

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 $\sim 65^\circ$ 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 $\sim 6^\circ$ to 8° in the grain boundary octonion (GBO) sense (multiply by 2 to convert to misorientation 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 criterion. We suggest that property data with no more than $\pm \sim 3\%$ error and point sets with GBs that are no more than $\sim 3\text{--}4^\circ$ apart should be used and then paired with high-fidelity interpolation strategies. Finally, in terms of dynamic material behavior, geodesic paths through 5DOF space for Ni suggest that, under appropriate conditions, a certain low-energy $\Sigma 7$ GB may transform into the frequently observed $\Sigma 3$ 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 mi-

crostructures 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

*Corresponding author.

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

GBs that are near each other in that space.

Examples of such spontaneous energy minimizing 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], or when 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 submanifolds (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 inclination¹. 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 lengths² and find that previous estimates are likely too high for low-noise, computational GBE calculations while on par when the low-noise constraint is removed. 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

¹Even 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.

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

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 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, followed by a brief comparison with the traditional GBO metric (Section S1). A 2-sphere analogue to Voronoi fundamental zone grain boundary octonions (VFZ-GBOs) is provided in Figure 1. The methods are based on functions and scripts from (github.com/sgbaird-5dof/interp), and we refer the reader to Baird et al. [36] 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 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 the correlation length, numerical optimization is performed via gradient-descent-based maximization of a likelihood function that depends on the correlation length, l , and noise, σ , parameters [40].

2.4. Visualizing 5DOF Paths

geodesic paths between two GBs in a VFZ are obtained via coordinate interpolation constrained to a hyperspherical arc. The geodesic 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 S3.1 for further details on the methods used to estimate the intrinsic uncertainty of the Fe simulation dataset.

3. Results and Discussion

Measuring and/or calculating the properties of GBs are generally time-intensive efforts. When seeking to collect data to establish a GB structure-property model, it is therefore desirable to determine the minimum number of measurements/calculations needed. To answer this question, three pieces of information are necessary: (1) the size and shape of the space, (2) how close the points must be to one another, and (3) the uncertainty of the data.

To obtain the first piece of information, we need to determine the size and shape of a VFZ. To obtain the second piece of information, we need to determine the distance over which the property of interest is correlated. If the crystallographic distance between nearest neighbors (NNs) in a collection of GBs is larger than the correlation length, then there will be large gaps between points where

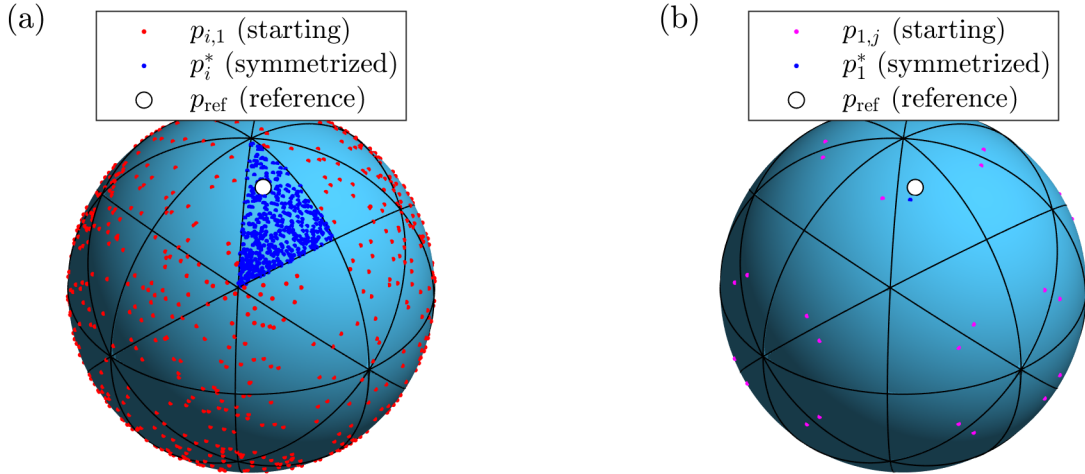


Figure 1: (a) Symmetrization of many points relative to a fixed reference point (white circle) into a 3D Cartesian VFZ point set (dark blue points). (b) To further illustrate, identical representations of a single input point (magenta points) is symmetrized (dark blue point) relative to a fixed reference point (white circle), demonstrating that only one symmetrized point is found within the borders (black lines) of each of the Voronoi cells (light blue spherical triangle). In the case of VFZ-GBOs, the same procedure applies to the U(1)-symmetrized 6-sphere. Reproduced with permission from Baird, S. G.; Homer, E. R.; Fullwood, D. T.; Johnson, O. K. Computational Materials Science 2021, 200, 110756 [36].

predictions/interpolation will be unreliable. If, on the other hand, the NN distances are smaller than the correlation length, then predictions between the measured/calculated values can be expected to be reasonable. Finally, estimates of uncertainty can be obtained by comparing the distribution of repeated measurements. We use VFZ-GBOs to explore structure-property relations in the 5DOF space of GBs. To make our aims more complete, we summarize these points in 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 of GB energy? (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 finding local minima (Section 3.5).

3.1. Dimensions of a VFZ

The maximum principal component of a particular³ singular value decomposition transformed O_h cubic VFZ is $\sim 65^\circ$. 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 (d_Ω) for a set of 20 000 VFZ-GBOs. These are the diagonal entries of the “S” matrix in the singular value decomposition.

Dimension	d_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

³This VFZ was used to analyze the dimensions and is distinct from the VFZs used for the Ni and Fe datasets.

By performing principal component analysis, we find that $\sim 99.6\%$ of the variance (spatial dispersion) 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 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.

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 GBs in a VFZ-GBO set affects, on average, how close GBs are to their 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 $1 d_\Omega^\circ$. Figure S1 shows the empirical semivariograms together with the analytical

fits used to obtain the values of the respective correlation lengths. The change in concavity suggests 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 $7.5 d_\Omega^\circ$ for both the semivariogram and gradient methods. For the Fe dataset, we observe $6.2665 d_\Omega^\circ$ for the semivariogram method and $8.3073 d_\Omega^\circ$ for the gradient method, the average of the two being $7.3 d_\Omega^\circ$. Thus both datasets seem to have correlation lengths of about $7.5 d_\Omega^\circ$. This is 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° ($7.5 d_\Omega^\circ$) 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

Table 3: Correlation lengths obtained via semivariogram and gradient based approaches in units of GBO distance (multiply by 2 to get misorientation 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.

	$l : d_\Omega [^\circ]$	
	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

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 $1 d_\Omega^\circ$ or less).

Using the correlation lengths, l , obtained via the semivariogram of the input data, Figure 2 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 $l' = 0 d_\Omega^\circ$ for perfect correlation ($\rho = 1$) to $l' \approx 23 d_\Omega^\circ$ and $l' \approx 19 d_\Omega^\circ$ for essentially⁴ zero correlation in the Ni and Fe datasets, respectively.

If low noise is assumed for Ni, the correlation length drops to ~ 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.

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

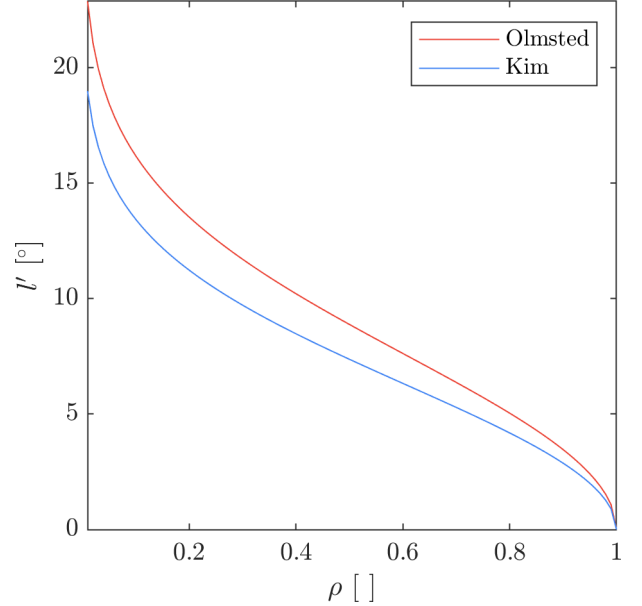


Figure 2: Global length scale of correlation, l' , in units of GBO distance ($d_\Omega [^\circ]$) as a function of correlation strength, ρ , for the Ni (Olmsted) and Fe (Kim) datasets. Multiply GBO distance by 2 to convert to misorientation angle.

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 in units of $d_\Omega (^\circ)$, signal standard deviation (Jm^{-2}), constant basis function (Jm^{-2}), and input property standard deviation (Jm^{-2}), respectively.

Mat.	Fix σ	$\sigma_L (^\circ)$	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

⁴Technically the correlation does not go to exactly zero until the length scale is infinite. The values listed here are for a correlation of $\rho = 0.01$.

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

the correlation lengths of 10° [42] and 15° [37] previously reported⁵ 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 misorientation 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 $\sim 1.9^\circ$ 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 length⁶. 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

⁵Both of these lengths are based on results from Olmsted et al. [37].

⁶The 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.

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?
- 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 semi-variogram 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 $2 d_\Omega^\circ$ 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 $\Sigma 5$ in the Ni dataset Section S2.3.

Table 6 provides the local correlation lengths obtained 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 $\Sigma 3$ 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 $\Sigma 9$ in the Ni dataset, and the

Table 6: Local correlation lengths in the vicinity of specific low- Σ GBs, obtained via the semivariogram method in units of octonion distance. The fit of the $\Sigma 5$ GB for the Ni dataset was sufficiently poor that we do not report a corresponding correlation length.

	$l : d_{\Omega} [^{\circ}]$	
	Ni	Fe
$\Sigma 3$	7.1460	7.0921
$\Sigma 5$	—	16.0440
$\Sigma 7$	9.6533	10.1493
$\Sigma 9$	17.2601	10.7792
$\Sigma 11$	10.2027	11.6011

$\Sigma 5$ 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 threshold for the maximum angular deviation that can be accommodated by a coincident site lattice boundary before losing coincidence:

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

where $\Delta\theta$ is the angular deviation threshold, $\theta_0 = 15^{\circ}$ 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 the 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 3 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

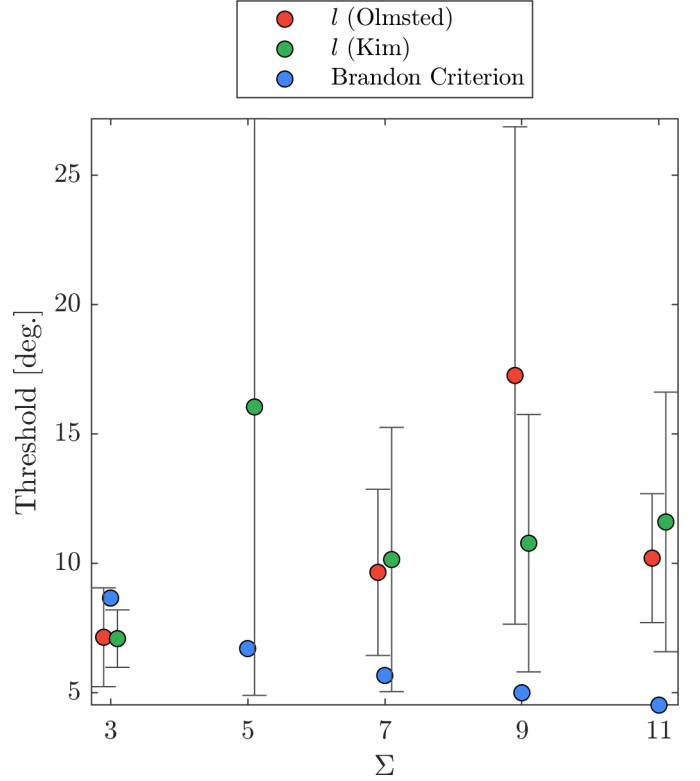


Figure 3: Comparison of the Brandon criterion [43] for low- Σ GBs (in units of $^{\circ}$) to the local correlation lengths obtained in this work (in units of d_{Ω}°). Conversion to consistent units would unnecessarily separate the data without changing the trends, and since we are most interested in 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.

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 explain why previous researchers found Sigma to be a poor predictor of GB energy [44].

3.2.3. Density and Distribution of Points

Figure 5 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 be-

tween $(10.7175 \pm 0.3684)^\circ$ and $(2.6479 \pm 0.2254)^\circ$, respectively (d_Ω). 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 $(2.87020 \pm 0.69117)^\circ$ (Figure 4a) while the average 100-th NN distance is within 10° (Figure 4b). While not shown, we also found that the sampling scheme used in this work (sampling misorientation and boundary plane normal as opposed to sampling GBOs directly) produced no qualitative differences in the observed distributions.

How large is the local region of influence for a 5DOF interpolation model? Another way of phrasing this question is: how many GBs are necessary to sample in the space to obtain a model that’s accurate enough? NN distances give context to these questions: if 50 000 GBs are randomly sampled, then on average, these GBs are $(2.6479 \pm 0.2254)^\circ$ apart from their NN. In the context of correlation lengths of $\sim 7\text{--}8^\circ$ (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 $\sim 7.5^\circ$. 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 GBs which are further apart than $\sim 7.5^\circ$ 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 lim-

itation by imposing prior information: a “scaffolding” 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

The VFZ-GBO framework allows us to probe the quality of the datasets that are used which provides additional context for the accuracy of the models that we report. Using 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; in other words, the observable model accuracy will not improve unless the input data is less noisy (Section S3.2)
- the predictions likely exhibit overprediction bias relative to the true minimum for a given GB; in other words, accurate resolution of GBE cusps and noisy data are incompatible (Section S3.3)
- future availability of multiple metastable state GBEs is anticipated to greatly improve the model performance; in other words, even with a large quantity of data, poor quality severely hampers the utility of the data (Section S3.4)

3.4. 5DOF Paths and GB Behavior

We constructed 5D structure-property models for GB energy (i.e. GB energy landscapes), with quantified uncertainty, for the Ni [37] and Fe [41] datasets previously mentioned using the GPR method of `interp5DOF` (<https://github.com/sgbaird-5DOF/interp>). The performance and fidelity of these models were discussed in previous work [36]. Here we use these models to explore geodesic paths through the resulting GB energy landscapes,

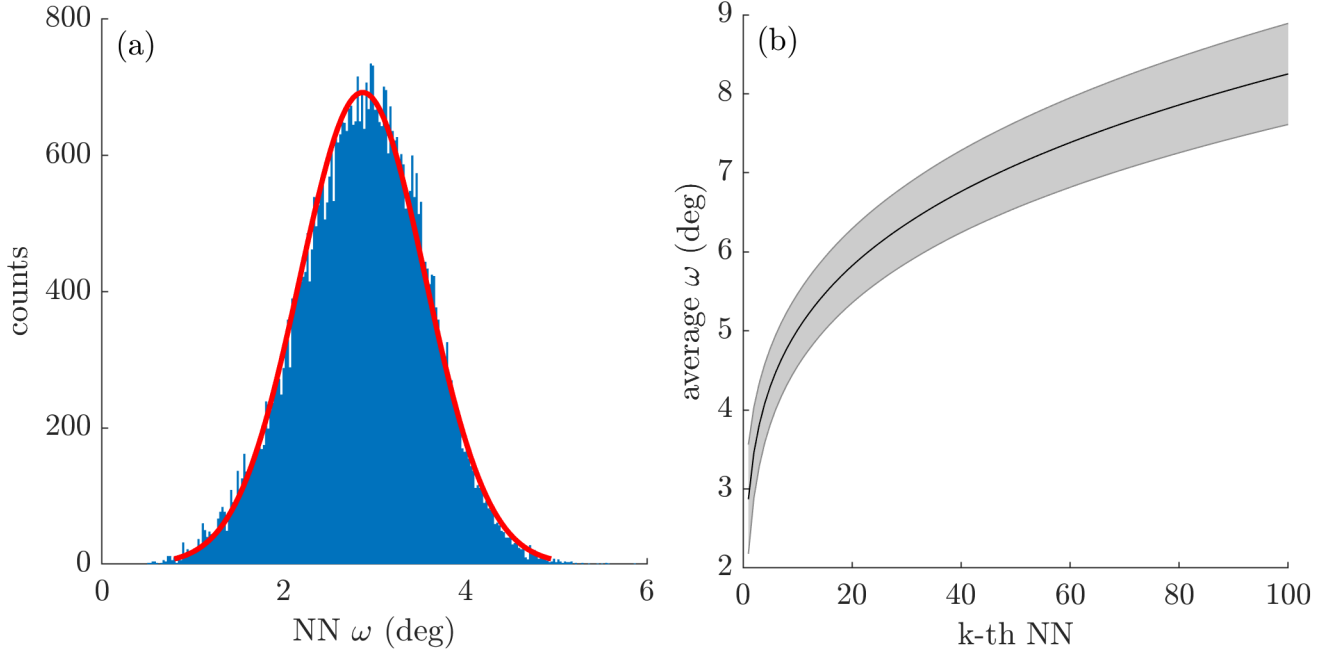


Figure 4: (a) Histogram of NN GBO distances (d_{Ω}) in a VFZ-GBO set of 50 000 points. The average NN distance was $(2.87020 \pm 0.69117)^{\circ}$. (b) The average k-th nearest neighbor distances demonstrate that NNs up to the ~ 10 -th NN fall within a tolerance of $\sim 5^{\circ}$. Standard deviation uncertainty bars using approximately 10 trial runs are also shown.

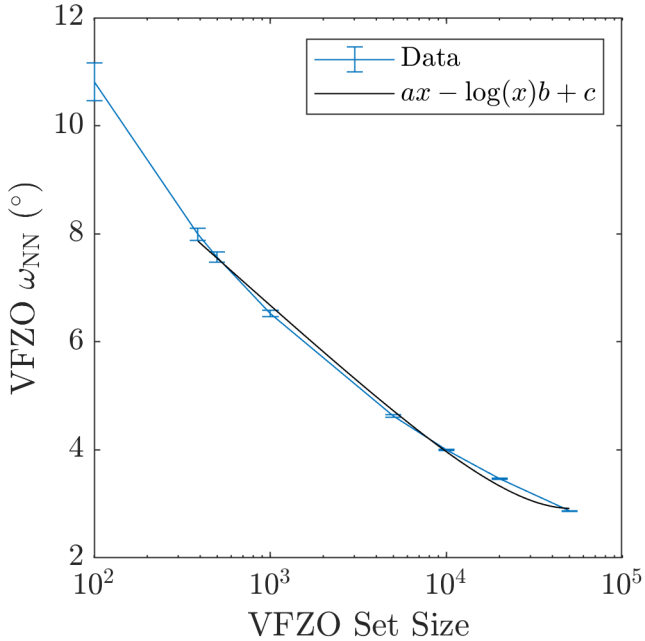


Figure 5: NN VFZ-GBO (ω_{NN}) distances ($^{\circ}$) versus VFZ-GBO set size out of 70-80 random VFZ-GBO sets per set size and a fit to $ax - \log(x)b + c$ where $a = 2.5025 \times 10^{-5}$, $b = 1.27396$, $c = 15.4499$, x represents set size, and $388 \leq x \leq 50000$.

between certain low- Σ GBs, and discuss the implications for microstructure evolution.

We choose the $\Sigma 5$, $\Sigma 7$, $\Sigma 9$, and $\Sigma 11$ GBs with the lowest GBE⁷; we then visualize geodesic paths as “tunnel” plots⁸ in a VFZ between each of these and the global minimum $\Sigma 3$ coherent-twin GB (Figures 6 and 7). This is performed for both the BRK and VFZ-GPR models. 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 data near each point along the path.

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,

⁷The 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 S1 and Table S2, respectively.

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

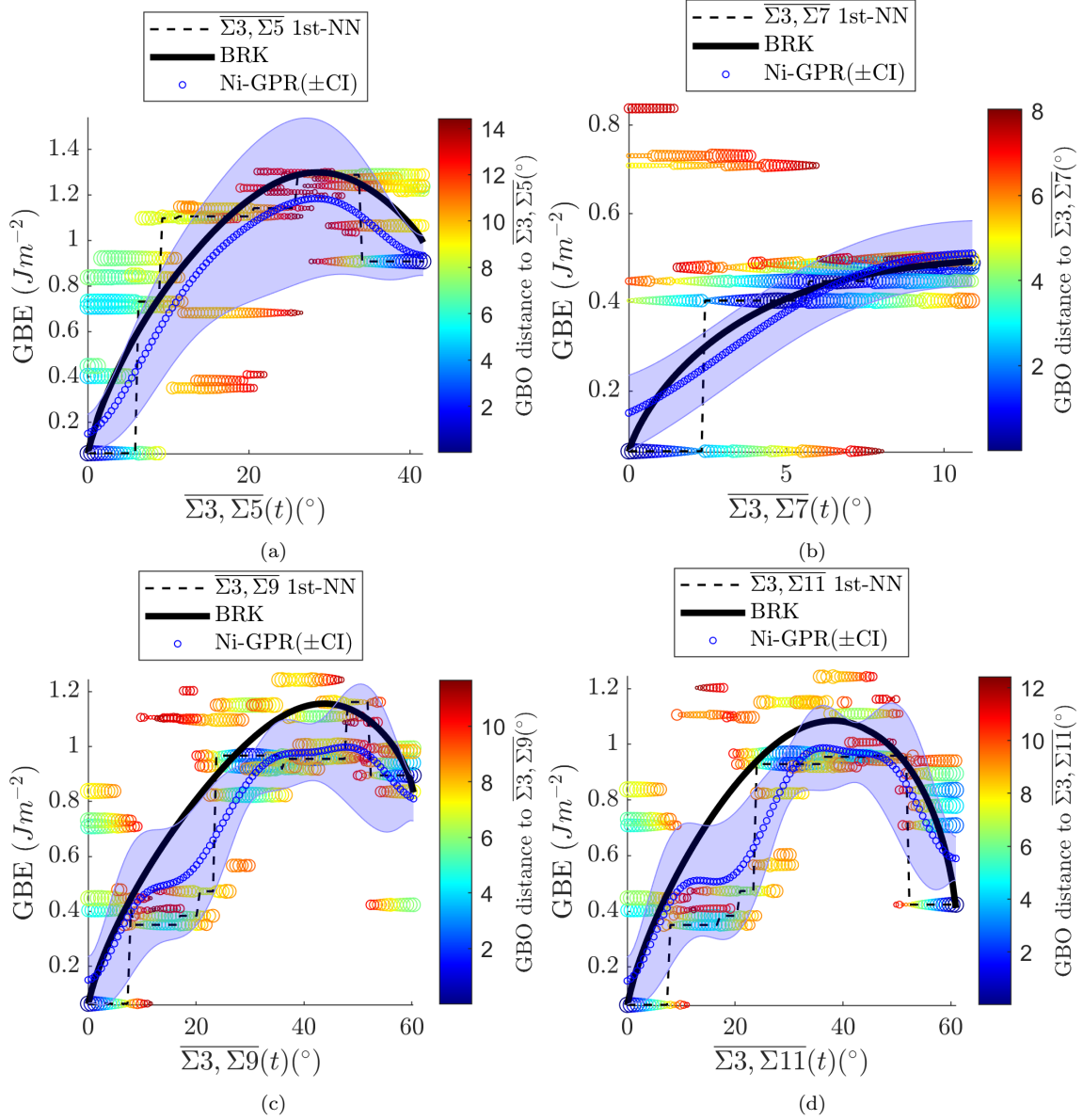


Figure 6: “Tunnel” plots of GBEs along geodesic paths in a VFZ between the $\Sigma 3$ coherent-twin boundary and minimum GBE (a) $\Sigma 5$, (b) $\Sigma 7$, (c) $\Sigma 9$, and (d) $\Sigma 11$ GBs within the Ni Olmsted et al. [37] dataset. GBE values are plotted for the BRK and GPR models which both 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 geodesic 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.

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.

From another perspective, any transformation will depend on both the driving force and the activation energy. For example, near the transformation temperature, nucleation is slow because the driving force is small despite having fast diffusion.

Likewise, at large undercoolings nucleation will be slow, but this time due to sluggish diffusion despite a large driving force. The GB energy landscape tells us about the driving force for such a GB crystallographic transformation, 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 pre-

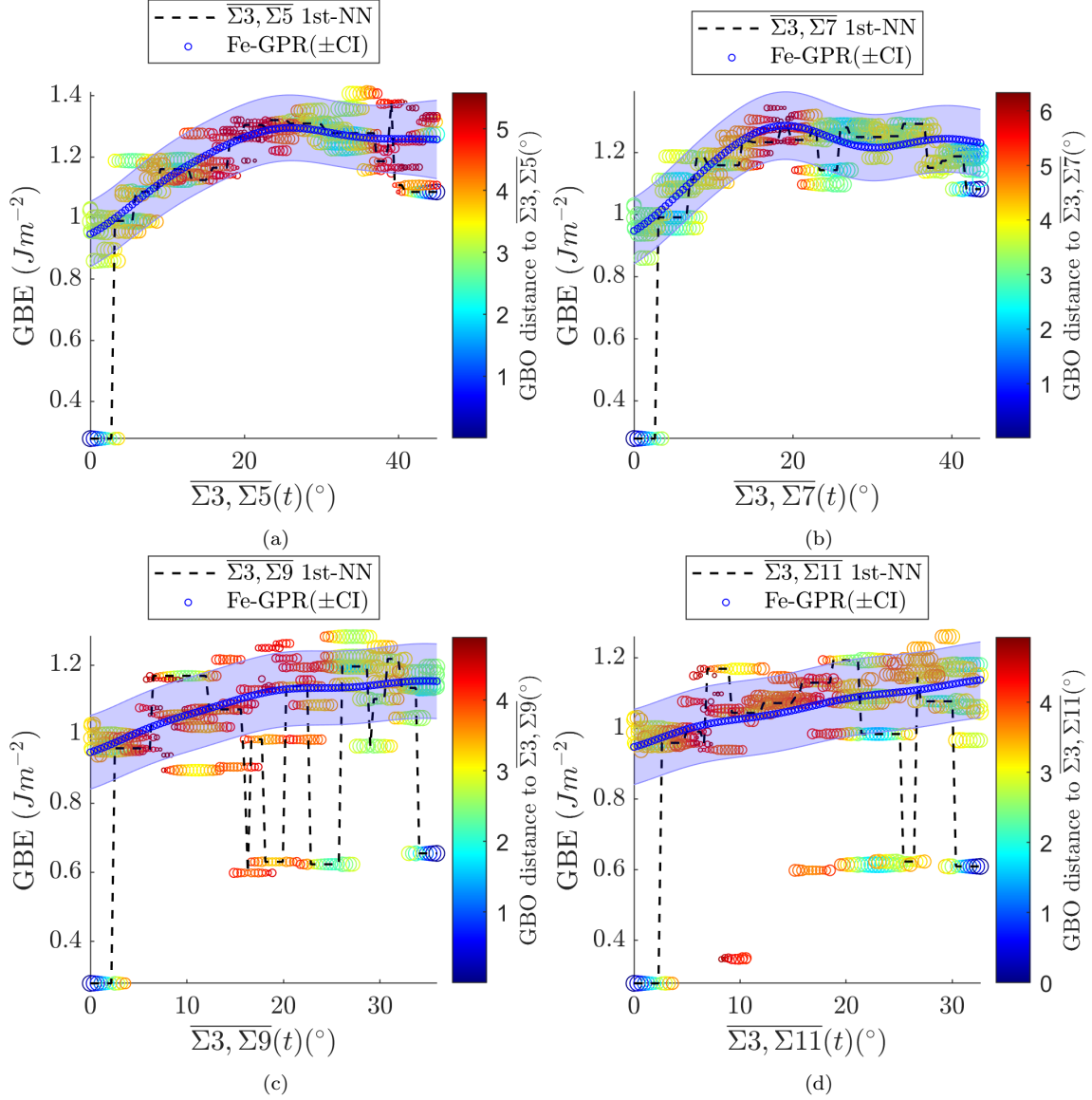


Figure 7: GBEs along geodesic paths in a VFZ between the minimum GBE $\Sigma 3$ and minimum GBE (a) $\Sigma 5$, (b) $\Sigma 7$, (c) $\Sigma 9$, and (d) $\Sigma 11$ GBs for the Fe Kim et al. [41] dataset. A GPR model trained on all 58 604 Fe Kim et al. [41] simulation datapoints was used. 95% confidence 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 geodesic 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.

dict whether and how fast a given crystallographic transformation might occur.

For the GPR model trained on the Ni Olmsted et al. [37] dataset, we observe that the path from $\Sigma 7$ to $\Sigma 3$ is strictly downhill in energy in Figure 6b while the geodesic paths between $\Sigma 3$ and $\Sigma 5$, $\Sigma 9$, $\Sigma 11$ are separated by energy barriers (Figures 6a, 6c and 6d). This indicates that for grain growth

systems governed by GBE, 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 are available (e.g. if both grain rotation and plane reorientation are possible, such as during high-temperature plastic deformation, recrystallization, or in nanocrystalline materials). In contrast, for a $\Sigma 5$, $\Sigma 9$, or $\Sigma 11$ GBs to transform

into $\Sigma 3$ coherent-twins, either an energy barrier must be overcome, or a different (perhaps more circuitous) trajectory through the 5DOF space must be taken.

For the GPR model trained on the Ni data (Figure 6), 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 $8\text{--}14^\circ$ relative to the geodesic 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, these trajectories traverse regions of the VFZ that are sparsely populated with data.

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 $\sim 6^\circ$ 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 S3.2) 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 can be estimated and used in grain growth simulations.

Because distance overestimations exist in the standard 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.

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 distance distributions are Gaussian and plotted these as a function of set size. Additionally, we found the GBE correlations to be Gaussian. We determined the maximum principal component of a particular O_h VFZ to be $\sim 65^\circ$.

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 be interesting to see the VFZ framework applied for other distance metrics (see Morawiec [45] 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 S3.1). 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 $\sim 1.9^\circ$. Plotting of geodesic paths between low-Sigma GBs of interest reveal that a $\Sigma 7$ cusp has a monotonically de-

creasing path towards the coherent-twin $\Sigma 3$ cusp, whereas a $\Sigma 5$, $\Sigma 9$, and $\Sigma 11$ cusps do not necessarily share this same type of monotonically decreasing path within 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

The authors thank Ian Chesser, Toby Francis, Victoria Baird, Brandon Snow, and José Niño for useful discussions. This work was supported by the National Science Foundation under Grant No. 1610077. This work was supported in part through computational resources provided by Brigham Young University’s Office of Research Computing.

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 1–4, 7, 9, 11, 13, 14

BRK Bulatov Reed Kumar 6, 7, 9–11

FZ fundamental zone 1, 13, 14

GB grain boundary 1–14

GBE grain boundary energy 1–3, 5–7, 9–13

GBO grain boundary octonion 1–4, 6, 7, 9, 10

GPR Gaussian process regression 5–7, 9–13

NN nearest neighbor 3–5, 8–13

VFZ Voronoi fundamental zone 1–7, 9–14

VFZ-GBO Voronoi fundamental zone grain boundary octonion 3–5, 8–10, 13

References

- [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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/10.1007/s11661-005-0333-7).

- [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](https://doi.org/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](https://doi.org/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](https://doi.org/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 $\langle 001 \rangle$ tilt grain boundaries in tin bicrystals, *Scripta Metallurgica* 18 (1984) 207–211. doi:[10.1016/0036-9748\(84\)90509-X](https://doi.org/10.1016/0036-9748(84)90509-X).
- [13] D. A. Molodov', U. Czubyko', G. Gottstein', L. S. Shvindlerman', MOBHYTY 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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/10.1016/j.actamat.2005.11.036).
- [24] S. Weissmann, On the Mechanism of Recrystallization of Aluminum, *Advances in X-Ray*

- Analysis 2 (1958/ed) 47–70. doi:[10.1154/S0376030800000422](https://doi.org/10.1154/S0376030800000422).
- [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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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:[10.1179/msc.1977.11.5.156](https://doi.org/10.1179/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:[10.1080/14786436708229736](https://doi.org/10.1080/14786436708229736).
- [31] M. Zelin, A. Mukherjee, Geometrical aspects of superplastic flow, *Materials Science and Engineering: A* 208 (1996) 210–225. doi:[10.1016/0921-5093\(95\)10080-6](https://doi.org/10.1016/0921-5093(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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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.2021.110756](https://doi.org/10.1016/j.commatsci.2021.110756).
- [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](https://doi.org/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](https://doi.org/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:[10.1088/0965-0393/22/3/034004](https://doi.org/10.1088/0965-0393/22/3/034004).

- [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](https://doi.org/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](https://doi.org/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](https://doi.org/10.1007/s10853-006-0665-8).
- [45] 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](https://doi.org/10.1016/j.actamat.2019.09.032).

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	1
S2 Semivariograms for Estimating Global and Local Correlation Lengths	2
S2.1 Semivariogram Method	2
S2.2 Global Correlation Lengths	3
S2.3 Local Correlation Lengths	4
S3 Fe Input Dataset Characteristics	4
S3.1 Methods and Results	7
S3.2 Intrinsic Uncertainty	7
S3.3 Overprediction Bias	8
S3.4 Improving on Existing Datasets	8
S4 Gridded Sampling for Numerical Differentiation	10
S5 GBs Used for Path Visualization	10
Glossary	10

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

*Corresponding author.

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

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 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 octonions¹ 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 energies (GBEs) 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

$$f(\mathbf{m}) \propto \exp \left(-\frac{1}{2}(\mathbf{m} - \mathbf{m}_0)^\top \mathbf{C}_{\mathbf{m}_0}^{-1}(\mathbf{m} - \mathbf{m}_0) \right) \quad (\text{S1})$$

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

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

(sometimes called the squared exponential) kernel:

$$\mathbf{C}_{\mathbf{m}_0}(x_i, x_j) = \sigma^2 e^{-\frac{(x_i - x_j)^\top (x_i - x_j)}{2l^2}} \quad (\text{S2})$$

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

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:

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

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 $E(x_j)$ 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

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

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

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

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-

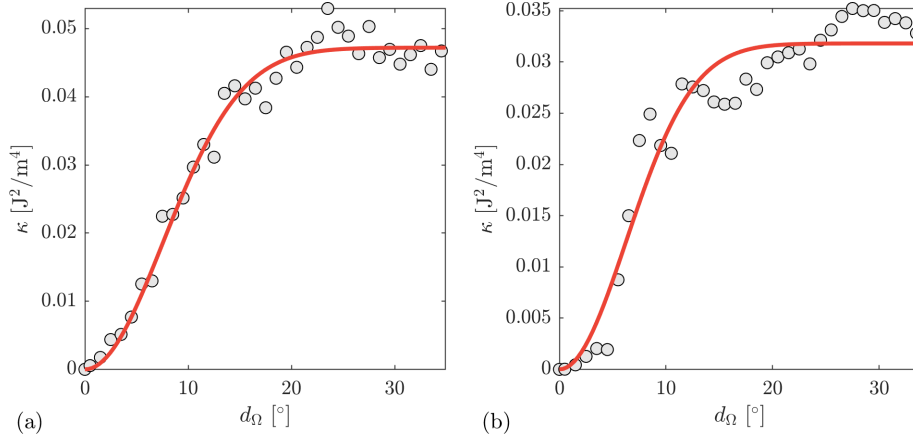


Figure S1: Empirical semivariograms (markers) for the (a) Ni and (b) Fe datasets. Solid lines show the fits of the analytical semivariogram models (Eq. S4).

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.

S2.3. Local Correlation Lengths

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 $\Sigma 5$ 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 Dataset Characteristics

We describe the methods and results used to evaluate the quality of the noisy Fe dataset (Section S3.1) and discuss intrinsic uncertainty (Section S3.2) and overprediction bias (Section S3.3). Finally, we offer suggestions on how to improve on existing datasets (Section S3.4).

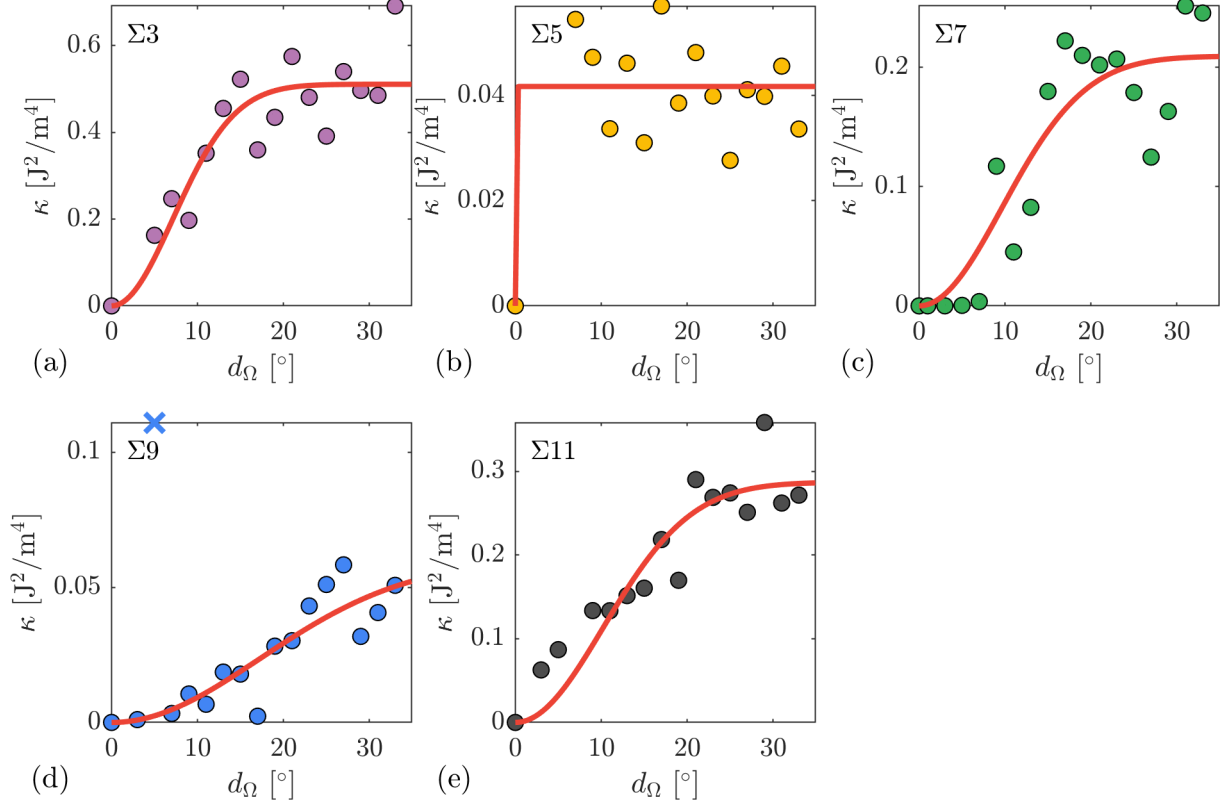


Figure S2: Local empirical semivariograms (markers), respectively centered at various low- Σ GBs for the Ni dataset. Solid lines show the fits of the analytical semivariogram models. In (d) the one point marked with an \times was considered an outlier and was excluded from the fit.

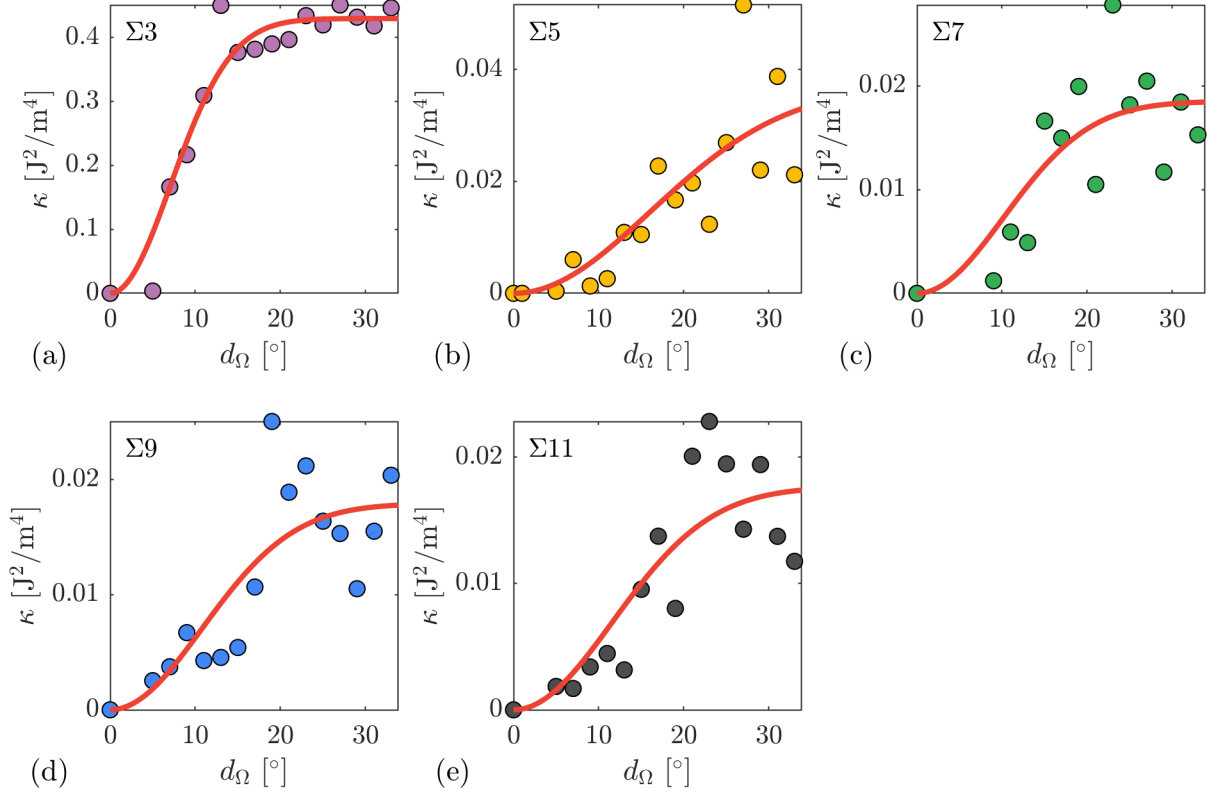


Figure S3: Local empirical semivariograms (markers), respectively centered at various low- Σ GBs for the Fe dataset. Keys for which GBs each of these correspond to in the original papers are given for Fe and Ni in [Tables S1](#) and [S2](#), respectively. Solid lines show the fits of the analytical semivariogram models.

S3.1. Methods and Results

Of the $\sim 60\,000$ GBs² 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 increments³ 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 sets⁴ (see Figure S4 for a degeneracy histogram). Thus, on average there is a degeneracy of approximately four per set of degenerate GBs.

By comparing GBE values of (unintentionally⁵) 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 J m^{-2} and 0.2625 J m^{-2} , respectively, indicating that a repeated Fe GB simulation from [4] can vary by as much as 0.525 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 $0.065\,29\text{ J m}^{-2}$ and $0.061\,90\text{ J m}^{-2}$, respectively, which is an approximate measure of the intrinsic 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 S3.2) 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 GBE 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).

S3.2. Intrinsic Uncertainty

We estimate the intrinsic uncertainty of an Fe 0 K molecular statics simulation dataset to be $0.065\,29\text{ J m}^{-2}$ and $0.061\,90\text{ J m}^{-2}$ 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 ($0.055\,035\text{ J m}^{-2}$) and mean absolute error ($0.039\,185\text{ J m}^{-2}$)

²The “no-boundary” GBs (i.e. GBs with close to 0 J m^{-2} GBE) were removed before testing for degeneracy.

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

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

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

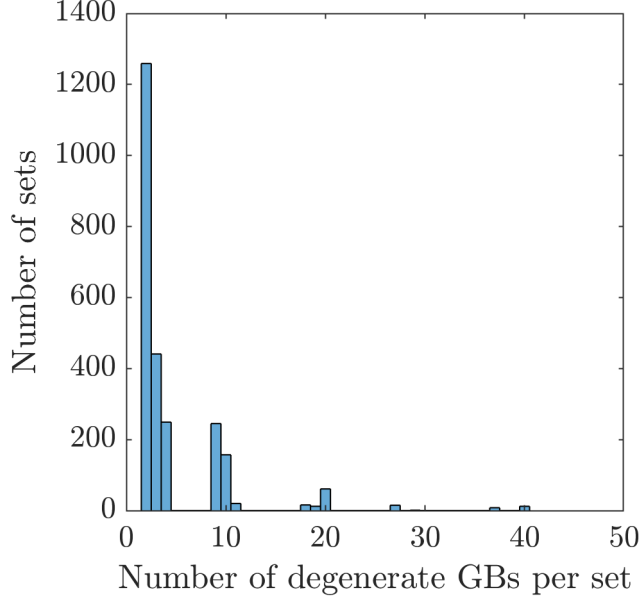


Figure S4: Histogram of number of sets vs. number of degenerate GBs per set for the Fe simulation dataset [4]. Most sets have a degeneracy of fewer than 5.

than the uncertainty estimates. This indicates that the uncertainty itself is somewhat overestimated⁶. The fact that both model and uncertainty metrics are relatively close and the prediction [1] 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”.

S3.3. 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 [6], not the minimum GBE blue curve of the same chart. See [6] for an in-depth treatment of equilibrium and metastable GBE.

S3.4. 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 Gaussian process regression model in predicting either

⁶The 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 Gaussian process regression models when the noise in the input data is approximately Gaussian).

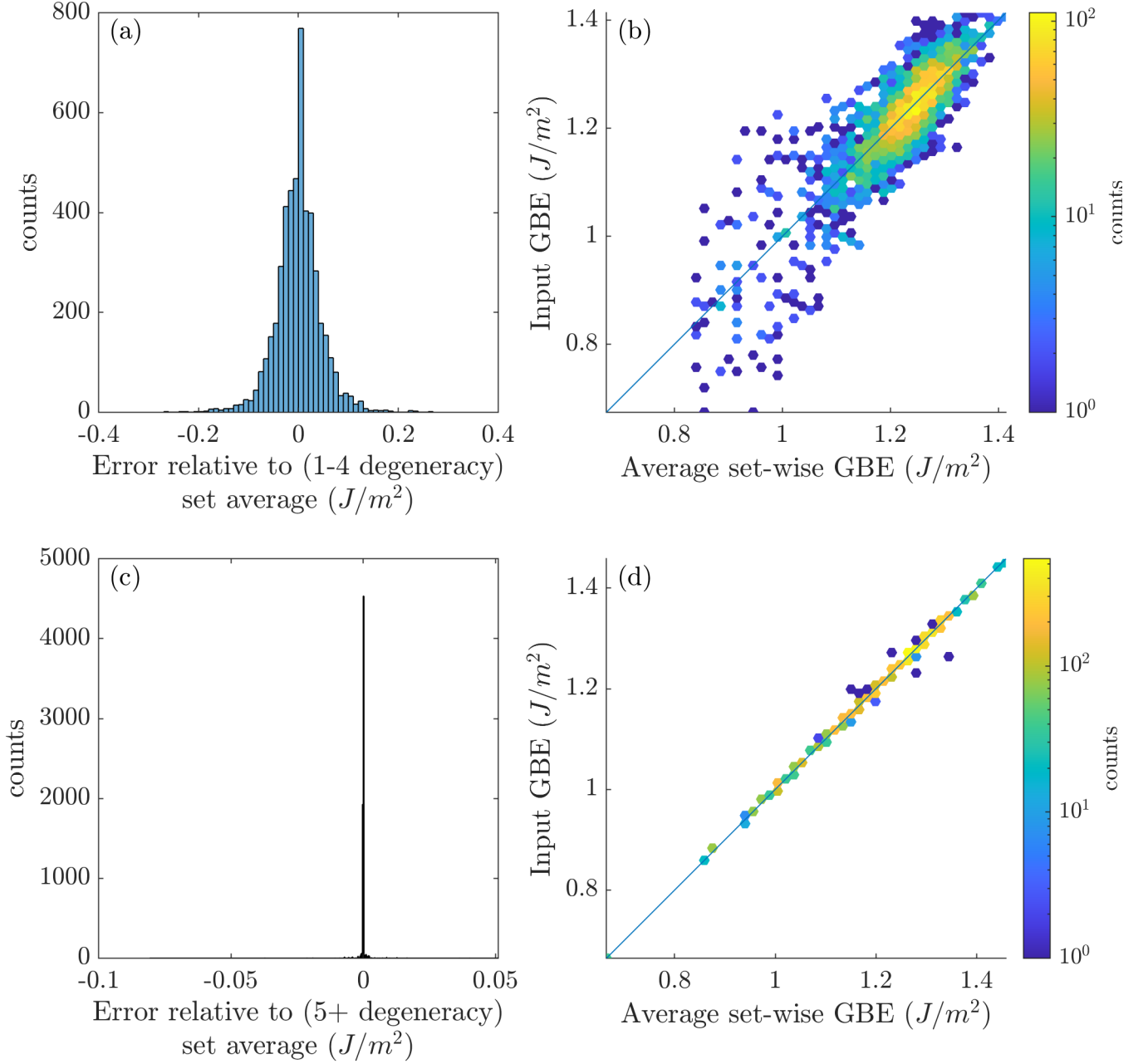


Figure S5: Degenerate GBs sets are split into those with a degeneracy of fewer than 5 and greater than or equal to 5 and plotted as (a) and (c), respectively) error histograms and (b) and (d), respectively) hexagonally-binned parity plots. Large degenerate sets tend to have very low error, whereas small 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.

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 [7].

S4. Gridded Sampling for Numerical Differentiation

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 dimensions⁷. Fortunately, most of 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 S1 and Table S2 for the Olmsted et al. [8] and Kim et al. [4] datasets, respectively.

Table S1: Minimum Σ (Sigma) GBs and corresponding IDs used for path visualization within the original Olmsted et al. [8] dataset.

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

Glossary

GB grain boundary 1, 2, 4, 6–11

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

Table S2: 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

GBE grain boundary energy [2](#), [7](#), [8](#), [10](#)

GBO grain boundary octonion [1](#), [2](#)

VFZ Voronoi fundamental zone [1-3](#), [7](#), [10](#)

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](https://doi.org/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](https://doi.org/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](https://doi.org/10.1016/j.scriptamat.2011.03.020).
- [6] 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](https://doi.org/10.1016/j.actamat.2015.11.035).
- [7] 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](https://doi.org/10.1038/s41563-020-00879-z).
- [8] 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](https://doi.org/10.1016/j.actamat.2009.04.015).