



# Automatic identification of slip pathways in ductile inorganic materials by combining the active learning strategy and NEB method



Jun Luo <sup>1,2</sup>, Tao Fan<sup>1</sup>, Jiawei Zhang <sup>1,2</sup>✉, Pengfei Qiu<sup>1,2</sup>, Xun Shi <sup>1,2</sup>✉ & Lidong Chen<sup>1,2</sup>

Ductile inorganic semiconductors have recently received considerable attention due to their metal-like mechanical properties and potential applications in flexible electronics. However, the accurate determination of slip pathways, crucial for understanding the deformation mechanism, still poses a great challenge owing to the complex crystal structures of these materials. In this study, we propose an automated workflow based on the interlayer slip potential energy surface to identify slip pathways in complex inorganic systems. Our computational approach consists of two key stages: first, an active learning strategy is utilized to efficiently and accurately model the interlayer slip potential energy surfaces; second, the climbing image nudged elastic band method is employed to identify minimum energy pathways, followed by comparative analysis to determine the final slip pathway. We discuss the validity of our selected feature vectors and models across various material systems and confirm that our approach demonstrates robust effectiveness in several case studies with both simple and complicated slip pathways. Our automated workflow opens a new avenue for the automatic identification of the slip pathways in inorganic materials, which holds promise for accelerating the high-throughput screening of ductile inorganic materials.

Recently discovered ductile inorganic non-metallic materials have attracted widespread attention due to their significant potential in flexible electronics applications<sup>1–5</sup>. Despite the promise of these materials, the number of identified ductile inorganic semiconductors remains limited, with only a handful having been experimentally and theoretically characterized<sup>6–10</sup>. The scarcity of ductile inorganic semiconductors has drawn significant attention to the need for systematic exploration and discovery of new inorganic materials with desirable mechanical properties, where high-throughput computational screening has emerged as a powerful strategy<sup>11</sup>. By high-throughput screening with first-principles calculations, one recent attempt by Gao et al. has been successful in identifying several potential candidates with plastic deformability from binary 2D van der Waals crystals<sup>12</sup>. In the high-throughput screening process, the most critical aspect, especially when it comes to understanding and predicting ductile behavior in inorganic materials, is the calculation of slip energy<sup>6,13</sup>. The key to calculating slip energy lies in identifying the optimal slip pathway. In Gao's work, the

minimum energy slip pathway and the corresponding energy barrier were determined manually, which becomes inefficient for inorganic materials with complex slip pathways. Overall, to date, there is still a lack of an automated workflow for the determination of slip pathways or slip systems in complex ductile inorganic non-metallic materials.

Slip is the fundamental mechanism of plastic deformation in crystalline materials. When a material undergoes mechanical stress, atoms in the crystal lattice move relative to one another along specific crystallographic planes, known as slip planes. The combination of a slip direction and plane is referred to as a “slip system”<sup>14,15</sup>. The energy required to drive this movement is termed the generalized stacking fault energy (GSFE)<sup>16</sup>, or simply, the slip energy. Calculating the slip energy provides insight into how easily a material deforms plastically and, by extension, its ductility. For brittle materials, the energy barriers are often too high to allow significant slip, leading to fracture instead of plastic deformation. Conversely, in ductile materials, the slip pathways offer low

<sup>1</sup>State Key Laboratory of High Performance Ceramics and Superfine Microstructure, Shanghai Institute of Ceramics, Chinese Academy of Sciences, 200050 Shanghai, China. <sup>2</sup>Center of Materials Science and Optoelectronics Engineering, University of Chinese Academy of Sciences, 100049 Beijing, China.

✉ e-mail: [jiaweizhang@mail.sic.ac.cn](mailto:jiaweizhang@mail.sic.ac.cn); [xshi@mail.sic.ac.cn](mailto:xshi@mail.sic.ac.cn)

energy barriers, allowing atoms to slide past each other more readily under stress. Usually, the more slip systems a material has, the easier it undergoes slip, leading to better ductility.<sup>17</sup>

In metals, due to the relatively simple arrangement of atoms, the potential slip systems have been identified<sup>18</sup>. Digital Image Correlation (DIC) combined with Scanning Electron Microscope (SEM) technology has been successfully applied in the automated identification of activated slip systems during the process of plastic deformation<sup>19–23</sup>. However, the complex elemental compositions and atomic arrangements complicate the slip pathways in inorganic materials, rendering this method ineffective. Compared to traditional experimental methods, energy-based approaches offer an alternative and effective means of determining slip systems<sup>24</sup>. The grown density functional theory (DFT)<sup>25,26</sup> has greatly facilitated the construction of slip surfaces<sup>27–29</sup>. Furthermore, the advent of machine learning (ML) has reduced the costs and accelerated the speed of the process<sup>30–33</sup>. Studies have shown that slip systems typically represent minimum energy pathways (MEPs) upon the slip surface<sup>34</sup>. Finding MEPs is a common and important problem in theoretical chemistry. Many different methods have been presented to find MEPs<sup>35,36</sup>. Among these, the nudged elastic band (NEB) method, which has been extensively employed, is robust, efficient, and easily parallelizable<sup>37,38</sup>. Hence, the combination of ML and NEB offers a potential

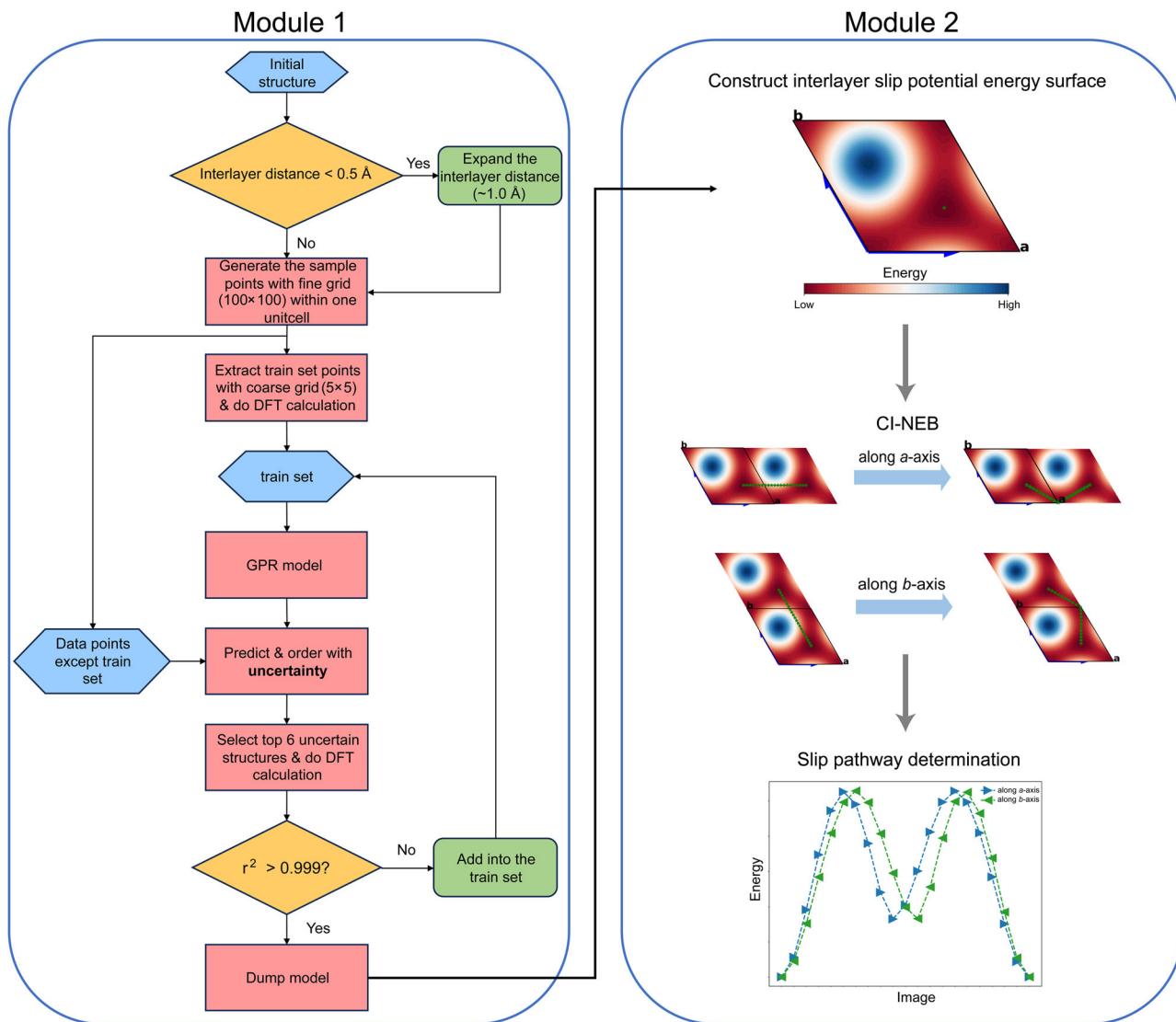
solution to the automatic identification of the slip pathways in complex inorganic materials.

In this paper, for complex inorganic material systems, we propose a workflow based on interlayer slip potential energy surfaces to automatically and rapidly identify potential slip pathways. Active learning strategies and the climbing image nudged elastic band (CI-NEB) method are utilized in the workflow. Our model and workflow have been tested in several archetypical materials. The results confirm the reliability of our workflow. Our work solves the problem of identifying slip pathways in complex inorganic materials and offers valuable support for the further high-throughput screening of ductile inorganic materials.

## Results

### Workflow

Figure 1 shows the schematic of our workflow, which is based on the implementation of “interlayer slip potential energy” (ISPE). While similar concepts have been reported in other literature<sup>39–42</sup>, our work uniquely incorporates ISPE. In a perfect crystal, the rigid slip of one group of atoms relative to another leads to variations in the system’s energy. We define ISPE as the energy computed from the post-slip structural configuration. Our methodology based on ISPE involves two modules. Module 1 implements



**Fig. 1 | The schematic of the workflow for the automated identification of slip pathways.** This workflow consists of two modules: Module 1 primarily explains the training of interlayer slip potential energy surfaces using an active learning strategy,

while Module 2 focuses on identifying slip pathways based on the interlayer slip potential energy surfaces through the CI-NEB method.

an active learning strategy that effectively samples the slip plane and learns the ISPE surfaces. Module 2 leverages the trained ISPE surfaces to automatically detect potential slip pathways using the CI-NEB method, ultimately selecting the final pathway based on the energy profile. The following sections will provide a comprehensive explanation of our process.

In Module 1, the initial structure must be a layered configuration capable of slip. Considering the definition of the slip system, we constrain the slipping to occur within a single plane, disregarding any slipping perpendicular to the plane. To simplify subsequent calculations, we fix the slip plane within the *ab* plane of the unit cell, with the *c*-axis oriented perpendicular to the *ab* plane. Additionally, to prevent atoms from becoming too close during the slip process, which could lead to unrealistic structures, we perform an initial check. If any slip configurations result in interlayer atomic distances that are less than 0.5 Å, we will expand the interlayer distance by inserting a vacuum layer with a length of around 1.0 Å between the layers before performing the slip process. The rationalization for this operation is discussed in detail in the Supplementary Materials.

Once the initial slip structure is prepared, we uniformly sample the slip plane along the *ab* axes. To obtain a sufficiently fine grid of sampling points, we divide the slip plane into  $100 \times 100$  sampling points, partitioning it into 99 segments along both the *a* and *b* axes. A subset of these sampling points is then extracted for DFT static calculations using a coarser grid, such as  $5 \times 5$ , to generate the initial training set. Based on this initial training set, we train a Gaussian process regression (GPR) model to produce a preliminary model.

Using this preliminary model, we predict the ISPE at the remaining sampling points, excluding those in the training set, and calculate the uncertainty at each point. The points with the highest uncertainties are ranked, and the 6 points with the greatest uncertainty are selected for further DFT static calculations to determine their energies. These points form the test set, and the model is evaluated using the  $r^2$  score. If the score is below the predefined threshold (such as 0.999), the test points are added to the training set, the model is retrained, and new high-uncertainty points are selected for the next iteration. This process is repeated until the model converges. During this process, we actively learn the ISPE surfaces based on uncertainty, aiming to achieve a high-quality model using minimal computational resources.

In Module 2, based on the trained ISPE surface model, our first task is to locate the global minimum energy point on the ISPE surface. In most cases, the initial point of the slip process corresponds to a local energy minimum on the ISPE surface, but it is not necessarily the global minimum. This depends on the specific system. For example, in the case of InSe with the space group  $P6_3/mmc$ , we found that the global energy minimum on its ISPE surface does not coincide with the initial position. A recent study has reported similar findings<sup>43</sup>. To address this issue, we have implemented a two-step procedure in our code to accurately locate the global energy minimum. This involves first using a trained model to roughly identify an initial point, followed by optimization from this point to ensure the global energy minimum is found.

The initial state and final state of the slip process is then fixed at the global minimum energy points of the neighboring unit cells. We then explore possible slip pathways in different directions. Specifically, we search within neighboring unit cells along both the *a*-axis and *b*-axis. By using linear interpolation, two initial pathways are generated. These two pathways are then refined using the CI-NEB method to identify the two MEPs.

The NEB is a chain-of-state method. The process of projecting the force components perpendicular to the path while allowing the spring force to act only along the path is known as “nudging”<sup>37</sup>. The force on the *i*th image is

$$\vec{F}_i = -\vec{\nabla}V(\vec{R}_i)|_{\perp} + \vec{F}_i^{\parallel} \cdot \hat{\tau}_{\parallel} \hat{\tau}_{\parallel}$$

Where  $\vec{R}_i$  denotes as the position vector of the *i*th image in the scalar field  $V$ ,  $\hat{\tau}_{\parallel}$  is the unit tangent vector to the path and  $\vec{\nabla}V(\vec{R}_i)|_{\perp}$

$= \vec{\nabla}V(\vec{R}_i) - \vec{\nabla}V(\vec{R}_i) \cdot \hat{\tau}_{\parallel} \hat{\tau}_{\parallel}$ . This decoupling of the path relaxation from its discrete representation is critical for ensuring convergence to the MEP<sup>44</sup>. The CI-NEB method constitutes a small modification to the NEB method<sup>45</sup>. Specifically, after a few iterations with the regular NEB, the full force of the image with the highest energy is expressed as:

$$\vec{F}_{i_{\max}} = -\vec{\nabla}V(\vec{R}_{i_{\max}}) + 2\vec{\nabla}V(\vec{R}_{i_{\max}}) \cdot \hat{\tau}_{\parallel} \hat{\tau}_{\parallel}$$

The subscript  $i_{\max}$  refers to the index of the image along the path that corresponds to the highest energy. In this way, information about the shape of the MEP is retained, but a rigorous convergence to a saddle point is also obtained. This additional feature does not add any significant computational effort.

Since we are using the CI-NEB method, the choice of spring constants and the number of images is not particularly stringent. To further ensure the global minimum energy pathway, we compare the energy barriers along each of the two MEPs and select the path with the lower energy barrier as the final slip pathway. The structure corresponding to the energy barrier along this pathway is used to compute the maximum slip energy of this material.

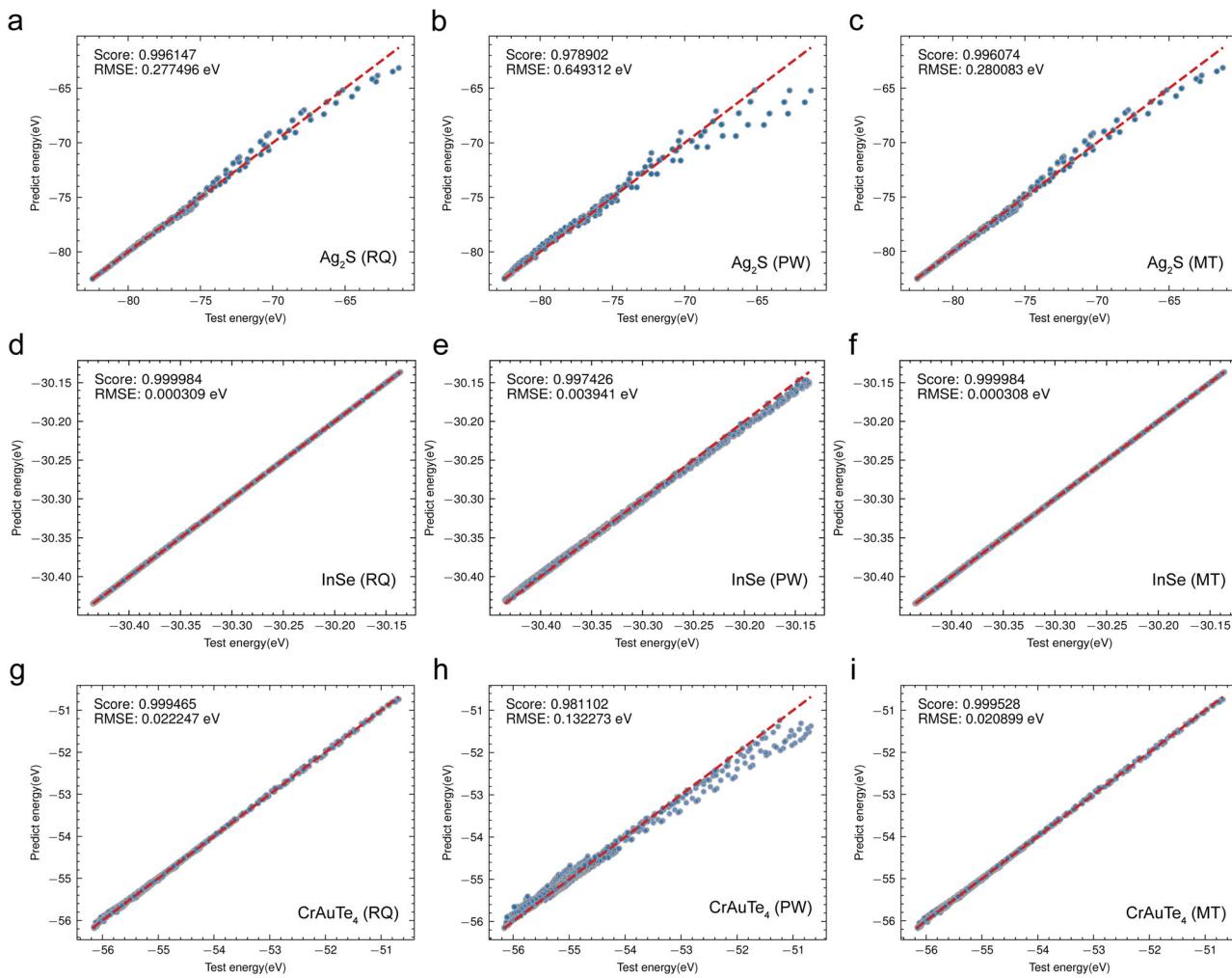
## Model training

**Feature vector.** For the active learn algorithms to learn the ISPE related to their corresponding slip structures, a numerical representation of the slip structures is required, as opposed to the character strings we normally use to identify different slip structures. Since our focus is on ISPE, we propose using the fractional slip vector (FSV) of the crystal as the preferred numerical representation of the slip structures. The concept of FSV has been widely employed in visualizing the stacking energy surface of van der Waals layers. It bears resemblance to the well-established concept of the “Burgers vector”<sup>46–49</sup>. However, compared to “Burgers vector”, FSV focuses on the interlayer slip behaviors within a single unit cell, and is derived from the lattice vector, denoted as  $\vec{a}$ ,  $\vec{b}$  and  $\vec{c}$ . For example, the FSV of  $(f_a, f_b, f_c)$  means that a portion of the atoms in the unit cell slips a distance of  $f_a \cdot |\vec{a}|$  along the *a*-axis,  $f_b \cdot |\vec{b}|$  along the *b*-axis, and  $f_c \cdot |\vec{c}|$  along the *c*-axis, relative to another portion of atoms. Since the slip process is constrained to the *ab* plane, a two-dimensional FSV of  $(f_a, f_b)$  is employed as the feature vector. However, in specific scenarios where the energy range of the ISPE surfaces exceeds 40 eV, we have observed that the FSV becomes ineffective. Therefore, we have also developed two additional feature vectors to address these exceptional cases. See the discussion and Fig. S1 in Supplementary Materials for details.

**ML model.** After extensive testing and a thorough literature review<sup>50,51</sup>, we find that the GPR model, with the advantages of uncertainty quantification, flexibility for small datasets, smoothness, and robustness<sup>52,53</sup>, is the most suitable for the active learning of ISPE. However, the performance of the GPR model is highly correlated with the kernel functions. In this study, we have tested the performance of the model using the Rational Quadratic (RQ) kernel<sup>54</sup>, Pairwise (PW) kernel, and Matérn (MT) kernel<sup>55</sup>.

Figure 2 shows our results. DFT static calculations are initially performed in a dense  $51 \times 51$  grid across the entire slip surface to obtain ISPEs for the test set. Subsequently, a uniform  $11 \times 11$  subset from this grid is extracted to train the model. Model testing is conducted on the  $\text{Ag}_2\text{S}$ , InSe, and  $\text{CrAuTe}_4$  systems. The energy span for the  $\text{Ag}_2\text{S}$  system is approximately 20 eV, about 0.5 eV for InSe, and around 5 eV for  $\text{CuAuTe}_4$ . The results indicate that, under the same kernel function, as the energy span of the system increases, the model’s performance deteriorates, with a gradual decrease in the  $r^2$  score and an increase in the root mean squared error (RMSE).

Furthermore, we evaluated three different kernel functions for each material system. The results demonstrate that, regardless of the system, the MT kernel consistently performs the best. The RQ kernel exhibits slightly



**Fig. 2 | Predicted ISPEs within the GPR model using different kernel functions.** The correlation between the DFT calculating ISPEs (Test energy) and model predicting ISPEs (Predict energy) of Ag<sub>2</sub>S (a–c), InSe (d–f), and CrAuTe<sub>4</sub> (g–i), the

corresponding kernel functions respectively are Rational Quadratic kernel, Pairwise kernel using Laplacian metric and Matérn kernel with  $\nu = 2.5$ . 51 × 51 dense ISPE surface for Ag<sub>2</sub>S, InSe, and CrAuTe<sub>4</sub> is used for the test.

lower performance compared to the MT kernel but remains competitive, while the PW kernel shows the lowest performance. Overall, the MT and RQ kernels perform at a comparable level. Notably, the parameter  $\nu$  in the MT kernel is crucial; our detailed test reveals that a  $\nu$  value of approximately 2.5 optimizes model performance. In summary, for various tested systems, the selected feature vectors and kernel functions consistently yield model scores exceeding 0.99, demonstrating the reliability and effectiveness of our model. We have discussed the influence of these three kernel functions on CI-NEB results in the supplementary materials.

### Case studies

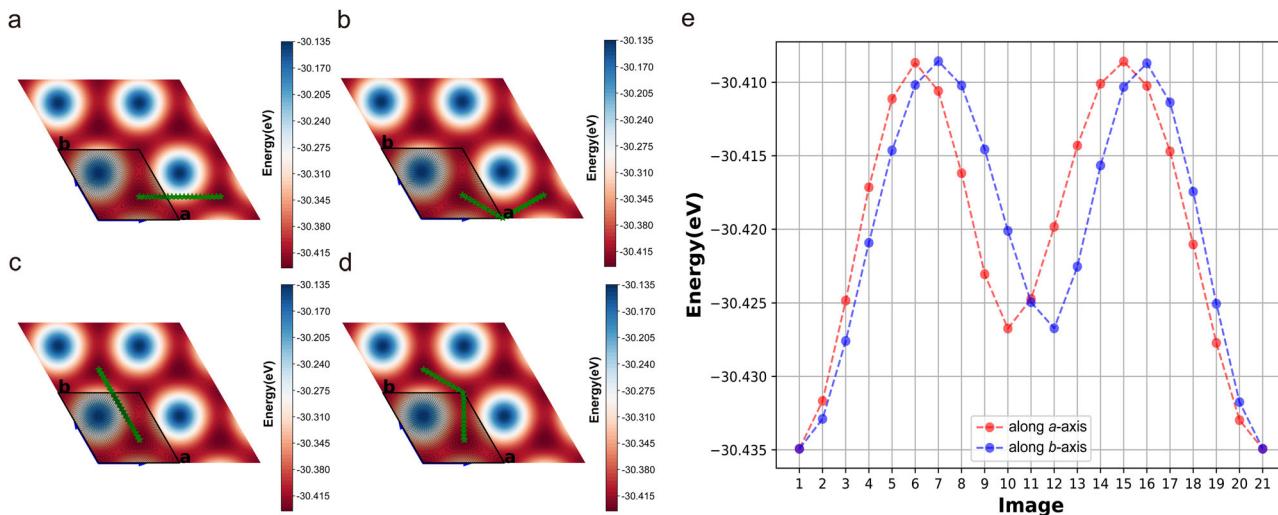
To validate our workflow, we first apply it to the InSe system with a relatively simple slip pathway. The algorithm for Module 1 is deployed on the AiiDA platform<sup>56,57</sup>, where calculations are automatically submitted and managed through AiiDA for model training. Thanks to the integration of active learning algorithms, we are able to efficiently select sampling points across the ISPE surface, resulting in rapid convergence of the calculations. Subsequently, we test the model using the pre-calculated 51 × 51 grid points. The results show an  $r^2$  of 0.999998 and an RMSE of  $9.4 \times 10^{-5}$  eV, indicating that the final trained model performed exceptionally well.

Building on the strategy from Module 2, we then conduct CI-NEB simulations based on the learned ISPE surface. Figure 3 presents our results. In this case, the initial pathways are generated by evenly interpolating a straight line between the two lowest energy points in adjacent cells along

both the  $a$ -axis and  $b$ -axis directions. Figure 3a, c shows the initial pathways along the  $a$ -axis and  $b$ -axis. Figure 3b, d shows the identified pathways after CI-NEB running, which locate MEPs correctly. Figure 3e shows the energy of each image along the MEPs. The resulting energy profiles indicate that the energy barriers along both pathways are identical, confirming that the two pathways are completely equivalent. From these curves, we can then identify the location of the energy barrier during the slip process, which corresponds to the 6th image along the pathway of the  $a$ -axis, representing the FSV at (0.844, 0.156).

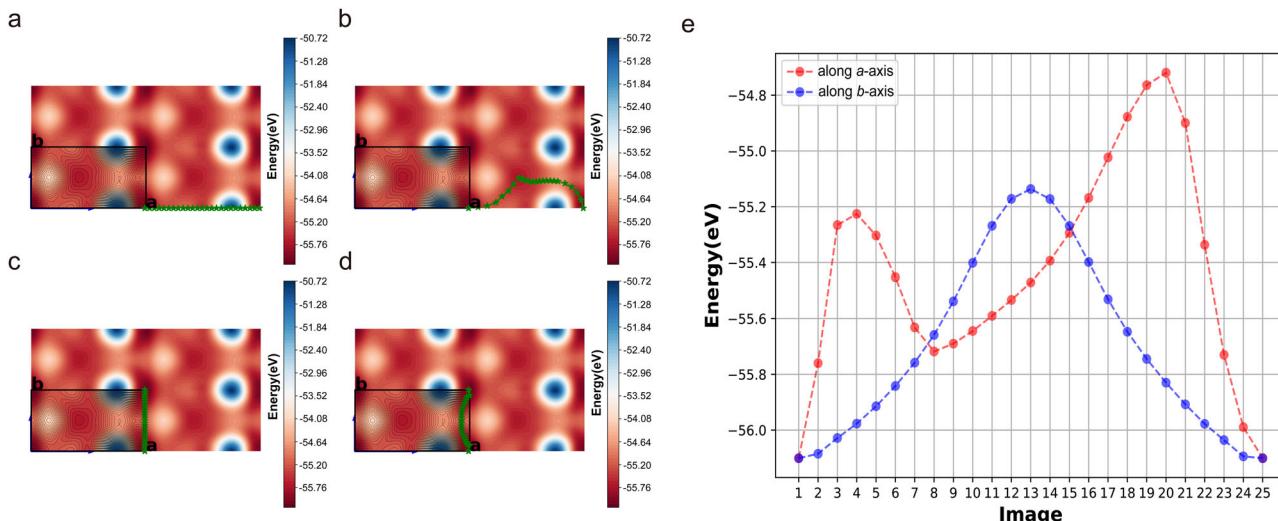
Furthermore, we have tested our workflow on the CrAuTe<sub>4</sub> system. Compared to InSe, CrAuTe<sub>4</sub> has a more complex interlayer structure, leading to a more intricate slip plane, and its ISPE surface operates on a larger energy span than that of InSe. However, thanks to the efficient active learning strategy, the calculations converged after approximately 20 iterations. We then tested the trained model using the previously computed 51 × 51 grid points, obtaining an  $r^2$  of 0.9972 and an RMSE of 0.050 eV, indicating a reasonably high accuracy of the model.

Subsequently, CI-NEB simulations are performed utilizing the learned ISPE surface model, with search directions along both the  $a$ -axis and  $b$ -axis. Figure 4 displays our findings. Figure 4a, c illustrates the initial pathway configurations analogous to the InSe system. Figure 4b, d illustrates the slip pathways, or MEPs, obtained from the simulations, corresponding to the  $a$ -axis and  $b$ -axis, respectively. We can observe that, in contrast to the InSe system, the slip pathways in this system are notably more complex, no



**Fig. 3 | Identification of slip pathways for InSe.** The initial pathway setups along the *a*-axis (a) and *b*-axis (c). The identified pathways after running the CI-NEB simulations along the *a*-axis direction (b) and *b*-axis direction (d), respectively. The

filled contours represent the intralayer slip potential surface trained using the active learning strategy we proposed. e The energy as a function of the image index along the slip pathways of *a*-axis direction (red) and *b*-axis direction (blue) for InSe.



**Fig. 4 | Identification of slip pathways for complex CrAuTe<sub>4</sub>.** The initial pathway setups along the *a*-axis (a) and *b*-axis (c). The identified non-linear pathways after running the CI-NEB simulations along the *a*-axis direction (b) and *b*-axis direction (d), respectively. The filled contours represent the intralayer slip potential surface

trained using the active learning strategy. e The energy as a function of the image index along the slip pathways of *a*-axis direction (red) and *b*-axis direction (blue) for CrAuTe<sub>4</sub>.

longer following a simple linear trajectory. Figure 4e presents the energy associated with each image along the slip pathways, highlighting a notable difference in energy barriers. The pathway along the *a*-axis exhibits a considerably higher energy barrier than that along the *b*-axis. Consequently, we identify the MEP along the *b*-axis as the optimal slip pathway, with the 13th image located at the position of maximum slip energy, corresponding to the FSV at (0.928, 0.496).

It is worth noting that, due to the complexity of the slip plane, Fig. 4b indicates that a potential slip pathway may have been missed, located on the opposite side of the identified pathway. To resolve this, additional initial pathways should be set to ensure that all possible slip pathways are captured. See Fig. S2 in supplementary materials for our test results.

## Discussion

In summary, we have developed a robust automated workflow for the identification of slip pathways in ductile inorganic materials, successfully

integrating an active learning strategy with the CI-NEB method. Our workflow comprises two key modules: the first leverages uncertainty-based sampling to construct the ISPEs within the unit cell, while the second employs the ISPE surfaces to perform the slip pathway identification using the CI-NEB method. This innovative approach allows for the rapid and efficient identification of MEPs and facilitates the determination of slip structures associated with varying energy barriers.

The performance of our machine learning descriptors and models was rigorously evaluated across several material systems, including Ag<sub>2</sub>S, InSe, and CrAuTe<sub>4</sub>. Notably, the ML model demonstrated exceptional sensitivity to the energy span of the ISPE surfaces, maintaining a high degree of accuracy ( $r^2 > 0.99$ ) for spans within 20 eV. Our kernel function analysis further reinforced the reliability of our approach, with both the RQ and MT kernels yielding satisfactory fitting results.

The comprehensive testing of our computational workflow has been conducted on InSe and CrAuTe<sub>4</sub> systems with simple and complicated slip

pathways, respectively, confirming its efficacy and reliability in identifying slip pathways. The outcomes highlight the potential of our methodology in advancing the understanding of slip mechanisms in ductile inorganic materials. This work paves the way for future high-throughput screening research to explore new inorganic non-metal materials with plastic deformability.

## Methods

### ML model test

GPR is employed as our ML model. It is a non-parametric model that utilizes a Gaussian Process (GP) prior for regression analysis of data. A GP is characterized as a collection of random variables, where any finite subset exhibits a multivariate normal distribution. In GPR, we initially define a prior GP that encapsulates our initial comprehension of the shape of the function. Upon observing the data, we update this prior through Bayesian inference to obtain a posterior distribution, which reflects our comprehension of the function given the known data. Assuming we have a set of training data  $(X, y)$ , where  $X$  represents the feature vector and  $y$  denotes the corresponding target variables, follows GP:

$$y \sim \mathcal{GP}(\mu(X), \sigma(X, X'))$$

Where  $\mu(X)$  represents the mean function, while  $\sigma(X, X')$  denotes the kernel function (or covariance function). Once the distribution is established, to predict new input  $X^*$ , we can obtain the posterior distribution through Bayesian inference:

$$y^* | X^*, X, y \sim \mathcal{N}(\mu^*, \Sigma^*)$$

Where  $\mu^*$  represents the predicted mean value, and  $\Sigma^*$  denotes the predicted variance value. The predicted  $\mu^*$  provides an estimate for the new input data, while the  $\Sigma^*$  quantifies the uncertainty associated with this prediction. A larger  $\Sigma^*$  indicates lower confidence in the prediction, whereas a smaller variance signifies higher confidence. Our active learning process is precisely based on this uncertainty.

For the construction of the ML model, two-dimensional FSV was selected as  $X$ , and the ISPE was selected as  $y$ . GPR algorithm provided by scikit-learn package<sup>58</sup> was used. Before training, the built-in regularization method of the algorithm was applied to regularize the ISPEs.

To evaluate the kernels, we computed the ISPE surfaces within  $51 \times 51$  grid points for the  $\text{Ag}_2\text{S}$ , InSe, and  $\text{CrAuTe}_4$  systems, respectively. For  $\text{Ag}_2\text{S}$ , we selected a two-layer zigzag slab structure containing 24 atoms, the construction of which has been detailed in our previous work<sup>13</sup>. Given that, in certain slip configurations, the interlayer atomic distances could become extremely short, causing a sharp increase in system energy and model failure, we inserted a vacuum layer with a length of 1.0 Å between the two atomic layers and used this modified structure as our initial configuration. To obtain slip structures for the calculations of ISPE, atoms in the bottom layer were fixed, while atoms in the top layer displaced along the  $a$ -axis and  $b$ -axis in a certain displacement within the unit cell. We uniformly divided the displacements along the  $a$ -axis and  $b$ -axis of the unit cell into 50 equal segments, thereby constructing a  $51 \times 51$  grid. Pymatgen package<sup>59</sup> was used to generate the slip structures for each grid point. For InSe and  $\text{CrAuTe}_4$ , a two-layer slab structure was constructed based on the structure from The Materials Project (mp-20485 and mp-12743)<sup>60</sup>. The way to construct slip structures was the same as  $\text{Ag}_2\text{S}$ .

For these slip structures, DFT static calculations were performed to obtain sufficient data, which forms our dataset. Vienna Ab initio Simulation Package (VASP) with a plane-wave basis set and projector-augmented wave (PAW)<sup>61,62</sup> pseudopotentials were used. For the three systems, the exchange-correlation functional was approximated by the Perdew–Burke–Ernzerhof (PBE) formulation of the generalized gradient approximation (GGA)<sup>63</sup>. In these calculations, van der Waals corrections were considered using the DFT-D3 method with Becke-Johnson damping function<sup>64</sup>. For  $\text{Ag}_2\text{S}$ , InSe, and  $\text{CrAuTe}_4$ , the plane-wave cutoff energy used in the calculations for all was 400 eV, and a gamma-centered k-point with the mesh of  $3 \times 3 \times 1$ ,  $11 \times 11 \times 1$  and  $9 \times 5 \times 1$  was utilized, respectively.

The calculated  $51 \times 51$  ISPEs formed our test set. We uniformly extracted  $11 \times 11$  ISPEs from the test set to create the training set for GPR model training. We initially experimented with a variety of kernel functions; however, only three of them provided relatively reasonable results. Therefore, we focused our testing on these three models. They were the Rational Quadratic kernel, Pairwise kernel, and Matérn kernel.

### Setups for the case studies

For InSe, we utilized the AiiDA platform<sup>56,57</sup> in combination with an active learning strategy to train the ISPE surface. The entire slip plane was divided into a  $100 \times 100$  grid. A sparse uniform  $5 \times 5$  grid of points was chosen for DFT calculations, forming the initial training set of the model. During the active learning process, the 6 grid points with the highest uncertainty were selected for each new round of DFT calculations. Throughout the iterations, we set a convergence threshold of  $r^2 = 0.999$ . For the CI-NEB simulation, the method built in ASE<sup>65</sup> was employed. The  $k$  of springs was set 1.0 eV/Å<sup>2</sup>, and 21 images were used to identify MEP. The MDMin algorithm built in the software was used to optimize the pathways. In the optimization process, we set the maximum number of optimization steps to 3000. The optimization was terminated when the maximum force on these images fell below 0.01 eV/Å, at which point the results were output. For  $\text{CrAuTe}_4$ , the parameters remained the same as InSe except for the use of 25 images.

### Data availability

The data that support the findings of this study are available from the corresponding authors upon reasonable request.

### Code availability

The code of our workflow is available in the compressed file in the supplementary information.

Received: 8 October 2024; Accepted: 25 December 2024;

Published online: 19 February 2025

## References

1. Yang, Q. Y. et al. Flexible thermoelectrics based on ductile semiconductors. *Science* **377**, 854–858 (2022).
2. Zhu, Y. et al. High performance full-inorganic flexible memristor with combined resistance-switching. *ACS Appl. Mater. Inter.* **14**, 21173–21180 (2022).
3. Zhao, X.-F. et al. A fully flexible intelligent thermal touch panel based on intrinsically plastic  $\text{Ag}_2\text{S}$  semiconductor. *Adv. Mater.* **34**, 2107479 (2022).
4. Chen, H. Y. et al. Room-temperature plastic inorganic semiconductors for flexible and deformable electronics. *InfoMat* **3**, 22–35 (2021).
5. Cao, T., Shi, X.-L. & Chen, Z.-G. Advances in the design and assembly of flexible thermoelectric device. *Prog. Mater. Sci.* **131**, 101003 (2023).
6. Shi, X. et al. Room-temperature ductile inorganic semiconductor. *Nat. Mater.* **17**, 421–426 (2018).
7. Wei, T. R. et al. Exceptional plasticity in the bulk single-crystalline van der Waals semiconductor InSe. *Science* **369**, 542 (2020).
8. Oshima, Y., Nakamura, A. & Matsunaga, K. Extraordinary plasticity of an inorganic semiconductor in darkness. *Science* **360**, 772–774 (2018).
9. Deng, T. et al. Plastic/ductile bulk 2D van der Waals single-crystalline  $\text{SnSe}_2$  for flexible thermoelectrics. *Adv. Sci.* **9**, e2203436 (2022).
10. Wu, Y. J. et al. Twisted-layer boron nitride ceramic with high deformability and strength. *Nature* **626**, 779–784 (2024).
11. Xu, Y. et al. High-throughput calculations of magnetic topological materials. *Nature* **586**, 702–707 (2020).
12. Gao, Z. et al. High-throughput screening of 2D van der Waals crystals with plastic deformability. *Nat. Commun.* **13**, 7491 (2022).
13. Luo, J., Gao, Z. Q., Zhang, J. W., Shi, X. & Chen, L. D. Comparative study of atomic models of plastic deformation mechanism in ductile inorganic semiconductors: A theoretical discussion. *J. Appl. Phys.* **136**, 125102 (2024).

14. Nguyen, K., Zhang, M., Amores, V. J., Sanz, M. A. & Montáns, F. J. Computational modeling of dislocation slip mechanisms in crystal plasticity: a short review. *Crystals* **11**, 42 (2021).
15. Bridier, F., McDowell, D. L., Villechaise, P. & Mendez, J. Crystal plasticity modeling of slip activity in Ti-6Al-4V under high cycle fatigue loading. *Int. J. Plasticity* **25**, 1066–1082 (2009).
16. Hu, Y.-J., Sundar, A., Ogata, S. & Qi, L. Screening of generalized stacking fault energies, surface energies and intrinsic ductile potency of refractory multicomponent alloys. *Acta Mater.* **210**, 116800 (2021).
17. Groves, G. W. & Kelly, A. Independent slip systems in crystals. *Philos. Mag.* **8**, 877–887 (1963).
18. Tu, P., Zheng, Y. H., Zhuang, C., Zeng, X. Q. & Zhu, H. A high-throughput computation framework for generalized stacking fault energies of pure metals. *Comp. Mater. Sci.* **159**, 357–364 (2019).
19. Vermeij, T., Peerlings, R. H. J., Geers, M. G. D. & Hoefnagels, J. P. M. Automated identification of slip system activity fields from digital image correlation data. *Acta Mater.* **243**, 118502 (2023).
20. Gibson, J. S. K. et al. Finding and characterising active slip systems: a short review and tutorial with automation tools. *Materials* **14**, 407 (2021).
21. Chen, Z. & Daly, S. H. Active slip system identification in polycrystalline metals by digital image correlation (DIC). *Exp. Mech.* **57**, 115–127 (2016).
22. Hu, H., Briffod, F., Shiraiwa, T. & Enoki, M. Automated slip system identification and strain analysis framework using high-resolution digital image correlation data: application to a bimodal Ti-6Al-4V alloy. *Int. J. Plasticity* **166**, 103618 (2023).
23. Sperry, R. et al. Comparison of EBSD, DIC, AFM, and ECCI for active slip system identification in deformed Ti-7Al. *Mater. Charact.* **173**, 110941 (2021).
24. Petryk, H. & Kursa, M. Energy approach to the selection of deformation pattern and active slip systems in single crystals. *Eur. J. Mech. A* **104**, 105040 (2024).
25. Giustino, F. *Materials Modelling Using Density Functional Theory: Properties and Predictions*. (Oxford University Press, 2014).
26. Jones, R. O. Density functional theory: Its origins, rise to prominence, and future. *Rev. Mod. Phys.* **87**, 897–923 (2015).
27. Benoit, M., Tarrat, N. & Morillo, J. Density functional theory investigations of titanium  $\gamma$ -surfaces and stacking faults. *Model. Simul. Mater. Sci.* **21**, 015009 (2013).
28. Lu, G., Kioussis, N., Bulatov, V. V. & Kaxiras, E. Generalized-stacking-fault energy surface and dislocation properties of aluminum. *Phys. Rev. B* **62**, 3099–3108 (2000).
29. Groß, A. & Sakong, S. Ab initio simulations of water/metal interfaces. *Chem. Rev.* **122**, 10746–10776 (2022).
30. Deringer, V. L., Caro, M. A. & Csányi, G. Machine learning interatomic potentials as emerging tools for materials science. *Adv. Mater.* **31**, 1902765 (2019).
31. Kulichenko, M. et al. Uncertainty-driven dynamics for active learning of interatomic potentials. *Nat. Comput. Sci.* **3**, 230–239 (2023).
32. Bassman Oftelie, L. et al. Active learning for accelerated design of layered materials. *NPJ Comput. Mater.* **4**, 74 (2018).
33. Sheng, Y. et al. Active learning for the power factor prediction in diamond-like thermoelectric materials. *NPJ Comput. Mater.* **6**, 171 (2020).
34. Rasmussen, T. et al. Atomistic determination of cross-slip pathway and energetics. *Phys. Rev. Lett.* **79**, 3676–3679 (1997).
35. Henkelman, G., Jóhannesson, G. & Jónsson, H. in *Theoretical Methods in Condensed Phase Chemistry* (ed Steven D. Schwartz) 269–302 (Springer Netherlands, 2002).
36. Saccani, S., Filippi, C. & Moroni, S. Minimum energy pathways via quantum Monte Carlo. *J. Chem. Phys.* **138**, 084109 (2013).
37. Jónsson, H., Mills, G. & Jacobsen, K. W. In *Classical and Quantum Dynamics in Condensed Phase Simulations* 385–404 (World Scientific, 1998).
38. Kolsbjer, E. L., Groves, M. N. & Hammer, B. An automated nudged elastic band method. *J. Chem. Phys.* **145**, 094107 (2016).
39. Wang, B., De Leon, N., Weinberger, C. R. & Thompson, G. B. A theoretical investigation of the slip systems of Ta2C. *Acta Mater.* **61**, 3914–3922 (2013).
40. Dai, S., Xiang, Y. & Srolovitz, D. J. Structure and energetics of interlayer dislocations in bilayer graphene. *Phys. Rev. B* **93**, 085410 (2016).
41. Xu, L., Ma, T.-b., Hu, Y.-z. & Wang, H. Molecular dynamics simulation of the interlayer sliding behavior in few-layer graphene. *Carbon* **50**, 1025–1032 (2012).
42. Leven, I., Maaravi, T., Azuri, I., Kronik, L. & Hod, O. Interlayer Potential for Graphene/h-BN heterostructures. *J. Chem. Theory Comput.* **12**, 2896–2905 (2016).
43. Ma, Y. et al. Remarkable plasticity and softness of polymorphic InSe van der Waals crystals. *J. Materomics* **9**, 709–716 (2023).
44. Henkelman, G. & Jónsson, H. Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points. *J. Chem. Phys.* **113**, 9978–9985 (2000).
45. Henkelman, G., Uberuaga, B. P. & Jónsson, H. A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *J. Chem. Phys.* **113**, 9901–9904 (2000).
46. Dennis, M. R. On the Burgers vector of a wave dislocation. *J. Opt. A* **11**, 094002 (2009).
47. Lubarda, V. A. Dislocation Burgers vector and the Peach–Koehler force: a review. *J. Mater. Res. Technol.* **8**, 1550–1565 (2019).
48. Gurtin, M. E. & Needleman, A. Boundary conditions in small-deformation, single-crystal plasticity that account for the Burgers vector. *J. Mech. Phys. Solids* **53**, 1–31 (2005).
49. Baggiooli, M., Kriuchevskyi, I., Sirk, T. W. & Zaccione, A. Plasticity in amorphous solids is mediated by topological defects in the displacement field. *Phys. Rev. Lett.* **127**, 015501 (2021).
50. Deringer, V. L. et al. Gaussian process regression for materials and molecules. *Chem. Rev.* **121**, 10073–10141 (2021).
51. Manzhos, S. & Ihara, M. The loss of the property of locality of the kernel in high-dimensional Gaussian process regression on the example of the fitting of molecular potential energy surfaces. *J. Chem. Phys.* **158**, 044111 (2023).
52. Nguyen-Tuong, D., Seeger, M. & Peters, J. Model learning with local Gaussian process regression. *Adv. Robot.* **23**, 2015–2034 (2009).
53. Schulz, E., Speekenbrink, M. & Krause, A. A tutorial on Gaussian process regression: modelling, exploring, and exploiting functions. *J. Math. Psychol.* **85**, 1–16 (2018).
54. Duvenaud, D. Z. *Automatic Model Construction with Gaussian Processes*. <https://www.cs.toronto.edu/~duvenaud/thesis.pdf> (2014).
55. Rasmussen, C. E. & Williams, C. K. I. *Gaussian Processes for Machine Learning*. <https://doi.org/10.7551/mitpress/3206.001.0001> (The MIT Press, 2005).
56. Huber, S. P. et al. AiiDA 1.0, a scalable computational infrastructure for automated reproducible workflows and data provenance. *Sci. Data* **7**, 300 (2020).
57. Uhrin, M., Huber, S. P., Yu, J., Marzari, N. & Pizzi, G. Workflows in AiiDA: engineering a high-throughput, event-based engine for robust and modular computational workflows. *Comp. Mater. Sci.* **187**, 110086 (2021).
58. Pedregosa, F. et al. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
59. Ong, S. P. et al. Python Materials Genomics (pymatgen): a robust, open-source Python library for materials analysis. *Comp. Mater. Sci.* **68**, 314–319 (2013).
60. Jain, A. et al. Commentary: the materials project: a materials genome approach to accelerating materials innovation. *APL Mater.* **1**, 011002 (2013).
61. Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **50**, 17953–17979 (1994).
62. Kresse, G. & Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169–11186 (1996).

63. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
64. Grimme, S., Ehrlich, S. & Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **32**, 1456–1465 (2011).
65. Hjorth Larsen, A. et al. The atomic simulation environment—a Python library for working with atoms. *J. Phys.* **29**, 273002 (2017).

## Acknowledgements

This work was supported by the National Natural Science Foundation of China (Nos. 91963208, 52232010, and 52122213), the Talent Plan of Shanghai Branch, Chinese Academy of Sciences (No. CASSHB-QNPD-2023-003), and Shanghai Government (Nos. 23JC1404000 and 23ZR1472800). We thank the computational resource provided by the Supercomputer Center in Shanghai Institute of Ceramics for DFT calculations in this study.

## Author contributions

J. Luo and J. Zhang proposed the workflow. J. Luo designed the code, performed the calculations, and analyzed and processed the data. J. Luo and T. Fan wrote the original paper. J. Zhang, T. Fan and X. Shi reviewed and edited the paper. J. Zhang and L. Chen supervised the project. J. Zhang and X. Shi acquired funding for this project. J. Zhang and P. Qiu provided scientific insight. All authors contributed to the discussions.

## Competing interests

The authors declare no competing interests. Prof. L. Chen serves as an editor of this journal and had no role in the peer-review or decision to publish this paper.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41524-025-01531-7>.

**Correspondence** and requests for materials should be addressed to Jiawei Zhang or Xun Shi.

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

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

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025