2018-04-01

Quantifying Grain Boundary Atomic Structures Using the Smooth Overlap of Atomic Positions

Jonathan Lake Priedeman
Brigham Young University

Quantifying Grain Boundary Atomic Structures Using the Smooth Overlap of Atomic Positions Descriptor

Jonathan Lake Priedeman

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Science

Eric R. Homer, Chair
David T. Fullwood
Oliver K. Johnson

Department of Mechanical Engineering

Brigham Young University

ABSTRACT

Quantifying Grain Boundary Atomic Structures Using the Smooth Overlap of Atomic Positions Descriptor

Jonathan Lake Priedeman

Department of Mechanical Engineering, BYU

Master of Science

In this work, the relationship between grain boundary crystallography and grain boundary atomic structure is examined, using [1 0 0] - symmetric tilt grain boundaries in nickel. The structural unit model is used as a benchmark to evaluate the atomic structure description capacities of an emerging structural descriptor, the local environment representation, which itself is a refinement of the also-emergent Smooth Overlap of Atomic Positions (SOAP) descriptor. We show that the local environment representation encodes both the information of the structural unit model and additional information, such as distortion in the structural units and the arrangement of the structural units at the interface. The use of the local environment representation permits the use of a visualization tool known as SPRING to represent structural similarities between grain boundaries. With the SPRING representation, we produce objective evidence of a relationship between crystallography and atomic structure, at least for [1 0 0] - symmetric tilt grain boundaries.

Keywords: grain boundaries, atomic structure, SOAP descriptor, structural unit model

ACKNOWLEDGMENTS

I would like to thank Conrad Rosenbrock for providing me with code and instruction regarding the application of the SOAP descriptor to grain boundaries. I also thank Dr. Oliver Johnson for providing knowledge of the SPRING visualization tool to this project; it was a significant contribution to the overall goals of the investigation. I cannot thank Dr. Eric Homer enough for the direction, feedback, and mentoring that he has provided as my advisor during my time at BYU. I am fortunate to have been able to work with such great men. And finally, I acknowledge the unsung contributions of my wife, Katie, the light of my life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

NOMENCLATURE

a_0	Lattice Parameter
$\vec{a}$, $\vec{b}$	SOAP vectors
ε	Equivalence threshold
l_{max}	Number of angular basis functions in the spherical harmonics
n_{max}	Number of radial basis functions
r_{cut}	Cutoff Radius
σ	Gaussian width
$s_{\vec{a},\,\vec{b}}$	Similarity between SOAP vectors $\vec{a}$, $\vec{b}$

CHAPTER 1. INTRODUCTION

1.1 Grain Boundary Characterization Methods

Grain boundaries (GBs) are interfaces in polycrystalline materials that exert significant influence on bulk material properties. For example, increasing the fraction of $\Sigma 3^n$ GBs in a material has been correlated with improvements in ductility [1], strength [2], cracking resistance [3], and corrosion resistance [4]. Even so, an understanding of GB structure-property relationships remains elusive. The relationship between GB crystallographic structure and GB atomic structure remains, at best, incompletely explored, with previous authors exploring specific regions of the GB space [5–17].

In part, this is due to the size of the space: three macroscopic degrees of freedom for the misorientation between the intersecting grains [18], two macroscopic degrees of freedom to define the boundary plane of intersection [19]. These five macroscopic parameters represent the full crystallographic characterization of a GB. Six microscopic degrees of freedom to govern factors such as lattice shifts [20, 21] are used to search for the minimum energy atomic structure (atomically, complete characterization of a GB requires 3·N degrees of freedom for N atoms at the boundary). However, it is often assumed that any given GB will assume the lowest energy configuration, so the microscopic degrees of freedom are ignored once the minimum is determined [14, 20, 21].

The characterization activities of GBs are another obstacle. Experiments tend to be tedious [22–25], and resolution of all five macroscopic parameters is not always possible. Atomistic simulation is computationally expensive, with the largest catalogs of simulated GBs numbering less than 500 in both FCC and BCC GBs [20, 26].

Rosenbrock *et al.* recently applied machine learning to predict GB properties from GB atomic structures [27]. This recent work offers the exciting possibility of rapid and accurate predictions of GB properties. Ideally, a machine learning-based, GB structure-property prediction tool should (1) predict GB properties given GB crystallographic structures, because experiment

captures crystallography and not atomic structure (the latter actually controls the properties); and (2) offer insight into the prediction process rather than only operating as a black box [27].

1.2 Atomic Structure Descriptors

To close the loop between crystallography and properties, the relationship between crystallography and atomic structure must be explored so that the latter can be predicted given the former. Describing GB atomic structures, however, is a complex task. For a GB with N interfacial atoms, a structure characterization approach must describe the positions of the atoms, which is a $3 \times N$ dimensional space. As there is no limit on the number of atoms possible in a GB, a structural descriptor is typically applied to make characterizing the GB's atomic structure more tractable. A structural descriptor will typically determine the subunits present in the GBs examined, so that the GBs can be discussed and examined in terms of the subunits, rather than the $3 \times N$ complete atomic structure. The ideal structural descriptor would have the following characteristics:

- Easily interpretable
- Easily visualized
- Enable comparison between GBs
- Invariant with respect to structural symmetries, rotations, and permutations
- Accommodates structural perturbations
- Smoothly varying structural description (small changes in position have corresponding small changes in the descriptor, or Lipschitz continuity with a small constant; smooth functions are inherently differentiable, so the descriptor will also be differentiable)
- Applicable to general, three-dimensional GB structures
- Amenable to automated structural description
- Capture connectivity between interfacial atoms and/or the subunits
- Automatic discovery and incorporation of new atomic structures and/or subunits

Finding a structural descriptor that satisfies each item from this list of characteristics is a tall order. The structural unit model [5–17] is most widely used atomic structure descriptor in the literature, and will be the structural metric employed in this work for comparisons against emerging descriptors. The structural unit model identifies patterns in quasi-two-dimensional GBs as structural units. The structural unit model's strengths lie in the easy interpretation and visualization of the results. While it might be said that the model is invariant with respect to structural symmetries and rotations, this invariance depends on the skill of the researcher executing the description, as GB structures are almost entirely described manually. Furthermore, the structural unit model is not easily extended to general, three-dimensional GBs and is not numerical in nature.

We acknowledge other GB atomic structural description methods available in the literature, but we only briefly explore their strengths and/or weaknesses. The approaches include the centrosymmetry parameter [28], polyhedral template matching, [29], Voronoi cell topology [30], packing of polyhedra [31, 32], and dislocation arrays [33]. The centrosymmetry parameter [28] and polyhedral template matching [29] easily identify crystal structures (FCC, HCP, etc.), but are not currently well adapted to distinguish differences between non-crystalline atomic arrangements, such as those at GBs (polyhedral template matching would require pre-determined atomic arrangement templates in order to classify a GB). The dislocation arrays approach [33] places dislocations in GBs to accommodate the disorientation between the two crystals. While this is easily interpreted and visualized, this tool is limited to low-angle GBs in which the dislocations do not appreciably overlap. Voronoi cell topology [30, 34] and polyhedra packing [31, 32] use Voronoi tessellation to determine the subunits in the atomic structure. They are easily visualized, applicable to general, three-dimensional GBs, and already automated; however, at the present time these methods require predefined subunits in order to classify GB atomic structures.

Here we present two emerging approaches to describing GB atomic structure. The first descriptor is obtained from a tool called the Smooth Overlap of Atomic Positions (SOAP) descriptor, developed by Bartók *et al*. for improved representation of atomic environments [35]. The second descriptor is a refinement of the SOAP metric that finds a limited set of environments (identified as unique local atomic environments) that can be used to represent a given GB. The second descriptor is labeled as the local environment representation.

The SOAP descriptor represents atomic positions around a reference atom with three-dimensional Gaussian density distributions placed at each neighbor atom. A basis set of radial and angular distribution functions in the spherical harmonics are fit to this density map, and the coefficients of the fit are compiled into a vector (referred to here as the SOAP vector) that quantifies the local atomic environment.

The SOAP descriptor has some inherent advantages we perceive to be useful for describing GB atomic structure [27], and we repeat those advantages here. First, SOAP was not designed for a specific crystalline lattice system and hence requires no accommodations for application to a new lattice system or atomic environments possessing no defined structure. Second, a given SOAP description is invariant to translations, rotations, and permutations of atoms. Third, SOAP varies smoothly and is differentiable.

The SOAP descriptor has weaknesses. SOAP's structural descriptions currently are not easily interpreted or visualized, in contrast to the structural unit model. Because of its flexibility and numerical nature, the SOAP descriptor will always produce a description of structures provided to it; the utility of these descriptions must be examined and verified by the researcher.

The local environment representation classifies each atom in a GB by a unique local atomic environment. This classification elucidates trends in atomic structures by reducing atomic descriptions to a single number for each atom, rather than a vector (with generally in excess of a thousand elements). This reduction permits easier visualization of atomic structures and facilitates interpretation. To represent entire GB atomic structures, the local environment representation determines the fractions of the unique environments present, allowing for easy comparison between boundaries.

Means to elucidate the physics (and prevent a black box) must also be explored. Such means might include the presence of particular structural features in the GBs that correlate to certain GB behaviors, characteristics, or properties. In this work, for example, we will explore the correlations between the local environment representation and the structural unit model, via the presence of specific unique local atomic environments occurring with corresponding structural units.

This investigation seeks to explore the relationship between GB crystallographic descriptions and GB atomic structures within the limited context of simulated nickel [1 0 0] - symmetric tilt

GBs, using the structural unit model and the local environment representation. The atomic structure trends of these GBs will be nothing new, as this class of GBs has historically been amenable to analysis using the structural unit model. The [1 0 0] - symmetric tilt GBs will thus allow us to compare the structural unit model and the local environment representation against our proposed list of ideal characteristics of an atomic structure descriptor, without requiring the reader to simultaneously adjust to novel atomic structures. We detail the various tools (simulated bicrystal GBs, the structural unit model, the SOAP descriptor, the local environment representation, and SPRING visualization tool) that this investigation leverages; some emphasis is given to the SOAP descriptor and the local environment representation, as we explore improvements to them. These tools are then synthesized and compared in our pursuit of both the ideal structural descriptor as well as the relationship between crystallography and atomic structure, so that the foundation for continued exploration of GB crystallographic-atomic structure relationships via machine learning is laid.

CHAPTER 2. METHODOLOGY

2.1 Grain Boundary Data Set

This investigation deals with the [1 0 0] - symmetric tilt GBs. There are 126 simulated nickel GBs in the dataset; 26 are drawn from the catalog of 388 nickel GBs created by Olmsted, Foiles and Holm [20], 1 is drawn from the unpublished work of Erickson, and 99 new GB structures are simulated as part of this investigation. Taken together, these 126 simulated GBs represent all of the [1 0 0] - symmetric tilt GBs having a Σ value under 400. These boundaries are identified in the text according to their [1 0 0] - symmetric tilt angle (tilt angle) to one decimal place (in degrees). A complete description of the GBs is found in Appendix A, with the tilt angle, CSL value, and boundary plane normal in the boundary plane normal fundamental zone, as described in [19, 36]. Commonly studied GBs will be identified with CSL value and boundary plane normal, as appropriate.

The 26 GBs from Olmsted *et al.* were constructed as orthogonal, bicrystal simulation cells with shrink-wrapped boundary conditions in the direction of the boundary plane normal and periodic boundary conditions in the plane of the GB. These GBs fit within simulation cells having a maximum periodic side length of $15/2 \cdot a_0$, with a_0 as the lattice parameter. The Foiles-Hoyt embedded atom method potential [37] and the LAMMPS atomistic simulation software [38] were used to determine the minimum energy configuration of these GBs. Further detail on the creation of these GBs can be found in [20].

The one hundred GBs created by Erickson and this work are constructed according to the grain boundary construction and minimization methodology of Homer [21], which is based on the methodology of Olmsted [20]. But, rather than create all possible CSL GBs that fit within specified dimensional constraints, we create GBs with specific misorientations and boundary plane orientations (no dimensional constraints are imposed). The bicrystal simulation cells are constructed using the CSL lattice to determine the dimensions of the cell; since we do not enforce the orthogo-

nality requirement, the vectors from the CSL lattice are applied to the simulation cell, resulting in a triclinic cell.

2.2 Structural Unit Model

The structural unit model identifies patterns in the atomic arrangements in quasi-two-dimensional GBs as structural units. Rittner and Seidman explain the process of characterizing the boundary structure as marking lines between atoms where the FCC structure appreciably breaks down [11]. These lines serve to assist in identifying the patterns occurring at the GB; a set of unique patterns forms the structural units to classify the structure of the boundary [11]. The classification of a GB using the structural unit model requires the researcher to visually identify and label the structural units present.

There are various structural unit identification conventions in the literature [6, 7, 9–17]; we follow the convention of Han *et al*. [17] in order to follow recent work, even though [17] examined BCC tungsten, rather than an FCC crystal system. The A-structural unit corresponds to the unit cell from the $(1\ 1\ 0)$ plane in FCC, the B-structural unit corresponds to the unit cell of the $(1\ 0\ 0)$ plane in FCC, and the I-structural unit is composed of an A-structural unit sandwiched between two B-structural units. The C-structural unit has been often referred to as a "kite" structure in contexts outside of the structural unit model, or an E-structural unit by other workers [11, 12, 15, 16]. The U-structural unit (U indicating undefined) is any region of the GB where a three-dimensional structure occurs, rather than a strictly quasi-two-dimensional structure.

2.3 Smooth Overlap of Atomic Positions Descriptor

As noted in the introduction, the SOAP descriptor is an emergent technique for describing the geometrical arrangement of atoms around a central, reference atom (the local atomic environment). First, the SOAP descriptor places three-dimensional Gaussian densities at the positions of the atoms (to accommodate uncertainty). Second, the geometrical arrangement of the atoms (represented by the Gaussian densities) is captured by fitting a set of angular spherical harmonic and radial basis functions to the three-dimensional density distribution. The resulting coefficients

of the fit are compiled into a vector, known as a SOAP vector, that describes the local atomic environment of the central, reference atom.

The description of a GB's atomic structure using the SOAP descriptor strives to capture only the atoms near the interface, rather than the whole computational cell. Using common neighbor analysis [28], the atoms in the GB are classified as having a bulk crystalline structure (e.g., FCC) or non-FCC crystal structure. The atoms having non-FCC crystal structure are used to find an approximate width for the interface. A padding of bulk atoms is included on either side of the approximate width; it is these atoms (those of the boundary and both paddings) whose local atomic environments are described using the SOAP descriptor. The resulting set of SOAP vectors is the raw representation of the atomic structure of the grain boundary by the SOAP method [27].

The SOAP descriptor has several tunable parameters that allow the researcher to adjust the representation to capture desired information [35, 39, 40]. The four parameters of mention for this work are the cutoff radius (r_{cut}), the number of angular spherical harmonic basis functions (l_{max}) and radial basis functions (n_{max}), and the width of the Gaussian densities (σ) placed at the position of each atom; the SOAP descriptor considers all possible index values m for each index l in the spherical harmonics. The variable r_{cut} controls the size of the environment to be represented; a small r_{cut} will only capture positions in the nearest vicinity while a large r_{cut} will quantify longer-range behavior. The padding used to augment the approximate width of the grain boundary (see the previous paragraph) has a thickness of r_{cut}. The variables l_{max} and n_{max} influence the accuracy of the representation as captured by the spherical harmonic and radial basis expansion; larger values of l_{max} and n_{max} will increase the accuracy for higher computational cost. The variable σ influences the smoothness of the descriptor [40]. The SOAP parameters used for this work are listed in Table 2.1.

Table 2.1: SOAP Parameters used for this work

Parameter	Value	
Cutoff Radius (r_{cut})	3.25	Å
Number of Angular Basis Functions (l_{max})	12	
Number of Radial Basis Functions (n_{max})	12	
Gaussian Width (σ)	0.5	Å

2.4 Local Environment Representation

We reiterate that the SOAP vector describes the geometric positioning of the atoms around the reference atom, or the reference atom's local atomic environment. As otherwise identical environments may differ by small perturbations, we account for these minute differences with the use of a similarity metric and an equivalence threshold. The similarity metric employed in this work is found in Eq. 2.1,

$$s_{\vec{a},\,\vec{b}} = 1 - \sqrt{\vec{a}\cdot\vec{a} + \vec{b}\cdot\vec{b} - 2\times\vec{a}\cdot\vec{b}} \tag{2.1}$$

where $s_{\vec{a},\,\vec{b}}$ is the similarity value and $\vec{a}$, $\vec{b}$ are the SOAP vectors being compared. The similarity value will have values from 0 to 1. The greater the value of $s_{\vec{a},\,\vec{b}}$, the greater the similarity of the two environments, and vice versa. If $s_{\vec{a},\,\vec{b}} > \varepsilon$, then $\vec{a}$ and $\vec{b}$ are considered equivalent, meaning the two represented local atomic environments are, for our intents and purposes, the same. In this work, we apply a value of $\varepsilon = 0.9975$.

We use the similarity metric and the equivalence threshold to refine the raw SOAP description into the local environment representation. Applying the similarity metric and equivalence threshold to a set of SOAP vectors will partition them into disjoint subsets, with the constituent SOAP vectors of a subset being equivalent (this partitioning is described in greater detail in Section 2.4.1). To simplify the description of the entire set, a representative environment is chosen to represent each subset, with the restriction that there is no equivalency between any representative environments. These representative environments, or unique local atomic environments, represent the basic "building blocks" of the atomic structures being described [27].

Because we are building upon published literature [27], we first apply the SOAP descriptor to all 388 of Olmsted's GBs along with the one hundred new GBs (which are [1 0 0] - symmetric tilts). By doing so, we produce 66 unique local atomic environments. For the 388 GBs alone there are 62 unique environments, and 4 additional unique environments were required to completely describe the atomic environments of the one hundred new GBs. Because the set of unique local atomic environments used to describe all of Olmsted's GBs are nearly sufficient to describe the one hundred new GBs (only 4 new unique environments were required), we conclude that the unique local atomic environments required to represent a reduced set of GBs are applicable to representing

GBs with a more complex atomic structure (e.g. larger Σ GBs). For the 126 [1 0 0] - symmetric tilt GBs that are the focus of this work, only 55 of the 66 unique local atomic environments is required for complete description. We also note that there are 494,495 GB atoms for the 126 GBs we examine.

To quantify the similarity between the unique local atomic environments, we plot the similarity of each environment to all other environments on a heat map in Fig. 2.1. The color scheme used in this work sorts the unique local atomic environments by their similarity to FCC, and it might be tempting to think of the environments as one-dimensional deviations from a perfect FCC structure, given the linear nature of the color progression. Fig. 2.1 refutes this notion: the unique local atomic environments are an independent set, varying uniquely from FCC and each other. In other words, if two unique local atomic environments were to have the same similarity to FCC, they would still be different from one another, and this is captured in the comparison of any two environments in the set. As mentioned, the unique local atomic environments are portrayed in the atomic structure plots of this paper with color, so that each atom is colored according to its unique local atomic environment classification. The environments are ordered according to their similarity to perfect FCC, so that cooler colors correspond to environments that have greater similarity to perfect FCC than environments represented by warmer colors.

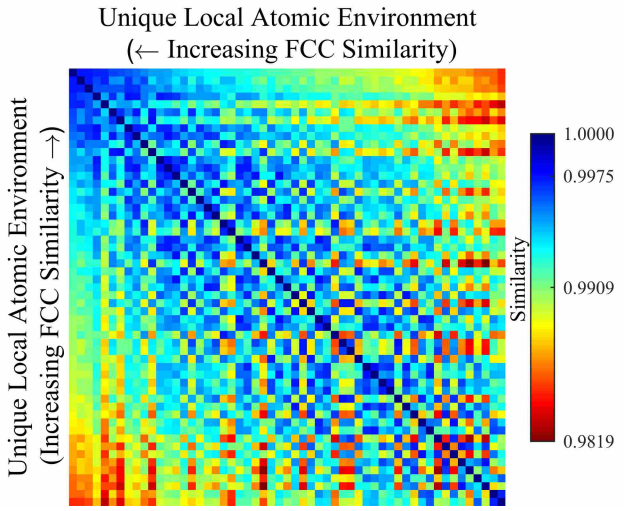

Figure 2.1: Heat map depicting unique local atomic environment similarities, using the color scheme to the right of the figure. The lack of a smooth variation in similarity across the heat map indicates that the unique local atomic environments vary uniquely from FCC and each other (rather than deviating in one dimension from FCC).

Since we choose unique local atomic environments to represent each of the disjoint subsets of local atomic environments, it possible to describe each GB's atomic structure in terms of the unique local atomic environments—building blocks—used to construct them. In other words, a given GB is described by the fraction of its atoms that are classified by each of the 55 unique local atomic environments. This characterization technique is known as the local environment representation [27]. This approach greatly simplifies the dimensionality of the GB atomic structure space: instead of 494,495 $\times$ 3 degrees of freedom (from describing the N = 494,495 atoms), we have only 55 total [27].

For the purposes of this work, we treat each unique environment as a dimension in a high-dimensional space we will identify as the local environment representation space. The fractions of the unique environments in a GB describe that GB's location in the local environment representation space. We also regard the local environment representation as a substitute for a genetic description of the GBs, allowing us to use genomic visualization tools (such as SPRING, see Section 2.6) to examine the atomic structural relationships between GBs.

2.4.1 Improving Unique Local Atomic Environment Selection

The original methodology for selecting the unique local atomic environments sequentially iterated through the total local atomic environment set, compiling the set of unique local atomic environments and classifying the local atomic environments with one iteration. In addition, there was no consistency in the choice of the unique local atomic environment used for classification; the first unique local atomic environment that satisfied the similarity criterion (see Section 2.4) was chosen.

To improve the consistency of classification, the methodology is modified in the current work. Rosenbrock's approach compiled the set of unique local atomic environments *and* classified environments in the same iteration [27]. In other words, the set of unique local atomic environments available for classification early in the iteration was smaller than the set at the end of the iteration, making the method very dependent upon the ordering of the local atomic environments analyzed. This work splits the two steps into two separate iterations: first identify the unique local atomic environments, and then classify the local atomic environments by the unique set.

First, the method iterates through the set and compiles the set of unique local atomic environments by comparing each environment to the current list of unique local atomic environments. If the considered environment has no similarity to any of the unique local atomic environments (i.e. $s_{\vec{a},\,\vec{b}} < \varepsilon$ for all current unique local atomic environments), it is added to the unique environment set, otherwise the unique environment set is left unchanged. This process is illustrated in the cartoons of Fig. 2.2(a) and (b). In Fig. 2.2(a), a local environment is compared to the current unique environment set; as there are no unique environments to which the local environment is similar, the local environment is added to the unique environment set (Fig. 2.2(b)).

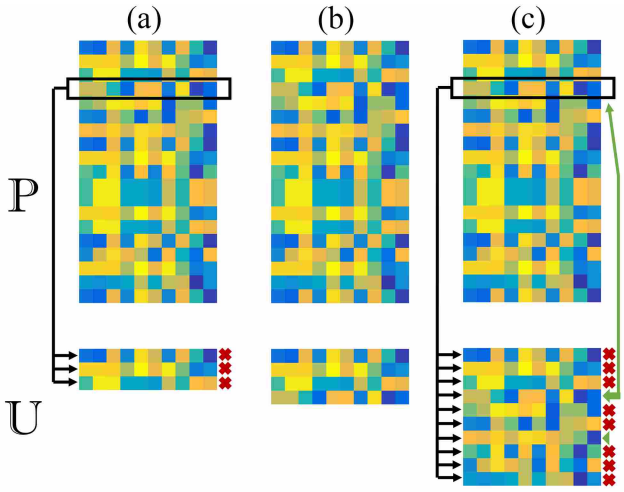

Figure 2.2: Cartoon describing the compilation of the unique local atomic environment set (see parts (a) and (b)) and the classification of the local atomic environments by the unique environment set (see part (c)). Set P indicates the total local atomic environment set; set U denotes the set of unique local atomic environments.

Second, the method iterates through the set with the purpose of classifying each local atomic environment by a unique local atomic environment. Each local environment is compared to the now-fixed set of unique environments, using the similarity metric ($s_{\vec{a},\,\vec{b}}$, Eq. 2.1) and equivalence threshold (ε) of Section 2.4; the local environment is classified by the unique environment to which the local environment has the *greatest* similarity.

Implementing these changes to the unique environment compilation/classification methodology significantly improved the consistency of the unique environment classification. It was after this methodology adjustment was made that the 2, 3, 5, and 6 positions in the C-structural unit (see

3.1.2) achieved a nearly constant unique environment classification. Further improvements could undoubtedly be made; the current method is not deterministic, and depends to some degree on the ordering of the local environments during the compilation step. Given the time constraints and success of the results, we consider the current approach to be appropriate for this work.

2.5 SOAP Parameter Selection

We also undertake a small investigation into the selection of the SOAP parameters. These parameters are critical to capturing the necessary information about the environment, as poor parameter selection will cause the descriptor to capture at best, unimportant, or at worst, interfering, information.

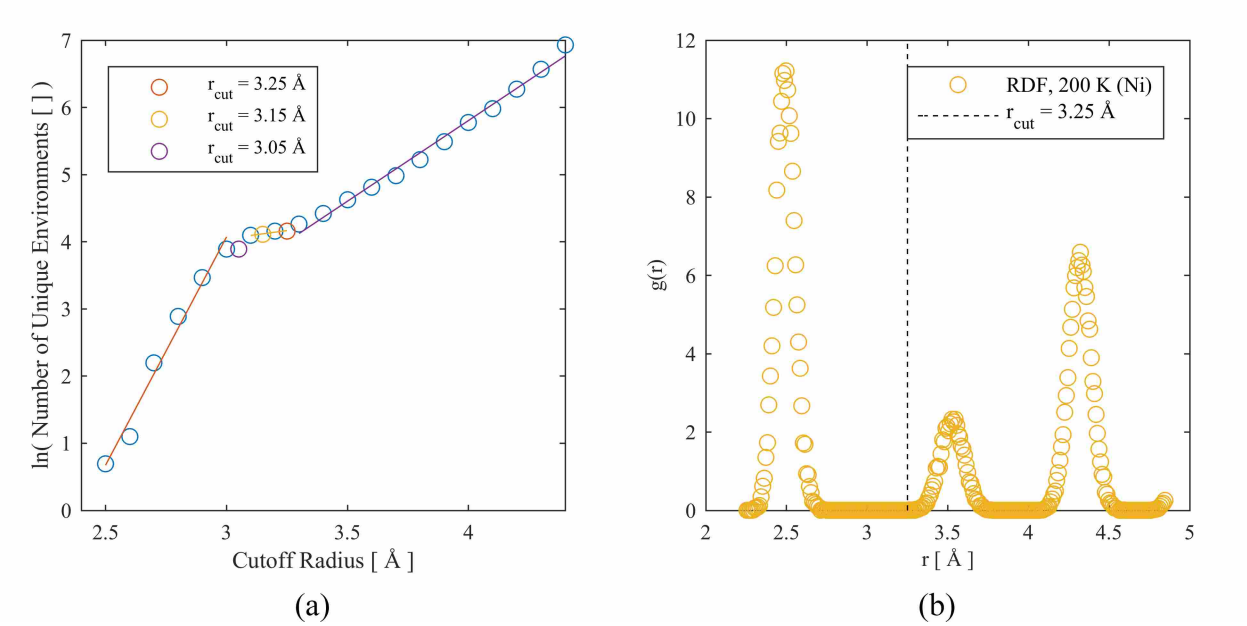

Figure 2.3: (a) Variation of the number of unique local atomic environments as a function of the cutoff radius (r_{cut}). (b) The radial distribution function in FCC nickel at 200 K, with the chosen cutoff radius marked by a dashed line. The selection of the cutoff radius allows for the SOAP descriptor to capture information almost out to the second "ring" of nearest neighboring atoms (the second spike in the RDF).

To investigate the cutoff radius, we classify the atomic structures of the 388 GBs from Olmsted's dataset, using a range of cutoff radii. For this analysis, the other SOAP parameters are: $l_{max} = 18$, $n_{max} = 18$, $\sigma = 0.5$ Å. We plot the number of unique local atomic environments against

the cutoff radius (r_{cut}) used in the analysis in Fig. 2.3(a). As expected, the number of environments increases as r_{cut} increases, since there are more ways to uniquely position the atoms in the local environment. However, in the range of 3.05 Å $< r_{cut} <$ 3.25 Å, the number of unique local atomic environments is almost constant. We identify this brief plateau as the ideal range for r_{cut}, since the number of environments appeared to be independent of r_{cut}; we select the upper end of this range (r_{cut} = 3.25 Å) to be as inclusive as possible. Visual comparison of the unique local atomic environment classification of the C-structural units for r_{cut} = 3.05, 3.15, and 3.25 Å indicated that r_{cut} = 3.25 Å yielded the greatest consistency in the classification of these C-structural units.

There is some physical meaning to a cutoff radius of 3.25 Å. The lattice parameter a_0 has a value in nickel of a_0 = 3.52 Å at 0 K, with the first ring of nearest neighbors at 2.49 Å and the second ring of nearest neighbors at 3.52 Å. At r_{cut} = 3.25 Å, we capture atom positions that would almost be in the second ring of nearest neighbors (in perfect FCC). We illustrate this point with Fig. 2.3(b), which depicts the occurrence of atoms as a function of radius in nickel at 200 K.

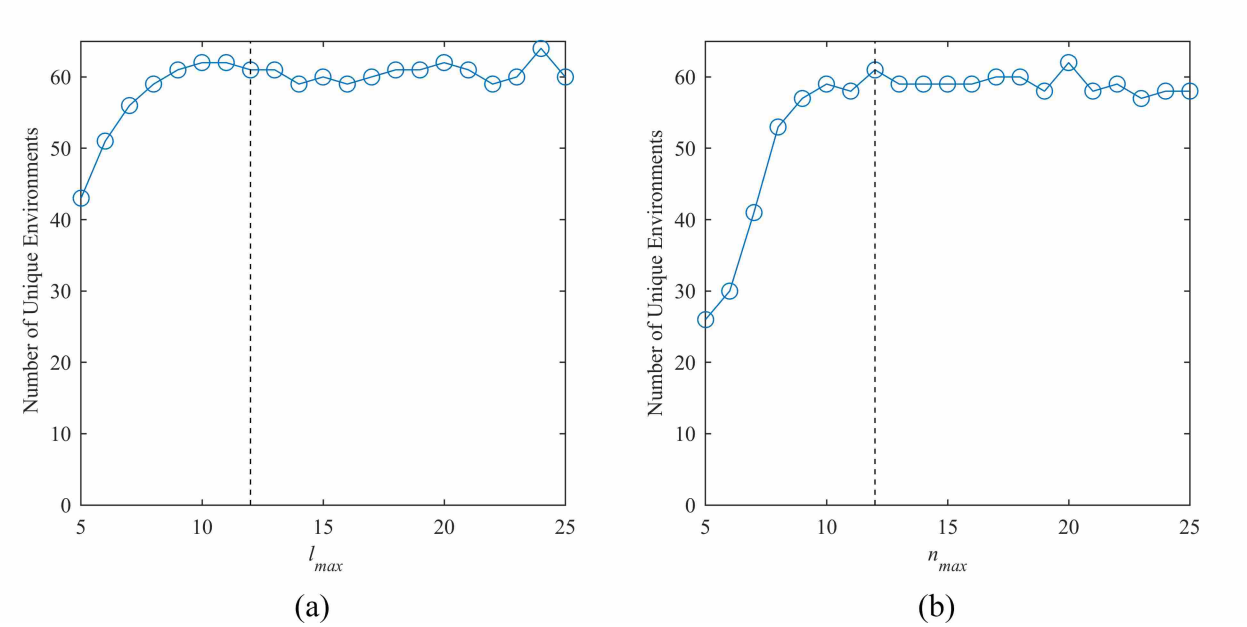

Figure 2.4: (a) Plot of the number of unique local atomic environments as a function of the number of angular basis functions in the spherical harmonics (l_{max}). The dashed line indicates the selected value. (b) Plot of the number of unique local atomic environments as a function of the number of radial basis functions in the spherical harmonics (n_{max}). The dashed line indicates the selected value.

To investigate the number of angular basis functions in the spherical harmonics required for accurate characterization of the local atomic environments, we determine the unique local atomic environments of the 388 GBs from Olmsted's data set, using a range of values for the number of basis functions. We plot the number of unique local atomic environments against the number of angular basis functions (l_{max}) in Fig. 2.4(a). For this analysis, the other SOAP parameter values are: r_{cut} = 3.25 Å, n_{max} = 9, σ = 0.5 Å. At high values of l_{max}, the number of unique local atomic environments appears to have asymptotic behavior. To fall within this asymptotic range, we choose an l_{max} value of l_{max} = 12, which is the smallest value for l_{max} that gives consistent results.

We also investigate the number of radial basis functions (n_{max}) required, following the procedure for l_{max}, and shown visually in Fig. 2.4(b). For this analysis, the other SOAP parameter values are: r_{cut} = 3.25 Å, l_{max} = 9, σ = 0.5 Å. To fall within the asymptotic behavior, we select n_{max} = 12.

Finally, we examine Gaussian width, σ. Since σ influences the smoothness of the descriptor [40], we also vary the number of basis functions (treating l_{max} and n_{max} as a single variable) to attempt to observe any changes in the number of unique local atomic environments produced. The results are plotted in Fig. 2.5. Since there was not a compelling argument to changing the σ value, it was left at σ = 0.5 Å.

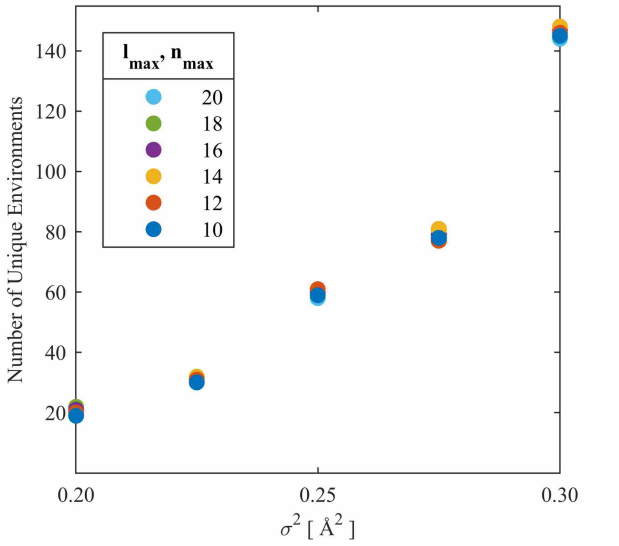

Figure 2.5: Plot of the number of unique local atomic environments as a function of the Gaussian width, for several different l_{max}, n_{max} values.

This parameter selection effort is not a thorough sampling, but is a more thoughtful than the previous application of the local environment representation to GBs (which was trying to prove that the local environment representation was capable of useful GB atomic structure description in the first place) [27]. These new parameters have the following benefits (compared to the previously used parameters in [27]):

- Consistent patterns in the local environment representation of the C-structural unit (from the structural unit model).
- A smaller set of unique local atomic environments has good prediction capability. Rosenbrock repeated his work of [27] using our SOAP parameters; he saw slightly better prediction of GB properties in the repeated simulation over the original. This slight improvement comes even as the machine uses less than half the number of unique local atomic environments and one third the number of coefficients in the SOAP vectors (1,015 instead of $\sim$3,000) compared to the original input.
- Disjoint unique local atomic environments.
- The local environment representation is less sensitive to the choice of SOAP parameters (e.g. r_{cut}).

2.6 SPRING Representation

We borrow a high-dimensional visualization tool from the field of biology to display the relationships between GBs based on their local environment representation. This tool, known as SPRING, was designed for visualizing gene expression data in single cells, so that researchers could grasp the relationships between differentiated cells (e.g. stem cells and their derivatives). SPRING constructs a directed k-nearest neighbor graph in the high-dimensional (fifty-five-dimensional) local environment representation space, and then displays a particular two-dimensional projection that preserves proximity of individual points (i.e. points that are close together/far apart in the fifty-five-dimensional space remain close together/far apart in the two-dimensional projection) using a force-directed layout algorithm [41].

The first step of the SPRING visualization determines the edge list, or proximity of points in the fifty-five-dimensional space. The edge list is an inventory of the connections between the nodes (each node corresponds to a GB). The concept of connection, in this context, signifies similarity in the local atomic representation of the two GBs, or atomic structural similarity. To compile the edge list, the GB local environment representation data is filtered, removing outliers and non-contributing unique environments. The GBs are then described in terms of the principal components of the local environment representation of the GB set. The Euclidean distances between all of the GBs in the principal component space are calculated, and the k-nearest neighbors (for each GB) are determined from this Euclidean distance matrix (for this work, k = 6). A GB may have more than k edges associated with it — GBs other than its k-nearest neighbors may view it as being "close" to them, pushing the number of edges beyond k (in graph theory terminology, the out-degree will be k, the in-degree might be above zero, and the total-degree can thus exceed k). Groups of nodes with large numbers of intra-group connections are known as clusters, and represent subsets of GBs with high structural similarity.

The second step of the SPRING visualization uses a force-directed algorithm to effect the dimensionality reduction. The layout of the network is achieved by minimizing the energy in a system representing the network: the system stipulates there is universal repulsion between all nodes, and attractive forces assigned between edge-connected nodes. These attractions are all of the same character (instead of depending on factors such as the Euclidean distance in hyperdimensional space). Further details can be found in [42], which implements a force-directed algorithm very similar to that of SPRING.

SPRING has two advantages that we wish to exploit: the visualization methodology captures the clustering of grain boundary atomic structures *and* preserves the structural relationships (or proximity) between the boundaries (and the clusters) [41]. By capturing the cluster and connection information, the SPRING visualization tool provides researchers with a sense of the similarities between nodes: spatial proximity directly correlates with structural similarity. Despite these advantages, this technique is often useful only as a preliminary exploration of high-dimensional space, allowing researchers to identify the regions that deserve further scrutiny.

CHAPTER 3. GB ATOMIC STRUCTURE: DESCRIPTOR COMPARISON AND RELATIONSHIP TO CRYSTALLOGRAPHY

3.1 [1 0 0] - Symmetric Tilt GB Atomic Structures

A representative set of atomic structures that this work investigates are plotted in Fig. 3.1, to summarize the atomic structures of the entire set of [1 0 0] - symmetric tilt GBs. The main plot of Fig. 3.1 depicts GB energy against tilt angle. The subplots depict the periodic atomic structures of selected GBs, and are labeled with the tilt angle of the corresponding GB. A line connects each subplot with the respective GB on the energy-tilt angle plot. The inset of Fig. 3.1 contains example structural units for easy reference.

The atoms of the atomic structure subplots are colored according to the similarity of the unique local atomic environment classifications to FCC, with the color scale plotted in the upper right corner of Fig. 3.1 (this coloring is not a full representation of the classification data, since it is a one-dimensional similarity to a single environment). Since symmetric tilt GBs lead to quasi-two-dimensional atomic structures, each atom in the plot represents a column of atoms (except in the U-structural unit regions). We use the unique local atomic environment classification of the first atom of each column as being representative of the whole column. The structural units are outlined in gray, with no structural unit labels (e.g. A, B) to prevent intrusive annotation.

3.1.1 Structural Unit Model Representation

We first discuss the evolution of the [1 0 0] - symmetric tilt GB atomic structures using the structural unit model, with Fig. 3.1 as our guide. At low tilt angles, the atomic structures consist of C-structural units evenly dispersed in a matrix of B-structural units (see 9.5^{o}, 10.4^{o}, and 21.0^{o} of Fig. 3.1). We also observe regions where the quasi-two-dimensionality of the GBs is lost. These regions are labeled as U-structural units, with U denoting undefined. As the misorientation angle increases, the separation between the C-structural units decreases (note the number of B-structural

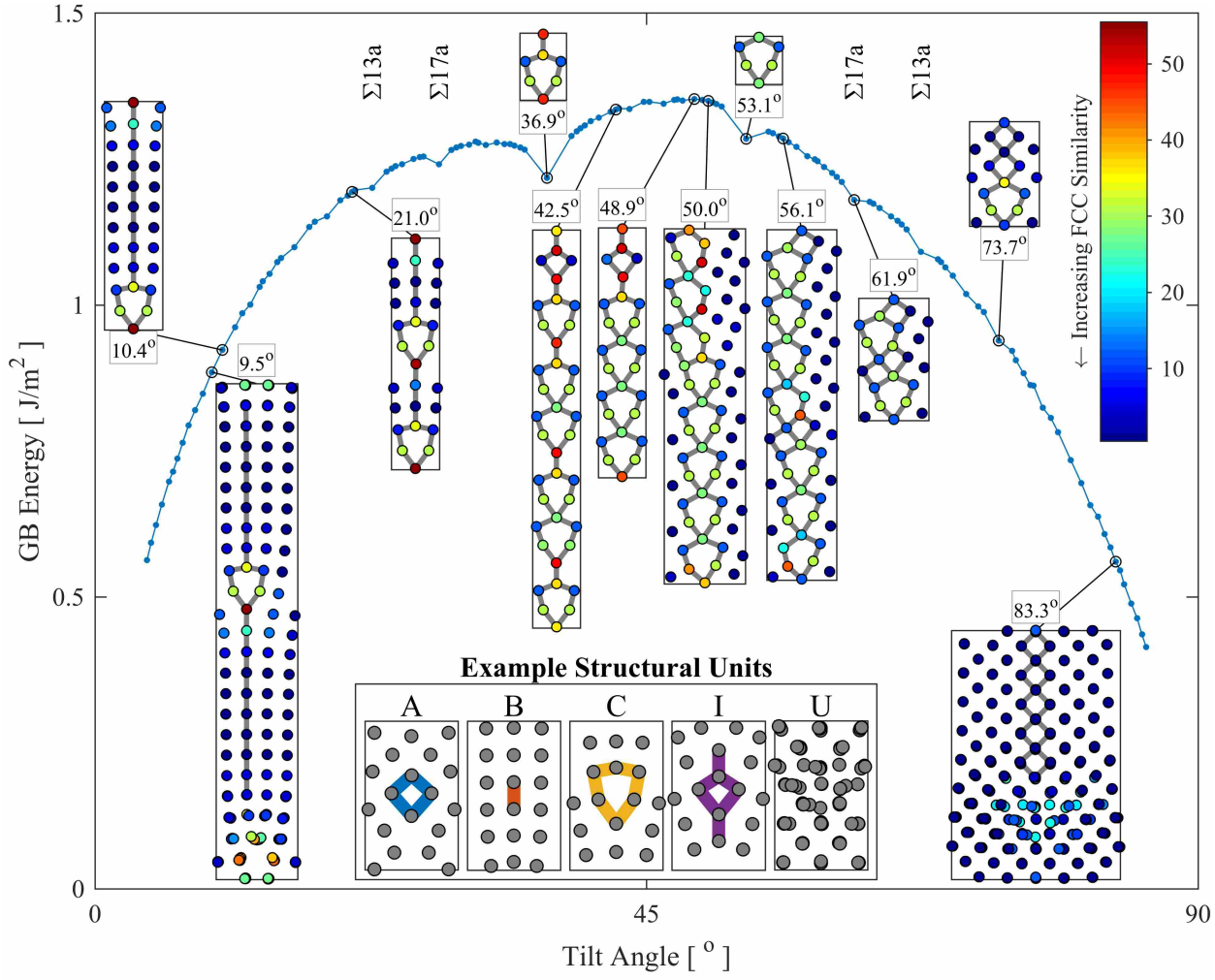

Figure 3.1: Summary of [1 0 0] - symmetric tilt GB atomic structures used in this work. The main plot is GB energy against tilt angle; the subplots present representative atomic structures. The atoms are colored according to their unique local atomic environment classification, with a color scale in the top right corner of the figure. Structural units are outlined in gray, with structural unit conventions provided in the bottom center of the figure.

units between the C-structural units for 10.4^o and 21.0^o in Fig. 3.1). Eventually, the C-structural units alternate with B-structural units (36.9^o).

Between 36.9^o and 53.1^o (which are the much studied Σ5 symmetric tilt boundaries), the boundaries undergo a change from boundaries composed of B- and C-structural units to boundaries composed of C-structural units only. Chains of C-structural units start to emerge (compare 42.5^o, 48.9^o, and 50.0^o) as the B-structural units disappear. The I-structural unit emerges in this region,

only to disappear as C-structural units impinge on one another (compare 48.9° and 50.0°). The impinging C-structural units eventually relax to form a perfect chain of C-structural units (53.1°).

Beyond 53.1°, the A-structural unit becomes increasingly predominant, causing sub-chains of C-structural units to again occur (56.1°); eventually individual C-structural units remain (61.9° and 73.7°). At high misorientation angles, U-structural units replace the C-structural units (83.3°). We stress that these results are not novel; many previous investigators have examined the atomic structures of grain boundaries [5–8, 11–17], particularly the [1 0 0] - symmetric tilts [5, 6, 9, 10, 15, 17]. To our knowledge, our results explore the [1 0 0] - symmetric tilts with greater resolution in the [1 0 0] - symmetric tilt angle than previous work.

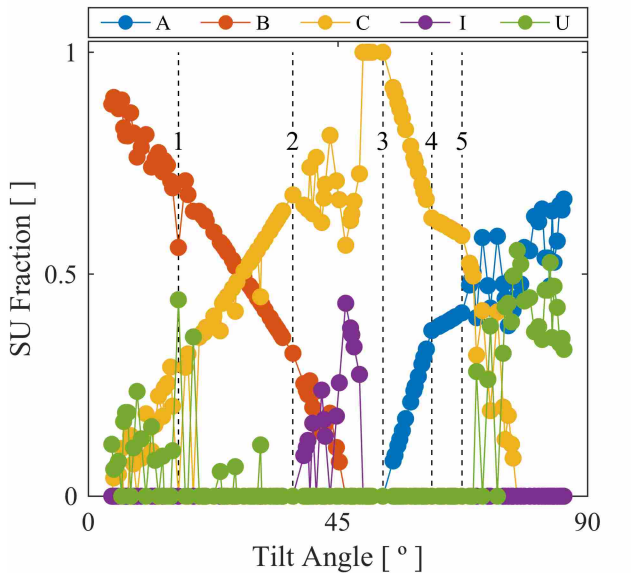

Figure 3.2: Structural unit length fractions in the periodic atomic structures of the [1 0 0] symmetric tilt grain boundaries. The labeled GBs are 1) Σ25*a* (0 $\overline{1}$ 7), 2) Σ5 (0 $\overline{1}$ 3), 3) Σ5 (0 $\overline{2}$ 1), 4) Σ17*a* (0 $\overline{5}$ 3), and 5) Σ13*a* (0 $\overline{3}$ 2). The reader is referred to Fig. 3.1 for structural unit conventions.

To numerically summarize the structural unit model representation of the [1 0 0] - symmetric tilts, we plot the structural unit content of the GBs with respect to misorientation angle in Fig. 3.2. The structural unit content is quantified as the fractional length of each structural unit within each GB's periodic atomic structure. Fig. 3.2 tells us about the relative occurrence of the structural units in each boundary but little else. Some investigators have devised or used schemes for representing the arrangements of structural units in GBs, using symbols and structural unit

identifiers to convey the connectivity of the structural units at the boundary (e.g. |AB·AB| [17], |6(D.D)/DC| [14]) [6, 12–14, 17]. Usually these approaches are discrete, if quantitative in nature, making numerical examination of structural unit arrangements difficult due to the lack of a smoothly varying structural descriptor.

3.1.2 Local Environment Representation and the C Structural Unit

In Fig. 3.1, the majority of the C-structural units appear to have the same pattern of unique local atomic environment classification. Given the occurrence of the C-structural unit throughout the [1 0 0] - symmetric tilt space, we focus on the relationship between the local environment representation and the structural unit model using the C-structural unit.

Since a C-structural unit is composed of multiple columns of atoms, we establish an atomic position identification convention in Fig. 3.3(a), numbering the atom column positions 1 through 6. We identify the C-structural units within the periodic structure for each GB, and compile the viewing plane positions of the atom columns of these C-structural units for analysis. We also compile the unique local atomic environment classifications for the atom columns, taking classification of the top atom–the visible atom–of each column to be representative of the classifications of the remaining atoms in the column (only 0.28% of remaining atoms differ in classification from the top atom).

To examine the trends of the C-structural unit with respect to tilt angle, we plot the atoms of each C-structural unit of our data set, treating each C-structural unit as a planar entity (see Fig. 3.1), and align all of the units at the 1 atom position. Fig. 3.3(b) illustrates this visual of all the C-structural unit atoms plotted as described. For GBs having more than one unique C-structural unit in their periodic structure (e.g. 50.0^{o} in Fig. 3.1), the atom positions of each unique unit are plotted with a different symbol (e.g. circle, square, triangle). The atom positions are colored according to their unique local atomic environment classification. The atom position markers are plotted with some transparency to enhance the visibility of the variability within atomic positions, at the cost of some ability to differentiate between different unique local atomic environments. The areas enclosed by the 36.9^{o} and 53.1^{o} (the $\Sigma 5$) GBs' C-structural units are shaded gray. As there is a general tilt angle-position trend for the 2, 3, 5, and 6 positions, we highlight the general increase

in tilt angle with small arrows that correspond to the color of the predominant environment for these positions.

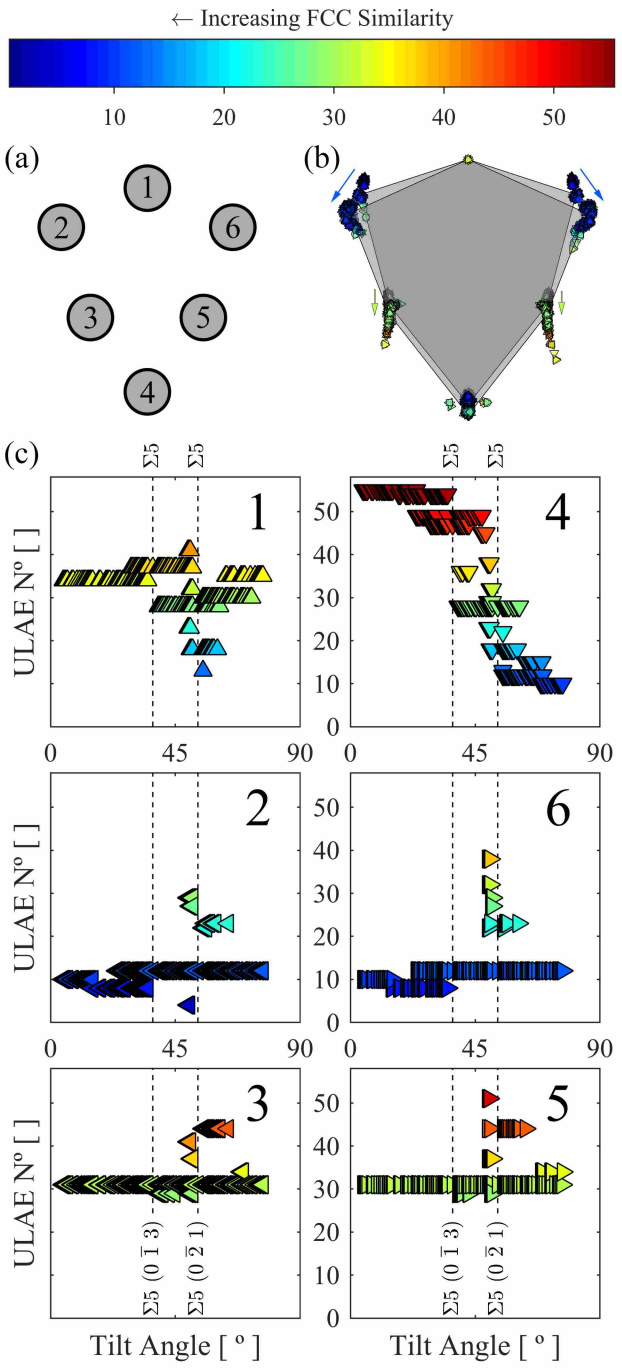

Figure 3.3: Unique local atomic environments trends in the C-structural unit. (a) portrays an atomic position numbering cartoon. (b) plots unique C-structural unit atoms, with atoms colored according to their unique local atomic environment classification (represented by the color scale at the top of the figure. The colored arrows near the 2, 3, 5, and 6 positions indicate the trend for atom position with respect to tilt angle. (c) plots the unique local atomic environment classification against tilt angle for the individual atom positions of the C-structural unit. The tilt angles of the Σ5 $(0\ \bar{1}\ 3$ and $(0\ \bar{2}\ 1)$ GBs are denoted with dashed lines.

We now refer the reader to Fig. 3.3(c), which serves to elucidate the atomic environment classifications occurring in the C-structural units. Each of the six subplots of Fig. 3.3(c) depicts the unique local atomic environment number of the atoms present as a function of the tilt angle, for each of the different C-structural unit atomic positions (see Fig. 3.3(a)). To facilitate comparisons, the symbols retain their coloring according to the color scale of Fig. 3.3. For reference, the dashed lines correspond to the two Σ5 symmetric tilt GBs.

For the 2, 3, 5, and 6 C-structural unit atom positions, we observe that the atoms for these positions are of a nearly uniform unique environment classification. What we mean is, the unique environment classification of these atom positions is nearly independent of tilt angle. There are, however, notable deviations in the GBs around the Σ5 $(0\ \overline{2}\ 1)$ GB that we attribute to the atomic structures attempting to achieve the continuous chain of C-structural units present in the Σ5 $(0\ \overline{2}\ 1)$ GB. The approximations in these near-Σ5 $(0\ \overline{2}\ 1)$ GBs involve structural units present in these boundaries that we will identify as distorted C-structural units, or distorted kites. We define these units by the linearity of the 4-3-2 or 4-5-6 positions of the C-structural unit and overall symmetry of the C-structural unit. If the 3 or 5 position deviates enough to cause asymmetry in the C-structural unit and a noticeable angle to occur in the 4-3-2 or 4-5-6 connections, we refer to this C-structural unit as a distorted kite.

Fig. 3.4 presents boundaries with representative distorted C-structural units. In Fig. 3.4, each boundary is identified at the top of its plot by the tilt angle, with the 53.1^{o} GB being the Σ5 $(0\ \overline{2}\ 1)$ boundary. The atoms within the periodic structures of Fig. 3.4 follow the unique local atomic environment color scheme for the atoms used throughout this work; repeats of the periodic structures shown are plotted in gray. The C-structural units within the periodic structure of each GB are numbered for referencing. For the 50.0^{o} GB, the distorted kites are the 1, 2 , 3, and 8 units; for the 51.1^{o} GB, the 6 and 12 kites are distorted; and the 55.3^{o} GB has distorted kites at 5 and 10. We attribute the occurrence of the distorted kite structures to be a result of the structure attempting to achieve the simply connected chain of C-structural units (the 53.1^{o} GB), and as noted, the local environment representation manifests changes that reflect these distorted structures.

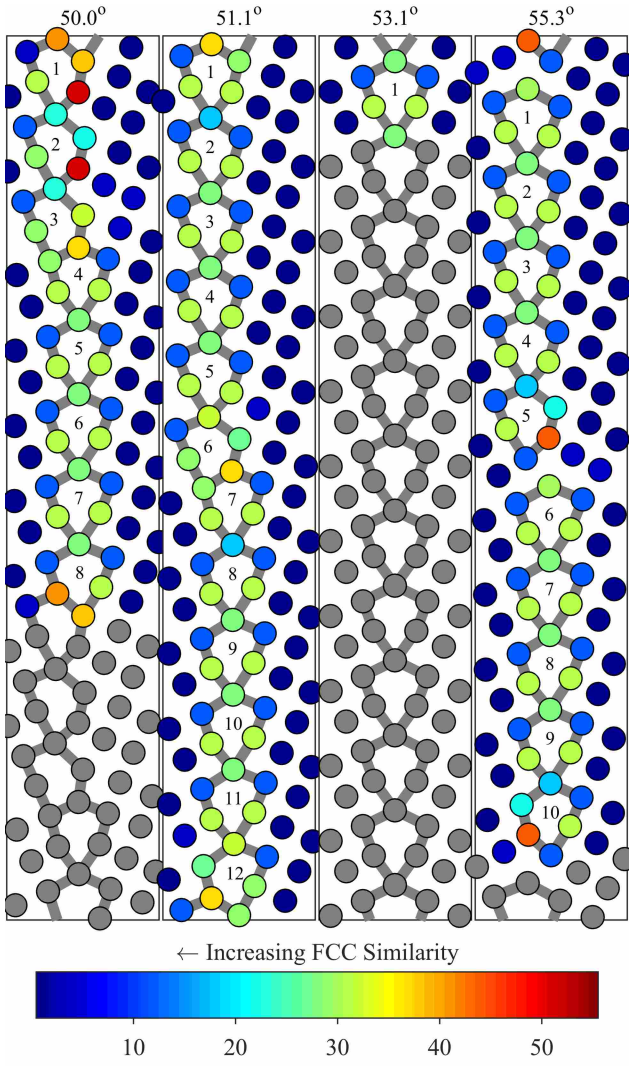

Figure 3.4: Representative structures with distorted C-structural units. Each GB is labeled with its tilt angle above the plot. The atoms within the periodic structure are colored according to their unique local atomic environment classification, with a color scale underneath the plot; atoms in repeat periods are colored gray. The periodic C-structural units are outlined in gray and numbered for reference. The 1, 2, 3, and 8 C-structural units of the 50.0° GB are distorted, the 6 and 12 C-structural units of the 51.1° GB are distorted, and the 5 and 10 C-structural units of the 55.3° GB are distorted.

We highlight specifically the presence of an environment at the junction of impinging C-structural units (as seen in the 50.0° GB: the 1 atom of C-structural unit 1 and the 3 atom of C-structural unit 8, see Fig. 3.4). This environment occurs when the 1 position references the 3 or 5 position (see Fig. 3.3(a)) of the C-structural unit above it.

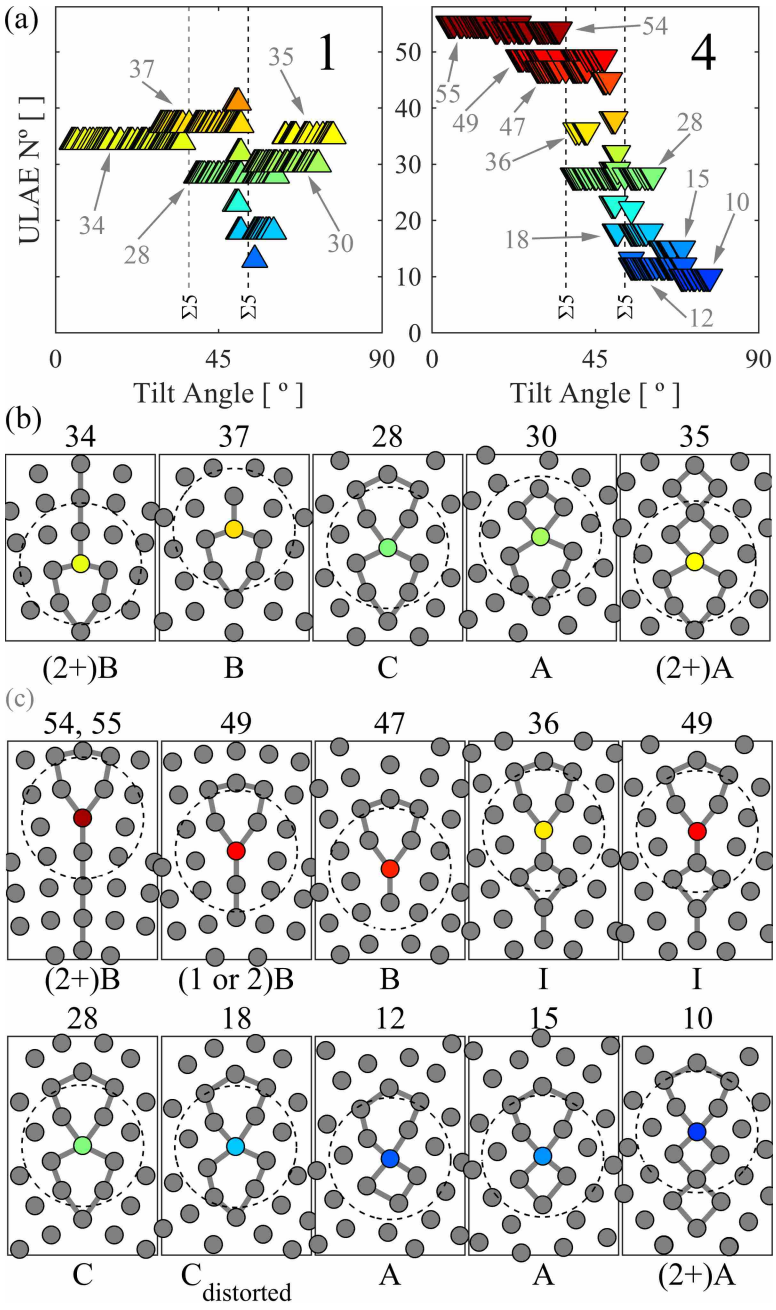

Figure 3.5: Structural unit arrangements and unique local atomic environment classifications for the 1 and 4 positions of the C-structural unit. In (a), we repeat part of Fig. 3.3(c) and annotate specific unique local atomic environment numbers. In (b), representative atomic structures occurring at the 1 position are plotted. Each atomic structure is labeled with the typical unique local atomic environment number of the atom at the 1 position (which is colored according to its unique local atomic environment classification (see Fig. 3.3), with all other atoms plotted in gray. The structural units are outlined in gray (see the inset of Fig. 3.1), and the structural unit arrangements intersecting at the 1 position are detailed below each plot. We denote r_{cut} with dashed lines. (c) presents representative atomic structures for the major unique local atomic environments occurring at the 4 position, and follows the plotting conventions of (b).

The reasons for the other deviations in the unique environment classification of the 2, 3, 5, and 6 positions are currently unexplained (for example, see the classifications of the 2 and 6 positions at low tilt angle). From the atomic structures, the differing atoms manifest no readily discernible differences with those atoms that follow the general trend for the corresponding atom position.

While the 2, 3, 5, and 6 atom positions of the C-structural unit remain largely the same throughout the range of [1 0 0] - symmetric tilt angles, the 1 and 4 atom positions manifest consistent changes that reflect the structural unit arrangements at the boundary. Fig. 3.5 is a visual aid to explaining these correlations. In Fig. 3.5(a), we repeat the plots for the 1 and 4 positions of Fig. 3.3(c), and include annotations that explicitly indicate the unique local atomic environment numbers. In Fig. 3.5(b), we plot representative atomic structures containing the major unique local atomic environments occurring at the 1 position. Fig. 3.5(c) plots representative atomic structures containing the major unique local atomic environments occurring at the 4 position. Under each representative atomic structure in Fig. 3.5(b-c), we include the structural units above (for atoms in position 1) or below (for atoms in position 4) the unique local atomic environment (e.g. (2+)B indicates that unique local atomic environment 54 occurs when a C-structural unit is joined below with a chain of 2 or more B-structural units).

In Fig. 3.5, we see that the major unique local atomic environments found at C-structural unit atom positions 1 and 4 correspond to specific structural unit arrangements (with unique local atomic environment 49 being the exception, correlating to 1 or 2 B-structural units and the I-structural unit). Specific structural unit arrangements above and below the C-structural unit also tend to coincide with specific unique local atomic environments, although there are several arrangements that appear to map to two unique local atomic environments (e.g. (2+)B with unique local atomic environments 54 and 55; B with unique local atomic environments 49 and 47; I with unique local atomic environments 36 and 49). In other words, the unique local atomic environments inherently encodes long-distance information about the atomic structure of GBs. Furthermore, in the case of the local environment representation of the [1 0 0] - symmetric tilts, the presence of a given unique local atomic environment in a GB signals that a certain structure, such as a specific structural unit, occurs in that boundary. In fact, it encodes information we cannot readily see, as unique local atomic environments 54 and 55 appear identical to us, but are different according to

the local atomic environment representation. This may be an artifact of choosing an equivalence threshold (ε) that is too stringent; a relaxed value of ε may combine the two unique local atomic environments.

In sum, the local environment representation is not always as easily interpreted as the structural unit model, but has other desired characteristics for an ideal atomic structure descriptor (e.g. easily visualized with atom coloring, invariance with respect to symmetries and rotations, accommodates structural perturbations, and most importantly, can be objectively applied to any simple or complex atomic arrangement).

3.1.3 SPRING Representation

Using the SPRING visualization method to evaluate the similarity of the GB atomic structures to each other, we construct the representation that appears in Fig. 3.6. We will refer to this figure (and its derivatives) as the GB atomic structure similarity network. The color scheme is plotted as a color scale to the right of the plot, and is indicative of the [1 0 0] - symmetric tilt angle.

Fig. 3.6 allows the reader to rapidly visualize the globally smooth variation of the tilt angle within the GB atomic structure similarity network. The nodes, which represent GBs, are numbered instead of labeled with their tilt angle to simplify visualization, but the ordering increases with tilt angle. The plot highlights the smooth variation, but also highlights that some GBs do have local deviations in spite of the global trend (e.g. the 25^{th} GB is similar to and appears in a cluster of GBs in the teens). Fig. 3.6 strongly implies a relationship between atomic structure and crystallography of the [1 0 0] - symmetric tilts GBs. This relationship is not necessarily surprising considering that one can see it visually, but this method is objective and can readily be applied to other GB structures to examine similarity/dissimilarity and their connection to properties or macroscopic structure.

The local environment representation provides additional information regarding GB atomic structure. To illuminate the connection between the local environment representation and the structural unit model, we duplicate the GB atomic structure similarity network of Fig. 3.6, but instead color the nodes with the structural unit content in Fig. 3.7. GBs colored white have no presence of the corresponding structural unit, to effectively distinguish between the GBs that have the structural unit present and those that do not.

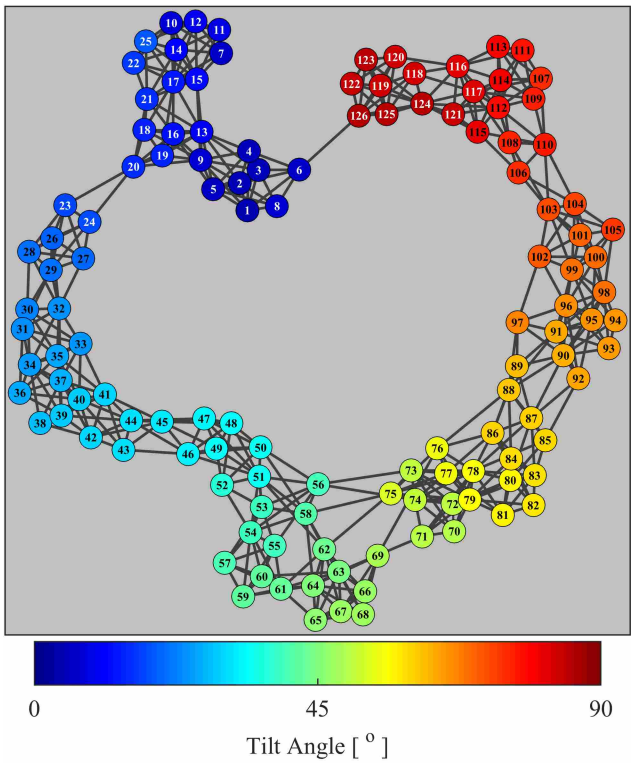

Figure 3.6: Force-directed visualization of the [1 0 0] symmetric tilt GBs using SPRING. The points representing the GBs are colored according to their [1 0 0] misorientation angle, as depicted in the color scale to the right of the plot. The lines (or 'edges') between the points indicate a nearest-neighbor relation. The coloring of the node (GB) labels serves only to improve readability.

Fig. 3.7 provides a strong indication that the SOAP descriptor is a superior method for quantifying GB atomic structures. Despite having only information about the unique local atomic environment content in each GB, the GB atomic structure similarity network clusters the GBs such that there is segregation of the structural units within the network. Furthermore, this process happens with little human input. This algorithmic process is in contrast with the structural unit model, which requires manual identification of the structural units and (sometimes) elaborate descriptions of the structural unit ordering at the interface. These results imply that the SOAP descriptor incorporates all of the information of SUM with additional knowledge of the atomic structure.

3.1.4 Unique Local Atomic Environments and the U-Structural Unit

We can also use the local environment representation and the SPRING tool to examine three-dimensional structures, given the success here with the U-structural unit. The U-structural

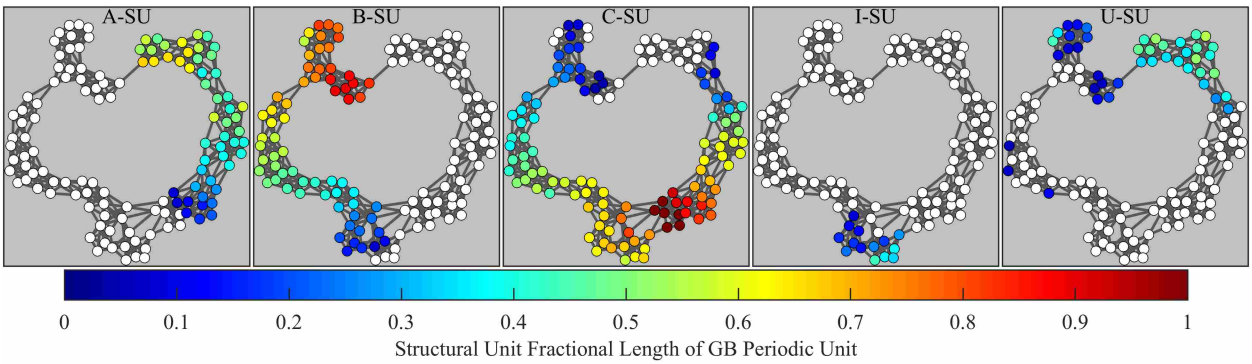

Figure 3.7: SPRING visualization of the [1 0 0] symmetric tilt GBs. The points representing the GBs are colored according to the structural unit content in the GBs, as depicted in the color scale to the right of the plot.

units are not really structural units, but simply regions that have no quasi-two-dimensionality and therefore could not be readily interpreted using the structural unit model characterization of a GB (traditional structural unit model descriptions are only used on quasi-two-dimensional structures, not fully three-dimensional atomic structures). The local environment representation (based on the SOAP descriptor) is easily extended to three-dimensional structures, and we readily obtain classifications for the atoms in the U-structural unit. Furthermore, by applying the SPRING visualization tool we are able to identify trends atomic structure across the [1 0 0] tilt angle, even when the quasi-two-dimensional structures break down and turn into completely three-dimensional structures. The classification of the U-structural units with the local environment representation provides additional evidence that local environment representation is a capable tool for numerical description of GB atomic structure.

CHAPTER 4. CONCLUSIONS

In this work we have examined the simulated atomic structures of [1 0 0] - symmetric tilt GBs in nickel using the structural unit model and the local environment representation. We have compared the two descriptors to evaluate the capability of the local environment representation. We have demonstrated that the local environment representation is capable of connecting GB crystallographic structure and atomic structure in an intuitive way that is automated. The major contributions from this work are summarized in the following list:

- We discover that complex GBs use the same atomic structures as simple GBs, evidenced by the addition of only 4 new unique local atomic environments to classify one hundred new GBs.

- We demonstrate that the use of the SOAP descriptor and the concept of unique local atomic environments allows us to fully describe the atomic structures of the [1 0 0] - symmetric tilt GBs; these tools are readily applicable to any complex GB structure.

- The uniform, geometrical character of the the C-structural units (excluding the distorted units), tied with the near-constant unique local atomic environment classifications of the 2, 3, 5 and 6 atom positions in these units indicates that the local environment representation is capturing the information of the structural unit model.

- The correlations between the unique local atomic environment classifications of the 1 and 4 atom positions and the structural units above and below the C-structural unit, respectively, point to the local environment representation encoding information regarding the arrangement of structural units in the boundary.

- The local environment representation of GB structures, coupled with the SPRING visualization tool, yields a map of the similarities between [1 0 0] - symmetric tilt GB structures

known as the GB atomic structure similarity network. This network traces the evolution of [1 0 0] - symmetric tilt GB atomic structures in a way readily understood by researchers.

- Coloring the nodes of the GB atomic structure similarity network according to their [1 0 0] - symmetric tilt angle produces a globally smooth evolution of tilt angle across the network. This smooth evolution implies a strong relationship between [1 0 0] - symmetric tilt angle and atomic structure.

- Coloring the nodes of the GB atomic structure similarity network to portray the relative strength of structural units in each boundary reveals that the structural units are confined to specific regions within the network, indicating that the local environment representation captures the information of the structural unit model and information regarding the arrangement of the structure units at the interface.

- Given the success of the local environment representation in characterizing the U-structural units, the local environment representation is a tool that can provide meaningful insight into three-dimensional GBs in addition to quasi-two-dimensional boundaries.

We also summarize the comparison of the structural unit model and the local environment representation in Table 4.1, where a checkmark indicates that the descriptor has a particular property; "User" denotes that the tool's possession of a particular property depends on the researcher executing the atomic structure classification.

Finally, we reiterate the motivations for this project: to lay the foundation for future prediction of GB structure-property relationships by connecting crystallography and atomic structure and exploring means for elucidating GB physics. We have seen that the unique local atomic environments correlate to specific structural features that may have connection with GB physics (as Rosenbrock observes [27]) and that there are ways to objectively connect crystallography and atomic structure. We anxiously await further additions of GBs to the dataset for potential correlations of more crystallographic parameters.

Table 4.1: Summary of Ideal Structural Descriptor Properties and Structural Descriptors. The structural unit model is abbreviated as SUM and the local environment representation as LER.

Property	SUM	LER
Easily Visualized	✓	✓
Easily Interpreted	✓	
Comparison	✓	✓
Invariance	User	✓
Perturbations	User	✓
Smoothly Varying		✓
Complex 3D GBs Structures		✓
Automation		✓
Connectivity	User	
Subunit Discovery	User	✓

REFERENCES

[1] Chiba, A., Hanada, S., Watanabe, S., Abe, T., and Obana, T., 1994. "Relation Between Ductility and Grain Boundary Character Distributions in Ni3Al." *Acta Metallurgica et Materialia,* **42**(5), pp. 1733–1738. 1

[2] Zhang, Y., Wang, J., Shan, H., and Zhao, K., 2015. "Strengthening high-stacking-fault-energy metals via parallelogram nanotwins." *Scripta Materialia,* **108**, pp. 35–39. 1

[3] Watanabe, T., Tsurekawa, S., Zhao, X., and Zuo, L., 2009. *The Coming of Grain Boundary Engineering in the 21st Century*. Springer London, London, ch. 4, pp. 43–82. 1

[4] Shimada, M., Kokawa, H., Wang, Z. J., Sato, Y. S., and Karibe, I., 2002. "Optimization of grain boundary character distribution for intergranular corrosion resistant 304 stainless steel by twin-induced grain boundary engineering." *Acta Materialia,* **50**(9), pp. 2331–2341. 1

[5] Frost, H., Ashby, M., and Spaepen, F., 1982. "A catalogue of [100], [110], and [111] symmetric tilt boundaries in face-centered cubic hard sphere crystals.." *Harvard Division of Applied Sciences*, pp. 1–211. 1, 3, 20

[6] Sutton, A. P., and Vitek, V., 1983. "On The Structure of Tilt Grain Boundaries In Cubic Metals I. Symmetrical Tilt Boundaries." *Philosophical Transactions of the Royal Society A,* **309**(1506), pp. 1–36. 1, 3, 7, 20, 21

[7] Sutton, A. P., and Vitek, V., 1983. "On The Structure of Tilt Grain Boundaries in Cubic Metals II. Asymmetrical Tilt Boundaries." *Philosophical Transactions of the Royal Society A,* **309**(1506), pp. 37–54. 1, 3, 7, 20

[8] Sutton, A. P., and Vitek, V., 1983. "On The Structure of Tilt Grain Boundaries In Cubic Metals. III. Generalizations of The Structural Study and Implications For The Properties of Grain Boundaries." *Philosophical Transactions of the Royal Society A,* **309**(1506), pp. 55–68. 1, 3, 20

[9] Wang, G. J., Sutton, A., and Vitek, V., 1984. "A COMPUTER SIMULATION STUDY OF <001> AND <111> TILT BOUNDARIES: THE MULTIPLICITY OF STRUCTURES." *Acta Metallurgica,* **32**(7), pp. 1093–1104. 1, 3, 7, 20

[10] Balluffi, R., and Bristowe, P., 1984. "On the Structural Unit/Grain boundary Dislocation Model for Grain Boundary Structure." *Surface Science,* **144**(1), pp. 28–43. 1, 3, 7, 20

[11] Rittner, J. D., and Seidman, D. N., 1996. "110 Symmetric Tilt Grain-Boundary Structures in Fcc Metals With Low Stacking-Fault Energies." *Physical Review B,* **54**(10), pp. 6999–7015. 1, 3, 7, 20

[12] Tschopp, M. A., Tucker, G. J., and McDowell, D. L., 2007. "Structure and free volume of <1 1 0> symmetric tilt grain boundaries with the E structural unit." *Acta Materialia,* **55**(11), pp. 3959–3969. 1, 3, 7, 20, 21

[13] Tschopp, M. A., and McDowell, D. L., 2007. "Structural unit and faceting description of Σ3 asymmetric tilt grain boundaries." *Journal of Materials Science,* **42**(18), pp. 7806–7811. 1, 3, 7, 20, 21

[14] Tschopp, M. A., and McDowell, D. L., 2007. "Structures and energies of Σ3 asymmetric tilt grain boundaries in copper and aluminium." *Philosophical Magazine,* **87**(22), pp. 3147–3173. 1, 3, 7, 20, 21

[15] Tschopp, M. A., and McDowell, D. L., 2007. "Asymmetric tilt grain boundary structure and energy in copper and aluminium." *Philosophical Magazine,* **87**(25), pp. 3871–3892. 1, 3, 7, 20

[16] Spearot, D. E., 2008. "Evolution of the E structural unit during uniaxial and constrained tensile deformation." *Mechanics Research Communications,* **35**, pp. 81–88. 1, 3, 7, 20

[17] Han, J., Vitek, V., and Srolovitz, D. J., 2017. "The grain-boundary structural unit model redux." *Acta Materialia,* **133**, pp. 186–199. 1, 3, 7, 20, 21

[18] Patala, S., Mason, J. K., and Schuh, C. A., 2012. "Improved representations of misorientation information for grain boundary science and engineering." *Progress in Materials Science,* **57**(8), pp. 1383–1425. 1

[19] Patala, S., and Schuh, C. A., 2013. "Symmetries in the representation of grain boundary-plane distributions." *Philosophical Magazine,* **93**(5), pp. 524–573. 1, 6

[20] Olmsted, D. L., Foiles, S. M., and Holm, E. A., 2009. "Survey of computed grain boundary properties in face-centered cubic metals: I. Grain boundary energy." *Acta Materialia,* **57**(13), pp. 3694–3703. 1, 6

[21] Homer, E. R., 2015. "Investigating the mechanisms of grain boundary migration during recrystallization using molecular dynamics." *IOP Conference Series: Materials Science and Engineering,* **89**, p. 012006. 1, 6

[22] Amiri-Hezavhe, A., and Balluffi, R. W., 1993. "Apparatus for producing ultraclean bicrystals by the molecular beam epitaxy growth and ultrahigh vacuum bonding of thin films." *Review of Scientific Instruments,* **64**(10), pp. 2983–2992. 1

[23] Schwarz, S. M., Houge, E. C., Giannuzzi, L. A., and King, A. H., 2001. "Bicrystal growth and characterization of copper twist grain boundaries." *Journal of Crystal Growth,* **222**, pp. 392–398. 1

[24] Heinemann, S., Wirth, R., and Dresen, G., 2001. "Synthesis of feldspar bicrystals by direct bonding." *Physics and Chemistry of Materials,* **28**(10), pp. 685–692. 1

[25] Gorkaya, T., Molodov, D. A., and Gottstein, G., 2009. "Stress-driven migration of symmetrical h 1 0 0 i tilt grain boundaries in Al bicrystals." *Acta Materialia,* **57**, pp. 5396–5405. 1

[26] Ratanaphan, S., Olmsted, D. L., Bulatov, V. V., Holm, E. A., Rollett, A. D., and Rohrer, G. S., 2015. "ScienceDirect Grain boundary energies in body-centered cubic metals." *Acta Materialia,* **88**, pp. 346–354. 1

[27] Rosenbrock, C. W., Homer, E. R., Csányi, G., and Hart, G. L. W., 2017. "Discovering the building blocks of atomic systems using machine learning: application to grain boundaries." *npj Computational Materials,* **3**(May), pp. 1–7. 1, 2, 4, 8, 9, 11, 16, 32

[28] Honeycutt, J. D., and Andemen, H. C., 1987. "Molecular Dynamics Study of Melting and Freezing of Small Lennard-Jones Clusters." *J. Phys. Chem,* **91**(24), pp. 4950–4963. 3, 8

[29] Larsen, P. M., Schmidt, S., and Schiøtz, J., 2016. "Robust structural identification via polyhedral template matching." *Modelling and Simulation in Materials Science and Engineering,* **24**, p. 055007. 3

[30] Lazar, E. A., 2018. "VoroTop: Voronoi cell topology visualization and analysis toolkit." *Modelling and Simulation in Materials Science and Engineering,* **26**(1), p. 015011. 3

[31] Ashby, M., Spaepen, F., and Williams, S., 1978. "The Structure of Grain Boundaries Described as a Packing of Polyhedra." *Acta Metallurgica,* **26**(11), pp. 1647–1663. 3

[32] Banadaki, A. D., and Patala, S., 2017. "A three-dimensional polyhedral unit model for grain boundary structure in fcc metals." *npj Computational Materials*, pp. 1–13. 3

[33] Bishop, G., and Chalmers, B., 1968. "A coincidence Ledge Dislocation description of grain boundaries." *Scripta Metallurgica,* **2**(2), pp. 133–139. 3

[34] Lazar, E. A., Han, J., and Srolovitz, D. J., 2015. "Topological framework for local structure analysis in condensed matter." *Proceedings of the National Academy of Sciences of the United States of America,* **112**(43), pp. 5769–5776. 3

[35] Bartók, A. P., Kondor, R., and Csányi, G., 2013. "On representing chemical environments." *Physical Review B - Condensed Matter and Materials Physics,* **87**(18), pp. 1–16. 3, 8

[36] Homer, E. R., Patala, S., and Priedeman, J. L., 2015. "Grain Boundary Plane Orientation Fundamental Zones and Structure-Property Relationships.." *Scientific reports,* **5**, p. 15476. 6

[37] Foiles, S. M., and Hoyt, J. J., 2006. "Computation of grain boundary stiffness and mobility from boundary fluctuations." *Acta Materialia,* **54**, pp. 3351–3357. 6

[38] Plimpton, S. J., 1995. "Fast Parallel Algorithms for Short-Range Molecular Dynamics." *Journal of Computational Physics,* **117**(1), pp. 1–19. 6

[39] Szlachta, W. J., Bartók, A. P., and Csányi, G., 2014. "Accuracy and transferability of Gaussian approximation potential models for tungsten." *Physical Review B - Condensed Matter and Materials Physics,* **90**(10), pp. 1–6. 8

[40] Deringer, V. L., and Csányi, G., 2017. "Machine learning based interatomic potential for amorphous carbon." *Physical Review B,* **95**(9), pp. 1–15. 8, 15

[41] Weinreb, C., Wolock, S., and Klein, A. M., 2017. "Gene expression SPRING : a kinetic interface for visualizing high dimensional single-cell expression data." *Bioinformatics*, pp. 1–3. 16, 17

[42] Jacomy, M., Venturini, T., Heymann, S., and Bastian, M., 2014. "ForceAtlas2 , a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software." *PLoS ONE,* **9**(6), pp. 1–12. 17

APPENDIX A. GRAIN BOUNDARY DATA SET

Table A.1: [1 0 0] - Symmetric Tilt GB Crystallographic Data.

Beginning of Table A.1			
[1 0 0] - STGB θ	Σ	BPFZ Normal	MFZ θ
4.242	365a	$(0\ \overline{1}\ 27)$	4.242
4.581	313a	$(0\ \overline{1}\ 25)$	4.581
4.979	265a	$(0\ \overline{1}\ 23)$	4.979
5.453	221a	$(0\ \overline{1}\ 21)$	5.453
6.026	181a	$(0\ \overline{1}\ 19)$	6.026
6.360	325a	$(0\ \overline{1}\ 18)$	6.360
6.733	145a	$(0\ \overline{1}\ 17)$	6.733
7.153	257a	$(0\ \overline{1}\ 16)$	7.153
7.628	113a	$(0\ \overline{1}\ 15)$	7.628
8.171	197a	$(0\ \overline{1}\ 14)$	8.171
8.797	85a	$(0\ \overline{1}\ 13)$	8.797
9.527	145b	$(0\ \overline{1}\ 12)$	9.527
10.389	61a	$(0\ \overline{1}\ 11)$	10.389
11.421	101a	$(0\ \overline{1}\ 10)$	11.421
12.018	365b	$(0\ \overline{2}\ 19)$	12.018
12.680	41a	$(0\ \overline{1}\ 9)$	12.680
13.420	293a	$(0\ \overline{2}\ 17)$	13.420
13.686	317a	$(0\ \overline{3}\ 25)$	13.686
14.250	65a	$(0\ \overline{1}\ 8)$	14.250
14.863	269a	$(0\ \overline{3}\ 23)$	14.863

Continuation of Table A.1			
[1 0 0] - STGB θ	Σ	BPFZ Normal	MFZ θ
15.189	229a	$(0\ \overline{2}\ 15)$	15.189
16.260	25a	$(0\ \overline{1}\ 7)$	16.260
17.492	173a	$(0\ \overline{2}\ 13)$	17.492
17.945	185b	$(0\ \overline{3}\ 19)$	17.945
18.925	37a	$(0\ \overline{1}\ 6)$	18.925
20.016	149a	$(0\ \overline{3}\ 17)$	20.016
20.610	125a	$(0\ \overline{2}\ 11)$	20.610
20.983	377a	$(0\ \overline{5}\ 27)$	20.983
21.239	265b	$(0\ \overline{3}\ 16)$	21.239
22.620	13a	$(0\ \overline{1}\ 5)$	22.620
23.777	377c	$(0\ \overline{4}\ 19)$	23.777
24.190	205b	$(0\ \overline{3}\ 14)$	24.190
24.530	277b	$(0\ \overline{5}\ 23)$	24.530
25.058	85b	$(0\ \overline{2}\ 9)$	25.058
25.989	89a	$(0\ \overline{3}\ 13)$	25.989
26.481	305a	$(0\ \overline{4}\ 17)$	26.481
26.785	233b	$(0\ \overline{5}\ 21)$	26.785
28.072	17a	$(0\ \overline{1}\ 4)$	28.072
29.069	389b	$(0\ \overline{7}\ 27)$	29.069
29.487	193b	$(0\ \overline{5}\ 19)$	29.487
29.863	241b	$(0\ \overline{4}\ 15)$	29.863
30.510	65b	$(0\ \overline{3}\ 11)$	30.510
31.048	349d	$(0\ \overline{5}\ 18)$	31.048
31.284	337b	$(0\ \overline{7}\ 25)$	31.284
31.891	53b	$(0\ \overline{2}\ 7)$	31.891
32.779	157b	$(0\ \overline{5}\ 17)$	32.779
33.398	109a	$(0\ \overline{3}\ 10)$	33.398

Continuation of Table A.1			
[1 0 0] - STGB θ	Σ	BPFZ Normal	MFZ θ
33.855	289c	$(0\ \overline{7}\ 23)$	33.855
34.205	185d	$(0\ \overline{4}\ 13)$	34.205
34.708	281d	$(0\ \overline{5}\ 16)$	34.708
35.051	397f	$(0\ \overline{6}\ 19)$	35.051
36.870	5	$(0\ \overline{1}\ 3)$	36.870
38.880	325d	$(0\ \overline{6}\ 17)$	38.880
39.308	221d	$(0\ \overline{5}\ 14)$	39.308
39.598	353e	$(0\ \overline{9}\ 25)$	39.598
39.966	137d	$(0\ \overline{4}\ 11)$	39.966
40.450	205e	$(0\ \overline{7}\ 19)$	40.450
41.112	73c	$(0\ \overline{3}\ 8)$	41.112
42.075	97c	$(0\ \overline{5}\ 13)$	42.075
42.501	373g	$(0\ \overline{7}\ 18)$	42.501
42.741	305g	$(0\ \overline{9}\ 23)$	42.741
43.603	29a	$(0\ \overline{2}\ 5)$	43.603
44.760	169d	$(0\ \overline{7}\ 17)$	44.760
45.240	169d	$(0\ \overline{12}\ 5)$	44.760
46.397	29a	$(0\ \overline{7}\ 3)$	43.603
47.259	305g	$(0\ \overline{16}\ 7)$	42.741
47.499	373g	$(0\ \overline{25}\ 11)$	42.501
47.925	97c	$(0\ \overline{9}\ 4)$	42.075
48.888	73c	$(0\ \overline{11}\ 5)$	41.112
49.550	205e	$(0\ \overline{13}\ 6)$	40.450
50.034	137d	$(0\ \overline{15}\ 7)$	39.966
50.402	353e	$(0\ \overline{17}\ 8)$	39.598
50.692	221d	$(0\ \overline{19}\ 9)$	39.308
51.120	325d	$(0\ \overline{23}\ 11)$	38.880

Continuation of Table A.1			
[1 0 0] - STGB θ	Σ	BPFZ Normal	MFZ θ
53.130	5	$(0\ \overline{2}\ 1)$	36.870
54.949	397f	$(0\ \overline{25}\ 13)$	35.051
55.292	281d	$(0\ \overline{21}\ 11)$	34.708
55.795	185d	$(0\ \overline{17}\ 9)$	34.205
56.145	289c	$(0\ \overline{15}\ 8)$	33.855
56.602	109a	$(0\ \overline{13}\ 7)$	33.398
57.221	157b	$(0\ \overline{11}\ 6)$	32.779
58.109	53b	$(0\ \overline{9}\ 5)$	31.891
58.716	337b	$(0\ \overline{16}\ 9)$	31.284
58.952	349d	$(0\ \overline{23}\ 13)$	31.048
59.490	65b	$(0\ \overline{7}\ 4)$	30.510
60.137	241b	$(0\ \overline{19}\ 11)$	29.863
60.513	193b	$(0\ \overline{12}\ 7)$	29.487
60.931	389b	$(0\ \overline{17}\ 10)$	29.069
61.928	17a	$(0\ \overline{5}\ 3)$	28.072
63.215	233b	$(0\ \overline{13}\ 8)$	26.785
63.519	305a	$(0\ \overline{21}\ 13)$	26.481
64.011	89a	$(0\ \overline{8}\ 5)$	25.989
64.942	85b	$(0\ \overline{11}\ 7)$	25.058
65.470	277b	$(0\ \overline{14}\ 9)$	24.530
65.810	205b	$(0\ \overline{17}\ 11)$	24.190
66.223	377c	$(0\ \overline{23}\ 15)$	23.777
67.380	13a	$(0\ \overline{3}\ 2)$	22.620
68.761	265b	$(0\ \overline{19}\ 13)$	21.239
69.017	377a	$(0\ \overline{16}\ 11)$	20.983
69.390	125a	$(0\ \overline{13}\ 9)$	20.610
69.984	149a	$(0\ \overline{10}\ 7)$	20.016

Continuation of Table A.1			
[1 0 0] - STGB θ	Σ	BPFZ Normal	MFZ θ
71.075	37a	$(0\ \overline{7}\ 5)$	18.925
72.055	185b	$(0\ \overline{11}\ 8)$	17.945
72.508	173a	$(0\ \overline{15}\ 11)$	17.492
73.740	25a	$(0\ \overline{4}\ 3)$	16.260
74.811	229a	$(0\ \overline{17}\ 13)$	15.189
75.137	269a	$(0\ \overline{13}\ 10)$	14.863
75.750	65a	$(0\ \overline{9}\ 7)$	14.250
76.314	317a	$(0\ \overline{14}\ 11)$	13.686
76.580	293a	$(0\ \overline{19}\ 15)$	13.420
77.320	41a	$(0\ \overline{5}\ 4)$	12.680
77.982	365b	$(0\ \overline{21}\ 17)$	12.018
78.579	101a	$(0\ \overline{11}\ 9)$	11.421
79.611	61a	$(0\ \overline{6}\ 5)$	10.389
80.473	145b	$(0\ \overline{13}\ 11)$	9.527
81.203	85a	$(0\ \overline{7}\ 6)$	8.797
81.829	197a	$(0\ \overline{15}\ 13)$	8.171
82.372	113a	$(0\ \overline{8}\ 7)$	7.628
82.847	257a	$(0\ \overline{17}\ 15)$	7.153
83.267	145a	$(0\ \overline{9}\ 8)$	6.733
83.640	325a	$(0\ \overline{19}\ 17)$	6.360
83.974	181a	$(0\ \overline{10}\ 9)$	6.026
84.547	221a	$(0\ \overline{11}\ 10)$	5.453
85.021	265a	$(0\ \overline{12}\ 11)$	4.979
85.419	313a	$(0\ \overline{13}\ 12)$	4.581
85.758	365a	$(0\ \overline{14}\ 13)$	4.242