Automated atomistic simulations of dissociated dislocations with *ab initio* accuracy

L. Mismetti [1,*] and M. Hodapp [2,†]

[1]*École Polytechnique Fédérale de Lausanne (EPFL), Lausanne 1015, Switzerland*
[2]*Materials Center Leoben Forschung GmbH (MCL), Leoben 8700, Austria*

In a previous work [M. Hodapp and A. Shapeev, Mach. Learn.: Sci. Technol. **1**, 045005 (2020)], we proposed an algorithm that fully automatically trains machine-learning interatomic potentials (MLIPs) during large-scale simulations, and successfully applied it to simulate screw dislocation motion in body-centered-cubic tungsten. The algorithm identifies local subregions of the large-scale simulation region where the potential extrapolates, and then constructs periodic configurations of 100–200 atoms out of these nonperiodic subregions that can be efficiently computed with plane-wave density functional theory (DFT) codes. In this work, we extend this algorithm to dissociated dislocations with arbitrary character angles and apply it to partial dislocations in face-centered-cubic aluminum. Given the excellent agreement with available DFT reference results, we argue that our algorithm has the potential to become a universal way of simulating dissociated dislocations in face-centered-cubic and possibly other materials, such as hexagonal-closed-packed magnesium, and their alloys. Moreover, it can be used to construct reliable training sets for MLIPs to be used in large-scale simulations of curved dislocations.

I. INTRODUCTION

Dislocations in metals are atomistic defects and, therefore, simulation methods that operate on the atomic scale are necessary in order to predict their behavior with the best possible accuracy. Simulating a dislocation requires configurations of at least 100–200 atoms and the currently most accurate method that allows one to simulate such configurations is considered to be Kohn-Sham density functional theory (DFT). To date, the arguably most popular implementations of DFT use plane-wave basis sets, as implemented in codes such as VASP [1]; for instance, the Materials Project database [2] is almost entirely based on plane-wave DFT calculations. With such configurations, dislocations in body-centered-cubic (bcc) metals can be simulated as dipoles, subject to periodic boundary conditions, because dislocations in bcc metals have a compact core structure (see, e.g., [3,4]). For dislocations in face-centered-cubic (fcc) metals, the situation is different: fcc dislocations dissociate into partial dislocations, with a splitting distance of several times the magnitude of the Burgers vector b (cf. [5]). For example, for nickel, the partial splitting is around $8–10b$ (see, e.g., [6]), which requires supercells of at least 2000–3000 atoms to accommodate a dislocation dipole. Presently, such a large number of atoms still appears to be prohibitively expensive using plane-wave DFT because plane-wave DFT scales cubically with the number of particles. However, to be predictive, simulation methods must be able to simulate the partial splitting very accurately as it influences other mechanisms, such as dislocation cross slip. To that end,

e.g., several groups [7–9] used quantum mechanics/molecular mechanics (QM/MM) methods that resolve only the dislocation core fully atomistically and use elasticity in the far field. Other works, by Iyer *et al.* [10] and Das and Gavini [11], have used more efficient orbital-free (OF-DFT) methods that allow one to simulate much larger cells with more than 5000 atoms.

However, none of the previous methods presently appears to be tractable to solve more complex problems that require millions of atoms, e.g., dislocations in high-entropy alloys, or long curved dislocations. QM/MM methods using plane-wave DFT as the QM model would still require too many atoms to resolve the dislocation core; moreover, they require additional buffer and vacuum regions to be included in the DFT supercell to ensure accurate force fields in the vicinity of the dislocation core. In general, OF-DFT methods have, to date, not been able to accurately predict the energetics of dislocations in alloys (e.g., [12]). Newer state-of-the-art implementations of Kohn-Sham DFT with finite-element basis sets offer a high scalability compared to plane-wave DFT, allowing for configurations with 10 000–100 000 atoms, while achieving the same accuracy (cf. [13,14]). However, they still require supercomputers to simulate such a large number of atoms.

In this work, we approach the problem of simulating fcc dislocations with machine-learning interatomic potentials (MLIPs) [15–24], parametrized using plane-wave DFT calculations. Contrary to empirical interatomic potentials, which are in general not quantitatively accurate, MLIPs have a flexible functional form that allows one to systematically approximate local DFT energies down to the usual noise in numerical DFT codes [18]. In practice, it is observed that electronic effects in many metals decay sufficiently fast so that properties that are relevant for computational metallurgy,

---

*laura.mismetti@epfl.ch
†Corresponding author: maxludwig.hodapp@mcl.at

such as dislocation core structures, core energies, etc., are not quantitatively impacted when replacing DFT with an interatomic interaction model. Next-generation MLIPs even include nonlocal interactions or charges (e.g., [24–26]) that allow one to account for more complex quantum mechanical effects if, e.g., the assumption of local interactions does not hold. All of this makes MLIPs a very promising candidate for predictive large-scale simulations for dislocations in alloys.

One of the success stories of MLIPs is their ability to generalize to large systems: they can be trained on sets of configurations of a few-hundred atoms, yet they still make excellent predictions when simulating much larger configurations of tens of thousands of atoms (see, e.g., [27–29], and for some recent mathematical analysis of this observation, see [30,31]). On the other hand, one of the major challenges in constructing a good MLIP is the construction of such a training set. The vast majority of works are focusing on developing so-called general-purpose potentials that must be trained on huge datasets containing various defects (vacancies, stacking faults, grain boundaries, etc.) [32–36]. Such potentials are predictive as long as their range of application is not too far from the training data, but not necessarily DFT accurate in regions where it has not been trained (e.g., [37]).

However, for alloy design, it is more tractable to use constitutive models for a mechanical property of interest (strength, ductility, hardness, etc.) that depend on a set of descriptors, e.g., the interaction energy of dislocation with a solute, to screen for the best possible alloy. Therefore, it is also highly desirable to develop efficient *special-purpose* potentials in order to keep the training set small sized and the potential sufficiently reliable for that specific descriptor. However, it is difficult to construct such a training set because it is, as outlined above, usually not feasible to simulate extended defects, such as dislocations, with plane-wave DFT alone.

To that end, we have developed an active learning algorithm for large-scale simulations of dislocations using moment tensor potentials (MTPs) [38]. Our algorithm uses D-optimality to measure the per-atom uncertainty—the extrapolation grade—of the MTP in the simulation region, and extracts those subregions in which the extrapolation grade exceeds some threshold. Then, our algorithm completes these local (in general, nonperiodic) subregions to *periodic configurations* that are suitable to be computed with plane-wave DFT. We have applied this method to simulate screw dislocation motion in bcc tungsten and obtained excellent agreement with reference DFT results.

In this work, we extend this algorithm to general dissociated dislocations with arbitrary character angles. The new challenge, compared to our previous work on compact dislocations, is to extract *several* subregions from the dislocation core and complete them to periodic configurations of 100–200 atoms. To that end, we develop a method that extracts two regions, i.e., one around each partial dislocation core, and completes them to two periodic configurations with a net partial Burgers vector of zero in which no artificial neighborhoods occur at the periodic cell boundaries that could possibly degrade the accuracy of the MTP. Using our algorithm, we train several MTPs on dislocations in fcc aluminum and

compare their core structures and core energies to available DFT results from the literature.

II. METHODOLOGY

Before explaining our methodology, we fix some notation. Let $\{r_i\}_{i=1,...,N} = \{r_i\}$ be an arbitrary (possibly nonperiodic) configuration of N atoms, which represents the system at a specific step of the atomistic simulation. The neighborhood of the ith atom is the set $\mathcal{N} = \{r_{ij}\}$ of all relative distances $r_{ij} = r_j - r_i$ between r_i and atoms r_j within a few lattice spacings. We assume that the total energy Π of $\{r_i\}$ is partitioned into per-atom contributions $\mathcal{E} = \mathcal{E}(\mathcal{N})$ such that

$$\Pi = \Pi(\{r_i\}) = \sum_{\mathcal{N} \in \{r_i\}} \mathcal{E}(\mathcal{N}). \tag{1}$$

A. Moment tensor potentials

We model the per-atom energies with moment tensor potentials (MTPs) [18,39], which are based on a linear combination of basis function B_α,

$$\mathcal{E}(\mathcal{N}) = \sum_\alpha \theta_\alpha B_\alpha(\mathcal{N}), \tag{2}$$

where the θ_α's are free parameters. The basis functions are built from scalar contractions of the moment tensors,

$$M_{\mu,\nu}(\mathcal{N}) = \sum_{r_{ij} \in \mathcal{N}} \sum_n c_{\mu n} f_n(|r_{ij}|) (\underbrace{r_{ij} \otimes \cdots \otimes r_{ij}}_{\nu \text{ times}}), \tag{3}$$

where the $c_{\mu n}$'s are additional nonlinear parameters, and the f_n's are Chebysev radial basis functions that smoothly go to zero at the potential cutoff. In (3), μ and ν define the moment tensor level, lev $M_{\mu,\nu} = 2 + 4\mu + \nu$. An MTP with a given level lev$_{\text{MTP}}$ is then constructed using all basis functions that can be obtained from all scalar contractions of the $M_{\mu,\nu}$'s that satisfy lev$_{\text{MTP}} \geqslant \sum_{i=0}^{N_m}$ lev M_{μ_i,ν_i}, where N_m is the number of moment tensors included in the contraction. For example, for an MTP of level 16, there are 92 basis functions, and for an MTP of level 18, there are 163 basis functions.

The parameters θ_α and $c_{\mu n}$ are obtained by minimizing the MTP's predictions for energies, forces, and stresses, with respect to data coming from density functional theory (DFT) calculations. For further details regarding the training, we refer the reader to Appendix A.

B. Active learning

Suppose now that we are running a simulation with our MTP that has been trained on configurations from some training set $\mathcal{T}$. During the simulation, we come across new neighborhoods $\mathcal{N}^*$ *not present* in the training set and in order to assess whether we should better add the configuration containing those neighborhoods to our training set, we need active learning. Active learning is a method of judging whether or not to add $\mathcal{N}^*$ to the training set based on some scalar model uncertainty γ [40]. The algorithm that computes γ is called the "query strategy," and there are a number of query strategies that have been successfully adapted for MLIPs. For example, Behler [41] and Zhang *et al.* [42] proposed a query strategy based on query-by-committee for neural network potentials,

in which γ is the standard deviation between different model predictions. Podryabinkin and Shapeev [43] have proposed D-optimal design for MTPs, in which γ has the meaning of an extrapolation of the potential. Jinnouchi *et al.* [44] and Vandermause *et al.* [45] proposed Bayesian active learning for the Gaussian process-based potentials, in which γ is the predictive variance, which is naturally built into Gaussian process regression.

1. D-optimal selection of training configurations

In this work, we use the D-optimality criterion to select the training configurations, which is well established for MTPs. Hence, we consider γ as the extrapolation grade per atom in the following.

In order to compute γ, assume for a moment an MTP that has m coefficients and an active set with m neighborhoods. We then define the $m \times m$ Jacobian

$$\underline{\underline{A}} = \begin{pmatrix} \frac{\partial \mathcal{E}(\mathcal{N}_1;\underline{\theta})}{\partial \theta_1} & \cdots & \frac{\partial \mathcal{E}(\mathcal{N}_1;\underline{\theta})}{\partial \theta_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial \mathcal{E}(\mathcal{N}_m;\underline{\theta})}{\partial \theta_1} & \cdots & \frac{\partial \mathcal{E}(\mathcal{N}_m;\underline{\theta})}{\partial \theta_m} \end{pmatrix}. \quad (4)$$

The extrapolation grade γ is then defined as the maximum change in the determinant of $\underline{\underline{A}}$ if we would replace any $\mathcal{N}_i$ with $\mathcal{N}^*$. Fortunately, we do not need to replace all $\mathcal{N}_i$ with $\mathcal{N}^*$ and compute all determinants individually, but can conveniently compute γ as follows:

$$\gamma = \max_i |c_i| \quad \text{with} \quad \underline{c} = \left(\frac{\partial \mathcal{E}(\mathcal{N}^*;\underline{\theta})}{\partial \theta_1} \quad \cdots \quad \frac{\partial \mathcal{E}(\mathcal{N}^*;\underline{\theta}))}{\partial \theta_m} \right)^{\mathsf{T}} \underline{\underline{A}}^{-1}.$$

In general, our training set contains many more neighborhoods than coefficients, so $\underline{\underline{A}}$ would be overdetermined. Therefore, we select those m neighborhoods that maximize linear independence between the column vectors of $\underline{\underline{A}}$ using the maxvol algorithm [46].

Now, in order to define whether to add a configuration $\{r_i\}^*$ to the training set, we compute the γ's for all $\mathcal{N} \in \{r_i\}^*$. Following [47],

$$\gamma \leqslant 1 \quad \text{indicates interpolation,}$$
$$1 < \gamma \leqslant 2 \quad \text{indicates accurate extrapolation,}$$
$$2 < \gamma \leqslant 10 \quad \text{indicates still reliable extrapolation,}$$
$$10 < \gamma \quad \text{indicates risky extrapolation.}$$

Then, if the maximum γ is higher than some threshold, we add this configuration to the training set.

2. Active learning algorithm for large-scale simulations of fcc dislocations

In general, we are interested in simulating atomic configurations $\{r_i\}$ with tens of thousands of atoms, or even more. While active learning can reliably detect extrapolative neighborhoods in such simulations, it is not trivial to construct training configurations containing those neighborhoods because we cannot afford even one single-point plane-wave DFT calculation for such a large number of atoms. In fact, we are only able to simulate small subsets of $\{r_i\}$ with ~100–200 atoms at a time. However, blindly simulating such

a—*generally nonperiodic*—subset with plane-wave DFT degrades the accuracy of the MTP because the corresponding training configuration (that *must* be subject to periodic boundary conditions) contains artificial neighborhoods at the cell boundaries (cf. [S2] in Fig. 1).

A general way that can, in principle, be applied to any large-scale simulation is to extract cluster configurations [48,49]. However, computing clusters with plane-wave DFT requires buffer and/or vacuum regions to be introduced at the cell boundaries that can heavily increase the computational cost. Instead of using buffer or vacuum regions, one could also optimize the atoms at the cell boundaries to mimic atomic neighborhoods close to the ones that occur in the large-scale configuration [38,48,50], but it is yet unclear how this should be done for general arrangements of atoms.

In [38], we have developed an efficient method for screw dislocations in bcc metals. In this method, a rectangular cluster around the dislocation core is completed to a periodic training configuration by symmetrizing the atomic positions at the periodic cell boundaries. Recently, Zhang *et al.* [51] showed that the same idea can also be applied to cracks.

In this work, we extend our algorithm for bcc screw dislocations to general dislocations in fcc metals that dissociate into two Shockley partial dislocations. The added difficulty is that a training configuration containing a full fcc dislocation would still require too many atoms, in particular for materials with a large separation between the two partials. Therefore, we extend our algorithm as follows. We first train the MTP on several bulk configurations, and configurations containing a stacking fault. Then we are left with *two regions* around each of the partial cores, $\{r_i\}^*_{\text{lead}}$ and $\{r_i\}^*_{\text{trail}}$, where the MTP potentially extrapolates (cf. Fig. 1, [S3]). If the extrapolation grade in those regions exceeds some threshold, we complete this configuration to a periodic one by symmetrizing *only* a rectangular configuration around each partial core—but *not* around the full dislocation. This method allows one to construct two potential training configurations, $\{r_i\}^{\text{tr}}_{\text{lead}}$ and $\{r_i\}^{\text{tr}}_{\text{trail}}$, that contain the neighborhoods from the large-scale simulation region, but do not suffer from artificial neighborhoods that do not occur in the large-scale simulation region. Details on how this procedure is implemented are postponed to Sec. II B 3.

The full algorithm is schematically depicted in Fig. 1 and outlined in the following.

[S0] Define the initial atomistic configuration $\{r_i\}$ that contains a full fcc dislocation composed of two Shockley partial dislocations. Set the index of the first iteration to $n_{\text{start}} = 1$.

[S1] Run the simulation for N iterations, starting from iteration n_{start}. In each iteration, $n = n_{\text{start}}, \ldots, n_{\text{start}} + N - 1$, compute the highest extrapolation grades in both $\{r_i\}^*_{\text{lead}}$ and $\{r_i\}^*_{\text{trail}}$.

[S2] Stop the simulation after N iterations—or if an extrapolation grade exceeds some threshold $\gamma_{\max}$. If any of the extrapolation grades computed in [S1] exceeds some threshold $\gamma_{\min}$, add the corresponding configuration to the *set of training candidates*.

[S3] If the set of training candidates is not empty, update the training set using the following query strategy:

[S3.1] Complete all configurations from the set of training candidates to periodic configurations using the method from Sec. II B 3.
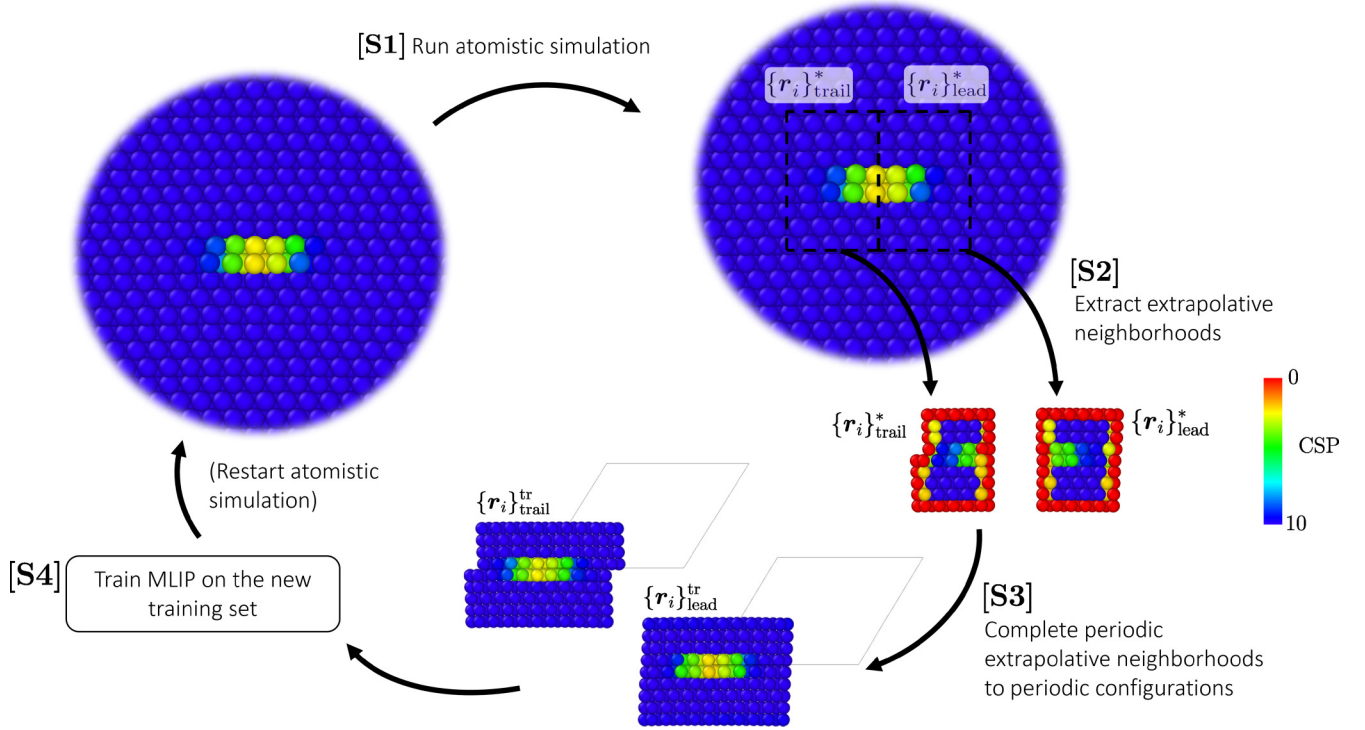
FIG. 1. Schematic illustration of the individual steps of the active learning algorithm for fcc dislocations presented in Sec. II B 2. The coloring of the atoms is due to the centrosymmetry parameter (CSP) [52], in order to visualize their deviation from the ideal crystal.

[S3.2] Move the configuration with the highest extrapolation grade from the set of training candidates to the *training set*.

[S3.3] Update the Jacobian $\underline{\underline{A}}$ (4) using the maxvol algorithm.

[S3.4] Recompute the extrapolation grades for all configurations that are left in the set of training candidates. If all grades are $< \gamma_{\min}$, go to [S4], otherwise go back to [S3.2].

[S4] If new configurations have been added to the training set, retrain the potential and go back to [S1] to restart the simulation from iteration n_{start}. Otherwise, set $n_{\text{start}} = n_{\text{start}} + N$ and go back to [S1] to continue the simulation.

Steps [S1]–[S4] are then repeated until convergence or until the maximum number of iterations is reached.

3. Completion of nonperiodic extrapolative neighborhoods to periodic training configurations

Since we want to use plane-wave DFT as our *ab initio* model, we need to apply periodic boundary conditions on our

training configurations. However, the extrapolative configurations containing the partial dislocations, $\{r_i\}_{\text{lead}}^*$ and $\{r_i\}_{\text{trail}}^*$, shown in Fig. 1, are not periodic and computing them with plane-wave DFT would degrade the reliability of the data and the accuracy of the MTP.

In [38], we have developed a method for completing such extrapolative neighborhoods to periodic configurations by symmetrizing the displacement field at the boundaries, and successfully applied it to simulate screw dislocation motion in bcc tungsten. The algorithm works as follows. First assume a straight dislocation along the z axis with the glide direction along the x axis. Let the displacement field of this dislocation be given by $\tilde{u}(x)$, with $x \in \mathbb{R}^3$, such that $\tilde{u}(r_{0,i}) = r_i - r_{0,i}$, where $r_{0,i}$ is the position of atom i in its reference (bulk) configuration, with the dislocation centered at $x = 0$. Next, we apply $\tilde{u}(x)$ only to the reference position of those atoms that lie in the rectangular region $[-L/2, L/2] \times [-L/2, L/2]$ with length L. Now, we mirror the displacements at the boundaries of this rectangular region and this procedure creates a displacement field in $[-3L/2, L/2] \times [-L/2, 3L/2]$,

$$u(x) = \begin{cases} \tilde{u}(x_1, x_2, x_3) & \forall x \in [-L/2, L/2] \times [-L/2, L/2], \\ \tilde{u}(-x_1 - L, x_2, x_3) & \forall x \in [-3L/2, -L/2] \times [-L/2, L/2], \\ \tilde{u}(x_1, -x_2 - L, x_3) & \forall x \in [-L/2, L/2] \times [L/2, 3L/2], \\ \tilde{u}(-x_1 - L, -x_2 - L, x_3) & \forall x \in [-3L/2, -L/2] \times [L/2, 3L/2], \end{cases} \tag{5}$$

which is *fully periodic* (cf. Fig. 2).

This method is general and can, in principle, be applied to any dislocation with arbitrary Burgers vector $b = (b_{\text{edge}} \quad 0 \quad b_{\text{screw}})^{\mathsf{T}}$ (and possibly even to other types of defects), as shown in Fig. 2. As done in [38], we can additionally exploit the fact that the displacement field is close-to symmetric up to a constant shift $b/2$ so that it suffices to consider training
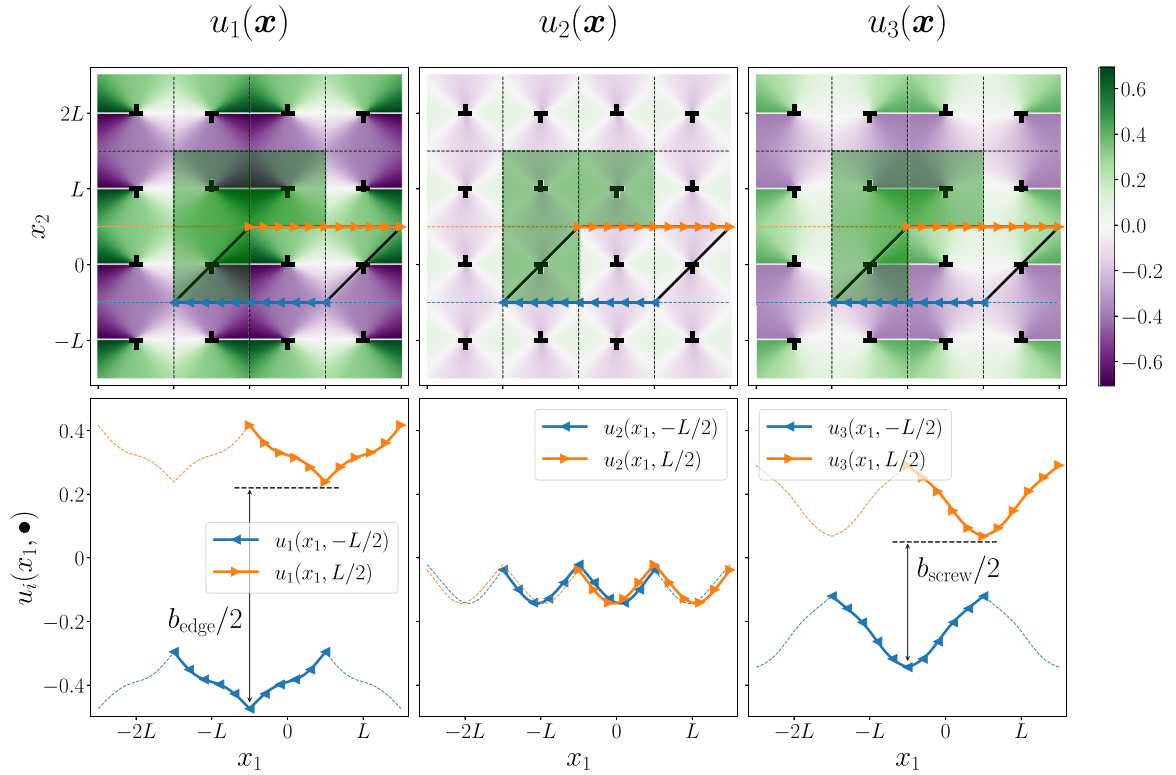
FIG. 2. Illustration of the solution (5) that is periodic over a rectangular region containing four linear elastic dislocations. Due to symmetry, the solution is even periodic up to a constant given by half of the magnitude of the Burgers vector over the smaller triclinic cell that contains only two dislocations.

configurations composed of atoms located in the triclinic region,

$$\left\{ \boldsymbol{x} = \sum_i \alpha_i \boldsymbol{v}_i + \boldsymbol{c} \,\middle|\, \alpha \in [0, 1] \right\} \qquad \text{with } \boldsymbol{c} = (-L \quad 0 \quad -L/2)^\mathsf{T}, \tag{6}$$

spanned by the cell vectors

$$\boldsymbol{v}_1 = (2L \quad 0 \quad 0)^\mathsf{T}, \quad \boldsymbol{v}_2 = (L + b_{\text{edge}}/2 \quad L \quad b_{\text{screw}}/2)^\mathsf{T}, \quad \boldsymbol{v}_3 = (0 \quad 0 \quad a_3)^\mathsf{T},$$

where a_3 is the periodic lattice spacing in the direction of the dislocation line. In the case of elasticity, the displacement $\boldsymbol{u}$ is exactly periodic over the triclinic region; for a formal proof of this statement, the reader is referred to [38] (Appendix B therein).

The problem we are considering here is more complicated since there are two (partial) dislocations in our large-scale configuration. This requires some additional steps to be performed on top of the previous method, which we explain in detail below.

(1) First, we detect the positions of the two partial dislocations since they may move along the glide plane during energy minimization. We do this by minimizing the difference between the atomistic solution and the elastic solution with respect to the positions of the two partial dislocations.

(2) Next, we compute the displacement field around each of the partial cores and apply it to some rectangular subset of the reference configuration, as shown in Fig. 3(a).

(3) Now, we extend the length of this region to $2L$ and mirror the displacements along the y axis according to (5).

This procedure creates a stacking fault in the center of the cell, but otherwise no artificial neighborhoods occur in the vicinity of the cell boundaries, except for the top and bottom layers [Fig. 3(b)].

(4) In the final step, we modify the periodic cell vectors to remove the artificial neighborhoods at the top and bottom layers. Since the atomistic solution is not exactly symmetric, we compute the shift as the difference of the minima of the solutions on the top and bottom layers,

$$\Delta \boldsymbol{u} = \min_{x_1', z_1'} \boldsymbol{u}(x_1', L/2, z_1') - \min_{x_1'', z_1''} \boldsymbol{u}(x_1'', -L/2, z_1'').$$

The cell vector $\boldsymbol{v}_2$ is then given by

$$\boldsymbol{v}_2 = (L + \Delta u_1 + d \quad L \quad \Delta u_3)^\mathsf{T},$$

where d is some additional shift that is required to match the stacking sequence and depends on the size of the cell; if $\boldsymbol{u}$ would be the elastic solution, then $\Delta \boldsymbol{u} = -\boldsymbol{b}/2$, as in (6). In practice, the final configuration with the modified cell vector $\boldsymbol{v}_2$ does not contain any artificial neighborhoods [cf. Fig. 3(c)].
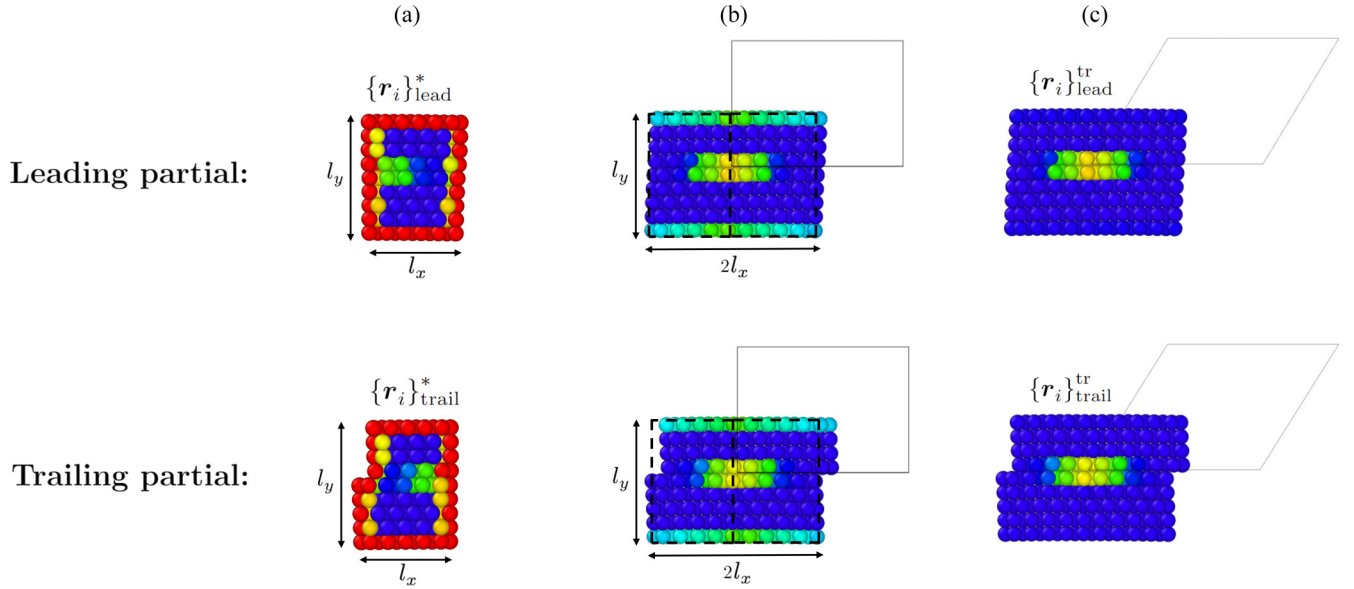
FIG. 3. Steps of the process of completing extrapolative neighborhoods around both partial dislocations to periodic training configurations. (a) Extraction of the extrapolative neighborhoods from the large-scale configuration $\{r_i\}$. (b) Mirroring of the displacement field along the y axis in order to construct a configuration that is periodic along the dislocation glide direction. (c) Adjustment of the triclinic cell vectors to remove any artificial neighborhoods near the cell boundaries.

In each iteration, we thus create two training candidate configurations containing two dislocations that have equal and opposite partial Burgers vectors (cf. Fig. 4). Of course, the configurations themselves are not physically meaningful; however, we point out that we do *not* attempt to run a full simulation on them, but *only* single-point DFT calculations. Hence, if the influence of electronic interactions decays sufficiently fast, these configurations are suitable for training a local model of interatomic interaction. We will show in the following section that the algorithm outlined above allows one to construct DFT-accurate MLIPs for fcc dislocations with any character angle.

We further remark that constructing training sets in this way is supported by the recent analysis of Ortner and Wang [30], who proved that the MLIP's error in a large-scale simulation with respect to a local DFT model converges exponentially with the size of the periodic training configurations (provided that the MLIP approximates the training set

sufficiently well). This implies that our way of training *directly on neighborhoods extracted from the large-scale simulations* should lead to a nearly optimal training set.

III. COMPUTATIONAL RESULTS

A. Training protocol

We now validate the proposed algorithm for dislocations in fcc aluminum that dissociate into $1/6\langle 112 \rangle$ Shockley partial dislocations. We apply the training algorithm to three types of dislocations: edge dislocations, screw dislocations, and mixed dislocations with a character angle of $30°$. Our large-scale simulation region is a cylindrical configuration of atoms with radius $35b$, where b is the magnitude of the Burgers vector. The simulation region contains ~ 9500 atoms for the edge and the mixed dislocations and ~ 5500 atoms for the screw dislocation. Outside the simulation region, we use Dirichlet boundary conditions, that is, we set the displacement of the
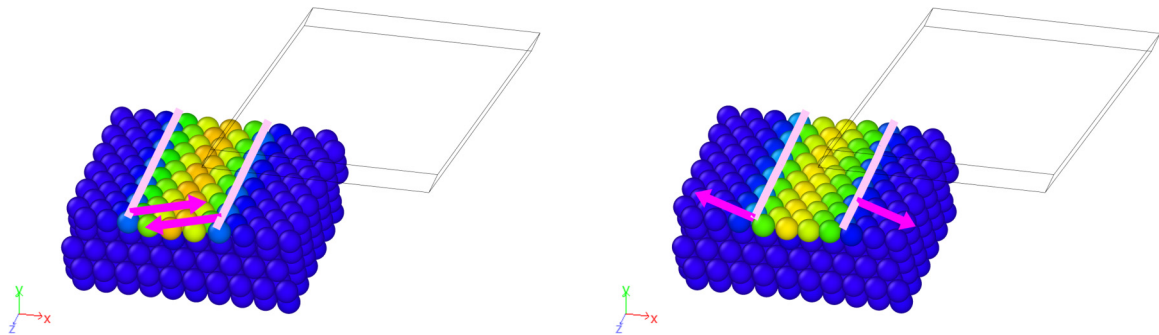

FIG. 4. Final training configurations $\{r_i\}^*_{\text{lead}}$ (left) and $\{r_i\}^*_{\text{trail}}$ (right) for an edge dislocation. The dislocations have been detected using the dislocation extraction algorithm (DXA) [53], implemented in OVITO [54]. The net Burgers vector is zero over each configuration, which follows from mirroring the configuration at the yz plane (cf. Fig. 3). Note that for a better visualization, the configuration has been extended in the z direction; the actual training configuration is confined to the triclinic cell.

TABLE I. Mean absolute errors (MAE) and root mean square errors (RMSEs) of the level-18 MTP that has been trained on configurations computed with DFT containing, e.g., edge, screw, and mixed dislocations that were found by our active learning algorithm, as described in Sec. III A. The training errors are close to the accuracy limit of interatomic potentials, indicating that the MTP can simultaneously predict the core structures of those three dislocations.

Quantity	MAE	RMSE
Energy (eV/atom)	4.2×10^{-4}	7.0×10^{-4}
Forces (eV/Å)	2.9×10^{-2}	3.6×10^{-2}
Stress (GPa)	3.1×10^{-1}	4.8×10^{-1}

atoms to the linear elastic solution of the corresponding dislocation. As our *ab initio* model, we use DFT with plane-wave basis sets, as implemented in the Vienna Ab initio Simulation Package (VASP) [1]. The corresponding simulation parameters are given in Sec. III B.

We perform structural relaxation using the fast inertial relaxation engine (FIRE) [55], as implemented in the atomic simulation environment (ASE) [56]. We consider a configuration as relaxed when the maximum absolute force on an atom is less than 10^{-3} eV/Å.

For the training simulations, we use level-16 MTPs. Prior to running the simulations, we train those MTPs on 10 32-atom bulk configurations and five 72-atom configurations containing a stacking fault (including random perturbations of the atoms). We then start all our simulations with an initial partial splitting of $3.5b$. During the simulation, we detect the two partial dislocations, compute the extrapolation grades of the atoms in the vicinity of the partial cores, and construct the potential training configurations if one per-atom extrapolation grade exceeds a threshold of $\gamma_{\min} = 2$; after $N = 30$ iterations, or if one per-atom extrapolation grade exceeds $\gamma_{\max} = 10$, we stop the simulation and update the training set according to step [S3] of the algorithm presented in Sec. II B 2. The training configurations contain 180 atoms, in the case of the edge and mixed dislocations, and 126 atoms, in the case of the screw dislocation, respectively.

Upon convergence, the training sets of the three MTPs contained, in total, 24 configurations for the edge dislocation, 44 configurations for the screw dislocation, and 30 configurations for the mixed dislocation. Hence, training the three potentials only required 68 single-point DFT calculations in total, which underlines the efficiency of the proposed training algorithm. Moreover, the training simulations can be run in parallel; each of them required ∼5–10 hours on a single 128-core node on the Vienna Scientific Cluster.

After training the three MTPs, we combine the training data into one big training set that now contains 68 configurations. We then train a level-18 MTP on this combined training set and rerun the simulations with *active learning switched off*. The training errors of this MTP are shown in Table I. The accuracy for per-atom energies is excellent, well below 1 meV/atom and, therefore, close to the limit of accuracy that can be achieved with interatomic potentials (cf. [57]). The force and stress errors are also within the range of high

TABLE II. Partial splitting distances (in Å) for the edge and screw dislocations predicted by the MTP and various DFT methods; the splitting distances are computed using three different methods: the Nye tensor method, differential displacements, and the DXA. Overall, the agreement between MTP and DFT is very good, in particular, in view of the uncertainty about the distance between two atomic planes' spacing that is inherent in all dislocation detection methods.

Dislocation	Nye tensor		Differential displacements			DXA
	MTP	QM/MM	MTP	QM/MM	OF-DFT	MTP
Edge	6.6	7.0 [7]	14.3	9.5 [7]	12.8 [10]	15.0
Screw	4.6	5.0 [7]	8.2	7.5 [7]	8.2 [11]	8.6

accuracy, i.e., of the order of 10 meV/Å for forces and of the order of 10^{-1} GPa for stresses.

In the following, we analyze the core structure and dislocation energies predicted by this level-18 MTP. Prior to training on DFT, we have tested our algorithm by training on embedded atom method (EAM) potentials. The corresponding results are in excellent agreement with the exact solutions and can be found in Appendix C. Considering that the DFT training errors (Table I) are not significantly worse than the EAM training errors (Table IV) is one indication that our training algorithm captures the right training data, and the MTP trained on DFT data with our active learning algorithm should also give very accurate results.

B. DFT calculations

In the following, we compare the core structures of the edge and screw dislocations, predicted by our MTP, with the core structures predicted by the quantum mechanics/molecular mechanics (QM/MM) method of Woodward *et al.* [7]. We have used a DFT setup similar to theirs, namely, we have used projector-augmented wave pseudopotentials [58,59] within the generalized gradient approximation [60]. Moreover, we use an energy cutoff of 480 eV, a Gaussian smearing of 0.08 eV, and a minimum $\boldsymbol{k}$-point spacing of 0.15 Å^{-1}. Electronic relaxation is performed using the preconditioned minimal residual method. We consider a configuration as converged when the energy difference between two subsequent iterations is less than 5×10^{-7} eV. With these parameters, we obtain a lattice constant of 4.042 Å. For setting up the boundary conditions using the linear elastic solution of a dislocation [5], we have used a finer grid of $\boldsymbol{k}$ points ($36 \times 36 \times 36$) to accurately predict the elastic constants, in agreement with real experiments (cf. Table III).

TABLE III. Cubic elastic constants (in GPa) predicted by DFT, MTP, OF-DFT (isotropic Voigt average), and real experiments.

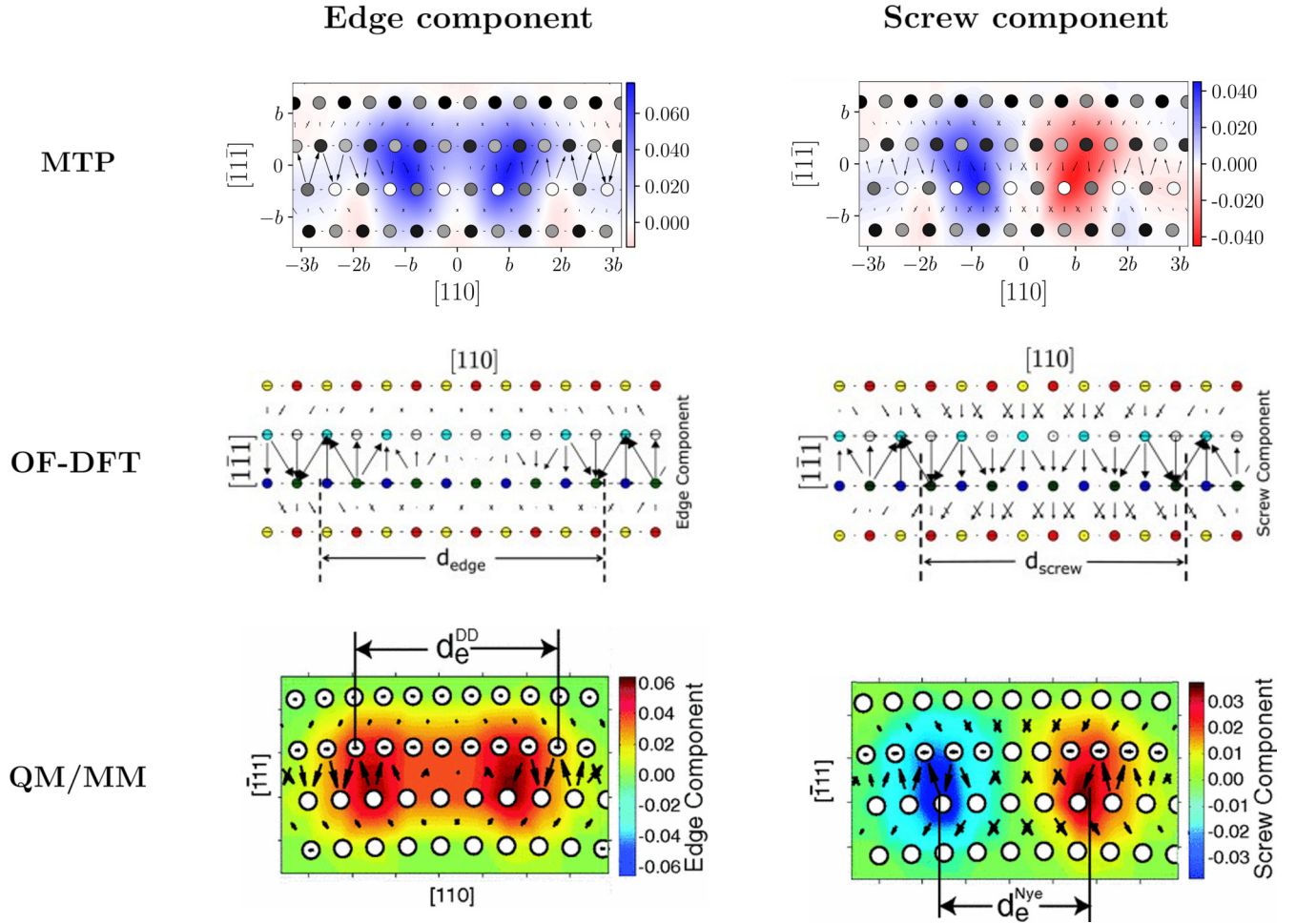| Component	DFT	MTP	OF-DFT [10,11]	Experiments [71]
C_{11}	113.4	112.0	95.3	116.3
C_{12}	60.0	63.4	51.3	64.8
C_{44}	32.9	30.4	22.0	30.9

FIG. 5. Visualization of the edge dislocation core structure predicted by the MTP, QM/MM [7], and OF-DFT [10], using the edge and screw components of the Nye tensor, and differential displacements. Overall, the agreement between the MTP and DFT core structures is very good, and the splitting distances between the partial dislocations closely coincide, up to the usual uncertainty of about the distance between two atomic planes. The QM/MM and OF-DFT figures are reproduced from [7] and [10] with permission.

In addition, we compare the dislocation core structures and dislocation energies, predicted by our MTP, to the orbital-free DFT (OF-DFT) method of Iyer *et al.* [10] and Das and Gavini [11] that uses the local density approximation for the exchange-correlation energy [61], and the Goodwin-Needs-Heine pseudopotential [62]. This OF-DFT method uses the Wang-Govind-Carter kinetic-energy functional [63], which has been shown to be in good agreement with (Kohn-Sham) DFT for bulk properties and vacancy formation energies of aluminum (see, e.g., [64]).

C. Core structure and partial splitting

We first compare the core structures of the edge and screw dislocations, predicted by our MTP, with the core structures predicted by pure-DFT methods, namely, the quantum mechanics/molecular mechanics (QM/MM) method of Woodward *et al.* [7], and the orbital-free DFT (OF-DFT) method of Iyer *et al.* [10] and Das and Gavini [11]. For comparison, we use the Nye tensor methodology of [65] and differential displacements (cf. [66]) in order to estimate the splitting distance between the partial dislocations.

From Figs. 5 and 6, it follows that the Nye tensor distributions are comparable for both the edge and the screw dislocations. The splitting distance of the edge dislocation is computed by taking the distance between the locations of the two extrema of the screw component of the Nye tensor along the glide direction, and the splitting distance of the screw dislocation is computed by taking the distance between the locations of the two extrema of the edge component of the Nye tensor along the glide direction. Our results for both the edge and the screw coincide with those computed by Woodward *et al.* [7] (cf. Table II), who used the same methodology for computing the Nye tensor.

Using differential displacements, the agreement between MTP, QM/MM, and OF-DFT is very good for the screw dislocation. For the edge dislocation, the MTP splitting is close to the OF-DFT splitting (up to the numerical uncertainty of approximately a distance of atomic planes in the glide direction due to the discreteness of the problem), but is slightly larger than the QM/MM prediction. While this difference is still small, and well within the range of tolerable deviation, it motivates further discussion. One possibility for the difference can be the size of the DFT region used in QM/MM. The size
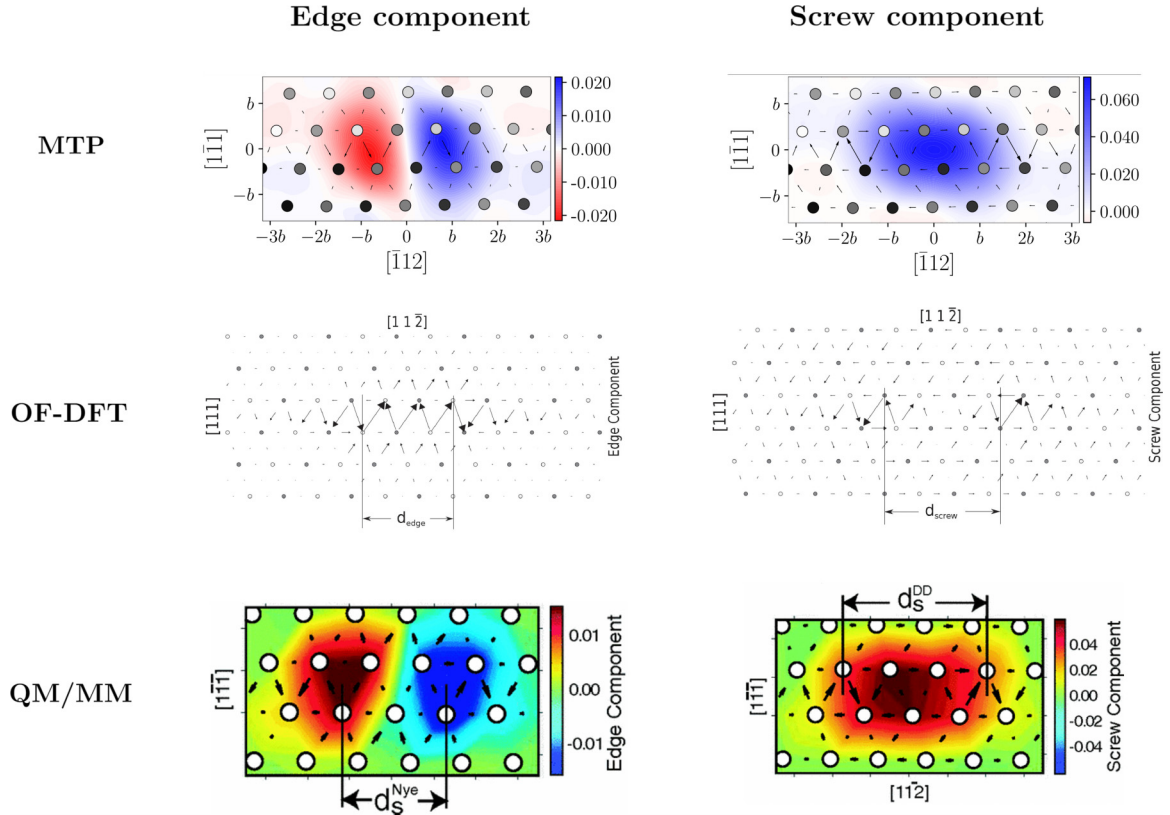
FIG. 6. Visualization of the screw dislocation core structure predicted by the MTP, QM/MM [7], and OF-DFT [11], using the edge and screw components of the Nye tensor, and differential displacements. Overall, the agreement between the MTP and DFT core structures is very good, and the splitting distances between the partial dislocations remarkably coincide. The QM/MM and OF-DFT figures are reproduced from [7] and [11] with permission.

of the DFT region in [7] is less than one-hundred atoms, while ours and OF-DFT contain more than one-thousand atoms. In general, for coupled methods such as QM/MM, there can be non-negligible spurious effects on the motion of the dislocation up to $\sim$3–4b from the boundary [67], and those effects may also influence the partial splitting. Olmsted *et al.* [68] have analyzed the spurious boundary stress on dislocations and their results show that this boundary stress is much larger for edge dislocations than for screw dislocations. This result supports our argument of attributing the differences in the partial splitting to the size of the DFT region since the splitting distances coincide when computed using the Nye tensor method, which uses the screw component of the Nye tensor, and by the fact that the splitting distances coincide across different model predictions for the screw dislocation, where the edge components of the partial dislocations are much smaller than the screw components.

The different results found in the literature, e.g., Lu *et al.* [69] report a partial splitting of 5.6 Å for the edge dislocation, show that this topic seems not to be completely settled yet. However, we anticipate that our methodology of using MLIPs and active learning, in addition to emerging mathematical analysis [30], now provides a systematic *and* tractable way for analyzing dislocation core structures in fcc metals.

We have also computed a 30°-mixed dislocation using our MTP. The predicted core structure and partial splitting is in agreement with the previous results for the edge and screw

dislocations, i.e., the partial splitting is in between the partial splitting for the edge and screw dislocations. The corresponding results are reported in Appendix B.

D. Energy differences

In order to validate whether our MTP can predict energy differences, we compute the dislocation energy $\Delta\Pi = \Delta\Pi(R)$, i.e., the difference between the energy of a configuration with and without the dislocation in a cylindrical region with radius R around the dislocation. Our results for the edge and screw dislocations and those reported in [10,11] for OF-DFT are shown in Fig. 7.

For the edge dislocation, the agreement between the MTP and OF-DFT dislocation energies is very good. The small difference between $4 \leqslant R/b \leqslant 10$ could be due to the boundary conditions used for OF-DFT that are set up with the isotropic elastic constants, which differ from the anisotropic ones (cf. Table III). This is supported by the fact that the OF-DFT curve is approximately linear in $\ln(R/b)$ in the range $4 \leqslant R/b \leqslant 10$ and its slope, i.e., the energy factor [5], is very close to that of the MTP curve in this interval. Another reason could be due to electronic effects present up to $R/b \approx 10$, as argued in [10], that cannot be captured with our local MTP.

For the screw dislocation, the MTP and OF-DFT results differ qualitatively for small R/b in the range $3 \leqslant R/b \geqslant 7$. Again, this could be due to the isotropic boundary conditions
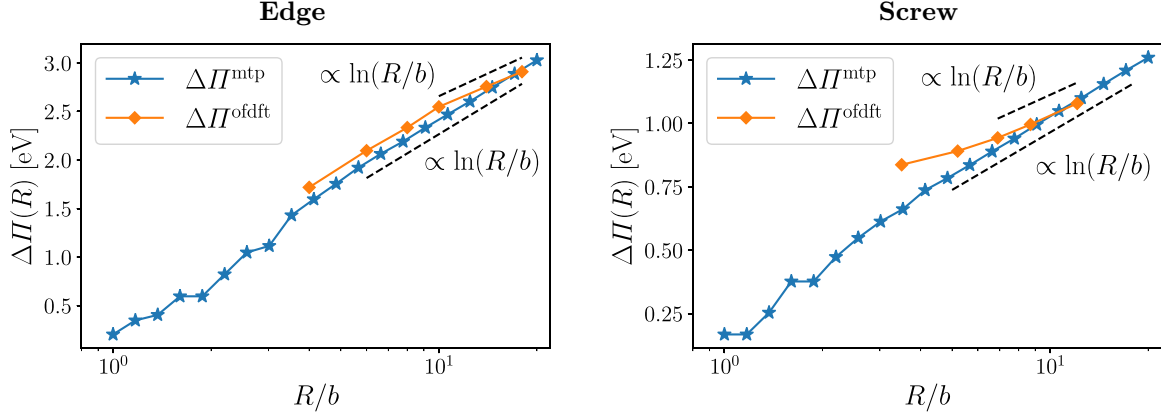
Edge   **Screw**

FIG. 7. Dislocation energies for the edge and screw dislocations predicted by the MTP and OF-DFT. The agreement between the MTP and OF-DFT is very good and the far-field behavior corresponds to the dislocation energy predicted by linear elasticity (dashed lines) up to some constant core energy.

used for OF-DFT. Another reason could be the smaller simulation regions used for OF-DFT; in [10,11], the simulation region was gradually increased with increasing R, so $\Delta\Pi^{\text{ofdft}}$ corresponds to the total energy of a configuration with radius R, whereas we use a simulation region of $35b$ and sum up the per-atom energies up to some R. As shown in [70], the difference between isotropic and anisotropic boundary conditions can have a non-negligible influence on dislocations close to the boundary, even for materials such as aluminum that are only weakly anisotropic. Moreover, there can be electronic effects up to 7 R/b, as pointed out in [11], that cannot be reproduced with a local MLIP. The behavior for smaller R/b could be similar to the results from [12], which reported oscillating dislocation energies close to the dislocation core for screw dislocations in magnesium using a QM/MM method.

Most importantly, in both cases, the dislocation energies for OF-DFT and MTP agree well when the OF-DFT curve enters the linear elastic regime. The asymptotic scalings predicted by the MTP and OF-DFT differ slightly since the isotropic elastic constants used to set up the boundary conditions for OF-DFT deviate ∼10–30% from experiments (cf. Table III). On the other hand, the elastic constants predicted by the MTP are in excellent agreement with DFT and the experimental ones. This implies that the MTP can predict reliable core energies.

We have also computed the dislocation energy for the 30°-mixed dislocation. As for the partial splitting, the results are in agreement with the results for the edge and screw in the sense that the dislocation energy for the mixed dislocation lies in between those two limiting cases. In the future, it would be interesting to also compare those results with the dislocation energies predicted by the QM/MM methods (e.g., [12]) for further validation.

IV. CONCLUDING REMARKS

A. Discussion and potential applications

Constructing a MLIP which is general and predictive is commonly referred to as the problem of transferability. One successful methodology to solve the problem of transferability is active learning, but it is impractical for large-scale problems

because, in this case, one cannot afford even one single-point plane-wave DFT calculation, in general. One important example for such large-scale problems is dislocations in fcc materials. The particular difficulty that arises in fcc materials is that the dislocations typically split into two Shockley partial dislocations and, depending on the material, their separation can be so large that simulating the full dislocation with plane-wave DFT alone would be infeasible; for example, Deng *et al.* [36] used the configurations of 7200 atoms for simulating screw dislocations in copper with interatomic potentials in a quadrupole arrangement.

To overcome this limitation, we have developed an active learning algorithm for training MTPs during large-scale simulations of dislocations in fcc materials. Our active learning algorithm only extracts a small cluster of atoms around each partial core, computes the per-atom extrapolation grade of each atom, and, if one of the grades is larger than some threshold, completes the cluster to a periodic configuration of 100–200 atoms that can be conveniently computed with plane-wave DFT. We have validated our algorithm by simulating dislocations in fcc aluminum, and the MTP, trained with our algorithm, was able to reproduce all existing DFT results available in the literature for the core structure, the partial splitting, and the dislocation core energy.

Hence, we anticipate that our algorithm can be readily used to construct reliable MLIPs for dislocations to be used for applications that can go beyond the scope of validity of the existing empirical interatomic potentials, for example:

(a) One possible application of our algorithm is dislocations in alloys. This, of course, may require larger training configurations than used here to accommodate solutes and/or impurities such as hydrogen along the dislocation line, but still appears feasible. For example, our algorithm could be used to construct MLIPs for computing dislocation-solute interactions that are inputs to solute-strengthening models (cf. [72]). This MLIP can then be used within workflows for predictive high-throughput screening (e.g., [73]) to find the strongest alloy over a wide range of compositions.

(b) Our algorithm can be readily used to construct training sets for MLIPs to be used in large-scale simulations of curved dislocations. This is supported by a recent work of Wang *et al.*
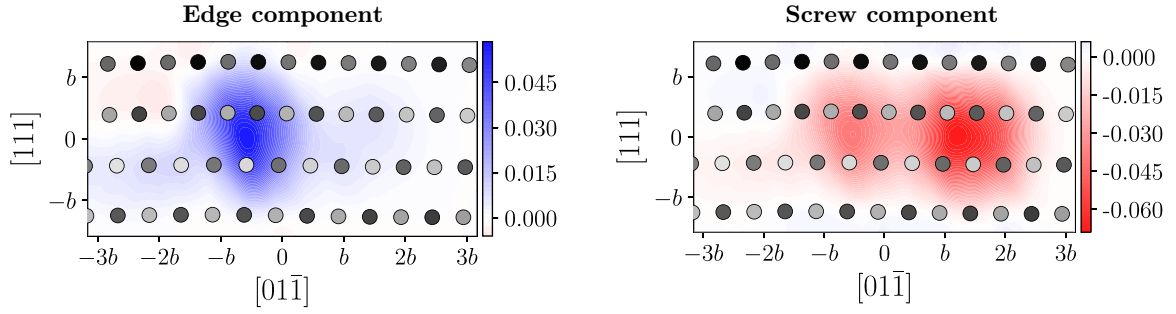
FIG. 8. Visualization of the $30°$-mixed dislocation core structure predicted by the MTP using the edge and screw components of the Nye tensor.

[31], which showed that for bcc metals, a MLIP trained on a few configurations that contain only straight dislocations with different character angles can be used to predict the energies of dislocation loops.

(c) Moreover, we anticipate that the proposed algorithm is not limited to fcc, but also to dissociated dislocations in other materials, such as, e.g., basal dislocations in hexagonal-closed-packed magnesium.

(d) Another potential application where we envision our algorithm to become valuable is magnetic materials. Including spins into the functional form of interatomic potentials doubles the size of a neighborhood and this significantly increases the complexity of constructing a good training set because the configurational space is now much larger—but active learning can come to the rescue. Recently, we approached the problem of training magnetic MTPs using a DFT method in which spins can be constrained in the same way as atomic positions [74]. Within this approach, the active learning algorithm developed here can potentially be applied verbatim to dislocations in magnetic materials.

B. How to simulate multidefect interactions?

Finally, we remark that presently, our algorithm cannot be readily applied to problems for which defects cannot be sufficiently isolated, such as dislocations cutting through precipitates. For such problems, we currently recommend to first train a potential on configurations that appear to be relevant
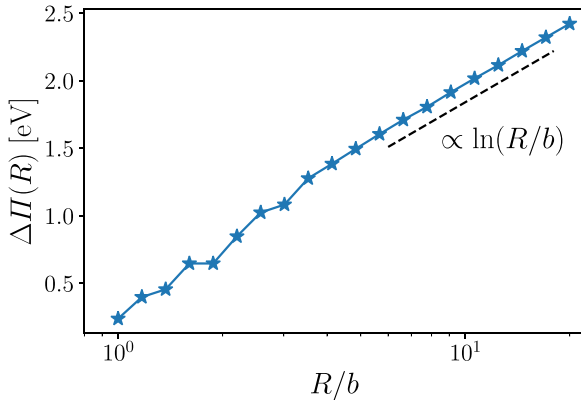

FIG. 9. Dislocation energies for the mixed dislocation predicted by the MTP. The far-field behavior corresponds to the dislocation energy predicted by linear elasticity (dashed lines) up to some constant core energy.

for the physical problem to be solved, e.g., bulk configurations, stacking faults, antiphase boundaries, etc., or on an existing dataset, e.g., from Ref. [75]. Afterwards, we recommend to run several simulations of straight dislocations and add additional configurations with high uncertainties to the training set using our algorithm. When starting the actual simulation of, say, a dislocation interacting with precipitate, one may leave active learning switched on and, if the uncertainty becomes too high in parts of the simulation region, run plane-wave DFT calculations on those parts using clusters [48], or periodic completion of fragments [76], and add them to the training set. Those methods can be more expensive than ours and pose the danger of introducing artificial neighborhoods polluting the training set, but if the initial training set is already diverse enough, such calculations would be only rarely triggered. In the future, another approach could be training on many small systems combined with training on a few large systems using more efficient *ab initio* methods such as those DFT methods that use finite-element basis sets to improve the MLIP's predictions for large arrangements of defects.

TABLE IV. Errors of the level-16 MTP that has been trained on configurations computed with the EAM containing, edge, screw, and mixed dislocations, that were found by our active learning algorithm, as described in Appendix C.

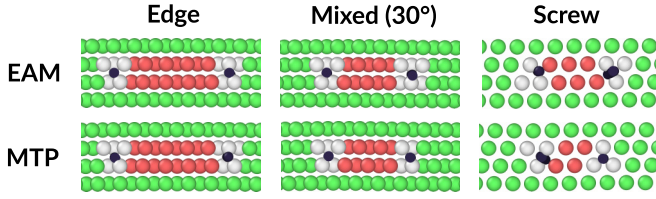| Quantity	MAE	RMSE
Energy (eV/atom)	1.0×10^{-4}	1.9×10^{-4}
Forces (eV/Å)	3.5×10^{-3}	5.9×10^{-3}
Stress (GPa)	2.0×10^{-2}	2.7×10^{-2}

FIG. 10. Visualization of the dislocation cores predicted by EAM and MTP using the common neighbor analysis and the DXA.

TABLE V. DXA partial splitting distances (in Å) for the EAM reference model and a level-16 MTP.

Dislocation	EAM	MTP
Edge	15.2	14.7
Mixed (30°)	11.1	10.7
Screw	8.9	8.1

Promotion Agency (FFG), and the federal states of Styria, Upper Austria, and Tyrol.

APPENDIX A: MTP TRAINING

Suppose we are given a training set $\mathcal{T} = \{\{\boldsymbol{r}_i\}_j, \Pi_j^{\mathrm{qm}}, \{\boldsymbol{f}_i\}_j^{\mathrm{qm}}, \boldsymbol{\sigma}_j^{\mathrm{qm}}\}_{j=1,\ldots,M}$ that contains M atomic configurations $\{\boldsymbol{r}_i\}_j$ and its associated quantum mechanical energies Π_j^{qm}, forces $\{\boldsymbol{f}_i\}_j^{\mathrm{qm}}$, and stresses $\boldsymbol{\sigma}_j^{\mathrm{qm}}$. We then identify the MTP coefficients $\mathcal{C} = \{\theta_\alpha, c_{\mu n}\}$ by minimizing the loss functional,

$$\mathcal{L}(\mathcal{T}, \mathcal{C}) = \sum_{j=1}^{M} \left\{ w_{\mathrm{e}} \left[\Pi_j^{\mathrm{mtp}}(\mathcal{C}) - \Pi_j^{\mathrm{qm}} \right]^2 \right.$$
$$+ w_{\mathrm{f}} \left[\sum_{i=1}^{N} \left\| \boldsymbol{f}_i^{\mathrm{mtp}}(\mathcal{C}) - \boldsymbol{f}_i^{\mathrm{qm}} \right\|^2 \right]$$
$$\left. + w_{\sigma} \left\| \boldsymbol{\sigma}_j^{\mathrm{mtp}}(\mathcal{C}) - \boldsymbol{\sigma}_j^{\mathrm{qm}} \right\|^2 \right\}, \qquad \text{(A1)}$$

with respect to $\mathcal{C}$, with the energies, forces, and stresses weighted as follows:

$$w_{\mathrm{e}} = 1, \quad w_{\mathrm{f}} = 1\text{e-}02, \quad w_{\sigma} = 1\text{e-}03.$$

To minimize (A1), we use SCIPY's BFGS solver. For our initial training, we use a limit of 500 iterations; for retraining during active learning, we use a limit of 200 iterations; and for retraining the potential on the entire training set after gathering all data, we use a limit of 2000 iterations (cf. Sec. III A).

APPENDIX B: MTP PREDICTIONS FOR THE MIXED DISLOCATION

The core structure of the 30°-mixed dislocation predicted by the MTP that has been trained according to Sec. III A is shown in Fig. 8. We have computed the splitting distance of the mixed dislocation by taking the distance between the locations of the two maxima of the screw component of the Nye tensor along the glide direction. Using this method, the splitting distance is 5.4 Å, which is in between the values for the edge (6.6 Å) and the screw (4.6 Å). Using the DXA, the partial splitting of the mixed dislocation is 10.8 Å, which is also in between the values for the edge (15 Å) and the screw (8.6 Å).

The dislocation energy for the mixed dislocation is shown in Fig. 9. As for the partial splitting, the result is in agreement with the dislocation energies for the edge and screw dislocations (cf. Fig. 7).

APPENDIX C: VALIDATION OF THE TRAINING ALGORITHM USING EAM AS A REFERENCE MODEL

For validation, we also ran our training algorithm using the EAM potential of Ercolessi and Adams [77] as a reference model instead of the expensive DFT. Following our training protocol from Sec. III A, we first trained three separate level-16 MTPs on the edge, screw, and mixed dislocations, respectively. Upon convergence, the training sets of the three MTPs contained 22 configurations for the edge dislocation, 33 configurations for the screw dislocation, and 28 configurations for the mixed dislocation. After training the three MTPs, we combine the training data into one big training set that now contains 53 configurations. We then train a level-16 MTP on this combined training set and rerun the simulations with active learning switched off.
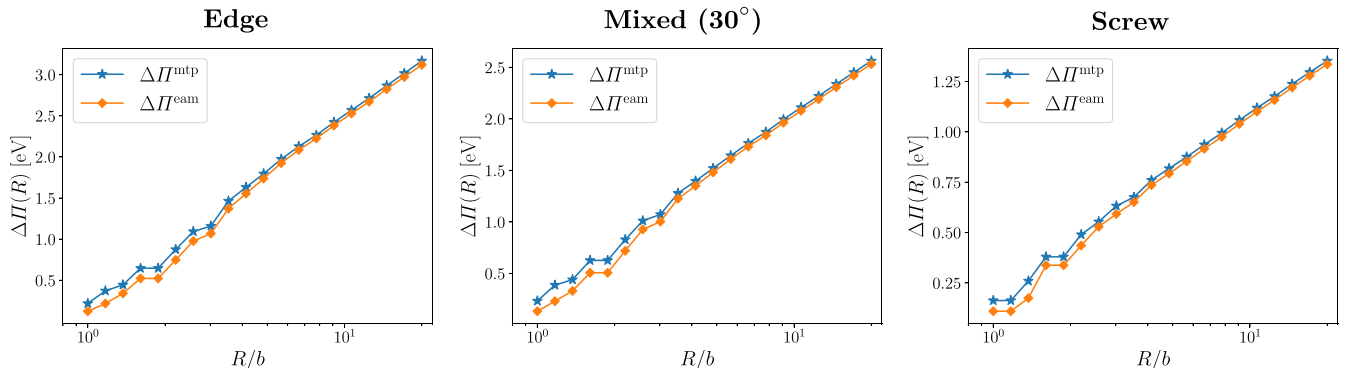

FIG. 11. Dislocation energies for the edge, mixed, and screw dislocations predicted by the MTP and the EAM potential.

The training errors, shown in Table IV, are very low, which is not surprising since an MTP is able to approximate an EAM potential exactly.

The core structures predicted by the MTP and the EAM potential are visualized in Fig. 10 using the common neighbor analysis (CNA) [78]. For the edge and mixed dislocations, there is no visible difference between MTP and EAM and the partial splitting distances coincide (cf. Table V). For the screw dislocation, the MTP core appears to be slightly more narrow than the EAM core, but note that the CNA assignment of structure types is very sensitive to small atomic displacements; so, both cores can be considered to be in good agreement.

The corresponding dislocation energies are shown in Fig. 11. Again, the quantitative agreement between MTP and EAM is very good in the vicinity of the dislocation core; in the far field, they even agree almost exactly. Moreover, the far-field behavior corresponds to the dislocation energy predicted by linear elasticity (dashed lines) up to some constant core energy.

[1] G. Kresse and J. Furthmüller, Efficient iterative schemes for *ab initio* total-energy calculations using a plane-wave basis set, Phys. Rev. B **54**, 11169 (1996).

[2] See https://materialsproject.org.

[3] J. R. K. Bigger, D. A. McInnes, A. P. Sutton, M. C. Payne, I. Stich, R. D. King-Smith, D. M. Bird, and L. J. Clarke, Atomic and electronic structures of the 90° partial dislocation in silicon, Phys. Rev. Lett. **69**, 2224 (1992).

[4] L. Romaner, C. Ambrosch-Draxl, and R. Pippan, Effect of rhenium on the dislocation core structure in tungsten, Phys. Rev. Lett. **104**, 195503 (2010).

[5] J. P. Hirth and J. Lothe, *Theory of Dislocations* (Wiley, New York, 1982).

[6] B. A. Szajewski, A. Hunter, D. J. Luscher, and I. J. Beyerlein, The influence of anisotropy on the core structure of Shockley partial dislocations within FCC materials, Model. Simul. Mater. Sci. Eng. **26**, 015010 (2018).

[7] C. Woodward, D. R. Trinkle, L. G. Hector, and D. L. Olmsted, Prediction of dislocation cores in aluminum from density functional theory, Phys. Rev. Lett. **100**, 045507 (2008).

[8] A. M. Z. Tan, C. Woodward, and D. R. Trinkle, Dislocation core structures in Ni-based superalloys computed using a density functional theory based flexible boundary condition approach, Phys. Rev. Mater. **3**, 033609 (2019).

[9] F. Bianchini, A. Glielmo, J. R. Kermode, and A. De Vita, Enabling QM-accurate simulation of dislocation motion in $\gamma - $Ni and $\alpha - $Fe using a hybrid multiscale approach, Phys. Rev. Mater. **3**, 043605 (2019).

[10] M. Iyer, B. Radhakrishnan, and V. Gavini, Electronic-structure study of an edge dislocation in Aluminum and the role of macroscopic deformations on its energetics, J. Mech. Phys. Solids **76**, 260 (2015).

[11] S. Das and V. Gavini, Electronic structure study of screw dislocation core energetics in aluminum and core energetics informed forces in a dislocation aggregate, J. Mech. Phys. Solids **104**, 115 (2017).

[12] Y. Dan and D. R. Trinkle, First-principles core energies of isolated basal and prism screw dislocations in magnesium, Mater. Res. Lett. **10**, 360 (2022).

[13] P. Motamarri, S. Das, S. Rudraraju, K. Ghosh, D. Davydov, and V. Gavini, DFT-FE—A massively parallel adaptive finite-element code for large-scale density functional theory calculations, Comput. Phys. Commun. **246**, 106853 (2020).

[14] S. Das, B. Kanungo, V. Subramanian, G. Panigrahi, P. Motamarri, D. Rogers, P. Zimmerman, and V. Gavini, Large-scale materials modeling at quantum accuracy: Ab Initio simulations of quasicrystals and interacting extended defects in metallic alloys, in *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis* (ACM, New York, 2023), pp. 1–12.

[15] J. Behler and M. Parrinello, Generalized neural-network representation of high-dimensional potential-energy surfaces, Phys. Rev. Lett. **98**, 146401 (2007).

[16] A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons, Phys. Rev. Lett. **104**, 136403 (2010).

[17] A. Thompson, L. Swiler, C. Trott, S. Foiles, and G. Tucker, Spectral neighbor analysis method for automated generation of quantum-accurate interatomic potentials, J. Comput. Phys. **285**, 316 (2015).

[18] A. V. Shapeev, Moment tensor potentials: A class of systematically improvable interatomic potentials, Multiscale Model. Simul. **14**, 1153 (2016).

[19] K. T. Schütt, P.-J. Kindermans, H. E. Sauceda, S. Chmiela, A. Tkatchenko, and K.-R. Müller, SchNet: A continuous-filter convolutional neural network for modeling quantum interactions, in *Proceedings of the 31st International Conference on Neural Information Processing Systems* (Curran Associates Inc., Red Hook, NY, 2017), pp. 992–1002.

[20] J. S. Smith, O. Isayev, and A. E. Roitberg, ANI-1: An extensible neural network potential with DFT accuracy at force field computational cost, Chem. Sci. **8**, 3192 (2017).

[21] G. P. P. Pun, R. Batra, R. Ramprasad, and Y. Mishin, Physically informed artificial neural networks for atomistic modeling of materials, Nat. Commun. **10**, 2339 (2019).

[22] R. Drautz, Atomic cluster expansion for accurate and transferable interatomic potentials, Phys. Rev. B **99**, 014104 (2019).

[23] S. Batzner, A. Musaelian, L. Sun, M. Geiger, J. P. Mailoa, M. Kornbluth, N. Molinari, T. E. Smidt, and B. Kozinsky, E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials, Nat. Commun. **13**, 2453 (2022).

[24] S. Takamoto, S. Izumi, and J. Li, TeaNet: Universal neural network interatomic potential inspired by iterative electronic relaxations, Comput. Mater. Sci. **207**, 111280 (2022).

[25] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, Neural message passing for quantum chemistry, in *Proceedings of the 34th International Conference on Machine Learning - Volume 70* (ACM Digital Library, New York, 2017), pp. 1263–1272.

[26] J. Behler and G. Csányi, Machine learning potentials for extended systems: A perspective, Eur. Phys. J. B **94**, 142 (2021).

[27] M. Stricker and W. A. Curtin, Prismatic slip in magnesium, J. Phys. Chem. C **124**, 27230 (2020).

[28] M. Poul, L. Huber, E. Bitzek, and J. Neugebauer, Systematic atomic structure datasets for machine learning potentials: Application to defects in magnesium, Phys. Rev. B **107**, 104103 (2023).

[29] J. A. Meziere, Y. Luo, Y. Xia, L. K. Béland, M. R. Daymond, and G. L. W. Hart, Accelerating training of MLIPs through small-cell training, J. Mater. Res. **38**, 5095 (2023).

[30] C. Ortner and Y. Wang, A framework for a generalization analysis of machine-learned interatomic potentials, Multiscale Model. Simul. **21**, 1053 (2023).

[31] Y. Wang, S. Patel, and C. Ortner, A theoretical case study of the generalization of machine-learned potentials, Comput. Meth. Appl. Mech. Eng. **422**, 116831 (2024).

[32] D. Dragoni, T. D. Daff, G. Csányi, and N. Marzari, Achieving DFT accuracy with a machine-learning interatomic potential: Thermomechanics and defects in bcc ferromagnetic iron, Phys. Rev. Mater. **2**, 013808 (2018).

[33] G. P. P. Pun, V. Yamakov, J. Hickman, E. H. Glaessgen, and Y. Mishin, Development of a general-purpose machine-learning interatomic potential for aluminum by the physically informed neural network method, Phys. Rev. Mater. **4**, 113807 (2020).

[34] A. M. Goryaeva, J. Dérès, C. Lapointe, P. Grigorev, T. D. Swinburne, J. R. Kermode, L. Ventelon, J. Baima, and M.-C. Marinica, Efficient and transferable machine learning potentials for the simulation of crystal defects in bcc Fe and W, Phys. Rev. Mater. **5**, 103803 (2021).

[35] D. Marchand and W. A. Curtin, Machine learning for metallurgy IV: A neural network potential for Al-Cu-Mg and Al-Cu-Mg-Zn, Phys. Rev. Mater. **6**, 053803 (2022).

[36] F. Deng, H. Wu, R. He, P. Yang, and Z. Zhong, Large-scale atomistic simulation of dislocation core structure in face-centered-cubic metal with Deep Potential method, Comput. Mater. Sci. **218**, 111941 (2023).

[37] M. R. Fellinger, A. M. Z. Tan, L. G. Hector, Jr., and D. R. Trinkle, Geometries of edge and mixed dislocations in bcc Fe from first-principles calculations, Phys. Rev. Mater. **2**, 113605 (2018).

[38] M. Hodapp and A. Shapeev, *In operando* active learning of interatomic interaction during large-scale simulations, Mach. Learn.: Sci. Technol. **1**, 045005 (2020).

[39] K. Gubaev, E. V. Podryabinkin, G. L. Hart, and A. V. Shapeev, Accelerating high-throughput searches for new alloys with active learning of interatomic potentials, Comput. Mater. Sci. **156**, 148 (2019).

[40] B. Settles, Active learning literature survey, http://digital.library.wisc.edu/1793/60660.

[41] J. Behler, Representing potential energy surfaces by high-dimensional neural network potentials, J. Phys.: Condens. Matter **26**, 183001 (2014).

[42] L. Zhang, D.-Y. Lin, H. Wang, R. Car, and W. E, Active learning of uniformly accurate inter-atomic potentials for materials simulation, Phys. Rev. Mater. **3**, 023804 (2019).

[43] E. V. Podryabinkin and A. V. Shapeev, Active learning of linearly parametrized interatomic potentials, Comput. Mater. Sci. **140**, 171 (2017).

[44] R. Jinnouchi, F. Karsai, and G. Kresse, On-the-fly machine learning force field generation: Application to melting points, Phys. Rev. B **100**, 014105 (2019).

[45] J. Vandermause, S. B. Torrisi, S. Batzner, Y. Xie, L. Sun, A. M. Kolpak, and B. Kozinsky, On-the-fly active learning of interpretable Bayesian force fields for atomistic rare events, npj Comput. Mater. **6**, 20 (2020).

[46] S. A. Goreinov, I. V. Oseledets, D. V. Savostyanov, E. E. Tyrtyshnikov, and N. L. Zamarashkin, *How to Find a Good Submatrix* (World Scientific, Singapore, 2010), pp. 247–256.

[47] I. S. Novikov, K. Gubaev, E. V. Podryabinkin, and A. V. Shapeev, The MLIP package: Moment tensor potentials with MPI and active learning, Mach. Learn.: Sci. Technol. **2**, 025002 (2021).

[48] E. V. Podryabinkin, A. G. Kvashnin, M. Asgarpour, I. I. Maslenikov, D. A. Ovsyannikov, P. B. Sorokin, M. Y. Popov, and A. V. Shapeev, Nanohardness from first principles with active learning on atomic environments, J. Chem. Theory Comput. **18**, 1109 (2022).

[49] L. C. Erhard, J. Rohrer, K. Albe, and V. L. Deringer, Modelling atomic and nanoscale structure in the silicon–oxygen system through active machine learning, Nat. Commun. **15**, 1927 (2024).

[50] C. Woodward and S. I. Rao, Flexible *ab initio* boundary conditions: Simulating isolated dislocations in bcc Mo and Ta, Phys. Rev. Lett. **88**, 216402 (2002).

[51] L. Zhang, G. Csányi, E. van der Giessen, and F. Maresca, Atomistic fracture in bcc iron revealed by active learning of Gaussian approximation potential, npj Comput. Mater. **9**, 217 (2023).

[52] C. L. Kelchner, S. J. Plimpton, and J. C. Hamilton, Dislocation nucleation and defect structure during surface indentation, Phys. Rev. B **58**, 11085 (1998).

[53] A. Stukowski, V. V. Bulatov, and A. Arsenlis, Automated identification and indexing of dislocations in crystal interfaces, Model. Simul. Mater. Sci. Eng. **20**, 085007 (2012).

[54] A. Stukowski, Visualization and analysis of atomistic simulation data with OVITO—The open visualization tool, Model. Simul. Mater. Sci. Eng. **18**, 015012 (2010).

[55] E. Bitzek, P. Koskinen, F. Gähler, M. Moseler, and P. Gumbsch, Structural relaxation made simple, Phys. Rev. Lett. **97**, 170201 (2006).

[56] A. Hjorth Larsen, J. Jørgen Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dułak, J. Friis, M. N. Groves, B. Hammer, C. Hargus, E. D. Hermes, P. C. Jennings, P. Bjerre Jensen, J. Kermode, J. R. Kitchin, E. Leonhard Kolsbjerg, J. Kubal, K. Kaasbjerg, S. Lysgaard, J. Bergmann Maronsson, T. Maxson, T. Olsen, L. Pastewka, A. Peterson, C. Rostgaard, J. Schiøtz, O. Schütt, M. Strange, K. S. Thygesen, T. Vegge, L. Vilhelmsen, M. Walter, Z. Zeng, and K. W. Jacobsen, The atomic simulation environment—a Python library for working with atoms, J. Phys.: Condens. Matter **29**, 273002 (2017).

[57] Y. Zuo, C. Chen, X. Li, Z. Deng, Y. Chen, J. Behler, G. Csányi, A. V. Shapeev, A. P. Thompson, M. A. Wood, and S. P. Ong, Performance and cost assessment of machine learning interatomic potentials, J. Phys. Chem. A **124**, 731 (2020).

[58] P. E. Blöchl, Projector augmented-wave method, Phys. Rev. B **50**, 17953 (1994).

[59] G. Kresse and D. Joubert, From ultrasoft pseudopotentials to the projector augmented-wave method, Phys. Rev. B **59**, 1758 (1999).

[60] J. P. Perdew, K. Burke, and M. Ernzerhof, Generalized gradient approximation made simple, Phys. Rev. Lett. **77**, 3865 (1996).

[61] J. P. Perdew and A. Zunger, Self-interaction correction to density-functional approximations for many-electron systems, Phys. Rev. B **23**, 5048 (1981).

[62] L. Goodwin, R. J. Needs, and V. Heine, A pseudopotential total energy study of impurity-promoted intergranular embrittlement, J. Phys.: Condens. Matter **2**, 351 (1990).

[63] Y. A. Wang, N. Govind, and E. A. Carter, Orbital-free kinetic-energy density functionals with a density-dependent kernel, Phys. Rev. B **60**, 16350 (1999).

[64] K. M. Carling and E. A. Carter, Orbital-free density functional theory calculations of the properties of Al, Mg and Al Mg crystalline phases, Model. Simul. Mater. Sci. Eng. **11**, 339 (2003).

[65] C. S. Hartley and Y. Mishin, Representation of dislocation cores using Nye tensor distributions, Mater. Sci. Eng.: A **400-401**, 18, (2005).

[66] V. Vitek, Theory of the core structures of dislocations in body-centered-cubic metals, Cryst. Lattice Defects **5**, 1 (1974).

[67] M. Dewald and W. A. Curtin, Analysis and minimization of dislocation interactions with atomistic/continuum interfaces, Model. Simul. Mater. Sci. Eng. **14**, 497 (2006).

[68] D. L. Olmsted, K. Y. Hardikar, and R. Phillips, Lattice resistance and Peierls stress in finite size atomistic dislocation simulations, Model. Simul. Mater. Sci. Eng. **9**, 215 (2001).

[69] G. Lu, E. B. Tadmor, and E. Kaxiras, From electrons to finite elements: A concurrent multiscale approach for metals, Phys. Rev. B **73**, 024108 (2006).

[70] M. Hodapp, G. Anciaux, J.-F. Molinari, and W. Curtin, Coupled atomistic/discrete dislocation method in 3D Part II: Validation of the method, J. Mech. Phys. Solids **119**, 1 (2018).

[71] J. Vallin, M. Mongy, K. Salama, and O. Beckman, Elastic constants of aluminum, J. Appl. Phys. **35**, 1825 (1964).

[72] G. P. M. Leyson, W. A. Curtin, L. G. Hector, and C. F. Woodward, Quantitative prediction of solute strengthening in aluminium alloys, Nat. Mater. **9**, 750 (2010).

[73] F. Moitzi, L. Romaner, M. Hodapp, A. V. Ruban, and O. E. Peil, Ab initio framework for deciphering trade-off relationships in multi-component alloys, arXiv:2311.12642.

[74] A. S. Kotykhov, K. Gubaev, M. Hodapp, C. Tantardini, A. V. Shapeev, and I. S. Novikov, Constrained DFT-based magnetic machine-learning potentials for magnetic alloys: A case study of Fe-Al, Sci. Rep. **13**, 19728 (2023).

[75] See http://www.materialscloud.org.

[76] F. N. Jalolov, E. V. Podryabinkin, A. R. Oganov, A. V. Shapeev, and A. G. Kvashnin, Mechanical properties of single and poly-crystalline solids from machine learning, arXiv:2309.15868.

[77] F. Ercolessi and J. B. Adams, Interatomic potentials from first-principles calculations: The force-matching method, Europhys. Lett. **26**, 583 (1994).

[78] J. Dana. Honeycutt, and H. C. Andersen, Molecular dynamics study of melting and freezing of small Lennard-Jones clusters, J. Phys. Chem. **91**, 4950 (1987).