbitrary linear combinations of the lattice vectors in the plane to eventually generate two roughly equal supercells of similar shape with a desired value of the estimated mismatch based on a ratio of the areas of the resulting lattices. We also constrain the maximum area of the supercells to obtain a Pareto optimal solution in terms of lattice mismatch and system size. This requires a series of trial runs of the algorithm with different values of the maximum area, followed by an analysis of the resulting lattice mismatch. Usually, large values of the maximum area lead to structures with low mismatch but a large number of atoms. It is therefore necessary to choose appropriate parameters before starting the calculations, depending on the desired level of accuracy and the available computational resources. As the supercells are generated, the algorithm connects the 2D layer to a substrate at a given gap distance.

Data availability

All the data produced in this work is available at the Materials Cloud Archive[114]. This includes the training set with Mo–S/Al_2O_3 structures, the trained MTP potential, and the results of the evolutionary search.

Code availability

The code will be merged into the next release of the USPEX code (https://uspex-team.org/).

Received: 26 September 2024; Accepted: 13 July 2025;

References

1. Novoselov, K. S. et al. Electric field in atomically thin carbon films. *Science.* **306**, 666–669 (2004).
2. Wang, Q. H., Kalantar-Zadeh, K., Kis, A., Coleman, J. N. & Strano, M. S. Electronics and optoelectronics of two-dimensional transition metal dichalcogenides. *Nat. Nanotechnol.* **7**, 699 (2012).
3. Castellanos-Gomez, A. et al. Isolation and characterization of few-layer black phosphorus. *2D Mater.* **1**, 025001 (2014).
4. Naclerio, A. E., Kidambi, P. R., Naclerio, A. E. & Kidambi, P. R. A review of scalable hexagonal boron nitride (h-BN) synthesis for present and future applications. *Adv. Mater.* **35**, 2207374 (2023).
5. Anasori, B. & Thakur, A. Accelerating 2D materials discovery. *Science.* **383**, 1182–1183 (2024).
6. Björk, J., Zhou, J., Persson, P. O. & Rosen, J. Two-dimensional materials by large-scale computations and chemical exfoliation of layered solids. *Science.* **383**, 1210–1215 (2024).
7. Haastrup, S. et al. The Computational 2D Materials Database: high-throughput modeling and discovery of atomically thin crystals. *2D Materials* **5**, (2018).
8. Gjerding, M. N. et al. Recent progress of the Computational 2D Materials Database (C2DB). *2D Mater.* **8**, 044002 (2021).
9. Novoselov, K. S. et al. Two-dimensional gas of massless Dirac fermions in graphene. *Nature.* **438**, 197–200 (2005).
10. Gupta, A., Sakthivel, T. & Seal, S. Recent development in 2D materials beyond graphene. *Prog. Mater. Sci.* **73**, 44–126 (2015).
11. Shao, C., Yu, X., Yang, N., Yue, Y. & Bao, H. A review of thermal transport in low-dimensional materials under external perturbation: effect of strain, substrate, and clustering. *Nanoscale Microscale Thermophys. Eng.* **21**, 201–236 (2017).
12. Manzeli, S., Ovchinnikov, D., Pasquier, D., Yazyev, O. V. & Kis, A. 2D transition metal dichalcogenides. *Nat. Rev. Mater.* **2**, 17033 (2017).
13. Iannaccone, G., Bonaccorso, F., Colombo, L. & Fiori, G. Quantum engineering of transistors based on 2D materials heterostructures. *Nat. Nanotechnol.* **13**, 183–191 (2018).
14. Novoselov, K. S. et al. A roadmap for graphene. *Nature.* **490**, 192–200 (2012).
15. Geim, A. K. & Grigorieva, I. V. Van der Waals heterostructures. *Nature.* **499**, 419–425 (2013).
16. Novoselov, K. S., Mishchenko, A., Carvalho, A. & Neto, A. H. C. 2D materials and van der Waals heterostructures. *Science* **353** (2016).
17. Liu, Y. et al. Van der Waals heterostructures and devices. *Nat. Rev. Mater.* **1**, 1–17 (2016).
18. Liu, Z., Lau, S. P. & Yan, F. Functionalized graphene and other two-dimensional materials for photovoltaic devices: device design and processing. *Chem. Soc. Rev.* **44**, 5638–5679 (2015).
19. Ahn, G. H. et al. Strain-engineered growth of two-dimensional materials. *Nat. Commun.* **8**, 1–8 (2017).
20. Yan, F. et al. Toward high-performance photodetectors based on 2d materials: strategy on methods. *Small Methods.* **2**, 1700349 (2018).
21. Wu, Y., Wang, L., Li, H., Dong, Q. & Liu, S. Strain of 2D materials via substrate engineering. *Chin. Chem. Lett.* **33**, 153–162 (2022).
22. Hus, S. M. & Li, A.-P. P. Spatially-resolved studies on the role of defects and boundaries in electronic behavior of 2D materials. *Prog. Surf. Sci.* **92**, 176–201 (2017).
23. Rhodes, D., Chae, S. H., Ribeiro-Palau, R. & Hone, J. Disorder in van der Waals heterostructures of 2D materials. *Nat. Mater.* **18**, 541 (2019).
24. Badaroglu, M. More Moore. *IEEE International Roadmap for Devices and Systems Outbriefs*. 01-38 (*IEEE*, 2021).
25. Mak, K. F. & Shan, J. Photonics and optoelectronics of 2D semiconductor transition metal dichalcogenides. *Nat. Photonics.* **10**, 216–226 (2016).
26. Das, S., Pandey, D., Thomas, J. & Roy, T. The role of graphene and other 2D materials in solar photovoltaics. *Adv. Mater.* **31**, 1802722 (2019).
27. Schaibley, J. R. et al. Valleytronics in 2D materials. *Nat. Rev. Mater.* **1**, 1–15 (2016).
28. Shi, L. & Zhao, T. Recent advances in inorganic 2D materials and their applications in lithium and sodium batteries. *J. Mater. Chem. A.* **5**, 3735–3758 (2017).
29. Luo, B., Liu, G. & Wang, L. Recent advances in 2D materials for photocatalysis. *Nanoscale.* **8**, 6904–6920 (2016).
30. Mannix, A. J., Kiraly, B., Hersam, M. C. & Guisinger, N. P. Synthesis and chemistry of elemental 2D materials. *Nat. Rev. Chem.* **1**, 1–15 (2017).
31. Dean, C. et al. Graphene based heterostructures. *Solid State Commun.* **152**, 1275–1282 (2012).
32. Kim, K. K. et al. Synthesis and characterization of hexagonal boron nitride film as a dielectric layer for graphene devices. *ACS Nano.* **6**, 8583–8590 (2012).
33. Wang, D. & Sundararaman, R. Substrate effects on charged defects in two-dimensional materials. *Phys. Rev. Mater.* **3**, 083803 (2019).
34. Fan, Y., Liu, X., Wang, J., Ai, H. & Zhao, M. Silicene and germanene on InSe substrates: structures and tunable electronic properties. *Phys. Chem. Chem. Phys.* **20**, 11369–11377 (2018).
35. Yankowitz, M., Ma, Q., Jarillo-Herrero, P. & LeRoy, B. J. van der Waals heterostructures combining graphene and hexagonal boron nitride. *Nat. Rev. Phys.* **1**, 112–125 (2019).
36. Rashid, A., Lazarev, M., Kazeev, N., Novoselov, K. & Ustyuzhanin, A. Review on automated 2D material design. *2D Materials* **11**, 032002 (2024).
37. Sarikurt, S., Kocabaş, T. & Sevik, C. High-throughput computational screening of 2D materials for thermoelectrics. *J. Mater. Chem. A.* **8**, 19674–19683 (2020).
38. Zhang, X., Chen, A. & Zhou, Z. High-throughput computational screening of layered and two-dimensional materials. *Wiley. Interdiscip. Rev.: Comput. Mol. Sci.* **9**, e1385 (2019).
39. Mounet, N. et al. Two-dimensional materials from high-throughput computational exfoliation of experimentally known compounds. *Nat. Nanotechnol.* **13**, 246–252 (2018).

40. Zhuang, H. L. & Hennig, R. G. Computational discovery, characterization, and design of single-layer materials. *JOM.* **66**, 366–374 (2014).
41. Zhou, X. F. et al. Semimetallic two-dimensional boron allotrope with massless Dirac fermions. *Phys. Rev. Lett.* **112**, 085502 (2014).
42. Dong, B. et al. New two-dimensional phase of tin chalcogenides: candidates for high-performance thermoelectric materials. *Phys. Rev. Mater.* **3**, 013405 (2019).
43. Miao, N. et al. Computational prediction of boron-based MAX phases and MXene derivatives. *Chem. Mater.* **32**, 6947–6957 (2020).
44. Popov, Z. I. et al. Novel two-dimensional boron oxynitride predicted using the USPEX evolutionary algorithm. *Phys. Chem. Chem. Phys.* **23**, 26178–26184 (2021).
45. Zhou, X.-F. et al. Two-dimensional magnetic boron. *Phys. Rev. B.* **93**, 85406 (2016).
46. Gu, T., Luo, W. & Xiang, H. Prediction of two-dimensional materials by the global optimization approach. *Wiley. Interdiscip. Rev.: Comput. Mol. Sci.* **7**, e1295 (2017).
47. Borlido, P., Steigemann, C., Lathiotakis, N. N., Marques, M. A. & Botti, S. Structural prediction of two-dimensional materials under strain. *2D Mater.* **4**, 045009 (2017).
48. Wang, Y. et al. An effective structure prediction method for layered materials based on 2D particle swarm optimization algorithm. *J. Chem. Phys.* **137**, 224108 (2012).
49. Penev, E. S., Marzari, N. & Yakobson, B. I. Theoretical prediction of two-dimensional materials, behavior, and properties. *ACS Nano.* **15**, 5959–5976 (2021).
50. Revard, B. C., Tipton, W. W., Yesypenko, A. & Hennig, R. G. Grand-canonical evolutionary algorithm for the prediction of two-dimensional materials. *Phys. Rev. B.* **93**, 054117 (2016).
51. Dong, R., Song, Y., Siriwardane, E. M. & Hu, J. Discovery of 2D materials using transformer network-based generative design. *Adv. Intell. Syst.* **5**, 2300141 (2023).
52. Zeni, C. et al. MatterGen: a generative model for inorganic materials design. Preprint at https://arxiv.org/abs/2312.03687v2 (2023).
53. Song, Y., Siriwardane, E. M., Zhao, Y. & Hu, J. Computational discovery of new 2D materials using deep learning generative models. *ACS Appl. Mater. Interfaces*. **13**, 53303–53313 (2021).
54. Agarwal, A., Goverapet Srinivasan, S. & Rai, B. Data-driven discovery of 2d materials for solar water splitting. *Front. Mater.* **8**, 679269 (2021).
55. Singh, A. K., Zhuang, H. L. & Hennig, R. G. Ab initio synthesis of single-layer III-V materials. *Phys. Rev. B: Condens. Matter Mater. Phys.* **89**, 245431 (2014).
56. Oganov, A. R. & Glass, C. W. Crystal structure prediction using ab initio evolutionary techniques: principles and applications. *J. Chem. Phys.* **124**, 244704 (2006).
57. Oganov, A. R., Lyakhov, A. O. & Valle, M. How evolutionary crystal structure prediction works and why. *Acc. Chem. Res.* **44**, 227–237 (2011).
58. Lyakhov, A. O., Oganov, A. R., Stokes, H. T. & Zhu, Q. New developments in evolutionary structure prediction algorithm USPEX. *Comput. Phys. Commun.* **184**, 1172–1182 (2013).
59. Plimpton, S. Fast parallel algorithms for short-range molecular dynamics. *J. Comput. Phys.* **117**, 1–19 (1995).
60. Thompson, A. P. et al. LAMMPS—a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Comput. Phys. Commun.* **271**, 108171 (2022).
61. Shapeev, A. V. Moment tensor potentials: a class of systematically improvable interatomic potentials. **14**, 1153–1173 https://doi.org/10.1137/15M1054183 (2016).
62. Zuo, Y. et al. Performance and Cost Assessment of Machine Learning Interatomic Potentials. *J. Phys. Chem. A*. **124**, 731–745 (2020).
63. Lysogorskiy, Y. et al. Performant implementation of the atomic cluster expansion (PACE) and application to copper and silicon. *npj Comput. Mater.* **7**, 1–12 (2021).
64. Novikov, I. S., Gubaev, K., Podryabinkin, E. V. & Shapeev, A. V. The MLIP package: moment tensor potentials with MPI and active learning. *Mach. Learn.: Sci. Technol.* **2**, 025002 (2020).
65. Podryabinkin, E. et al. MLIP-3: Active learning on atomic environments with moment tensor potentials. *J. Chem. Phys.* **159**, 8 (2023).
66. Podryabinkin, E. V., Tikhonov, E. V., Shapeev, A. V. & Oganov, A. R. Accelerating crystal structure prediction by machine-learning interatomic potentials with active learning. *Phys. Rev. B.* **99**, 64114 (2019).
67. Gubaev, K., Podryabinkin, E. V., Hart, G. L. W. & Shapeev, A. V. Accelerating high-throughput searches for new alloys with active learning of interatomic potentials. *Computational Mater. Sci.* **156**, 148–156 (2019).
68. Kostiuchenko, T., Körmann, F., Neugebauer, J. & Shapeev, A. Impact of lattice relaxations on phase transitions in a high-entropy alloy studied by machine-learning potentials. *npj Computational Mater. 2019 5:1* **5**, 1–7 (2019).
69. Kostiuchenko, T., Ruban, A. V., Neugebauer, J., Shapeev, A. & Körmann, F. Short-range order in face-centered cubic VCoNi alloys. *Phys. Rev. Mater.* **4**, 113802 (2020).
70. Nyshadham, C. et al. Machine-learned multi-system surrogate models for materials prediction. *npj Comput. Mater.* **5**, 1–6 (2019).
71. Mortazavi, B. et al. Exploring phononic properties of two-dimensional materials using machine learning interatomic potentials. *Appl. Mater. Today.* **20**, 100685 (2020).
72. Mortazavi, B. et al. Machine-learning interatomic potentials enable first-principles multiscale modeling of lattice thermal conductivity in graphene/borophene heterostructures. *Mater. Horiz.* **7**, 2359–2367 (2020).
73. Mortazavi, B. et al. First-principles multiscale modeling of mechanical properties in graphene/borophene heterostructures empowered by machine-learning interatomic potentials. *Adv. Mater.* **33**, 2102807 (2021).
74. Oganov, A. R., Pickard, C. J., Zhu, Q. & Needs, R. J. Structure prediction drives materials discovery. *Nat. Rev. Mater.* **4**, 331–348 (2019).
75. Zhu, Q., Li, L., Oganov, A. R. & Allen, P. B. Evolutionary method for predicting surface reconstructions with variable stoichiometry. *Phys. Rev. B.* **87**, 195317 (2013).
76. Lepeshkin, S. V., Baturin, V. S., Uspenskii, Y. A. & Oganov, A. R. Method for simultaneous prediction of atomic structure and stability of nanoclusters in a wide area of compositions. *J. Phys. Chem. Lett.* **10**, 102–106 (2018).
77. Núñez-Valdez, M., Allahyari, Z., Fan, T. & Oganov, A. R. Efficient technique for computational design of thermoelectric materials. *Comput. Phys. Commun.* **222**, 152–157 (2018).
78. Kvashnin, A. G., Oganov, A. R., Samtsevich, A. I. & Allahyari, Z. Computational search for novel hard chromium-based materials. *J. Phys. Chem. Lett.* **8**, 755–764 (2017).
79. Bushlanov, P. V., Blatov, V. A. & Oganov, A. R. Topology-based crystal structure generator. *Comput. Phys. Commun.* **236**, 1–7 (2019).
80. Fredericks, S., Parrish, K., Sayre, D. & Zhu, Q. PyXtal: a Python library for crystal structure generation and symmetry analysis. *Comput. Phys. Commun.* **261**, 107810 (2021).
81. Dumcenco, D. et al. Large-area epitaxial monolayer MoS2. *ACS nano.* **9**, 4611–4620 (2015).
82. Cao, T. et al. Valley-selective circular dichroism of monolayer molybdenum disulphide. *Nat. Commun. 2012 3:1.* **3**, 1–5 (2012).
83. Ganatra, R. & Zhang, Q. Few-layer MoS2: a promising layered semiconductor. *ACS Nano.* **8**, 4074–4099 (2014).

84. Subbaiah, Y. P. V., Saji, K. J. & Tiwari, A. Atomically thin MoS2: a versatile nongraphene 2D material. *Adv. Funct. Mater.* **26**, 2046–2069 (2016).
85. Kruglov, I. A. et al. Crystal structure prediction at finite temperatures. *npj Comput. Mater. 2023 9:1.* **9**, 1–8 (2023).
86. Sukhanova, E. V. et al. 2D-Mo3S4 phase as promising contact for MoS2. *Appl. Surf. Sci.* **589**, 152971 (2022).
87. Ryou, J., Kim, Y. S., Santosh, K. C. & Cho, K. Monolayer MoS_2 bandgap modulation by dielectric environments and tunable bandgap transistors. *Sci. Rep. 2016 6:1.* **6**, 1–8 (2016).
88. Splendiani, A. et al. Emerging photoluminescence in monolayer MoS2. *Nano Lett.* **10**, 1271–1275 (2010).
89. Ramasubramaniam, A. Large excitonic effects in monolayers of molybdenum and tungsten dichalcogenides. *Phys. Rev. B - Condens. Matter Mater. Phys.* **86**, 115409 (2012).
90. Komsa, H. P. & Krasheninnikov, A. V. Effects of confinement and environment on the electronic structure and exciton binding energy of MoS2 from first principles. *Phys. Rev. B : Condens. Matter Mater. Phys.* **86**, 241201 (2012).
91. Shi, H., Pan, H., Zhang, Y. W. & Yakobson, B. I. Quasiparticle band structures and optical properties of strained monolayer MoS2 and WS2. *Phys. Rev. B - Condens. Matter Mater. Phys.* **87**, 155304 (2013).
92. Zhu, D. et al. Capture the growth kinetics of CVD growth of two-dimensional MoS2. *npj 2D Mater. Appl.* **1**, 1–8 (2017).
93. Song, J. G. et al. Layer-controlled, wafer-scale, and conformal synthesis of tungsten disulfide nanosheets using atomic layer deposition. *ACS Nano.* **7**, 11333–11340 (2013).
94. Tan, L. K. et al. Atomic layer deposition of a MoS2 film. *Nanoscale.* **6**, 10584–10588 (2014).
95. Jin, Z., Shin, S., Kwon, D. H., Han, S. J. & Min, Y. S. Novel chemical route for atomic layer deposition of MoS_2 thin film on SiO_2/Si substrate. *Nanoscale.* **6**, 14453–14458 (2014).
96. Kong, D. et al. Synthesis of MoS2 and MoSe2 films with vertically aligned layers. *Nano Lett.* **13**, 1341–1347 (2013).
97. Laskar, M. R. et al. Large area single crystal (0001) oriented MoS2. *Appl. Phys. Lett.* **102** (2013).
98. Lee, Y. et al. Synthesis of wafer-scale uniform molybdenum disulfide films with control over the layer number using a gas phase sulfur precursor. *Nanoscale.* **6**, 2821–2826 (2014).
99. Liu, H. F., Wong, S. L. & Chi, D. Z. CVD growth of MoS2-based two-dimensional materials. *Chem. Vap. Depos.* **21**, 241–259 (2015).
100. Reuter, K. Ab initio thermodynamics and first-principles microkinetics for surface catalysis. *Catal. Lett.* **146**, 541–563 (2016).
101. Rau, H., Kutty, T. R. & Guedes De Carvalho, J. R. Thermodynamics of sulphur vapour. *J. Chem. Thermodyn.* **5**, 833–844 (1973).
102. Kresse, G. & Furthmüller, J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mater. Sci.* **6**, 15–50 (1996).
103. Kresse, G. & Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B.* **54**, 11169 (1996).
104. Kresse, G. & Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B.* **59**, 1758 (1999).
105. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865 (1996).
106. Ganose, A. et al. atomate2 https://github.com/materialsproject/atomate2 (2024).
107. Hinuma, Y., Pizzi, G., Kumagai, Y., Oba, F. & Tanaka, I. Band structure diagram paths based on crystallography. *Comput. Mater. Sci.* **128**, 140–184 (2017).
108. Perdew, J. P. et al. Restoring the density-gradient expansion for exchange in solids and surfaces. *Phys. Rev. Lett.* **100**, 136406 (2008).
109. Togo, A. & Tanaka, I. First principles phonon calculations in materials science. *Scr. Mater.* **108**, 1–5 (2015).
110. Togo, A. First-principles phonon calculations with phonopy and Phono3py. **92** https://doi.org/10.7566/JPSJ.92.012001 (2022).
111. Togo, A., Chaput, L., Tadano, T. & Tanaka, I. Implementation strategies in phonopy and phono3py. *J. Phys.: Condens. Matter.* **35**, 353001 (2023).
112. Zur, A. & McGill, T. C. Lattice match: an application to heteroepitaxy. *J. Appl. Phys.* **55**, 378–386 (1984).
113. Ong, S. P. et al. Python Materials Genomics (pymatgen): a robust, open-source Python library for materials analysis. *Computational Mater. Sci.* **68**, 314–319 (2013).
114. Mazitov, A. et al. Substrate-aware computational design of two-dimensional materials. https://doi.org/10.24435/materialscloud:8q-a1 (2024).

Acknowledgements

I.A.K. thanks grant RSF No. 24-73-10055 for financial support of the development of ML interatomic potential for Mo-S system and calculation of dynamical properties of newly found materials. A.V.A. is supported by the Ministry of Science and Higher Education (FSMG-2025-0005). D.G.K. acknowledges financial support from the federal budget of the Russian Ministry of Science and Higher Education (No.125020401357-4). A.R.O. gratefully acknowledges support from Russian Science Foundation (grant 24-43-00162). K.S.N acknowledges support from the Ministry of Education, Singapore under Research Centre of Excellence award to the Institute for Functional Intelligent Materials, I-FIM (project No. EDUNC-33-18-279-V12) and under the Tier 3 program (MOE-MOET32024-0001), as well as by the National Research Foundation, Singapore under its AI Singapore Programme (AISG Award No: AISG3-RP-2022-028).

Author contributions

A.M. implemented the ASCT algorithm and the extension for the USPEX code for predicting the 2D layers on substrates, performed the training of the MLIP for Mo–S/Al2O3 system and prediction of the stable 2D Mo–S structures on Al2O3 substrate, calculated the phonon and electronic band structures, and performed the thermochemical stability analysis. I.A.K. helped with developing the MLIP for Mo-S system and calculated the anharmonic phonon density of states. A.V.Y., D.G.K., and A.R.O. directed the development of the USPEX code extension and advised on the thermochemical calculations. A.V.A., V.S.V., and K.S.N. guided the project and assisted in the analysis and interpretation of the results. All authors contributed to the preparation of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41524-025-01754-8.

Correspondence and requests for materials should be addressed to Arslan Mazitov.

Reprints and permissions information is available at http://www.nature.com/reprints