

Modeling extensive defects in metals through classical potential-guided sampling and automated configuration reconstruction

Shuang, Fei; Liu, Kai; Ji, Yucheng; Gao, Wei; Laurenti, Luca; Dey, Poulumi

DOI

[10.1038/s41524-025-01599-1](https://doi.org/10.1038/s41524-025-01599-1)

Publication date

2025

Document Version

Final published version

Published in

npj Computational Materials

Citation (APA)

Shuang, F., Liu, K., Ji, Y., Gao, W., Laurenti, L., & Dey, P. (2025). Modeling extensive defects in metals through classical potential-guided sampling and automated configuration reconstruction. *npj Computational Materials*, 11(1), Article 118. <https://doi.org/10.1038/s41524-025-01599-1>

Important note

To cite this publication, please use the final published version (if applicable). Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.

<https://doi.org/10.1038/s41524-025-01599-1>

Modeling extensive defects in metals through classical potential-guided sampling and automated configuration reconstruction

Fei Shuang¹ ✉, Kai Liu¹, Yucheng Ji^{1,2}, Wei Gao^{3,4}, Luca Laurenti⁵ & Poulumi Dey¹ ✉

Extended defects such as dislocation networks and general grain boundaries are ubiquitous in metals, and accurate modeling these extensive defects is crucial to elucidate their deformation mechanisms. However, existing machine learning interatomic potentials (MLIPs) often fall short in adequately describing these defects, as their large characteristic scales exceed the computational limits of first-principles calculations. To address this challenge, we present a computational framework combining a defect genome constructed via empirical interatomic potential-guided sampling, with an automated reconstruction technique that enables accurate first-principles modeling of general defects by converting atomic clusters into periodic configurations. The effectiveness of this approach was validated through simulations of nanoindentation, tensile deformation, and fracture in BCC tungsten. This framework enhances the modeling accuracy of extended defects in crystalline materials and provides a robust foundation for advancing MLIP development by leveraging defect genomes strategically.

In the intersecting realms of computational chemistry, materials science, and mechanics, machine learning has made substantial strides, particularly through the development of machine learning interatomic potentials (MLIPs). These innovative tools have transformed atomic-scale simulations by facilitating the accurate modeling of complex material behavior with quantum-level accuracy^{1–6}. A diverse array of MLIPs leveraging unique descriptors has emerged. For instance, Neural Network Potentials (NNPs) employ the Atom-Centered Symmetry Functions (ACSF)⁷, while Gaussian Approximation Potentials (GAP) utilize the Smooth Overlap of Atomic Positions (SOAP)⁸, and Spectral Neighbor Analysis Potentials (SNAP) depend on bispectrum components⁹. Additionally, both Moment Tensor Potential (MTP)¹⁰ and Atomic Cluster Expansion (ACE)¹¹ have found extensive applications across various materials, from metals and alloys to 2D materials and complex systems like carbon and silicon^{12–14}. The computational accuracy-efficiency analysis reveals that while GAP achieves the highest accuracy, it lacks computational efficiency¹⁵. In contrast, MTP and

ACE present a more balanced profile of accuracy and computational demand, with ACE being notably faster than MTP^{14,16,17}.

An essential element in developing MLIPs is constructing a good training database that adequately covers the configuration space relevant to the intended simulations. Typically, this database is curated by domain knowledge, such as ground-state structures, structures deformed under various elastic strains, ab initio molecular dynamics (AIMD) configurations at different temperatures, and simple defects such as vacancy, interstitial, crack tip, and dislocation core¹⁸. Advancements in machine learning have driven the adoption of on-the-fly active learning (OTF-AL)^{19–21}, a sophisticated approach that accelerates the development of MLIPs by efficiently selecting new configurations exclusively for necessary DFT calculations. The effectiveness of OTF-AL strategies hinges on the ability to calculate similarities among configurations by comparing their local atomic environments (LAEs). MTP with OTF-AL, utilizing the D-optimality criterion and the MaxVol algorithm, exemplifies one of the most widely employed MLIPs,

¹Department of Materials Science and Engineering, Faculty of Mechanical Engineering, Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands. ²Beijing Advanced Innovation Center for Materials Genome Engineering, National Materials Corrosion and Protection Data Center, Institute for Advanced Materials and Technology, University of Science and Technology Beijing, 100083 Beijing, China. ³J. Mike Walker'66 Department of Mechanical Engineering, Texas A&M University, College Station, Texas, TX, 77843, USA. ⁴Department of Materials Science & Engineering, Texas A&M University, College Station, Texas, TX, 77843, USA. ⁵Delft Centre of System and Control (DCSC), Faculty of Mechanical Engineering, Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands. ✉e-mail: F.Shuang@tudelft.nl; P.dey@tudelft.nl

which has been used in numerous studies^{22,23}. Other active learning strategies, such as uncertainty-driven dynamics²⁴ and hyperactive learning²⁵, also significantly contribute to efficiently sampling the configuration space.

Two primary challenges persist in the construction of databases for the development of MLIPs. The first challenge is the comprehensive coverage of defect. For metals, it is relatively straightforward to train MLIPs that can reproduce lattice constants, elastic constants, vacancy formation energy, and energies of simple grain boundaries. However, it remains unclear whether existing MLIPs can effectively model extended defects and their interactions, even if some of them are deemed for general-purpose^{18,26–29}. This uncertainty primarily stems from the lack of quantification for unknown defects. Although OTF-AL has emerged as a promising strategy to accelerate MLIP development, existing frameworks struggle to achieve comprehensive defect sampling across all loading conditions, notably in computationally demanding simulations involving large-scale atomic systems and extended time scales. Qi et al. suggested that an ideal strategy would be to efficiently generate and sample configuration space before conducting any DFT calculations³⁰. They used M3GNet³¹ as the engine for effective and efficient database generation for atomic hydrogen diffusion in titanium hydride system. This strategy however is unsuitable for modeling extended defects in metals due to the low accuracy and computational efficiency of universal MLIPs in large-scale simulations.

The second challenge involves the accurate modeling of extended defects. For instance, while current MLIPs can handle local defects such as vacancies, screw dislocations, and simple grain boundaries, they often exclude extended defects such as dislocation nucleation and multiplication, general grain boundaries (GBs), and interactions among them³². This exclusion is due to the inability of these extended defects to fit within the small periodic configurations typically used in DFT calculations. One workaround has been to extract a non-periodic atomic cluster with a spherical shape from large-scale simulations. To restore periodic boundary conditions required for plane-wave DFT calculations, a sufficiently large vacuum layer is added to prevent image interaction, as implemented in the MLIP-3 package²³. This strategy has been applied in nanoindentation simulations³³ and elastic constant calculations of polycrystals³⁴ for diamond. Nevertheless, this approach can introduce irrelevant free surfaces or other boundaries³⁵. To overcome these limitations, new strategies have been developed to maintain periodic boundary conditions without additional vacuum layers. For instance, Zhang et al. created a periodic crack-tip DFT

cell through duplication and rotation operations³⁶. Hodapp et al. utilized a screw dislocation in BCC W as an example to construct a periodic configuration³⁷, while Mismetti et al. designed periodic configurations containing leading and trailing parts of edge dislocation in face-centered cubic aluminum (FCC Al)³⁸. However, these methods fall short in handling aforementioned extended defects.

In this study, we introduce a generalized framework for developing MLIPs specifically designed to model extensive defects in metals. This framework integrates two key components: (1) Empirical Interatomic Potentials-Guided Sampling (EIP-GS), and (2) Periodic Configuration Construction via Grand Canonical Monte Carlo simulations (PCC-GCMC). These components are crucial for overcoming the challenges previously discussed. The EIP-GS method is designed to enhance defect sampling capabilities during large-scale simulations under a wide range of loading conditions. The PCC-GCMC technique, on the other hand, converts non-periodic atomic clusters of selected defects into periodic configurations without the need for vacuum layers. This conversion is essential to ensure that these configurations are compatible with standard plane-wave DFT calculations. Focusing on body-centered cubic tungsten (BCC W), our analysis underscores the critical role of the EIP-GS and PCC-GCMC techniques in enhancing the predictive accuracy and reliability of MLIPs for large-scale simulations. Ultimately, we have developed a set of MLIPs optimized for BCC W, capable of simulating a wide range of plastic deformations. We demonstrate the capabilities of our MLIPs through polycrystal tension and nanoindentation simulations. The exceptionally low uncertainty of new MLIPs underscores the robustness and adaptability of our approach across diverse simulation challenges.

Results

MLIP development framework for modeling extensive defects

In this study, we introduce a generalized framework designed to develop versatile MLIPs for accurately modeling extensive defects in metals, as depicted in Fig. 1. This framework is implemented through a sequential three-step process. Initially, a comprehensive pool of defect structures is generated using large-scale simulations based on well-established EIPs, notably the Embedded-Atom Method (EAM) potentials for metals. These potentials have been instrumental for over three decades, enabling the rapid simulation of complex defects across millions of atoms, albeit with lower accuracy compared to DFT. These capabilities are indispensable for

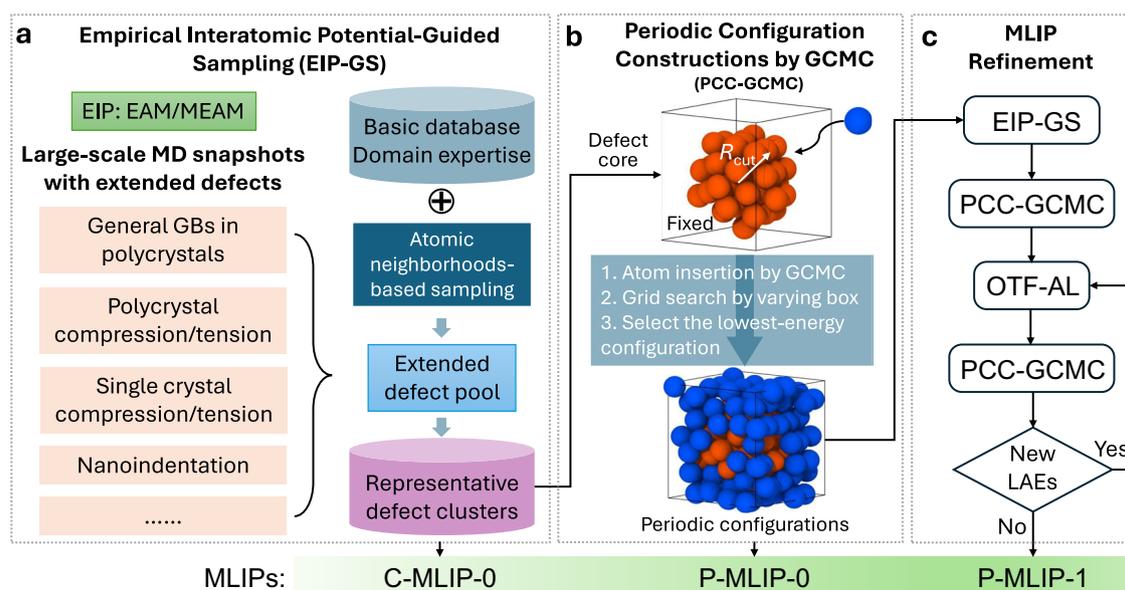


Fig. 1 | The framework of versatile MLIP development for modeling extensive defects. **a** Empirical interatomic potential-guided sampling (EIP-GS).

b Construction of periodic configurations from non-periodic atomic clusters

through precise atom insertion using grand canonical Monte Carlo simulations (PCC-GCMC). **c** MLIP refinement by combining EIP-GS, on-the-fly (OTF-AL), and PCC-GCMC.

simulating a broad spectrum of deformation scenarios, encompassing general GBs in polycrystals, as well as compression, tension, shear, and nanoindentation tests in both polycrystalline and single-crystal systems (Fig. 1a). For a particular metal (BCC W in this study), we employed EAM potentials to generate a diverse set of LAEs for various defects, addressing phenomena such as grain boundary relaxation and deformation, dislocation nucleation from material inside and exposed surfaces, dislocation multiplication, and complex dislocation-grain boundary interactions. The details of these MD simulations are presented in the “METHODS” section. Subsequently, the D-optimality criterion-based algorithm implemented in MLIP-3²³ is used to select representative LAEs according to the local atomic neighborhoods. This strategy meticulously evaluates the extensive pool of generated configurations against an established basic dataset (obtained through domain knowledge, as detailed in Supplementary Note 1), enabling the precise selection of the most representative LAEs. This process yields distinct atomic clusters with typical size of 100–200 atoms. For these clusters, vacuum layers of 8 Å thickness are added in all three dimensions of the simulation box to minimize image interaction. We develop an MLIP, named C-MLIP-0, combining the basic dataset with these cluster configurations. It should be noted that this C-MLIP-0 is very similar to the outcome of MLIP-3²³, with the only difference is that OTF-AL is used in MLIP-3, while EIP-GS is used in our approach.

In the second step, our PCC-GCMC strategy is employed to convert all clusters from the first step into periodic configurations, as shown in Fig. 1b. For each atomic cluster, we begin by initializing the simulation box size slightly larger than the atomic coordinates in all directions. Subsequently, new atoms (depicted in blue in Fig. 1b) are inserted into the box via GCMC simulations, continuing until no additional atoms can be accommodated. EAM potentials are used in GCMC simulations. During this insertion phase, the core atoms (shown in red) remain fixed to preserve the original LAE of the center atom. After the insertion is complete, both the coordinates of the newly inserted blue atoms and the box size are relaxed simultaneously. This step is crucial to prevent highly distorted LAEs among the newly inserted atoms. This procedure is iteratively repeated, with the box size gradually increasing to enable the insertion of atoms into lower energy states. By selecting only the lowest-energy configurations from PCC-GCMC, we ensure that all LAEs in these configurations are physically pertinent to the

central LAE. The resulting new DFT dataset, combined with the basic dataset, is utilized to train a foundational model, termed P-MLIP-0, designed to be suitable for simulating common defects.

The final stage, Step 3, is the MLIP refinement phase, as illustrated in Fig. 1c. For specific applications not addressed in the initial EIP-GS of Step 1, such as crack propagation and radiation damage, further EIP-GS and PCC-GCMC operations are required, utilizing appropriate EIPs. A limitation of EIP-GS is its potential bias in generating necessary LAEs due to the inherent bias of EIPs. Consequently, MLIP-based large-scale simulations may generate unforeseen LAEs. To ensure the transferability of MLIPs across diverse applications, we implement the traditional OTF-AL process here for specific cases. All generated atomic clusters are converted into periodic configurations via PCC-GCMC and reincorporated into the training dataset. This process is repeated until no new LAEs are produced by OTF-AL. The final version of our MLIP, named P-MLIP-1, represents the culmination of our development efforts and is designed to robustly model a wide range of defect phenomena.

Inconsistent data in periodic configuration and cluster with vacuum layers

First, we investigate the potential issues arising from the use of clusters with vacuum layers in DFT calculations. In Fig. 2, we compare the DFT results for a periodic configuration and an isolated cluster with vacuum layers. The periodic configuration consists of a relaxed BCC supercell, where all atoms exhibit vanishing forces due to the equilibrium conditions (such as atom A in Fig. 2a). A spherical cluster with a radius of 9 Å is then extracted from the relaxed BCC supercell, as illustrated in Fig. 2b. Since the central atom of the cluster retains the same LAE as in the pristine BCC lattice, we expect its force to be zero. However, the results in Fig. 2c reveal a non-zero force magnitude of 18.7 meV/Å for the central atom B. Atom C, the first nearest neighbor of atom B, exhibits a force magnitude of 11.7 meV/Å, while the second nearest neighbor atom D shows a significantly larger force of 527.5 meV/Å. We further compare the LAEs of atoms A–D by plotting their SOAP (Smooth Overlap of Atomic Positions) vectors in Fig. 2d. We use the SOAP descriptor³⁹ parameters set as $r_{\text{cut}} = 5\text{Å}$, $n_{\text{max}} = 12$, and $l_{\text{max}} = 10^{40}$. The SOAP analysis confirms that atoms A and B have identical environments, while atoms C and D exhibit environments very close to those of A and B.

Fig. 2 | Inconsistent DFT results in a periodic configuration and an isolated cluster. **a** A relaxed BCC W supercell containing 256 atoms. **b** An isolated spherical cluster cut from a relaxed BCC W supercell. **c** Magnitude of atomic forces in the core region of the cluster. Only the center atom and its first two neighbors are color-coded by the force magnitude. Atom B is the center atom, while atoms C and D represent the first and second neighbors, respectively. **d** Comparison of SOAP vectors for the atomic environments of atoms A, B, C, and D.

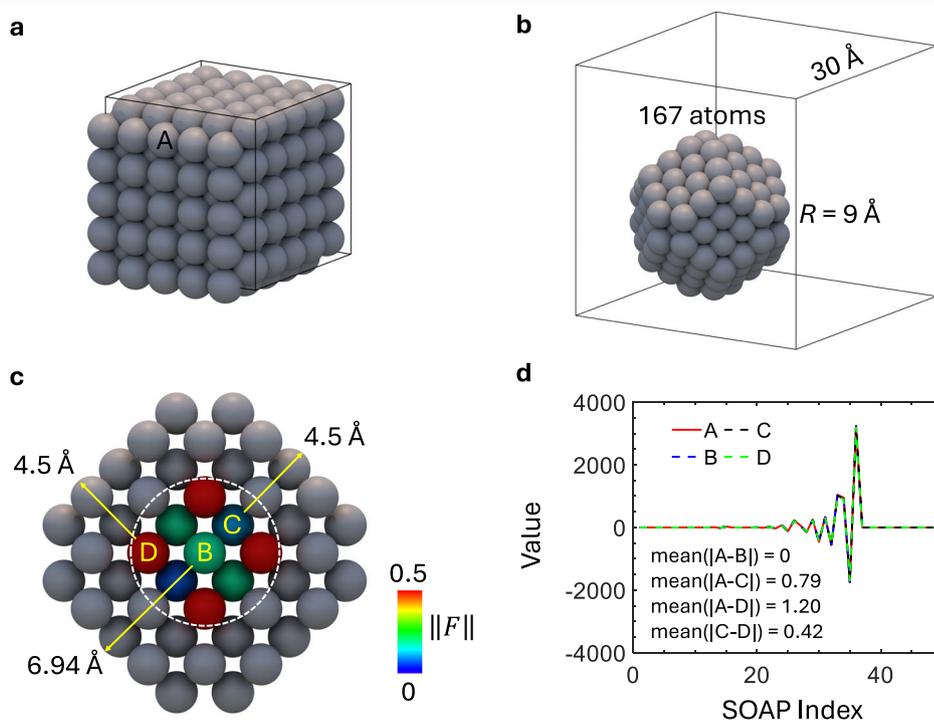
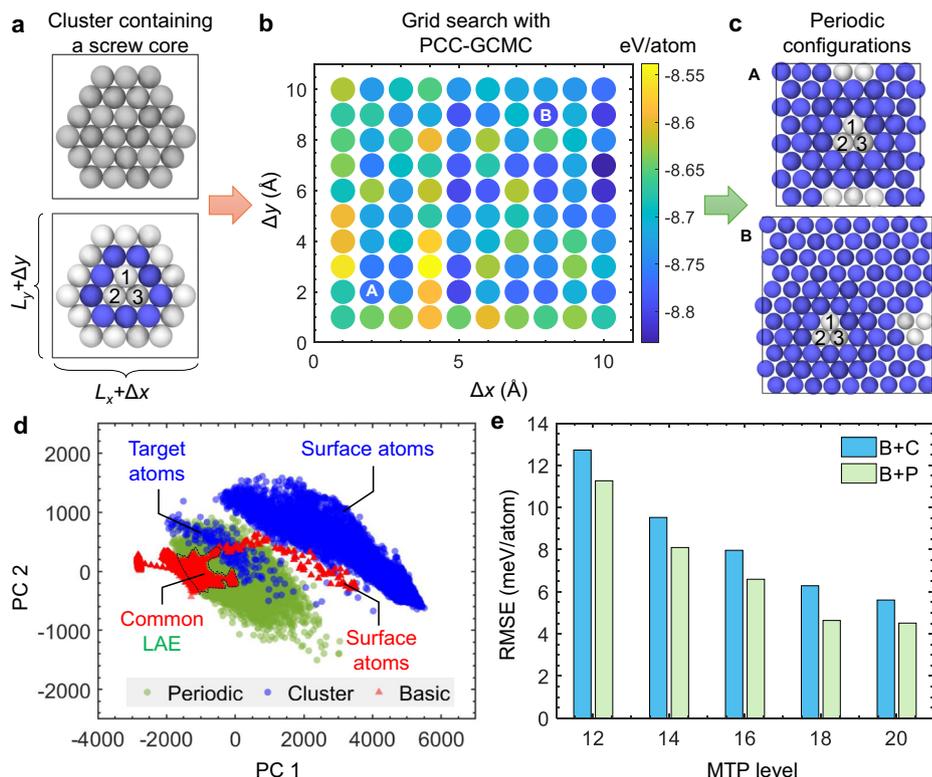


Fig. 3 | Significance of PCC-GCMC in defect structure generation. **a** Atomic cluster containing a screw dislocation core. The lower panel shows atoms color-coded by common neighbor analysis (CNA): blue denotes pristine BCC lattice atoms, while white highlights disordered atoms at the defect core and surface. **b** PCC-GCMC-based grid search for energetically stable periodic configurations, with each data point colored by the system's average energy (lower energy = darker blue). **c** Representative periodic configurations (A and B) reconstructed via PCC-GCMC, demonstrating its capability to generate DFT-compatible configurations from defect clusters. **d** Principal component analysis (PCA) of different datasets. **e** MTP training performance comparison between datasets B+C and B+P, evaluated across increasing model complexity levels (levels 12–20).



However, the force calculations performed on the cluster with vacuum layers in Fig. 2c yield inconsistent results. These findings highlight a critical discrepancy between DFT results obtained from periodic configurations and those from clusters with vacuum layers. Such inconsistencies can pose significant challenges for MLIP training, as a single atomic environment may correspond to distinct atomic forces.

The atomic forces on the central atom in an ideal BCC W structure exhibit marked differences between periodic and truncated cluster-vacuum models due to contrasting boundary conditions and interaction truncation. In the periodic model, forces arise from long-range Hellmann-Feynman contributions of all periodic images, preserving lattice symmetry and yielding near-zero net forces at equilibrium positions. In contrast, the cluster model restricts interactions to finite atomic neighbors while introducing surface-induced symmetry breaking. This truncation eliminates long-range electronic screening and imposes asymmetric force distributions from edge atoms, resulting in non-vanishing residual forces on the central atom. These force discrepancies correlate directly with asymmetric charge density redistributions near the cluster surface and disrupted delocalization patterns, as visualized through comparative CHGCAR analysis of charge density gradients as shown in Fig. S2.

Significance of PCC-GCMC

In this section, we discuss the significance of PCC-GCMC in modeling complex defects. To demonstrate the functionality of PCC-GCMC, we employ the example of a screw dislocation core in BCC W, as illustrated in Fig. 3a. The atomic cluster includes a screw dislocation core with three compact core atoms, labeled 1–3. This configuration maintains periodic boundary conditions along the z -direction, with free boundary conditions along the x and y directions. The PCC-GCMC process is repeated with varying box dimensions, Δ_x and Δ_y , from 1 to 10 Å. Fig. 3b shows the average potential energies of each configuration constructed via PCC-GCMC. Notably, configurations with potential energies below -8.8 eV/atom, depicted in blue, approach the cohesive energy of -8.9 eV/atom which are nearly close to a perfect BCC W. Fig. 3c highlights two representative configurations. In the top panel, we observe a periodic configuration

featuring a new non-compact core crossing the periodic boundary along the y -axis dislocation, with the rest of the atoms conforming to the perfect BCC lattice structure, excluding the original screw core. The bottom panel reveals a periodic configuration that includes a screw dislocation dipole, akin to typical screw dislocation dipole configurations with a tilted box. Both configurations depicted in Fig. 3c are generated entirely through automated processes. All newly inserted atoms either form a pure BCC lattice or constitute relevant LAEs, making them suitable for DFT calculation and MLIP training.

It should be noted that we strictly enforce a 140-atom system size threshold in our PCC-GCMC methodology to ensure inserted atoms physically occupy vacancy regions through energy-minimized configuration selection. While larger atomic configurations would enable more physically representative models, particularly for achieving defect-free bulk regions surrounding isolated defects (e.g., the screw dislocation core shown in Fig. 2), DFT-based approaches remain computationally intractable for atomistic-scale defect modeling at these system sizes. This fundamental limitation necessitates our optimized balance between physical fidelity and computational feasibility.

All plastic deformations utilized in EIP-GS are illustrated in Fig. S1. Large-scale simulations are conducted using the EAM-Zhou potential⁴¹. Following the procedure outlined in Fig. 1a, we generate both cluster and periodic datasets. We systematically compare three datasets: (1) the cluster dataset (Fig. 1a), (2) the periodic dataset (Fig. 1b), and (3) the basic dataset curated through domain knowledge. These comparisons employ the Smooth Overlap of Atomic Positions (SOAP) descriptor³⁹ with parameters $r_{\text{cut}} = 5$ Å, $n_{\text{max}} = 12$, and $l_{\text{max}} = 10$ ⁴⁰. The results, visualized in Fig. 3d through principal component analysis (PCA), reveal distinct patterns. The basic dataset predominantly occupies $PC_1 < 0$, with an extended arm representing surface atoms with $PC_1 > 0$. The cluster dataset bifurcates into two groups: a minor cluster overlapping with the core of the basic dataset, and a major cluster positioned above its extended arm. This dominant group corresponds to surface atoms near free surfaces, while the smaller group reflects defect-related LAEs at cluster centers. The size disparity between groups indicates that only a minority of LAEs (central atoms) directly

influence defect modeling, whereas most surface atoms do not contribute meaningfully to force/energy calculations of bulk atoms.

Unlike the cluster dataset, the periodic dataset forms a single cohesive group containing all target atoms from the cluster dataset. There is an overlapping region between the periodic and basic datasets, marked by a dashed line, suggesting shared common LAEs. This overlap signifies that the periodic configurations created via our PCC-GCMC method exhibit characteristics similar to those identified in ref. 42, with data points in this region likely corresponding to the perfect BCC lattice. Additionally, other data points in the periodic dataset represent inserted atoms from the GCMC process and atoms close to the surfaces in the original clusters. Although these LAEs are not explicitly defined in the initial EIP-GS (such as the non-compact screw core discovered in Fig. 3c), they may prove instrumental in exploring other types of unknown plastic deformations, which we will discuss further in the following section. Another interesting observation is that the LAEs of defects are not included in the basic dataset (3d), indicating that the existing MLIPs obtained through domain knowledge are unsuitable for modeling general defects.

Next, we proceed to train C-MLIP-0 and P-MLIP-0, using the cluster dataset (C) and the periodic dataset (P), respectively, with both datasets integrated into the basic dataset (B). Training is performed using the MTP framework. The performance of these models is evaluated by comparing the root mean square error (RMSE) of configuration energies, as shown in Fig. 3(e). Consistently, the B+P combination outperforms B+C across all MTP levels. The superior performance of B+P can be attributed to the simpler and more consistent nature of the periodic dataset, which avoids the inclusion of complex and irrelevant LAEs, such as surface atoms, that are present in the cluster dataset (illustrated in Fig. 3d). More importantly, DFT calculations of clusters with vacuum layers may introduce inconsistencies in the data compared to the basic dataset (B), which primarily consists of periodic configurations. These inconsistencies possibly arise from the truncation of long-range interactions and the introduction of surface effects in the cluster model, which are absent in the periodic case (Fig. 2). This finding highlights the first significant advantage of using the periodic dataset for MLIP training: by eliminating irrelevant LAEs and data inconsistencies, the training process becomes more robust, leading to more accurate and reliable MLIPs.

Uncertainty quantification across diverse plastic deformations

Our datasets, i.e., B+C and B+P, are designed to accommodate a wide range of defect simulations including dislocations, GBs, dislocation nucleation and multiplication, and dislocation-GB interactions. To verify the transferability of our datasets, we utilize per-atom uncertainty quantification available in MLIP-3²³ (the extrapolation grade γ) to evaluate various defects in BCC W. Values of γ ranging from 0 to 1 signify interpolation, while values greater than 1 imply extrapolation. Figure 4a–c display three common types of dislocation loops encountered in the plastic deformation of BCC metals: a $\langle 100 \rangle$ interstitial loop, a $\langle 111 \rangle$ glide loop, and a $\langle 111 \rangle$ climb loop. Given that the performance outcomes of the basic dataset combined with cluster one (B+C) and the basic dataset with periodic one (B+P) are comparable, we present only the results for B+C for clarity. Notably, using the basic dataset (B), all atoms within the $\langle 100 \rangle$ interstitial loop and those near the edge segments of the $\langle 111 \rangle$ glide loop, along with some atoms in the $\langle 111 \rangle$ climb loop, exhibit γ values significantly greater than 1, indicating a higher level of uncertainty (left panels of Fig. 4a–c). In contrast, all atoms within these dislocation loops, when analyzed using the B+C dataset, show γ values less than 1. This demonstrates our datasets' enhanced effectiveness in capturing complex dislocations, with energies and forces of all LAEs obtainable through interpolation.

Additionally, we assess the transferability of our models regarding general GBs in polycrystals and their temperature-induced evolution. We calculate the extrapolation grade γ for all the atoms in a polycrystal relaxed by a MEAM potential⁴³ at 300 K and 2000 K. It should be noted that our EIP-GS in the initial EIP-GS only uses the EAM-Zhou potential⁴¹, such that the GBs obtained by MEAM may have different LAEs. Fig. 4d illustrates the polycrystal at 300 K, with each atom color-coded according to the γ value

obtained from B+C. Notably, all atoms in this representation exhibit γ values less than 1, indicating low uncertainty and high confidence in the predictions. Fig. 4e and f display histogram plots of the γ values for this polycrystal calculated using different datasets: B, B+C, and B+P at 300 K and 2000 K. The analysis reveals that the dataset B predicts γ values greater than 2 for atoms at both 300 K and 2000 K, indicating higher uncertainty. In contrast, γ values calculated by datasets B+C and B+P are consistently below 1 at 300 K. However, at 2000 K, there are a few exceptions with values of γ slightly higher than 2. This indicates that although B+C and B+P perform well at lower temperatures, at higher temperatures the uncertainty escalates but stays manageable. These results underscore the high transferability and effectiveness of generated defect genome in simulating the behavior of general GB and the thermodynamics of polycrystals.

We further evaluate the uncertainty quantification of our datasets during a continuous nanoindentation simulation, illustrated in Fig. 4g–i. Fig. 4g and h display the distribution of defect atoms color-coded by γ values at the maximum depth (4 nm), calculated using B and B+C, respectively. For the case of B, several regions exhibit γ values above 1, particularly in regions involving dislocation interactions and surface regions, as indicated by red arrows in Fig. 4g. In contrast, all defect atoms analyzed with B+C demonstrate γ values lower than 1 (Fig. 4h), indicating significantly lower uncertainty. Additionally, we calculate the maximum γ values (γ_{\max}) among all defect atoms for each snapshot throughout the nanoindentation process. Comparisons of γ_{\max} for B, B+C, and B+P reveal that B consistently shows γ_{\max} values above 2, reflecting high uncertainty throughout the nanoindentation process. Conversely, both B+C and B+P maintain γ values around 1. These results indicate that our new datasets encompasses all the plastic LAEs during nanoindentation, including dislocation nucleation from open surfaces and complex multiplication beneath the indenter. It should be noted that datasets B+C and B+P exhibit comparable performance in uncertainty quantification analysis in Fig. 4. This is because the key LAEs for these defects are included in the core atoms of both datasets. However, this does not imply that dataset P, constructed using PCC-GCMC, is unnecessary. Firstly, DFT calculations of clusters introduce inconsistencies into the training data, as highlighted earlier in Fig. 2. Secondly, the periodic configurations in the dataset P enhance the extrapolation capability of the resulting MLIP. This advantage will be discussed in detail later.

MLIP refinement

It should be noted that all LAEs discussed above are generated using the EAM-Zhou potential⁴¹. The applicability of the MLIP-0 models developed based on these LAEs for modeling extended defects in large-scale simulations, particularly within MLIP simulations, remains uncertain. To address this issue, we employ the traditional on-the-fly AL (OTF-AL) implemented in MLIP-3²³. This process allows us to monitor the extrapolation grade γ during MLIP-based simulations and to identify unknown LAEs. These activities correspond to Step 3 in Fig. 1c.

We focus on the high-temperature relaxation of a polycrystal. Fig. 4e shows that our dataset, constructed via the initial EIP-GS, includes all LAEs in a polycrystal relaxed by the MEAM potential at 300 K. However, Fig. 4f reveals that these LAEs do not fully represent those at 2000 K, making EIP-GS ineffective for high temperatures. To address this, we use OTF-AL to model polycrystal relaxation at 300 K, 900 K, 1200 K, and 2000 K, starting with the dataset B+P. In the first AL loop, no new configurations are generated at 300 K and 900 K. However, the simulation at 1200 K is terminated due to atom loss, leading to the selection of 21 new configurations with $\gamma > 2$. These are converted into periodic configurations via PCC-GCMC and added to the training database to develop an updated MLIP. In the next loop, relaxation simulations proceed smoothly across all temperatures, with 25 additional configurations identified at 2000 K. This ensures a comprehensive defect genome for modeling general GBs in random polycrystals, all embedded in periodic configurations via PCC-GCMC. These configurations form the dataset OTF-AL-GB. To assess the uncertainty of our MLIP trained on B+P+OTF-AL-GB, we use ensemble learning (detailed in the METHODS section) to evaluate a polycrystal

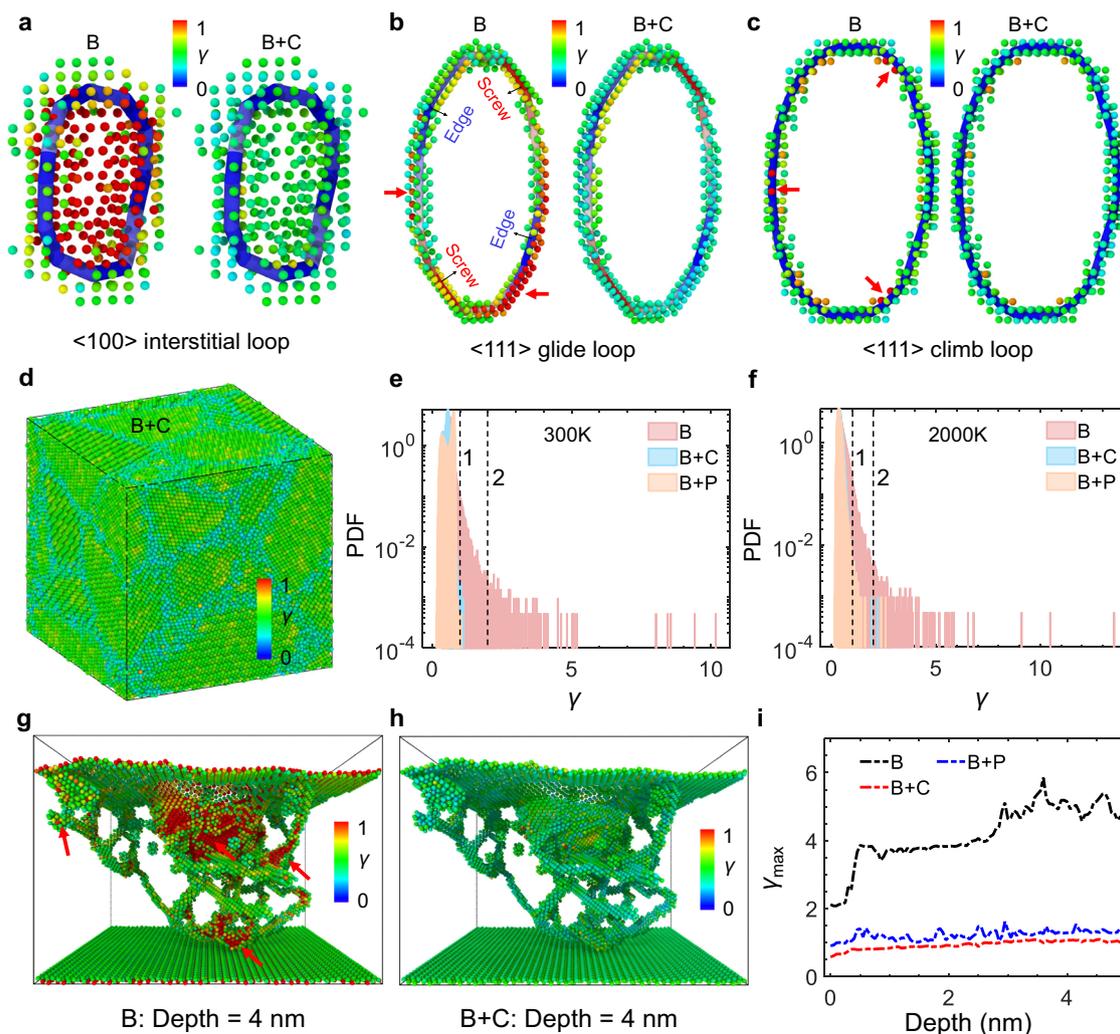


Fig. 4 | Uncertainty quantification of various defects in BCC W based on different dataset. **a–c** Three types of dislocation loop, $\langle 100 \rangle$ interstitial loop, $\langle 111 \rangle$ glide loop, and $\langle 111 \rangle$ climb loop. Left panels are results from B, and right panels are results from B+C. Blue lines are edge dislocation segments, and red lines are screw segments. **d** Polycrystal model. **e, f** The probability density function (PDF) of extrapolation grade

(γ) at 300 K and 2000 K. **g, h** Defect structures at the maximum indentation depth from B and B+C. **i** Evolution of the maximum extrapolation grade (γ_{\max}) during the whole indentation process. All atoms in (**a, b, c, d, g, h**) are color-coded by γ . Red arrows in (**b, c, g**) indicate the atoms with high uncertainty. A value of γ between 0 and 1 indicates interpolation, while a γ greater than 1 suggests extrapolation.

relaxed at 300 K and 2000 K (Fig. 5). The results show significantly reduced uncertainty, with most force standard deviations (σ_F) below 0.2 eV/Å, comparable to MLIP force prediction errors (Table S2). This confirms the high accuracy of the new MLIP in modeling general GBs.

Additionally, we refine our MLIP model by incorporating crack propagation, as detailed in Supplementary Note 1. Our approach successfully captures LAEs associated with diverse fracture mechanisms, as demonstrated in Fig. S2 using both EAM and MEAM potentials. Finally, we consolidate all datasets in Table S1 and train hierarchical MTPs alongside a high-dimensional ACE potential. Due to the excellent training accuracy of ACE, good performance for fundamental properties of BCC W, and computational efficiency (Supplementary Note 2 and Table S2), we select ACE as the final MLIP for large-scale MD simulations. The training performance of the ACE potential is summarized in Fig. S4. Further validation is provided by comparing the ACE-predicted phonon spectrum with the DFT results in Fig. S5, which shows excellent agreement. The DFT reference data are sourced from a recent study⁴⁴, reinforcing the reliability of our ACE model.

Direct validation through comparison with DFT calculations

To demonstrate the superiority of our approach and the resulting MLIP, we validate MLIPs trained on datasets B, B+C, B+P (Fig. 3), and the final

composite dataset against new DFT calculations, while benchmarking against two existing W datasets: the 2014 and 2019 datasets^{45,46}. Six ACE potentials are trained and evaluated spanning five distinct validation scenarios: (1) GBs relaxed at 300 K, 900 K, and 2000 K (equilibrium structures from⁴⁷), (2) severe compression-induced deformations and phase transitions (FCC/HCP lattices, amorphous phases; Fig. S10), (3) random 2D polycrystals, (4) 3D polycrystals generated via Voronoi tessellation, and (5) crack propagation configurations from³⁶. Representative configurations and energy/force errors comparing MLIP predictions with DFT references are shown in Fig. 6. For GBs (task 1), nine symmetric configurations exhibit temperature-dependent structural evolution, while compression (task 2) reveals pressure-driven transformations detailed in Fig. S10. In tasks 3–4 (random 2D/3D polycrystals) and task 5 (crack systems), the MLIPs are rigorously evaluated across multifaceted defect environments, probing their robustness under extreme structural complexity.

For validation tasks 1–4 in Fig. 6a, focusing on general grain boundaries (GBs), datasets B+C and B+P consistently outperform dataset B across all scenarios. A notable exception occurs in energy comparisons for thermally perturbed simple GBs, where B achieves marginally lower errors than B+C and B+P. However, B exhibits significantly higher force RMSE compared to B+C and B+P, underscoring the critical role of defect datasets in force

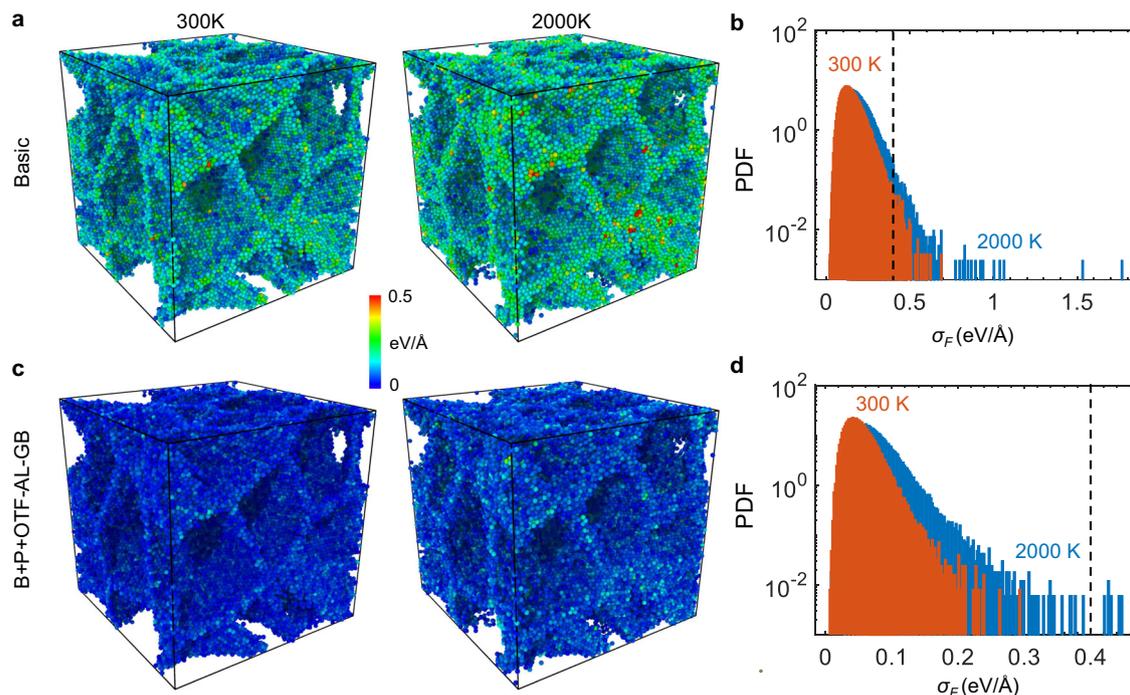
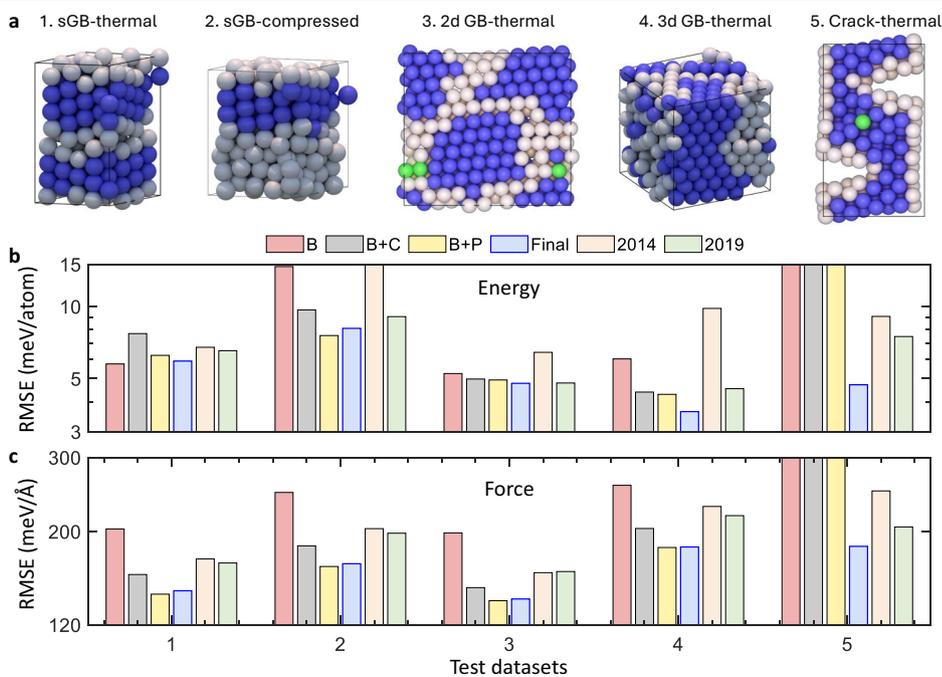


Fig. 5 | Uncertainty quantification based on ensemble learning for a polycrystal. **a** Results obtained using the basic dataset (B). **b** Probability density function (PDF) of the force uncertainty (σ_F) at temperatures of 300 K and 2000 K, considering only non-BCC atoms. **c** Results obtained using the combined dataset (B+P+OTF-GB). **d** Probability density function (PDF) of σ_F at 300 K and 2000 K, derived from the combined dataset, with only non-BCC atoms considered.

Fig. 6 | Direct validation of MLIPs for grain boundaries (GBs) and crack propagation through comparison with DFT-calculated energies and forces. **a** Structures including a simple GB under thermal perturbation and compressive loading, random 2D and 3D polycrystals, and crack propagation. Atoms in (a) are color-coded by common neighbor analysis (CNA), where blue represents perfect BCC lattices and white indicates defects or GBs. Performance of various MLIPs trained on different DFT datasets in terms of (b) energy and (c) force predictions.

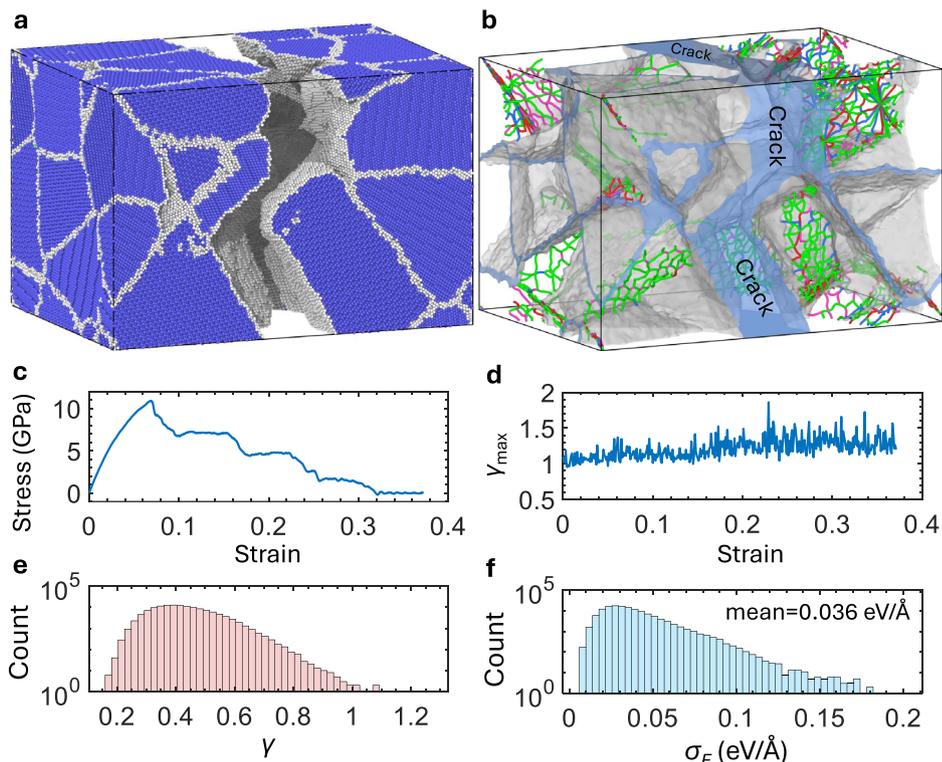


prediction. These results confirm that incorporating cluster (C) or periodic (P) configurations enhances MLIP accuracy. Remarkably, B+P surpasses B+C in both energy and force accuracy, particularly for compressed simple GBs. This superiority stems from two factors: (1) PCC-GCMC introduces additional atoms as shown in Fig. 3d, enriching the representation of critical LAEs missing in B+C, and (2) periodic configurations mitigate data inconsistencies inherent to clusters with vacuum layers, as demonstrated in Fig. 2. These findings collectively validate the superiority of PCC-GCMC-

derived periodic configurations over conventional cluster-based approaches.

Additionally, our final dataset demonstrates superior performance compared to the 2014 dataset⁴⁵ and the 2019 dataset⁴⁶. The latter dataset is widely adopted for large-scale applications such as radiation damage studies⁴⁸. Notably, our approach achieves higher force prediction accuracy across all validation tasks. With only 39 crack-related configurations (Fig. S1 and Table S1), our dataset outperforms the 2019 dataset in crack

Fig. 7 | Tensile deformation of BCC W polycrystals. a Final fracture morphology after tensile loading. **b** Dislocation network analysis and crack initiation pathways. **c** Macroscopic stress-strain response under uniaxial tension. **d** Evolution of maximum extrapolation grade (γ_{\max}) during tensile deformation. **e** Distribution of extrapolation grade (γ) across the fractured microstructure in (a). **f** Distribution of force standard deviation (σ_F) for the fractured configuration in (a).



propagation modeling, underscoring its exceptional data representativeness. This performance advantage persists even when tested on complex deformation mechanisms, highlighting the robustness of our curated atomic environments. The strength of our dataset stems from its systematic inclusion of diverse atomic configurations spanning multiple deformation modes (dislocation glide, phase transitions, fracture) in Fig. 1a, ensuring broad transferability to different material systems and extreme loading conditions. Combined with its compact size and reduced computational overhead, these features make our approach uniquely suited for both fundamental studies of material behavior and large-scale simulations of engineering-relevant scenarios.

MLIP applications in polycrystal tension

To demonstrate the high reliability of our final MLIP model in modeling general defects in BCC W, we have conducted MD simulations of polycrystal tension at room temperature (300 K). The results are presented in Fig. 7. The MD snapshot at the end of the simulation, shown in Fig. 7a, reveals a classical intergranular fracture pattern. Dislocation analysis in Fig. 7b indicates that very few dislocations are present within the grains, while dislocation networks primarily result from GB deformation. The stress-strain curve in Fig. 7c demonstrates that the W polycrystal exhibits very limited plasticity. The decreasing stress is attributed to crack initiation and propagation, with the sample fracturing completely at a strain of 0.32. These results align with previous experimental findings^{49–51}, confirming that BCC W polycrystals are highly brittle at room temperature and that intergranular fracture is the dominant deformation mechanism.

Importantly, we monitor the extrapolation grade (γ_{\max}) throughout the tension simulation, as shown in Fig. 7d. It is evident that all atomic environments encountered during the simulation have counterparts in our final dataset, highlighting the comprehensiveness of the dataset. Furthermore, we provide a detailed uncertainty quantification analysis for the MD snapshot in Fig. 7a. The distribution of γ for each atom in Fig. 7e and the force standard deviation in Fig. 7f and Fig. S7 demonstrate that our MLIP is highly reliable, with minimal uncertainty in force predictions. These

findings further validate the robustness and transferability of our MLIP for modeling complex deformation scenarios in BCC W.

Additionally, we have conducted comprehensive validation tasks to demonstrate the high reliability of our MLIP across different scenarios, including crack initiation (Fig. S8), nanoindentation into a single crystal (Fig. S9), nanoindentation into a bicrystal containing a twin boundary (Fig. S10), and structural transitions of GBs (Fig. S11). In all cases, the results exhibit very low extrapolation grade (γ) values and minimal force standard deviation, further confirming the robustness and accuracy of our MLIP in modeling diverse deformation mechanisms and defect behaviors.

Discussion

Over the past decade, MLIPs have emerged as a cornerstone in computational science, significantly impacting fields such as chemistry, materials science, and mechanics^{1–6}. Various frameworks employing a range of atomic descriptors, including ASCF⁷, SNAP⁹, moment tensor¹⁰, and ACE¹¹, have been developed. These frameworks have implemented diverse training strategies, from linear regression to neural networks, primarily focusing on enhancing the accuracy and computational efficiency of MLIPs. Such advancements are crucial, as they significantly improve the ability of MLIPs to model complex atomic interactions both accurately and efficiently, thereby paving the way for groundbreaking discoveries in multiple scientific domains. However, the construction of a robust training dataset is equally essential. The quality and comprehensiveness of the data used for training directly impact the effectiveness and reliability of MLIPs. Despite these developments, the scrutiny applied to most MLIPs remains insufficient. These potentials are often deployed in large-scale applications without undergoing rigorous validation across diverse conditions or against a wide spectrum of atomic configurations. Consequently, due to inadequately comprehensive training datasets, many existing MLIPs fail to accurately model general defects, thereby limiting their applicability in complex simulations.

Our approach demonstrates significant advancements over previous studies^{36–38}. While these earlier methods are restricted to simple defects and often rely on manual operations, such as rotation and merging in ref. 36,

they also require elastic solutions for simple defects to construct atomistic positions, as seen in refs. 37,38. In contrast, our methodology is uniquely capable of addressing complex defects, including mixed dislocations and general GBs, where prior approaches prove inadequate. An alternative strategy for developing a general-purpose MLIP involves systematic exploration of all possible space groups in random crystal structures, which generates the RANDSPG dataset^{52,53}. Although MLIPs trained on the RANDSPG dataset achieve accurate modeling of simple GBs in Mg and general GBs in BCC Fe⁵³, recent findings reveal substantial limitations. Specifically, high extrapolation grade ($\gamma > 2$) and significant force errors are observed in the polycrystal deformations of BCC Fe⁵⁴, as reported by the same authors in the previous study⁵³. In comparison, our MLIPs exhibit markedly lower uncertainty ($\gamma < 2$) in polycrystal tension even with significant fracture (Fig. 7e) and nanoindentation (Fig. S9 and S10). These results underscore the insufficiency of the RANDSPG dataset for modeling extensive defects and highlight the superior robustness of our approach.

While MLIP-3 enables OTF-AL with MTPs to address similar challenges²³, our framework integrating EIP-GS and PCC-GCMC delivers key advancements. EIP-GS significantly reduces the computational expense of OTF-AL, which improves MLIP training efficiency. Conventional OTF-AL struggles with large-scale (million-atom) defect systems due to slow MLIP inference speeds¹⁵, especially for rare events such as diffusion and phase transitions. By leveraging EIP-GS with classical potentials (e.g., EAM, which is at least 100 times faster than MTPs), we bypass this bottleneck. Additionally, OTF-AL requires iterative retraining cycles³⁰—whether using linear regression or neural networks—which remain computationally costly. Although our method retains OTF-AL, it operates as a lightweight safety check. As evidenced in Table S1, only 72 of 348 configurations derive from OTF-AL, and only one retraining is required, underscoring its minimal role in our workflow.

Our PCC-GCMC method significantly advances MLIP development for modeling extended defects, serving as a critical enhancement to both EIP-GS and OTF-AL. Unlike conventional approaches in frameworks such as MLIP-3²³—which rely on vacuum layers to mitigate image interactions—PCC-GCMC eliminates both data inconsistencies and computational inefficiencies inherent to cluster-based methods. These limitations manifest in two critical ways: (1) Data inconsistencies from vacuum-containing clusters degrade model performance during both training (Fig. 3e) and inference (Fig. 6), and (2) surface-dominated LAEs in clusters underutilize critical defect-environment information (Fig. 3a). PCC-GCMC addresses these issues by generating periodic configurations directly compatible with DFT calculations, ensuring seamless integration into training datasets. This approach circumvents artificial surface effects while enabling systematic sampling of bulk-like defect environments. The inserted atoms either reconstruct pristine lattices/defects (Fig. 3c) or reveal previously uncharacterized defect configurations, significantly enhancing MLIP transferability (Fig. 6). By completely bypassing vacuum-layer approximations, PCC-GCMC provides a robust framework for capturing complex defect interactions essential to predictive large-scale simulations.

Finally, we demonstrate that our framework is generalizable to diverse crystalline materials for modeling extended defects, as illustrated in Fig. 8. The methodology applies universally across common crystal structures—including FCC, BCC, HCP, and diamond-cubic systems—and

leverages the vast repository of EIPs from resources, such as the NIST Interatomic Potentials Repository (NIST-IPR) <https://www.ctcms.nist.gov/potentials> and OpenKIM⁵⁵. These platforms collectively provide over 644 potentials (predominantly EIPs) spanning EAM, modified EAM (MEAM), bond-order (BOP), Stillinger-Weber (SW), Tersoff, and reactive force field (ReaxFF) formulations. While EIPs exhibit larger force/energy errors compared to DFT, their physics-driven consistency has fueled computational materials science for three decades⁵⁶, which remains valuable despite the MLIP revolution. Our framework could harness these EIPs through three synergistic strategies: (1) lattice-constant-rescaled adaptation of isostructural material potentials, (2) mechanical-property-optimized parameterization^{57,58}, and (3) average-atom potential generation for random alloys⁵⁹. This enables the systematic creation of defect genome libraries for crystalline materials by sampling diverse LAEs across loading conditions. Critically, the approach transforms historical EIP investments into MLIP training assets, combining the interpretability of physical models with the accuracy of machine learning. The resulting LAE diversity—particularly for rare defect configurations—positions our methodology as both computationally efficient and physically insightful for next-generation materials simulation.

It is worth noting that extending PCC-GCMC to crystalline compounds requires careful adaptation of the GCMC algorithm to handle multiple atomic species. Nevertheless, PCC-GCMC remains applicable in certain scenarios. For high-entropy alloys, where elements are often interchangeable, PCC-GCMC can insert atoms of a single element during sampling and later substitute them with target elements. This approach prioritizes the generation of diverse structural environments while deferring the chemical complexity to post-processing steps. For non-exchangeable compounds like FeO, PCC-GCMC can insert atoms of a single element (e.g., Fe) around the oxide core, producing atomic environments relevant to interfaces (e.g., Fe/FeO boundaries). These environments are crucial for studying processes such as oxidation.

In summary, this work introduces an advanced framework for developing MLIPs by integrating EIP-GS, PCC-GCMC, and OTF-AL. This framework enables automated, ab initio-accurate, large-scale atomistic modeling of general defects in metals. The developed MLIPs are highly transferable and are based on comprehensive datasets that facilitate simulations of complex behaviors, including general GBs in polycrystals, dislocation nucleation and interactions, dislocation-GB interactions, and crack propagation in BCC W. Rigorous uncertainty quantification analyses, employing D-optimality criterion and ensemble learning techniques, confirm the accuracy of our models in these scenarios. This approach proves versatile enough to be generalized for modeling extensive defects in other BCC metals, FCC metals, and high-entropy alloys, demonstrating its broad applicability across various material systems.

Methods

First principles calculations

We utilize the Vienna Ab initio Simulation Package (VASP) to perform first-principles calculations of all new configurations necessary for MLIP development⁶⁰. A gradient-corrected functional in the Perdew-Burke-Ernzerhof (PBE) form is used to describe the exchange and correlation interactions⁶¹. Electron-ion interactions are treated within the projector-augmented-wave (PAW) method, using the standard PAW pseudopotentials provided by VASP. The energy convergence criterion is set to 10^{-6} eV for electronic self-consistency calculations. The plane-wave cutoff energy is chosen to be 520 eV. The KPOINTS are generated by VASPKIT⁶², based on the Monkhorst-Pack scheme, with a consistent density of $2\pi \times 0.03 \text{ \AA}^{-1}$ across the entire dataset. For clusters with vacuum layers, only a single KPOINT is employed.

Empirical interatomic potentials

In our simulations, we utilize both the Embedded-Atom Model (EAM) and the Modified Embedded-Atom Model (MEAM) to perform large-scale molecular dynamics simulations of body-centered cubic tungsten (BCC W) under various loading conditions. Specifically, we employ the EAM-Zhou

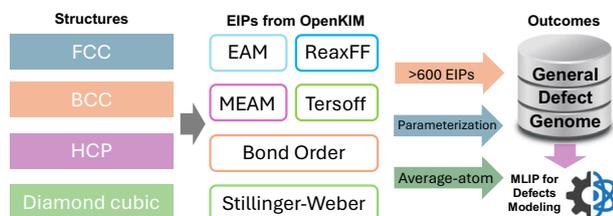


Fig. 8 | Implications of this study on MLIP development for modeling extensive defects in general crystalline materials.

potential⁴¹ for simulating the relaxation, compression, and tension of W polycrystals, as well as single-crystal compression, tension, and nanoindentation processes. Additionally, the MEAM potential referenced in ref. 43 is applied to model crack propagation in single-crystal W.

Atomistic simulations and analysis

We employ the Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS) package for all atomistic simulations⁶³. For visualizing atomic configurations and post-processing results, such as extracting dislocation structures, we use OVITO⁶⁴. Additionally, AtomsK is used to generate polycrystalline structures⁶⁵. These packages provide a comprehensive suite of tools to support our study of the mechanical properties and plastic deformation mechanisms of BCC W. All MD simulation details are shown in Supplementary Note 3.

Machine learning potential development

We employ the MLIP-2 package²² and pacemaker¹⁶ to develop MTPs and ACE potential for BCC W. MLIP-2 utilizes moment tensor descriptors and applies linear regression to train the machine learning model, enabling it to predict the energy, force, and stress of atomistic systems. In our study, the weights of energy, force, and stress are set as 1, 0.01 and 0, respectively. For ACE, we utilize the highly nonlinear potential energy function with the complex form of $E_i = \varphi + \sqrt{\varphi} + \sum_i \varphi^f$, where $E_i = \varphi + \sqrt{\varphi} + \sum_i \varphi^f$, $f_i \in \{\frac{1}{8}, \frac{1}{4}, \frac{3}{8}, \frac{3}{4}, \frac{7}{8}, 2\}$, obtained from ref. 26. For the expansion of the atomic properties, we used 72 basis functions with 801 parameters. Bessel functions are used as the radial basis. The ratio between the force and the energy weights during the fitting, κ , is set as 0.01. For optimization, we use the BFGS algorithm for 4000 steps. The cutoff distance is set as 5 Å for both MTP and ACE.

More importantly, we utilize MLIP-3²³ to perform direct sampling based on the MaxVol algorithm from extensive simulations conducted with EAM or MEAM potentials. This method was originally developed for the OFT-AL of atomic environments, incorporating a cycle of MTP-based simulations, DFT calculations, and subsequent MTP retraining. In this study, to expedite the process, we exclusively utilize the EIP-GS to identify representative defects from a comprehensive pool of large-scale simulations. This approach and its implications are elaborated in Section 3.

We employ the regular on-the-fly active learning (OTF-AL) implemented in MLIP-3²³ to sample new local atomic environments (LAEs) during high temperature relaxation of polycrystals and crack propagation. This methodology utilizes an extrapolation grade γ , which quantifies the deviation of atomic configurations from those in the training set. The process begins by training an initial moment tensor potential (MTP) on a baseline dataset. This MTP is then deployed to simulate the target system under application-specific conditions such as mechanical loading. During simulations, configurations with $\gamma > \gamma_{\text{break}} = 10$ trigger immediate termination to avoid unreliable predictions, while those with $\gamma > \gamma_{\text{select}} = 2$ are flagged for DFT validation and subsequent inclusion in the updated training set. The MTP is iteratively retrained until all configurations in production simulations are maintained $\gamma < \gamma_{\text{break}}$ ensuring robustness across target deformation regimes.

Uncertainty quantification

We adopt two methodologies for uncertainty quantification. The first utilizes atomic descriptors, specifically the extrapolation grade γ , as implemented in MLIP-3²³ and similarly applied in ACE³⁵, based on the D-optimality criterion and the MaxVol algorithm. A value of γ between 0 and 1 indicates interpolation, while a γ between 1 and 2 indicates reliable extrapolation, and greater than 10 means risky extrapolation.

The second employs ensemble learning, where we train six independent MTPs on the same dataset to compute the standard deviation of atomic force:

$$\sigma_F = \sqrt{\sum_{i \in \{x,y,z\}} \sigma(F_{\alpha,i}^{(1..N)})^2} \quad (1)$$

Here, $F_{\alpha,i}^{(1..N)}$ is the i -th component of the force acting on atom α as obtained by one of the 5 committee members (MTPs) for a particular configuration; σ denotes the standard deviation. It should be noted that while both approaches are valuable for uncertainty quantification, they fulfill different objectives. Each method can identify new local atomic environments (LAEs), but the extrapolation grade, γ , is specifically designed for this purpose and is particularly useful for OTF-AL. However, merely having a good sampling of LAE does not ensure the effectiveness of an MLIP. It is essential to implement a robust MLIP framework that can effectively train these datasets. Therefore, we integrate both methods in our work to leverage their distinct advantages.

Data availability

Data is provided within the manuscript or supplementary information files.

Code availability

All simulations are executed using open-source software LAMMPS. The machine learning force field was trained and validated by the MLIP package²². All source codes are available at the GitHub repository: <https://github.com/ufsf/ML-Defect-Modeling>.

Received: 7 January 2025; Accepted: 8 April 2025;

Published online: 03 May 2025

References

- Mishin, Y. Machine-learning interatomic potentials for materials science. *Acta Materialia* **214**, 116980 (2021).
- Deringer, V. L. et al. Machine learning interatomic potentials as emerging tools for materials science. *Adv. Mater.* **31**, 1902765 (2019).
- Unke, O. T. et al. Machine learning force fields. *Chem. Rev.* **121**, 10142–10186 (2021).
- Wang, T. et al. Unraveling dislocation-based strengthening in refractory multi-principal element alloys. *npj Comput. Mater.* **10**, 143 (2024).
- Fedik, N. et al. Extending machine learning beyond interatomic potentials for predicting molecular properties. *Nat. Rev. Chem.* **6**, 653–672 (2022).
- Deringer, V. L. et al. Gaussian process regression for materials and molecules. *Chem. Rev.* **121**, 10073–10141 (2021).
- Behler, J. Atom-centered symmetry functions for constructing high-dimensional neural network potentials. *J. Chem. Phys.* **134**, 74106 (2011).
- Bartók, A. P., Payne, M. C., Kondor, R. & Csányi, G. Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons. *Phys. Rev. Lett.* **104**, 136403 (2010).
- Thompson, A., Swiler, L., Trott, C., Foiles, S. & Tucker, G. Spectral neighbor analysis method for automated generation of quantum-accurate interatomic potentials. *J. Comput. Phys.* **285**, 316–330 (2015).
- Shapeev, A. V. Moment tensor potentials: A class of systematically improvable interatomic potentials. *Multiscale Modeling Simul.* **14**, 1153–1173 (2016).
- Drautz, R. Atomic cluster expansion for accurate and transferable interatomic potentials. *Phys. Rev. B* **99**, 014104 (2019).
- Yin, S. et al. Atomistic simulations of dislocation mobility in refractory high-entropy alloys and the effect of chemical short-range order. *Nat. Commun.* **12**, 4873 (2021).
- 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).
- Lysogorskiy, Y. et al. Performant implementation of the atomic cluster expansion (pace) and application to copper and silicon. *npj Comput. Mater.* **7**, 97 (2021).
- Zuo, Y. et al. Performance and cost assessment of machine learning interatomic potentials. *J. Phys. Chem. A* **124**, 731–745 (2020).
- Bochkarev, A. et al. Efficient parametrization of the atomic cluster expansion. *Phys. Rev. Mater.* **6**, 013804 (2022).

17. Dusson, G. et al. Atomic cluster expansion: Completeness, efficiency and stability. *J. Comput. Phys.* **454**, 110946 (2022).
18. Bartók, A. P., Kermode, J., Bernstein, N. & Csányi, G. Machine learning a general-purpose interatomic potential for silicon. *Phys. Rev. X* **8**, 041048 (2018).
19. Podryabinkin, E. V. & Shapeev, A. V. Active learning of linearly parametrized interatomic potentials. *Comput. Mater. Sci.* **140**, 171–180 (2017).
20. Zhang, L., Lin, D.-Y., Wang, H., Car, R. & E, W. Active learning of uniformly accurate interatomic potentials for materials simulation. *Phys. Rev. Mater.* **3**, 023804 (2019).
21. Vandermause, J. et al. On-the-fly active learning of interpretable bayesian force fields for atomistic rare events. *npj Comput. Mater.* **6**, 20 (2020).
22. 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 (2021).
23. Podryabinkin, E., Garifullin, K., Shapeev, A. & Novikov, I. Mlip-3: Active learning on atomic environments with moment tensor potentials. *J. Chem. Phys.* **159**, 84112 (2023).
24. Kulichenko, M. et al. Uncertainty-driven dynamics for active learning of interatomic potentials. *Nat. Comput. Sci.* **3**, 230–239 (2023).
25. van der Oord, C., Sachs, M., Kovács, D. P., Ortner, C. & Csányi, G. Hyperactive learning for data-driven interatomic potentials. *npj Comput. Mater.* **9**, 168 (2023).
26. Erhard, L. C., Rohrer, J., Albe, K. & Deringer, V. L. Modelling atomic and nanoscale structure in the silicon-oxygen system through active machine learning. *Nat. Commun.* **15**, 1927 (2024).
27. Deringer, V. L., Caro, M. A. & Csányi, G. A general-purpose machine-learning force field for bulk and nanostructured phosphorus. *Nat. Commun.* **11**, 5461 (2020).
28. Pun, G. P. P., Yamakov, V., Hickman, J., Glaessgen, E. H. & Mishin, Y. Development of a general-purpose machine-learning interatomic potential for aluminum by the physically informed neural network method. *Phys. Rev. Mater.* **4**, 113807 (2020).
29. Meng, F.-S. et al. General-purpose neural network interatomic potential for the α -iron and hydrogen binary system: Toward atomic-scale understanding of hydrogen embrittlement. *Phys. Rev. Mater.* **5**, 113606 (2021).
30. Qi, J., Ko, T. W., Wood, B. C., Pham, T. A. & Ong, S. P. Robust training of machine learning interatomic potentials with dimensionality reduction and stratified sampling. *npj Comput. Mater.* **10**, 43 (2024).
31. Chen, C. & Ong, S. P. A universal graph deep learning interatomic potential for the periodic table. *Nat. Comput. Sci.* **2**, 718–728 (2022).
32. Freitas, R. & Cao, Y. Machine-learning potentials for crystal defects. *MRS Commun.* **12**, 510–520 (2022).
33. Podryabinkin, E. V. et al. Nanohardness from first principles with active learning on atomic environments. *J. Chem. Theory Comput.* **18**, 1109–1121 (2022).
34. Jalilov, F. N., Podryabinkin, E. V., Oganov, A. R., Shapeev, A. V. & Kvashnin, A. G. Mechanical properties of single and polycrystalline solids from machine learning. *Adv. Theory Simul.* **7**, 2301171 (2024).
35. Lysogorskiy, Y., Bochkarev, A., Mrovec, M. & Drautz, R. Active learning strategies for atomic cluster expansion models. *Phys. Rev. Mater.* **7**, 043801 (2023).
36. Zhang, L., Csányi, G., van der Giessen, E. & Maresca, F. Atomistic fracture in bcc iron revealed by active learning of gaussian approximation potential. *npj Comput. Mater.* **9**, 217 (2023).
37. Hodapp, M. & Shapeev, A. In operando active learning of interatomic interaction during large-scale simulations. *Mach. Learn. Sci. Technol.* **1**, 045005 (2020).
38. Mismetti, L. & Hodapp, M. Automated atomistic simulations of dissociated dislocations with ab initio accuracy. *Phys. Rev. B* **109**, 094120 (2024).
39. Barták, A. P., Kondor, R. & Csányi, G. On representing chemical environments. *Phys. Rev. B* **87**, 184115 (2013).
40. Himanen, L. et al. Dscribe: Library of descriptors for machine learning in materials science. *Comput. Phys. Commun.* **247**, 106949 (2020).
41. Zhou, X. W., Johnson, R. A. & Wadley, H. N. G. Misfit-energy-increasing dislocations in vapor-deposited co/nife multilayers. *Phys. Rev. B* **69**, 144113 (2004).
42. Kong, L. et al. Overcoming the size limit of first principles molecular dynamics simulations with an in-distribution substructure embedding active learner. *arXiv* (2023).
43. Hiremath, P., Melin, S., Bitzek, E. & Olsson, P. A. Effects of interatomic potential on fracture behaviour in single- and bicrystalline tungsten. *Comput. Mater. Sci.* **207**, 111283 (2022).
44. Pan, J. et al. Atomic cluster expansion interatomic potential for defects and thermodynamics of cu-w system. *J. Appl. Phys.* **136**, 155108 (2024).
45. Szlachta, W. J., Barták, A. P. & Csányi, G. Accuracy and transferability of gaussian approximation potential models for tungsten. *Phys. Rev. B* **90**, 104108 (2014).
46. Byggmástar, J., Nordlund, K. & Djurabekova, F. Gaussian approximation potentials for body-centered-cubic transition metals. *Phys. Rev. Mater.* **4**, 093802 (2020).
47. Zheng, H. et al. Grain boundary properties of elemental metals. *Acta Materialia* **186**, 40–49 (2020).
48. Liu, J., Byggmástar, J., Fan, Z., Qian, P. & Su, Y. Large-scale machine-learning molecular dynamics simulation of primary radiation damage in tungsten. *Phys. Rev. B* **108**, 054312 (2023).
49. Gludovatz, B., Wurster, S., Hoffmann, A. & Pippin, R. A study into the crack propagation resistance of pure tungsten. *Eng. Fract. Mech.* **100**, 76–85 (2013).
50. Gaganidze, E., Rupp, D. & Aktaa, J. Fracture behaviour of polycrystalline tungsten. *J. Nucl. Mater.* **446**, 240–245 (2014).
51. Levin, Z. S., Srivastava, A., Foley, D. C. & Hartwig, K. T. Fracture in annealed and severely deformed tungsten. *Mater. Sci. Eng.: A* **734**, 244–254 (2018).
52. Poul, M., Huber, L., Bitzek, E. & Neugebauer, J. Systematic atomic structure datasets for machine learning potentials: Application to defects in magnesium. *Phys. Rev. B* **107**, 104103 (2023).
53. Ito, K., Yokoi, T., Hyodo, K. & Mori, H. Machine learning interatomic potential with dft accuracy for general grain boundaries in α -fe. *npj Comput. Mater.* **10**, 255 (2024).
54. Ito, K., Yokoi, T., Hyodo, K. & Mori, H. Transferability of machine-learning interatomic potential to α -Fe nanocrystalline deformation. *Int. J. Mech. Sci.* **291–292**, 110132 (2025).
55. Tadmor, E. B., Elliott, R. S., Sethna, J. P., Miller, R. E. & Becker, C. A. The potential of atomistic simulations and the knowledgebase of interatomic models. *JOM* **63**, 17–17 (2011).
56. Becker, C. A., Tavazza, F., Trautt, Z. T. & de Macedo, R. A. B. Considerations for choosing and using force fields and interatomic potentials in materials science and engineering. *Curr. Opin. Solid State Mater. Sci.* **17**, 277–283 (2013).
57. Daw, M. S. & Chandross, M. Simple parameterization of embedded atom method potentials for fcc metals. *Acta Mater.* **248**, 118771 (2023).
58. Rajan, V. P., Warner, D. H. & Curtin, W. A. An interatomic pair potential with tunable intrinsic ductility. *Model. Simul. Mater. Sci. Eng.* **24**, 025005 (2016).
59. Varvenne, C., Luque, A., Náhring, W. G. & Curtin, W. A. Average-atom interatomic potential for random alloys. *Phys. Rev. B* **93**, 104201 (2016).
60. 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).
61. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
62. Wang, V., Xu, N., Liu, J.-C., Tang, G. & Geng, W.-T. VASPKIT: A user-friendly interface facilitating high-throughput computing and analysis using vasp code. *Comput. Phys. Commun.* **267**, 108033 (2021).

63. Thompson, A. P. et al. Lammmps - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Comput. Phys. Commun.* **271**, 108171 (2022).
64. Stukowski, A. Visualization and analysis of atomistic simulation data with ovito-the open visualization tool. *Model. Simul. Mater. Sci. Eng.* **18**, 015012 (2010).
65. Hirel, P. Atomsk: A tool for manipulating and converting atomic data files. *Comput. Phys. Commun.* **197**, 212–219 (2015).

Acknowledgements

This work was sponsored by Nederlandse Organisatie voor Wetenschappelijk Onderzoek (The Netherlands Organization for Scientific Research, NWO) domain Science for the use of supercomputer facilities. The authors also acknowledge the use of DelftBlue supercomputer, provided by Delft High Performance Computing Center (<https://www.tudelft.nl/dhpc>).

Author contributions

F.S. performed DFT calculations, developed the machine learning interatomic potentials, performed the atomistic simulations, and wrote the first draft of the manuscript, K.L. and F.S. proposed the PCC-GCMC approach. F.S., L.L. and P.D. conceptualized the project and designed the research. F.S., K.L., Y.J., G.W., L.L. and P.D. analyzed the simulation data. All the authors contributed to the interpretation of the data. All authors reviewed 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/s41524-025-01599-1>.

Correspondence and requests for materials should be addressed to Fei Shuang or Poulumi Dey.

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