

# Five Degree-of-Freedom Property Interpolation of Arbitrary Grain Boundaries via Voronoi Fundamental Zone Framework

Sterling G. Baird<sup>a,\*</sup>, Eric R. Homer<sup>a</sup>, David T. Fullwood<sup>a</sup>, Oliver K. Johnson<sup>a</sup>

<sup>a</sup>*Department of Mechanical Engineering, Brigham Young University, Provo, UT 84602, USA*

---

## Abstract

We introduce the Voronoi fundamental zone (VFZ) framework which is useful for grain boundary (GB) structure-property models and gaining insights about the nature of a five degree-of-freedom (5DOF) fundamental zone (FZ) for both cubic and non-cubic symmetries and potentially alloys. We cover the methods associated with the VFZ framework. The VFZ framework offers an advantage over other 5DOF based property interpolation methods because directly computed Euclidean distances approximate the original grain boundary octonion (GBO) distance with significantly reduced runtime ( $\sim 7$  CPU minutes vs. 153 CPU days for a  $50\,000 \times 50\,000$  pairwise-distance matrix). We perform grain boundary energy (GBE) interpolation for a non-smooth validation function on sets of up to 50 000 GBs using four interpolation methods: barycentric interpolation, Gaussian process regression (GPR) or Kriging, inverse-distance weighting (IDW), and nearest neighbor (NN) interpolation. The best performance was achieved with GPR which results in a reduction of the root mean square error (RMSE) by 83.0% relative to RMSE of a constant, average model. This error is comparable to the minimum expected uncertainty associated with reconstruction of noise-free experimental polycrystalline data. We then use GPR to interpolate simulated bi-crystal datasets for Fe and Ni and demonstrate better than (34.4% vs. 21.2% improvement) and similar (57.6% vs. 56.4% improvement) performance to prior work, respectively. The noise and non-uniform sampling in the Fe dataset make it difficult to resolve low GBE (i.e. cusps) and to validate the model. We resolve the Ni dataset cusps with high accuracy, but uncertainty may be high in regions far from the input data. The trade-offs between noise, dataset size, sampling scheme, and repeat measurements must be carefully managed. We provide a vectorized, parallelized, MATLAB interpolation function (`interp5DOF.m`) and related routines ([github.com/sgbaird-5dof/interp](https://github.com/sgbaird-5dof/interp)) which can be applied to future datasets for a variety of GB properties and to better understand 5DOF FZs. For example, we find that the maximum dimension of an  $O_h$  VFZ is  $\sim 66.5^\circ$ .

*Keywords:* Grain Boundary, Structure-Property Model, Interpolation, Octonion, Machine Learning

---

## 1. Introduction

### 1.1. Motivation

High fidelity GB structure-property models can accelerate the design and understanding of materials for GB engineering applications such as

grain growth (GBE [1], mobility [2], and grain rotation [3–6]), stress-corrosion cracking (diffusivity [7, 8], solubility [9], and segregation [10]) [11–15], strength [16–18], ceramics [19, 20], electronics [21, 22], and thermoelectrics [23]. With the increased use of nanomaterials [16, 22], GBs take on increasingly larger roles as the GB volume fraction becomes significant; this is complicated by the fact that properties of GBs can span orders of magnitude depending on the five macroscopic degrees of

---

\*Corresponding author.

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

freedom [24–26] as well as the three microscopic degrees of freedom [27, 28]. However, the mentioned studies generally only consider a binary classification of GBs or variation of a few degrees of freedom which represents a small “slice” of the full grain boundary character (GBC) space. Recent advances in high-throughput simulation [26, 29–35], experimental characterization [26, 34, 36–38], and availability of rich GB datasets [33, 35, 39–47] warrant high-fidelity structure-property models capable of handling large amounts of input data to aid in the aforementioned applications.

### 1.2. Prior Work

In prior work, a number of strategies have been developed for predicting<sup>1</sup> 5DOF GB properties from experimental or simulated data. Because different works use different validation functions and data, it is difficult to objectively compare their performance. To facilitate meaningful comparisons, in addition to quoting absolute performance in terms of RMSE or mean absolute error (MAE), we will also report the percent reduction in error compared to a constant-valued control model whose value is chosen as the mean of the respective input data.

Detle et al. [48] builds on the work of Bulatov et al. [49] by predicting the location and number of cusps rather than assuming this a-priori, and using sequential sampling (a popular design of experiments technique often combined with machine learning) to provide better accuracy with fewer datapoints. Additionally, they provide an equation for uncertainty of predicted values (Equation 7 of [48]) and suggest that their approach is general to all alloys and non-cubic crystal systems. While their proposed methods are applicable to 5DOF (as a consequence of being based on Bulatov et al. [49]), they only present results in 1D and 2D subspaces and to the authors’ knowledge do not have a publicly available implementation (i.e. codebase).

Several researchers have taken the approach of discretizing unsymmetrized 5DOF GBC space, and then using a least squares objective function and

gradient descent to fit a piecewise-constant function, resulting in 5DOF grain boundary energy distributions for nickel [41], yttria [50], and copper [44] based on experimentally characterized 3D microstructures. Additionally, Fortran scripts to perform the processing and analysis are available at [http://mimp.materials.cmu.edu/~gr20/Grain\\_Boundary\\_Data\\_Archive/](http://mimp.materials.cmu.edu/~gr20/Grain_Boundary_Data_Archive/). In a follow-up approach, Shen et al. [51] provided a non-parametric reconstruction technique of polycrystalline data and corresponding Python/C++ code (<https://github.com/Yufeng-shen/TJ2GBE>) that provides a higher fidelity estimation of GBE. They suggest that this technique can then be paired with a 5DOF interpolation technique<sup>2</sup>.

Restrepo et al. [52] used an artificial neural network (ANN) and approximately 17 000 and 51 000 Fe bicrystal simulations from Kim et al. [39] as training and validation data, respectively, to achieve MAEs of  $0.0486 \text{ J m}^{-2}$  and approximately  $0.09 \text{ J m}^{-2}$  in the best fitted ANNs for randomly selected and special GBs, respectively. If a constant, average value (i.e. average of the input GBEs) was chosen as the model, the MAE would be  $0.0617 \text{ J m}^{-2}$ , implying that predictions of randomly selected GBs were improved by 21.2% relative to this simple, control model. Others have combined machine learning approaches with large lists of macroscopic and microscopic descriptors [53, 54].

Recently, a new GB representation, GBOs, was reported [55] and tested [56]. The GBO representation is valuable for a number of applications. Most relevant to the present work is the resulting distance metric. The GBO distance metric offers an advantage over other metrics in that it “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” [55]. In this context, Francis et al. [55] derived octonion Spherical Linear Inter-

<sup>1</sup>We use the term “predict” throughout this work to refer to interpolation, inference, and/or extrapolation as some approaches can individually involve multiple prediction types.

<sup>2</sup>We believe the VFZ interpolation methods introduced in our work to be especially applicable.

polation (oSLERP) and provided examples showing that oSLERP produces smooth, minimum distance paths through GB character space between two arbitrary GBs. Additionally, MATLAB and EMSOFT implementations were made available at [https://github.com/ichesser/GB\\_octonion\\_code](https://github.com/ichesser/GB_octonion_code) and <https://github.com/EMsoft-org/EMsoft>, respectively.

Laplacian kernel regression (similar to IDW) involving scaled pairwise distance matrices was later used with GBOs to predict properties of arbitrary GBs from a set of known values [56]. Using  $k$ -fold cross validation with  $k = 10$  for 388 Ni GBE simulations [43] and an optimized scaling parameter, a RMSE of  $0.0977 \text{ J m}^{-2}$  was obtained compared to a constant, average model RMSE of  $0.2243 \text{ J m}^{-2}$  (56.4% improvement). Due to computation time of pairwise distance matrices, this approach is currently “limited to datasets with several thousand or fewer” GBs [56].

### 1.3. Voronoi Fundamental Zone Framework

The VFZ interpolation framework introduced in this work offers an advantage over other methods because it is defined as a VFZ point set in a manifold<sup>3</sup> for which directly computed, scaled Euclidean distances approximate the original GBO distance given by Francis et al. [55]. This advantage is manifest in the ability to triangulate a mesh using standard routines (e.g. quickhull [58]) and interpolate using barycentric coordinates or machine learning methods such as GPR. Building on prior work on GBOs [55, 56], we create a VFZ point set by obtaining a set of GBOs minimized with respect to Euclidean distance and an arbitrary reference GBO after considering all symmetrically equivalent octonions (SEOs). Because GBOs are guaranteed to reside on the surface of a hypersphere [55] (a type of Riemannian manifold) a point set which locally resembles Euclidean space is the result (Section 2.1.3). Our implementation of the VFZ strategy utilizes GBOs; however, it may be possible to

apply the same strategy using other distance metrics that have been proposed in the literature (for a comparison of many different metrics, see Morawiec [57]). Below we provide the detailed description of our methods (Section 2), followed by numerical test results (Section 3).

We also provide a vectorized, parallelized implementation of the VFZ framework and related functions. These are contained in what we will refer to as the VFZ repository, which is available at [github.com/sgbaird-5dof/interp](https://github.com/sgbaird-5dof/interp). In what follows, when we refer to built-in MATLAB functions, we refer to them with parentheses as in `interp1()`. When we refer to functions in the VFZ repository, we do so with the `.m` extension as in `interp5DOF.m` unless specifying the usage with arguments as in `interp5DOF(qm,nA,qm2,nA2,y)`.

## 2. Methods

We describe methods related to the VFZ framework (Section 2.1.1), generation of random GBs (Section S0.2), and four different GB property interpolation schemes (Section 2.2). We also describe details regarding two simulated literature datasets that we use (Section 2.3).

### 2.1. The Voronoi Fundamental Zone Framework

The core operations of the VFZ framework are:

1. generating GBOs (Section 2.1.1)
2. mapping GBOs into a VFZ (Section 2.1.2)
3. calculating distances within the VFZ (Section 2.1.3)

#### 2.1.1. Defining the Voronoi Fundamental Zone

Three degree-of-freedom FZs have typically been defined using linear inequalities (e.g. the orientation [59] and misorientation [59, 60] FZs). Instead of using linear inequalities<sup>4</sup>, we take a numerical approach to define what we will call a VFZ.

<sup>3</sup>“In mathematics, a manifold is a topological space that locally resembles Euclidean space” [57]. By removing the Euclidean approximation in the VFZ framework and the renormalization of GBOs, the metric is intrinsic [57].

<sup>4</sup>If desired, linear inequalities can be obtained for a VFZ by determining a Voronoi tessellation’s junction points (similar to what is shown in Figure 1 by e.g. `voronoin()`), transforming to 6D Cartesian coordinates via a singular value decomposition transformation and defining the bounded region by e.g. MATLAB FEX function `vert2lcon.m`.

To define a VFZ, an arbitrary, fixed, low-symmetry reference GBO is chosen ( $o_{\text{ref}}$ ) 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  $o_{\text{ref}}$  than any of its symmetric images<sup>5</sup>. However, use of the VFZ does not require its explicit construction. Rather, practical calculations require only the selection of the single point  $o_{\text{ref}}$  (which completes the definition of the VFZ), followed by mapping of query points into the VFZ by comparison of their SEOs with  $o_{\text{ref}}$ .

To illustrate the process of mapping points into the VFZ, we describe a 3D Cartesian analogue (Figure 1) to a 7D Cartesian non-degenerate (i.e. U(1) degeneracy removed) representation of a VFZ, which produces the standard stereographic triangle [61] for the  $O_h$  point group. A set of 500 points ( $p_i, i \in [1, 500]$ ) randomly scattered on the surface of the 2-sphere comprise the data (red points in Figure 1a). A random point,  $p_{\text{ref}}$ , also on the surface of the 2-sphere, is chosen as the reference point (white circle). In this illustration,  $O_h$  or  $m\bar{3}m$  point group rotations are used as symmetry operators,  $S_j, j \in [1, N_p]$ , where  $N_p$  is the twice the number of proper rotations due to inversion symmetry and  $N_p = 48$  for the  $O_h$  point group. For each data point, 48 symmetrically equivalent representations ( $p_{i,j}^{\text{sym}} = S_j(p_i), j \in [1, 24]$ ) are produced by applying each of the relevant symmetry operators. After calculating the Euclidean distance between  $p_{\text{ref}}$  and  $p_{i,j}^{\text{sym}}$ , the point ( $p_i^*$ ) closest to  $p_{\text{ref}}$  is chosen and retained as the unique representative of  $p_{i,j}^{\text{sym}}$ . As illustrated in Figure 1a, the projected points  $p_i^*$  (dark blue points) all fall in the VFZ without ever having to construct or define it analytically. We call this group of projected points a *VFZ point set*. Note also that there is only one  $p_i^*$  in the VFZ for each  $p_{i,j}^{\text{sym}}$  (see Figure 1b).

<sup>5</sup>We also refer to lower-dimensional representations of the 8D Cartesian VFZ as VFZs (described in Section 2.1.2) and describe which dimensionality we are referring to as appropriate.

To calculate<sup>6</sup> the distance between a given GBO, and the reference GBO, we employ the standard 8D Euclidean distance

$$d_E(o_A, o_B) = \left( \sum_{k=1}^8 (o_{A,k} - o_{B,k})^2 \right)^{1/2} \quad (1)$$

where  $o_{A,k}$  and  $o_{B,k}$  represent the  $k$ -th element of normalized GBOs  $o_A$ , and  $o_B$ , respectively.

Euclidean distance is an approximation to the true geodesic arc length on  $\mathbb{S}^7$ , which is given by

$$d_S(o_A, o_B) = \cos^{-1}(o_A \cdot o_B) \quad (2)$$

where  $\cdot$  is the dot product,  $\cos^{-1}$  is the inverse cosine operator, and  $o_A$  and  $o_B$  are each normalized and  $d_S \simeq d_E$  (Figure S1). In [55], the original GBO distance metric was defined by

$$d_\Omega(o_A, o_B) = 2 \cos^{-1}(o_A \cdot o_B) \quad (3)$$

where  $o_A$  and  $o_B$  are each normalized and  $d_\Omega$  can be seen to be simply twice the geodesic arc length:  $d_\Omega = 2d_S$ . Thus,  $d_E \simeq \frac{1}{2}d_\Omega$ .

The definition of  $d_\Omega$  has certain aesthetic benefits in that it mirrors the definition of a misorientation angle,  $\omega_{AB}$ , between two crystal orientations in the quaternion parameterization:  $\omega_{AB} = 2 \cos^{-1}(q_A \cdot q_B)$ .

Our choice to use  $d_E$  instead of  $d_S$  or  $d_\Omega$  is motivated by the fact that it enables the use of standard algorithms, for a variety of operations, that require or assume Euclidean distances. See Section S1 for further justification of this choice.

For applications other than interpolation which require precise quantification of high-dimensional volume, a mapping between Euclidean-approximated volumes and true volumes

<sup>6</sup>VFZ-GBO Euclidean, hyperspherical arc length, and GBO distances are computed via VFZ repository function `GBdist4.m` which is used in the symmetrization function `get_octpairs.m` and an example of ensemble VFZ-GBO distance calculations is given in `plotting.m`.



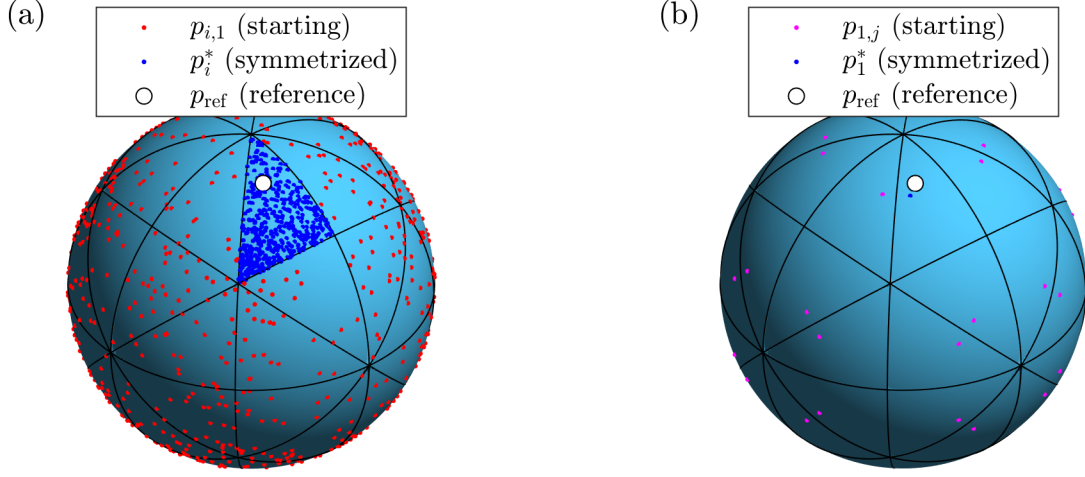


Figure 1: (a) 3D Cartesian analogue to a non-degenerate 7D Cartesian representation of U(1)-symmetrized GBOs and Voronoi fundamental zone grain boundary octonions (VFZ-GBOs) (VFZ-GBOs are inherently U(1)-symmetrized) which demonstrates the symmetrization of many points relative to a fixed reference point (white circle). This produces a 3D Cartesian VFZ point set (dark blue points). (b) To further illustrate, 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) of each of the Voronoi cells (light blue). The Voronoi tessellation is defined by the symmetric images of the reference point, and the spherical Voronoi diagram for this illustration is constructed using a modified version of [62].

may be necessary<sup>7</sup> or the Euclidean approximation may be removed altogether<sup>8</sup>. The latter combined with removal of the renormalization of GBO allows for the (non-ensembled) VFZ-GBO metric to be intrinsic (see Morawiec [57] for an in-depth treatment of intrinsicity).

The expectation that a single, unique SEO will be found (within numerical tolerance and given a low-symmetry reference GBO<sup>9</sup>) is verified by several manual tests. If multiple SEOs are found, a warning is issued internally within the symmetrization sub-routine `get_octpairs.m` [64] that is part of the `interp5D0F.m` package. Similar numerical tests reveal that inappropriately selecting a high-symmetry reference GBO to (attempt to) define a VFZ results in many degenerate minimum distance SEOs, with the identity ( $\{1, 0, 0, 0, 0, 0, 0, 0\} \in \mathbb{R}^8$ )

<sup>7</sup>We have not tested to what extent a Euclidean-approximated volume will differ from the true volume; however, Euclidean-approximated volumes can be obtained by using triangulation methods [63] (i.e. via `convhulln()`).

<sup>8</sup>i.e. by setting the `dtype` argument of `GBdist4.m` to 'omega' rather than 'norm'.

<sup>9</sup>The probability that a randomly generated GBO will fall exactly on a high-symmetry boundary vanishes in the limit of infinite precision.

[55] giving the highest degeneracy.

### 2.1.2. Mapping GBs to the Voronoi Fundamental Zone

As described above in the 3D analogy, with a reference GBO chosen ( $o_{\text{ref}}$ ), and consequently the VFZ defined (Section 2.1.1), a GBO is mapped into the VFZ by finding among its SEOs the one that is closest to  $o_{\text{ref}}$  according to  $d_E$  (Eq. (1)). This is performed for all input and prediction points with respect to  $o_{\text{ref}}$ , and the result is a VFZ input and prediction point set.

### 2.1.3. Distance Calculations in the Voronoi Fundamental Zone

Euclidean distances are an accurate approximation of arc length distances in a VFZ as shown in Figure S1. This also demonstrates that the maximum dimension of an  $O_h$  VFZ is  $\sim 66.5^\circ$ .

However, when compared with the traditional GBO distance [55], due to the presence of low-symmetry GBs near the exterior of a VFZ, some GB pairs will exhibit larger Euclidean or arc length distances than is truly representative (see e.g. Figure 2a). In other words, moving “past” the low-symmetry border of a VFZ will result in an instan-

taneous relocation to a possibly distant point in the VFZ that in reality is highly correlated.

This is a limitation of the VFZ framework, which generates a VFZ with low-symmetry GBs at the borders in contrast to typical FZs [61, 65]. While defining a FZ with high-symmetry GBs at the borders (especially mirror-symmetry GBs) will certainly increase interpolation accuracy, the favorable interpolation results presented in this work are obtained because overestimation is infrequent within a small correlation length (e.g.  $10^\circ$  [42]). We expect that smaller datasets will exhibit more frequent overestimation than larger datasets. See [Section S4](#) for details on how overestimation imposes erroneous “sparseness” and how this limitation can be overcome through ensemble or data augmentation methods.

In addition to their use for distance calculations alone, ensembles of VFZ-GBO sets can be employed with interpolation methods to increase overall interpolation accuracy, but there is a computational cost (e.g. approximately  $10\times$  using an ensemble of 10 VFZ-GBO sets). For 50 000 input points, use of an ensemble with 10 VFZ-GBO sets decreases RMSE and MAE from  $0.0241 \text{ J m}^{-2}$  and  $0.0160 \text{ J m}^{-2}$  to  $0.0187 \text{ J m}^{-2}$  and  $0.0116 \text{ J m}^{-2}$ , respectively (single trial run). We expect these overall accuracy improvements occur because GBE predictions near the exterior of the VFZ where data may be sparse are improved. Ensemble interpolation results as a function of ensemble size and parity plots for mean, median, minimum, and maximum functions applied to the ensemble are shown in [Figure S5](#) and [Figure S6](#), respectively. Further details of ensemble interpolation are given in [Section S5](#).

#### 2.1.4. Comparison with Traditional Octonion Framework

We summarize differences between the VFZ framework and the traditional GBO metric ([Table 1](#)) and give examples that illustrate the computational complexity of each approach.

The construction of the VFZ dramatically reduces the computational burden of pairwise distance calculations. We illustrate the mechanism by which this reduction is achieved with an example in [Section S2](#).

## 2.2. Interpolation in the Voronoi Fundamental Zone Framework

With the VFZ framework established, it is possible to define interpolation schemes over the VFZ to predict properties of new GBs from the known properties of other GBs. For one application of interest to us, it is necessary to evaluate multiple different functions over a fixed set of input and prediction GBOs. We first summarize a barycentric interpolation method that we developed [63] to efficiently accomplish this specialized task by pre-computing the interpolation weights (which remain fixed when only the function being evaluated changes). We then present adaptations of three other interpolation methods—GPR ([Section 2.2.2](#)), IDW ([Section 2.2.3](#)), NN ([Section 2.2.4](#))—that are useful for general applications (an additional interpolation method—Gaussian process regression mixture (GPRM)—which we developed specifically for a non-uniformly distributed, noisy, simulation dataset is described in [Section S6.3](#)). Usage instructions for the VFZ-GBO repository can be found at the GitHub page ([github.com/sgbaird-5dof/interp](https://github.com/sgbaird-5dof/interp)) and in [Section S0.1](#).

### 2.2.1. Barycentric Interpolation

Barycentric coordinates are a coordinate system that locates a point within a simplex by assigning “masses” or weights to the corners of the simplex such that the point of interest is the barycenter (i.e. center of mass). Interpolation can then be performed by taking a dot product of each of the corner weights with the corner properties. We provided a detailed description of our implementation of barycentric interpolation for the 7-sphere (especially w.r.t. GBOs) in Baird et al. [63].

### 2.2.2. Gaussian Process Regression

GPR or Kriging uses the notion of similarity between points to fit Gaussian processes (random variables) to data based on prior information and provides uncertainty information in addition to interpolated or inferred values. For a general treatment of GPR, see Rasmussen and Williams [66]. We use MATLAB’s built-in func-

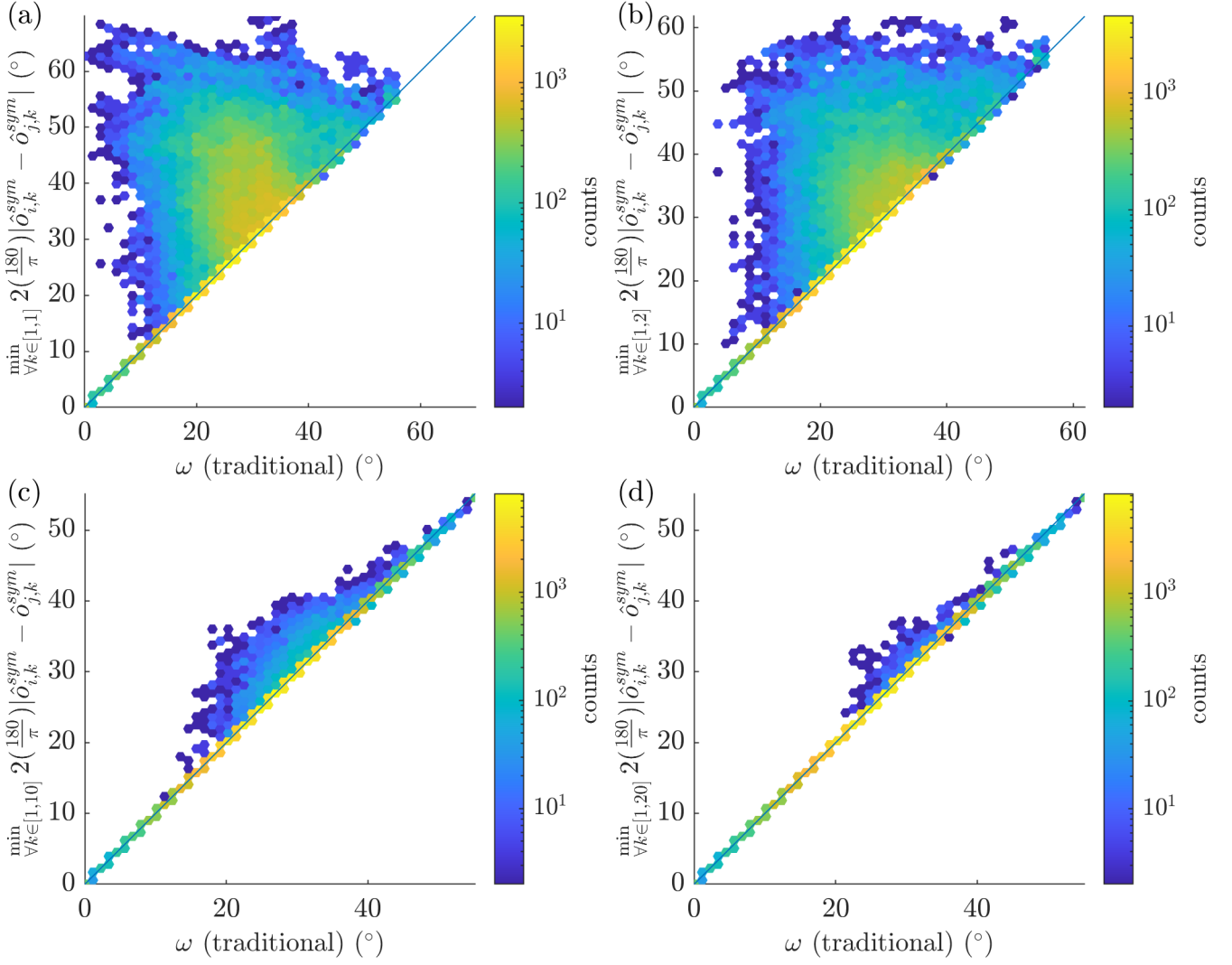


Figure 2: Hexagonally binned parity plots of pairwise distances of 388 Ni bicrystals [43]. Euclidean distance approximation is converted to GBOs ( $x_{i,j,k} = 2 \left( \frac{180}{\pi} \right) |\hat{o}_{i,k}^{\text{sym}} - \hat{o}_{j,k}^{\text{sym}}|$ ) for comparison with the traditional GBO metric [56]. The minimum distance among an ensemble of VFZ-GBO sets ( $\min_{k \in [1, k_{\text{max}}]} x_{i,j,k}$ ) is used for (a) 1, (b) 2, (c) 10, and (d) 20 VFZ-GBO sets. As the number of VFZ-GBO sets increases, the correlation between the Euclidean distance and the traditional GBO distance improves.

tion, `fitrgp()`, with all default parameters<sup>10</sup> except that a fully independent conditional approximation is used (`PredictMethod = 'fic'`) regardless of the number of input points. We assume a Euclidean approximation of the VFZ (see Section 2.1.3 and Figure S1). A slower, more accurate, and more memory-intensive prediction method that doesn't use sparse approximation

(`PredictMethod = 'exact'`) is also available (Section 3.2).

### 2.2.3. Inverse-distance Weighting Interpolation

IDW interpolation applies a weighted average to points within a neighborhood of a query point to obtain an interpolated value. `interp5DOF.m` implements a simple IDW approach based on [67]. A default radius of influence of  $r = \sqrt{2}\mu$  is used, where  $\mu$  represents the mean NN distance, and where GBO distance is approximated by the Euclidean distance or 2-norm (see Section 2.1.3, and

<sup>10</sup>MATLAB R2020b was used for the Fe simulation dataset, all other results employed MATLAB R2019b, the latest installed version on our computing cluster.

Table 1: Comparison between Voronoi fundamental zone grain boundary octonion and traditional GBO frameworks. \*6D Cartesian representation used only for mesh triangulation efficiency in barycentric interpolation and \*7D Cartesian representation only required for barycentric interpolation. 7D Cartesian representation is also implemented (though not required) for GPR, NN, and IDW. For pairwise distance complexity,  $N_p$  is the number of proper rotations ( $N_p = 24$  for  $m\bar{3}m$  face-centered cubic point group) and  $L$  is the number of GBs.

Property	Traditional	This Work
Symmetrizing Distance	GBO	VFZ Euclidean
Dimensionality	8D Cartesian	6*/7*/8D Cartesian
Bounded by FZ	No	Yes
Pairwise Distance Complexity	$O(N_p^2 L^2)$	$O(N_p^2 L)$
Rotation Convention	Passive	Active

Figure S1). NN interpolation (Section 2.2.4) is used for a given query point when there are no input points in the radius of influence.

#### 2.2.4. Nearest Neighbor Interpolation

NN interpolation takes the nearest input point relative to a query point and assigns the value of the NN input point to the query point. This is implemented via the built-in MATLAB function `dsearchn()` using a Euclidean approximation of GBO distance (see Section 2.1.3, and Figure S1).

#### 2.3. Literature Datasets

In addition to performing validation tests of the VFZ framework, we also describe results in which we apply it to actual GB property data available from literature sources [42, 68]. We describe details related to the retrieval and processing of two molecular statics (MS) simulation datasets from the literature. We describe GPR applied to Fe (Section S6.1) and Ni (Section S6.2) simulations, as well as a specialized GPRM model applied to Fe to address non-uniformity and noise concerns (Section S6.3).

### 3. Results and Discussion

To illustrate the utility of the VFZ framework for one application, namely interpolation, we compare the (i) accuracy (Section 3.1), and (ii) efficiency (Section 3.2) of the four previously described interpolation methods implemented over the VFZ with each other and with existing methods from the literature (see Section 1). For these tests, we use the

5DOF GB energy function by Bulatov et al. [49] (trained on Ni bicrystal simulation data [42]) as a validation function which we refer to as the Bulatov Reed Kumar (BRK) function.

Following this validation study, we also demonstrate VFZ-GBO GPR interpolation applied to a large, noisy, MS Fe bicrystal simulation dataset [68] and a small, low-noise, MS Ni bicrystal simulation dataset [42] (Section 3.4.1), to evaluate performance on real GB property data.

#### 3.1. Interpolation Accuracy

Accuracy of GPR, barycentric, NN, and IDW interpolation methods are given w.r.t. the BRK validation function (Section 3.1.1). Context is given to these error metrics through comparison with a constant-valued control model (Section 3.1.2) and the uncertainty associated with experimental and simulated datasets (Section 3.1.3).

##### 3.1.1. Accuracy of Four Interpolation Methods

Figure 4 provides hexagonally binned parity plots (`parityplot.m` via modified version of [69]) for each of the four interpolation methods using 50 000 input GBs. Results for 388 and 10 000 GBs are given in Figure S3 and Figure S4, respectively.

All of the methods permit successful interpolation, and the highest density region in all cases falls squarely on the parity line. The GPR and barycentric results show a slight asymmetry such that low energy values are overpredicted more often than they are underpredicted. The width of



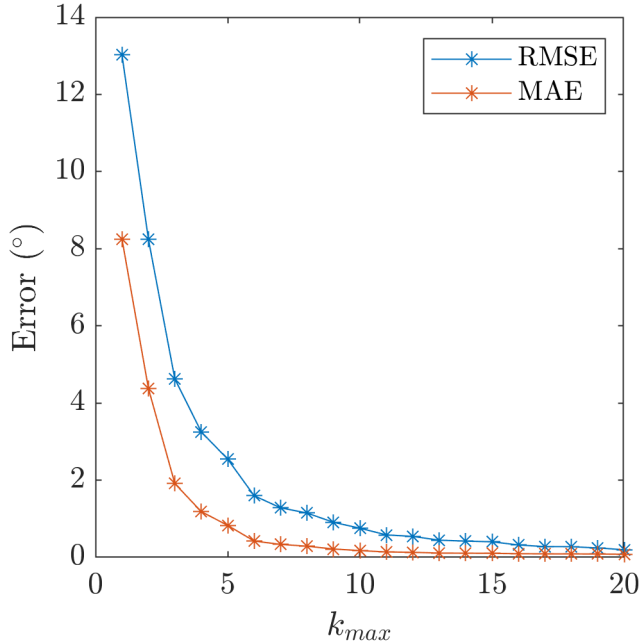


Figure 3: RMSE and MAE of pairwise distance errors for 388 Ni bicrystals [42] of scaled Euclidean distance approximation relative to the traditional GBO metric [56] (compare with Figure 2). The minimum distance among an ensemble of VFZ-GBO sets ( $\min_{\forall k \in [1, k_{max}]} x_{i,j,k}$ , where  $x_{i,j,k}$  is the scaled Euclidean distance) is taken, iteratively adding consecutive sets up to  $k_{max} = 20$ . As the number of VFZ-GBO sets increases, RMSE and MAE between the scaled Euclidean distance approximation and the traditional GBO distance decreases.

the point clouds provides a qualitative indication of the dispersion in the prediction errors, and the logarithmically scaled color indicates the frequency of errors of a given magnitude. As can be seen, the vast majority of errors are very small (the highest density—yellow region—is concentrated on the line of parity). Quantitative measures of the overall accuracy are presented for RMSE (Table 2) and MAE (Table 3), and will be discussed in detail below (see `get_errmetrics.m`).

As shown in Tables 2 and 3, of the four interpolation methods from this work, GPR has the lowest error, both in terms of RMSE and MAE, while NN has the highest error. Compared to a constant valued control model, GPR interpolation reduced the prediction RMSE by 83 %, which outperforms all of the interpolation methods in this work with respect to accuracy, as well as those considered from the literature. After GPR the next most accu-

rate methods are barycentric, IDW, and NN. We also note that the RMSE interpolation error for the GPR and barycentric methods is comparable to the minimum achievable noise-free experimental interpolation error which is the estimated error in experimental data (Section 3.1.3).

The accuracy of the predictions made using the VFZ methods depends on the VFZ-GBO set size and distribution. Figure 5 compares the prediction accuracy for each of the 4 methods to the constant valued control model, as a function of the number of input VFZ-GBOs (`ninputpts`). As expected, higher density VFZ-GBO sets result in lower error, but eventually give diminishing returns. Moreover, the standard deviations produced via multiple runs are tightly constrained and generally shrink as the VFZ-GBO set size increases.

GPR consistently gives lower error than the other three interpolation methods for all VFZ-GBO set sizes. NN interpolation produces the worst error of the four methods, but is better than a constant valued control model (i.e. average of the input GBEs) so long as `ninputpts` exceeds a few hundred input points.

It is worthwhile to note that both GPR and IDW are kernel-based in that a model parameter controls the size of the region that can influence the interpolation results. In the GPR case, this is automatically calculated via an internal fitting routine of `fitrgp()`. NN distances can lead to insight about correlation lengths in a given VFZ-GBO set and are used in the IDW implementation (future work). For IDW, the radius of influence is set to  $r = \sqrt{2}\mu$ , where  $\mu$  is the mean NN distance. It is likely that better tuning of the kernel parameters in these two methods (such as use of built-in hyperparameter optimization in the case of `fitrgp()`) could further decrease their interpolation errors. Additionally, for GPR, use of the 'exact' `PredictMethod` or a larger fully independent conditional ('fic' `PredictMethod`) set size will also likely reduce interpolation error.

By contrast, barycentric interpolation automatically adjusts its effective region of influence because the size of the simplices in the mesh decreases as the number of vertices increases. More uniformly distributed meshes (such as obtained via constrained

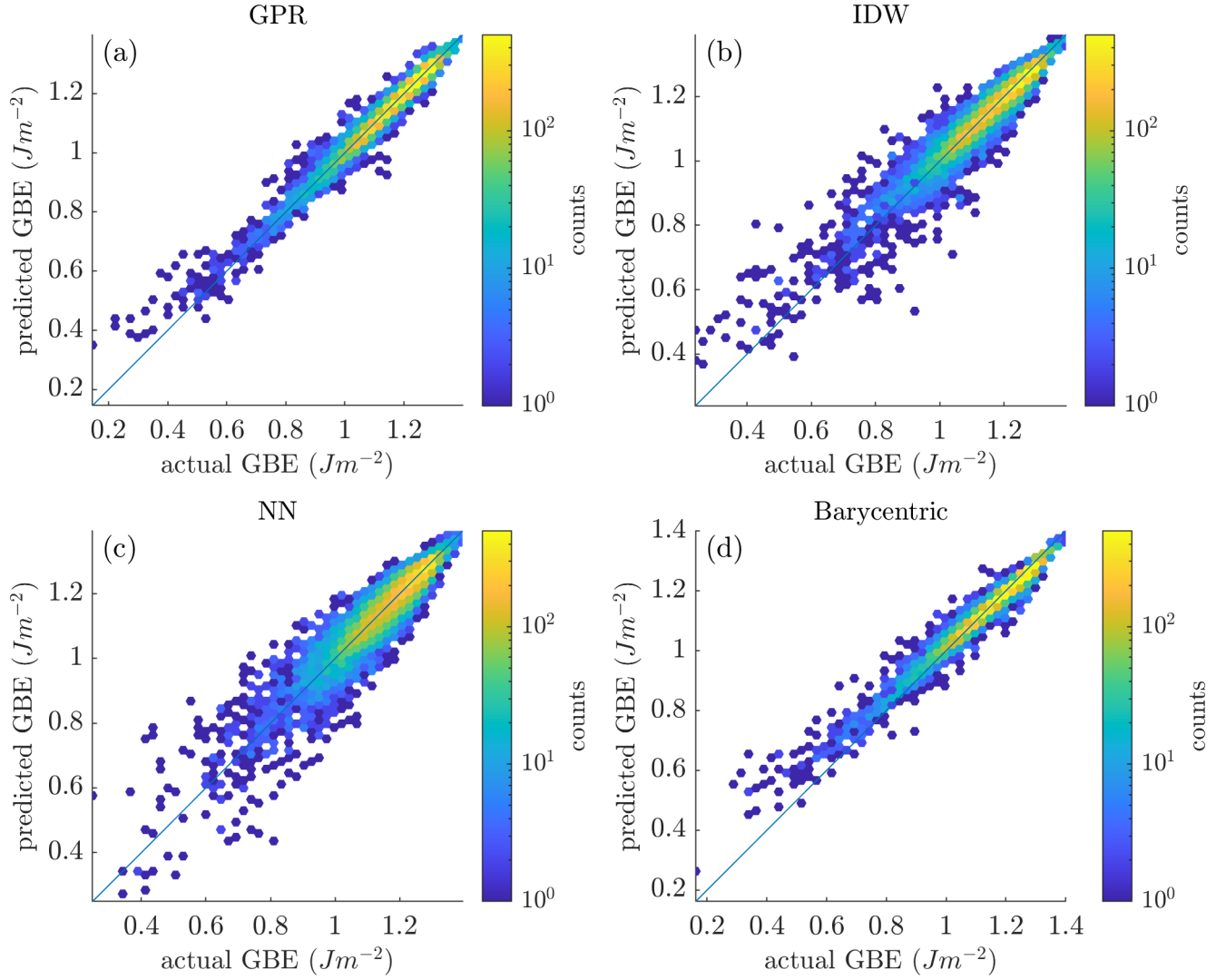


Figure 4: Hexagonally binned parity plots for 50 000 input and 10 000 prediction GBOs formed via pairs of a random cubochorically sampled quaternion and a spherically sampled random boundary plane normal. Interpolation via (a) GPR, (b) IDW, (c) NN, and (d) barycentric coordinates. BRK GBE function for face-centered cubic Ni [49] was used as the test function.

Table 2: Comparison of average interpolation RMSE (approximately 10 trial runs) for each interpolation method in the present work, using 50 000 points in the definition of the VFZ and GBEs obtained by evaluating the BRK validation function by Bulatov et al. [49] at these points. A constant model (Cst, Avg RMSE), whose value was chosen to be the mean of the input GBE was used as a control. The last two columns represent the reduction ( $\downarrow$ ) in RMSE in absolute units of  $\text{J m}^{-2}$  and % relative to the control model, respectively.

Method	Distance	Dataset	# GBs	RMSE ( $\text{J m}^{-2}$ )	Cst, Avg RMSE ( $\text{J m}^{-2}$ )	RMSE $\downarrow$ ( $\text{J m}^{-2}$ )	RMSE $\downarrow$ (%)
GPR	VFZ	BRK	50 000	0.0218	0.1283	0.1065	83
Barycentric	VFZ	BRK	50 000	0.0238	0.1283	0.1045	81.4
IDW	VFZ	BRK	50 000	0.0356	0.1283	0.0927	72.3
NN	VFZ	BRK	50 000	0.0445	0.1283	0.0838	65.3

Table 3: Comparison of average interpolation MAE (approximately 10 trial runs) for each interpolation method in the present work, using 50 000 points in the definition of the VFZ and GBEs obtained by evaluating the BRK validation function ([49]) at these points. A constant model (Cst, Avg MAE), whose value was chosen to be the mean of the input GBE was used as a control. The last two columns represent the reduction ( $\downarrow$ ) in MAE in absolute units of  $\text{J m}^{-2}$  and % relative to the control model, respectively.

Method	Distance	Dataset	# GBs	MAE ( $\text{J m}^{-2}$ )	Cst, Avg MAE ( $\text{J m}^{-2}$ )	MAE $\downarrow$ ( $\text{J m}^{-2}$ )	MAE $\downarrow$ (%)
GPR	VFZ	BRK	50 000	0.0145	0.0955	0.081	84.8
Barycentric	VFZ	BRK	50 000	0.0145	0.0955	0.081	84.8
IDW	VFZ	BRK	50 000	0.0225	0.0955	0.073	76.4
NN	VFZ	BRK	50 000	0.0307	0.0955	0.0648	67.9

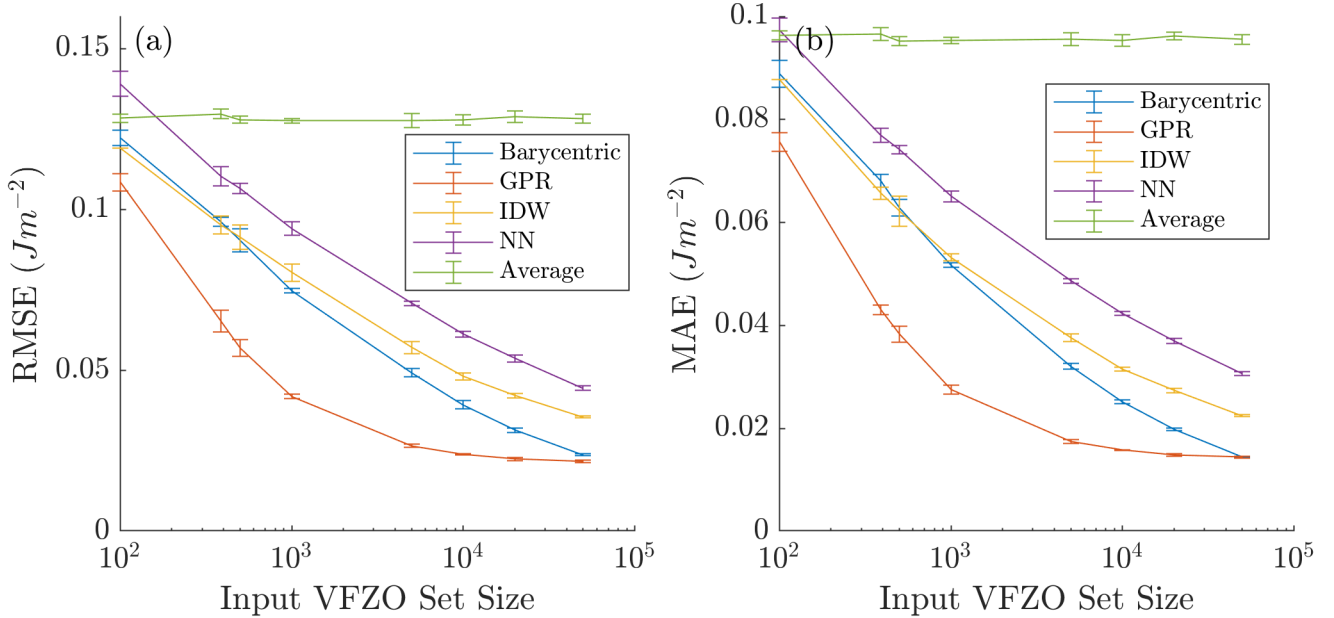


Figure 5: (a) Average RMSE and (b) average MAE vs. number of input points for (planar) barycentric (blue), GPR (orange), IDW (yellow), and NN (purple) interpolation for approximately 10 random runs with different input and prediction points. Standard deviations of approximately 10 runs are also included. Compare with approximately  $0.1283 \text{ J m}^{-2}$  and  $0.0955 \text{ J m}^{-2}$  RMSE and MAE, respectively, for a constant, average model (green) using the average of the input properties (approximately  $1.16 \text{ J m}^{-2}$ ).

optimization [70, 71]) will likely result in lower, more uniform interpolation error, especially for this simplex-based approach which can exhibit high-aspect ratio facets and non-intersections outside the bounds of the mesh [63]. While the barycentric interpolation error is always higher than GPR for the considered set sizes, at 50 000 VFZ-GBOs, the errors of GPR and barycentric interpolation are nearly identical.

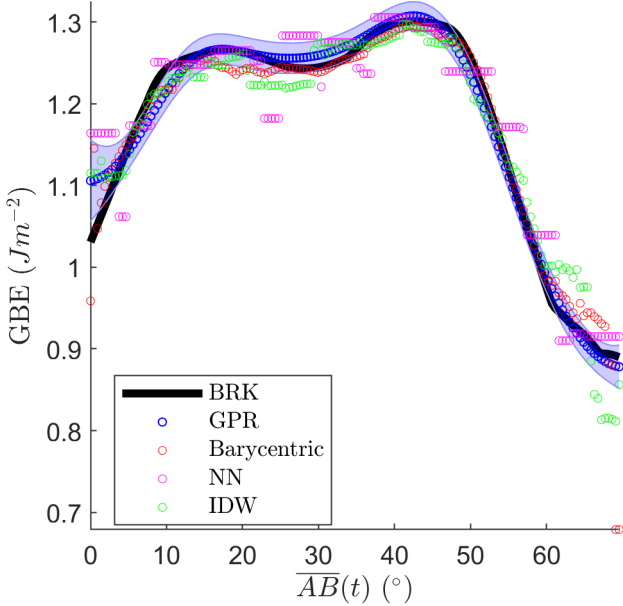


Figure 6: Predictions of GPR (blue circles), barycentric (red circles), NN (magenta circles), and IDW (green circles) as a function of distance along a 1D arc ( $\overline{AB}$ ) between two VFZ-GBOs ( $A$  and  $B$ ). The true, underlying BRK function is also shown (black line). 50 000 random input VFZ-GBOs were generated and used for each of the models. 150 equally spaced points between  $A$  and  $B$  obtained via oSLERP [55] were used as prediction points. GPR uncertainty standard deviation is plotted as shaded error band.

### 3.1.2. Constant-Valued Control Models

To aid in objective interpretation of the error metrics, comparison is made to a constant valued control model, whose value is chosen to be the average of  $y$  (approximately  $1.16 \text{ J m}^{-2}$  in the limit of  $n_{\text{inputpts}} \rightarrow \infty$ ) resulting in RMSE and MAE values of approximately 0.1283 and  $0.0955 \text{ J m}^{-2}$ . This comparison with the relevant constant-valued function gives a sense of the complexity and variability of the validation function and allows for a

more objective comparison between differing works. For example, the RMSE for the relevant constant function compared to the validation function employed for the ANN interpolation method in [52] is  $0.0854 \text{ J m}^{-2}$ ; in contrast, the RMSE for the relevant constant function compared to the BRK validation function used in this work is  $0.1302 \text{ J m}^{-2}$  (see Table 2). This suggests that the BRK validation function is more complex and therefore less well approximated by a constant than the validation function used to test the ANN interpolation method in [52]. Consequently, the improved performance of the present methods (see Section 3.4.1 and Tables 5 and 6) is even more notable in that the validation function employed here is more difficult to interpolate.

### 3.1.3. Experimental and Simulation Error

To give further context to the results of this and prior works, it is useful to consider what the intrinsic error is for typical GB property data. This provides an idea of the minimum possible interpolation error, since one cannot reliably *detect* lower error in the interpolation than already exists in the observed data itself.

One such estimate for error is furnished by the work of Shen et al. [51], who introduced a non-discretizing approach to extract relative GB energies from polycrystalline samples using the locally optimal block preconditioned conjugate gradient method. Their approach utilizes regularization imposed on triple junction equilibrium equations and k-nearest neighbor distances. Using 60 000 triple junctions (180 000 GBs) and a custom, non-smooth validation function they obtained GBE RMSE values of  $0.0076 \text{ J m}^{-2}$  and  $0.0277 \text{ J m}^{-2}$  for GBE values greater than  $0.9 \text{ J m}^{-2}$  and less than  $0.9 \text{ J m}^{-2}$ , respectively. This suggests that an optimistic estimate for the error in noise-free<sup>11</sup> experimental GBE data obtained using such a method is on the order of  $0.0076 \text{ J m}^{-2}$  to  $0.0277 \text{ J m}^{-2}$ , which

<sup>11</sup>These errors are based on Figure 8 from Shen et al. [51], which employed synthetic triple junctions with a custom validation function, rather than experimental data. While the authors did also consider the addition of noise, we use the noise-free results as an estimate of the best-case scenario.



also serves as an estimate of the minimum achievable noise-free experimental interpolation error for any of the interpolation methods described here. Again comparing the relevant constant-valued control model<sup>12</sup> to the validation function employed by Shen et al. [51], we calculate a RMSE and MAE of  $0.0976 \text{ J m}^{-2}$  and  $0.0466 \text{ J m}^{-2}$ , respectively. This implies that the validation model used by Shen et al. [51] is also simpler<sup>13</sup> than the BRK validation model employed in the present work.

### 3.2. Interpolation Efficiency

Below, we present interpolation efficiency results in terms of computational runtime and memory for the four interpolation schemes used in this work (Section 3.2.1). Additionally, an in-depth treatment of the improved symmetrization runtime (separate from interpolation runtime) relative to the original GBO metric is given (Section 3.2.2).

#### 3.2.1. Efficiency of Four Interpolation Methods

We discuss runtime and memory requirements for barycentric, GPR, IDW, and NN interpolation methods. Computational runtimes of the various methods are shown in Table 4.

Barycentric interpolation takes the longest, in spite of the fact that it is the only parallelized method by default (not accounted for in Table 4). In other words, since 12 cores were used to obtain these runtime results, the total runtime across all cores is much higher compared with the other methods; however, it is possible that other methods used multi-threading via built-in vectorized functions. The long computation times of barycentric interpolation result primarily from the large number of facets present in a high-dimensional mesh triangulation and the interconnectedness of facets with respect to each other.

GPR is fast compared to barycentric interpolation; however, the entire process has to be reevaluated (in the current implementation) if the input points (i.e. VFZ-GBOs) or input property values (i.e. GBEs) change (typically referred to as predictors/features and responses, respectively, in the machine learning community). On the other hand, barycentric interpolation is fast if the triangulation and intersections are pre-computed and only input property values change (`interp_bary_fast.m`), but slow if the input or prediction points change, which requires recomputing the triangulation and intersections. Additionally, GPR is the second-longest in terms of runtime.

NN and IDW interpolation have vectorized implementations and are much simpler than the barycentric and GPR methods. Consistent with expectations, NN and IDW exhibit almost negligible runtimes. It should also be noted that barycentric interpolation has much higher memory requirements than GPR, NN, and IDW due to the need to store large matrices. If using an exact prediction scheme, then GPR also has high memory requirements for large VFZ-GBO sets. For 50 000 input points with sufficient RAM (e.g.  $\sim 32 \text{ GB}$ ) and 12 cores available, the 'exact' method runtime is  $(535.1 \pm 392.6)$  seconds. However, because the fully independent conditional approximation is always used in this work, memory requirements are similar to NN and IDW.

Because the default implementation of IDW uses a radius cut-off, the distance and weight matrices can be stored as sparse objects, dramatically reducing both the final memory storage requirements and computational complexity of this method. We expect that a k-nearest neighbor approach would produce similar results both in terms of runtime and error when a relatively uniform sampling of GBC is obtained.

#### 3.2.2. Symmetrization Runtime Comparison with Traditional Octonion Metric

In addition to the interpolation runtime of the methods just presented, it is valuable to consider the runtime of the VFZ symmetrization step (not included in Table 4). The symmetrization step is at the core of the VFZ framework and is a key

<sup>12</sup>We use the mean of the true GBEs from their validation function to define the constant-valued control model instead of the mean of the input GBEs because the latter does not exist for polycrystalline data.

<sup>13</sup>Shen et al. [51] used 8 cusps of varying depths and widths based on the Read-Shockley model and unity GBE everywhere else.

Table 4: Comparison of average runtime (s) for 10 trials for barycentric, GPR, IDW, and NN interpolation methods for various input VFZ-GBO set sizes using 12 cores and evaluated on 10 000 prediction VFZ-GBOs. Because GPR, IDW, and NN method defaults do not use `parfor` loops but may have internal multi-core vectorization, it is unclear to what extent the number of cores affects the runtime of methods other than barycentric interpolation. VFZ-GBO symmetrization runtime was not included; however, symmetrization of 50 000 GBOs takes approximately 76 seconds on 6 cores (Intel i7-10750H, 2.6 GHz) and is a common step in every interpolation method (i.e. it is fundamental to the VFZ framework). We used the BRK validation function for GBE [49].

VFZ-GBO Set Size	Runtime (s)			
	Barycentric	GPR	IDW	NN
100	191.8 $\pm$ 19.57	0.4187 $\pm$ 0.4342	0.034 $\pm$ 0	0.0367 $\pm$ 0.0041
388	388.4 $\pm$ 18.84	0.943 $\pm$ 0.3481	0.0904 $\pm$ 0.0224	0.0705 $\pm$ 0.0129
500	455.7 $\pm$ 55.28	0.6104 $\pm$ 0.3138	0.1352 $\pm$ 0.0364	0.0724 $\pm$ 0.0051
1000	536.5 $\pm$ 35.26	1.743 $\pm$ 0.9464	0.1948 $\pm$ 0.0395	0.1203 $\pm$ 0.0184
5000	998.9 $\pm$ 54.48	5.216 $\pm$ 0.4816	0.8726 $\pm$ 0.1529	0.9277 $\pm$ 0.2418
10 000	1516 $\pm$ 56.59	5.609 $\pm$ 0.8756	1.631 $\pm$ 0.3915	0.8938 $\pm$ 0.1717
20 000	2526 $\pm$ 119.5	11.45 $\pm$ 3.29	3.191 $\pm$ 0.4752	1.275 $\pm$ 0.3423
50 000	5743 $\pm$ 361.3	13.69 $\pm$ 4.05	7.635 $\pm$ 1.872	3.817 $\pm$ 0.5884

to its overall performance. It is a common step for both (i) distance calculations and (ii) all of the interpolation methods presented here.

Directly computed, scaled Euclidean and arc length distances in the VFZ framework approximate the original GBO distance by Francis et al. [55], and the calculation speed is even higher than explicit GBO distance calculations using the original GBO distance. For example, 50 000 GBOs can be symmetrized into VFZ-GBOs in approximately 76 seconds using 6 cores (`get_octpairs.m`), and the corresponding 50 000  $\times$  50 000 pairwise-distance matrix can be computed in approximately 10 seconds (`pdist()`), giving a total runtime of  $\sim$  86 seconds (466 total CPU seconds). Compared to the original GBO metric distance calculations [56] in the Fortran-based EMSOFT package [72], this represents an improvement in computational speed by  $\sim$  5 orders of magnitude using our MATLAB implementation in the VFZ repository [64].

Improvement per distance calculation per core of the VFZ repository is about  $4 \times 10^5$  relative to the EMSOFT [72] metric of 26 minutes using 8 cores for a 388  $\times$  388 pairwise distance matrix. This EMSOFT timing information is directly reported in Chesser et al. [56]. In other words, computation of a 50 000  $\times$  50 000 using the traditional GBO metric and EMSOFT implementation would take approximately 6.6 CPU years (or 153 CPU days by apply-

ing the isometry equation in Section 7 of Morawiec [57]). Since most interpolation methods will depend on computing new distances, probing the model at new GBs will also be expensive. For example, it would take at minimum  $\sim$  30 CPU days (after isometry equivalence has been applied) to perform property interpolation for 10 000 prediction GBs assuming the pairwise-distance matrix relative to 50 000 input GBOs needs to be computed. This presents an issue for iterative simulations (e.g. mesoscale grain growth) in which 1000's of new GB segments would need to be sampled at each time step. By contrast, property values for 10 000 new GBs would be sampled in our approach in  $\sim$  90 CPU seconds. For perspective, a phase-field simulation might have 10 000 or more time steps with thousands of GBs Kim et al. [68], Dimokrati et al. [73]. Recently, Miyoshi et al. [74] presented Reed-Shockley anisotropic 3D phase-field grain growth results for initially 3 125 000 grains with as many as 125 000 time steps to reach  $\sim$  10 000 final grains. Performing such a simulation with even the efficient VFZ framework would require 56 CPU years for the property sampling

alone <sup>14</sup>.

The significant speed up of the VFZ framework stems from the fact that SEOs only need to be considered once per GB,  $O(L)$ , rather than once per distance calculation,  $O(L^2)$ , and that SEOs only need to be considered once in a GB pair,  $O(N_p^2)$ , rather than for every combination between the two GBs,  $O(N_p^4)$ . The SEO computation complexity is thus  $O(N_p^2 L)$ , a significant improvement compared with the original SEO complexity of  $O(N_p^4 L^2)$  [56], where  $N_p$  is the number of proper rotations of the crystallographic point group ( $N_p = 24$  for  $m\bar{3}m$  face-centered cubic point group) and  $L$  is the number of GBs.

Empirically, to compute a pairwise-distance matrix for  $L = 50\,000$  GBs using the VFZ repository [64], the full  $O(N_p^2 L)$  symmetrization operations take about  $76 \text{ seconds} \times 6 \text{ cores} = 456 \text{ seconds}$  of CPU time, whereas the subsequent pairwise-distance computation is  $O_{\text{pd}}(L^2)$  and takes approximately  $10 \text{ seconds}$  for a  $50\,000 \times 50\,000$  matrix. Even though  $O(N_p^2 L) \ll O_{\text{pd}}(L^2)$ , the symmetrization step takes far more time than the pairwise distance calculation (even for large  $L$ ) because of the cost of generating SEOs. Because Euclidean distances—which can be computed faster than trigonometric inverse functions—are employed, and built-in, vectorized MATLAB functions are utilized, there is a further speed enhancement in the VFZ-GBO approach.

### 3.3. Interpolation Along a 1D Arc

To provide a visual illustration of the property predictions, Figure 6 shows the predicted GBE for each of the four interpolation methods as a function of distance along a 1D arc ( $\overline{AB}$ ) between two VFZ-GBOs,  $A$  and  $B$ . Approximate coordinates for  $A$  and  $B$  are given in Table 1, and each intermediate point between  $A$  and  $B$  resides on the surface of a hypersphere. The 150 intermediate points were obtained using oSLERP [55]. Each model

used its own set of 50 000 random input VFZ-GBOs with GBE sampled via the BRK validation function. The two VFZ-GBOs were chosen by taking the furthest apart pair out of 20 000 VFZ-GBOs which thus approximates the largest dimension of the VFZ where each endpoint is close to the true VFZ exterior.

Comparison of the predictions from the four interpolation methods with the true values of the BRK validation function along this 1D path shows that all methods yield reasonable agreement with the true model. The GPR and barycentric methods appear to agree most with the true model, followed by IDW and NN. The NN method shows the piecewise-constant (stair-step) artifact typical of NN methods. We also note that while the fidelity of the predictions is quite good for all methods in the interior of the VFZ, the performance does degrade at the extreme limits of the VFZ (note the deviations at the left and right limits of Figure 6). This effect seems to be particularly pronounced for the barycentric method, and much less so for the GPR method.

We believe this is the first<sup>15</sup> plot of a GB property continuously interpolated between two arbitrary GBs (i.e. neither residing entirely in a single misorientation fundamental zone nor a single boundary plane fundamental zone). Such visualizations can naturally be extended to 2D and 3D by plotting colored points in a triangle or tetrahedron, respectively, all of which (1D, 2D, and 3D) represent small “slices” of the GBC space.

### 3.4. Literature Datasets

In addition to validation results (Section 3.1), we also apply the VFZ framework to real GB property data from two sources in the literature. This allows more direct comparison to previous methods as well as demonstration of the performance

<sup>14</sup>For such an application, a GPU implementation of the VFZ framework, batch implementation of the SEO considerations, directly tracking GBs movement within a VFZ, and/or other approaches would likely be necessary to make the problem more tractable.

<sup>15</sup>OSLERP results from Francis et al. [55] plots GB *structure* continuously between two GBs, Chesser et al. [56] performs cross-validation on the simulated Olmsted Ni GBs, and [57] plots distances between GBs on a geodesic with another GB. The results in these works are distinct from what is presented here: a plot of continuously interpolated GBEs between two arbitrary GBs.

of the the VFZ framework for typical MS data. Specifically, we present GPR interpolation results for MS Fe and Ni simulation datasets and compare them with prior work (Section 3.4.1). Finally, because GPR overestimates the low GBE for the non-uniformly distributed, noisy Fe simulation dataset, we also provide results for an adaptation called the GPRM model that compensates for this effect (Section 3.4.2).

### 3.4.1. Comparison with Prior Work

The GPR interpolation method of the present work was used with the same number of input GBs as was supplied in Restrepo et al. [52] for Fe (17176) and Chessser et al. [56] for Ni (388) to provide a more consistent comparison with prior work. For Fe, the remainder of the simulation data was used for testing, consistent with Restrepo et al. [52], except that zero-energy GBs and degenerate GBs were treated differently as described in Section S6.1. For Ni, a leave-one-out cross validation scheme was used, consistent with Chessser et al. [56].

Hexagonally binned parity plots for the Fe and Ni simulation datasets are shown in Figure S7d and Figure S9, respectively. RMSE and MAE comparisons along with improvement relative to a constant, average model are given in Table 5 and Table 6, respectively.

For the Fe case, we see a larger improvement than prior work likely due to our incorporation of GB symmetry, which was not considered in Restrepo et al. [52]. For the Ni case, there is a slight improvement relative to prior work, indicating that accuracy is similar to the original GBO metric while maintaining the significant computational benefits of the VFZ framework.

Since the BRK validation function is also an interpolation function on the Ni simulation data, GPR within the VFZ framework and the BRK function results are directly compared via parity plot in Figure 7.

For the BRK and GPR interpolations, MAE is  $0.00975 \text{ J m}^{-2}$  and  $0.03626 \text{ J m}^{-2}$ , respectively. Likewise, RMSE is  $0.01727 \text{ J m}^{-2}$  and  $0.04972 \text{ J m}^{-2}$ , respectively. From Figure 7a, we see that low GBE is predicted more accurately and high GBE less accurately with BRK interpolation

vs. GPR in the VFZ framework. Without access to the original fitting routines used to produce the BRK function, we have not performed leave-one-out cross validation which would allow for a safer model evaluation (i.e. one in which fair results are less likely due to overfitting). Leave-one-out cross validation results for the GPR case are, however, shown in Figure S9, indicating that the model performs much worse in such a data-limited regime at points the model has never seen before.

### 3.4.2. Gaussian Process Regression Mixture Applied to Metastable Fe Simulation Data

In addition to GPR, a GPRM model (Figure S7) based on a sigmoid mixing function (Figure S8) is used to better predict low GBE values of the non-uniformly distributed, noisy Fe dataset (Section S6.3)<sup>16</sup>. GPRM interpolation results for the Fe GBE simulations [68] are shown in Figure 8, where approximate coordinates for the GBOs *A* and *B* in Figure 8b are given in Table 2. The noise in the data makes it difficult to resolve low GBE cusps and to validate the model (i.e. we have no ground truth with which to compare).

## 4. Conclusion

In this work, we presented the VFZ framework for (i) computing distances between GBs and (ii) predicting the properties of GBs from existing measurements. We found that distance calculations in the VFZ framework are dramatically more computationally efficient than traditional methods at the expense of infrequent, large distance overestimation which can be addressed through ensemble techniques at a small computational cost as discussed in Section 2.1.3.

We also tested a barycentric interpolation method, and adapted three other interpolation methods for use in the VFZ framework. We

<sup>16</sup>Some alternatives: including no-boundary octonions may likewise improve low GBE performance, but possibly at the expense of poorer predictive accuracy for high GBE. Specifying no noise in input data will cause the model to pass through the low GBE, but is an unrealistic assumption for the Fe simulation data.



Table 5: Comparison of interpolation MAE (1 trial run) for 0K molecular statics (MS) datasets. A constant model (Cst, Avg MAE), whose value was chosen to be the mean of the input GBE was used as a control. The last two columns, MAE  $\downarrow$  ( $\text{J m}^{-2}$ ) and MAE  $\downarrow$  (%), represent the reduction in MAE in units of  $\text{J m}^{-2}$  and % relative to the control model, respectively. Non-sym refers to distances calculated in Restrepo et al. [52] without regard for crystal symmetries.

Method	Distance	Dataset	# GBs	MAE ( $\text{J m}^{-2}$ )	Cst, Avg MAE ( $\text{J m}^{-2}$ )	MAE $\downarrow$ ( $\text{J m}^{-2}$ )	MAE $\downarrow$ (%)
GPR	VFZ	MS Fe	17 176	0.0405	0.0617	0.0212	34.4
ANN [52]	Non-sym	MS Fe	17 176	0.0486	0.0617	0.0131	21.2
LKR [56]	GBO	MS Ni	388	—	0.1752	—	—

Table 6: Comparison of interpolation RMSE (1 trial run) for 0K molecular statics (MS) datasets. A constant model (Cst, Avg RMSE), whose value was chosen to be the mean of the input GBE was used as a control. The last two columns, RMSE  $\downarrow$  ( $\text{J m}^{-2}$ ) and RMSE  $\downarrow$  (%), represent the reduction in RMSE in units of  $\text{J m}^{-2}$  and % relative to the control model, respectively. Non-sym refers to distances calculated in Restrepo et al. [52] without regard for crystal symmetries.

Method	Distance	Dataset	# GBs	RMSE ( $\text{J m}^{-2}$ )	Cst, Avg RMSE ( $\text{J m}^{-2}$ )	RMSE $\downarrow$ ( $\text{J m}^{-2}$ )	RMSE $\downarrow$ (%)
ANN [52]	Non-sym	MS Fe	17 176	—	0.0854	—	—
GPR	VFZ	MS Ni	388	0.0951	0.2243	0.1292	57.6
LKR [56]	GBO	MS Ni	388	0.0977	0.2243	0.1266	56.4

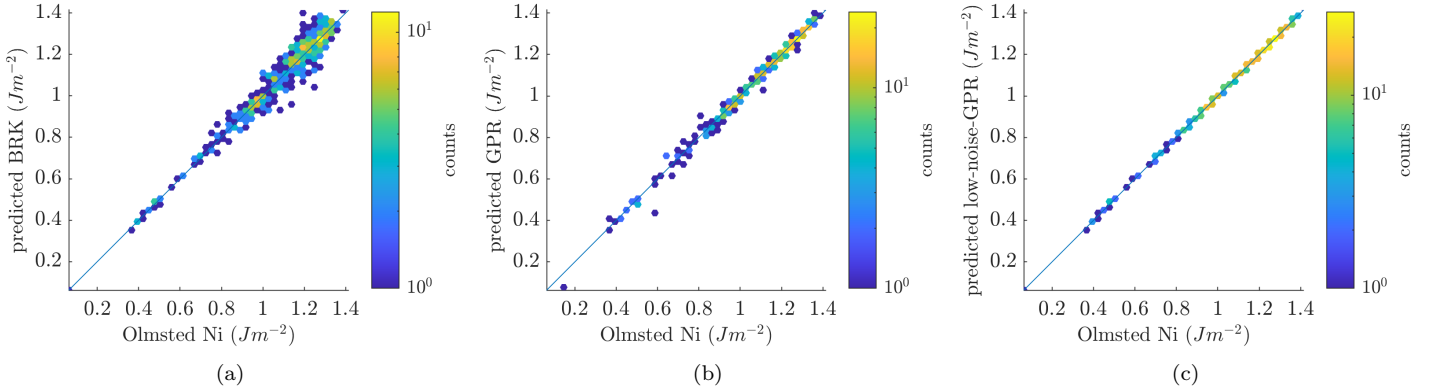


Figure 7: Resubstitution parity plots for (a) BRK model, (b) GPR, and (c) GPR constrained to negligible property noise vs. Olmsted Ni simulation input data.

provide an easy-to-use, versatile implementation of our methods through an interpolation function `interp5DOF.m` written in MATLAB ([github.com/sgbaird-5dof/interp](https://github.com/sgbaird-5dof/interp), [64]) and many companion functions in the VFZ repository. This approach is general and can be applied to any crystal system (i.e. any of the 32 crystallographic point groups

can be selected by the parameter `pgnum`<sup>17</sup>). The

<sup>17</sup>While our testing focused on cubic point group symmetry, symmetry operators for other point group symmetries were provided in the TutorialCode/crystal\_symmetry\_ops directory of [github.com/ichesser/GB-octonion-code](https://github.com/ichesser/GB-octonion-code) (as of commit: f57f9be). Other point groups (in particular those which are noncentrosymmetric) may give rise to differently shaped/larger VFZs for which a Euclidean distance approximation will have the worst case error of 2 vs. the true value of  $\pi$  which represent the furthest Euclidean and arc length distances on a unit hypersphere, respectively. The distance type of `GBdist4.m` can be changed from 'norm' to 'omega' to address this issue. We plan to investigate symmetries other than cubic in future work.

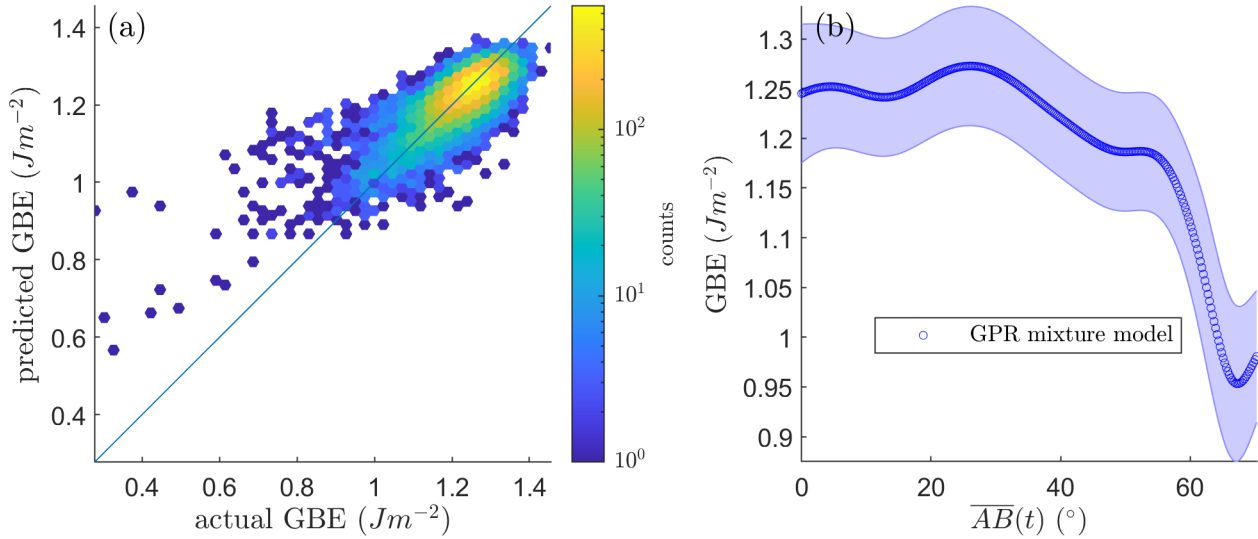


Figure 8: Interpolation results for a large Fe simulation database [68] using 46 883 input GBs and 11 721 prediction GBs in an 80%/20% split and a GPRM model to better approximate low GBEs. Use of a GPRM model predicts low GBE better than the standard GPR model (compare with Figure S7d). (a) Hexagonally binned parity plot of the GPR mixing model with RMSE and MAE of  $0.055\,035\,\text{J m}^{-2}$  and  $0.039\,185\,\text{J m}^{-2}$ , respectively, relative to typical, constant average models of  $0.0854\,\text{J m}^{-2}$  and  $0.0617\,\text{J m}^{-2}$ , respectively. (b) Predictions of GPRM model (blue circles) as a function of distance along a 1D arc ( $\overline{AB}$ ) between two VFZ-GBOs ( $A$  and  $B$ ).

methods described here may be applicable to other distance metrics (see Morawiec [57] for a comprehensive summary of metrics). We also developed a GPRM model specifically for better low GBE prediction using a non-uniformly distributed, noisy dataset.

Of the interpolation methods that we present in this work, GPR provided the highest accuracy predictions. It also provided higher accuracy predictions than any of the methods in the literature. The GPR interpolation errors (50 000 VFZ-GBOs) for the BRK validation model are about 2.4 times the intrinsic error that would be expected from reconstruction of noise-free, experimental polycrystalline data via locally optimal block preconditioned conjugate gradient [51] (180 000 GBs) with their simpler validation model. We recommend the GPR interpolation method for the VFZ framework for most applications because it provides the best combination of accuracy and speed, handles input noise, and has built-in uncertainty quantification; however, the other methods can meet niche needs. For example, barycentric interpolation enables rapid and accurate predictions when the function to be evaluated changes, but the input and

prediction GBs remain fixed.

We anticipate that the VFZ framework and corresponding implementation will benefit numerous applications related to GB structure and properties, including facilitating GB structure-property model development, enabling efficient surrogate modeling of GB properties, and larger scale iterative simulations that require repetitive evaluation of computationally expensive structure-property models.

In addition to its potential for surrogate models, we believe it will be instructive to probe properties for a 5DOF FZ. For example, we found that the maximum dimension of an  $O_h$  VFZ is  $\sim 66.5^\circ$ . Other topics of interest are:

- correlation lengths for various GB properties and crystallographic symmetries within a FZ
- upper and lower bounds for metastable properties
- paths through 5DOF space w.r.t. low-Sigma GBs of interest
- FZ shapes

- analytical definitions of a 5DOF FZ

We plan to explore such properties of a 5DOF FZ in future work.

## 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, Writing - Review & Editing. **David Fullwood:** Funding acquisition, Writing - Review & Editing. **Eric Homer:** Funding acquisition, Writing - Review & Editing

## Appendix A Active vs. Passive Convention

Misorientation quaternions are represented in the active sense<sup>18</sup>:

$$q_m = q_A^{-1} q_B \quad (4)$$

where  $q_m$ ,  $q_A$ , and  $q_B$  represent the misorientation quaternion, orientation quaternion of grain A in the sample frame, and orientation quaternion of grain B in the sample frame, respectively. The  $^{-1}$  operator denotes a unit quaternion inverse (identical to conjugation of a unit quaternion). Quaternion multiplication is given by equation 23 of [75]

$$pq \equiv (p_0 q_0 - \mathbf{p} \cdot \mathbf{q}, q_0 \mathbf{p} + p_0 \mathbf{q} + P \mathbf{p} \times \mathbf{q}) \quad (5)$$

where  $q_0$  and  $p_0$  are scalar components of the quaternions, and  $\mathbf{q}$  and  $\mathbf{p}$  are the vector components.

In this work, we use the convention that  $P = 1$  throughout the various operations in the VFZ repository ( $P \equiv \text{epsijk}$ ) and highly encourage interested readers to refer to Rowenhorst et al. [75] to understand the redefined versions of quaternion multiplication, quaternion rotation, nuances associated with use of active vs. passive conventions, etc. Boundary plane unit normals are expressed pointing away from grain A and in the reference frame of grain A (i.e. the outward-pointing normal convention).

## Glossary

**5DOF** five degree-of-freedom 1, 2, 8, 18, 19

**ANN** artificial neural network 2, 12, 17

**BRK** Bulatov Reed Kumar 8, 10–18

**FZ** fundamental zone 1, 3, 6, 8, 18, 19

**GB** grain boundary 1–3, 5, 6, 8, 10–12, 14–18

**GBC** grain boundary character 2, 13, 15

**GBE** grain boundary energy 1–3, 6, 9–18

**GBO** grain boundary octonion 1–10, 13, 14, 16, 17

**GPR** Gaussian process regression 1, 3, 6, 8–18

**GPRM** Gaussian process regression mixture 6, 8, 16, 18

**IDW** inverse-distance weighting 1, 3, 6–15

**LKR** Laplacian kernel regression 17

**MAE** mean absolute error 2, 6, 9, 11–13, 16–18

**MS** molecular statics 8, 16, 17

**NN** nearest neighbor 1, 6–15

<sup>18</sup>The passive convention is used in [55]

**oSLERP** octonion Spherical Linear Interpolation 2, 3, 12, 15

**RMSE** root mean square error 1–3, 6, 9–13, 16–18

**SEO** symmetrically equivalent octonion 3–5, 15

**VFZ** Voronoi fundamental zone 1–11, 13–19

**VFZ-GBO** Voronoi fundamental zone grain boundary octonion 4–9, 12–15, 18

## Data Availability

The raw data required to reproduce these findings are available to download from <https://doi.org/10.1088/0965-0393/22/3/034004>, [https://github.com/ichesser/GB\\_octonion\\_code](https://github.com/ichesser/GB_octonion_code), and <https://github.com/sgbaird-5DOF/interp>. The processed data required to reproduce these findings are available to download from <https://figshare.com/articles/dataset/14405924>.

## References

- [1] S. Jin, M. Huang, Y. Kwon, L. Zhang, B. W. Li, S. Oh, J. Dong, D. Luo, M. Biswal, B. V. Cunnig, P. V. Bakharev, I. Moon, W. J. Yoo, D. C. Camacho-Mojica, Y. J. Kim, S. H. Lee, B. Wang, W. K. Seong, M. Saxena, F. Ding, H. J. Shin, R. S. Ruoff, Colossal grain growth yields single-crystal metal foils by contact-free annealing, *Science* 362 (2018) 1021–1025. doi:[10.1126/science.aao3373](https://doi.org/10.1126/science.aao3373).
- [2] J. E. Brandenburg, L. A. Barrales-Mora, D. A. Molodov, On migration and faceting of low-angle grain boundaries: Experimental and computational study, *Acta Materialia* 77 (2014) 294–309. doi:[10.1016/j.actamat.2014.06.006](https://doi.org/10.1016/j.actamat.2014.06.006).
- [3] 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 Materials* 14 (2015) 691–695. doi:[10.1038/nmat4311](https://doi.org/10.1038/nmat4311).
- [4] 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).
- [5] 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).
- [6] L. G. Ware, D. H. Suzuki, K. R. Wicker, Z. C. Cordero, Grain boundary plane manipulation in directionally solidified bicrystals and tricrystals, *Scripta Materialia* 152 (2018) 98–101. doi:[10.1016/j.scriptamat.2018.03.047](https://doi.org/10.1016/j.scriptamat.2018.03.047).
- [7] J. Li, A. Oudriss, A. Metsue, J. Bouhattate, X. Feaugas, Anisotropy of hydrogen diffusion in nickel single crystals: The effects of self-stress and hydrogen concentration on diffusion, *Scientific Reports* 7 (2017) 45041. doi:[10.1038/srep45041](https://doi.org/10.1038/srep45041).
- [8] A. Oudriss, J. Creus, J. Bouhattate, E. Conforto, C. Berziou, C. Savall, X. Feaugas, Grain size and grain-boundary effects on diffusion and trapping of hydrogen in pure nickel, *Acta Materialia* 60 (2012) 6814–6828. doi:[10.1016/j.actamat.2012.09.004](https://doi.org/10.1016/j.actamat.2012.09.004).
- [9] A. Metsue, A. Oudriss, X. Feaugas, Hydrogen solubility and vacancy concentration in nickel single crystals at thermal equilibrium: New insights from statistical mechanics and ab initio calculations, *Journal of Alloys and Compounds* 656 (2016) 555–567. doi:[10.1016/j.jallcom.2015.09.252](https://doi.org/10.1016/j.jallcom.2015.09.252).
- [10] S. Huang, D. Chen, J. Song, D. L. McDowell, T. Zhu, Hydrogen embrittlement of grain boundaries in nickel: An atomistic study, *npj Computational Materials* 3 (2017) 1–8. doi:[10.1038/s41524-017-0031-1](https://doi.org/10.1038/s41524-017-0031-1).
- [11] S. Xia, H. Li, T. G. Liu, B. X. Zhou, Applying grain boundary engineering to Alloy 690



- tube for enhancing intergranular corrosion resistance, *Journal of Nuclear Materials* 416 (2011) 303–310. doi:[10.1016/j.jnucmat.2011.06.017](https://doi.org/10.1016/j.jnucmat.2011.06.017).
- [12] M. J. Demkowicz, A threshold density of helium bubbles induces a ductile-to-brittle transition at a grain boundary in nickel, *Journal of Nuclear Materials* 533 (2020) 152118. doi:[10.1016/j.jnucmat.2020.152118](https://doi.org/10.1016/j.jnucmat.2020.152118).
- [13] J. P. Hanson, A. Bagri, J. Lind, P. Kenesei, R. M. Suter, S. Gradečak, M. J. Demkowicz, Crystallographic character of grain boundaries resistant to hydrogen-assisted fracture in Ni-base alloy 725, *Nature Communications* 9 (2018) 1–11. doi:[10.1038/s41467-018-05549-y](https://doi.org/10.1038/s41467-018-05549-y).
- [14] S. Jothi, S. V. Merzlikin, T. N. Croft, J. Andersson, S. G. Brown, An investigation of micro-mechanisms in hydrogen induced cracking in nickel-based superalloy 718, *Journal of Alloys and Compounds* 664 (2016) 664–681. doi:[10.1016/j.jallcom.2016.01.033](https://doi.org/10.1016/j.jallcom.2016.01.033).
- [15] X. Zhou, D. Marchand, D. L. McDowell, T. Zhu, J. Song, Chemomechanical Origin of Hydrogen Trapping at Grain Boundaries in fcc Metals, *Physical Review Letters* 116 (2016) 1–33. doi:[10.1103/PhysRevLett.116.075502](https://doi.org/10.1103/PhysRevLett.116.075502).
- [16] W. Huang, M. Shishehbor, N. Guarín-Zapata, N. D. Kirchhofer, J. Li, L. Cruz, T. Wang, S. Bhowmick, D. Stauffer, P. Manimunda, K. N. Bozhilov, R. Caldwell, P. Zavattieri, D. Kisailus, A natural impact-resistant bicontinuous composite nanoparticle coating, *Nature Materials* 19 (2020) 1236–1243. doi:[10.1038/s41563-020-0768-7](https://doi.org/10.1038/s41563-020-0768-7).
- [17] Y. M. Wang, T. Voisin, J. T. McKeown, J. Ye, N. P. Calta, Z. Li, Z. Zeng, Y. Zhang, W. Chen, T. T. Roehling, R. T. Ott, M. K. Santala, P. J. Depond, M. J. Matthews, A. V. Hamza, T. Zhu, Additively manufactured hierarchical stainless steels with high strength and ductility, *Nature Materials* 17 (2018) 63–71. doi:[10.1038/nmat5021](https://doi.org/10.1038/nmat5021).
- [18] N. Y. C. Lin, M. Bierbaum, P. Schall, J. P. Sethna, I. Cohen, Measuring nonlinear stresses generated by defects in 3D colloidal crystals, *Nature Materials* 15 (2016) 1172–1176. doi:[10.1038/nmat4715](https://doi.org/10.1038/nmat4715).
- [19] D. Yin, C. Chen, M. Saito, K. Inoue, Y. Ikuhara, Ceramic phases with one-dimensional long-range order, *Nature Materials* 18 (2019) 19–23. doi:[10.1038/s41563-018-0240-0](https://doi.org/10.1038/s41563-018-0240-0).
- [20] Y. Guan, W. Li, Y. Gong, G. Liu, X. Zhang, J. Chen, J. Gelb, W. Yun, Y. Xiong, Y. Tian, H. Wang, Analysis of the three-dimensional microstructure of a solid-oxide fuel cell anode using nano X-ray tomography, *Journal of Power Sources* 196 (2011) 1915–1919. doi:[10.1016/j.jpowsour.2010.09.059](https://doi.org/10.1016/j.jpowsour.2010.09.059).
- [21] I. V. Vlassiouk, Y. Stehle, P. R. Pudasaini, R. R. Unocic, P. D. Rack, A. P. Baddorf, I. N. Ivanov, N. V. Lavrik, F. List, N. Gupta, K. V. Bets, B. I. Yakobson, S. N. Smirnov, Evolutionary selection growth of two-dimensional materials on polycrystalline substrates, *Nature Materials* 17 (2018) 318–322. doi:[10.1038/s41563-018-0019-3](https://doi.org/10.1038/s41563-018-0019-3).
- [22] Y. Han, M.-Y. Li, G.-S. Jung, M. A. Marsalis, Z. Qin, M. J. Buehler, L.-J. Li, D. A. Muller, Sub-nanometre channels embedded in two-dimensional materials, *Nature Materials* 17 (2018) 129–133. doi:[10.1038/nmat5038](https://doi.org/10.1038/nmat5038).
- [23] J. Sun, J. Yu, Y. Guo, Q. Wang, Enhancing power factor of SnSe sheet with grain boundary by doping germanium or silicon, *npj Computational Materials* 6 (2020) 1–6. doi:[10.1038/s41524-020-00368-6](https://doi.org/10.1038/s41524-020-00368-6).
- [24] O. K. Johnson, L. Li, M. J. Demkowicz, C. A. Schuh, Inferring grain boundary structure–property relations from effective property measurements, *Journal of Materials Science* 50 (2015) 6907–6919. doi:[10.1007/s10853-015-9241-4](https://doi.org/10.1007/s10853-015-9241-4).

- [25] C.-C. Yang, A. Rollett, W. Mullins, Measuring relative grain boundary energies and mobilities in an aluminum foil from triple junction geometry, *Scripta Materialia* 44 (2001) 2735–2740. doi:[10.1016/S1359-6462\(01\)00960-5](https://doi.org/10.1016/S1359-6462(01)00960-5).
- [26] J. Zhang, W. Ludwig, Y. Zhang, H. H. B. Sørensen, D. J. Rowenhorst, A. Yamanaka, P. W. Voorhees, H. F. Poulsen, Grain boundary mobilities in polycrystals, *Acta Materialia* 191 (2020) 211–220. doi:[10.1016/j.actamat.2020.03.044](https://doi.org/10.1016/j.actamat.2020.03.044).
- [27] 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).
- [28] J. Wei, B. Feng, R. Ishikawa, T. Yokoi, K. Matsunaga, N. Shibata, Y. Ikumura, Direct imaging of atomistic grain boundary migration, *Nature Materials* (2021). doi:[10.1038/s41563-020-00879-z](https://doi.org/10.1038/s41563-020-00879-z).
- [29] R. Bostanabad, A. T. Bui, W. Xie, D. W. Apley, W. Chen, Stochastic microstructure characterization and reconstruction via supervised learning, *Acta Materialia* 103 (2016). doi:[10.1016/j.actamat.2015.09.044](https://doi.org/10.1016/j.actamat.2015.09.044).
- [30] E. R. Homer, High-throughput simulations for insight into grain boundary structure-property relationships and other complex microstructural phenomena, *Computational Materials Science* 161 (2019) 244–254. doi:[10.1016/j.commatsci.2019.01.041](https://doi.org/10.1016/j.commatsci.2019.01.041).
- [31] S. Jothi, T. N. Croft, S. G. Brown, Multi-scale multiphysics model for hydrogen embrittlement in polycrystalline nickel, *Journal of Alloys and Compounds* 645 (2015) S500–S504. doi:[10.1016/j.jallcom.2014.12.073](https://doi.org/10.1016/j.jallcom.2014.12.073).
- [32] H. Pirgazi, On the alignment of 3D EBSD data collected by serial sectioning technique, *Materials Characterization* 152 (2019) 223–229. doi:[10.1016/j.matchar.2019.04.026](https://doi.org/10.1016/j.matchar.2019.04.026).
- [33] H. Pirgazi, K. Glowinski, A. Morawiec, L. A. Kestens, Three-dimensional characterization of grain boundaries in pure nickel by serial sectioning via mechanical polishing, *Journal of Applied Crystallography* 48 (2015) 1672–1678. doi:[10.1107/S1600576715017616](https://doi.org/10.1107/S1600576715017616).
- [34] A. Speidel, R. Su, J. Mitchell-Smith, P. Dryburgh, I. Bisterov, D. Pieris, W. Li, R. Patel, M. Clark, A. T. Clare, Crystallographic texture can be rapidly determined by electrochemical surface analytics, *Acta Materialia* 159 (2018) 89–101. doi:[10.1016/J.ACTAMAT.2018.07.059](https://doi.org/10.1016/J.ACTAMAT.2018.07.059).
- [35] H. Zheng, X. G. Li, R. Tran, C. Chen, M. Horton, D. Winston, K. A. Persson, S. P. Ong, Grain boundary properties of elemental metals, *Acta Materialia* 186 (2020) 40–49. doi:[10.1016/j.actamat.2019.12.030](https://doi.org/10.1016/j.actamat.2019.12.030). [arXiv:1907.08905](https://arxiv.org/abs/1907.08905).
- [36] R. Keinan, H. Bale, N. Gueninchault, E. Lauridsen, A. Shahani, Integrated imaging in three dimensions: Providing a new lens on grain boundaries, particles, and their correlations in polycrystalline silicon, *Acta Materialia* 148 (2018) 225–234. doi:[10.1016/J.ACTAMAT.2018.01.045](https://doi.org/10.1016/J.ACTAMAT.2018.01.045).
- [37] M. Seita, M. Volpi, S. Patala, I. McCue, C. A. Schuh, M. V. Diamanti, J. Erlebacher, M. J. Demkowicz, A high-throughput technique for determining grain boundary character non-destructively in microstructures with through-thickness grains, *Npj Computational Materials* 2 (2016) 16016. URL: <http://dx.doi.org/10.1038/npjcompumats.2016.16>.
- [38] B. Winiarski, A. Gholinia, K. Mingard, M. Gee, G. Thompson, P. Withers, Broad ion beam serial section tomography, *Ultramicroscopy* 172 (2017) 52–64. doi:[10.1016/j.ultramic.2016.10.014](https://doi.org/10.1016/j.ultramic.2016.10.014).
- [39] 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).

- [40] S. Li, L. Yang, C. Lai, Atomistic simulations of energies for arbitrary grain boundaries. Part I: Model and validation, *Computational Materials Science* 161 (2019) 330–338. doi:[10.1016/j.commatsci.2019.02.003](https://doi.org/10.1016/j.commatsci.2019.02.003).
- [41] J. Li, S. J. Dillon, G. S. Rohrer, Relative grain boundary area and energy distributions in nickel, *Acta Materialia* 57 (2009) 4304–4311. doi:[10.1016/j.actamat.2009.06.004](https://doi.org/10.1016/j.actamat.2009.06.004).
- [42] 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).
- [43] 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).
- [44] V. Randle, G. S. Rohrer, H. M. Miller, M. Coleman, G. T. Owen, Five-parameter grain boundary distribution of commercially grain boundary engineered nickel and copper, *Acta Materialia* 56 (2008) 2363–2373. doi:[10.1016/j.actamat.2008.01.039](https://doi.org/10.1016/j.actamat.2008.01.039).
- [45] D. M. Saylor, A. Morawiec, B. L. Adams, G. S. Rohrer, Misorientation dependence of the grain boundary energy in magnesia, *Interface Science* 8 (2000) 131–140. doi:[10.1023/A:1008764219575](https://doi.org/10.1023/A:1008764219575).
- [46] D. M. Saylor, A. Morawiec, G. S. Rohrer, The relative free energies of grain boundaries in magnesia as a function of five macroscopic parameters, *Acta Materialia* 51 (2003) 3675–3686. doi:[10.1016/S1359-6454\(03\)00182-4](https://doi.org/10.1016/S1359-6454(03)00182-4).
- [47] L. Yang, C. Lai, S. Li, Atomistic simulations of energies for arbitrary grain boundaries. Part II: Statistical analysis of energies for tilt and twist grain boundaries, *Computational Materials Science* 162 (2019) 268–276. doi:[10.1016/j.commatsci.2019.03.010](https://doi.org/10.1016/j.commatsci.2019.03.010).
- [48] H. Dette, J. Gösmann, C. Greiff, R. Janisch, Efficient sampling in materials simulation - Exploring the parameter space of grain boundaries, *Acta Materialia* 125 (2017) 145–155. doi:[10.1016/j.actamat.2016.11.044](https://doi.org/10.1016/j.actamat.2016.11.044).
- [49] 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).
- [50] S. J. Dillon, G. S. Rohrer, Characterization of the grain-boundary character and energy distributions of yttria using automated serial sectioning and ebsd in the FIB, *Journal of the American Ceramic Society* 92 (2009) 1580–1585. doi:[10.1111/j.1551-2916.2009.03064.x](https://doi.org/10.1111/j.1551-2916.2009.03064.x).
- [51] Y. F. Shen, X. Zhong, H. Liu, R. M. Suter, A. Morawiec, G. S. Rohrer, Determining grain boundary energies from triple junction geometries without discretizing the five-parameter space, *Acta Materialia* 166 (2019) 126–134. doi:[10.1016/j.actamat.2018.12.022](https://doi.org/10.1016/j.actamat.2018.12.022).
- [52] S. E. Restrepo, S. T. Giraldo, B. J. Thijsse, Using artificial neural networks to predict grain boundary energies, *Computational Materials Science* 86 (2014) 170–173. doi:[10.1016/j.commatsci.2014.01.039](https://doi.org/10.1016/j.commatsci.2014.01.039).
- [53] M. Guziewski, D. Montes de Oca Zapiain, R. Dingreville, S. P. Coleman, Microscopic and Macroscopic Characterization of Grain Boundary Energy and Strength in Silicon Carbide via Machine-Learning Techniques, *ACS Applied Materials & Interfaces* 13 (2021) 3311–3324. doi:[10.1021/acsami.0c15980](https://doi.org/10.1021/acsami.0c15980).
- [54] C. Hu, Y. Zuo, C. Chen, S. Ping Ong, J. Luo, Genetic algorithm-guided deep learning of grain boundary diagrams: Addressing the challenge of five degrees of freedom, *Materials Today* 38 (2020) 49–57. doi:[10.1016/j.mattod.2020.03.004](https://doi.org/10.1016/j.mattod.2020.03.004). [arXiv:2002.10632](https://arxiv.org/abs/2002.10632).
- [55] 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).
- [56] I. Chesser, T. Francis, M. De Graef, E. Holm, Learning the grain boundary manifold: Tools for visualizing and fitting grain boundary properties, *Acta Materialia* 195 (2020) 209–218. doi:[10.1016/j.actamat.2020.05.024](https://doi.org/10.1016/j.actamat.2020.05.024).
- [57] 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).
- [58] C. B. Barber, D. P. Dobkin, H. Huhdanpaa, The quickhull algorithm for convex hulls, *ACM Transactions on Mathematical Software* 22 (1996) 469–483. doi:[10.1145/235815.235821](https://doi.org/10.1145/235815.235821).
- [59] A. Heinz, P. Neumann, Representation of orientation and disorientation data for cubic, hexagonal, tetragonal and orthorhombic crystals, *Acta Crystallographica Section A* 47 (1991) 780–789. doi:[10.1107/S0108767391006864](https://doi.org/10.1107/S0108767391006864).
- [60] H. Grimmer, A unique description of the relative orientation of neighbouring grains, *Acta Crystallographica Section A* 36 (1980) 382–389. doi:[10.1107/S0567739480000861](https://doi.org/10.1107/S0567739480000861).
- [61] S. Patala, C. A. Schuh, Symmetries in the representation of grain boundary-plane distributions, *Philosophical Magazine* 93 (2013) 524–573. doi:[10.1080/14786435.2012.722700](https://doi.org/10.1080/14786435.2012.722700).
- [62] B. Luong, Voronoi Sphere, MATLAB Central File Exchange, 2020. URL: <https://www.mathworks.com/matlabcentral/fileexchange/40989-voronoi-sphere>.
- [63] S. G. Baird, E. R. Homer, D. T. Fullwood, O. K. Johnson, Barycentric Interpolation on the 7-Sphere, *Computer Aided Geometric Design* (Under Review) 8.
- [64] S. Baird, O. Johnson, Five Degree-of-Freedom (5DOF) Interpolation, 2020. URL: [github.com/sgbaird-5dof/interp](https://github.com/sgbaird-5dof/interp).
- [65] E. R. Homer, S. Patala, J. L. Priedeman, Grain Boundary Plane Orientation Fundamental Zones and Structure-Property Relationships, *Scientific Reports* 5 (2015) 1–13. doi:[10.1038/srep15476](https://doi.org/10.1038/srep15476).
- [66] C. E. Rasmussen, C. K. I. Williams, Gaussian Processes for Machine Learning, Adaptive Computation and Machine Learning, MIT Press, Cambridge, Mass, 2006.
- [67] A. Tovar, Inverse distance weight function, MATLAB Central File Exchange, 2020. URL: <https://www.mathworks.com/matlabcentral/fileexchange/46350-inverse-distance-weight-function>.
- [68] 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 and Simulation in Materials Science and Engineering* 22 (2014) 034004. doi:[10.1088/0965-0393/22/3/034004](https://doi.org/10.1088/0965-0393/22/3/034004).
- [69] G. Bean, Hexscatter, MATLAB Central File Exchange, 2020. URL: <https://www.mathworks.com/matlabcentral/fileexchange/45639-hexscatter-m>.
- [70] E. D. Dolan, J. J. More, T. S. Munson, Benchmarking Optimization Software with COPS 3.0, Technical Report, Argonne National Laboratory (ANL), United States, 2004. doi:[10.2172/834714](https://doi.org/10.2172/834714).
- [71] MATLAB Optimization Toolbox, Constrained Electrostatic Nonlinear Optimization, Problem-Based, 2020. URL: <https://www.mathworks.com/help/optim/ug/constrained-electrostatic-problem-based-optimization.html>.
- [72] M. De Graef, EMSOft, 2020. doi:[10.5281/zenodo.3489720](https://doi.org/10.5281/zenodo.3489720).



- [73] A. Dimokrati, Y. Le Bouar, M. Benyoucef, A. Finel, S-PFM model for ideal grain growth, *Acta Materialia* 201 (2020) 147–157. doi:[10.1016/j.actamat.2020.09.073](https://doi.org/10.1016/j.actamat.2020.09.073).
- [74] E. Miyoshi, T. Takaki, S. Sakane, M. Ohno, Y. Shibuta, T. Aoki, Large-scale phase-field study of anisotropic grain growth: Effects of misorientation-dependent grain boundary energy and mobility, *Computational Materials Science* 186 (2021) 109992. doi:[10.1016/j.commatsci.2020.109992](https://doi.org/10.1016/j.commatsci.2020.109992).
- [75] D. Rowenhorst, A. D. Rollett, G. S. Rohrer, M. Groeber, M. Jackson, P. J. Konijnenberg, M. De Graef, Consistent representations of and conversions between 3D rotations, *Modelling and Simulation in Materials Science and Engineering* 23 (2015) 083501. doi:[10.1088/0965-0393/23/8/083501](https://doi.org/10.1088/0965-0393/23/8/083501).