LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

LLNL-JRNL-505013

Structure identification methods for atomistic simulations of crystalline materials

A. Stukowski

October 12, 2011

Modelling and Simulation in Materials Science and Engineering

Disclaimer

This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes.

Structure identification methods for atomistic simulations of crystalline materials

Alexander Stukowski

Lawrence Livermore National Laboratory, Livermore, CA 94550, USA

E-mail: alexander@stukowski.de

Abstract. We discuss existing and new computational analysis techniques to classify local atomic arrangements in large-scale atomistic computer simulations of crystalline solids. This article includes a performance comparison of typical analysis algorithms such as *Common Neighbor Analysis*, *Centrosymmetry Analysis*, *Bond Angle Analysis*, *Bond Order Analysis*, and *Voronoi Analysis*. In addition we propose a simple extension to the *Common Neighbor Analysis* method that makes it suitable for multi-phase systems. Finally, we introduce a new structure identification algorithm, the *Neighbor Distance Analysis*, that is designed to identify atomic structure units in grain boundaries.

This paper has been published in *Modelling and Simulation in Materials Science and Engineering* 20 (2012), 045021.

1. Introduction

Atomistic simulation methods such as molecular dynamics (MD), molecular statics, and Monte Carlo schemes are routinely used to study crystalline materials at the atomic scale. In many cases crystal defects play a critical role in materials behavior, and their identification in the simulation data is essential for the understanding of materials properties. Classical atomistic simulation models, however, do not keep track of crystal defects explicitly. These models are governed by a Hamiltonian or other rules which determine the trajectories of individual particles. Therefore, crystal defects and defect-free crystal regions must be recovered from the generated particle-position datasets in a post-processing step to enable the interpretation of simulation results.

For this purpose, many computational analysis methods have been developed in the past. Their task is to assign a structural type to each atom or particle based on an analysis of its local environment. Most such methods attempt to match a local structure to an idealized one (such as fcc or bcc), and measure how closely they fit. This information can then be used to color particles for visualization purposes or to quantify the occurrence of different crystalline phases and defects in a simulation. Another important application is filtering the simulation data on the fly to reduce it to a manageable amount, e.g. by storing only particles with an atypical environment.

Our goal is to give an overview of current computational analysis techniques, as they are offered by many visualization tools and simulation codes, and as they are employed in many recent simulation studies described in the literature. In particular we will review the most commonly used structure characterization methods for molecular dynamics simulations of crystalline solids:

- (a) energy filtering,
- (b) centrosymmetry parameter analysis (CSP) [1],
- (c) bond order analysis [2],
- (d) common neighbor analysis (CNA) [3],
- (e) bond angle analysis (BAA) [4], and
- (f) Voronoi analysis.

In addition to describing the respective strengths and weaknesses of these methods, we introduce two new identification schemes:

- (g) adaptive common neighbor analysis (a-CNA) and
- (h) neighbor distance analysis (NDA).

The adaptive CNA is a simple extension of the common neighbor analysis method to improve the characterization of multi-phase systems. The computationally more expensive NDA is targeted at the classification of complex structural environments as they occur inside crystal defects such as grain boundaries.

As part of this paper we have implemented all discussed algorithms for benchmarking purposes. We provide the source code of this analysis tool at http://asa.ovito.org/ as a reference, to facilitate the use of the described techniques, and to foster their advancement by the research community.

2. Existing analysis methods

Here, we focus on analysis techniques for simulation studies of crystalline solids only. A broader overview of structural characterization methods and shape matching algorithms for general particle systems has recently been given by Keys et al. [5].

2.1. General considerations

One can name several features that an ideal structure characterization technique should provide:

- Accuracy - The method should be able to correctly distinguish several structural environments solely based on the local arrangement of atoms and independent of the crystal orientation (rotational and translational invariance).
- Robustness - The algorithm should assign a local structure to most particles in the system and avoid errors arising from small displacements of particles from their equilibrium or symmetry positions.
- Computational efficiency - Since the local structure characterization needs to be performed for every particle in a system, and possibly at high frequency as part of an on-the-fly analysis, the computational cost is an important factor.
- Simplicity - Because wide-spread use of a method requires an algorithm that is easy to implement and understand.
- Universality - Ideally, the set of reference structures recognized by the method is not hard-coded into the algorithm and can be easily extended by the user.

Note that the first two requirements are in conflict with each other: A low sensitivity to atomic displacements usually comes at the price of a reduced capability of the identification method to distinguish similar structures. Some of the methods discussed here allow the user to explicitly control this tradeoff between accuracy and robustness. In general one wants to avoid any wrong classifications, i.e. *false positives* as well as *false negatives*, in the structure recognition process.

The techniques discussed in this paper can be divided into two sets. Methods (a)-(c) quantify the similarity of a given atomic arrangement to a particular reference structure. A positive match is detected by comparing the computed similarity measure to a threshold chosen by the user. A high threshold increases the robustness (and the chance of false positives) while a low threshold increases the sensitivity (and the chance of false negatives). The aim of the second group of methods is to distinguish between several reference structures and to uniquely assign a type to each particle in the system (with the possibility of assigning no type at all if the local atomic arrangement deviates too much from all of the reference structures). These structure identification methods are usually based on a discrete signature that is calculated from the particle positions, and which identifies the arrangement unambiguously.

2.2. Energy filtering

The potential energy of an atom can be used as a simple indicator to decide whether it forms a perfect lattice with its neighbors. Given that atoms which are part of a crystal defect are usually higher in energy than the perfect lattice (the ground state), one can detect defective atoms by using a simple threshold criterion: Atoms having

a potential energy above the threshold are considered defect atoms, while low-energy atoms are classified as regular crystalline atoms.

Several shortcomings of this method have contributed to the fact that it is rarely used nowadays. The atomic energy levels of perfect lattice atoms and metastable defects can easily overlap due to degeneracies, elastic strain energy, or thermal energy. Then the discrimination between the different structural states becomes impossible. Moreover, the potential energy of individual atoms is specific to the employed interaction model, and, for quantum mechanical descriptions and some interatomic potentials, is not defined at all. This is why one prefers purely structural analysis methods, which characterize the spatial arrangement of atoms without reference to the interatomic interaction laws.

2.3. Centrosymmetry parameter

The centrosymmetry property of some lattices (e.g. fcc and bcc) can be used to distinguish them from other structures such as crystal defects where the local bond symmetry is broken. Kelchner et al. [1] have developed a metric, the so-called centrosymmetry parameter (CSP), that quantifies the local loss of centrosymmetry at an atomic site, which is characteristic for most crystal defects. The CSP of an atom having N nearest neighbors is defined as

$$\mathrm{CSP} = \sum_{i=1}^{N/2} \left|\mathbf{r}_i + \mathbf{r}_{i+N/2}\right|^2 \tag{1}$$

where $\mathbf{r}_i$ and $\mathbf{r}_{i+N/2}$ are vectors from the central atom to a pair of opposite neighbors. Practical ways of finding these pairs are described in [6] and in the accompanying documentation of the visualization program AtomEye [7] and the molecular dynamics code LAMMPS [8]. The latter uses the following scheme: There are $N(N-1)/2$ possible neighbor pairs (i, j) that can contribute to above formula. The sum of two bond vectors, $|\mathbf{r}_i + \mathbf{r}_j|^2$, is computed for each, and only the $N/2$ smallest are actually used to compute the CSP. For centrosymmetric lattice sites, they will be pairs of neighbor atoms in symmetrically opposite positions with respect to the central atom; hence the $i + N/2$ notation in formula 1. The CSP is close to zero for regular sites of a centrosymmetric crystal and becomes non-zero for defect atoms. The number of nearest neighbors taken into account is $N = 12$ for fcc and $N = 8$ for bcc.

The main advantage of the CSP is that it is only marginally affected by elastic distortions of the crystal. In particular, any affine deformation of the lattice does not change its degree of centrosymmetry at all. The CSP is, however, sensitive to random thermal displacements of atoms. Being only a scalar measure, the CSP's capability of discriminating between different defect types is rather weak. The noise induced by thermal displacements and inhomogeneous elastic strain may well dominate any characteristic differences between defect structures. Most notably, the method can only be applied to the class of centrosymmetric lattices (which does not include hcp, for example), and it provides no means of distinguishing multiple centrosymmetric crystal phases.

The user needs to choose a proper threshold to distinguish defect atoms from perfect lattice atoms. At elevated temperatures the distribution of CSP values in a perfect crystal becomes broader, and may begin to overlap with the characteristic range of crystal defects. To show this, we have measured the CSP distributions in a

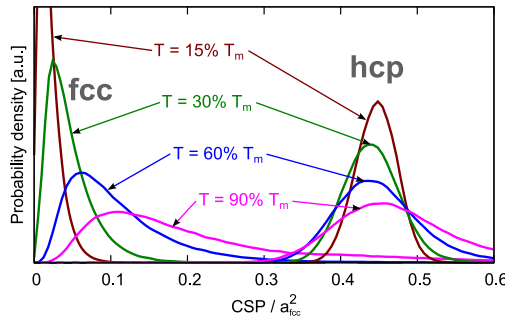

Figure 1. Distribution of CSP values (normalized by the square of the lattice parameter) measured at various homologous temperatures (in a Cu crystal). The CSP values have been sampled from atoms located in the defect-free fcc lattice and in a fcc stacking fault (corresponding to an hcp-like arrangement of neighbors with broken centrosymmetry).

perfect fcc Cu crystal and inside an intrinsic stacking fault at various temperatures (Fig. 1). While at low homologous temperature the two distributions are well-separated, the identification of perfect fcc atoms becomes less reliable at high temperature (above 60% of the melting temperature). Note that other crystal defects such as Shockley partial dislocations exhibit CSP values lower than those of stacking fault atoms.

2.4. Bond order analysis

Given a central atom, we can project its near neighbor bonds to a unit sphere. Based on these projected vectors one can define a set of local bond order parameters, also known as *Steinhardt order parameters* [2], that are rotationally invariant combinations of spherical harmonics. The bond order parameters exhibit characteristic values for each crystal structure, allowing us to discriminate between them.

Given the N neighbor vectors $\mathbf{r}_1 \ldots \mathbf{r}_N$ of a central atom, the l-th order parameter is defined according to Steinhardt as

$$Q_l = \sqrt{\frac{4\pi}{2l+1} \sum_{m=-l}^{+l} |q_{lm}|^2} \tag{2}$$

with

$$q_{lm} = \frac{1}{N} \sum_{i=1}^{N} Y_l^m(\mathbf{r}_i). \tag{3}$$

Here, the complex functions $Y_l^m(\mathbf{r}) = Y_l^m(\theta, \varphi)$ are the spherical harmonics, whose evaluation is computationally expensive. Note that the set of local bond order parameters, $\{Q_l\}$ with $l = 1, 2, 3, \ldots$, is invariant under rotations of the coordinate system (meaning that it is independent of the crystal's orientation). Bond order parameters up to $l = 3$ are zero for lattices with cubic symmetry, and one usually takes into account the values of Q_4 and Q_6 to discriminate between fcc, hcp and bcc phases. ‡

‡ One obtains the following reference values for the perfect lattices: $Q_4^{\mathrm{fcc}} = 0.191$, $Q_6^{\mathrm{fcc}} = 0.575$ and $Q_4^{\mathrm{hcp}} = 0.097$, $Q_6^{\mathrm{hcp}} = 0.485$ and $Q_4^{\mathrm{bcc}} = 0.036$, $Q_6^{\mathrm{bcc}} = 0.511$.

The parameter set (Q_4, Q_6) can be used to measure the structural order of a particle system when averaged over all atoms. Hence, bond order parameters are often used in computational studies of crystallization to determine the fractions of crystalline and liquid phases. Crystal deformation and thermal fluctuations, however, smear out the order parameter distributions [9]. Thus, to assign a particular structure type to a particle, one needs to define non-overlapping regions in the Q_4-Q_6 parameter plane [10] for all crystal phases considered. The choice of these regions is arbitrary, and, to our knowledge, no generally accepted scheme for the classification of bond order parameters exists so far.

2.5. Common neighbor analysis

Structure analysis methods that employ more complex, high-dimensional signatures to characterize arrangements of atoms are usually better in discriminating between several structures. A popular method of this type is the common neighbor analysis (CNA) [3, 11]. Unlike the CSP and the local bond order parameters, the CNA does not directly take into account the spatial vectors pointing from the central atom to its neighbor. Instead, a characteristic signature is computed from the topology of bonds that connect the surrounding neighbor atoms.

Usually, two atoms are said to be (near-)neighbors, or bonded, if they are within a specified cutoff distance r_{cut} of each other. For densely packed structures (fcc and hcp) the cutoff distance is set to be halfway between the first and second neighbor shell, giving for fcc

$$r_{\mathrm{cut}}^{\mathrm{fcc}} = \frac{1}{2}\left(\sqrt{1/2}+1\right) a_{\mathrm{fcc}} \simeq 0.854\, a_{\mathrm{fcc}}, \tag{4}$$

where a_{fcc} is the lattice constant of the crystal structure. For the bcc lattice, two neighbor shells need to be taken into account, and atoms are considered to be bonded with their first- and second-nearest neighbors:

$$r_{\mathrm{cut}}^{\mathrm{bcc}} = \frac{1}{2}\left(1+\sqrt{2}\right) a_{\mathrm{bcc}} \simeq 1.207\, a_{\mathrm{bcc}}. \tag{5}$$

To assign a local crystal structure to an atom, three characteristic numbers are computed for each of the N neighbor bonds of the central atom: The number of neighbor atoms the central atom and its bonded neighbor have in common, n_{cn}; the total number of bonds between these common neighbors, n_b; and the number of bonds in the longest chain of bonds connecting the common neighbors, n_{lcb}. This yields N triplets (n_{cn}, n_b, n_{lcb}), which are compared to a set of reference signatures to assign a structural type to the central atom (Table 1).

fcc ($N = 12$)	hcp ($N = 12$)	bcc ($N = 14$)	cubic diamond ($N = 16$)
12 × (421)	6 × (421)	8 × (666)	12 × (543)
	6 × (422)	6 × (444)	4 × (663)

Table 1. CNA signatures of common crystal structures. For example: An hcp-coordinated atom has six bonds of (421) type and six of (422) type. That is, any two near-neighbors in a hcp crystal have exactly four common neighbors, which are interconnected by two bonds. And the longest continuous chain these two bonds form is either of length one or two (in six cases each).

The *common neighborhood parameter* [12] should be mentioned as an alternative analysis method, which was proposed by Tsuzuki et al. to combine the strengths of both the CNA and CSP methods. The CNA has also been extended to binary atomic systems by taking the chemical species of common neighbors into account as an additional criterion [13]. This extension enables the identification of simple binary structures such as $L1_0$, $L1_2$ etc.

2.6. Bond angle analysis

The bond angle analysis has been developed by Ackland and Jones [4] to distinguish fcc, hcp and bcc coordination structures. From the N bond vectors of the central atom, an eight-bin histogram of the $N(N-1)/2$ bond angle cosines, $\cos\theta_{ijk}$, is computed first. Here, θ_{ijk} denotes the angle formed by the central atom i, and two of its neighbors, j and k. The obtained histogram is then further evaluated using a set of heuristic decision rules to determine the most likely structure type. These rules have been optimized by the authors such that a robust discrimination of the most important crystal structures is archived. The number of neighbors used to calculate the bond angle distribution is determined adaptively by employing a cutoff radius that is proportional to the average distance of the six nearest neighbors.

2.7. Voronoi analysis

The Voronoi decomposition [14] can serve as a geometric method to determine the near neighbors of a particle (i.e. its coordination number) by considering the faces of the Voronoi polyhedron enclosing the particle. Furthermore, the geometric shape of the Voronoi polyhedron reflects the characteristic arrangement of near neighbors. For this reason the Voronoi decomposition has been employed in simulation studies of liquids and glasses to analyze various properties of their atomic structure [15, 16].

To effectively characterize the arrangement of near neighbors, the computed Voronoi polyhedron for a particle is translated into a compact signature by counting the number of polygonal facets having three, four, five and six vertices/edges. This yields a vector of four integers, (n_3, n_4, n_5, n_6), that identifies the structural type. For instance the Voronoi polyhedron of an fcc lattice atom is equivalent to the fcc Wigner-Seitz cell and comprises 12 facets with four vertices each. Thus the corresponding signature for fcc is (0,12,0,0). The polyhedron of a bcc atom has facets with four and six vertices, and the corresponding signature is (0,6,0,8).

Even though the Voronoi method has been used numerous times for the analysis of particle systems without long-range order such as liquids and glasses, it has rarely been applied to simulations of crystalline materials. One reason is that singular Voronoi vertices, which are adjacent to more than three facets, and which occur in the Voronoi decomposition of some highly symmetric crystalline packings such as fcc and hcp, will dissociate into multiple vertices as soon as the atomic coordinates are only slightly perturbed. This dramatically changes the Voronoi polyhedra and the computed signatures [17], making the identification of such crystal structures nontrivial.

In our implementation we use the following approach to mitigate the problem of singularities in the Voronoi decomposition of fcc crystals: First, the conventional Voronoi polyhedra are constructed (using the VORO++ code library [18]). When counting the number of edges of a Voronoi facet, we skip edges which are shorter than

a certain threshold. Thus, a singular vertex, which may have dissociated into multiple vertices due to perturbations, will still be counted as one. Small facets with less than three edges above the threshold are completely ignored. The edge threshold is set to 30% of the polyhedron's maximum radius.

It should be pointed out that the described sensitivity of the Voronoi method to perturbations of the atomic coordinates, in addition to the high computational cost of the Voronoi polyhedron construction, render its application to crystalline systems rather unattractive. Remarkably, the Voronoi method, in its simplest form described here, is not capable of discriminating hcp-coordinated atoms from fcc atoms because the corresponding Voronoi polyhedra have a (0,12,0,0) signature in both cases.

3. New analysis methods

We now describe two new methods, which can provide superior analysis results for some applications. The *adaptive common neighbor analysis* is a simple extension of the standard CNA method, which adds some convenience on the user's side and the ability to analyze multi-phase systems. The *neighbor distance analysis* (NDA), in contrast, is a completely new algorithm that employs a more complex signature to identify a wider range of atomic arrangements.

3.1. Adaptive common neighbor analysis

The common neighbor analysis method described in section 2.5 is one of the most frequently used structure identification methods for atomistic simulation studies of fcc, hcp, and bcc crystal plasticity. It provides efficient and unambiguous classification of local atomic arrangements, making it possible to effectively distinguish crystal defects from undisturbed lattice atoms. The only parameter required is the cutoff radius, which determines the maximum separation of near-neighbors, and which must be chosen according to the crystal phase under consideration (cf. Eqs. 4 and 5).

In the case of multi-phase systems, however, the choice of the cutoff parameter is no longer well-defined. In many cases, for instance a fcc-bcc bicrystal simulation, one cannot specify a global cutoff radius that fits all phases equally well. We therefore propose to pick the cutoff radius individually for each atom and in dependence of the reference structure we want to compare it to. We implement this approach, which we refer to as *adaptive common neighbor analysis* (a-CNA), as follows.

Given a central atom to be analyzed, we first generate the list of $N_{\max}$ *nearest* neighbors and sort it by distance. $N_{\max}$ is the maximum required number of neighbors for all considered reference structures, e.g. $N_{\max} = 16$ for the set of structures listed in Table 1. One can generate such a nearest neighbor list either by means of a k-d tree data structure [19] and a recursive k-th nearest neighbor query algorithm [20], or by sorting a pre-existing neighbor list that has already been generated on the basis of an excessively large cutoff radius (for instance, to compute the interatomic forces in a molecular dynamics simulation).

To test whether the local coordination structure matches an fcc crystal, we take only the first $N_{\mathrm{fcc}} = 12$ entries from the sorted neighbor list. The average distance of these 12 nearest neighbors provides a local length scale, analogous to the approach used in the bond angle analysis. That is, we can define a local cutoff radius, which is specific to the current atom and used for matching with the fcc reference structure

only:

$$r_{\mathrm{cut}}^{\mathrm{local}}(\mathrm{fcc}) = \frac{1+\sqrt{2}}{2} \cdot \frac{\sum_{j=1}^{12} |\mathbf{r}_j|}{12}. \tag{6}$$

The local cutoff is subsequently used to determine the "bonding" between the 12 nearest neighbors and to compute the CNA signature as usual. If the signature does not conform to fcc, the algorithm proceeds with testing against the next candidate structure. For the bcc structure, for instance, the 14 nearest neighbors must be taken into account and a local cutoff is computed as

$$r_{\mathrm{cut}}^{\mathrm{local}}(\mathrm{bcc}) = \frac{1+\sqrt{2}}{2} \left[\frac{2}{\sqrt{3}} \cdot \frac{\sum_{j=1}^{8} |\mathbf{r}_j|}{8} + \frac{\sum_{j=9}^{14} |\mathbf{r}_j|}{6} \right]. \tag{7}$$

Here, the local length scale is determined from the eight nearest neighbors in the sorted neighbor list (forming the first shell) and the successive six neighbors (forming the second shell). Their average distances are weighted accordingly to yield the local cutoff radius that lies halfway between the second and third bcc coordination shell.

The computational cost of the adaptive CNA increases with the number of reference structures to be tested. In practice, however, the analysis is only 25% more expensive than the standard CNA when identifying fcc, hcp, and bcc atoms. This cost is in most cases offset by the convenience of a parameterless method (no cutoff radius) and the superior analysis results provided. To demonstrate the strength of the adaptive CNA we have applied it to a simulation of the Fe-Cu multi-phase alloy. In the simulation study, a combination of Monte Carlo sampling (variance-constrained semi-grandcanonical ensemble [21]) and conventional MD time integration was used to determine the equilibrium structure of Cu-rich precipitates in a Fe-rich bcc matrix. The equilibrium distribution of Cu atoms at a prescribed temperature is found via Monte Carlo transmutation steps, while alternating MD steps allow the positional degrees of freedom to relax simultaneously. This enables structural phase transformations to occur in the simulation. Starting off from a random distribution of Cu atoms in the bcc-Fe matrix, the Cu atoms precipitate to form a spherical particle. At certain conditions, the crystal structure of the cluster changes from bcc to a multiply-twinned 9R structure (herringbone structure) [22] as shown in figure 2. The system has been quenched to zero temperature to remove thermal displacements prior to the structure analysis.

The results of the conventional common neighbor analysis strongly depend on the cutoff parameter used. A cutoff that is suitable for identifying the 9R phase is given by Eq. 4 and the lattice constant of fcc-Cu, while for the identification of the bcc-Fe phase one would apply formula 5. Varying the cutoff between these limiting cases lets the observed bcc-9R interface slide and makes the precipitate appear smaller or larger. In all cases, the CNA will be unable to assign a structural type to the atoms right at the interface since their coordination does not match to either of the reference structures. The adaptive CNA overcomes this problem by computing a cutoff radius for each atom and taking into account the local dilatation. Virtually every atom in the bcc-9R interface is identified as crystalline by the a-CNA, giving even slightly better analysis results than the bond angle analysis, which was specifically designed for applications like this.

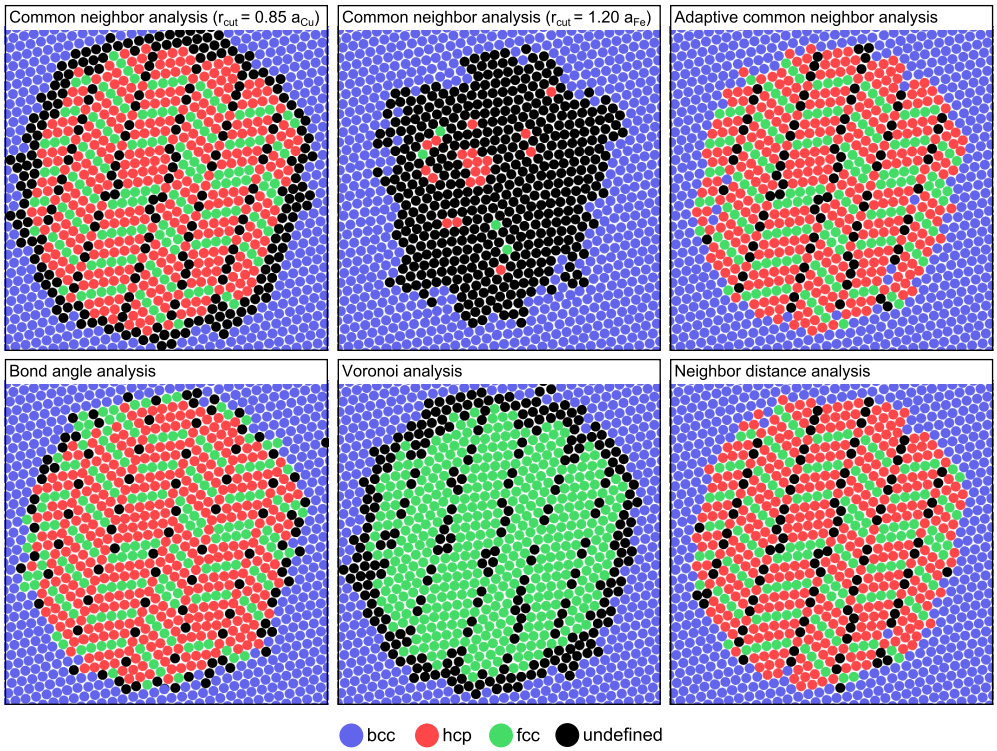

Figure 2. Cross-sectional view of a Cu-rich 9R precipitate in bcc-Fe (viewing direction $\langle 111 \rangle_{\mathrm{bcc}}$). The same simulation snapshot has been analyzed with several coordination structure identification methods discussed in this paper. Atom colors indicate the local structural type as classified by the algorithms.

3.2. Neighbor distance analysis

In section 2 we described several structure matching methods that all exploit structural symmetries in some way. Instead of directly comparing the actual atomic positions to a set of reference coordinates, they condense the particle coordinates into characteristic signatures which are invariant under rotation, and which can easily be compared. This transformation is essentially what makes the identification process efficient and robust (see [5] for an in-depth discussion). Note that, at the same time, this data reduction usually results in some insensitivity to elastic deformations: Small perturbations of the atomic positions do not change the calculated signature.

In general, however, the atomic arrangements found in the core regions of crystal defects such as grain boundaries may not exhibit any symmetries or order (e.g. discrete neighbor shells). It might therefore be more difficult to reduce their description to a small, but unambiguous set of characteristic numbers (and even less so to a scalar signature like the CSP). Thus, in such a case, one has to resort to more extensive types of signatures, as we will propose it in the following. Here, we will introduce the *neighbor distance analysis* (NDA), a new structure identification method that aims at situations where the coordination structure of atoms is lacking any particular symmetries or shell structure that could be exploited, as it is often the case in crystal defect cores.

Let us assume that a reference coordination structure (which we want to search for in the simulation data) is specified in terms of the list of bond vectors $(\mathbf{R}_1, \ldots, \mathbf{R}_N)$ connecting the central atom with its N nearest neighbors (with N being a freely selectable parameter). The coordination pattern for fcc lattice atoms, for instance, would consist of $N = 12$ neighbors, with the reference vectors $(\mathbf{R}_1, \ldots, \mathbf{R}_N)$ comprising the $1/2\,\langle 110\rangle$ vector family. We assume that this list of vectors is ordered according to their distance from the central atom such that $R_1 \leq \ldots \leq R_N$.

Given an atom to be analyzed and to be tested against the reference pattern described above, we first determine its N nearest neighbor vectors, $(\mathbf{r}_1, \ldots, \mathbf{r}_N)$, and sort them according to their magnitude (such that $r_1 \leq \ldots \leq r_N$). Obviously, this is not sufficient to associate the actual neighbor bonds with their counterparts in the reference pattern: The bond lengths may be perturbed by thermal displacements, and the ordering can be non-unique if neighbors are arranged on shells. Despite that, we may compute a local hydrostatic scale factor, λ, from the two sorted bond lists:

$$\lambda = \frac{1}{N} \sum_{i=1}^{N} (R_i / r_i)\,. \tag{8}$$

This scale factor relates the lattice constant of the reference structure (which is arbitrary, and may be chosen to be unity) to that of the actual crystal, which depends on factors such hydrostatic stress, temperature, and chemical composition.

The one-to-one mapping between the reference vectors $(\mathbf{R}_1, \ldots, \mathbf{R}_N)$ and the actual neighbor vectors $(\mathbf{r}_1, \ldots, \mathbf{r}_N)$, as we still need to determine it, can be expressed in terms of a permutation $\sigma = (\mathbf{r}_{\sigma(1)}, \ldots, \mathbf{r}_{\sigma(N)})$ of the original neighbor list. Note that multiple equivalent permutations may exist due to symmetries of the coordination structure.

How is the permutation mapping σ determined? For this we define a new type of signature that is based on the linear distance $d_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$ between two neighbors i and j of the central atom, which is invariant under rotation. Hence, we give this approach the name *neighbor distance analysis* (NDA). We need to consider that particle positions may be displaced due to thermal vibrations or elastic distortions of the crystal. Let the maximum allowed deviation of an atom from its equilibrium position be given by a user-definable parameter $\delta_{\max}$. Then the test structure $(\mathbf{r}_1, \ldots, \mathbf{r}_N)$ matches the reference pattern if at least one mapping σ exists such that the condition

$$\underbrace{|\mathbf{R}_i - \mathbf{R}_j| - \delta_{\max}}_{d_{ij}^{\min}} \;\leq\; \underbrace{\lambda \left|\mathbf{r}_{\sigma(i)} - \mathbf{r}_{\sigma(j)}\right|}_{d_{ij}} \;\leq\; \underbrace{|\mathbf{R}_i - \mathbf{R}_j| + \delta_{\max}}_{d_{ij}^{\max}} \tag{9}$$

is fulfilled for all $N(N-1)/2$ neighbor pairs. That is, all rescaled distances must lie in the corresponding intervals of the reference structure. This condition is illustrated in figure 3(a).

To find a valid permutation map σ that fulfills condition 9 (or to confirm the non-existence), up to $N!$ possible permutations of the neighbors must tested (figure 3(b)). To avoid a fully exhaustive search, the search space can, however, be considerably reduced by pruning the combinatorial search tree and employing a backtracking algorithm [23]. As an additional optimization step prior to the full combinatorial search we perform an early rejection test on the entire coordination structure by sorting both the list of pair-wise distances, $\{d_{ij}\}$, and the list of distance ranges, $\{[d_{ij}^{\min}, d_{ij}^{\max}]\}$, in ascending order (figure 3(c)). If any of the distances falls outside

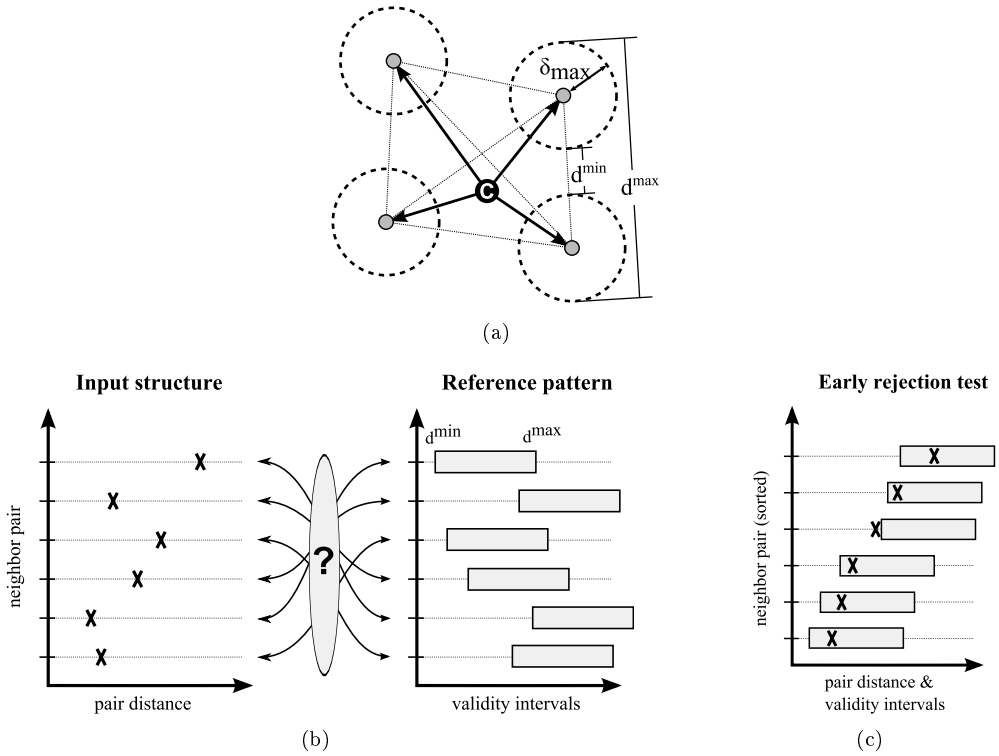

Figure 3. (a) Schematic picture of a low-symmetry coordination structure. Dashed circles indicate the maximum distance a neighbor may deviate from its equilibrium position. This yields six min-max constraints on the mutual distances between the four neighbors in the example. (b) For a positive match, a permutation of the neighbors must exist such that the actual distances fall into the intervals of the reference pattern. (c) By sorting the neighbor distances and the reference intervals, a quick rejection test can be performed without knowledge of the actual neighbor-to-reference mapping.

the corresponding admissible range, no valid permutation map can exist and the test structure does not match the reference pattern.

The user needs to specify two control parameters for each NDA reference pattern: The number of nearest neighbors to be taken into account (N) and the maximum admissible displacement ($\delta_{\max}$). N must be at least three, should include complete shells, and, apart from that, be as small as possible for best efficiency.

The maximum admissible displacement $\delta_{\max}$ determines the tolerance of the identification process. In general one wants to use a large $\delta_{\max}$ to make the recognition of structures robust at high temperatures or in the presence of strong elastic distortions. On the other hand, an excessively large $\delta_{\max}$ parameter may lead to *false positives* when testing against multiple, only slightly different coordination patterns.

Note that we proposed the NDA primarily for identifying defective coordination structures that cannot be handled well with existing methods. In simple cases (such as perfect fcc, hcp, or bcc lattices), the conventional techniques such as the CNA are the more economic choice. One important advantage of the NDA, however, is

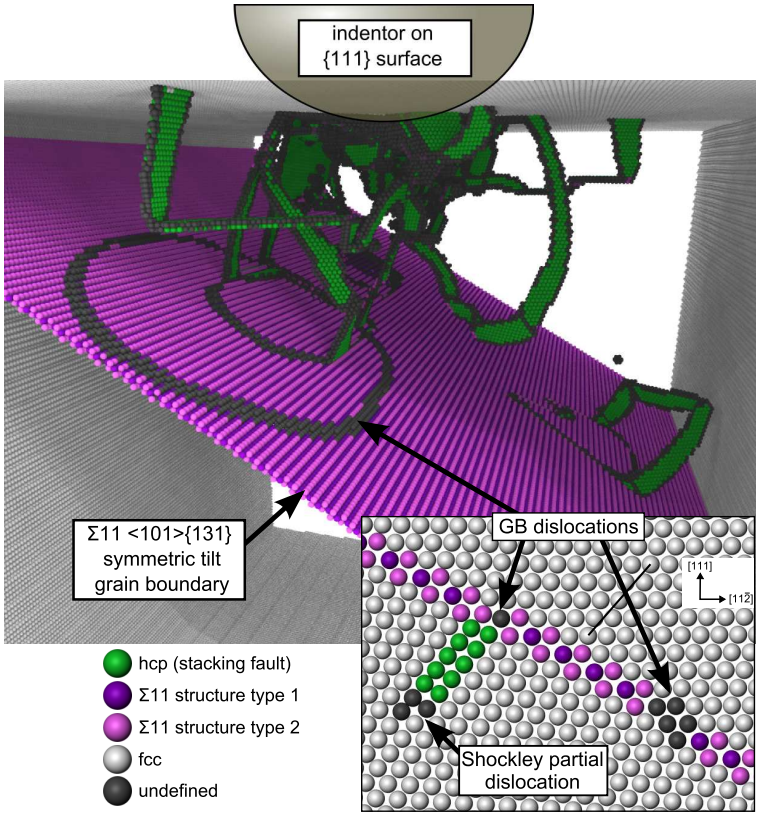

Figure 4. Molecular dynamics simulation of a nanoindentation experiment. The bicrystal contains a Σ11 ⟨101⟩ [113] symmetric tilt boundary that acts as a barrier for lattice dislocations. The neighbor distance analysis was used to identify undisturbed grain boundary regions (purple atoms), thereby revealing the cores of secondary grain boundary dislocations (dark gray atoms), which are the product of reactions of lattice dislocations with the boundary.

its capability to identify a wide range of coordination structures. In contrast to the bond angle analysis method, for instance, which employs hard-coded decision rules, the catalog of reference structures recognized by the NDA can be easily extended. The user simply has to provide a set of perfect reference coordinates, from which the NDA signature for an atom can be automatically generated.

We demonstrate this for a molecular dynamics simulation of a Σ11 ⟨101⟩ [113] symmetric tilt grain boundary [24] in fcc aluminum. This low-energy grain boundary (GB) is composed of repeating structural units, which can be well identified with the NDA. Two different coordination structures occur in a perfect Σ11 ⟨101⟩ [113] GB. Hence, the reference pattern catalog contains two GB-specific structures in addition to the perfect fcc and hcp signatures. The latter is needed to identify atoms in fcc instrinsic stacking faults, which exhibit an hcp-like coordination structure. The maximum displacement parameter is set to 21% of the nearest neighbor distance, i.e. $\delta_{\max} = 0.21 R_1$.

The simulation setup consists of a bicrystal with a single Σ11 GB and a nanometer-sized spherical indentor tip. The prismatic dislocation loops nucleated

	Dimensionality of signature	Computational cost factor
Atomic energy	1	-
Centrosymmetry parameter	1	1
Common neighbor analysis	$3N$	3
Adaptive common neighbor analysis	$3N$	4
Bond angle analysis	8	4
Bond order analysis	2	100
Voronoi analysis	4	50
Neighbor distance analysis	$N(N-1)/2$	20

Table 2. Comparison of structure identification methods discussed in this article. The signature dimensionality is the number of components of the vector that identifies a particular coordination structure, and which is used for matching within the given analysis framework. N denotes the number of near neighbors of a central particle. The cost factors are expressed relative to the computational cost of calculating the centrosymmetry parameter and do *not* include time spent on generating neighbor lists. Note that timings are approximate, depend on the input dataset, and are based on our own, optimized implementations of the algorithms, which we provide for reference.

beneath the indentor interact with the GB (absorption, transmission, and re-emission). Figure 4 shows the NDA analysis results visualized with OVITO [25]. In the large picture, fcc atoms have been removed to reveal all crystal defects. One can observe GB dislocation loops gliding in the $\Sigma 11$ boundary. The inset shows a cross-section of the symmetric tilt GB with two secondary GB dislocations. Grain boundary dislocations are clearly visible because the characteristic structure of the GB is disturbed inside their cores (dark gray atoms).

4. Comparison

The various structure identification methods discussed in this article employ different types of descriptors or signatures to identify atomic coordination structures. In general, the classification of a structure is not based on the particle coordinates themselves but rather on a derived descriptor. The size of this signature differs for each analysis method as shown in Table 2. While the centrosymmetry parameter is a scalar quantity, the neighbor distance analysis method takes into account all pair-wise distances between the N neighbor atoms. In general, the capability of a characterization method to discriminate between a wide range of structures requires a signature with a sufficient number of degrees of freedom.

We have implemented all analysis algorithms discussed in this article within a single computer code framework to facilitate the comparison between them. The code is made available for download at the website http://asa.ovito.org/. This may be useful for researchers that wish to further explore comparisons between the methods or for someone trying to understand the details of the implementations. To measure and compare their computational costs we applied all discussed methods to the Fe-Cu dataset shown in figure 2. For those analysis algorithms that assign a structural type to each atom, we have included fcc, hcp, and bcc as possible candidates. Calculating the centrosymmetry parameter is the least expensive analysis, and we have taken it as reference for our timings. Accordingly, the computation time per atom of the other

methods (Table 2) is expressed in terms of multiples of this reference time.

5. Outlook

Note that the list of methods discussed here is not exhaustive. Indeed, there exist many more methods, with new ones still appearing, such that a truly exhaustive study is beyond our finite capabilities. The aim of the present work was to focus on often-cited techniques that are routinely used in current simulation studies.

Filtering methods such as the CNA or the CSP are efficient and convenient techniques that serve well in the visualization and interpretation of datasets obtained from molecular dynamics simulations of simple systems with fcc, hcp, or bcc structure. All available structure identification methods have several limitations in common though, which should be addressed by future work. By taking into account only near neighbors of a central atom, the described methods are effectively limited to simple lattices with a monatomic basis, where the characterization of the short-range structure around individual atoms is sufficient. For identification of complex lattices with multiple atoms per primitive cell (such as 9R) one needs to take into account the medium-range order of atoms. The same applies to the automated identification of structured crystal defects such as coherent grain boundaries with large Σ, whose characteristic structural units may comprise many atoms with each having a different local environment.

Furthermore, the sensitivity of structure recognition methods to perturbations of the particle positions is a problem that hampers the analysis of systems at high temperature or under large deformation. While the effect of thermal displacements can, in many cases, be effectively mitigated by the use of time-averaged particle positions or by quenching the system using a steepest-descent technique, non-uniform lattice strains can easily interfere with the identification of coordination structures. The reason is that most structure signatures used to identify atomic arrangements are invariant only under rotation but not under arbitrary affine deformations.

So far, the structure characterization techniqes described in this article are primarily used to filter simulation datasets to reveal crystal defects for visualization purposes. In addition, they are employed to estimate crystal defect densities in MD simulations (e.g. fcc stackings faults and twin boundaries [26], or dislocations [27]). More recently, however, they have become integral parts of several sophisticated analysis and simulation methods. Examples for such applications are the characterization of dislocation lines via an automated Burgers circuit analysis [28], the mapping of a crystal to a stress-free configuration to separate elastic from plastic deformation [29], and the automated construction of a catalog of structural motives for the efficient discovery of transition events in self-learning kinetic Monte Carlo simulations [30]. Such applications usually require more than simple classification of local atomic arrangements. For instance, to determine the local crystallographic directions in a crystal [31], it is necessary to map all neighbors of the central particle to the reference pattern in a one-to-one fashion (as it is already performed by the neighbor distance analysis described in section 3.2), and to determine the list of all equivalent neighbor permutations, which correspond to the elements of the point symmetry group of the structure at hand.

Acknowledgments

The author thanks Paul Erhart and Tomas Oppelstrup for helpful discussions. This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

References

[1] C. L. Kelchner, S. J. Plimpton, and J. C. Hamilton. Dislocation nucleation and defect structure during surface indentation. *Phys. Rev. B*, 58(17):11085, 1998.

[2] P. J. Steinhardt, D. R. Nelson, and M. Ronchetti. Bond-orientational order in liquids and glasses. *Phys. Rev. B*, 28:784–805, Jul 1983.

[3] J. D. Honeycutt and H. C. Andersen. Molecular dynamics study of melting and freezing of small Lennard-Jones clusters. *J. Phys. Chem.*, 91(19):4950–4963, 1987.

[4] G. J. Ackland and A. P. Jones. Applications of local crystal structure measures in experiment and simulation. *Phys. Rev. B*, 73(5):054104, 2006.

[5] A. S. Keys, C. R. Iacovella, and S. C. Glotzer. Characterizing complex particle morphologies through shape matching: Descriptors, applications, and algorithms. *J. Comp. Phys.*, 230(17):6438–6463, 2011.

[6] V. V. Bulatov and W. Cai. *Computer Simulations of Dislocations.* Oxford University Press, 2006.

[7] J. Li. AtomEye: an efficient atomistic configuration viewer. *Model. Simul. Mater. Sci. Eng.*, 11(2):173–177, 2003. http://mt.seas.upenn.edu/Archive/Graphics/A/.

[8] S. Plimpton. Fast parallel algorithms for short-range molecular dynamics. *J. Comp. Phys.*, 117(1):1, 1995. Software available at http://lammps.sandia.gov/.

[9] W. Lechner and C. Dellago. Accurate determination of crystal structures based on averaged local bond order parameters. *The Journal of Chemical Physics*, 129(11):114707, 2008.

[10] C. Desgranges and J. Delhommelle. Crystallization mechanisms for supercooled liquid Xe at high pressure and temperature: Hybrid Monte Carlo molecular simulations. *Phys. Rev. B*, 77:054201, Feb 2008.

[11] D. Faken and H. Jonsson. Systematic analysis of local atomic structure combined with 3d computer graphics. *Comput. Mater. Sci.*, 2(2):279–286, 1994.

[12] H. Tsuzuki, P. S. Branicio, and J. P. Rino. Structural characterization of deformed crystals by analysis of common atomic neighborhood. *Comput. Phys. Commun.*, 177(6):518–523, 2007.

[13] N. Lümmen and T. Kraska. Common neighbour analysis for binary atomic systems. *Modelling Simul. Mater. Sci. Eng.*, 15(3):319–334, 2007.

[14] G. Z. Voronoi. *J. Reine Angew. Math.*, 134:199, 1908.

[15] J. L. Finney. Random packings and structure of simple liquids. 1. geometry of random close packing. *Proceedings of the Royal Society of London A.*, 319:479, 1970.

[16] A. Okabe, B. Boots, K. Sugihara, and S. N. Chiu. *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams.* John Wiley, Chichester, 2nd edition, 2000.

[17] C. S. Hsu and A. Rahman. Interaction potentials and their effect on crystal nucleation and symmetry. *The Journal of Chemical Physics*, 71(12):4974–4986, 1979.

[18] C. H. Rycroft. VORO++: A three-dimensional Voronoi cell library in C++. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 19(4):041111, 2009.

[19] J. L. Bentley. Multidimensional binary search trees used for associative searching. *Commun. ACM*, 18:509–517, 1975.

[20] J. H. Friedman, J. L. Bentley, and R. A. Finkel. An algorithm for finding best matches in logarithmic expected time. *ACM Trans. Math. Softw.*, 3:209–226, 1977.

[21] B. Sadigh, P. Erhart, A. Stukowski, A. Caro, E. Martinez, and L. Zepeda-Ruiz. Scalable parallel Monte Carlo algorithm for atomistic simulations of precipitation in alloys. *Phys. Rev. B*, 85:184203, 2012.

[22] F. Ernst, M. W. Finnis, D. Hofmann, T. Muschik, U. Schönberger, U. Wolf, and M. Methfessel. Theoretical prediction and direct observation of the 9R structure in Ag. *Phys. Rev. Lett.*, 69(4):620–623, 1992.

[23] D. E. Knuth. *The Art of Computer Programming*, volume 4A: Combinatorial Algorithms. Upper Saddle River, Addison-Wesley Professional, 2011.

[24] M. de Koning, R.J Kurtz, V.V. Bulatov, C.H. Henager, R.G. Hoagland, W. Cai, and M. Nomura. Modeling of dislocation–grain boundary interactions in FCC metals. *Journal of Nuclear Materials*, 323:281–289, 2003.

[25] A. Stukowski. Visualization and analysis of atomistic simulation data with OVITO – the Open Visualization Tool. *Modelling Simul. Mater. Sci. Eng.*, 18:015012, 2010. Software available at http://ovito.org/.

[26] A. Stukowski, K. Albe, and D. Farkas. Nanotwinned fcc metals: strengthening versus softening mechanisms. *Phys. Rev. B*, 82:224103, 2010.

[27] F. Sansoz. Atomistic processes controlling flow stress scaling during compression of nanoscale face-centered-cubic crystals. *Acta Materialia*, 59(9):3364 – 3372, 2011.

[28] A. Stukowski and K. Albe. Extracting dislocations and non-dislocation crystal defects from atomistic simulation data. *Modelling Simul. Mater. Sci. Eng.*, 18(8):085001, 2010.

[29] A. Stukowski and A. Arsenlis. On the elastic-plastic decomposition of crystal deformation at the atomic scale. *Modelling Simul. Mater. Sci. Eng.*, 20:035012, 2012.

[30] F. El-Mellouhi, N. Mousseau, and L. J. Lewis. Kinetic activation-relaxation technique: An off-lattice self-learning kinetic Monte Carlo algorithm. *Phys. Rev. B*, 78:153202, Oct 2008.

[31] C. S. Hartley and Y. Mishin. Characterization and visualization of the lattice misfit associated with dislocation cores. *Acta Materialia*, 53(5):1313–1321, 2005.