

## Journal Pre-proof

Learning the grain boundary manifold: tools for visualizing and fitting grain boundary properties

I. Chesser, T. Francis, M. De Graef, E.A. Holm

PII: S1359-6454(20)30363-3  
DOI: <https://doi.org/10.1016/j.actamat.2020.05.024>  
Reference: AM 16034



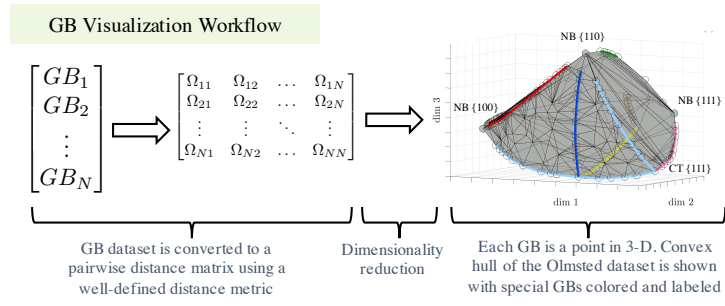
To appear in: *Acta Materialia*

Received date: 20 September 2019  
Revised date: 8 May 2020  
Accepted date: 10 May 2020

Please cite this article as: I. Chesser, T. Francis, M. De Graef, E.A. Holm, Learning the grain boundary manifold: tools for visualizing and fitting grain boundary properties, *Acta Materialia* (2020), doi: <https://doi.org/10.1016/j.actamat.2020.05.024>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2020 Published by Elsevier Ltd on behalf of Acta Materialia Inc.



# Learning the grain boundary manifold: tools for visualizing and fitting grain boundary properties

I. Chesser<sup>a</sup>, T. Francis<sup>b</sup>, M. De Graef<sup>a</sup>, E.A. Holm<sup>a,\*</sup>

<sup>a</sup>*Department of Materials Science and Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, USA*

<sup>b</sup>*Materials Department, University of California Santa Barbara, Santa Barbara CA, 93106-5050, USA*

---

## Abstract

With the proliferation of grain boundary data in materials science from both experiments and simulations, tools are needed to explore the five dimensional space of grain boundaries and visualize and fit structure property relationships along arbitrary paths through this space. In this work, we leverage a recently developed geodesic metric for grain boundaries to visualize the global geometry of grain boundary datasets and fit grain boundary energy to macroscopic grain boundary geometry. It is found that the 5D connectivity of the 388 grain boundary Olmsted dataset can be visualized via dimensionality reduction in 5D with a high degree of interpretability. Furthermore, after selectively adding new grain boundaries to the dataset, these visualizations suggest new global features of grain boundary space, including the existence of a grain boundary fundamental zone with well defined subsets of high symmetry boundaries along faces. Geodesic sampling is shown to be an effective tool to extend grain boundary datasets to new regions of the 5D space. Finally, a simple grain boundary energy kernel regression model with only one fitting parameter is demonstrated to predict grain boundary energy in the Olmsted dataset to within 10% RMSE.

*Keywords:* , ...

---

## 1. Introduction

Grain boundaries have five macroscopic crystallographic degrees of freedom: three correspond to the relative rotation, or misorientation, between grains, and two correspond to the unit normal inclination of the interface plane separating the grains. These crystallographic parameters may be  
 5 measured in serial-sectioning EBSD or 3D-XRD experiments [1, 2]. A wealth of grain boundary

---

\*E.A. Holm, Department of Materials Science and Engineering, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213-3890, USA.

*Email addresses:* [ichesser@andrew.cmu.edu](mailto:ichesser@andrew.cmu.edu) (I. Chesser), [tfrancis@ucsb.edu](mailto:tfrancis@ucsb.edu) (T. Francis), [degraeef@cmu.edu](mailto:degraeef@cmu.edu) (M. De Graef), [eaholm@cmu.edu](mailto:eaholm@cmu.edu) (E.A. Holm)

data may also be obtained from atomistic simulations. The canonical Olmsted survey computed the energy and mobility of 388 crystallographically diverse grain boundaries via molecular statics and dynamics simulations [3, 4]. With the proliferation of grain boundary data in materials science from both experiments and simulations, tools are needed to explore the five dimensional space of grain boundaries and visualize and fit structure property relationships along arbitrary paths through this space.

Current representations for grain boundary structure-property relationships are limited by the lack of an interpolation scheme between arbitrary boundaries. Consider two well known grain boundaries: the  $\Sigma 3$  (111) coherent twin boundary, the lowest energy and most common grain boundary in most FCC metals, and the frequently simulated  $\Sigma 5$  (210) boundary. It is unclear how to define a shortest path between these grain boundaries in the full five dimensional space. Yet structure property relationships for grain boundaries depend on such paths: Grain boundary properties are often represented as a function of misorientation angle for a fixed tilt or twist axis or as a function of boundary plane inclination for fixed misorientation [5, 6, 7, 8]. The distance between two boundaries in these high symmetry subspaces is implicitly given by the difference in tilt/twist angles or by the angular distance between unit normal boundary plane inclinations. However, it is not obvious whether a path on a given subspace is the shortest possible path, or geodesic, between two grain boundaries in the full 5D space.

A particularly convenient representation for grain boundaries, the grain boundary octonion representation, allows for interpolation between arbitrary pairs of grain boundaries along geodesics [9]. Furthermore, it allows for an unambiguous definition of shortest distances between grain boundaries in the 5D space for arbitrary point group symmetries [9]. Each grain boundary octonion (GBO) is a normalized 8-D vector on the 7 dimensional surface of a sphere ( $\mathbb{S}^7$ ). A GBO distance metric is defined on grain boundary space, the 5 dimensional subspace  $SO(3) \times SO(3)/U(1)$  of  $\mathbb{S}^7$ . Distance values given by the GBO metric match those of other properly symmetrized metrics on a set of reference cases [10, 11]. Unlike other distance functions, the GBO metric allows for computationally efficient interpolation between pairs of symmetrized grain boundaries along geodesics (great circles) on the surface of  $\mathbb{S}^7$ . Although such geodesics may deviate from the true space of grain boundaries, it has been shown that appropriately renormalized grain boundary octonions along such trajectories closely approximate geodesics on the true space of grain boundaries (the latter class of geodesics being difficult or intractable to compute) [9]. In this work, we will show that geodesic sampling is a convenient tool for extending grain boundary datasets to new regions of the



5D space.

Even though a variety of grain boundary metrics on the 5D space have previously been introduced, visualization methods have not been developed that leverage the ability to handle arbitrary grain boundary geometry. The GB character distribution and GB plane fundamental zone, though useful for visualizing structure property relationships, only consider variation in GB inclination [5, 6, 7, 8]. Standard tilt/twist trajectories only capture one parameter and apply only to high symmetry families of grain boundaries. The geodesic angle in interpolated GB trajectories generalizes the concept of misorientation angle to any shortest path between two grain boundaries. The ability to compute shortest distances between every pair of grain boundaries in a dataset allows visualization of the global connectivity of the dataset, either directly through projection methods (where each GB is represented as a pair of points in 5D) or indirectly via dimensionality reduction (where each GB is represented as a point in 3D).

5D metrics have yet to be incorporated into models for grain boundary properties. The well known Bulatov-Reed-Kumar (BRK) model for grain boundary energy only defines metrics on misorientation and boundary plane subspaces [12]. We show in this work that a 5D metric on the space of grain boundaries allows for simplification of grain boundary energy regression models. Likewise, models for grain boundary migration are usually parametrized only as a function of misorientation angle for high symmetry classes of grain boundaries or more recently as a function of boundary plane inclination [13]. We demonstrate that visualizations of mobility in the full 5D space help quickly assess whether grain boundaries have high or low mobilities relative to their local crystallographic neighborhood.

With the aim of integrating a geodesic metric into grain boundary visualization and property regression, our work is divided into two main topics:

1. *GB visualization*: how can we visualize the five dimensional grain boundary geometry in three dimensions?
2. *GB regression*: how can grain boundary metrics be incorporated into regression models for grain boundary properties?

We begin with a methods section that walks through construction of the pairwise grain boundary octonion distance matrix for grain boundary datasets, the starting point for our visualization and regression routines. We then pursue the geometric problem of GB visualization via dimensionality reduction. The global connectivity of a grain boundary dataset is approximated in  $\mathbb{R}^3$  via

classical multidimensional scaling. In the results section, we show that the 5D connectivity of the  
 70 388 grain boundary Olmsted dataset can be visualized in 5D in a physically meaningful and inter-  
 pretable manner. After selectively adding new grain boundaries to the dataset, these visualizations  
 suggest new global features of grain boundary space, including the existence of a grain boundary  
 fundamental zone with well defined subsets of high symmetry boundaries along faces and general  
 boundaries in the interior. We show that arbitrary grain boundaries within a dataset may be con-  
 75 nected via geodesic sampling and that these paths are well behaved under dimensionality reduction.  
 Finally, we show that an octonion-based kernel regression method with one fitting parameter can  
 predict grain boundary energy in the Olmsted dataset to within 10% RMSE.

## 2. Methods

In this section we describe an easily implemented method to visualize the global connectivity  
 80 of grain boundary datasets in 5D given a list of macroscopic crystallographic parameters. An  
 implementation of the method for the Olmsted grain boundary data may be found at the companion  
 Github page for this paper: [octonion code](#). We start with background on the theory of the GBO  
 representation. Next, we describe construction of the pairwise geodesic distance matrix, the starting  
 point for dimensionality reduction and regression. Several dimensionality reduction techniques are  
 85 reviewed with an emphasis on classical multidimensional scaling.

### 2.1. The grain boundary octonion representation

Representation of grain boundary data is not arbitrary. Just as different representations for  
 orientations have different advantages (for instance, quaternions versus Euler angles [14]), different  
 representations for grain boundaries impact the feasibility and computational efficiency of defining  
 90 paths in GB-space [15]. The traditional definition of grain boundaries as the product space of  
 a boundary plane unit normal vector and a misorientation  $\mathbb{S}^2 \times SO(3)$  mixes two very different  
 quantities, a rotation and a vector, which makes the definition of a geodesic metric non-intuitive.  
 In order to alleviate these issues, Olmsted proposed a representation for grain boundaries that  
 mixes two rotations with respect to a reference frame attached to the grain boundary plane [16].  
 95 This framework naturally suggests a geodesic metric: the Riemannian metric on  $SO(3) \times SO(3)$ .  
 However, the specification of a grain boundary with two rotations has an intrinsic symmetry — the  
 rotational freedom of the reference frame in the plane (which we refer to as the “ $U(1)$  symmetry”)  
 — with respect to which the Riemannian metric cannot be analytically minimized. The octonion

metric correctly determines the angular distances between grain boundaries with a common normal or misorientation (cases for which the angular distance is known) and closely approximates the geodesic metric on  $SO(3) \times SO(3)$  for all grain boundary pairs while maintaining the ability to be analytically minimized with respect to the  $U(1)$  symmetry. This leads to the ability to compute symmetrized geodesic paths in grain boundary configuration space, avoiding the need to only consider grain boundary paths along high symmetry subspaces.

We represent each grain boundary in a given dataset as a unit grain boundary octonion (GBO) in  $\mathbb{S}^7$ , the 7-sphere embedded in  $\mathbb{R}^8$ . Each GBO is generated by concatenating two rotation quaternions expressed in the boundary plane reference frame:  $o = (q_A, q_B)$ . Since each quaternion itself has unit norm, this octonion has norm  $\sqrt{2}$ , so we introduce a normalization factor:

$$o = \frac{1}{\sqrt{2}}(q_A, q_B) \rightarrow \|o\|^2 = \frac{1}{2} (\|q_A\|^2 + \|q_B\|^2) = 1. \quad (1)$$

The representation now has six degrees of freedom. It is important to follow a fixed set of conventions in defining grain orientations relative to the boundary plane reference frame. A routine converting the traditional grain boundary descriptor in terms of the boundary normal  $\hat{\mathbf{n}}$  and the misorientation  $p_m$  to the GBO representation can be found in the Supplementary Material of [9]. There are four different cartesian reference frames involved: the sample frame, the two rotated grain frames, and the grain boundary frame, for which the selection of the axes *in* the boundary plane constitutes the extra degree of freedom associated with  $U(1)$  symmetry necessary to recover a 5D space. By convention, the grain boundary plane normal direction lies along the  $z$  axis. All input grain boundaries must be rotated such that this convention is satisfied. Example code transforming the Olmsted boundaries to the octonion representation is given in the octonion code repo.

By definition, the geodesic arc length  $\Omega$  on  $\mathbb{S}^n$  is a metric on the unit  $n$ -sphere embedded in  $\mathbb{R}^{n+1}$ . The arc length between two GBOs can thus be computed as:

$$\Omega = 2 \arccos |o_1 \cdot o_2|. \quad (2)$$

We write the geodesic arc length as:

$$\Omega = \Omega(o_1, o_2) = \Omega((q_A, q_B), (q_C, q_D)), \quad (3)$$

The geodesic arc length  $\Omega$  naturally satisfies the conditions to be a metric on  $\mathbb{S}^7$  and can be taken as a “distance” between the two GBOs. For normalized octonions, we have:

$$\Omega_{AB,CD} = 2 \arccos \frac{1}{2} \left| \cos \frac{\omega_{AC}}{2} + \cos \frac{\omega_{BD}}{2} \right| = 2 \arccos \frac{1}{2} |q_A \cdot q_C + q_B \cdot q_D|. \quad (4)$$

At this point, the distance metric is un-symmetrized: symmetrically equivalent grain boundary  
 125 pairs may lead to different values of  $\Omega$ . The GBOM angle  $\Omega$  must be minimized over four types  
 of grain boundary symmetries: grain exchange symmetry,  $U(1)$  symmetry, the fact that the space  
 of quaternions is a double cover of  $SO(3)$ , and crystal symmetry. Full details of this process are  
 described in [9]. Since cubic crystal symmetry is relevant to computation of pairwise distances in  
 the Olmsted dataset, we describe its implementation below.

130 Consider a material with rotational point group  $\mathcal{G}$  of order  $N_p$ ; the symmetry elements of  
 this group can be expressed as (passive) rotation quaternions  $S_i$  with  $i \in [1, N_p]$ . The set of  
 symmetrically equivalent orientations for a given quaternion  $q$  can be computed as  $\{S_i q, i \in [1, N_p]\}$ .  
 The true GBO pair misorientation angle  $M$  is given by:

$$M(o_1, o_2) = M((q_A, q_B), (q_C, q_D)) = \min \{\Omega_{ijkl}, (i, j, k, l) \in [1, N_p]\}. \quad (5)$$

where

$$\Omega_{ijkl} = \Omega((S_i q_A, S_j q_B), (S_k q_C, S_l q_D)) \quad (6)$$

135 The cardinality of this set is equal to  $N_p^4$  (assuming that grain exchange symmetry is already  
 taken care of in the computation of  $\Omega$  [9]). The minimum value in this set  $\Omega^* = M(o_1, o_2)$  produces  
 the smallest geodesic arc length between symmetrically equivalent GBO pairs. For the cubic FCC  
 Ni grain boundaries in the Olmsted dataset,  $\mathcal{G} = 432$  and  $N = 24$ .

Parallellized octonion distance calculations with arbitrary crystallographic symmetry have been  
 140 implemented in the open source software package EMsoft [17].

## 2.2. Pairwise distance matrix construction

Consider a grain boundary dataset consisting of  $L$  grain boundaries represented by unsym-  
 metrized octonions; for the Olmsted dataset,  $L = 388$ . A pairwise distance matrix  $\mathcal{P}_{ij}$  is con-  
 structed from the minimum GBOM angles  $\Omega_{ij}^* = \Omega^*(o_i, o_j)$ , where  $i, j \in [1, L]$ . The pairwise  
 145 distance matrix is positive and symmetric (positive definite) with zeros along the diagonal. Matrix  
 construction is  $O(N_p^4 L^2)$  and is therefore computationally demanding. Pairwise distance matrix  
 construction is implemented in the EMsoft program *EMGBOdM*. Construction of a  $388 \times 388$   
 pairwise distance matrix for the Olmsted dataset takes 26 minutes on 8 cores.

For large datasets with thousands of grain boundaries, it may be possible to define a "scaffold-  
 150 ing" of GB space with a relatively small dataset from which other boundaries are locally computed.

Analytical projections of the grain boundary manifold to low dimensional spaces are also worth pursuing [9].

### 2.3. Dimensionality reduction

Given a pairwise distance matrix  $\mathcal{P}_{ij}$ , we seek to find the embedding of grain boundary coordinates  $x_1, x_2, \dots, x_N \in \mathbb{R}^d$  in euclidean space which minimizes a suitably defined reconstruction error term. For visualization, we choose  $d = 3$ . We consider two related dimensionality reduction techniques: multidimensional scaling (MDS) and kernel principal component analysis (kPCA) [18, 19].

MDS refers to a family of techniques for embedding pairwise distance data in low dimensions. Classical MDS (cMDS) expresses reconstruction error minimization  $|x_i - x_j| \approx p_{ij} \in \mathcal{P}_{ij}$  as an eigenvalue problem.

$$\mathcal{B} = -(H\mathcal{P})(H\mathcal{P})^T = V^T \Lambda V \quad (7)$$

The distance matrix  $\mathcal{P} = \mathcal{P}_{ij}$  is centered about its mean with a centering matrix  $H$ . The spectral decomposition of the squared distance matrix  $(H\mathcal{P})(H\mathcal{P})^T$  gives a diagonal matrix of eigenvalues  $\Lambda$  and corresponding eigenvectors  $V$ . Eigenvalues are sorted from largest to smallest  $\lambda_1 \geq \lambda_2 \dots \geq \lambda_L$ . The squared distance matrix can be reconstructed in  $d$  dimensions as  $\mathcal{B}_d = V_d^T \Lambda_d V_d$  where  $\Lambda_d = \text{diag}(\lambda_1 \dots \lambda_d)$  and  $V_d$  is an  $L \times d$  matrix of eigenvectors of  $\mathcal{B}_d$ .

Reduced grain boundary coordinate data  $\mathcal{X}_d$  can be recovered as  $\mathcal{B}_d = \mathcal{X}_d \mathcal{X}_d^T$ , resulting in reduced grain boundary coordinates  $\mathcal{X}_d = V_d \lambda_d^{1/2}$ . The amount of variance explained by the  $d$ -th dimension is simply the ratio of the  $d$ -th eigenvalue with the sum of all positive eigenvalues.

If an input distance matrix contains euclidean distances, the original data can be reconstructed perfectly using an eigenbasis with the intrinsic dimensionality of the space. Recall that grain boundary distance data represented by symmetrized octonions lies on the manifold  $SO(3) \times SO(3)/U(1)$  with cubic symmetry. Since this space is highly nonlinear, not all eigenvalues in eqn 7 will be positive, and it will be impossible to perfectly reconstruct the data with a finite number of dimensions using the standard Euclidean distance metric [20, 21]. Nevertheless, three dimensions are sufficient to capture 45 percent of the variance in the pairwise Olmsted distance data, as shown in Fig. 1a. We will show that even 45 percent variance comes with a surprising amount of interpretability with respect to global grain boundary geometry. Recent analysis of cMDS applied to pairwise distance

180 data on the  $n$ -sphere  $\mathbb{S}^n$  suggests that in the limit of uniform sampling of distances on  $\mathbb{S}^n$ , cMDS eigenfunctions are related to spherical harmonic functions [20, 21].

Out of a variety of dimensionality reduction methods tested on the Olmsted data, we have chosen to focus on cMDS because of its analytical form and interpretable results. In the following, we compare cMDS to other dimensionality reduction methods, most notably kernel PCA (kPCA). It is worth emphasizing that our method starts with precomputed geodesic distances – dimensionality reduction techniques such as Isomap first estimate geodesic distances from a manifold (e.g. an S-shape) embedded in  $\mathbb{R}^n$  [22] and then apply multidimensional scaling to the geodesic distance matrix. cMDS is a closed form optimization problem. Other variants of MDS, such as metric MDS, employ a stochastic minimization procedure that is prone to getting stuck in local minima and gives different results on different runs. We highlight the following observations with respect to the pairwise geodesic distance matrix for the Olmsted data:

1. Dimensionality reduction results for kernel PCA (kPCA) are the same as those for cMDS for a proper choice of kernel function and tuning parameter  $\lambda$ . Both techniques are known to capture the global geometry of datasets in low dimensions.
- 195 2. cMDS and kPCA give an analytical and repeatable standard to which other (more complex) dimensionality reduction techniques can be compared.

Kernel PCA and kernel regression apply a nonlinear “feature” transformation to the pairwise distance matrix. In this work, we use the Laplace kernel  $\mathcal{K}_{ij} = \exp(-\lambda \mathcal{P}_{ij})$ . The Laplace kernel has the benefit of being positive definite for distances defined on  $n$ -spheres, unlike other common choices of kernel [23]. The tuning parameter  $\lambda$  may be fit to grain boundary properties via kernel regression, as discussed in Section 3.

### 3. Results and Discussion

#### 3.1. Visualizing the global geometry of grain boundary datasets

In this section we demonstrate that the 5D connectivity of the 388 grain boundary Olmsted dataset can be visualized via dimensionality reduction in 5D with physically meaningful and interpretable results. Because each grain boundary is represented as a single point in 3D, this allows for direct comparison of grain boundaries with different misorientation axes, angles and boundary plane inclination, unlike traditional visualizations of grain boundary properties. New grain boundaries can easily be added, including perfect crystals, random general grain boundaries, or high symmetry

boundaries. Visualization of an augmented dataset of nearly 600 grain boundaries suggests several interesting global features of grain boundary space, including the existence of a grain boundary fundamental zone with well defined subsets of high symmetry boundaries along faces.

### 3.1.1. Visualizing and augmenting the Olmsted dataset

Dimensionality reduction is performed on the  $388 \times 388$  pairwise cubic GBO distance matrix via cMDS, resulting in a  $3 \times 388$  matrix of reduced coordinates for the Olmsted dataset. These coordinates capture 45% of the variance in the data, as shown in Fig 1a. This amount of variance is sufficient to capture a large amount of structure in benchmark datasets such as the MNIST digits dataset [24]. Although the intrinsic dimensionality of GB space is five, fifteen dimensions (15 coordinates per GB) are necessary to explain even 80% of variance in the data. This is consistent with the fact that non-euclidean pairwise distances on the n-sphere cannot be perfectly reconstructed by euclidean distances in  $\mathbb{R}^n$  with finite  $n$  [21]. Nevertheless, contrary to intuition, we find that a 3D euclidean dimensionality reduction recovers and extends common knowledge of 5D cubic grain boundary geometry.

Dimensionality reduction results are given for a high symmetry subset of boundaries in the Olmsted dataset in Fig. 1. Each point represents a grain boundary. We will argue that the connectivity of grain boundary classes in the reduced space unifies common knowledge in the grain boundary literature. Each highlighted trajectory is a well known grain boundary geodesic. The red trajectory, for instance, spans the  $\langle 100 \rangle$  symmetric tilt (ST) boundaries. When energy is plotted as a function of geodesic angle along this path (Fig. 1c), the canonical energy profile for  $\langle 100 \rangle$  STs is recovered. In the literature, a total tilt angle is often given which is two times the geodesic angle, spanning 0 to 90 degrees. The factor of two is a result of the space of quaternions (isomorphic to  $SU(3)$ ) being a double cover of the space of rotations  $SO(3)$  [25].

The three reduced dimensions may loosely be interpreted as tilt and twist angles. For instance, geodesics for three classes of twist boundaries ( $\{100\}$ ,  $\{110\}$  and  $\{111\}$  in orange, pink and green) vary primarily along the second dimension. These geodesics run approximately perpendicular to four classes of symmetric tilt boundaries ( $\langle 100 \rangle$ ,  $\langle 110 \rangle$ ,  $\langle 111 \rangle$   $\{110\}$  ST and  $\langle 111 \rangle$   $\{112\}$  ST in red, light blue, dark blue and brown) that vary primarily along the first and third dimensions. The frontmost set of symmetric tilt boundaries (red, light blue, dark blue) comprise a curved triangular surface that maps onto the standard stereographic triangle (SST) for symmetric tilt boundaries with arbitrary boundary plane inclinations (plotted by Wolf in [26]). For high symmetry paths in

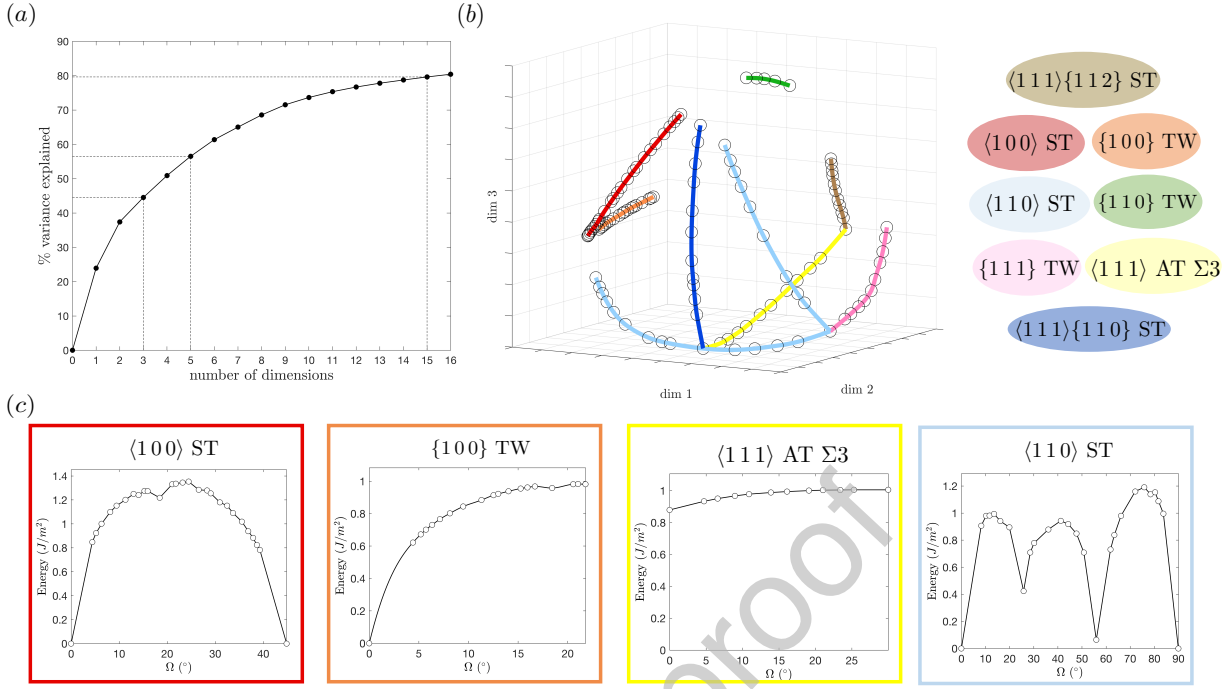


Figure 1: (a) Cumulative variance in the pairwise distance matrix explained with the number of dimensions of the ordered eigenbasis constructed by cMDS (b) High symmetry geodesics are highlighted in color in the reduced space. These three dimensions explain the most variance in the pairwise distance data (c) GB energy along four selected geodesic paths. Points with zero energy correspond to perfect crystals along the geodesic paths.

the Olmsted dataset, the reduced space is therefore interpretable and unifies common knowledge. The first and third reduced dimensions are related to tilt angles, while the second dimension is related to a twist angle.

More general paths through grain boundary space complicate the interpretation of the reduced axes. The  $\Sigma 3 \langle 111 \rangle$  asymmetric tilt (AT) boundaries have both tilt and twist character (connecting two incoherent twin boundaries with different boundary plane inclinations), but correspond to a geodesic in the reduced space that varies primarily along the second dimension, interpreted as a twist angle. The reduced space contains information about *all five* degrees of freedom, and it is therefore unsurprising that the three dimensions do not have precise interpretations. In the next section, grain boundary energy will be analyzed on a boundary plane fundamental zone embedded in the reduced space.

The  $\langle 110 \rangle$  ST boundary geodesic (light blue) has several special features that deserve further commentary. The geodesic changes direction abruptly at the coherent twin boundary (CT,  $\Sigma 3(111)$ , explicitly labeled in Fig. 2a). The CT boundary is the deepest minimum in the GBE



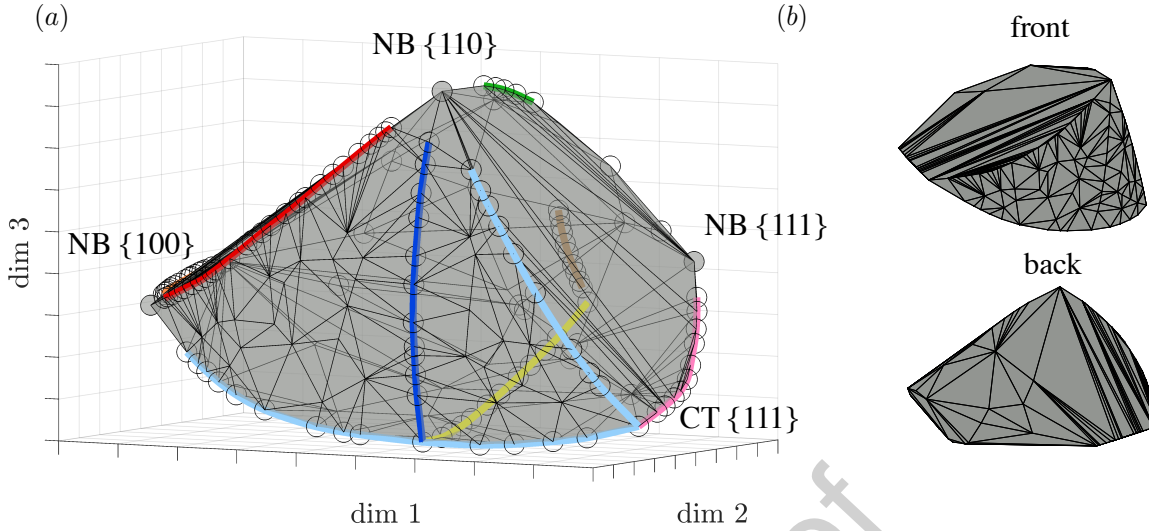


Figure 2: Convex hull plotted for the reduced space coordinates of the Olmsted dataset. (a) The Olmsted data set has been augmented with three perfect crystals (trivial no boundary (NB) singularities) which appear at vertices of the convex hull and correspond to different high symmetry orientations. We hypothesize the existence of a grain boundary fundamental zone in the 5D space with high symmetry boundaries on the surface and general boundaries contained within the volume. More thorough sampling of the 5D space is needed to test this hypothesis. b) different perspectives of the convex hull defined by the Olmsted dataset.

255 landscape for FCC metals other than perfect crystals [3, 27], as can be seen in the GB geodesic energy plot for  $\langle 110 \rangle$  ST boundaries (at  $54.8^\circ$ ). The boundary was classified by Bulatov as a 5-GROF, or true minimum in energy with respect to variation in all 5 degrees of freedom [12]. The  $\langle 110 \rangle$  ST geodesic also passes through the incoherent twin boundary (ICT,  $\Sigma 3(112)$ , intersection of light blue, dark blue and yellow trajectories). The incoherent twin boundary is a local minimum in GB energy along the  $\langle 111 \rangle \{110\}$  ST and  $\langle 111 \rangle$  AT  $\Sigma 3$  geodesics and we believe it is also a weak local minimum along the  $\langle 110 \rangle$  ST geodesic which is not resolved by the angular sampling in the Olmsted survey ( $35.2^\circ$  on  $\langle 110 \rangle$  ST GB energy plot). The local minima is resolved in other GB energy calculations for FCC metals [27]. These results suggest that the incoherent twin boundary is at least a 3-GROF and could be a true minimum in the GB energy landscape. Similar analysis  
 260 can be performed for other grain boundaries at apparent cusps, such as the  $\Sigma 5(310)$  ( $25.2^\circ$  along  $\langle 110 \rangle$  ST) and  $\Sigma 11(113)$  ( $18.4^\circ$  along  $\langle 100 \rangle$  ST) boundaries.

The change in direction of the  $\langle 110 \rangle$  ST geodesic is reminiscent of an orientation trajectory leaving and re-entering a fundamental zone. Remarkably, considering that this is a 3D representation of a 5D space, it is found that a high symmetry subset of the Olmsted boundaries, including

all symmetric tilt boundaries, reside on the surface of the convex hull of the dataset, while general boundaries lie inside the delimited volume. 100 random grain boundaries obtained via cubochoric sampling of orientation space [28] were verified to lie inside the convex hull defined by the Olmsted dataset when added to the pairwise distance matrix. We hypothesize the existence of a fundamental zone in the 5D space that depends on point group symmetry. Analytical specification of the grain boundary fundamental zone via the octonion representation is a subject of future work.

Perfect crystals with no grain boundary (no boundary = NB) may be included in the reduced configuration space of grain boundaries by augmenting the Olmsted dataset and pairwise distance matrix. The appearance of multiple distinct NB points in the projection is a consequence of different grain boundary geodesics with decreasing disorientation angle arriving to perfect crystals with different boundary plane inclinations [29]. For example, in the reduced space, the red, orange, and light blue trajectories all meet at a  $\Sigma 1\{100\}$  boundary (denoted as NB  $\{100\}$  in Fig. 2a). This boundary is at the origin of the GB energy plots for  $\langle 100 \rangle$ ,  $\langle 110 \rangle$  ST and  $\langle 100 \rangle$  TW boundaries. Four trajectories meet at the  $\Sigma 1\{110\}$  boundary (labeled NB  $\{110\}$  in Fig. 2a). As noted in [9], including perfect crystals in the configuration space of GB octonions is necessary to define a manifold and interpolate between arbitrary boundaries. Thus, 200 perfect crystals are added to the Olmsted dataset including 100 low index orientations and 100 random orientations. A thin volumetric slice is formed by the perfect crystals, forming the closure of a well defined set in the reduced space (Fig. 3a). There is an apparent one-to-one correspondence between the the set of NB points on the surface of the convex hull (the NB facet) and the SST of possible boundary plane inclinations, though it is unclear exactly how the NB points in the interior project onto this surface. Disorientation angle increases away from the NB facet to a maximum disorientation of 62 degrees (Fig. 3b).

It is important to note that the ability to explain the full variance in the data is lost during dimensionality reduction, therefore we cannot fully reconstruct the 5D character of an arbitrary grain boundary using the location of the projected data point in the 5D space. Nevertheless, as discussed in the supplementary information, each grain boundary is distinct in the 3D visualization, and the distance between grain boundary points in the 3D Euclidean reconstruction is meaningful with respect to their geodesic distance (see (Fig. S1a)). There is local variation in how well the projected data explains the variance of the high dimensional data; the results are most reliable for pairs of boundaries with small and large, but not intermediate, pairwise distances. This explains the interpretability of closely spaced high symmetry boundaries along the surfaces of the delimiting

volume (Fig. 3a) and the smooth variation of misorientation angle moving out from the NB facet (Fig. 3b). In contrast, points interior to the convex hull are less interpretable. For example, iso-disorientation surfaces may be twisted, as shown by the  $\Sigma 3$  boundaries (all with a disorientation angle of 60 degrees) in Fig. S3b. It is possible that analysis of subsets of related boundaries would reveal structures that are lost in the global dimensionality reduction and visualization process.

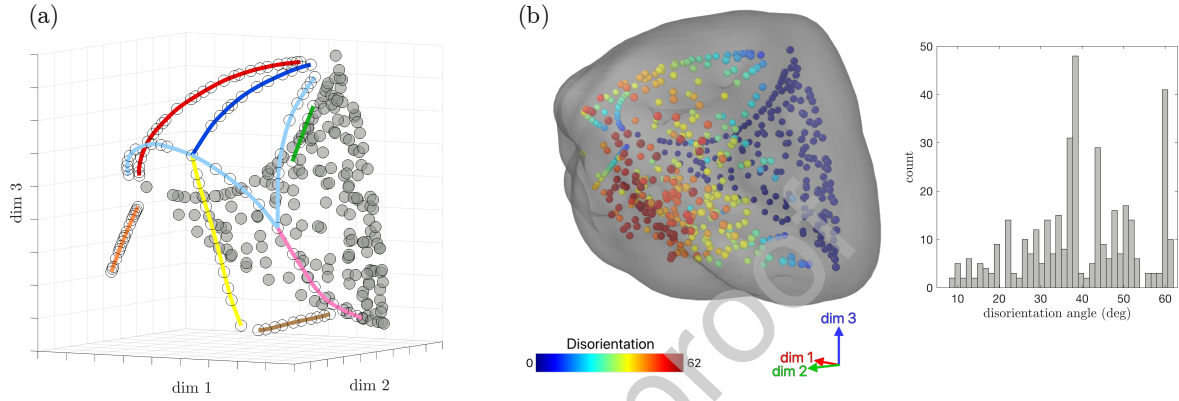


Figure 3: (a) 200 perfect crystals (grey) form the closure of a well-defined set in the reduced space – a slice with a triangular face along the convex hull (the NB facet) and a thin width along the second dimension. The NB facet is isomorphic to the SST of boundary plane inclinations. (b) Disorientation angle increases away from the NB facet along the second dimension. Points are colored by disorientation angle and rendered with a transparent mesh that expands a gaussian density around each point. The Mackenzie distribution for the Olmsted dataset (excluding perfect crystals) is shown on the right. Videos of rotating configurations for 2a and 2b are given in the supplementary information

### 3.1.2. Anisotropy of GB energy and mobility

The global grain boundary visualization method introduced in this work is a quick tool for surveying grain boundary properties across an entire dataset. On a more fundamental level, the geometry of grain boundary space itself may be connected to grain boundary properties, and visualization is a first step toward exploring such connections. To further examine the properties of special and general boundaries, we color points in the reduced representation by property as in Fig. 4. We examine grain boundary energy, mobility, and energy barriers to grain boundary motion in the reduced space. Methods for computing mobility and energy barriers are described in [4] and [30]. Mobility is computed at a fixed driving force for grain boundary motion (25 eV/atom) at 1400 K, while energy barriers are computed proportionally to the critical driving force required to initiate grain boundary motion at 100 K [30]. Capturing variation in the properties of 388 points is

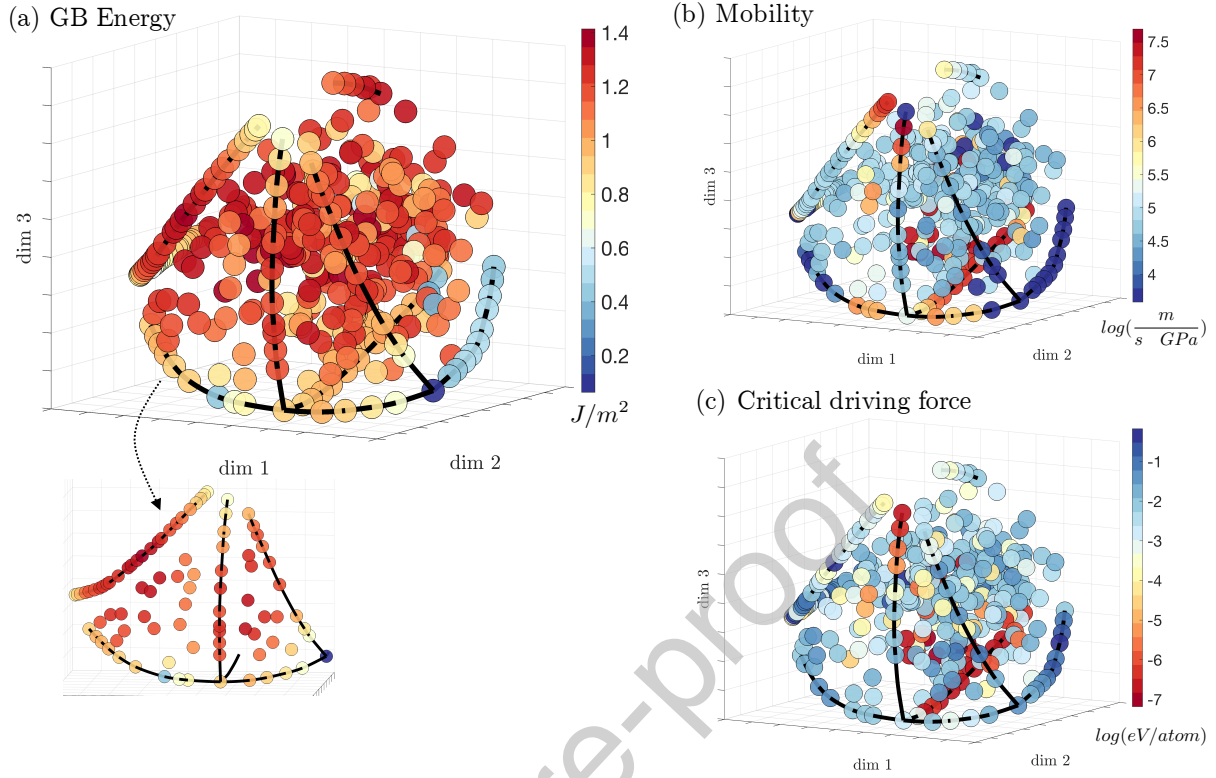


Figure 4: (a) GB energy for Ni in the reduced space. The inset shows symmetric tilt boundaries with varying plane inclination. (b) GB mobility data for Ni at 1400 K in the reduced space. Immobile boundaries are colored dark blue. (c) Critical driving force for grain boundary motion in Ni at 10 K generally varies inversely with grain boundary mobility (the colorbars in b and c are inverted). Videos of rotating configurations for 3a and 3b are given in the supplementary information

a visualization challenge – videos of rotating GB configurations are included in the Supplementary Information.

Many grain boundary energy trends highlighted in the original Olmsted survey can be seen compactly in the reduced space. The 388 boundary set is divided into three subspaces which are more clearly visualized: symmetric tilt boundaries, low  $\Sigma$  boundaries, and general boundaries. The curved triangular face of  $\langle 100 \rangle$ ,  $\langle 110 \rangle$ , and  $\langle 111 \rangle \{110\}$  symmetric tilt boundaries (shown at the bottom of Fig. 4a) depicts variation of grain boundary energy for general ST boundary plane inclinations. Grain boundary energy on the surface of general ST grain boundaries appears to be smooth apart from cusps along the  $\langle 100 \rangle$  and  $\langle 110 \rangle$  ST energy trajectories. Next, we consider low  $\Sigma$  boundaries. Fig. S2b depicts grain boundary energy for  $\Sigma 3$ ,  $\Sigma 5$ ,  $\Sigma 7$ ,  $\Sigma 9$  and  $\Sigma 11$  boundaries. Visually, some clustering of color is apparent corresponding to regions of low  $\Sigma$  boundaries with

similar energy. The distribution of  $\Sigma$  in the reduced space is not intuitive, with no clear pattern  
 330 separating boundaries of different  $\Sigma$ . We examine the embedding of the  $\Sigma 3$  BPFZ (Fig. S3a) in the  
 reduced space (Fig. S3b). Although the perimeter of the stereographic triangle in the  $\Sigma 3$  BPFZ is  
 warped in the reduced space, the space still captures apparently smooth variation in GBE, albeit  
 in a less interpretable manner. General boundaries also show relatively smooth variation of grain  
 boundary energy. Grain boundaries are classified as general in this work if they are not part of  
 335 high symmetry geodesics, STGBs, or low  $\Sigma$  boundaries. Clustering is apparent by grain boundary  
 energy (Fig. S2c) with distinct regions of low and high energy boundaries. Overall, the reduced  
 space enables us to explore visually the local and global variation in grain boundary energy in the  
 Olmsted dataset.

We can also gain insight into grain boundary mobility via colored reduced space representations  
 340 (Fig. 4b and c). It is well known from the Olmsted survey that many  $\Sigma 3$  boundaries have high mo-  
 bility. In contrast,  $\langle 111 \rangle$  TW boundaries including the coherent twin (CT) boundary are observed  
 to be immobile in molecular dynamics simulations at the tested driving force and temperature.  
 Both of these trends are shown in the reduced space along the second dimension at the bottom of  
 the figure. For these paths, the critical driving force at low temperature (10 K) can be meaningfully  
 345 linked to mobility [31]. While  $\Sigma 3$  boundaries have high mobility, low energy barriers to motion  
 and low critical driving force,  $\langle 111 \rangle$  TW boundaries have large energy barriers to motion and  
 are immobile for low driving forces. A general inverse relationship between mobility and critical  
 driving force is observed throughout the space. For  $\langle 110 \rangle$  ST boundaries, the relationship between  
 mobility and critical driving force appears to be more complex, likely due to transitions at finite T  
 350 involving multiple motion mechanisms which are not captured by a single energy barrier at 10 K.  
 The reduced representation is a tool for quickly identifying regions of grain boundary space which  
 are most interesting for further study of a given property.

### 3.2. Interpolation and regression

#### 3.2.1. Grain boundary interpolation via OSLERP

355 Grain boundary datasets can be extended to new regions of the 5D space in a directed manner.  
 The main tool used in this section is the geodesic sampling scheme introduced in [9] to smoothly  
 interpolate between arbitrary grain boundaries. First, we pick a pair of boundaries, for instance,  
 the frequently simulated  $\Sigma 5$  (210) boundary and the coherent twin boundary  $\Sigma 3$  (111) (marked  
 A and B on Fig. 5a). We then interpolate between the GBOs  $o_A$  and  $o_B$  along a geodesic path on

360  $\mathbb{S}^7$ :

$$r(t) = \text{oSLERP}(o_A, o_B; t) = \frac{\sin[(1-t)\theta]}{\sin\theta} o_A + \frac{\sin[t\theta]}{\sin\theta} o_B \quad (t \in [0, 1]; \theta = \frac{\Omega}{2}). \quad (8)$$

where oSLERP denotes octonions spherical linear interpolation. In this work, 10 equally spaced points in the interpolating variable  $t$  are used to produce 8 new octonions along the interpolated trajectory  $r(t)$  that are not included in the original dataset. As discussed in [9], these geodesics are a close approximation of those on  $SO(3) \times SO(3)/U(1)$  with a small and bounded error term. After  
 365 the new interpolated octonions are added to the dataset, the pairwise distance matrix is expanded to include this data. Dimensionality reduction is performed again.

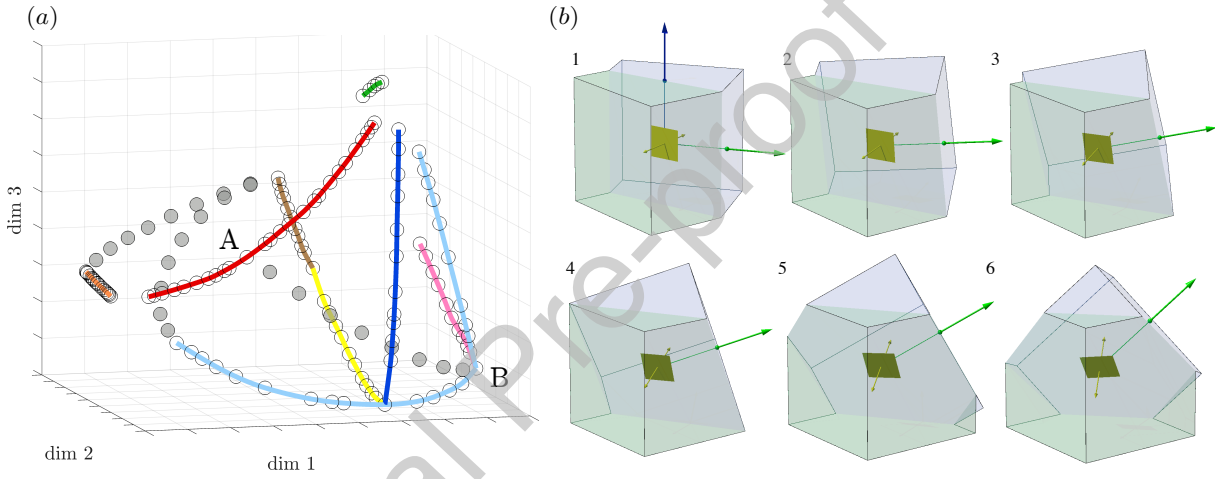


Figure 5: (a) Reduced space augmented with three geodesic trajectories. The labeled geodesic connect the  $\Sigma 5$  (210) boundary (A) and CT boundary (B). (b) Bicrystal configurations along the A-B geodesic (numbering is only meant to guide eye). Videos of rotating configurations for 5a are given in the supplementary information

Three geodesic paths are added to the reduced space in Fig. 5. With addition of geodesic paths, the overall space does not appear to change. These trajectories are well behaved in the sense that they smoothly transition between pairs of grain boundaries and remain within the convex  
 370 hull defined by the Olmsted dataset. Neighboring points on OSERP trajectories are separated by equal amounts in the reduced space, just as they are on the grain boundary manifold. The transition between the  $\Sigma 5$  (210) boundary (A) and CT boundary (B) is shown explicitly via bicrystals in Fig. 5b. The blue grain rotates along an axis in the plane of the grain boundary (and perpendicular to the tilt axis) and also gradually twists into the orientation of the CT boundary.  
 375 These visualizations were created with the *EMOSLERP* tool implemented in the EMsoft package [17]. Two other geodesics are shown that connect boundaries in high symmetry geodesics in the

Olmsted dataset. They explore regions of grain boundary space that are not well sampled by the Olmsted dataset. These visualizations verify that geodesic sampling is an effective tool for extending grain boundary datasets to new regions of the 5D space.

### 3.2.2. Kernel regression results

The Olmsted dataset provides a scaffolding of GB space to which properties can be fit [12]. The octonion metric and OSLERP allow us to move along arbitrary paths within the dataset, thus providing a valuable tool for grain boundary regression. Kernel regression has been performed on the Olmsted grain boundary energy data using a scaled pairwise distance matrix. A positive definite Laplace kernel exponentially scales each pairwise distance by a fixed parameter  $\lambda$ . The grain boundary energy estimate  $\tilde{\epsilon}_i$  is a weighted average of the energy of nearby grain boundaries in the training data.

$$\tilde{\epsilon}_i = \frac{\sum_{j \in \text{train}} e^{-\lambda \Omega_{ij}} \epsilon_j}{\sum_{j \in \text{train}} e^{-\lambda \Omega_{ij}}} \quad (9)$$

The optimal value of  $\lambda$  which minimizes the root mean squared error (RMSE) energy is selected via k-fold cross validation for  $k = 10$ .

$$RMSE(k) = \frac{1}{k} \sum_k \frac{1}{N_k} \sum_{i \in \text{test}(k)} \sqrt{(\tilde{\epsilon}_i - \epsilon_i)^2} \quad (10)$$

A value  $\lambda^* = 36$  gives an optimal RMSE energy of  $0.0977 \text{ J/m}^2$ , 9.8% error compared to an average grain boundary energy of  $1 \text{ J/m}^2$  (with  $\Omega_{ij}$  measured in radians). Per GB errors are distributed around a mean of zero with a standard deviation of  $0.099 \text{ J/m}^2$ . The tails of the per-error distribution contain large negative and positive errors with a maximum absolute error of  $0.5 \text{ J/m}^2$  for the coherent twin boundary (plotted at  $54.5^\circ$  in Fig. 6d and corresponding to a disorientation of  $60^\circ$  in Fig. 6b). The model only sometimes resolves cusp locations. For  $\langle 110 \rangle$  ST GBs, the two deepest cusps are resolved, albeit with large errors associated with incorrect cusp depths for the corresponding  $\Sigma 11$  and  $\Sigma 3$  boundaries. Shallower cusps, including the  $\Sigma 5 \langle 100 \rangle$  ST GB at 18 degrees, are not resolved, though an anomaly in the derivative of energy with respect to tilt angle is apparent. The model for grain boundary kernel regression is not physics aware – it is a local fitting function with one free parameter that is solely based on macroscopic grain boundary geometry and is not constrained to follow the Read-Shockley model for boundaries near perfect crystals or other cusps. This is in contrast to the BRK model, which includes an explicitly fit

grain boundary energy surface for all grain boundaries with  $\langle 100 \rangle$ ,  $\langle 110 \rangle$  and  $\langle 111 \rangle$  disorientation axis (including 3 perfect crystals) in the training data [12]. The BRK model achieves an RMSE of 5% on more general grain boundaries with an overhead of 41 geometry related fitting parameters for the initial fit, although only a single material-dependent parameter is required to calibrate it among isostructural materials (i.e. FCC metals) [12]. A SOAP descriptor based regression model for grain boundary energy is also able to achieve an RMSE of 5% (95 % accuracy) for the FCC Ni dataset [32, 33]. Though this model only uses 50% of the dataset at random for training, it has the drawback of requiring the fully relaxed boundary structure as input for SOAP vector construction (if the relaxed structure is known, the energy is known for a given interatomic potential). Our kernel regression model is able to achieve 10% RMSE only using macroscopic grain boundary geometry and a random set of grain boundary energies pre-calculated from atomistic simulations.

Systematic errors in energy estimation suggest room for improvement in our kernel regression model. For instance, low energies are systematically overestimated and high energies are systematically underestimated (Fig. 6a). As specific examples, the energy of low angle grain boundaries (disorientation less than  $15^\circ$ ) and  $\Sigma 3$  boundaries (disorientation of  $60^\circ$ ) are overestimated (Fig. 6b). No perfect crystals are included in these results. It is found that when the nearest perfect crystal to a given grain boundary is included in the energy estimate for that grain boundary, the global RMSE for the dataset increases to near 11%. If the perfect crystals are only included in the weighted average for low angle boundaries with disorientation less than  $15^\circ$ , RMSE improves to only 9.7% error. With the current functional form of the regression model, only small improvements appear to be possible by including the NB singularity. Given the relatively large  $\lambda^*$  value of 36, it is worth examining how many grain boundaries in the training data contribute to the estimated energy of a grain boundary in the test data. It is not the case that our model simply outputs the energy of the nearest neighbor in the training data – the average maximum energy contributed to a grain boundary by surrounding grain boundaries is  $0.15 \text{ J/m}^2$ . Most grain boundaries have meaningful contributions to their energy from more than 8 distinct grain boundaries. Overall, the kernel regression method achieves a reasonable RMSE energy by taking a weighted average over pairwise geodesic distances and a known subset of grain boundary energies. Further improvements to the model may be possible by more systematically accounting for the global geometry of the grain boundary fundamental zone (for example, by weighting contributions from the surface of the convex hull more than the interior).



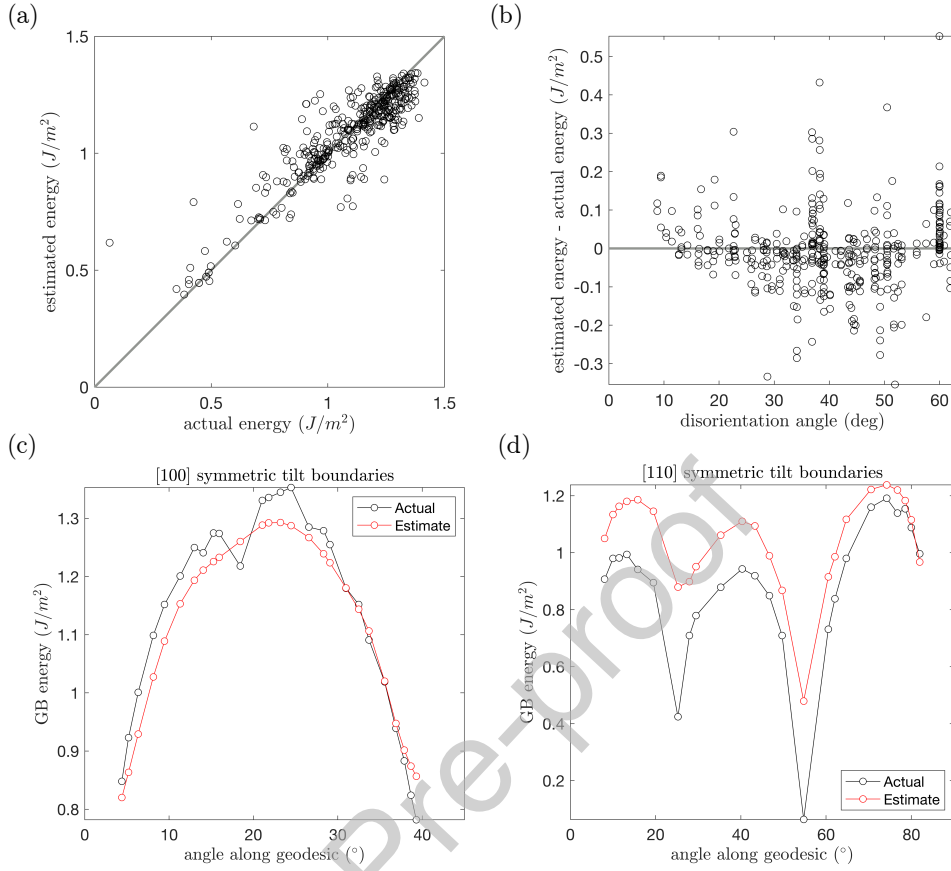


Figure 6: (a) Estimated energy over validation sets versus actual energy from molecular statics calculations ( $R^2 = 0.81$ ). (b) distribution of errors with disorientation angle. (c) and (d) show the energy estimates along geodesics (after leaving these boundaries out of the training and test data). The energy of [110] ST boundaries is systematically overestimated, though two cusp locations are captured (notably the coherent twin boundary at 54.5 deg). Shallow cusps along both geodesics are not resolved, though anomalies in the derivative may be present such as the  $\Sigma 5$   $\langle 100 \rangle$  ST GB at 18 degrees. These energy estimates do not incorporate NB points with zero energy.

#### 4. Summary and Conclusions

In this work we have used a geodesic metric for grain boundaries to visualize and fit grain boundary geometry and energy. Our contributions are:

1. An easily implemented and general method to visualize the global connectivity of grain boundary datasets in 5D given a list of macroscopic crystallographic parameters. The method is physically meaningful and interpretable but is currently limited to datasets with several thousand or fewer grain boundaries.
2. Several new observations about global grain boundary geometry. The delimiting volume formed by the nearly 600 GBs considered in this work suggests the existence of a grain

boundary fundamental zone with well defined subsets of high symmetry boundaries along faces and general boundaries in the interior. Perfect crystals and symmetric tilt boundaries form the closure of two well defined sets in the reduced space and make up two faces of this delimiting volume. The distribution of symmetric tilt and perfect crystal grain boundaries on each face is equivalent to the standard stereographic projection of boundary plane inclinations within each class.

3. Verification of grain boundary interpolation. We have shown that grain boundary datasets can be extended to new regions of the 5D space in a directed manner by connecting pairs of arbitrary grain boundaries with geodesics. These geodesics are well behaved under dimensionality reduction.
4. A kernel regression model for grain boundary energy currently achieves 10% RMSE using only one fitting parameter, a pairwise geodesic distance matrix, and a random subset of precomputed grain boundary energies as input.

## Acknowledgments

This work was supported by the National Science Foundation under award numbers GRFP-1252522 (IC) and DMR-1710186 (EAH, TF). The authors would also like to acknowledge financial support from the ONR Vannevar Bush Fellowship # N00014-16-1-2821 (MDG) and use of the microscopy and computational resources of the Materials Characterization Facility at Carnegie Mellon University supported by grant MCF-677785.

## References

- [1] W. Ludwig, A. King, P. Reischig, M. Herbig, E. M. Lauridsen, S. Schmidt, H. Proudhon, S. Forest, P. Cloetens, S. R. Du Roscoat, et al., New opportunities for 3d materials science of polycrystalline materials at the micrometre lengthscale by combined use of x-ray diffraction and x-ray imaging, *Materials Science and Engineering: A* 524 (1-2) (2009) 69–76.
- [2] R. Pokharel, J. Lind, S. F. Li, P. Kenesei, R. A. Lebensohn, R. M. Suter, A. D. Rollett, In-situ observation of bulk 3d grain evolution during plastic deformation in polycrystalline cu, *International Journal of Plasticity* 67 (2015) 217–234.
- [3] D. L. Olmsted, E. A. Holm, S. M. Foiles, Survey of computed grain boundary properties

in face-centered cubic metals?ii: Grain boundary mobility, *Acta Materialia* 57 (13) (2009) 3704–3713.

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

- [5] D. M. Saylor, B. S. El-Dasher, B. L. Adams, G. S. Rohrer, Measuring the five-parameter grain-boundary distribution from observations of planar sections, *Metallurgical and Materials Transactions A* 35 (7) (2004) 1981–1989.

- [6] G. S. Rohrer, D. M. Saylor, B. E. Dasher, B. L. Adams, A. D. Rollett, P. Wynblatt, The distribution of internal interfaces in polycrystals, *Zeitschrift für Metallkunde* 95 (4) (2004) 197–214.

- [7] E. R. Homer, S. Patala, J. L. Priedeman, Grain Boundary Plane Orientation Fundamental Zones and Structure-Property Relationships, *Scientific Reports* 5 (2015) 15476.

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

- [9] T. Francis, I. Chesser, S. Singh, E. A. Holm, M. De Graef, A geodesic octonion metric for grain boundaries, *Acta Materialia* 166 (2019) 135–147.

- [10] A. Morawiec, On distances between grain interfaces in macroscopic parameter space, *Acta Materialia* 181 (2019) 399–407.

- [11] A. Sutton, E. Banks, A. Warwick, The five-dimensional parameter space of grain boundaries, *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 471 (2181) (2015) 20150442.

- [12] V. V. Bulatov, B. W. Reed, M. Kumar, Grain boundary energy function for fcc metals, *Acta Materialia* 65 (2014) 161–175.

- [13] F. Abdeljawad, S. M. Foiles, A. P. Moore, A. R. Hinkle, C. M. Barr, N. M. Heckman, K. Hattar, B. L. Boyce, The role of the interface stiffness tensor on grain boundary dynamics, *Acta Materialia* 158 (2018) 440–453.

- [14] D. Rowenhorst, A. Rollett, G. Rohrer, M. Groeber, M. Jackson, P. J. Konijnenberg, M. De Graef, Consistent representations of and conversions between 3d rotations, *Modelling and Simulation in Materials Science and Engineering* 23 (8) (2015) 083501.
- [15] A. Morawiec, A new metric for the space of macroscopic parameters of grain interfaces, *Metallurgical and Materials Transactions A* 50 (9) (2019) 4012–4015.
- [16] D. Olmsted, A new class of metrics for the macroscopic crystallographic space of grain boundaries, *Acta Materialia* 57 (2009) 2793–2799.
- [17] S. Singh, F. Ram, M. De Graef, Emsoft: open source software for electron diffraction/image simulations, *Microscopy and Microanalysis* 23 (S1) (2017) 212–213.
- [18] T. F. Cox, M. A. Cox, *Multidimensional scaling*, Chapman and hall/CRC, 2000.
- [19] C. K. Williams, On a connection between kernel pca and metric multidimensional scaling, *Machine Learning* 46 (1-3) (2002) 11–19.
- [20] L. Kassab, Multidimensional scaling: Infinite metric measure spaces, *arXiv preprint arXiv:1904.07763*.
- [21] H. Kvinge, M. Blumstein, Letting symmetry guide visualization: multidimensional scaling on groups, *arXiv preprint arXiv:1812.03362*.
- [22] M. Balasubramanian, E. L. Schwartz, The isomap algorithm and topological stability, *Science* 295 (5552) (2002) 7–7.
- [23] A. Feragen, F. Lauze, S. Hauberg, Geodesic exponential kernels: When curvature and linearity conflict, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3032–3042.
- [24] L. Deng, The mnist database of handwritten digit images for machine learning research [best of the web], *IEEE Signal Processing Magazine* 29 (6) (2012) 141–142.
- [25] P. Woit, Woit, Bartolini, *Quantum theory, groups and representations*, Springer, 2017.
- [26] D. Wolf, Correlation between the energy and structure of grain boundaries in bcc metals. ii. symmetrical tilt boundaries, *Philosophical Magazine A* 62 (4) (1990) 447–464.

- [27] M. A. Tschopp, S. P. Coleman, D. L. McDowell, Symmetric and asymmetric tilt grain boundary structure and energy in cu and al (and transferability to other fcc metals), *Integrating Materials and Manufacturing Innovation* 4 (1) (2015) 11.
- [28] S. Singh, M. De Graef, Orientation sampling for dictionary-based diffraction pattern indexing methods, *Modelling and Simulation in Materials Science and Engineering* 24 (8) (2016) 085013.
- [29] J. W. Cahn, J. E. Taylor, Metrics, measures, and parametrizations for grain boundaries: a dialog, *Journal of Materials Science* 41 (23) (2006) 7669–7674.
- [30] T. Yu, S. Yang, C. Deng, Survey of grain boundary migration and thermal behavior in ni at low homologous temperatures, *Acta Materialia*.
- [31] I. Chesser, T. Yu, C. Deng, E. Holm, B. Runnels, A continuum thermodynamic framework for grain boundary motion, *Journal of Mechanics and Physics of Solids* 10 (5) (2019) 145–165.
- [32] C. W. Rosenbrock, E. R. Homer, G. Csányi, G. L. Hart, Discovering the building blocks of atomic systems using machine learning: application to grain boundaries, *npj Computational Materials* 3 (1) (2017) 1–7.
- [33] E. R. Homer, D. M. Hensley, C. W. Rosenbrock, A. H. Nguyen, G. L. Hart, Machine-learning informed representations for grain boundary structures, *Frontiers in Materials* 6 (2019) 168.

540 **Declaration of interests**

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: