Towards a Quantitative Cartography of the Grain Boundary Energy Landscape: Paths and Correlations

Sterling G. Baird[a,*], Eric R. Homer[a], David T. Fullwood[a], Oliver K. Johnson[a]

[a]*Department of Mechanical Engineering, Brigham Young University, Provo, UT 84602, USA*

Abstract

We apply a newly developed Voronoi fundamental zone (VFZ) framework to gain insights about grain boundary (GB) structure-property relationships in the five degree-of-freedom (5DOF) space of cubic GBs. We analyze the shape and size of a 5DOF fundamental zone (FZ), molecular statics energy uncertainty, property similarity of GBs that are crystallographically "close" (i.e. correlations), and energy pathways through 5DOF space. Considered together, these insights are important for managing trade-offs between accuracy, complexity, and design considerations for electron backscatter diffraction/serial sectioning, high-energy diffraction microscopy, molecular statics, and density-functional theory. In terms of the shape and size of a 5DOF FZ, we discover that a FZ is smaller than expected at only $\sim$65° in the largest principal component. Thus, a 10° difference between two GBs, which may have previously been considered small, is actually quite large. We represent a GB by five transformed Cartesian coordinates equipped with a Euclidean distance metric. Using this representation, we find that the FZ has a low aspect-ratio shape (i.e. width, length, height, etc. are similar) which is important for 5DOF numerical differentiation. Semivariogram and numerical optimization methods reveal that grain boundary energy (GBE) in Ni and Fe are globally correlated within $\sim$6° to 8° in the grain boundary octonion (GBO) sense (multiply by 2 to convert to ~~tilt~~ angle). For local correlation lengths of high-symmetry GBs of interest, we notice significant variation relative to global correlation lengths and an inverse relationship with the Brandon ~~criteria~~. We suggest that property data with no more than $\pm \sim$3 % error and point sets with GBs that are no more than $\sim$3−4° apart should be used and then paired with high-fidelity interpolation strategies. Finally, in terms of dynamic material behavior, direct paths through 5DOF space for Ni suggest that a certain low-energy Σ7 GB may transform into the frequently observed Σ3 coherent-twin GB which may be interesting to verify by experiment or simulation.

Keywords: grain boundary energy, five degree-of-freedom, structure-property model, machine learning, octonion

1. Introduction

When subjected to plastic deformation and/or elevated temperatures, the crystallographic character of grain boundaries (GBs) in polycrystalline microstructures can change to reduce the total system energy [1–5]. These energy minimizing transformations of GB crystallography constitute trajectories or paths along the GB energy landscape, which is a function over the 5-dimensional (5D) GB character space. Consequently, to understand, model, and predict such transformations and their effects on microstructure evolution, it is necessary to understand these paths and the relationships between GBs that are near each other in that space.

Examples of such spontaneous energy minimiz-

---

*Corresponding author.

 Email address: ster.g.baird@gmail.com (Sterling G. Baird)

Summary of comments: merged-gb-correlation - OKJ comments.pdf

Page:1

Number: 1 Author: sterg Subject: Highlight Date: 2021-09-29 00:20:46

Number: 2 Author: sterg Subject: Highlight Date: 2021-09-29 00:03:43

Number: 3 Author: Oliver Johnson Subject: Inserted Text Date: 2021-10-13 09:50:47
misorientation
Author: sterg Date: 2021-11-20 13:20:15
Accepted set by sterg

Number: 4 Author: Oliver Johnson Subject: Inserted Text Date: 2021-10-13 09:51:20
criterion
Author: sterg Date: 2021-11-20 13:17:59
Accepted set by sterg

Number: 5 Author: Oliver Johnson Subject: Inserted Text Date: 2021-10-13 09:53:08
, under appropriate conditions,
Author: sterg Date: 2021-11-20 13:20:31
Accepted set by sterg

ing changes to GB character include the rotation of the GB plane to achieve equilibrium at triple junctions [6, 7]—both those formed by the intersection of three GBs within a polycrystal and those formed between a GB and two free surfaces at the exterior of a microstructure. It is this phenomenon that is exploited to produce the quarter- and half-loop bicrystal geometries used in classic constant-driving-force GB migration experiments [8–13]. The inclination of the GB plane can also change during GB migration [REF], faceting into different combinations of GB planes to reduce energy [14], or upon the disappearance of a grain during grain growth (which requires establishment of new equilibrium configurations at the newly formed triple junctions). This class of energy minimizing changes to GB character constitute paths through the GB energy landscape that are restricted to two-dimensional sub-manifolds (the GB plane fundamental zone (FZ) for a fixed misorientation).

Another class of energy minimizing changes to GB character are illustrated by the phenomenon of grain rotation [15–24], which can occur during high-temperature plastic deformation [25–31] or recrystallization [32], as well as at lower temperatures in nanocrystalline materials [33]. The primary feature of grain rotation is the change in crystal orientation of one grain relative to its neighbors resulting in lower-energy GBs between them. However, when the misorientation between two grains changes there will also be an accompanying change to the GB plane inclination[1]. Consequently, such GB character transformations represent more general paths through the 5D GB character space that are not restricted to any misorientation or boundary plane subspaces.

Until recently, models for GB energy that depend on their full 5D crystallographic character were unavailable. Bulatov, Reed, and Kumar (BRK) [34] developed a fully 5D GB energy model by fitting a closed-form function to GB energies from a database [35] of 388 GBs in several materials.

This function has been employed in a variety of applications to study mesoscale microstructure phenomena like X [REF], Y [REF], and Z [REF]. We recently developed a general approach, called the Voronoi fundamental zone (VFZ) framework, for inferring GB structure-property models from GB structure-property databases [36].

In this work we present 5D GB structure-property models for FCC Ni and BCC Fe developed using the VFZ framework. We also study correlations in GB energy as a function of crystallographic distance. We use the VFZ framework to give context to GB property correlation lengths[2] and find that previous estimates are likely too high for low-noise, computational GBE calculations. Finally, we investigate general paths through the GB energy landscape and show qualitatively distinct types of relationships between important types of GBs.

2. Methods

We first review the VFZ construction and interpolation methods which have been published elsewhere [38] (Section 2.1). We describe tools to examine 5DOF results, including dimensionality reduction (Section 2.2) and techniques for estimating spatial correlation lengths (Section 2.3) and methods to visualize paths through the 5DOF space (Section 2.4). Finally, we describe literature datasets used in this work (Section 2.5).

2.1. The Voronoi Fundamental Zone Framework

The VFZ is a newly developed tool that allows for efficient and relatively accurate predictions of properties across the 5DOF space [38] by building on the recent GBO distance metric which "correctly determines the angular distances between GBs with a common normal or misorientation" 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" [39]. To best describe what

---

[1]Even if the GB plane remains fixed in the macroscopic reference frame, a change in the GB misorientation results in a change of the GB normal in the crystal reference frame, which is the physically relevant reference frame.

[2]Correlation length in the context of GBs has been described as the degree to which boundaries with the same macroscopic geometrical degrees of freedom in different materials have related properties [37].

Page:2

Number: 1 Author: Oliver Johnson Subject: Highlight Date: 2021-10-13 09:57:06

these placeholders need to be filled in.

Author: sterg Date: 2021-11-20 13:22:49

Deferred set by sterg

Author: sterg Subject: sterg Date: 2021-11-20 13:22:43

These were placeholders that you suggested, but without references. I don't have time for the literature search, so I can either remove them or if you have suggestions I can add them in.

Number: 2 Author: Oliver Johnson Subject: Inserted Text Date: 2021-10-13 09:58:41

new paragraph

Author: sterg Date: 2021-11-20 13:21:44

Accepted set by sterg

Number: 3 Author: Oliver Johnson Subject: Highlight Date: 2021-10-13 09:54:37

missing reference

Author: sterg Date: 2021-11-20 13:23:30

Deferred set by sterg

Author: sterg Subject: sterg Date: 2021-11-22 14:02:02

These were placeholders that you suggested, but without references. I don't have time for the literature search, so I can either remove them or if you have suggestions I can add them in.

Number: 4 Author: Oliver Johnson Subject: Inserted Text Date: 2021-10-13 09:55:10

or when

Author: sterg Date: 2021-11-20 13:23:11

Accepted set by sterg

Number: 5 Author: Oliver Johnson Subject: Highlight Date: 2021-10-13 10:03:22

Given our final results, I don't think this is true. We find global correlation lengths on the order of 6-8 octonion degrees, which corresponds to a misorientation angle of about 15 degrees of misorientation, which is the classical low- to high-angle threshold. Moreover, historical guesses for a correlation length were about 10 degrees of misorientation, so we actually find that they are correlated over larger distances than the what is estimated in the literature.

Author: sterg Date: 2021-11-20 13:25:38

Accepted set by sterg

Author: sterg Subject: sterg Date: 2021-11-22 14:01:51

I think this is true for the low-noise case as mentioned though. See Table 4. However, I'll mention that without the low-noise condition what you said.

Number: 6 Author: Oliver Johnson Subject: Highlight Date: 2021-10-13 10:00:54

This is not the correct definition. Rather the correlation length is the crystallographic distance over which GBs exhibit similar properties, for a single material. In other words it is the distance over which GBs with DIFFERENT macroscopic geometrical degrees of freedom, in the SAME material have similar properties.

Author: sterg Date: 2021-11-20 13:25:49

Accepted set by sterg

the VFZ framework is and how it all fits together, we summarize its requisite parts and steps for creating and defining a VFZ , mapping GBOs into the VFZ , distance calculations , and interpolation Section 2.1.[2] We also provide a brief comparison with the traditional GBO metric. The methods are based on functions and scripts from (`github.com/sgbaird-5dof/interp`), and we refer the reader to Baird et al. [38] for a detailed description of the methods and model.

2.2. Dimensionality Reduction

A singular value decomposition transformation is used to remove degenerate dimensions and rotate/align the Voronoi fundamental zone grain boundary octonion (VFZ-GBO) point clouds such that the 1st, 2nd, 3rd, etc. dimensions are progressively smaller.

Additionally, principal component analysis is applied and the variance explained by each dimension is extracted.

2.3. Correlation Lengths

We use semivariogram and numerical optimization methods to obtain several estimates of correlation lengths L from VFZ-GBO pairwise distances and GBEs for two GB datasets Section 2.5.

In the semivariogram method, we assume a stationary Gaussian kernel and limit the semivariogram to half of the maximum pairwise distance. We use a correlation strength of $\rho = 0.61$ for correlation lengths reported in this work; however, as the choice of correlation strength is arbitrary, one can use Eq. (S5) to determine the length scale corresponding to any specified correlation strength. See Section S2 for additional information.

In a separate method for estimating correlation length, numerical optimization is performed via gradient-descent–based maximization of a likelihood function that depends on ~~correlation length L and noise σ parameters~~ [40].

2.4. Visualizing 5DOF Paths

Direct paths between two GBs in a VFZ are obtained via coordinate interpolation constrained to a hyperspherical arc. The direct path in a VFZ is not always the minimum distance path (which may cross the borders of a VFZ). However, it is instructive to observe these paths because the minimum distance path in 5DOF space is not necessarily the path a GB will take during grain growth.

2.5. Literature Datasets

Ni [37] and Fe [41] GBE datasets from the literature are used. Intrinsic uncertainty for the Fe simulation data is estimated by the following steps:

1. Sort GBs into degenerate sets

2. Determine the average GBE for each degenerate set

3. Compare each of the degenerate GBs to the set-wise average GBE (root mean square error or mean absolute error)

See Section S3 for further details on methods used to estimate intrinsic uncertainty of the Fe simulation dataset.

3. Results and Discussion

We use the VFZ framework to explore structure-property relations in the 5DOF space of GBs. We address the following questions:

- How large is a 5DOF fundamental zone? What is it shaped like? (Section 3.1)

- How similar are the energies for GBs with similar crystallographic character (i.e. how correlated)? How close/densely spaced are randomly generated GBs? (Section 3.2)

- What is the uncertainty of molecular statics simulations? (Section 3.3)

- What can crystallographic paths in 5DOF space teach us about material behavior? (Section 3.4)

Finally, we discuss the potential for computing numerical derivatives with respect to 5DOF space for computing minimum energy paths and local minima (Section 3.5).

Number: 1   Author: Oliver Johnson   Subject: Highlight   Date: 2021-10-13 13:25:11

I would modify this a little bit to point out that crystallographic transformations (e.g. changes in the GB plane) will be governed not by the crystallographic distance between two points, but rather the topography of the GB energy landscape between them. The shortest distance between Provo and Heber is over the mountains, but I always drive the long way through the canyon.

Number: 2   Author: Oliver Johnson   Subject: Highlight   Date: 2021-11-22 14:02:44

1. This IS section 2.1, so I think this must be an outdated reference.2. I don't see anywhere in the paper where you actually include what you've described here (steps for creating a VFZ, mapping GBOs into the VFZ, etc.). So I think all of Section 2.1 may need updating. For the purposes of this paper, I think all that is needed is for the reader to have a conceptual understanding of what a VFZ is. I don't think the reader needs all of the gory details, for that you should just point them to the CMS paper. I think all you need to describe here is how a VFZ is defined, and that once created it facilitates the other things (distance calculations, interpolation, etc.), but you don't need to give the details about how those other things are done.

> **Author: sterg   Subject: sterg   Date: 2021-11-20 13:32:57**
>
> This was a dangling reference. I agree that we could give a few more details about the methods.

Number: 3   Author: sterg   Subject: Highlight   Date: 2021-09-29 00:04:35

Number: 4   Author: Oliver Johnson   Subject: Inserted Text   Date: 2021-10-13 13:19:22

, l,

Number: 5   Author: Oliver Johnson   Subject: Inserted Text   Date: 2021-10-13 13:26:47

of GB energy

Author: sterg   Date: 2021-11-20 13:34:43

Accepted set by sterg

Number: 6   Author: Oliver Johnson   Subject: Inserted Text   Date: 2021-10-13 13:20:21

-

Author: sterg   Date: 2021-11-20 13:27:49

Accepted set by sterg

Number: 7   Author: Oliver Johnson   Subject: Inserted Text   Date: 2021-10-13 13:21:03

correlation length, l, and noise parameters, sigma

Author: sterg   Date: 2021-11-20 13:28:58

Accepted set by sterg

Number: 8   Author: sterg   Subject: Note   Date: 2021-09-29 00:31:52

l and sigma should have parantheses around them

Number: 9   Author: Oliver Johnson   Subject: Inserted Text   Date: 2021-10-13 13:27:15

finding

Author: sterg   Date: 2021-11-20 13:34:36

Accepted set by sterg

3.1. Dimensions of a VFZ

The maximum principal component of a particular singular value decomposition transformed O_h cubic VFZ is $\sim$65°. The sizes for this first and the remaining 7 dimensions are given in Table 1.

Table 1: Dimension of singular value decomposition transformed coordinates (Dimension) and GBO dimension size (ω) for a set of 50 000 VFZ-GBOs. These are the diagonal entries of the "S" matrix in the singular value decomposition decomposition.

Dimension	ω (°)
1	65.03
2	63.24
3	58.8
4	53.7
5	46.34
6	6.651
7	5.821
8	2.436e-13

By performing principal component analysis, we find that $\sim$99.6 % of the variance is explained by the first 5 transformed coordinates of an O_h VFZ as shown in Table 2.

Table 2: Dimension of principal component analysis transformed coordinates (Dimension) and percent variance explained (v) for a set of 50 000 VFZ-GBOs. The first 5 dimensions cumulatively explain $\sim$99.6 % of the variance.

Dimension	v (%)
1	25.19
2	23.82
3	20.6
4	17.18
5	12.79
6	0.2631
7	0.1565
8	3.532e-28

What does it mean that the 5th dimension is $\sim$70 % the size of the 1st dimension? This gives an indication of the shape of the VFZ. The information-dense dimension sizes are approximately isotropic, indicating a space that is not a perfect hypersphere or hypercube, but somewhat hyper-elliptical or hyper-rectangular.

3.2. Correlation Lengths

We present global (Section 3.2.1) and local (Section 3.2.2) correlations for the Ni and Fe simulation datasets. We then describe how the number of GB in a VFZ affects, on average, how close GBs are to their nearest neighbors (NNs) (Section 3.2.3) in the context of correlation.

3.2.1. Global Correlation Lengths

Using the input data for each of the datasets, global correlation lengths were obtained via the semivariogram method described in Section 2.3 with a bin width of 1 d_Ω°. Figure S1 shows the empirical semivariograms together with the analytical fits used to obtain the values of the respective correlation lengths. The plateau at long distances suggest that GBE correlations for Ni and Fe are Gaussian in nature (Section S2.2).

In addition to the correlation length estimates obtained via the semivariogram method, we also computed correlation length estimates using the gradient-based `fitrgp()` method described in Section 2.3.

To better understand the Gaussian process regression (GPR) models that were trained on the GBE data, we also interrogated their correlation lengths (as opposed to the input data only). This was done by the semivariogram method, with the GBEs ($E(x_i)$ and $E(x_j)$ in Eq. (S3)) given by the GPR model (which is the mean of a collection of posterior models). This was done in two ways: (1) by evaluating the posterior models at the same locations as the input data, and (2) by evaluating the posterior models at randomly selected points (replications of which facilitate uncertainty quantification). The correlation lengths obtained by all of these methods are summarized in Table 3.

For the Ni dataset, the input data exhibits a correlation length of about 7.5 d_Ω° for both the semivariogram and gradient methods. For the Fe dataset, we observe 6.2665 d_Ω° for the semivariogram method and 8.3073 d_Ω° for the gradient method, the average of the two being 7.3 d_Ω°. Thus both datasets seem to have correlation lengths of about 7.5 d_Ω°. This is

Number: 1 Author: Oliver Johnson Subject: Inserted Text Date: 2021-10-13 17:01:29

The first 5 dimensions contain essentially all of the information, which is consistent with the fact that there should be only 5 independent crystallographic parameters to describe a GB. The fact that these 5 dimensions are of similar size indicates that the VFZ is roughly isometric and can be thought of as a hyper-rectangle or hyper-ellipse with only minor eccentricity.

Author: sterg Date: 2021-11-20 13:52:54

Accepted set by sterg

Number: 2 Author: Oliver Johnson Subject: Highlight Date: 2021-10-13 13:58:36

In this section you've only reported results for a single VFZ, is this for the Ni or the Fe? Presumably you had to create one for each of them. If so, how different are they? Or perhaps you only needed to make a single VFZ (and corresponding VFZO set), but you get 2 different GB energy landscapes for the respective data sets (seems more likely). But if this is the case, it should be articulated explicitly.

Author: sterg Date: 2021-11-20 13:53:02

Accepted set by sterg

Number: 3 Author: sterg Subject: Highlight Date: 2021-09-29 00:04:49

Number: 4 Author: Oliver Johnson Subject: Inserted Text Date: 2021-10-13 17:01:57

GBs

Author: sterg Date: 2021-11-20 13:53:33

Accepted set by sterg

Number: 5 Author: Oliver Johnson Subject: Inserted Text Date: 2021-10-13 17:02:20

VFZO set

Author: sterg Date: 2021-11-20 13:53:47

Accepted set by sterg

Number: 6 Author: Oliver Johnson Subject: Highlight Date: 2021-10-13 16:51:32

Need to make it clear what units this is, and standardize the notation.

Author: sterg Date: 2021-11-20 21:39:47

Accepted set by sterg

Number: 7 Author: Oliver Johnson Subject: Highlight Date: 2021-10-13 16:51:48

Need to make it clear what units this is, and standardize the notation.

Author: sterg Date: 2021-11-20 21:39:43

Accepted set by sterg

Number: 8 Author: sterg Subject: Highlight Date: 2021-09-29 00:05:01

Number: 9 Author: Oliver Johnson Subject: Highlight Date: 2021-10-15 10:52:02

This is not true, it is the change in concavity that is characteristic of Gaussian correlations.

Author: sterg Date: 2021-11-20 13:55:02

Accepted set by sterg

Number: 10 Author: Oliver Johnson Subject: Highlight Date: 2021-10-13 16:54:06

While this term comes directly from PCA, I think it will confuse the readers since we're not talking about property variation, rather this " variance" is simply the spatial dispersion of the points themselves. Perhaps we can just add a parenthetical clarification, like this: "...~99.6% of the variance (spatial dispersion) is explained..."

Author: sterg Date: 2021-11-20 13:54:32

Accepted set by sterg

3.1. Dimensions of a VFZ

The maximum principal component of a particular singular value decomposition transformed O_h cubic VFZ is $\sim 65°$. The sizes for this first and the remaining 7 dimensions are given in Table 1.

Table 1: Dimension of singular value decomposition transformed coordinates (Dimension) and GBO dimension size (ω) for a set of 50 000 VFZ-GBOs. These are the diagonal entries of the "S" matrix in the singular value decomposition decomposition.

Dimension	ω (°)
1	65.03
2	63.24
3	58.8
4	53.7
5	46.34
6	6.651
7	5.821
8	2.436e-13

By performing principal component analysis, we find that $\sim 99.6\%$ of the variance is explained by the first 5 transformed coordinates of an O_h VFZ as shown in Table 2.

Table 2: Dimension of principal component analysis transformed coordinates (Dimension) and percent variance explained (v) for a set of 50 000 VFZ-GBOs. The first 5 dimensions cumulatively explain $\sim 99.6\%$ of the variance.

Dimension	v (%)
1	25.19
2	23.82
3	20.6
4	17.18
5	12.79
6	0.2631
7	0.1565
8	3.532e-28

What does it mean that the 5th dimension is $\sim 70\%$ the size of the 1st dimension? This gives an indication of the shape of the VFZ. The information-dense dimension sizes are approximately isotropic, indicating a space that is not a perfect hypersphere or hypercube, but somewhat hyper-elliptical or hyper-rectangular.

3.2. Correlation Lengths

We present global (Section 3.2.1) and local (Section 3.2.2) correlations for the Ni and Fe simulation datasets. We then describe how the number of GB in a VFZ affects, on average, how close GBs are to their nearest neighbors (NNs) (Section 3.2.3) in the context of correlation.

3.2.1. Global Correlation Lengths

Using the input data for each of the datasets, global correlation lengths were obtained via the semivariogram method described in Section 2.3 with a bin width of 1 d_Ω°. Figure S1 shows the empirical semivariograms together with the analytical fits used to obtain the values of the respective correlation lengths. The plateau at long distances suggest that GBE correlations for Ni and Fe are Gaussian in nature (Section S2.2).

In addition to the correlation length estimates obtained via the semivariogram method, we also computed correlation length estimates using the gradient-based `fitrgp()` method described in Section 2.3.

To better understand the Gaussian process regression (GPR) models that were trained on the GBE data, we also interrogated their correlation lengths (as opposed to the input data only). This was done by the semivariogram method, with the GBEs ($E(x_i)$ and $E(x_j)$ in Eq. (S3)) given by the GPR model (which is the mean of a collection of posterior models). This was done in two ways: (1) by evaluating the posterior models at the same locations as the input data, and (2) by evaluating the posterior models at randomly selected points (replications of which facilitate uncertainty quantification). The correlation lengths obtained by all of these methods are summarized in Table 3.

For the Ni dataset, the input data exhibits a correlation length of about 7.5 d_Ω° for both the semivariogram and gradient methods. For the Fe dataset, we observe 6.2665 d_Ω° for the semivariogram method and 8.3073 d_Ω° for the gradient method, the average of the two being 7.3 d_Ω°. Thus both datasets seem to have correlation lengths of about 7.5 d_Ω°. This is

Table 3: Correlation lengths obtained via semivariogram and gradient based approaches in units of GBO distance (multiply by 2 to get tilt angle). For the semivariogram of the posterior mean model evaluated at random points, we used 10^3 random points and repeated this process 10 times. The corresponding values in this table represent the mean $\pm$ one standard deviation of those 10 replicates.

	$l : d_\Omega$ [°]	
	Ni	Fe
Semivariogram (Input Data)	7.5491	6.2665
Semivariogram (Posterior Mean at Input Points)	7.7316	7.5179
Semivariogram (Posterior Mean at Random Points)	8.8551 ± 0.1970	7.7082 ± 0.1746
Gradient (Input Data)	7.3995	8.3073

significant. For symmetric tilt GBs, the octonion distance metric corresponds to half of the difference in misorientation angles. That means that in the more traditional units of degrees of misorientation difference the observed correlation lengths are very close to 15°, which is the traditional low-angle GB threshold. Thus, the traditional 15° threshold appears to hold (on average) more generally than just for differentiating low- and high-angle GBs. Rather we find that on average, GBs that are within 15° (7.5 d_Ω°) will have well correlated energies regardless of their misorientations or GB planes (i.e. it is not restricted to symmetric tilt GBs, but holds across the entire GB character space).

The correlation lengths obtained by evaluating the GPR models are similar to those obtained directly from the input data, but they are slightly larger, suggesting that GB energies in the trained models are correlated over a slightly longer distance than the input data, though the difference is nearly negligible (about 1 d_Ω° or less).

Using the correlation lengths, l, obtained via the semivariogram of the input data, Figure 1 shows the length scale, l', corresponding to an arbitrary user-specified correlation strength, ρ. The length scales over which GB energies are correlated range from $l' = 0$ d_Ω° for perfect correlation ($\rho = 1$) to $l' \approx 23$ d_Ω° and $l' \approx 19$ d_Ω° for essentially[3] zero correlation in the Ni and Fe datasets, respectively.
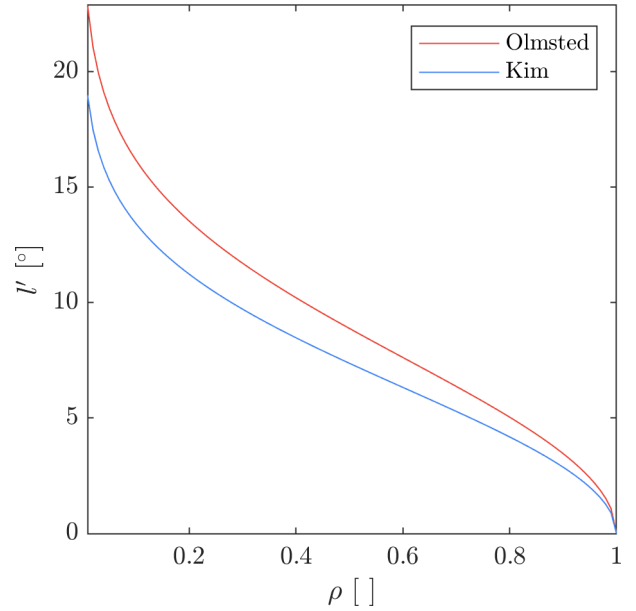

Figure 1: Global length scale of correlation, l', in units of GBO distance (d_Ω [°]) as a function of correlation strength, ρ, for the Ni (Olmsted) and Fe (Kim) datasets. Multiply GBO distance by 2 to convert to tilt angle.

---

[3]Technically the correlation does not go to exactly zero until the length scale is infinite. The values listed here are for a correlation of $\rho = 0.01$.

Number: 1 Author: sterg Subject: Highlight Date: 2021-09-29 00:05:09

Number: 2 Author: sterg Subject: Highlight Date: 2021-09-29 00:05:14

Number: 3 Author: sterg Subject: Highlight Date: 2021-09-29 00:05:27

Number: 4 Author: sterg Subject: Highlight Date: 2021-09-29 00:32:32

Table 4: Fitted parameters for two GPR models fitted to the 388 simulated Ni GBEs by Olmsted et al. [35] and fitted parameters for a GPR model trained on 80% of the Fe simulation data (46 883 GBs). The first Ni model allows σ to vary, whereas the second constrains σ to be fixed. Mat., σ_L, σ_F, β, and σ are the material (i.e. element), kernel length scale (°), signal standard deviation (Jm^{-2}), constant basis function (Jm^{-2}), and input property standard deviation (Jm^{-2}), respectively.

Mat.	Fix σ	σ_L (°)	σ_F	β	σ
Ni	no	7.3995	0.2049	1.0913	0.0321
Ni	yes	1.9354	0.201	1.1044	0.0001
Fe	no	8.3073	0.0716	1.2192	0.0562

GPR facilitates numerically solving for correlation lengths (i.e. how similar properties are for similar crystallographic character), which are determined to be $\sim$7.4° and 8.3° for the Ni and Fe datasets, respectively. If low noise is assumed for Ni, the correlation length drops to $\sim$2 degrees. See Table 4 for GPR parameters for each dataset. By contrast, if a GPR model is trained on a large set of 50 000 GBs sampled from the Bulatov Reed Kumar (BRK) model, the numerically optimized correlation length is 10.5° suggesting that the BRK model imposes a-priori information that GBs are more correlated than the data warrants. In other words, the BRK function is too smooth.

The two simulation datasets have distinct differences from each other, as summarized in Table 5.

Table 5: Comparison of Ni (Olmsted et al. [37]) and Fe (Kim et al. [41]) molecular statics simulation datasets. The differences in noise-levels results from whether multiple initial starting configurations were probed in search of a globally minimized configuration as opposed to using a single metastable configuration.

Property	Ni	Fe
Size	388	58604
Noise	Low	High
Symmetry	FCC	BCC

Despite these differences in terms of noise, dataset size, and crystal symmetry, it is interesting to see that the correlation lengths within a VFZ are similar for the two datasets. Both are lower than the correlation lengths of 10° [42] and 15° [37] previously reported[4] with respect to misorientation, and are about on par in terms of boundary plane normal owing to the fact that GBO distances for tilt angles are half the value of tilt angles.

It is reasonable to assume that the Ni data has low noise due to use of a global optimization strategy [37]; thus, the correlation length of $\sim$1.9° after imposing the low-noise condition suggests that *the Ni dataset may have an actual correlation length much smaller than previously reported.*

By contrast, the correlation length of a GPR model trained on many BRK GBEs remains relatively large at 10.4°. What does this suggest? The BRK model is smoothed more than the data warrants on its own. This has the following implications:

1. More sophisticated methods are required which do not impose mistaken a-priori information about the correlation length[5]. Instead, the data itself should suggest proper correlation lengths

2. Larger, low-noise datasets which span all 5DOF are necessary to be confident in structure-property paths that are not restricted to a single misorientation fundamental zone or boundary plane fundamental zone

We believe that the GPR model within the VFZ framework meets the requirements of point #1 and is capable of handling the ideal dataset proposed in point #2.

Some questions that remain are:

- Does the similarity between correlation lengths for FCC and BCC extend to non-cubic crystal symmetries?

---

[4]Both of these lengths are based on results from Olmsted et al. [37].

[5]The a-priori information that the BRK model imposes is that correlation lengths within a misorientation fundamental zone or boundary plane fundamental zone hold for arbitrary paths through 5DOF space and that moving from one subspace to another results in monotonic behavior.

Number: 1 Author: Oliver Johnson Subject: Highlight Date: 2021-10-16 18:04:47

- What are the differences in correlation length for other properties? (e.g. mobility)

It is possible that the correlation length will increase with the size of the VFZs, and we expect that the correlation length will depend on the property of interest.

3.2.2. Local Correlation Lengths

In addition to the global correlation lengths reported in Section 3.2.1, we also investigated correlation length scales locally in the vicinity of certain low-Σ GBs. This is accomplished by the same semivariogram process explained in Section 2.3, except that in Eq. (S3) we consider only pairs of GBs for which one of the GBs is the particular GB of interest (i.e. we fix one of the GBs). Because the resulting set of GB pairs is smaller than in the global analysis, we use larger bins with a width of 2 d_Ω° for the local empirical semivariograms.

Figures S2 and S3 show the local empirical semivariograms for the Ni and Fe datasets respectively. Each panel shows the semivariogram centered at a different low-Σ GB. While noisier than the global empirical semivariograms, reasonable fits were obtained for all except $\Sigma5$ in the Ni dataset Section S2.3.

Table 6 provides ~~the correlation local lengths ob~~tained for each of the low-Σ GBs for each of the datasets. The most noteworthy observation is that these correlation lengths are different from one another and different from the respective global correlation lengths. While the local correlation lengths for the $\Sigma3$ GBs are similar to the global values for both datasets, other GBs have correlation lengths more than twice as large as the global correlation length (the $\Sigma9$ in the Ni dataset, and the $\Sigma5$ in the Fe dataset). The fact that certain local correlation lengths differ from the global correlation length suggests that a non-stationary Gaussian kernel (which depends on both relative distance and absolute location) may be a more appropriate choice than the stationary Gaussian kernel (Section S2.3).

These local correlation lengths facilitate an interesting comparison with the canonical Brandon criterion [43]. The Brandon criterion provides a

Table 6: Local correlation lengths in the vicininty of specific low-Σ GBs, obtained via the semivariogram method in units of octonion distance. The fit of the $\Sigma5$ GB for the Ni dataset was sufficiently poor that we do not report a corresponding correlation length.

	$l : d_\Omega$ [°]	
	Ni	Fe
$\Sigma3$	7.1460	7.0921
$\Sigma5$	—	16.0440
$\Sigma7$	9.6533	10.1493
$\Sigma9$	17.2601	10.7792
$\Sigma11$	10.2027	11.6011

threshold for the maximum angular deviation that can be accommodated by a coincident site lattice boundary before losing coincidence:

$$\Delta\theta = \theta_0 \Sigma^{-1/2} \qquad (1)$$

where $\Delta\theta$ is the angular deviation threshold, $\theta_0 = 15°$ a constant corresponding to the low- to high-angle GB threshold and Σ is the reciprocal density of coincidence lattice points (i.e. the coincident site lattice number). The Brandon criterion predicts that amount of distortion that the GB can accommodate (e.g. via introduction of GB dislocations) before losing coincidence should *decrease* with Σ because the density of coincident sites decreases with Σ.

In contrast, Figure 2 shows that correlation lengths generally *increase* with Σ. That is to say that in general there is a trend of GB energies being correlated over longer length scales with increasing Σ. This seems reasonable since the deepest cusps in the GB energy landscape (corresponding to small correlation lengths) tend to correspond to the lowest-Σ GBs. This suggests that while strict coincidence may indeed be lost over smaller crystallographic deviations with increasing Σ, GB energies seem to be relatively insensitive to this loss of coincidence and instead seem to be increasingly correlated over larger length scales. This may also imply that the coincident site lattice model becomes decreasingly relevant for predicting GB energy as Σ increases [44].

Number: 1 Author: Oliver Johnson Subject: Highlight Date: 2021-11-20 21:06:28

General Comment: This section feels like a cut and paste. There's not enough context to really understand the motivation of this, what exactly you did, or to appreciate the results. It also repeats itself several times.Specific Comments:(1) The first sentence "GPR facilitates..." is unnecessary since you already explained this and gave the results a few paragraphs ago.(2) I don't know what it means to have low or high noise, nor why/how one would impose this after the fact. The noise in the data just is what it is, right? (3) I don't know what you mean by " ~1.9 deg...the Ni dataset may have an actual correlation length much smaller than previously reported." We just showed by 2 methods that data suggests a correlation length of around 7.5 oct.deg., so why are you telling me that it is 1.9 deg now? Without imposing any other assumptions on the data, the data shows 7.5 oct.deg.(4) Is Table 4 necessary? I think for the correlation length discussion, we only need to report the sigma_L column, and 2 of the 3 values in this column are already reported in Table 3. Perhaps we could just add 1 row to Table 3 to report the correlation length for Ni when low noise is assumed?(5) I think this section as a whole can be abbreviated and integrated better into the preceding discussion. I think the key points are: The BRK model has a correlation length of about 10.5 oct.deg. which is about 40% larger than what the data suggests, indicating that the BRK model is overly smooth. The correlation length obtained by the fitRGP method is sensitive to certain parameters, including the noise parameter. If you use a small noise parameter you find a much smaller correlation length for Ni (about 1.9 oct.deg). However, I'm not sure what to make of this last point or whether or not to include it, since it is not consistent with the correlation length found by another independent method (the semivariogram method)--I would be more inclined to say that assuming low noise is not consistent with the data, and providing a suggestion for how to pick a proper noise parameter value (if fitRGP picked the right value automatically in the first place though, I'm not sure we need to include this result or discussion, since it seems fitRGP handles it fine).

Author: sterg Subject: sterg Date: 2021-11-22 15:49:10

(1) that's fine(2, 3) It has to do with the global optimization used to get the GBE value. We talked about how Olmsted is low-noise, but Kim is high-noise, right? I think we've talked about this, but it has been a while. I'm not thinking it's low-noise because the correlation models suggest it has low-noise. Rather, it has low-noise because of how thorough Olmsted was with the global optimization to get those values. In other words, it's a-priori information that we can impose or leave out. When we impose it, the correlation length drops dramatically. The methods you included in the paper as far as I know didn't have an a-priori noise level set, so to me it's not conflicting information. If you impose a low-noise condition a-priori and the correlation lengths were still very large for semivariogram etc. then that's a different story. (4, 5) I'd rather keep the small noise parameter result and discussion. The low-noise assumption seems to conflict with how the standard, default methods we use interpret the data, but another way of stating this is that our a-priori information about the level of noise of the data is inconsistent with the level of noise predicted without that information. Something else about the low-noise assumption is that it respects cusps in the model.

Number: 2 Author: sterg Subject: Highlight Date: 2021-09-29 00:06:04

Number: 3 Author: Oliver Johnson Subject: Inserted Text Date: 2021-10-16 18:32:36

the

Author: sterg Date: 2021-11-20 13:57:31

Accepted set by sterg

Number: 4 Author: Oliver Johnson Subject: Inserted Text Date: 2021-10-16 18:30:14

the local correlation lengths

Author: sterg Date: 2021-11-20 13:57:47

Accepted set by sterg

Number: 5 Author: Oliver Johnson Subject: Highlight Date: 2021-10-16 18:35:38

I think we should phrase this differently. In previous work (e.g. the Olmsted papers), they showed that GB energy was poorly correlated with Sigma. So perhaps what we should say here is "This may explain why previous researchers found Sigma to be a poor predictor of GB energy [Olmsted papers]."

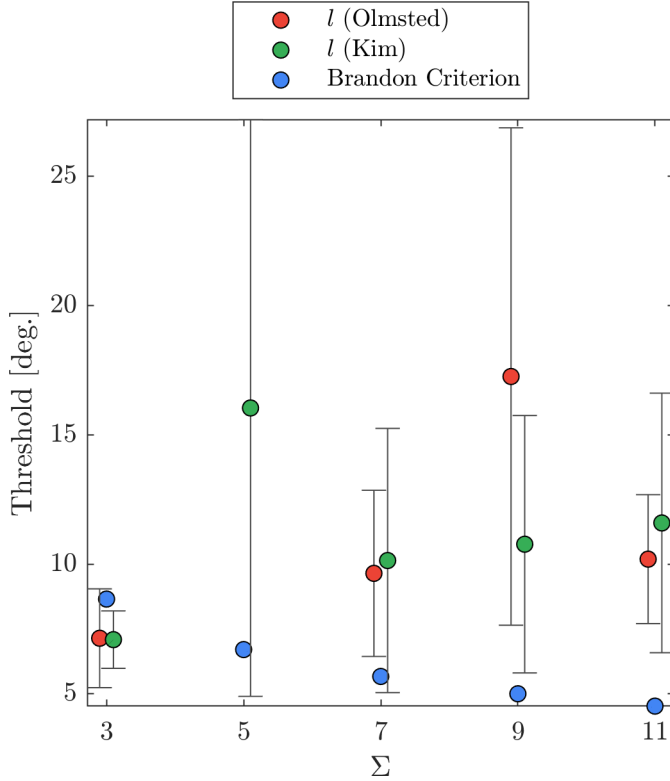**Author: sterg Date: 2021-11-20 13:59:16**

Accepted set by sterg

Figure 2: Comparison of the Brandon criterion [43] for low-Σ GBs (in units of °) to the local correlation lengths obtained in this work (in units of d_Ω°). Conversion to consistent units would unnecessarily separate the data without changing the trends, and since we are most interested in the uniqueness of the trends rather than the numerical values themselves, we leave the values in their native units. Error bars indicate 95% confidence intervals. There is no marker for the $\Sigma 5$ GB in the Ni (Olmsted) dataset because of the failure of the fit discussed above.

3.2.3. Density and Distribution of Points

Figure 4 illustrates how the VFZ-GBO average NN distance varies with the cardinality of the set (i.e. number of random VFZ-GBOs in the set). The average NN distance (over approximately 70 trials) of set sizes between 100 and 50 000 between $(10.7175 \pm 0.3684)°$ and $(2.6479 \pm 0.2254)°$, respectively. Additionally, a curve fit is given in the range of 388 to 50 000. For a specific 50 000 VFZ-GBO set, the NN GBO distance is $(2.870\,20 \pm 0.691\,17)°$ (Figure 3a) while the average 100-th NN distance is within 10° (Figure 3b).

How large is the local region of influence for a 5DOF interpolation model? These NN distances give context to this question: if 50 000 GBs are randomly sampled, then on average, these GBs are $(2.6479 \pm 0.2254)°$ apart from their NN. In the context of correlation lengths of $\sim 7-8°$ (Section 3.2.1), a random set of at least $\sim 10\,000$ GBs is required to reach a reasonable goal of half the correlation length (i.e. to have a reasonably performing structure-property model). By contrast, A set of 388 randomly sampled GBs will yield average NN distances of $\sim 7.5°$. The fact that this is approximately the same as the correlation lengths derived in this work suggest that this size of dataset is far too small. We demonstrate this via an example.

Assume the property of a particular GB which has been randomly sampled has never been measured before. What is available, however, is a set of property measurements for 388 other GBs. Our model suggests that GB which are further apart than $\sim 7.5°$ can't reveal much information about each other. So, how many GBs are in this local region of influence to be able to predict the property for the GB of interest? The answer is only one GB. To make matters worse, this GB happens to be at the borderline distance of containing relevant information (closer is better).

Correlation lengths can vary locally and are subject to some interpretation (Section 3.2.2); however, it's clear that without knowing more, the property prediction of the GB of interest might be OK, or it might be completely off. The BRK [34] model (which uses 388 GBs) works around this limitation by imposing prior information: a "scaffold-

Number: 1 Author: Oliver Johnson Subject: Highlight Date: 2021-10-18 10:32:08

I have had a nagging question about this: The procedure you are using to generate the VFZO set involves sampling misorientations cubochorically, and normals from the uniform distribution on the sphere. It's not clear to me that this will result in a uniform distribution of GBOs. One can produce uniform GBOs by the same simple procedure you use to sample points uniformly on the 2-sphere (i.e. generate 8 normally distributed numbers, then normalize, this will produce an octonion that is sampled from the uniform distribution on the 7-sphere). There may be a good reason to generate the VFZO set using your 2-step procedure, but I do wonder whether uniformly distributed GBOs would have the same NN distance distribution that you are reporting here for your 2-step procedure. I think it should be easy to get the NN distribution for the uniform GBOs, so I think it is worth testing this (you could plot the results for both methods in Figs. 3-4). If it turns out the results are identical, then you don't need to show both results, but I would still state that the distributions are identical.

Author: sterg Date: 2021-11-20 14:03:50

Accepted set by sterg

Author: sterg Subject: sterg Date: 2021-11-20 14:01:34

I already verified this as you mentioned, hence not needing to show both results. I'll mention that the distributions appear to be identical.

Number: 2 Author: Oliver Johnson Subject: Highlight Date: 2021-10-18 10:38:11

I'm not convinced of the value of this model. See my comment below in the figure caption.

Number: 3 Author: Oliver Johnson Subject: Highlight Date: 2021-10-18 10:41:16

I really like the general content of this discussion, but I think it should be put in the context of answering the question of how many GBs I need to sample the space. I have made a suggestion of how to revise this discussion to couch it in these terms in the outline I provided in my email.

Author: sterg Subject: sterg Date: 2021-11-20 21:09:54

I added the question as a 2nd sentence

Number: 4 Author: sterg Subject: Highlight Date: 2021-09-29 00:06:10

Number: 5 Author: Oliver Johnson Subject: Inserted Text Date: 2021-10-16 18:39:08

s

Author: sterg Date: 2021-11-20 14:05:11

Accepted set by sterg

Number: 6 Author: Oliver Johnson Subject: Highlight Date: 2021-10-18 10:46:44

I think this is an important enough point that you ought to also plot this as a function of the VFZO set cardinality, to show how many NNs will be within 0.5, 1, and 2 correlation lengths. This information will inform the person who wants to use this for practical purposes, to know how their decision about the VFZO set size will influence how many NNs will inform the predictions of a given prediction point.

Author: sterg Date: 2021-11-20 14:05:34

Rejected set by sterg

Author: sterg Subject: sterg Date: 2021-11-22 15:50:54

This could be done using the large distance matrix via figshare. gpr388-olmsted-dist.mat
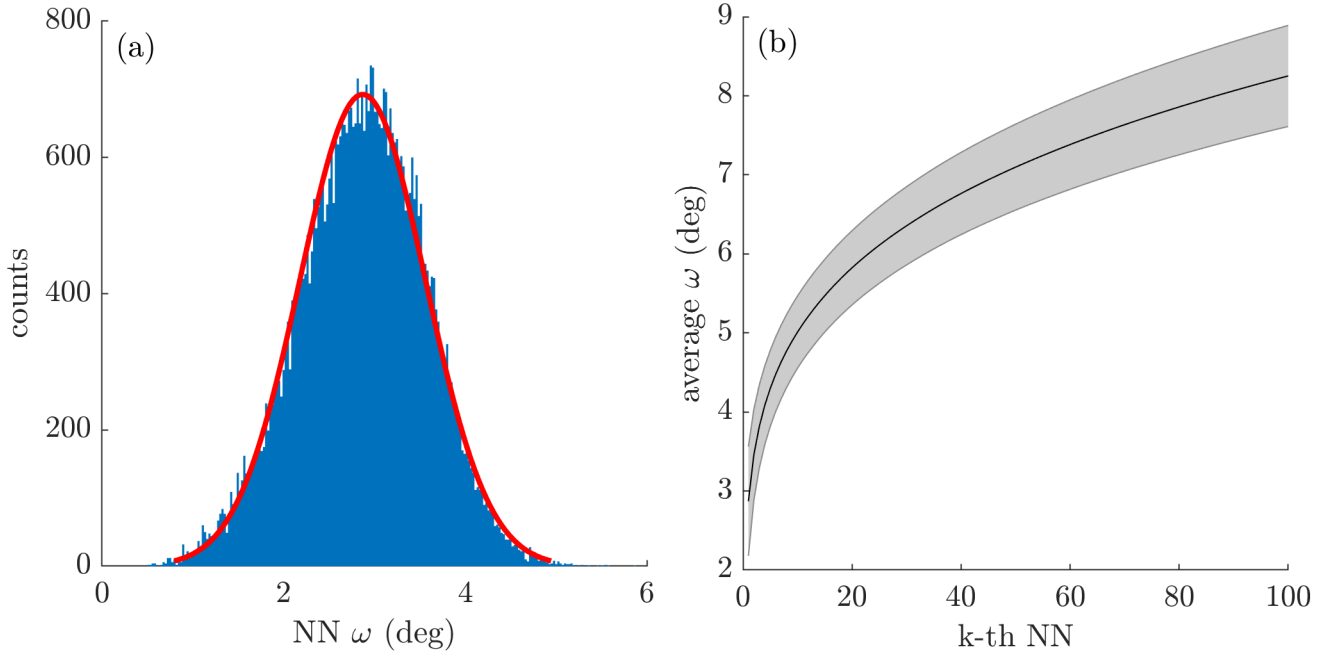https://figshare.com/s/0ac714343cac8b2a5935

Figure 3: (a) Histogram of NN GBO distances (ω) in a VFZ-GBO set of 50 000 points. The average NN distance was $(2.870\,20 \pm 0.691\,17)°$. (b) The average k-th nearest neighbor distances demonstrate that NNs up to the $\sim$10-th NN fall within a tolerance of $\sim$5°. Standard deviation uncertainty bars using approximately 10 trial runs are also shown.

ing" of GBs at high-symmetry cusps is used, and it is assumed that the function varies monotonically from point to point everywhere else. For a well-studied property (GBE) and a well-studied material system (FCC), this domain knowledge can be successfully baked in. But what if the property or material system is not so well understood? These are cases where the VFZ framework is especially useful.

3.3. Uncertainty of Noisy Molecular Statics Simulations

We analyze implications of results for a Gaussian process regression mixture model developed for a noisy, Fe molecular statics simulation dataset. We find that:

- the model error is on par with the intrinsic uncertainty of the data (Section 3.3.1)

- the predictions likely exhibit overprediction bias relative to the true minimum for a given GB (Section 3.3.2)

- future availability of multiple metastable state GBEs is anticipated to greatly improve the model performance (Section 3.3.3)
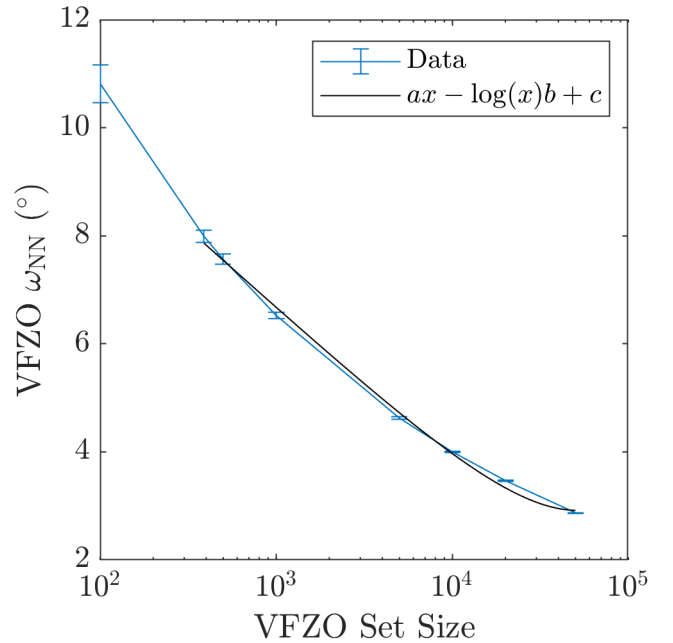

Figure 4: NN VFZ-GBO (ω_{NN}) distances (°) versus VFZ-GBO set size out of 70-80 random VFZ-GBO sets per set size and a fit to $ax - \log(x)b + c$ where $a = 2.5025 \times 10^{-5}$, $b = 1.27396$, $c = 15.4499$, x represents set size, and $388 \leq x \leq 50000$.

Number: 1 Author: Oliver Johnson Subject: Highlight Date: 2021-10-18 14:16:15

I think this section should be moved to supplementary material (and modified to give the reader more context and interpretation; right now I don't think the reader can appreciate what is being discussed or why it is relevant). Earlier in the manuscript, In the section where you introduce the GPR model, I would recommend pointing the reader to this content in the supplementary material for further reading about the fidelity of the model (if there are key points from this content that you think the reader needs to know in the main document you can summarize them there).

Author: sterg Date: 2021-11-20 21:33:41

Deferred set by sterg

Author: sterg Subject: sterg Date: 2021-11-22 15:51:06

I moved the subsections to supp and added some clarifying statements that should provide more context to the reader.

Number: 2 Author: Oliver Johnson Subject: Highlight Date: 2021-10-18 10:37:11

I'm not convinced that this model is useful. based on the trajectory, it looks like it would deviate significantly from the data below 4e2 (and why the line is truncated is not clear to me, since it seems to occlude this fact). I think the reader will get just as good of an estimate by eyeballing the data as by using this model.

Author: sterg Subject: sterg Date: 2021-11-22 15:52:18

See caption, the fit was between 388 and 50000. If you have time to modify the figure, I'm fine with removing the fit. However, I'm kind of at the point where unless figures have a glaring error, I'm inclined to leave them as-is due to time constraints

3.3.1. Intrinsic Uncertainty

We estimate the intrinsic uncertainty of an Fe $0\,K$ molecular statics simulation dataset to be $0.065\,29\,J\,m^{-2}$ and $0.061\,90\,J\,m^{-2}$ depending on whether root mean square error or mean absolute error estimates are used, respectively. Minimum and maximum error was $-0.2625\,J\,m^{-2}$ and $0.2625\,J\,m^{-2}$, respectively.

First, because only a single metastable state was used for each GBE simulation, both the training and validation data are subject to noise, consistent with a wide lateral spread of predictions and the intrinsic uncertainty estimation (Figure S5). The Fe simulation dataset Gaussian process regression mixture model gives lower root mean square error $(0.055\,035\,J\,m^{-2})$ and mean absolute error $(0.039\,185\,J\,m^{-2})$ than the uncertainty estimates. This indicates that the uncertainty itself is somewhat overestimated[6]. The fact that both model and uncertainty metrics are relatively close and the prediction [38] and uncertainty parity plots (Figure S5b) are similar suggests that the model is performing well. It also suggests that further improvements in the model relative to the "true" values will be "hidden", i.e. they will probably not manifest as lower root mean square error or mean absolute error nor as more tightly distributed parity plots despite improving performance "behind the scenes".

3.3.2. Overprediction Bias

Next, given the theoretical existence of a true minimum GBE for a given GB, the predictions which were based on metastable GBEs can be assumed to have an overprediction bias relative to the true minimum. On average, we expect this overprediction bias relative to the true minimum GBE (rather than the most likely metastable state) may be on the order of a few hundred $mJ\,m^{-2}$ and may vary as a function of true minimum GBE. In other words, the model obtained is probably an estimate

of the most likely metastable GBE rather than the true minimum GBE. This is akin to saying that we obtain from this data a model that approximates the non-equilibrium, Stillinger quenched red curve of Figure 4(c1) in [45], not the minimum GBE blue curve of the same chart. See [45] for an in-depth treatment of equilibrium and metastable GBE.

3.3.3. Improving on Existing Datasets

Finally, datasets where multiple metastable GBEs (e.g. 3-10 repeats) are provided for each GB will likely greatly improve the performance of the GPR model in predicting either the most likely metastable GBE (when all GBEs are considered) or the true minimum GBE (when only the minimum GBE is considered for each GB) and may even negate the need for a Gaussian process regression mixture approach. Thus, it is suggested that, where feasible, future large-scale GB bicrystal simulation studies report all property data for repeated trial runs rather than a single trial run or a single value from a set of trial runs. Ideally, data for the three additional microscopic degrees of freedom for GBs (which falls into the category of epistemic uncertainty in this work) would also be included. We believe it is likely that minimum energy paths (i.e. paths of steepest descent) in the GBE landscape depend on both macroscopic and microscopic degrees of freedom (in total, 8DOF) and could offer a more holistic view of GB behavior that better mimics and explains experimental grain growth observations. Indeed, it has been experimentally observed that at least some GB migration mechanisms involve structural transformations between equilibrium GBs via metastable states [46].

3.4. 5DOF Paths and GB Behavior

~~We analyze direct connections between low Sigma GBs of interest to the materials science community. U~~sing the set of 388 GBs defined by Olmsted et al. [37], we choose the $\Sigma 5$, $\Sigma 7$, $\Sigma 9$, and $\Sigma 11$ GBs with the lowest GBE ~~out of the Olm-~~

---

[6]The prediction error of a model typically cannot be less than the noise of the prediction data of a model even if the model is estimating the true prediction values with better accuracy than the noise (which is very possible and even expected with GPR models when the noise in the input data is approximately Gaussian).

Number: 1　Author: Oliver Johnson　Subject: Highlight　Date: 2021-10-18 14:09:49

I would remove this. It's not actually 8D, it is 5D for the crystallography and 3D for every atom in the GB so it's really 5D+N*3D. I think it is enough to just note that it would be good to have both macroscopic and microscopic DOFs reported.

Author: sterg　Date: 2021-11-20 14:15:08

Rejected set by sterg

Author: sterg　Subject: sterg　Date: 2021-11-22 15:55:30

My understanding of the microscopic DOFs as described in the cited paper is not on a per-atom basis, but using a construct related to CSL theory and 3 parameters. It's also fairly related to some of the work Lydia has been doing. I think I see where you're coming from in the perspective of being able to represent GBs in the context of the atomic DOFs (generally an intractable problem), but I think that's a separate issue from what I'm talking about. Do you have a suggestion on how to make it clear that I'm not referring to Cartesian atomic position DOFs?

Number: 2　Author: Oliver Johnson　Subject: Inserted Text　Date: 2021-10-20 10:02:42

Using a VFZO set containing X points, we constructed 5D structure-property models for GB energy (i.e. GB energy landscapes), with quantified uncertainty, for the Ni and Fe datasets previously mentioned using the GPR method of the interp5DOF toolbox (Section Y) [REF]. The performance and fidelity of these models were discussed in previous work [REF]. Here we use these models to explore geodesic paths through the resulting GB energy landscapes, between certain low-Σ GBs, and discuss the implications for microstructure evolution.

Author: sterg　Date: 2021-11-20 14:24:53

Accepted set by sterg

Number: 3　Author: Oliver Johnson　Subject: Inserted Text　Date: 2021-10-18 16:54:19

New Paragraph

Author: sterg　Date: 2021-11-20 14:24:16

Accepted set by sterg

Number: 4　Author: Oliver Johnson　Subject: Cross-Out　Date: 2021-10-18 16:55:34

Author: sterg　Date: 2021-11-20 14:24:23

Accepted set by sterg

sted et al. [37] Ni dataset[7] we then visualize ~~direct~~ paths as "tunnel" plots[8] in a VFZ between each of these and the global minimum $\Sigma 3$ coherent-twin GB (~~Figure 5~~). ~~This is performed for both the BRK and VFZ-GPR models.~~ ~~Likewise, we perform GPR for the Fe simulation dataset for GBs of interest and also visualize these using tunnel plots (Figure 6).~~

For the GPR model trained on the Ni Olmsted et al. [37] dataset, we observe that ~~$\Sigma 3$ and $\Sigma 7$ are connected (i.e. no activation energy barrier)~~ in Figure 5b while the ~~direct~~ paths between $\Sigma 3$ and $\Sigma 5$, $\Sigma 9$, $\Sigma 11$ are separated by energy barriers (Figures 5a, 5c and 5d). This indicates that for grain growth systems governed by GBE, ~~$\Sigma 7$ GBs may be more likely to transform into $\Sigma 3$ coherent-twin GBs, whereas if~~ $\Sigma 5$, $\Sigma 9$, or $\Sigma 11$ transform into $\Sigma 3$ coherent-twins, ~~this will need to occur either by overcoming the activation energy or via paths through 5DOF space other than the ones shown.~~ We note that while the GBE path is direct with respect to crystallographic distances, the actual rotation of a grain is not a barrier-free transformation and requires energy to induce the crystallographic change [REF]. Thus, when grain rotation and other non-spontaneous atomic transitions are involved, the true energy path can be thought of as a signal of local energy barriers associated with atomic movement superimposed on the 5DOF paths shown.

~~The~~ substantial vertical spread of the 1st-6th NNs (colored, variable-size circles) are consistent with the fact that many points have distances on the order of $8-14°$ relative to the ~~direct~~ path between the two GBs. For perspective, these distances are greater than the global correlation length of 7.4° (Section 3.2.1), for which a higher proportion of distant points from the models are also distant crystallographically (small red and yellow circles). In other words, ~~the local region of influence along the path is sparsely populated.~~

---

[7]The IDs that correspond to each of the low-Σ GBs used for path visualization for the Olmsted et al. [37] and Kim et al. [41] datasets are given in Table 1 and Table 2, respectively.

[8]Similar to traveling through a 1D tunnel while also looking at nearby points in the region close to the line in all directions.

For the GPR model trained on the Fe Kim et al. [41] dataset, the 1st NN path and the GPR path have substantial qualitative deviation from each other for $\overline{\Sigma 3, \Sigma 9}$ and $\overline{\Sigma 3, \Sigma 11}$ paths. While the GPR model suggests that $\overline{\Sigma 3, \Sigma 11}$ are connected, the 1st NN path suggests that there are local minima that are overestimated. Likewise, the 1st-6th NNs (colored, variable-size circles) for $\overline{\Sigma 3, \Sigma 5}$ and $\overline{\Sigma 3, \Sigma 7}$ follow the general trend of the models. We observe that all NNs distances up to the 6th NN are less than $\sim 6°$ from the paths of interest; this in turn is less than the global correlation length of 8.3° from Section 3.2.1 which indicates that local regions of influence along these paths in the Fe dataset are densely populated.

By contrast, some NNs have vastly different energies compared with the GPR model. This is because the noise of the simulated GBEs (Section 3.3.1) precludes the GPR model from simultaneously resolving sharp transitions and maintaining smooth behavior elsewhere. This is one of the drawbacks of using a noisy dataset in a diverse space (some regions with sharp cusps, others with shallow hills and valleys) with a global smoothness (i.e. correlation) length.

3.5. Potential for Numerical Derivatives

GB path visualizations in the VFZ framework suggest the ability to estimate numerical derivatives or gradients of GB properties without being restricted to a GB subspace (e.g. misorientation fundamental zone or boundary plane fundamental zone) which can be a useful mathematical construct for the GB community. For example, steepest descent paths and all local GBE minima can be estimated and used in grain growth simulations.

Because distance overestimation exist in the VFZ framework, use of ensembled VFZ-GBO interpolation or data augmentation may be necessary to mitigate discontinuity artifacts when crossing the exterior of a VFZ as discussed in Section 2.1. Alternatively, the "excess" points in a gridded sampling can act as a type of data augmentation and help to address this issue. We plan to explore these topics in future work. Section S4 contains further discussion of a gridded sampling approach.

Page:11

Number: 1 Author: Oliver Johnson Subject: Cross-Out Date: 2021-10-18 16:55:34

Author: sterg Date: 2021-11-20 14:25:28

Accepted set by sterg

Number: 2 Author: Oliver Johnson Subject: Highlight Date: 2021-10-20 10:11:59

I would move this before the discussion about the paths since it is more of a description of the results, whereas the paragraph about the downhill energy path is more of a discussion of the results so it should come afterwards.

Author: sterg Date: 2021-11-20 14:48:41

Accepted set by sterg

Number: 3 Author: Oliver Johnson Subject: Inserted Text Date: 2021-10-20 10:04:03

geodesic

Author: sterg Date: 2021-11-20 14:25:20

Accepted set by sterg

Number: 4 Author: Oliver Johnson Subject: Inserted Text Date: 2021-10-18 16:56:54

Figure 5 and 6

Author: sterg Date: 2021-11-20 14:26:23

Accepted set by sterg

Number: 5 Author: Oliver Johnson Subject: Cross-Out Date: 2021-10-18 16:57:00

Author: sterg Date: 2021-11-20 14:25:51

Accepted set by sterg

Number: 6 Author: Oliver Johnson Subject: Inserted Text Date: 2021-10-20 10:23:13

These tunnel plots show the predicted energy along the respective geodesic path, together with the 1st-6th nearest data points along the path so that one can compare the predictions to datapoints in the vicinity of the path. For the Ni dataset, we also plot the predictions of the BRK model along the same trajectories for comparison.

Author: sterg Date: 2021-11-20 14:31:56

Accepted set by sterg

Number: 7 Author: Oliver Johnson Subject: Inserted Text Date: 2021-10-18 17:03:50

the path from the $\Sigma 7$ GB to the $\Sigma 3$ GB is strictly downhill in energy

Author: sterg Date: 2021-11-20 14:28:33

Accepted set by sterg

Number: 8 Author: Oliver Johnson Subject: Inserted Text Date: 2021-10-20 10:04:16

geodesic

Author: sterg Date: 2021-11-20 14:26:28

Accepted set by sterg

Number: 9 Author: Oliver Johnson Subject: Cross-Out Date: 2021-10-18 17:09:01

Author: sterg Date: 2021-11-20 14:30:13

Accepted set by sterg

Number: 10 Author: Oliver Johnson Subject: Inserted Text Date: 2021-10-18 17:08:04

, it is possible that a $\Sigma 7$ GB may transform into a $\Sigma 3$ GB, if the mechanisms required to traverse this trajectory in the GB energy landscape

~~sted et al. [37] Ni dataset~~[7]; we then visualize ~~direct~~ paths as "tunnel" plots[8] in a VFZ between each of these and the global minimum $\Sigma3$ coherent-twin GB (~~Figure 5~~). ~~This is performed for both the BRK and VFZ-GPR models.~~ ~~Likewise, we perform GPR for the Fe simulation dataset for GBs of interest and also visualize these using tunnel plots (Figure 6).~~

For the GPR model trained on the Ni Olmsted et al. [37] dataset, we observe that ~~$\Sigma3$ and $\Sigma7$ are connected (i.e. no activation energy barrier)~~ in Figure 5b while the ~~direct~~ paths between $\Sigma3$ and $\Sigma5$, $\Sigma9$, $\Sigma11$ are separated by energy barriers (Figures 5a, 5c and 5d). This indicates that for grain growth systems governed by GBE, ~~$\Sigma7$ GBs may be more likely to transform into $\Sigma3$ coherent-twin GBs,~~ ~~whereas if~~ $\Sigma5$, $\Sigma9$, or $\Sigma11$ transform into $\Sigma3$ coherent-twins, ~~this will need to occur either by overcoming the activation energy or via paths through 5DOF space other than the ones shown.~~ We note that while the GBE path is direct with respect to crystallographic distances, the actual rotation of a grain is not a barrier-free transformation and requires energy to induce the crystallographic change [REF]. Thus, when grain rotation and other non-spontaneous atomic transitions are involved, the true energy path can be thought of as a signal of local energy barriers associated with atomic movement superimposed on the 5DOF paths shown.

~~The~~ substantial vertical spread of the 1st-6th NNs (colored, variable-size circles) are consistent with the fact that many points have distances on the order of $8-14°$ relative to the ~~direct~~ path between the two GBs. For perspective, these distances are greater than the global correlation length of $7.4°$ (Section 3.2.1), for which a higher proportion of distant points from the models are also distant crystallographically (small red and yellow circles). In other words, ~~the local region of influence along the path is sparsely populated.~~

---

[7]The IDs that correspond to each of the low-Σ GBs used for path visualization for the Olmsted et al. [37] and Kim et al. [41] datasets are given in Table 1 and Table 2, respectively.

[8]Similar to traveling through a 1D tunnel while also looking at nearby points in the region close to the line in all directions.

For the GPR model trained on the Fe Kim et al. [41] dataset, the 1st NN path and the GPR path have substantial qualitative deviation from each other for $\overline{\Sigma3, \Sigma9}$ and $\overline{\Sigma3, \Sigma11}$ paths. While the GPR model suggests that $\overline{\Sigma3, \Sigma11}$ are connected, the 1st NN path suggests that there are local minima that are overestimated. Likewise, the 1st-6th NNs (colored, variable-size circles) for $\overline{\Sigma3, \Sigma5}$ and $\overline{\Sigma3, \Sigma7}$ follow the general trend of the models. We observe that all NNs distances up to the 6th NN are less than $\sim6°$ from the paths of interest; this in turn is less than the global correlation length of $8.3°$ from Section 3.2.1 which indicates that local regions of influence along these paths in the Fe dataset are densely populated.

By contrast, some NNs have vastly different energies compared with the GPR model. This is because the noise of the simulated GBEs (Section 3.3.1) precludes the GPR model from simultaneously resolving sharp transitions and maintaining smooth behavior elsewhere. This is one of the drawbacks of using a noisy dataset in a diverse space (some regions with sharp cusps, others with shallow hills and valleys) with a global smoothness (i.e. correlation) length.

3.5. Potential for Numerical Derivatives

GB path visualizations in the VFZ framework suggest the ability to estimate numerical derivatives or gradients of GB properties without being restricted to a GB subspace (e.g. misorientation fundamental zone or boundary plane fundamental zone) which can be a useful mathematical construct for the GB community. For example, steepest descent paths and all local GBE minima can be estimated and used in grain growth simulations.

Because distance overestimation exist in the VFZ framework, use of ensembled VFZ-GBO interpolation or data augmentation may be necessary to mitigate discontinuity artifacts when crossing the exterior of a VFZ as discussed in Section 2.1. Alternatively, the "excess" points in a gridded sampling can act as a type of data augmentation and help to address this issue. We plan to explore these topics in future work. Section S4 contains further discussion of a gridded sampling approach.

re available (e.g. if both grain rotation and plane reorientation are possible, such as during high-temperature plastic deformation, recrystallization, or in nanocrystalline materials).

Author: sterg Date: 2021-11-20 14:32:18

Accepted set by sterg

Number: 11 Author: Oliver Johnson Subject: Inserted Text Date: 2021-10-18 17:09:27

GB to

Author: sterg Date: 2021-11-20 14:47:42

Accepted set by sterg

Number: 12 Author: Oliver Johnson Subject: Inserted Text Date: 2021-10-18 17:09:19

In contrast, for a

Author: sterg Date: 2021-11-20 14:34:15

Accepted set by sterg

Number: 13 Author: Oliver Johnson Subject: Inserted Text Date: 2021-10-18 17:10:50

either an energy barrier must be overcome, or a different (perhaps more circuitous) trajectory through the 5DOF space must be taken.

Author: sterg Date: 2021-11-20 14:40:17

Accepted set by sterg

Number: 14 Author: Oliver Johnson Subject: Highlight Date: 2021-10-18 17:17:08

I think perhaps the way we should phrase this is in terms of driving force and activation energy. Any transformation will depend on both the driving force AND the activation energy (e.g. near the transformation temperature nucleation will be slow because the driving force is small, even though diffusion is fast, and at large undercoolings even though the driving force is large nucleation will again be slow but this time because diffusion is slow). The GB energy landscape tells us about the driving force for such a GB crystallographic transformation (there is a lower-energy state that I can get to), but it does not describe the activation energies of the atomic processes by which that transformation is achieved. Both pieces of information are necessary to predict whether and how fast a given crystallographic transformation might occur.

Author: sterg Date: 2021-11-20 14:46:36

Accepted set by sterg

Author: sterg Subject: sterg Date: 2021-11-20 14:46:50

Added "from another perspective, ..."

Number: 15 Author: Oliver Johnson Subject: Highlight Date: 2021-10-18 16:58:48

missing reference

Author: sterg Date: 2021-11-20 14:41:32

Deferred set by sterg

Author: sterg Subject: sterg Date: 2021-11-22 15:56:46

I can either remove or add a REF if you find one. This was a suggestion from Dr. Homer

---

Number: 16 Author: Oliver Johnson Subject: Highlight Date: 2021-10-20 10:49:23

I'm not sure I totally understand what you're suggesting here. I get that having a continuous function allows me to find minima; however, I don't understand how I could use that in grain growth simulations. In a grain growth simulation the microstructure is what it is. It's evolution is influenced by the underlying GBE landscape, but I don't impose minima on it... Perhaps you're saying that I can interpret the results of anisotropic grain growth simulations in terms of a knowledge of the minima in the GBE landscape?

Author: sterg Date: 2021-11-20 14:50:01

Accepted set by sterg

Author: sterg Subject: sterg Date: 2021-11-20 14:49:58

removed the reference to local minima

Number: 17 Author: Oliver Johnson Subject: Inserted Text Date: 2021-10-20 10:25:59

For the GPR model trained on the Ni data (Figure 5), the

Author: sterg Date: 2021-11-20 14:35:43

Accepted set by sterg

sted et al. [37] Ni dataset[7]; we then visualize ~~direct~~ paths as "tunnel" plots[8] in a VFZ between each of these and the global minimum $\Sigma 3$ coherent-twin GB (~~Figure 5~~). ~~This is performed for both the BRK and VFZ-GPR models.~~ ~~Likewise, we perform GPR for the Fe simulation dataset for GBs of interest and also visualize these using tunnel plots (Figure 6).~~

For the GPR model trained on the Ni Olmsted et al. [37] dataset, we observe that ~~$\Sigma 3$ and $\Sigma 7$ are connected (i.e. no activation energy barrier)~~ in Figure 5b while the ~~direct~~ paths between $\Sigma 3$ and $\Sigma 5$, $\Sigma 9$, $\Sigma 11$ are separated by energy barriers (Figures 5a, 5c and 5d). This indicates that for grain growth systems governed by GBE, ~~$\Sigma 7$ GBs may be more likely to transform into $\Sigma 3$ coherent-twin GBs,~~ ~~whereas if~~ $\Sigma 5$, $\Sigma 9$, or $\Sigma 11$ transform into $\Sigma 3$ coherent-twins, ~~this will need to occur either by overcoming the activation energy or via paths through 5DOF space other than the ones shown.~~ We note that while the GBE path is direct with respect to crystallographic distances, the actual rotation of a grain is not a barrier-free transformation and requires energy to induce the crystallographic change [REF]. Thus, when grain rotation and other non-spontaneous atomic transitions are involved, the true energy path can be thought of as a signal of local energy barriers associated with atomic movement superimposed on the 5DOF paths shown.

~~The~~ substantial vertical spread of the 1st-6th NNs (colored, variable-size circles) are consistent with the fact that many points have distances on the order of $8-14°$ relative to the ~~direct~~ path between the two GBs. For perspective, these distances are greater than the global correlation length of 7.4° (Section 3.2.1), for which a higher proportion of distant points from the models are also distant crystallographically (small red and yellow circles). In other words, ~~the local region of influence along the path is sparsely populated.~~

For the GPR model trained on the Fe Kim et al. [41] dataset, the 1st NN path and the GPR path have substantial qualitative deviation from each other for $\overline{\Sigma 3, \Sigma 9}$ and $\overline{\Sigma 3, \Sigma 11}$ paths. While the GPR model suggests that $\overline{\Sigma 3, \Sigma 11}$ are connected, the 1st NN path suggests that there are local minima that are overestimated. Likewise, the 1st-6th NNs (colored, variable-size circles) for $\overline{\Sigma 3, \Sigma 5}$ and $\overline{\Sigma 3, \Sigma 7}$ follow the general trend of the models. We observe that all NNs distances up to the 6th NN are less than $\sim 6°$ from the paths of interest; this in turn is less than the global correlation length of 8.3° from Section 3.2.1 which indicates that local regions of influence along these paths in the Fe dataset are densely populated.

By contrast, some NNs have vastly different energies compared with the GPR model. This is because the noise of the simulated GBEs (Section 3.3.1) precludes the GPR model from simultaneously resolving sharp transitions and maintaining smooth behavior elsewhere. This is one of the drawbacks of using a noisy dataset in a diverse space (some regions with sharp cusps, others with shallow hills and valleys) with a global smoothness (i.e. correlation) length.

3.5. Potential for Numerical Derivatives

GB path visualizations in the VFZ framework suggest the ability to estimate numerical derivatives or gradients of GB properties without being restricted to a GB subspace (e.g. misorientation fundamental zone or boundary plane fundamental zone) which can be a useful mathematical construct for the GB community. For example, steepest descent paths and all local GBE minima can be estimated and used in grain growth simulations.

Because distance overestimation exist in the VFZ framework, use of ensembled VFZ-GBO interpolation or data augmentation may be necessary to mitigate discontinuity artifacts when crossing the exterior of a VFZ as discussed in Section 2.1. Alternatively, the "excess" points in a gridded sampling can act as a type of data augmentation and help to address this issue. We plan to explore these topics in future work. Section S4 contains further discussion of a gridded sampling approach.

---

[7]The IDs that correspond to each of the low-Σ GBs used for path visualization for the Olmsted et al. [37] and Kim et al. [41] datasets are given in Table 1 and Table 2, respectively.

[8]Similar to traveling through a 1D tunnel while also looking at nearby points in the region close to the line in all directions.

Number: 18 Author: Oliver Johnson Subject: Inserted Text Date: 2021-10-20 10:04:54

geodesic

Author: sterg Date: 2021-11-20 14:36:00

Accepted set by sterg

Number: 19 Author: Oliver Johnson Subject: Inserted Text Date: 2021-10-20 10:45:06

s

Author: sterg Date: 2021-11-20 14:39:37

Accepted set by sterg

Number: 20 Author: Oliver Johnson Subject: Inserted Text Date: 2021-10-20 10:45:22

standard

Author: sterg Date: 2021-11-20 14:47:33

Accepted set by sterg

Number: 21 Author: Oliver Johnson Subject: Inserted Text Date: 2021-10-20 10:24:53

these trajectories traverse regions of the VFZ that are sparsely populated with data.

Author: sterg Date: 2021-11-20 14:36:20

Accepted set by sterg

Number: 22 Author: Oliver Johnson Subject: Highlight Date: 2021-10-18 16:56:26

These do not point to the right tables. Table 1 and 2 are of the SVD stuff.

Author: sterg Date: 2021-11-20 14:39:13

Accepted set by sterg

Author: sterg Subject: sterg Date: 2021-11-22 15:56:56

Forgot to "renew" \thetable (supposed to point to Supp). Thanks for catching
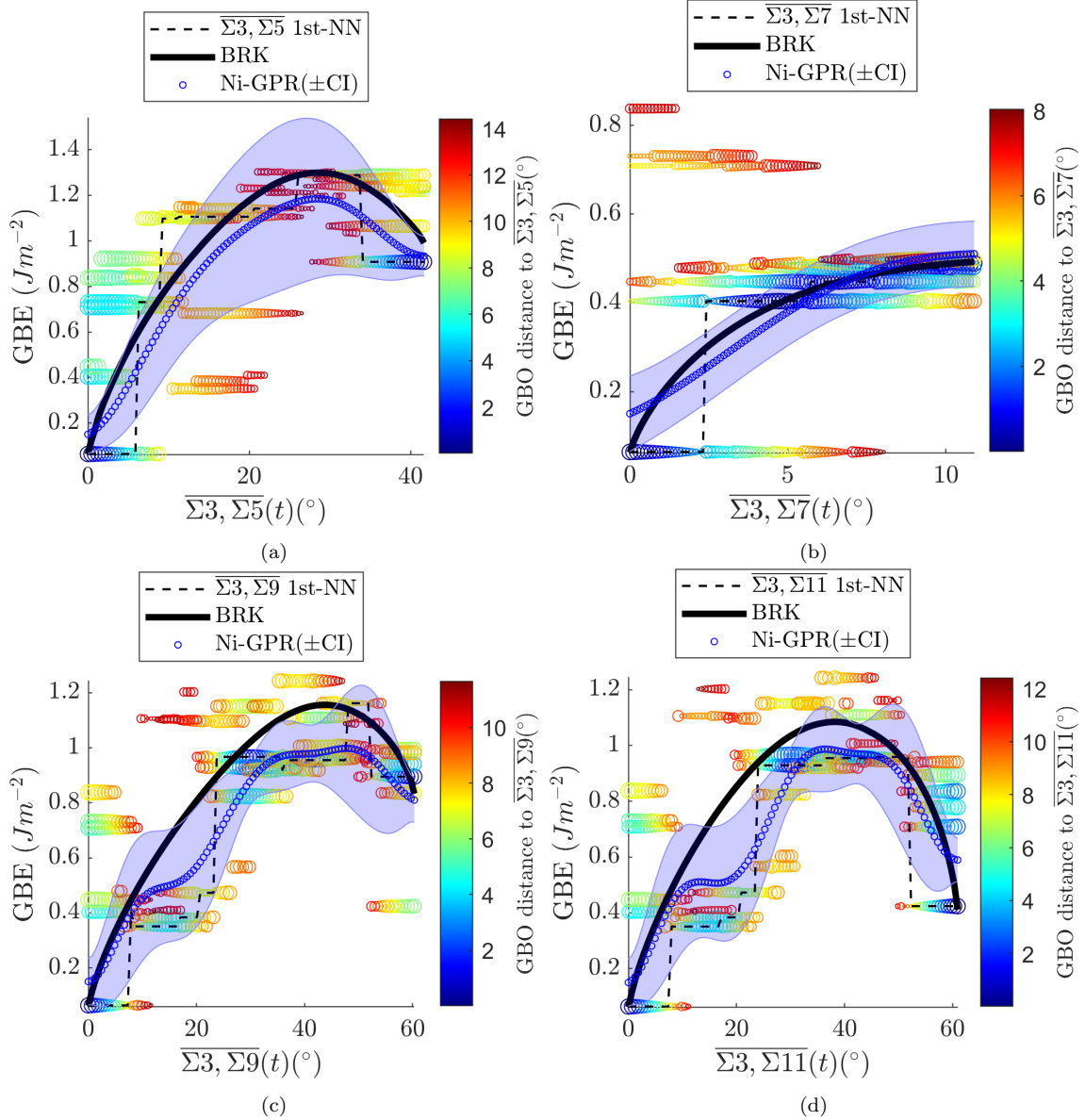
Figure 5: "Tunnel" plots of GBEs along direct paths in a VFZ between the $\Sigma 3$ coherent-twin boundary and minimum GBE (a) $\Sigma 5$, (b) $\Sigma 7$, (c) $\Sigma 9$, and (d) $\Sigma 11$ GBs within the Ni Olmsted et al. [37] dataset. GBE values are plotted for the BRK and GPR models which both used Olmsted et al. [37] as input data. 95% confidence intervals are plotted for the GPR model. A "tunnel" plot is formed by calculating up to the 6th NNs of the input data relative to the direct path formed between two GBs. The distances of the NNs relative to the arc are used to both color and size the markers on the plot; NNs which are closer to the arc are large, blue circles, whereas NNs which are further from arc are small, red circles. Additionally, the 1st NN path is plotted as a dashed line.

4. Conclusion

We applied the VFZ framework to learning more about the nature of a 5DOF FZ.

The increase of distance computation throughput and the development of a 5DOF VFZ with continuous coordinates enabled us to explore the nature of a 5DOF FZ. We found that symmetrized NN distances are Gaussian and plotted these as a function of set size. We determined the maximum principal component of a particular O_h VFZ to be $\sim 65°$.

Other point groups (in particular those which are noncentrosymmetric) may give rise to differently shaped/larger VFZs and for which the Euclidean approximation may need to be removed. It will

---

Number: 1 Author: Oliver Johnson Subject: Highlight Date: 2021-11-20 14:54:21

A few comments that I think would make these figures clearer:(1) I think that it is confusing that the GPR model uses circles just like the data. It would be better to use a solid blue line for the GPR model, and only use markers for the data.(2) I think that you need to explicitly give the GB plane for each of these GBs somewhere (not necessarily on the plot, but at least in the figure caption or in the text or something).(3) This is optional, but I would like to reiterate this suggestion with some additional explanation. In color theory there is a concept of perceptual uniformity. The jet colormap is not perceptually uniform, meaning that certain colors are more noticeable and carry more visual weight. In particular red is more noticeable than blue. The built-in parula colormap is a perceptually uniform colormap, but I think it's kind of ugly and the colors don't have an intuitive association for me. Moreover, when looking at the plot one has to interpret the meaning of the color in addition to the spatial location of the points which takes some time. I think it would be easier to digest if you instead used a colormap that associated a solid color (any color will do) with zero distance, and white with large distances, and then scaled linearly between (e.g. blue to white or red to white) that way points far from the path would appear nearly transparent which has the intuitive interpretation of not contributing much.

Author: sterg Date: 2021-11-20 14:58:43

Deferred set by sterg

Author: sterg Subject: sterg Date: 2021-11-22 18:01:53

(1) I agree with this. I'm fine if you'd like to make this edit. (2) I agree it would be better to have it in the text rather than refer to the original papers via a supporting information table, but the information is still available. Could you look up what these correspond to in the original papers? (3) I think if I saw the changes you mentioned, it would be easier to see how it affects the interpretability and what messages that are conveyed. In particular, I think something important about these figures is that it shows that the datasets are sparse and noisy for Ni and Fe, respectively.

Number: 2 Author: Oliver Johnson Subject: Inserted Text Date: 2021-10-20 10:05:13

geodesic

Author: sterg Date: 2021-11-20 14:58:51

Accepted set by sterg

Number: 3 Author: Oliver Johnson Subject: Inserted Text Date: 2021-10-20 10:05:21

geodesic

Author: sterg Date: 2021-11-20 14:58:55

Accepted set by sterg

Number: 4 Author: Oliver Johnson Subject: Highlight Date: 2021-10-20 13:42:07

You're conflating 2 separate concepts here. Distances are not Gaussian, it is the GB energy correlations that are Gaussian. Separately you found the NN distances as a function of VFZO set cardinality.

Author: sterg Date: 2021-11-20 14:59:45

Accepted set by sterg

Author: sterg Subject: sterg Date: 2021-11-22 15:57:33
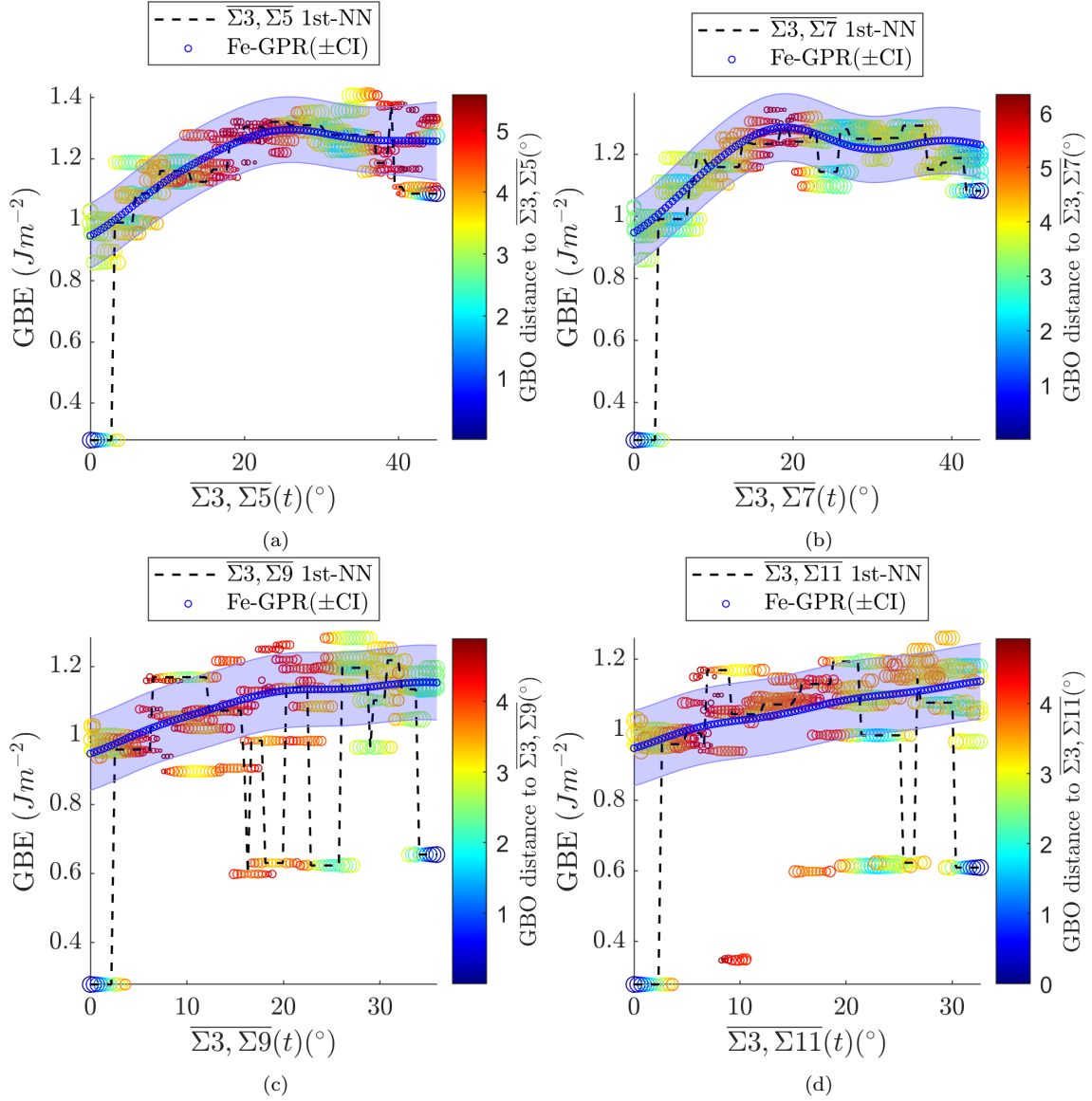
Tried to make the two distinct concepts clearer

Figure 6: GBEs along ~~direct~~ paths in a VFZ between the minimum GBE $\Sigma3$ and minimum GBE (a) $\Sigma5$, (b) $\Sigma7$, (c) $\Sigma9$, and (d) $\Sigma11$ GBs for the Fe Kim et al. [41] dataset. A GPR model trained on all 58 604 Fe Kim et al. [41] simulation datapoints was used. 95% confidence intervals are plotted for the GPR model. A "tunnel" plot is formed by calculating up to the 6th NNs of the input data relative to the ~~direct~~ path formed between two GBs, similar to traveling through a 1D tunnel while also looking at nearby points in the region close to the line in all directions. The distances of the NNs relative to the arc are used to both color and size the markers on the plot; NNs which are closer to the arc are large, blue circles, whereas NNs which are further from arc are small, red circles. Additionally, the 1st NN path is plotted as a dashed line.

be interesting to see the VFZ framework applied for other distance metrics (see Morawiec [47] for a comprehensive summary of metrics).

The interpolation errors for a Fe simulation dataset are on par with the intrinsic uncertainty of the dataset itself (Section S3). Analysis of the GPR fitting results indicates that the Ni and Fe simulation datasets have correlation lengths of 8.3° and 7.4°, respectively, but that when the Ni dataset is constrained to have low noise, the numerical correlation length drops to $\sim$1.9°. Plotting of ~~direct~~ paths between low-Sigma GBs of interest reveal that a $\Sigma7$ cusp has a monotonically decreasing path towards the coherent-twin $\Sigma3$ cusp, whereas a $\Sigma5$, $\Sigma9$, and $\Sigma11$ cusps do not necessarily share this same type of monotonically decreasing path within

13

Page:13

Number: 1 Author: Oliver Johnson Subject: Inserted Text Date: 2021-10-20 10:05:29
geodesic
Author: sterg Date: 2021-11-20 15:00:05
Accepted set by sterg

Number: 2 Author: Oliver Johnson Subject: Inserted Text Date: 2021-10-20 10:05:37
geodesic
Author: sterg Date: 2021-11-20 15:00:02
Accepted set by sterg

Number: 3 Author: Oliver Johnson Subject: Inserted Text Date: 2021-10-20 10:05:47
geodesic
Author: sterg Date: 2021-11-20 15:00:09
Accepted set by sterg

Number: 4 Author: sterg Subject: Note Date: 2021-09-29 00:28:42
Need to add a sentence or two here about the local and global correlations and the Brandon criteria.

a VFZ. We demonstrated that two cusps can be connected in 5DOF space.

In addition to its previous implementation for GB property interpolation [38], we anticipate the VFZ framework will continue to reveal important aspects of a 5DOF FZ and inform us about material behavior especially with respect to grain growth and other large scale time-dependent or iterative processes.

Acknowledgement

CRediT Statement

Sterling Baird: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization. **Oliver Johnson**: Supervision, Project administration, Funding acquisition, Conceptualization, Methodology, Formal analysis, Investigation, Writing - Original Draft, Writing - Review & Editing. **David Fullwood**: Funding acquisition, Writing - Review & Editing. **Eric Homer**: Funding acquisition, Writing - Review & Editing

Glossary

5DOF five degree-of-freedom 1–3, 6, 8, 11, 12, 14

BRK Bulatov Reed Kumar 6, 8, 11, 12

FZ fundamental zone 1, 12, 14

GB grain boundary 1–14

GBE grain boundary energy 1–4, 6, 9–13

GBO grain boundary octonion 1–6, 8, 9

GPR Gaussian process regression 4–6, 10–13

NN nearest neighbor 4, 8, 9, 11–13

VFZ Voronoi fundamental zone 1–4, 6, 7, 9, 11–14

VFZ-GBO Voronoi fundamental zone grain boundary octonion 3, 4, 8, 9, 11

References 1

[1] D. Bachurin, A. Nazarov, J. Weissmüller, Grain rotation by dislocation climb in a finite-size grain boundary, Acta Materialia 60 (2012) 7064–7077. doi:10.1016/j.actamat.2012.09.014.

[2] L. A. Barrales-Mora, J.-E. Brandenburg, D. A. Molodov, Impact of grain boundary character on grain rotation, Acta Materialia 80 (2014) 141–148. doi:10.1016/j.actamat.2014.07.049.

[3] K. E. Harris, V. V. Singh, A. H. King, GRAIN ROTATION IN THIN FILMS OF GOLD (????) 11.

[4] L. Klinger, Shape evolution by surface and interface diffusion with rigid body rotations, Acta Materialia (2011) 9.

[5] M. Upmanyu, D. Srolovitz, A. Lobkovsky, J. Warren, W. Carter, Simultaneous grain boundary migration and grain rotation, Acta Materialia 54 (2006) 1707–1719. doi:10.1016/j.actamat.2005.11.036.

[6] C. Herring, Surface tension as a motivation for sintering, in: W. E. Kingston (Ed.), The Physics of Powder Metallurgy, McGraw-Hill, New York, 1951, pp. 143–179.

[7] D. J. Rowenhorst, P. W. Voorhees, Measurements of the grain boundary energy and anisotropy in tin, Metall and Mat Trans A 36 (2005) 2127–2135. doi:10.1007/s11661-005-0333-7.

Number: 1 Author: sterg Subject: Highlight Date: 2021-09-29 00:08:14

Some of the references need to have journals added, etc. Ignore the ????

[8] J.-E. Brandenburg, L. Barrales-Mora, D. Molodov, G. Gottstein, Effect of inclination dependence of grain boundary energy on the mobility of tilt and non-tilt low-angle grain boundaries, Scripta Materialia 68 (2013) 980–983. doi:10.1016/j.scriptamat.2013.02.054.

[9] M. Furtkamp, G. Gottstein, D. Molodov, V. Semenov, L. Shvindlerman, Grain boundary migration in Fe–3.5% Si bicrystals with [001] tilt boundaries, Acta Materialia 46 (1998) 4103–4110. doi:10.1016/S1359-6454(98)00105-0.

[10] G. Gottstein, L. S. Shvindlerman, Grain Boundary Migration in Metals (????) 135.

[11] V. Ivanov, D. A. Molodov, L. S. Shvindlerman, G. Gottstein, Impact of Boundary Orientation on the Motion of Curved Grain Boundaries in Aluminum Bicrystals, MSF 467-470 (2004) 751–756. doi:10.4028/www.scientific.net/MSF.467-470.751.

[12] D. A. Molodov, B. B. Straumal, L. S. Shvindlerman, The effect of pressure on migration of ⟨001⟩ tilt grain boundaries in tin bicrystals, Scripta Metallurgica 18 (1984) 207–211. doi:10.1016/0036-9748(84)90509-X.

[13] D. A. Molodov', U. Czubayko', G. Gottstein', L. S. Shvindlerman', MOBHJTY OF <ill> TILT GRAIN BOUNDARIES IN THE VICINITY OF THE SPECIAL MISORIENTATION Z=7 IN BICRYSTALS OF PURE ALUMINIUM 32 (????) 6.

[14] A. Ramasubramaniam, V. B. Shenoy, On the evolution of faceted grain-boundary grooves by surface diffusion, Acta Materialia 53 (2005) 2943–2956. doi:10.1016/j.actamat.2005.03.013.

[15] A. Basak, A. Gupta, A two-dimensional study of coupled grain boundary motion using the level set method, Modelling Simul. Mater. Sci. Eng. 22 (2014) 055022. doi:10.1088/0965-0393/22/5/055022.

[16] N. Bernstein, The influence of geometry on grain boundary motion and rotation, Acta Materialia 56 (2008) 1106–1113. doi:10.1016/j.actamat.2007.11.002.

[17] H. Fujita, Direct Observation Subgrain-Growth of Cold-Worked Aluminium by means of Electron Microscopy, J. Phys. Soc. Jpn. 16 (1961) 397–406. doi:10.1143/JPSJ.16.397.

[18] Z. Huang, M. Bartels, R. Xu, M. Osterhoff, S. Kalbfleisch, M. Sprung, A. Suzuki, Y. Takahashi, T. N. Blanton, T. Salditt, J. Miao, Grain rotation and lattice deformation during photoinduced chemical reactions revealed by in situ X-ray nanodiffraction, Nature Mater 14 (2015) 691–695. doi:10.1038/nmat4311.

[19] H. Sharma, R. M. Huizenga, A. Bytchkov, J. Sietsma, S. E. Offerman, Observation of changing crystal orientations during grain coarsening, Acta Materialia 60 (2012) 229–237. doi:10.1016/j.actamat.2011.09.057.

[20] Z. Trautt, Y. Mishin, Capillary-driven grain boundary motion and grain rotation in a tricrystal: A molecular dynamics study, Acta Materialia 65 (2014) 19–31. doi:10.1016/j.actamat.2013.11.059.

[21] T. Uehara, N. Wakabayashi, Y. Hirabayashi, N. Ohno, An atomistic study of grain boundary stability and crystal rearrangement using molecular dynamics techniques, International Journal of Mechanical Sciences (2008) 10.

[22] M. Upmanyu, D. Srolovitz, A. Lobkovsky, J. Warren, W. Carter, Simultaneous grain boundary migration and grain rotation, Acta Materialia 54 (2006) 1707–1719. doi:10.1016/j.actamat.2005.11.036.

[23] M. Upmanyu, D. Srolovitz, A. Lobkovsky, J. Warren, W. Carter, Simultaneous grain boundary migration and grain rotation, Acta Materialia 54 (2006) 1707–1719. doi:10.1016/j.actamat.2005.11.036.

[24] S. Weissmann, On the Mechanism of Recrystallization of Aluminum, Advances in X-Ray

No Comments.

Analysis 2 (1958/ed) 47–70. doi:10.1154/S0 376030800000422.

[25] M. F. Ashby, G. H. Edward, J. Davenport, R. A. Verrall, Application of bound theorems for creeping solids and their application to large strain diffusional flow, Acta Metallurgica 26 (1978) 1379–1388. doi:10.1016/0001 -6160(78)90153-0.

[26] R. C. Gifkins, Grain-boundary sliding and its accommodation during creep and superplasticity, MTA 7 (1976) 1225–1232. doi:10.1007/ BF02656607.

[27] C. Herring, Diffusional Viscosity of a Polycrystalline Solid, Journal of Applied Physics 21 (1950) 437–445. doi:10.1063/1.1699681.

[28] D. Lee, The strain rate dependent plastic flow behavior of zirconium and its alloys, MT 1 (1970) 1607–1616. doi:10.1007/BF02642007.

[29] K. Matsuki, H. Morita, M. Yamada, Y. Murakami, Relative motion of grains during superplastic flow in an Al-9Zn-1 wt.%Mg alloy, Metal Science 11 (1977) 156–163. doi:10.117 9/msc.1977.11.5.156.

[30] F. R. N. Nabarro, Steady-state diffusional creep, The Philosophical Magazine: A Journal of Theoretical Experimental and Applied Physics 16 (1967) 231–237. doi:10.1080/1478 6436708229736.

[31] M. Zelin, A. Mukherjee, Geometrical aspects of superplastic flow, Materials Science and Engineering: A 208 (1996) 210–225. doi:10.101 6/0921-5093(95)10080-6.

[32] J. C. M. Li, Possibility of Subgrain Rotation during Recrystallization, Journal of Applied Physics 33 (1962) 2958–2965. doi:10.1063/1. 1728543.

[33] L. Wang, J. Teng, P. Liu, A. Hirata, E. Ma, Z. Zhang, M. Chen, X. Han, Grain rotation mediated by grain boundary dislocations in nanocrystalline platinum, Nat Commun 5 (2014) 4402. doi:10.1038/ncomms5402.

[34] V. V. Bulatov, B. W. Reed, M. Kumar, Grain boundary energy function for fcc metals, Acta Materialia 65 (2014) 161–175. doi:10.1016/j. actamat.2013.10.057.

[35] 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 (2009) 3694–3703. doi:10.1016/j.actamat.2009.04.007.

[36] S. G. Baird, E. R. Homer, D. T. Fullwood, O. K. Johnson, Five degree-of-freedom property interpolation of arbitrary grain boundaries via Voronoi fundamental zone framework, Computational Materials Science 200 (2021) 110756. doi:10.1016/j.commatsci.20 21.110756.

[37] 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 (2009) 3704–3713. doi:10.1016/j.actamat.2009.04.015.

[38] S. G. Baird, E. R. Homer, D. T. Fullwood, O. K. Johnson, Five Degree-of-Freedom Property Interpolation of Arbitrary Grain Boundaries via Voronoi fundamental zone Framework, Computational Materials Science (Under Review) 26.

[39] 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. doi:10.1016/j.actamat.2018.12. 034.

[40] Exact GPR Method - MATLAB & Simulink, https://www.mathworks.com/help/stats/exact-gpr-method.html, ????

[41] H.-K. Kim, S. G. Kim, W. Dong, I. Steinbach, B.-J. Lee, Phase-field modeling for 3D grain growth based on a grain boundary energy database, Modelling Simul. Mater. Sci. Eng. 22 (2014) 034004. doi:10.1088/0965-0 393/22/3/034004.

No Comments.

[42] G. S. Rohrer, E. A. Holm, A. D. Rollett, S. M. Foiles, J. Li, D. L. Olmsted, Comparing calculated and measured grain boundary energies in nickel, Acta Materialia 58 (2010) 5063–5069. doi:10.1016/j.actamat.2010.05.042.

[43] D. G. Brandon, The structure of high-angle grain boundaries, Acta Metallurgica 14 (1966) 1479–1484. doi:10.1016/0001-6160(66)90168-4.

[44] A. H. King, S. Shekhar, What does it mean to be special? The significance and application of the Brandon criterion, J Mater Sci 41 (2006) 7675–7682. doi:10.1007/s10853-006-0665-8.

[45] J. Han, V. Vitek, D. J. Srolovitz, Grain-boundary metastability and its statistical properties, Acta Materialia 104 (2016) 259–273. doi:10.1016/j.actamat.2015.11.035.

[46] J. Wei, B. Feng, R. Ishikawa, T. Yokoi, K. Matsunaga, N. Shibata, Y. Ikuhara, Direct imaging of atomistic grain boundary migration, Nat. Mater. (2021). doi:10.1038/s41563-020-00879-z.

[47] A. Morawiec, On distances between grain interfaces in macroscopic parameter space, Acta Materialia 181 (2019) 399–407. doi:10.1016/j.actamat.2019.09.032.

No Comments.

Towards a Quantitative Cartography of the Grain Boundary Energy Landscape: Paths and Correlations: Supplementary Information

Sterling G. Baird[a,*], Eric R. Homer[a], David T. Fullwood[a], Oliver K. Johnson[a]

[a]*Department of Mechanical Engineering, Brigham Young University, Provo, UT 84602, USA*

Contents

S1. Brief Summary of VFZ Methods

We summarize the following aspects of the Voronoi fundamental zone (VFZ) framework:

- Creation and definition of a VFZ

- Mapping grain boundaries (GBs) to the VFZ

- Distance calculations

- Interpolation

- Comparison with traditional grain boundary octonion (GBO) metric

Each of these is described in greater detail in Baird et al. [1].

To define a VFZ, an arbitrary, fixed, low-symmetry reference GBO is chosen (o_{ref}) and for our use of GBOs, the VFZ is defined as the region of $\mathbb{S}^7$ (the unit 7-sphere in 8 dimensions) closer to o_{ref} than any of its symmetric images. If a low-symmetry GB is chosen, the point within a VFZ will be unique

---

Number: 1 Author: sterg Subject: Highlight Date: 2021-09-29 00:10:06

within numerical tolerance (and hence it is a true fundamental zone). Additionally, we use a Euclidean approximation to the true geodesic distance.

A GBO is composed of two quaternions, with the boundary plane normal in the $+z$ direction [2]. A GBO is mapped into a VFZ by calculating the pairwise distances between the reference GBO and each of the symmetrically equivalent octonions[1] and taking the symmetrically equivalent octonion closest to the reference GBO.

Once a GBO has been mapped into a VFZ, distance calculations proceed without further consideration of symmetrically equivalent octonions. The VFZ framework suffers from occasional, large distance overestimation which imposes a local sparseness of data and lead to poorer interpolation near the borders of a VFZ. However, this can be mitigated through ensemble or data augmentation techniques.

Note that in the original definition of the VFZ, the authors examined different interpolation techniques [1]. In the present work, we focus on Gaussian process regression which in our case imposes the assumption that crystallographically similar GBs share similar grain boundary energys within some correlation length. Gaussian process regression has the added benefit of built-in uncertainty quantification.

The primary differences between the VFZ framework and traditional GBO distance metric are that the VFZ framework is defined by a continuous set of points, exhibits occasional distance overestimation, uses a Euclidean approximation, and has a lower computational complexity.

S2. Semivariograms for Estimating Global and Local Correlation Lengths

S2.1. Semivariogram Method

First, we provide some background to put correlation lengths in context. The construction of Gaussian process regression models involves the combination of a prior distribution over the model space, with some set of observations and their quantified uncertainty. The result is a posterior distribution that provides the probability (density) of any particular model in light of the observed data and priors. In Gaussian process regression, as the name suggests, the priors are assumed to be Gaussian and therefore of the form

$$f(\mathbf{m}) \propto \exp\left(-\frac{1}{2}(\mathbf{m} - \mathbf{m}_0)^\mathsf{T} \mathbf{C}_{\mathbf{m}_0}^{-1}(\mathbf{m} - \mathbf{m}_0)\right) \tag{S1}$$

where $f(\mathbf{m})$ is the probability density of an arbitrary model, $\mathbf{m} = \mathbf{m}(x)$, and $\mathbf{m}_0 = \mathbf{m}_0(x)$ is the prior model (i.e. a guess as to what the true model ought to look like). The quantity $\mathbf{C}_{\mathbf{m}_0} = \mathbf{C}_{\mathbf{m}_0}(x_i, x_j)$ is the "kernel" or covariance function of the prior and it describes the prior (assumed) covariance between the values $\mathbf{m}(x_i)$ and $\mathbf{m}(x_j)$.

It is possible to use a wide variety of kernel types, depending on the prior information one may have about the physical phenomenon one is attempting to model, e.g. continuity, differentiability, anisotropy, stationarity, and length-scales of correlation. One of the most common kernels employed is the Gaussian (sometimes called the squared exponential) kernel:

$$\mathbf{C}_{\mathbf{m}_0}(x_i, x_j) = \sigma^2 e^{-\frac{(x_i - x_j)^\mathsf{T}(x_i - x_j)}{2l^2}} \tag{S2}$$

where σ, l, x, and $(\cdot)^\mathsf{T}$ represent signal standard deviation, length scale, VFZ coordinates, and transpose operator, respectively.

---

[1]Contrary to Francis et al. [2] which uses the passive convention for misorientation, we employ the active convention [1].

Number: 1 Author: sterg Subject: Highlight Date: 2021-09-29 00:10:15

The values of σ and l are typically estimated from the data. One approach, employed by the fitrgp() routine in MATLAB involves numerical optimization via gradient descent which maximizes the likelihood as a function of these parameters [3].

An alternative approach, adapted from geostatistical applications, involves calculation of the empirical semivariogram [REF], in which one bins the space of pairwise distances between all GBs and then calculates half the average pairwise variance of the corresponding property values in each bin:

$$\kappa(d_k) = \frac{1}{2N_k} \sum_{\substack{d_\Omega(x_i,x_j) \in \\ [d_k^-, d_k^+]}} |E(x_i) - E(x_j)|^2 \tag{S3}$$

where x_i and x_j are the crystallographic coordinates of GBs i and j, $d_\Omega(x_i, x_j)$ is the distance between them, and $E(x_i)$ and $E(x_i)$ are their respective energies. d_k is the location (distance) of the k-th bin center having left- and right-hand limits d_k^- and d_k^+, and N_k is the number of measurement pairs whose distance falls in the k-th bin. Due to limited sampling of large distances and the fact that the most informative part of the semivariogram is the region near $d_k = 0$, it is customary to limit the semivariogram to half of the maximum distance [REF]. The empirical semivariogram is then fit with an analytical model to obtain the parameters of the kernel (covariance) function taking advantage of the relationship

$$\kappa(d) = \sigma_f^2 - \mathbf{C}_{\mathbf{m}_0}(d) \tag{S4}$$

where we have made explicit the stationarity of the Gaussian kernel (i.e. that it depends only the distance between two points, not on their respective locations).

Having obtained a value for the length scale kernel parameter l, one can define a correlation length for the data. If the distance between two points is equal to l the kernel function indicates that their correlation will be equal to $\rho = \exp(-1/2) \approx 0.61$. However, one might reasonably want to know the length scale over which GB properties are correlated by a different amount. In general for a specified correlation strength, ρ, the corresponding correlation length is given by

$$l'(\rho) = l\sqrt{-2\ln \rho} \tag{S5}$$

We will refer to the parameter l as *the* correlation length, but one can use Eq. (S5) to determine the length scale corresponding to any specified correlation strength.

S2.2. Global Correlation Lengths

One of the most fundamental observations comes from the shape of the empirical semivariograms. As mentioned in Section 2.3, there are a wide variety of kernels that could be candidates for modeling correlations in different systems. Different kernels are used to capture different types of correlations, and each has a characteristic signature that can be observed in the semivariogram. For example, when a system exhibits exponential-type correlations, the semivariogram manifests this in the form of an exponential convergence towards an asymptotic constant value at long-distances. Linear and power-type correlations manifest an absence of a long-distance plateau. In the present system, there is a clear change in concavity in the semivariogram, this is a signature of correlations that are Gaussian in nature. Thus, we find that GB energy correlations in these systems are Gaussian, and should therefore be modeled using Gaussian kernels.
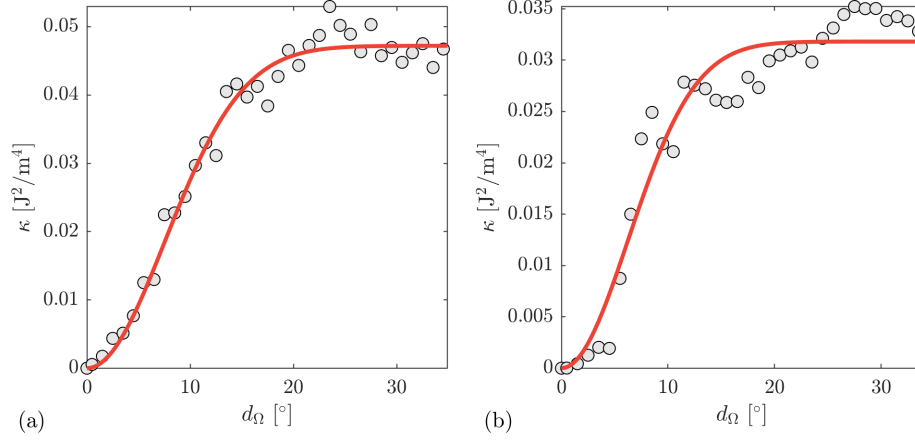
Number: 1 Author: sterg Subject: Highlight Date: 2021-09-29 00:10:23

4

Number: 1 Author: sterg Subject: Highlight Date: 2021-09-29 00:10:27

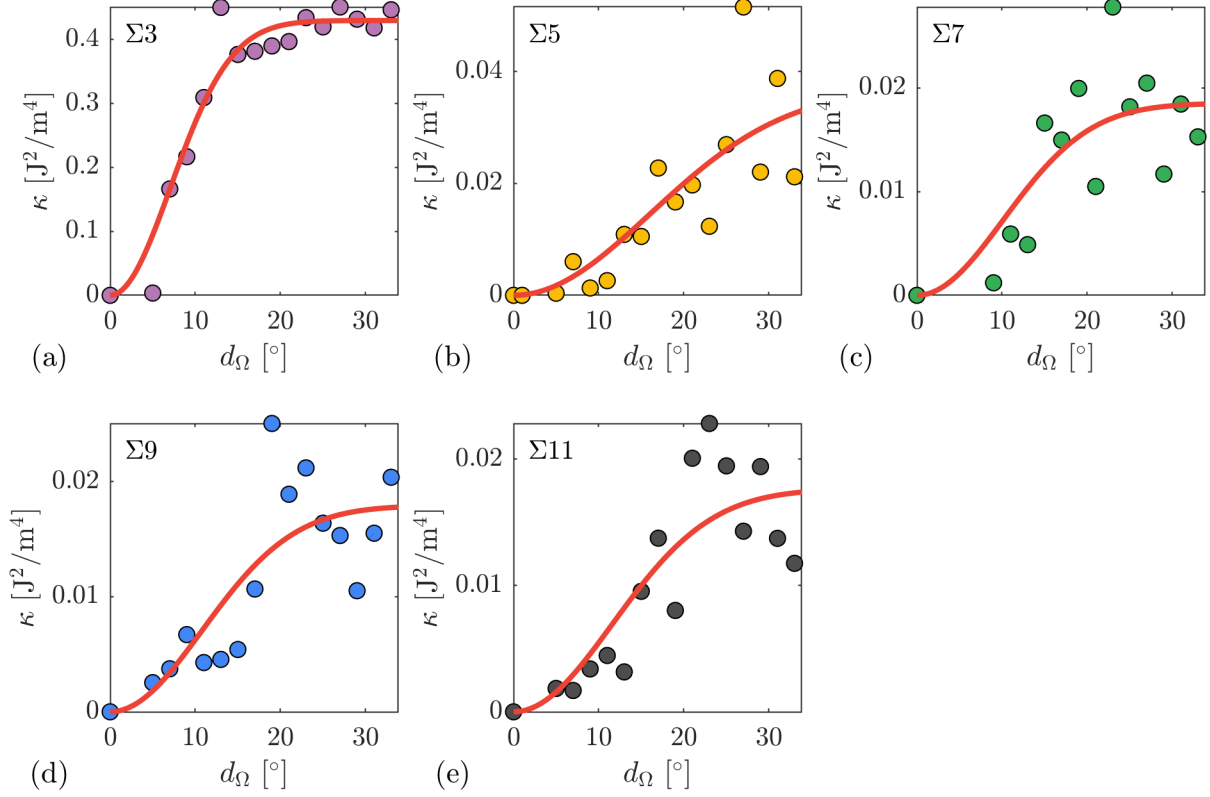Number: 2 Author: sterg Subject: Highlight Date: 2021-09-29 00:10:30

Number: 1 Author: sterg Subject: Highlight Date: 2021-09-29 00:10:35

The local empirical semivariograms are noisier than the global empirical semivariograms, likely due to considering fewer GB pairs. Nevertheless, reasonable fits were obtained for most of the GBs with the exception of the $\Sigma5$ GB in the Ni dataset. In many of the local semivariograms we again see the signature change in concavity suggesting that the local correlations in the vicinity of these GBs are also Gaussian in nature. However, there are some exceptions where the nature of the correlations is more ambiguous. We anticipate that this ambiguity could be resolved with datasets that are either larger (compared to the Ni dataset) or having less noise (compared to the Fe dataset) than those considered here. However, the local empirical semivariograms seem to be generally consistent with Gaussian-type correlations.

The traditional Gaussian kernel exhibits the property of stationarity, meaning that the covariance depends only on the distance between two points, not on their respective locations (i.e. $\mathbf{C}_{\mathbf{m}_0}(x_i, x_j) = \mathbf{C}_{\mathbf{m}_0}(d(x_i, x_j))$). The use of a stationary kernel implies a prior assumption that there is a single global correlation length that applies everywhere. The fact that we observe significant variation in correlation length across the GB character space suggests that it would be better to employ non-stationary kernels (this is why, when referring to the global correlation length results presented earlier, we were careful to say that the global correlation lengths hold "on average" across the space). In particular, due to the fact that the local semivariograms do seem to be generally consistent with Gaussian-type correlations, we suggest that the non-stationary version of the Gaussian kernel [REF] may be a reasonable choice. One additional potential benefit of employing non-stationary kernels might be improved resolution of cusps in the GB energy landscape.

S3. Fe Input Data Quality

Of the $\sim$60 000 GBs[2] in [4], $\sim$10 000 GBs were repeats that were identified by converting to Voronoi fundamental zone grain boundary octonions and applying VFZ repository function `avg_repeats.m`. In [4], mechanically selected GBs were those which involved sampling in equally spaced increments[3] for each five degree-of-freedom parameter, and a few thousand intentionally selected GBs (i.e. special GBs) were also considered. Of mechanically and intentionally selected GBs, 9170 and 112 are repeats, respectively, with a total of 2496 degenerate sets[4] (see Figure S4 for a degeneracy histogram). Thus, on average there is a degeneracy of approximately four per set of degenerate GBs.

By comparing grain boundary energy values of (unintentionally[5]) repeated GBs in the Fe simulation dataset [4], we can estimate the intrinsic error of the input data. For example, minimum and maximum deviations from the average value of a degenerate set are $-0.2625\,\mathrm{J\,m^{-2}}$ and $0.2625\,\mathrm{J\,m^{-2}}$, respectively, indicating that a repeated Fe GB simulation from [4] can vary by as much as $0.525\,\mathrm{J\,m^{-2}}$, though rare. Additionally, Root mean square error and mean absolute error values can be obtained within each degenerate set by comparing against the set mean. Overall root mean square error and mean absolute error are then obtained by averaging and weighting by the number of GBs in each degenerate set. Following this procedure, we obtain an average set-wise root mean square error and mean absolute error of $0.065\,29\,\mathrm{J\,m^{-2}}$ and $0.061\,90\,\mathrm{J\,m^{-2}}$, respectively, which is an approximate measure of the intrinsic

---

[2]The "no-boundary" GBs (i.e. GBs with close to $0\,\mathrm{J\,m^{-2}}$ grain boundary energy) were removed before testing for degeneracy.

[3]In some cases, this was equally spaced increments of the argument of a trigonometric function.

[4]A degenerate "set" is distinct from a Voronoi fundamental zone grain boundary octonion "set", the latter of which is often used in the main text.

[5]To our knowledge, the presence of repeat GBs were not mentioned in [4] or [5]
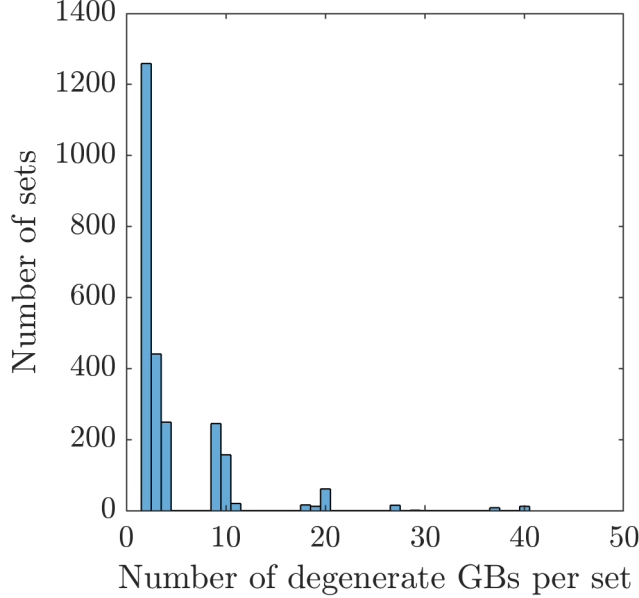
Page:23

missing reference (Dr. Johnson)

Figure S4: Histogram of number of sets vs. number of degenerate GBs per set for the Fe simulation dataset [4]. Most sets have a degeneracy of fewer than 5.

error of the data. Figure S5 shows histograms and parity plots of the intrinsic error. The overestimation of intrinsic error mentioned in the main text (Section 3.3.1) could stem from bias as to what type of GBs exhibit repeats based on the sampling scheme used in [4] and/or that many of the degenerate sets contain a low number of repeats (Figure S4).

Next, we see that by binning GBs into degenerate sets, most degenerate sets have a degeneracy of fewer than 5 Figure S4. We split the repeated data into sets with a degeneracy of fewer than 5 and greater than or equal to 5 and plot the errors (relative to the respective set mean) in both histogram form (Figure S5a and Figure S5c, respectively) and as hexagonally-binned parity plots (Figure S5b and Figure S5d, respectively). While heavily repeated GBs tend to give similar results, occasionally repeated GBs often have larger grain boundary energy variability. This could have physical meaning: Certain types of (e.g. high-symmetry) GBs tend to have less variation (i.e. fewer and/or more tightly distributed metastable states). However, it could also be an artifact of the simulation setup that produced this data (e.g. deterministic simulation output for certain types of GBs).

S4. Gridded Sampling for Numerical Differentation

An isotropically sized fundamental zone may be easier to uniformly discretize than a high aspect-ratio space (i.e. a fixed discretization length can be used across all dimensions). What this doesn't describe, however, is curvature. In order to create a gridded array, which is important for numerical differentiation, a hypercube with each primary axis oriented with Euclidean dimensions is to be preferred. As curvature or misalignment is introduced as may be expected with a VFZ point cloud, GBs outside of the VFZ will necessarily be sampled; this phenomena will be exaggerated in high dimensions[6]. Fortunately, most of

---

[6]For perspective, a discretization into 9 segments (10 points) along each dimension will have a spacing of $\sim 7°$ and require 1×10^5 grid points. In order to achieve a more reasonable grid spacing of $\sim 2°$, a minimum of ~ 24 discretizations (25 points) along each dimension is necessary and will produce $\sim 1 \times 10^7$ grid points.
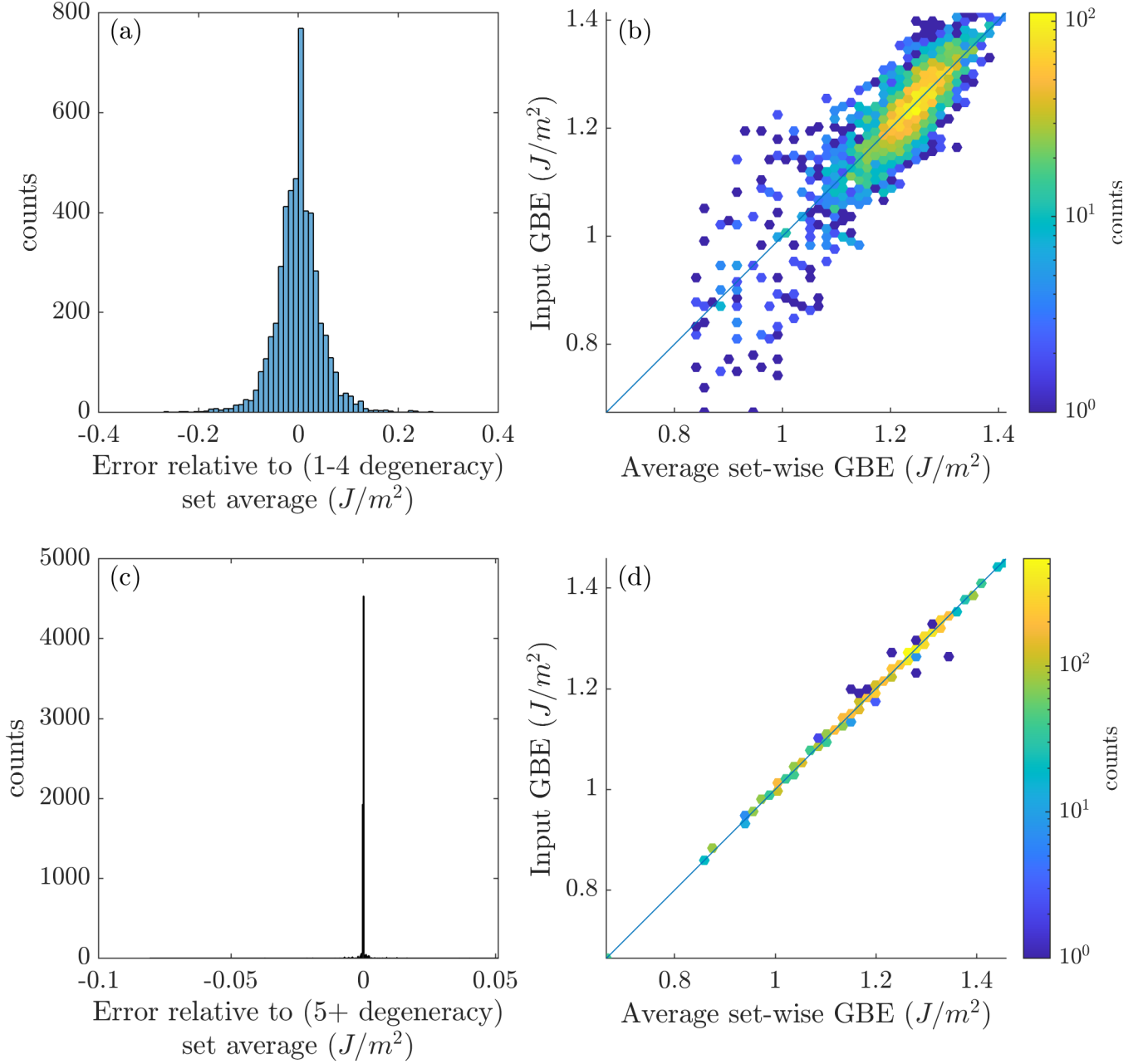
7

No Comments.

Figure S5: Degenerate GBs sets are split into those with a degeneracy of fewer than 5 and greater than or equal to 5 and plotted as ((a) and (c), respectively) error histograms and ((b) and (d), respectively) hexagonally-binned parity plots. Large degenerate sets tend to have very low error, whereas small degenerate sets tend to have higher error. In other words, GBs that are more likely to be repeated many times based on the sampling scheme in [4] tend to give similar results, whereas GBs that are less likely to be repeated often have larger variability in the simulation output. We do not know if this has physical meaning or is an artifact of the simulation setup.

No Comments.

the information is contained in the first five dimensions after singular value decomposition transformation (Section 3.1). Thus, the latter three dimensions can likely be ignored without substantially affecting e.g. an interpolation or numerical differentiation scheme.

S5. GBs Used for Path Visualization

See Table 1 and Table 2 for the Olmsted et al. [6] and Kim et al. [4] datasets, respectively.

Table 1: Minimum Σ (Sigma) GBs and corresponding IDs used for path visualization within the original Olmsted et al. [6] dataset.

Sigma	Olmsted ID
3	3
5	169
7	32
9	21
11	33

Table 2: Minimum Σ (Sigma) GBs and corresponding IDs used for path visualization within the original Kim et al. [4] dataset.

Sigma	Kim Special ID
3	7
5	162
7	259
9	315
11	406

Glossary

GB grain boundary 1, 2, 5–9

GBO grain boundary octonion 1, 2

VFZ Voronoi fundamental zone 1, 2, 6, 7

No Comments.

References

[1] S. G. Baird, E. R. Homer, D. T. Fullwood, O. K. Johnson, Five Degree-of-Freedom Property Interpolation of Arbitrary Grain Boundaries via Voronoi fundamental zone Framework, Computational Materials Science (Under Review) 26.

[2] 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. doi:10.1016/j.actamat.2018.12.034.

[3] Exact GPR Method - MATLAB & Simulink, https://www.mathworks.com/help/stats/exact-gpr-method.html, ????

[4] H.-K. Kim, S. G. Kim, W. Dong, I. Steinbach, B.-J. Lee, Phase-field modeling for 3D grain growth based on a grain boundary energy database, Modelling Simul. Mater. Sci. Eng. 22 (2014) 034004. doi:10.1088/0965-0393/22/3/034004.

[5] H. K. Kim, W. S. Ko, H. J. Lee, S. G. Kim, B. J. Lee, An identification scheme of grain boundaries and construction of a grain boundary energy database, Scripta Materialia 64 (2011) 1152–1155. doi:10.1016/j.scriptamat.2011.03.020.

[6] 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 (2009) 3704–3713. doi:10.1016/j.actamat.2009.04.015.

No Comments.