Microstructure as a Graph: Surrogates for Elastic and Failure Responses

A Dissertation Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Erdem Caliskan

August 2026

To my wife and family, with love and gratitude.

Acknowledgements

I would like to express my deepest gratitude to my advisor, Professor Reza Abedi, for his guidance, support, and mentorship throughout my Ph.D. journey. His insight, patience, and unwavering encouragement have shaped not only this dissertation but also my growth as a researcher and engineer. I am especially grateful for the trust he placed in me, the opportunities he created for me to grow, and the care he showed during both the challenging and successful moments of my doctoral studies.

I am deeply thankful to my co-advisors, Dr. Massimiliano (Max) Lupo Pasini and Professor Stephanie TerMaath, for their mentorship and support. I am especially grateful to Max for serving as a mentor during my time at Oak Ridge National Laboratory over the past two years. His guidance broadened my perspective and played an important role in my professional development. I also sincerely appreciate Professor TerMaath for her support and for serving on my committee.

I would also like to thank my dissertation committee members, Professor Sergei Kalinin and Professor Katherine Acton, for their time, thoughtful feedback, and valuable input.

I am grateful to Dr. Alex Plotkowski for his support and encouragement, and to Dr. Vladimir Protopopescu for his advice and for helping me strengthen my presentation skills. I also thank Amber Gillespie for her administrative support and kindness throughout my graduate studies.

I would like to acknowledge Professor Alireza Amirkhizi, with whom I had the opportunity to work during the first three years of my Ph.D. I am thankful for his

support and collaboration during that important part of my graduate journey. I also thank Dr. Weidi Wang and Willoughby Cheney for their collaboration and support.

I am very grateful to my group mates and colleagues for making this journey more meaningful. I especially thank Anik Das Anto, with whom I worked closely, and Ali Heidari Shirazi, who was also my roommate for two years. I also thank Ryan Connor, Ali Behbahani, Giang Huynh, and Hang Wang for their friendship, collegiality, and the many shared experiences along the way.

I would also like to thank my friends Ender, Izzet, Tarik, and Emre for their friendship, encouragement, and support over the years.

Most importantly, I would like to thank my family. To my parents, Metin and Ayfer, thank you for your love, sacrifices, and constant belief in me. Your support has been the foundation of everything I have achieved. To my wife, Alicia, thank you for your love, patience, understanding, and unwavering support throughout this journey. Your presence gave me strength in the most difficult times and made every success more meaningful. I would not have reached this point without you.

"All models are wrong, but some are useful." — George E. P. Box

Abstract

Microstructure-sensitive mechanical response prediction is central to the analysis of heterogeneous materials, but direct numerical simulation becomes expensive when many statistically distinct volume elements are required. This dissertation develops graph neural network (GNN) surrogates that represent microstructures as graphs and learn mappings from local topology, phase information, geometry, and crystallography to mechanical quantities of interest. Across three studies, the work advances graph-based surrogate modeling from anisotropic composite homogenization to nonlinear polycrystal response to microstructure-sensitive fatigue localization.

The first study focuses on two-dimensional fiber composites. Statistical volume elements are converted into compact topology-based graphs and used to predict stiffness and failure-related properties. Because small volume elements can remain anisotropic and statistically variable, the full stiffness tensor is learned rather than reducing the response to isotropic bulk and shear moduli. A mean-field-method-based tensor normalization is introduced to stabilize learning across extreme elastic contrast ratios, and Voronoi-derived node features improve prediction in small-data settings. The resulting GNNs accurately predict anisotropic stiffness, peak strength under phase-field damage, and brittle fracture initiation strength using compact representations relative to image-based models.

The second study extends the graph-surrogate framework to dual-phase ferrite–martensite polycrystals. Each statistical volume element is represented as a grain-adjacency graph with node features describing orientation, phase, grain size, and

composition, and edge features describing misorientation. A multitask GNN jointly predicts elastic constants, isotropic moduli, yield strengths, hardening descriptors, and full stress–strain trajectories across multiple martensite volume fractions and volume-element sizes. The model preserves pointwise accuracy and population-level variability, supporting reconstruction of heterogeneous block-wise response fields for mesoscale analyses.

The third study addresses fatigue indicator prediction in austenitic stainless steel under cyclic plasticity. A hierarchy-aware multitask graph surrogate predicts per-grain and per-volume statistics of Fatemi–Socie and plastic-work fatigue indicator parameters under high-cycle and low-cycle loading regimes. The framework combines calibrated uncertainty estimates with an extreme-value-theory bridge between grain-scale predictions and volume-level maxima, clarifying how grain-level hot-spot behavior propagates to volume-level fatigue measures.

Together, these chapters establish microstructure graphs as accurate, efficient, and extensible representations for surrogate modeling of heterogeneous material response across elastic, elastoplastic, failure, and fatigue regimes.

Table of Contents

List of Tables

List of Figures

Chapter 1

Graph Neural Networks for Mechanical Property Prediction of Composites

Reproduced from: Erdem Caliskan, Reza Abedi, and Massimiliano Lupo Pasini, "Graph Neural Networks for Mechanical Property Prediction of 2D Fiber Composites," Materials & Design, 2025, 114500. https://doi.org/10.1016/j.matdes.2025.114500.

Abstract

This work investigates the graph neural networks' (GNNs) ability to homogenize 2D fiber composite microstructures. We use different inhomogeneity and anisotropy indices to motivate and show that the Volume Elements (VEs) used in ML methods should ideally be far from their Representative Volume Element (RVE) size limit and, consequently, are notably anisotropic. Hence, training only the isotropic limit properties may not be acceptable. Another aspect is the need to normalize elastic stiffness values for ML, especially when high elastic contrast ratios are encountered between composite phases or in the material set. We introduce a normalization technique based on the mean-field method (MFM) to handle such high contrast ratios and train for the entire stiffness tensor. We show that the proposed GNN approaches exhibit high accuracy and efficiency compared to traditional methods and convolutional neural networks, utilizing unstructured graphs constructed from microstructure topology. Our model successfully predicts the stiffness tensor, peak strength under bulk damage, and brittle fracture initiation strength across diverse microstructure configurations while maintaining high accuracy even for extreme material contrasts and volume fractions. We also present a method to improve prediction accuracy for small dataset sizes using Voronoi partitioning.

1.1 Introduction

Composite materials derive their macroscopic behavior from complex interactions at the microscale. Directly resolving microstructural features (e.g., fibers, grains,

inclusions) in simulations demands element sizes on the order of these features, making computations prohibitively expensive. Homogenization circumvents this by averaging microscale responses to obtain effective properties at the macroscale Miehe and Bayreuther (2007).

Homogenization methods pioneered by Hill (1963, 1972); Mandel (1966); Ogden (1974) implicitly incorporate the effects of microstructural details, such as multiple material phases and defects, into the homogenized properties at a larger length scale. The *Volume Element* (VE) refers to the volume of material to derive the homogenized properties.

There are some distinctions between different kinds of VEs. The first kind is the *Representative Volume Element* (RVE). This corresponds to a repeating unit cell for a periodic domain or a sufficiently large domain that is statistically representative of the macroscopic body, independent of any *boundary condition* (BC) Ostoja-Starzewski (1998); Ostoja-Starzewski et al. (2007); Stapleton et al. (2016). The second kind of VE is referred to as the *Statistical Volume Element* (SVE) Ostoja-Starzewski (2006). SVEs are small enough that they are far from reaching the RVE limit based on one of several convergence criteria. SVEs were initially used to model the statistical variation of homogenized elastic properties Baxter and Graham (2000); Huyse and Maes (2001); Tregger et al. (2006). Another use is for fracture problems Garrard and Abedi (2020a,b); Yang et al. (2021), where the implied homogeneous and deterministic properties homogenized by RVEs often result in unsatisfactory results. The SVEs are often used in the context of the moving-window method Graham and Baxter (2001); Graham-Brady et al. (2003); Acton and Graham-Brady (2009a,b, 2010), where assigning a value at the center of each SVE addresses the homogeneity concern of the macroscopic domain. Voronoi VEs are a popular choice of SVEs as, by not cutting through the inclusions, they still provide accurate homogenized properties for small sizes Salmi et al. (2012); Acton et al. (2018, 2019); Abedi et al. (2023).

Different methods can be employed to obtain the bounds of the stiffness tensor value of a particulate composite. In their most basic form, the Voigt and Reuss

averaging formulas determine the upper- and lower-bounds of a multiphase composite by using only the volume fraction (v^I) and stiffness tensor ($\mathbf{C}$) of each phase. These bounds do not account for the topology of the phases, such as the shape of the particles. *Mean-field method* (MFM) refers to a group of (semi)analytical methods that, in addition to v^I and $\mathbf{C}$ of each phase, consider the shape of different phases. Some of the concerns with the MFMs are: 1) They do not consider size, positions, and proximity of phases; 2) Consequently, the MFMs only provides the RVE limit stiffness tensor; 3) Their use is mainly limited to simple inclusion geometric shapes; 4) The bounds are very loose, and estimates are inaccurate when the elastic contrast ratio of the phases is high.

Finally, the finite element method (FEM), among other computational methods, directly models the underlying continuum microstructural details and resolves all of the MFM concerns, although at a much higher computational cost. Furthermore, FEM can be used to account for material failure modes such as matrix failure, debonding, and delamination Orifici et al. (2008). For example, the phase-field damage model Miehe et al. (2010) is commonly used to model the quasi-brittle fracture of composite materials Singh and Pal (2021). However, these methods require a fine mesh and small time steps, which renders them even more computationally expensive.

Recently, *Machine Learning* (ML) methods have shown great promise in increasing the efficiency of numerical methods for modeling composites Mirkhalaf and Rocha (2024); Loh et al. (2024); Zhou and Hubert (2024); Feng et al. (2024); Gomez-Flores et al. (2024); Tasdemir et al. (2025). For instance, Pathan et al. (2019) uses a gradient-boosted tree regression model to predict a homogenized linear stiffness tensor of unidirectional fiber composites. Another commonly used technique for image-based regression of mechanical properties is *Convolutional Neural Networks* (CNNs). They have been employed to derive the stiffness tensor and its upper- and lower-bounds in Eidel (2023). CNNs, specifically those with the UNET Ronneberger et al. (2015) architecture, are also used for predicting peak strength under matrix

damage Gao et al. (2023). However, researchers noted that predicting peak strength from the microstructure image using a standard regression model failed to generalize, and they adopted a method to predict the peak strength from the damage field instead.

Despite their popularity, CNNs have limitations and drawbacks. For instance, they can only operate on regular grids or Euclidean data and do not incorporate the topology information of the microstructure. Moreover, they are inefficient in terms of computational resources, as they necessitate the storage of the entire image in the memory, which increases training time. Excessive memory usage also limits the batch size, which may decrease the accuracy Smith et al. (2017). To address these limitations, researchers introduced *Graph Neural Networks* (GNN) as an efficient alternative, as they can inherently operate on non-Euclidean data Bronstein et al. (2017). GNNs have been used for characterizing metamaterials Meyer et al. (2022), full-field structural dynamics predictions Li et al. (2023), and mechanical response prediction of hollow concrete structures Wang et al. (2025). In a recent study, GNNs have been adopted to predict the effective elastic moduli of rocks Chung et al. (2024). They converted the images into topological graphs using the Mapper algorithm Singh et al. (2007) and showed that GNNs outperform CNNs in terms of accuracy and computational efficiency.

Some general observations are made on the properties of a material and VE to be trained by ML methods. First, given that the ML model is expected to obtain different homogenized mechanical properties from different input elements, the VEs are technically SVEs as opposed to RVEs. The larger the VEs and the closer the RVE limit, the less statistical variation there is among samples. Second, even for a macroscopically isotropic material, one may need to treat it as anisotropic for the VEs used in ML. This is a consequence of the first item, and it is shown that, at least for a few macroscopically isotropic two-phase composites, the convergence of the stiffness tensor to its RVE limit occurs around the same VE size at which it can be accurately represented as isotropic Abedi et al. (2022, 2023). Third, as the contrast

ratio of elastic properties increases, or when considering a dataset with diverging phase contrast ratios, it is necessary to normalize the ML outputs for improved performance Singh and Singh (2020); Cabello-Solorzano et al. (2023). To address this issue, *Hashin-Shtrikman* (HS) bounds have been used to normalize the effective shear and bulk modulus of rocks for ML training Ahmad et al. (2023). However, their normalization method is limited to isotropic materials.

Multiple challenges remain to be addressed, mainly from the perspective of GNNs for mechanical properties. First, GNNs are shown to be more efficient than CNNs in predicting the response of various materials; see, for example, polycrystalline metals Dai et al. (2021) and rocks Chung et al. (2024). It is desired to investigate whether such advantages carry over to other materials, such as 2D fiber composites. Second, the images in the previous studies lack diverse statistical variations, which renders the test dataset statistically similar to the training dataset. We believe this arises from using VEs that are too close to the RVE size limit (item 1 above). Third, a recent review of GNNs Zhao et al. (2024) highlights that studies have overwhelmingly focused on materials with similar mechanical properties, resulting in minimal variation in predicted properties. A proper normalization scheme is required if composites with higher property contrast ratios are used. Following the previous discussion, for VE sizes that can be best modeled with ML, one cannot ignore anisotropy. Thus, normalization of the entire stiffness tensor beyond that of Ahmad et al. (2023) is sought. Fourth, previous methods require a large amount of ground truth data to generate an accurate surrogate model, while successfully training with small datasets remains a challenge. This is particularly important when considering the failure properties of a VE, given the nonlinear nature of the fracture and the computational expense associated with creating the dataset.

We utilize GNNs trained with topological graphs to predict the mechanical properties of two-phase composites with fibers. Our main contributions can be summarized as follows: First, we show that GNNs have superior prediction accuracy and efficiency compared to CNNs for predicting homogeneous properties from 2D

fiber composite microstructure images with diverse statistics and extreme material pairs. This is illustrated for elasticity and fracture responses using Micro2D and PFD datasets. Second, we motivate the need to address SVEs rather than RVEs in ML and show that at such VE sizes, the anisotropy of the stiffness tensor cannot be overlooked. Hence, rather than the common practice of predicting only bulk as shear moduli, the entire stiffness tensor should be predicted. Third, referring to the Micro2D dataset, we propose an effective DEM-based normalization scheme for the entire stiffness tensor. Fourth, we significantly improve the prediction accuracy for small datasets by employing additional node-level features from Voronoi cells. These four contributions directly address the four challenges mentioned above.

1.2 Homogeneity, isotropy, and normalization of stiffness tensor

1.2.1 Homogenized stiffness tensor

For 2D elasticity, the stiffness tensor $\mathbf{C}$ in Voigt notation relates engineering strain $\boldsymbol{\gamma}$ to engineering stress σ through,

$$\sigma = \mathbf{C}\boldsymbol{\gamma}, \quad \text{where} \quad \sigma = \begin{bmatrix} \sigma_{11} \\ \sigma_{22} \\ \sigma_{12} \end{bmatrix}, \ \boldsymbol{\gamma} = \begin{bmatrix} \epsilon_{11} \\ \epsilon_{22} \\ 2\epsilon_{12} \end{bmatrix}. \tag{1.1}$$

This 3×3 symmetric matrix provides a compact representation of the fourth-order elasticity tensor. The Voigt compliance matrix $\mathbf{D} = \mathbf{C}^{-1}$ provides the inverse map $\boldsymbol{\gamma} = \mathbf{D}\sigma$. The components of strain tensor are expressed in terms of displacement vector (U_1, U_2): $\epsilon_{11} = U_{1,1}$, $\epsilon_{22} = U_{2,2}$, $\epsilon_{12} = 0.5(U_{1,2} + U_{2,1})$, where the subscripts 1 and 2 refer to the direction of the components and partial derivatives.

For a multi-phase composite, each phase can have its own stiffness tensor. For the two-phase composites considered, we assume that the matrix and fibers are homogeneous and isotropic with stiffness tensors $\mathbf{C}^M$ and $\mathbf{C}^I$. The superscripts refer to material or homogenization scheme symbols. For example, the superscripts M and I are associated with the matrix and inclusion phases, respectively. The corresponding volume fractions in a VE are denoted by $v^M = 1 - v^I$ and v^I.

The objective is to determine the stiffness tensor $\mathbf{C}^*$ of a given VE. The Voigt $\mathbf{C}^V = v^M\mathbf{C}^M + v^I\mathbf{C}^I$ and Reuss $\mathbf{C}^R = \mathbf{D}^{R^{-1}}$ for $\mathbf{D}^V = v^M\mathbf{D}^M + v^I\mathbf{D}^I$ are loose upper- and lower-bounds for the $\mathbf{C}^*$, which, as evident, only take into account the stiffness and volume fractions of the two phases.* MFMs also consider the shape of the inclusions. These methods are discussed in detail in Willis (1981); Walpole (1966); Böhm (2004). For completeness and consistency with the notations, we provide a brief overview of different MFM bounds and estimates in the A.1. HS $\mathbf{C}^{\text{HS}+}$ and $\mathbf{C}^{\text{HS}-}$ are tighter bounds than Voigt-Reuss as they include additional information, such as the shape of inclusions. As for estimates, we consider Mori-Tanaka (MT) Mori and Tanaka (1973), self-consistent (SC) Hill (1965), and Differential Effective Medium (DEM) Berryman (1980) MFMs.

Two concerns of MFMs are: 1) only providing the RVE-limit solutions; 2) possibly having a low accuracy for such RVE-limit solutions. To explain this, let us assume that a VE is sufficiently large to be considered an RVE. Even for this VE, the stiffness matrices obtained with a more accurate (*e.g.,* FEM discussed below) and an MFM method may noticeably differ (concern 2), especially for large elastic contrast ratios between the phases. The reason for this is that MFMs still do not incorporate many aspects of the microstructure, such as size, positions, and proximity of phases. Additionally, in their most basic form, they are limited to simple shapes, such as ellipses, in 2D. The first concern is crucial in analyzing finite-size VEs. If the VEs are already large enough and close to the RVE limit, there is no need to develop ML

*The energy-based ordering of stiffness/compliance matrices is further discussed in Huet (1990); Ostoja-Starzewski (1998).

surrogate models to predict homogenized properties and, in general, the mechanical response of the VE. This is because the response of all VEs will be nearly identical. For the linear response, the accurate estimation of one VE stiffness tensor suffices, and there is no need for a surrogate model.

Other analytical solutions attempt to improve the accuracy of MFMs. For example, Torquato (1997, 1998) developed a third-order estimate of the stiffness tensor by incorporating three-point statistics. However, these methods become even more complex than classical MFMs and are still not suitable for providing SVE (as opposed to RVE) homogenized properties. Finally, FEM provides the most accurate solution among the methods discussed. It can be applied to VEs with complex microstructures and can include nonlinear effects (if needed, as in Section 1.4.3). *Displacement BC* (DBC), *Traction BC* (TBC), *Mixed BC* (MBC), and *Periodic BC* (PBC) are some of the common choices in FEM, as they satisfy the Hill-Mandel Hill (1967) condition (for the MBC, an orthotropic homogenized response is required Hazanov and Amieur (1995)). DBC and TBC provide upper- and lower-bounds for the homogenized stiffness tensor Hazanov and Huet (1994) that are even tighter than HS+ and HS- bounds. Table 1.1 summarizes the BCs used for each dataset.

1.2.2 Homogeneity and isotropy of elastic response

All the composites considered in this paper are isotropic at the RVE limit, as the two phases are isotropic, the inclusions are circular, and there is no angular bias in the positioning of the inclusions. The corresponding MFM solutions are also isotropic. By homogeneity, we refer to the statistical variation in the elastic response of different VEs. At small VEs, *i.e.,* when SVEs are considered, the homogenized stiffness matrices are both inhomogeneous (high variation across VEs) and anisotropic. At the RVE size limit, all VEs provide almost the same stiffness tensor, *i.e.,* a homogeneous response, and are isotropic as well.

The *Coefficient of Variation* (CoV) of the bulk modulus c_κ and different anisotropy indices will be used to quantify whether the VEs of a given size have reached such homogeneous and isotropic limits, respectively. For a generic anisotropic index, denoted by A, $A = 0$ corresponds to an isotropic response, and its value increases as the stiffness tensor becomes more anisotropic. We will use $\bar{A}$, the nominal mean of A over a population of VEs, as the VEs' measure of anisotropy. Thus, $c_\kappa = 0$ and $\bar{A} = 0$ correspond to complete homogeneity and isotropy. Refer to A.2 for more information.

The question we aim to address is whether it is justifiable to treat the material as isotropic in ML or if one needs to predict the entire stiffness tensor $\mathbf{C}$. Treating the material as inhomogeneous (*i.e.,* having notable variability across VEs) but isotropic is common in literature; see, for example, Ahmad et al. (2023); Chung et al. (2024). In contrast, we believe one needs to predict the entire stiffness matrix. This is supported by the studies in Abedi et al. (2022, 2023), where the homogeneity and isotropy limits are reached at almost the same VE size for different macroscopically isotropic two-phase composites. This aspect will be further investigated in Section 1.2.5 by using the values of c_κ and $\bar{A}$ from the Micro2D dataset.

1.2.3 Normalization of elasticity tensor

Using the absolute value of the input features or target labels is generally not preferred in ML methods, as higher values dominate the loss, and samples with lower values will not contribute to the surrogate model Singh and Singh (2020); Cabello-Solorzano et al. (2023). Thus, normalization methods such as min/max or standard normalizer are commonly used to overcome this issue. However, these methods may not be effective when normalizing the stiffness tensor due to its tensorial nature.

Recently, Ahmad et al. (2023) addressed this problem by representing the stiffness tensor with two scalar parameters, which, as discussed in A.2 are enough to identify an isotropic material. By ignoring the anisotropy of the tensor, one can represent $\mathbf{C}$

by the bulk modulus κ and shear modulus $\mu = C_{33}$ (although the presentation of μ from Eq. A.10 is preferred). The normalization is as follows Ahmad et al. (2023),

$$\kappa^* = (1 - f_\kappa)\kappa^{\mathrm{HS}-} + f_\kappa \kappa^{\mathrm{HS}+}, \tag{1.2a}$$

$$\mu^* = (1 - f_\mu)\mu^{\mathrm{HS}-} + f_\mu \mu^{\mathrm{HS}+}, \tag{1.2b}$$

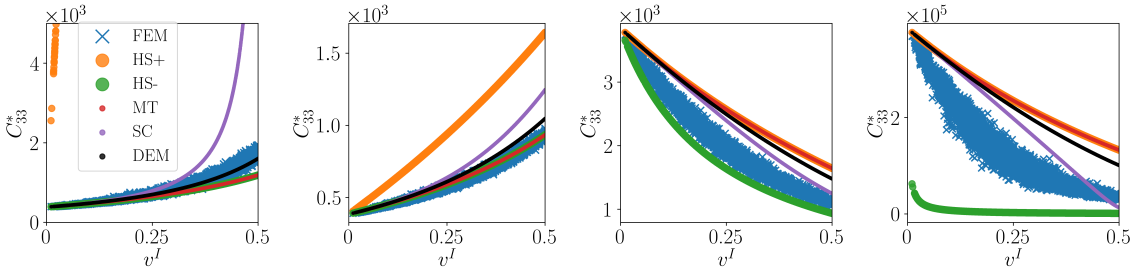
where HS− and HS+ stand for the lower and upper Hashin-Shtrikman bounds provided in A.1, and κ^* and μ^* are the ground truth homogenized values (obtained by the FEM in Ahmad et al. (2023)). The ML will directly work with normalization factors f_κ and f_μ, which are expected to be in the interval $[0, 1]$ due to the bounding nature of HS values.[†] Even when they fall slightly outside the interval, they still serve effectively as decent normalization factors for the ML.

Motivated by the discussion in Section 1.2.2, we propose a normalization for the entire stiffness tensor, as the VEs may be far from their RVE-limit isotropic response for the considered sizes. Normalizing individual values of $\mathbf{C}$ does not work properly, as some components, such as C_{13} and C_{23}, are significantly smaller than other components and have an RVE-limit of 0 (for a macroscopically isotropic material). Moreover, similar to κ^* and μ^* in Eq. (1.2), other components of $\mathbf{C}$ can vary by orders of magnitude for different materials and elastic contrast ratios. We propose the following normalization scheme,

$$\mathbf{C}^{\mathrm{norm}} = \mathbf{I} - (\mathbf{C}^{\mathrm{MFM},*})^{-1}\mathbf{C}^*, \tag{1.3}$$

where in our case, the ground truth homogenized stiffness tensor $\mathbf{C}^*$ is obtained by the FEM. The matrix $\mathbf{C}^{\mathrm{MFM},*}$ is a lower-accuracy estimate of this matrix, obtained

[†]Given that the VE may still not have reached the RVE limit, the factors can fall outside this interval. For example, the RVE limit bulk and shear moduli are reached from above and below for DBC and TBC, respectively Huet (1990); Hazanov and Huet (1994); Ostoja-Starzewski (1998). Also, the inclusions often have irregular shapes and do not exactly match the shapes used in the HS bounds. However, for the rock samples considered in Ahmad et al. (2023), the factors f_κ and f_μ remain inside $[0, 1]$



(a) $r_E = 1000$. **(b)** $r_E = 100$. **(c)** $r_E = 0.1$. **(d)** $r_E = 0.001$.

Figure 1.1: FEM and MFM values of the shear modulus ($\mu = C^*_{33}$) versus inclusion volume fraction v^I.

by one of the MFMs. $\mathbf{C}^{\text{norm}}$ represents a non-dimensional normalization matrix that measures the discrepancy between the ground truth and the MFM model; that is, $\mathbf{C}^{\text{norm}} = 0$ when $\mathbf{C}^{\text{MFM},*} = \mathbf{C}^*$. For ML, we first normalize $\mathbf{C}^*$ tensors to $\mathbf{C}^{\text{norm}}$ and develop the ML surrogate model based on $\mathbf{C}^{\text{norm}}$. Once, $\mathbf{C}^{\text{norm}}$ is predicted from an image, the *de-normalization* scheme takes $\mathbf{C}^{\text{norm}}$ to the predicted stiffness tensor of the VE from,

$$\mathbf{C}^{\text{ML},*} = \mathbf{C}^{\text{MFM},*} \left(\mathbf{I} - \mathbf{C}^{\text{norm}}\right). \tag{1.4}$$

For the MFMs, we consider the HS, MT, SC, and DEM methods. Their corresponding bulk and shear moduli are formulated in A.1. The next section compares these MFMs for normalizing the stiffness tensor.

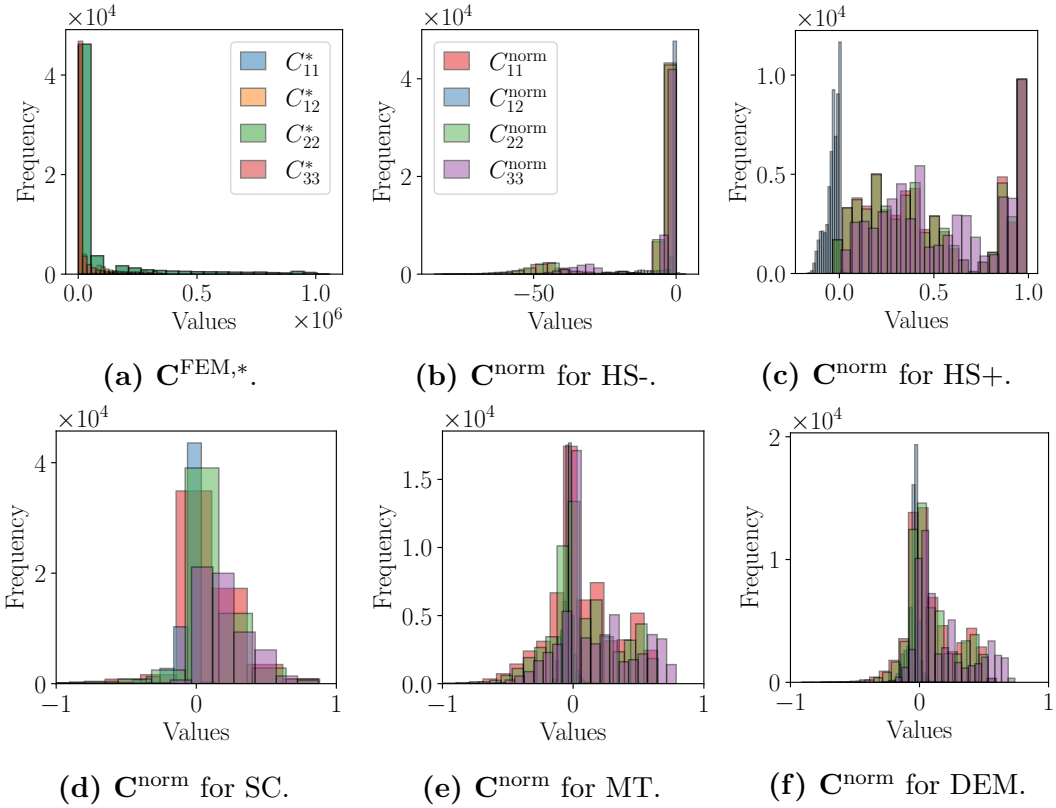
1.2.4 Performance of the normalization scheme

Figure 1.1 presents how the MFM HS-/HS+ bounds, and MT, SC, and DEM estimates compare with direct FEM calculations as a function of inclusion volume fraction (v^I). We chose the Micro2D dataset Robertson et al. (2024) for this analysis because it has a wide range of elastic contrast ratios from $r_E = 1000$ to $r_E = 0.001$. The Micro2D dataset provides the orthotropic (approximate) form of the stiffness tensor. Accordingly, only C^*_{11}, C^*_{22}, C^*_{12}, and C^*_{13} are nonzero and will be considered in the following analysis. Further details about this dataset are given in Section 1.3.1.

Figure 1.1a reveals pronounced discrepancies at the high contrast ratio of $r_E = 1000$ (stiff inclusions in a compliant matrix). HS+ provides exceptionally high values even for low v^I, while the SC overshoots starting from $v^I = 0.3$. The MT and DEM methods show better agreement with the FEM data for this extreme r_E value. In Figure 1.1b, at a moderate $r_E = 10$, all methods produce curves that are close to FEM results, with HS+ resulting in the farthest deviation from the FEM data. The fact that FEM values fall below the HS- is acceptable as the VE shear moduli can be far from the RVE limit for which the HS bounds apply. Figures 1.1c and 1.1d examine the inverse scenario with compliant inclusions in a stiff matrix at $r_E = 0.1$ and $r_E = 0.001$, respectively. For $r_E = 0.1$, the effective properties are close to the lower bound, and the other methods overestimate the stiffness. Whereas for $r_E = 0.001$, HS- significantly underestimates the stiffness for low v^I. Overall, both DEM and MT provide close estimates, although we believe the DEM estimates better match the FEM values across all r_E.

Figure 1.2 further investigates the performance of different MFMs. Figure 1.2a displays the ground truth FEM-derived stiffness tensor components (C^*_{11}, C^*_{12}, C^*_{22}, and C^*_{33}), showing highly concentrated distributions near zero with maximum values approaching 10^6. Subsequent panels show the nonzero components of $\mathbf{C}^{\text{norm}}$ from Eq. (1.3) using different MFMs.

Figure 1.2b shows that HS- yields predominantly negative values. This is expected, as by analyzing the decoupled normal ($C^*_{11} = C^*_{22}$ and C^*_{12}) and shear C^*_{33} components of the $\mathbf{C}^*$ tensor at its isotropic RVE limit, we expect the normalization correction factors to be less than zero for $C^*_{11} = C^*_{22}$ and C^*_{33}. However, the very low negative components of $\mathbf{C}^{\text{norm}}$ refer to the significant underestimation of $\mathbf{C}^*$ for the HS- bound, especially for $r_E = 0.001$ in Figure 1.1d. Figure 1.2c illustrates that HS+ produces more balanced distributions between 0 and 1. The values close to one correspond to the extremely stiff inclusions with $r_E = 1000$ in Figure 1.1a. Since HS bounds are defined for bulk and shear moduli (and extend to diagonal members of $\mathbf{C}^*$), C^*_{12}



(a) $\mathbf{C}^{\text{FEM},*}$. **(b)** $\mathbf{C}^{\text{norm}}$ for HS-. **(c)** $\mathbf{C}^{\text{norm}}$ for HS+.

(d) $\mathbf{C}^{\text{norm}}$ for SC. **(e)** $\mathbf{C}^{\text{norm}}$ for MT. **(f)** $\mathbf{C}^{\text{norm}}$ for DEM.

Figure 1.2: (a) Histograms for nonzero components of the FEM stiffness tensor $\mathbf{C}^*$. (b-f) Histograms of nonzero components of normalization matrix $\mathbf{C}^{\text{norm}}$ using different MFMs.

is not necessarily bounded. This explains the positive correction values for HS- and negative correction values for HS+ for this component.[‡]

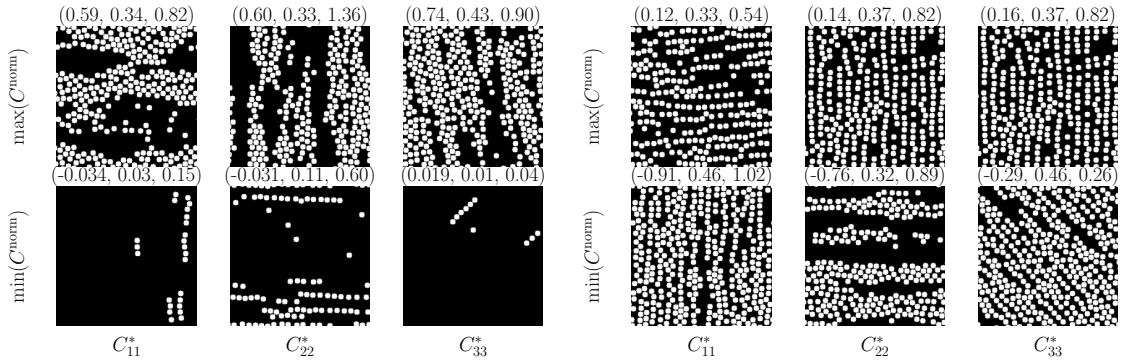
Figure 1.2d shows the SC results, which include negative values clustered near zero. There is a small set of small (up to -4) and large correction values that are not included in the range of values depicted. We believe that these values correspond to the underestimation (at high v^I) and overestimation of the SC method in Figures 1.1d and 1.1a, respectively. MT approximation generates distributions centered around zero with both positive and negative components, as shown in Figure 1.2e. DEM method, plotted in Figure 1.2f, yields a similar distribution around zero. These

[‡]Without going to details, as noted in Section 1.2, for VEs that are far from the RVE isotropic limit, the HS bounds for bulk and shear moduli may be violated especially if upper- and lower-bound DBC and TBC are used. It is noted that the PBC is used for Micro2D results.

normalization outcomes reveal that different MFMs result in distinctly different distributions for $\mathbf{C}^{\text{norm}}$.

For the well-ordered two-phase composites considered (*i.e.,* one phase simultaneously having the higher bulk and shear moduli), the MT estimate matches HS- when $r_E > 1$ (Figures 1.1a, 1.1b) and HS+ when $r_E > 1$ (Figures 1.1c, 1.1d). Given that the MT estimate matches one of the HS bounds, we believe the DEM method is better across all contrast ratios. This is particularly true for the extreme cases, such as $r_E = 1000$ and $r_E = 0.001$, where there is an even greater need to normalize the stiffness tensor. The SC significantly overestimated the shear (and bulk) moduli for the high contrast ratio of $r_E = 1000$ in Figure 1.1a. Moreover, in the case of voids ($r_E = 0$), the method fails beyond $v^I = 1/3$ in 2D due to the percolation of voids Torquato (1991). However, it is noted that we did not observe a percolation-based blow-up problem for small but nonzero $r_E = 0.001$ in Figure 1.1d. SC and MT yield similar distributions to the DEM method shown in Figure 1.2f. While MT distribution is the most similar to DEM, there is a small set of small (up to -4) values for SC which is attributed to overestimation and underestimation (at high v^I) of the SC method in Figures 1.1a and 1.1d, respectively.

The primary drawback of the DEM method is the necessity to solve an *Ordinary Differential Equation* (ODE), as illustrated in Eq. A.5 and Eq. A.6. Accordingly, the DEM is the most expensive among the MFMs at around 3,000 VE/s, while other methods, *i.e.,* MT, can be calculated at around 200,000 VE/s. Still, the computational cost of calculating the MFM values is trivial, taking less than a minute on the processor (Intel Xeon w7-3465X) for the entire dataset (59,844 samples) for all seven MFMs combined. Even for the DEM, the MFM calculations account for less than 1% of the overall training cost.



(a) $r_E = 0.001$. **(b)** $r_E = 1000$.

Figure 1.3: Microstructures with highest and lowest $\mathbf{C}^{\text{norm}}$ components for two extreme contrast ratios: $r_E = 0.001$ and $r_E = 1000$. Values in parenthesis are: $\mathbf{C}^{\text{norm}}$ component value, v^I, and $A^{\epsilon}_{\text{energy}}$.

1.2.5 Analysis of inhomogeneity and anisotropy of VEs

The MFM normalization provides a means to normalize $\mathbf{C}$ for a wide range of contrast ratios, facilitating ML. In addition, it can be used to understand which VEs have the highest discrepancy between the ground truth (FEM) stiffness tensor $\mathbf{C}^*$ and the MFM tensor (DEM used from hereon). Figure 1.3 illustrates VE examples with the highest and lowest diagonal components of $\mathbf{C}^{\text{norm}}$. For a given inclusion layout, the FEM $\mathbf{C}^*$ is the most anisotropic and the MFM is least accurate for extreme elastic contrast ratios. Hence, we only show examples from $r_E = 0.001$ in Figure 1.3a and $r_E = 1000$ in Figure 1.3b. We observe that the geometry layouts of the VEs with the highest $\mathbf{C}^{\text{norm}}$ components are highly anisotropic. For example, in Figure 1.3b, the inclusions form rather horizontal for C^*_{11}, vertical lines for C^*_{22}, and diagonal lines for C^*_{33}. We have included the corresponding $\mathbf{C}^*$ anisotropy indices in the figures. By referring to Figure 1.4b, we observe that these are rather the extreme cases of SVE anisotropy. Given that MFM estimates in this case correspond to an isotropic material, such high discrepancies through $\mathbf{C}^{\text{norm}}$ are well expected.

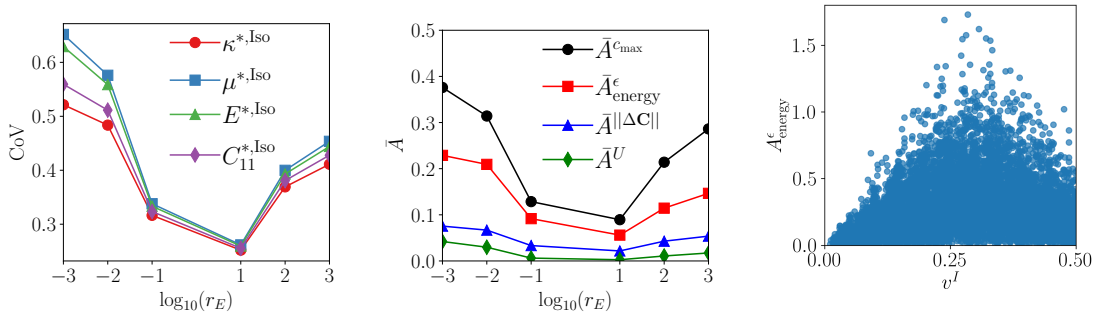
A distinct behavior is observed for the highest and lowest $\mathbf{C}^{\text{norm}}$ components in $r_E = 0.001$, depicted in Figure 1.3a. The highest $\mathbf{C}^{\text{norm}}$ components are observed in

v^I around 0.3, which are highly anisotropic. However, the lowest components have values near zero, although they are still highly anisotropic for such a low v^I. This is caused by the DEM method's tendency to overestimate stiffness for $r_E \ll 1$, as illustrated in Figures 1.1c and 1.1d. When a given method overestimates, such as HS+ as we discussed above, the upper bound of $\mathbf{C}^{\text{norm}}$ diagonals are equal to 1. To obtain a negative diagonal for $\mathbf{C}^{\text{norm}}$, the method should underestimate. However, for $r_E \ll 1$, the method almost always overestimates; thus, the lowest $\mathbf{C}^{\text{norm}}$ are obtained around 0, which correspond to very low v^I but highly anisotropic.

Figure 1.4 provides the VE population-level measures of inhomogeneity (statistical variation) and anisotropy. As expected, the highest cases of inhomogeneity and anisotropy correspond to extreme contrast ratios $r_E = 1000$ and $r_E = 0.001$. In addition to c_κ, the coefficients of variations of shear modulus $\mu^{*,\text{Iso}}$, elastic modulus $E^{*,\text{Iso}}$, and first component of stiffness matrix $C_{11}^{*,\text{Iso}}$ are provided in Figure 1.4a, as other measures of inhomogeneity.[§] Overall, by noting $c_\kappa > \varepsilon$ for the user-defined tolerance of $\varepsilon = 0.01$ (same tolerance is used in Kanit et al. (2003); Abedi et al. (2022, 2023) among others), one can claim that the VEs in the Micro2D are SVEs and have not reached their RVE limit, which holds even for the lowest statistical variations of 0.25 for $r_E = 10$. As discussed in Sections 1.1 and 1.2.1, this justifies the use of ML-based image-to-stiffness tensor prediction.

The anisotropy indices are shown in Figure 1.4b. For all r_E values, the highest anisotropy is observed for the (maximum) wave speed ($A^{c_{\max}}$), followed by elastic energy ($A^{\epsilon}_{\text{energy}}$) based indices. The indices based on the normalized discrepancy of stiffness tensor relative to its isotropic value ($A^{||\Delta\mathbf{C}||}$) and Voigt and Reuss isotropic stiffness tensors (A^U) provide notably lower anisotropy indices. This can be explained by the fact that the first two indices directly measure directional variability (θ dependency) of a quantity, whereas the latter two only use the isotropic stiffness tensor(s). Given the high sensitivity of $A^{\epsilon}_{\text{energy}}$ and the important role of energy,

[§]The latter three components are computed from the Voigt isotropic stiffness Eq. A.9, where the superscript V is dropped for convenience.



(a) Inhomogeneity indices. **(b)** Anisotropy indices. **(c)** Anisotropy index versus v^I.

Figure 1.4: Population-level (a) inhomogeneity (CoVs of elastic constants, statistical variation) and (b) anisotropy of Micro2D VEs versus r_E. (c) distribution of $A^{\epsilon}_{\mathrm{energy}}$ with v^I for $r_E = 0.001$.

we have used $A^{\epsilon}_{\mathrm{energy}}$ as the default anisotropy index; see, for example, Figures 1.3 and 1.4c. In the latter figure, we show that most anisotropic responses are observed for $v^I \in [0.2, 0.4]$ (especially 0.3), for which the fiber density is high enough to influence $\mathbf{C}$ and low enough to still generate the highly anisotropic geometric layouts shown in Figure 1.3.

Regardless of the choice, we observe that the anisotropy indices are either higher or of the same order of magnitude as the inhomogeneity index. The only exceptions would be the use of A^U for r_E values of 0.1 and 10, where the anisotropy indices are close to 0.006 and 0.003, respectively. This example further cements the argument from Section 1.2.2 and Abedi et al. (2022, 2023) that when a notable statistical variation is desired in the dataset (manifested in the form of inhomogeneity indices here), one cannot ignore the anisotropy of elasticity. Accordingly, surrogate ML models should ideally predict $\mathbf{C}$ rather than simply κ and μ. We strengthen this argument in A.3 by showing how the predictions worsen if only κ and μ are considered; for example, the worse case R^2 values across all $\mathbf{C}$ components and r_E can drop from above 0.98 to about 0.85–0.97. Moreover, the error of the worst-case SVEs can be doubled.

We summarize the role of r_E for the Micro2D dataset. Figure 1.4 shows that low contrast ratios (more compliant inclusions, $r_E < 1$) are always more inhomogeneous and anisotropic than the inverse contrast ratio (stiffer inclusions, $r_E > 1$); see for example, the results for $r_E = 0.001$ and $r_E = 1000$. This population level observation is extended to extreme VEs in terms of anisotropy indices in Figure 1.3. Interestingly, the predictions of the MFMs are also generally worse for lower contrast ratios, as shown in Figures 1.1 and 1.3. Thus, for a fiber composite with a given relative contrast ratio, the response for the compliant inclusion case is richer (more anisotropic and statistically variable), and the MFM is less accurate. Thus, a more accurate method is more necessary, such as FEM or a surrogate model.

In conclusion, we believe that the tensorial normalization Eq. (1.3) addresses key issues of bulk and shear normalization Eq. (1.2): 1) As shown in Abedi et al. (2022, 2023) and just shown for the Micro2D dataset, one often cannot assume that the VEs stiffness tensors are isotropic for the sizes that their response is sufficiently varied (inhomogeneous); 2) Even if $\mathbf{C}$ is close to the isotropic state, the normalization Eq. (1.2) is not expected to perform well at high contrast ratios: At high and low r_E, f_κ and f_μ will cluster around 0 and 1, respectively. This is specifically evident for the results in Figure 1.1a, where FEM solutions are much closer to the HS- bound for $r_E = 1000$. It is challenging to have a successful training when all VE values are clustered around 0 or 1. In contrast, the normalized values corresponding to DEM have a decent variation in Figure 1.2f.

1.3 Datasets and graph generation

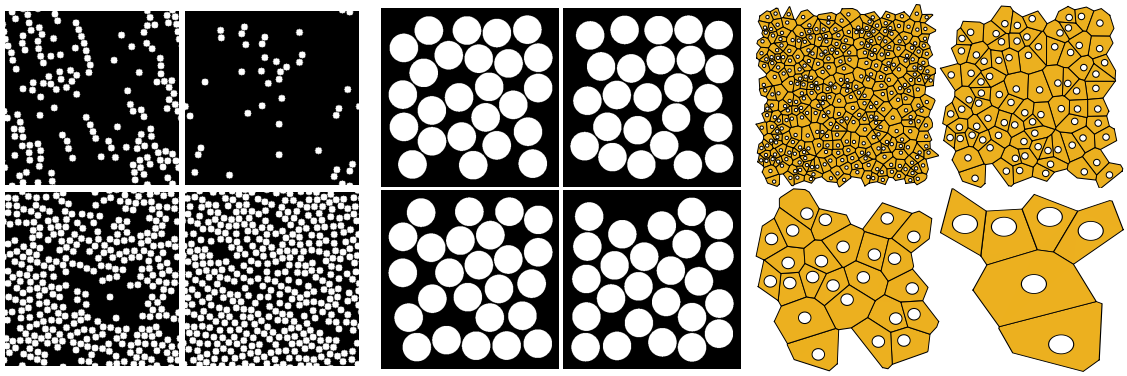
1.3.1 Datasets

In this work, three different datasets from the literature are considered. The datasets include two-phase composite images with fibers and will be used to evaluate the GNN's ability to predict both linear and nonlinear mechanical properties of such

structures. First, we will use a large dataset that describes diverse n-point statistics (Micro2D) Robertson et al. (2024) to predict the homogenized orthotropic stiffness tensor. Then, we will use a dataset generated with a phase-field damage (PFD) model Gao et al. (2023) to predict the peak strength resulting from bulk damage in the matrix material. Finally, we will investigate the generalization performance for initiation strength under interface damage using a smaller dataset Acton and Baxter (2018); Acton et al. (2018) that includes Voronoi SVEs of different sizes with homogenized stiffness values, which are predicted alongside fracture properties. The results obtained using this dataset are also used to illustrate that features calculated from Voronoi cells can improve prediction accuracy. Figure 1.5 shows some example microstructures, and Table 1.1 summarizes the properties of these three datasets.

Table 1.1: Summary of datasets.

	Micro2D Robertson et al. (2024)	PFD Gao et al. (2023)	Voronoi Acton and Baxter (2018)
Type of analysis	Linear	Nonlinear	Linear
Boundary condition	Periodic BC	Mixed BC	Mixed BC
2D assumption	Plane stress	Plane strain	Plane strain
Number of microstructures	9,974	1,000	400
r_E values	0.001, 0.01, 0.1, 10, 100, 1000	21.14	100
v^I range	0-0.5	0.5	0.1
Node features	Locations and r_E	Locations	Locations and Voronoi cell features
QoI	Stiffness	Peak strength	Stiffness and initiation strength
Challenge	Microstructure and r_E variety	Nonlinear QoI	Limited dataset size



(a) Micro2D **(b)** PFD **(c)** Voronoi

Figure 1.5: Example microstructures from three different datasets: (a) Micro2D, (b) bulk damage with phase-field damage (PFD), and (c) Voronoi SVEs with interface damage.

Micro2D

The Micro2D dataset Robertson et al. (2024) provides two-phase microstructures with diverse spatial (n-point) statistics. Traditional curation techniques rely solely on the v^I, which does not encapsulate the differences in the distribution of inclusions with respect to one another. In this work, microstructures are generated based on autocorrelations to introduce statistical variety in the dataset. Thus, the dataset contains inclusions with orientation and distribution biases, making it more challenging to develop a surrogate model that represents all structures. The effective properties of some microstructures generated in this dataset have values close to the theoretical bounds, which shows the statistical variety.

The dataset consists of 87,379 256×256 periodic microstructure images categorized into ten distinct classes based on their neighborhood distributions. In this study, we utilized only the VoidSmall class, which corresponds to a single-circle neighborhood with 9,989 microstructures. The dataset has been created with six different material contrast ratios, ranging from 10^{-3} to 10^{3} on a log-scale and a constant Poisson's ratio of 0.3 for both phases. Constitutive phases are considered isotropic, while the VEs are approximated as orhotropic. Considering different contrast ratios, the dataset comprises a total of 59,844 samples. Figure 1.5a illustrates example microstructures from this dataset.

Homogenized orthotropic elastic properties are computed with FEM under periodic boundary conditions under plane-stress. Three simulations are conducted for each material set: uniaxial tension in the X-direction, uniaxial tension in the Y-direction, and pure shear. During the simulations, a traction equivalent to 1 MPa of effective stress was applied on the driving nodes.

Bulk damage with phase-field damage (PFD)

We used the dataset provided by Gao et al. Gao et al. (2023) to predict the peak strength under the bulk damage scenario. In this dataset, the number of fibers (or

circular inclusions) for a given v^I is kept fixed, and the locations are varied. Inclusion centers are limited to a smaller square inside the microstructure to prevent damage initiation and propagation around inclusions close to the boundaries of VEs. The fiber is considered linear elastic, with a Young's modulus of 74 GPa and a Poisson's ratio of 0.2. In contrast, the matrix has Young's modulus of 3.5 GPa and Poisson's ratio of 0.35 but can also accumulate damage with a critical strain energy release rate of 10 J/m^2. The fiber-matrix interface is assumed to be perfectly bonded.

Gao et al. demonstrated that purely data-driven surrogates can suffer a marked loss of accuracy when they are evaluated on microstructures whose governing parameters (e.g., fiber radius or volume fraction) fall outside the range represented in the training set; their study showed the RMSRE for peak-load prediction almost doubled when the test radius (6 μm) was unseen during training and dropped back once a few representative samples were included in the training set, while broad, range-based sampling consistently yielded the best generalization performance. They also observed that the image-to-value model did not perform well in predicting peak loads. Thus, they used an image-to-image UNET model to predict the damage field, which is then fed into an image-to-value ML model to get the peak load.

Bulk damage in the matrix is considered using a phase-field damage model implemented in FEniCS Alnaes et al. (2015). The loading is applied axially, and the other sides are kept traction-free. Each simulation was run until the applied load dropped to 15% of the peak load. The dataset includes various v^I and fiber radius combinations, making it suitable for evaluating the ability of machine learning models to predict damage evolution and peak material strength across different microstructural configurations. Example microstructures from this dataset are plotted in Figure 1.5b. In this work, we only used the dataset labeled as Dataset A in their work, as they provide regression results for that dataset, which we will compare with the GNN regression results.

Initiation strength with Voronoi partitioning

To predict the homogenized stiffness and an *initiation* fracture strength based on interface damage, we utilized the dataset introduced by Acton and Baxter (2018); Acton et al. (2018), where researchers showed that Voronoi is a more effective method for partitioning than the traditional square grid approach for homogenization and statistical analysis. In this dataset, volume elements (VEs) are constructed by combining polygons defined by Voronoi partitioning, illustrated in Figure 1.5c. The combination is obtained by using square grids of varying sizes on a large microstructure, resulting in varying volume element sizes.

We used different numbers of SVEs from grids of different sizes, which is required to capture the statistical variance. RVEs, by definition, do not exhibit statistical variation between different samples, whereas SVEs may differ from each other significantly due to the small window size from which they are taken. We used 16 VEs from a 2×2 grid, 64 VEs from 4×4 and 8×8 grids, and 256 VEs from a 16×16 grid. Thus, 2×2 and 4×4 require four microstructures, while 8×8 and 16×16 require only one microstructure. The dataset contains a total of 400 VEs. Examples of resulting VE images from each grid size are given in Figure 1.5c.

The FEM simulations are conducted with linear elastic materials and a contrast ratio of 100. The v^I of the generated microstructures is targeted to be around 10%. The effective stiffness tensor is calculated by combining hydrostatic, pure shear, and simple shear loadings and is assumed to be fully anisotropic. Stress around every inclusion/matrix interface is tracked at 32 equidistant points to calculate the initiation strength under interface debonding conditions for normal and shear directions. This peak stress provides an *initiation* fracture strength, corresponding to the stress level at which debonding initiates. This strength is a good proxy for the peak strength of the VE when the VE failure is mostly brittle Acton et al. (2018).

The results obtained from the dataset illustrate three different takeaways in Section 1.4.4. First, multitask learning can simultaneously predict multiple quantities

of interest, such as linear stiffness tensor and fracture strength, thereby reducing computational cost. Second, smaller volume elements created from a bigger domain can be used for homogenization without many microstructure images. Finally, additional training labels computed from the geometry of the polygons of the Voronoi partitioning allow successful predictions for small dataset sizes.

1.3.2 Graph generation

Identifying inclusion centers

We reduce the input dimensionality of the microstructure images by using graph representations created from the center locations of the inclusions. This transformation of the input image to the graph captures the topology with a minimal number of input parameters. For instance, the PFD dataset consists of 512×512 images (262,144 dimensions), but our method requires only 25 dimensions—one for each fiber. The Micro2D dataset consists of 256×256 images (65,536 dimensions), which are mapped into varying numbers of nodes for each VE. The nodes in our graph representation correspond directly to inclusion centers, providing a compact representation of complex microstructures while preserving essential topological information.

Recently, the Mapper algorithm Singh et al. (2007) has been utilized to create graphs for rock samples Chung et al. (2024). This algorithm presents significant challenges due to its sensitivity to parameters and computational complexity. Its reliance on user-defined parameters, such as the lens function, number of bins, overlap ratio, and clustering method, introduces variability, as these choices heavily influence the resulting graph without a standardized selection method, often necessitating extensive trial and error Haşegan et al. (2023). Additionally, for large datasets, the algorithm's computational intensity, particularly in the clustering phase, can hinder scalability, posing practical limitations in big data contexts Tao and Ge (2025). Due to these drawbacks, we employ a more robust algorithm that leverages the circular, disconnected topology, which is well-suited for large-scale applications.

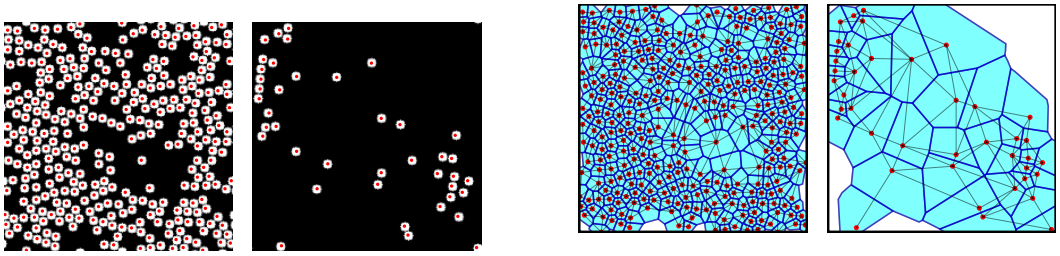
The images in the datasets are processed with OpenCV's Bradski (2000) *connectedComponentsWithStats* function to identify the center locations of the fibers. This method is suitable when the inclusions are separated from each other by the matrix material. As an alternative, distance transform and watershed algorithms can be used from the same library to address the separation limitation; however, the information on the volume of inclusions represented by a node needs to be calculated using a segmentation method, which would still be more straightforward than the Mapper algorithm. Although the latter technique may generalize GNNs to more arbitrarily shaped inclusions, fibers with a clear separation are a distinct enough class of microstructure and will be the only focus of this work.

For the Micro2D dataset, we processed the 9989 sample images in the VoidSmall class, which is suitable for the OpenCV function mentioned above. To verify that all centers are identified correctly, we calculated the v^I resulting from the topological graph and compared it with the v^I provided in the dataset. With this verification method, we identified 14 microstructures with connected inclusions for which the algorithm didn't work and thus were removed from the dataset.[¶] Example images from Micro2D with the identified center locations are plotted in Figure 1.6a. The PDF dataset didn't have any connected inclusions. Therefore, the algorithm is applied to all microstructures. Finally, the Voronoi dataset included the center locations, allowing them to be used directly.

Creating edges from a point cloud

Once the center locations are identified, we need to connect the nodes to create edges and complete the graph. In the traditional unstructured FEM mesh, areas with low proximity between geometric features are meshed with smaller elements to capture

[¶]Specifically, samples corresponding to indices [1434, 2229, 3013, 4921, 5670, 5956, 5989, 6250, 6270, 7425, 8099, 8294, 8361, 9358] are identified as having connected inclusions (indexing starts from 0). In addition, we noticed that the mechanical properties corresponding to $r_E = 10$ of index 5502 results in a $\mathbf{C}$ that is not positive definite and, thus, also removed from the set. So, in total, 15 microstructures were removed.



(a) Identified center locations (red) plotted on top of the microstructure images.

(b) The corresponding graphs after Delaunay triangulation, with Voronoi cells surrounding the generator points.

Figure 1.6: Creating graphs from microstructure images.

the stress concentrations. In contrast, a graph representation from the topology allows modeling the low proximity of the features by only having a short edge between inclusions, which is advantageous for graphs from topology since the resulting number of nodes can be several orders of magnitude less than a traditional FEM mesh. On the other hand, there is a significant increase in representational significance compared to structured grid-based methods, such as CNNs, since they do not put additional emphasis on the low proximity between geometric features. In contrast, edge length information is aggregated through the graphs in GNNs. We used the edge length and relative Cartesian coordinates of linked nodes as edge attributes (*Distance* and *LocalCartesian* transforms in PyTorch-Geometric Fey and Lenssen (2019)) for neural network architectures that support edge features, specifically PNA, as explained in Section 1.4.1.

In this study, we compared three different approaches for edge creation: radius-cutoff based (*RadiusGraph* in PyTorch-Geometric), k-Nearest Neighbors (kNN) Kramer (2013), and Delaunay triangulation. In the *RadiusGraph* method, edges are created between nodes that fall within a specified radial distance from each other. This method offers direct control over edge density through the radius parameter, but may result in disconnected subgraphs if the radius is too small. The k-nearest Neighbors method ensures each node connects to its k-closest neighbors, guaranteeing a minimum degree of connectivity for all nodes while maintaining local relationships.

Delaunay triangulation ensures that no node lies within the circumcircle of any triangle formed by three connected nodes, resulting in a unique graph for a given point cloud. This triangulation is also the dual graph of Voronoi partitioning. We used SciPy's Delaunay and Voronoi functions, which employ the Qhull library for triangulation and tessellation. Example graphs created with Delaunay triangulation for the Micro2D dataset are illustrated in Figure 1.6b.

We compared these methods for edge creation and their influence on the accuracy of the GNN models across the three datasets described in Section 1.3. The results show that *RadiusGraph* had the most variability in its results when trained with different cutoff radius values, as smaller values resulted in disconnected subgraphs. Long edges created with a large cutoff to capture long-range interactions didn't significantly benefit this problem. The other two methods, kNN and Delaunay, provided similar results that the number of edges in the Voronoi cells can explain. An illustration of the Voronoi partitioning with Delaunay triangulation is plotted in Figure 1.6b. The Delaunay triangulation graph corresponds to one-hop Voronoi neighbors, meaning the generator points are connected if the polygons share an edge. Since the average number of polygon edges in the dataset is around five and the maximum number of polygon edges is ten, creating a kNN graph with $k = 5$ yields similar graphs.

Both kNN and Delaunay constructions rely exclusively on local Euclidean proximity, so they capture essentially the same neighborhood information. In a 2D point set, the Delaunay triangulation links a generator point to every other generator whose Voronoi cell shares an edge with it; this degree is bounded above by the number of polygon edges and has an expected value of about six for a Poisson-like inclusion distribution Miles (1970), with a maximum of ten in our datasets. Selecting a number of neighbors k in kNN to match this typical degree (we used k=5 and k=10) means that each node is connected to almost exactly the same set of nearest neighbors as in the Delaunay graph. Because both methods favor short edges and ignore longer-range

interactions, they produce very similar edge-length distributions, node degrees, and ultimately identical message-passing paths for the GNN.

On the other hand, Voronoi polygons are ill-defined for the boundary nodes and require additional treatment, which is a disadvantage of the Delaunay method over the kNN method. For example, Micro2D features periodic microstructures that can be used to define cells that would otherwise be located at the boundary of the volume element. Alternatively, the volume elements in the Voronoi dataset are created by removing inclusions at the microstructure boundary. This does not affect the statistics since the original microstructure is significantly larger than the RVE limit. However, the PFD dataset lacks periodic microstructures and has a limited number of fibers, for which disregarding the boundary inclusions substantially skews the dataset. Furthermore, high v^I (0.5) volume elements resulted in concave polygons when the same Voronoi generation method was used. Thus, we used the kNN method to generate edges for the PFD dataset.

Voronoi node features

Similar to the edge features, another significant advantage of GNNs is that they enable additional node features to be embedded within the nodes in the graph. Node features have been used to extract topological features of graphs and embed them into a node vector space Hamilton (2020). For instance, node coordinates, degree, centrality, and pagerank can be added to the node vectors to represent geometric and topological properties Chung et al. (2024). In this work, we introduce the geometric properties of Voronoi cells as additional node embeddings to improve the prediction accuracy.

A Voronoi diagram partitions a plane into regions based on the distance to a specific set of points, where each region contains all points closer to its associated point than any other. This results in a unique polygonal tessellation, as shown in Figures 1.5c and 1.6b. As each polygon is unique, it contains information about the generator point (inclusion). We calculated 14 features associated with each cell and

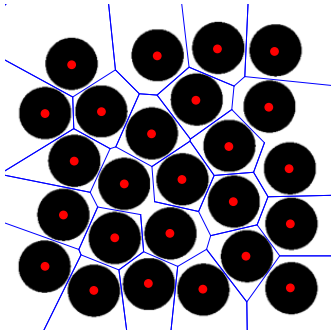


Figure 1.7: Voronoi partitioning of a microstructure from PFD dataset. Black and red dots correspond to inclusions and their centers, respectively. Blue lines represent the Voronoi cell boundaries.

added them to the node embedding vector along with the coordinates of the nodes. These additional features are area, perimeter, number of edges, centroidal (I_{uu} I_{vv}, I_{uv}) and polar (J) moments of inertia, principal moments with respect to origin (I_1, I_2), principle angles, and moments of inertia with respect to origin (I_{xx}, I_{yy}, I_{xy}). The distribution of these features for the Voronoi dataset is provided in Figure B.4.

While the Voronoi-based node embeddings substantially enrich the geometric description of the microstructure, two practical limitations should be acknowledged. (i) Boundary cells for sparse microstructures: When an SVE contains only a few inclusions, some Voronoi cells intersect the external boundary and formally extend to infinity. Because these *open* cells, illustrated in Figure 1.7, lack a well-defined polygon, geometric quantities such as centroid, area, and moments of inertia cannot be computed consistently, rendering the Voronoi features unusable for very low inclusion counts (especially if the microstructure is not periodic). (ii) Degenerate cells at high volume fractions: at high inclusion volume fractions (*e.g.*, $v^I \approx 0.50$), the limited free space forces many cells to become highly concave or self-intersecting after cropping to the SVE window. Such ill-defined polygons introduce numerical noise and occasional singularities in the calculation of shape descriptors, which can degrade model accuracy and stability. In these regimes, supplementary feature engineering or alternative tessellation strategies may be necessary to maintain predictive performance.

1.4 ML predictions for the mechanical properties

1.4.1 ML models

We utilized the HydraGNN software package to build and train the ML models, which is a distributed PyTorch implementation of multi-headed graph convolutional neural networks (GCNNs) Lupo Pasini et al. (2021). The software architecture can be divided into four distinct layers. First, the input layer contains the feature vectors representing nodes and edges, such as node coordinates and Voronoi features, as provided in Section 1.3.1, or edge lengths. Second, message-passing layers exchange information across nodes and edges in the graph, using this information to update the nodal and edge features. Third, global pooling layers aggregate node data into scalars for graph-level properties, such as the stiffness tensor, peak strength, and initiation strength. The final set of layers is the fully connected layers, or multi-task heads, which fork the architecture into separate branches, each predicting a specific property.

Message Passing Layers

For a set of nodes $\mathcal{V}$ and edges between nodes $\mathcal{E}$, a graph can be defined as $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. An edge $(u, v) \in \mathcal{E}$ connects nodes u and v, where $u, v \in \mathcal{V}$, $\mathcal{E} \in \mathcal{V} \times \mathcal{V}$. A graph is conveniently represented with and *adjacency matrix* $\mathbf{A} \in \mathbb{R}^{|\mathcal{V}|\times|\mathcal{V}|}$. The presence of an edge as an entry in this matrix is expressed as $\mathbf{A}[\mathrm{u}, \mathrm{v}] = 1$ if $(u, v) \in \mathcal{E}$ and $\mathbf{A}[\mathrm{u}, \mathrm{v}] = 0$ otherwise Hamilton (2020). The input graph $\mathcal{G}$ is accompanied by a set of node features $\mathbf{X} \in \mathbb{R}^{\mathrm{d}\times|\mathcal{V}|}$, where d is the number of node features, and edge features $\mathbf{E} \in \mathbb{R}^{\mathrm{d_e}\times|\mathcal{E}|}$, where d_e is the number of edge features. For each node $v_i \in V$ and each edge $e_{ij} \in \mathcal{E}$ between nodes u_i and v_j, the input layer encodes their initial feature vectors $\mathbf{X}_{\mathrm{i}}^{(0)} \in \mathbb{R}^{\mathrm{d}^{(0)}}$ and $\mathbf{E}_{\mathrm{ij}}^{(0)} \in \mathbb{R}^{\mathrm{d}_{\mathrm{e}}^{(0)}}$, respectively.

The message-passing layer aggregates and passes information from u's neighborhood $\mathcal{N}(u)$ to learn the representation vector of a node, also known as the *hidden*

embedding, h_u^k, or of a graph, h_G^k. Message-passing is the iterative update of h_u^k through aggregating information from $\mathcal{N}(u)$. For every k iteration, h_u^k is passed to k-hop neighbor as a message and can be expressed as:

$$m_{\mathcal{N}(u)}^{(k)} = \oplus^{(k)}(\{\mathrm{h}_v^{(k)}, \forall v \in \mathcal{N}(u)\}), \tag{1.5}$$

where $m_{\mathcal{N}(u)}^k$ is the message passed to node u from $\mathcal{N}(u)$ at k-th step through aggregation function $\oplus^{(k)}(\cdot)$, *i.e.,* neural network (NN). Then, $\mathrm{h}_u^{(k)}$ is updated with:

$$\mathrm{h}_u^{(k+1)} = \phi^{(k)}(\mathrm{h}_u^{(k)}, m_{\mathcal{N}(u)}^{(k)}), \tag{1.6}$$

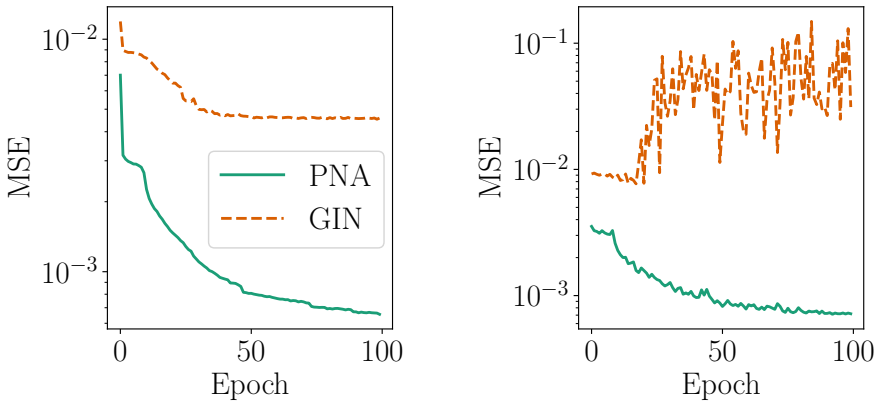
where $\phi^{(k)}(.)$ is the update function, *i.e.,* NN. Node embeddings can be used for node-level tasks such as classification. For graph-level tasks, a readout function is used at the final step K to calculate the graph embeddings Xu et al. (2018):

$$\mathrm{h}_G = \rho(\{\mathrm{h}_u^k | u \in \mathcal{G}\}), \tag{1.7}$$

where $\rho(\cdot)$ is the readout function.

The choice of $\oplus(\cdot)$ holds prominent importance in the performance of GNNs Lupo Pasini et al. (2023). For instance, Graph Isomorphism Network (GIN) Xu et al. (2019) and Principal Neighbourhood Aggregation (PNA) Corso et al. (2020) are two popular choices. GIN is widely used since it has been shown to perform as effectively as the Weisfeiler-Lehman Leman and Weisfeiler (1968) graph isomorphism test. This ensures that two topologically identical (isomorphic) graphs would provide the same result, and non-isomorphic graphs would yield a different result. The aggregate operation in GIN is expressed as:

$$\oplus^{\mathrm{GIN}}(\varepsilon^k, \mathrm{h}_u^{(k-1)}, \mathrm{h}_v^{(k-1)} : v \in \mathcal{N}(u)) = (1 + \varepsilon^k) \cdot \mathrm{h}_u^{(k-1)} + \sum_{v \in \mathcal{N}(u)} \mathrm{h}_v^{(k-1)}, \tag{1.8}$$



(a) Train set. **(b)** Test set.

Figure 1.8: Comparison of MSE loss history of GIN and PNA during training.

where ε is a learnable parameter or a fixed scalar. $\oplus^{\text{GIN}}(\cdot)$, which is used to update h_u^k with a multi-layer perceptron $\text{MLP}^{(k)}$, is referred to as *sum* aggregator. Other aggregators, such as *max* and *mean*, are also used in the literature Gilmer et al. (2017); Kipf and Welling (2016). On the other hand, PNA utilizes standard deviation (*std*), *mean*, *max*, and *min*, along with scalers. The aggregators are defined as the tensor product of the scalers and aggregators:

$$\oplus^{\text{PNA}} = \Big[\text{identity, amplification, attenuation}\Big]^{\text{T}} \otimes \Big[\text{mean, max, min, std}\Big]^{\text{T}}. \quad (1.9)$$

For brevity, we refer readers to Corso et al. (2020) for an in-depth discussion of the PNA aggregators and scalers. The PNA aggregator is used in the message-passing layer in Eq. (1.6) as:

$$\text{h}_u^{(k+1)} = \phi^{(k)}(\text{h}_u^{(k)}, \oplus^{\text{PNA}}(\text{NN}(\text{h}_\text{u}^{(\text{k})}, \text{E}_{\text{uv}}, \text{h}_\text{v}^{(\text{k})}))), \quad (1.10)$$

where E_{uv} is the edge feature, if present, and NN is a neural network.

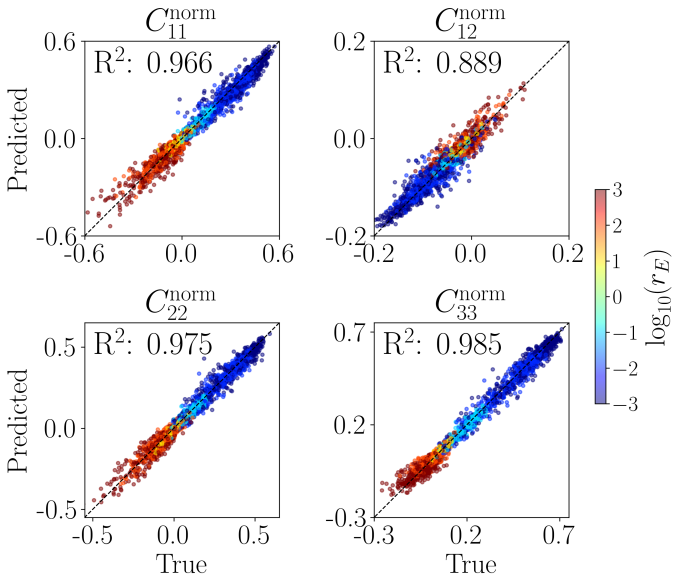
In this work, we compared GIN and PNA message-passing policies using the Micro2D dataset. We employed three message-passing layers (MPL), comprising

two shared layers and four head layers, each with 32 neurons and *tanh* activation function. Graph convolution layers are followed by batch normalization layers here and subsequently. The maximum number of k-hop neighbors is set to 14. The model is trained for 100 epochs using a batch size of 64, with AdamW optimizer Loshchilov and Hutter (2019), mean squared error (MSE) as the loss function, and adaptive learning rate (*ReduceLROnPlateau*) ranging from 10^{-6} to 10^{-4}. The model has around 50,000 learnable parameters. The train, validation, and test sets correspond to 90%, 5%, and 5% of the total dataset size, respectively. The hyperparameters used for this dataset are listed in Table B.5.

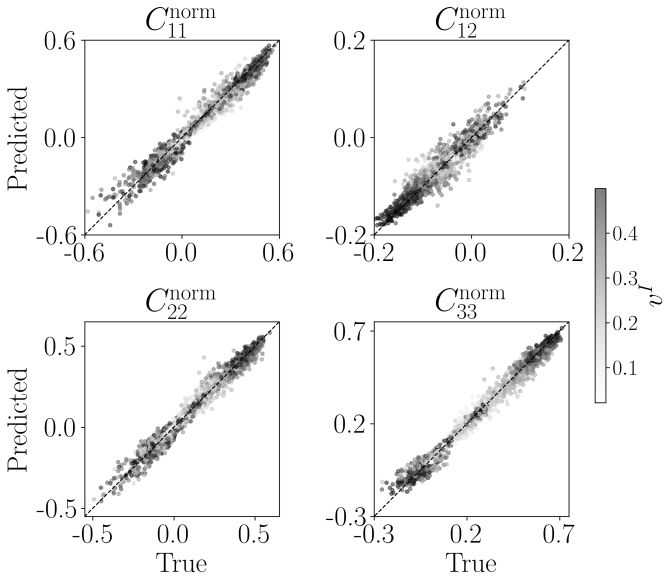
MSE histories are provided in Figure 1.8 for the train and test sets. The loss evolution indicates that GIN significantly overfits the data early during the training. However, PNA continues to learn from the data, as seen in the loss curve for the unseen test set throughout the training. Additionally, we didn't observe a bias in the parity plots for the train set (not provided here for brevity). The striking difference between these policies can be attributed to the more robust aggregator used in PNA and the incorporation of edge features. Due to the interaction of the inclusions, edge features, especially edge length, are very important for the response of the VE. Since the GIN policy does not utilize edge features and only uses the *sum* aggregator Xu et al. (2019), the train loss is higher (Figure 1.8a), and the model overfits the data (Figure 1.8b). Thus, we used PNA as the message-passing policy form hereafter.

1.4.2 Homogenization with linear elastic materials (Micro2D)

In this section, we present the numerical results obtained when HydraGNN is used for homogenization with linear elastic materials by using Micro2D. The target property that the HydraGNN model needs to predict is the $\mathbf{C}^{\text{norm}}$, as formulated in Section 1.2, since the dataset comprises different material pairs. Parity plots for the $\mathbf{C}^{\text{norm}}$ components colored with corresponding contrast ratios (r_E in log scale) are plotted in Figure 1.9a. The plot illustrates that $r_E \ll 1$ (more compliant inclusion) necessitates



(a) Values colored with $\log_{10}(r_E)$



(b) Values colored with v^I

Figure 1.9: Parity plots for the correction factors for different stiffness tensor components.

a positive correction factor for all components except the off-diagonal coupling term, indicating that DEM overestimates the stiffness (also discussed in Section 1.2.4). For this group, the predictions of $\mathbf{C}^{\text{norm}}$ are more accurate than those of $r_E \gg 1$. This discrepancy can be explained by the skewed distribution of the $\mathbf{C}^{\text{norm}}$ shown in Figure 1.2f.

$\mathbf{C}^{\text{norm}}$ *de-normalized* using Eq. (1.4) to calculate the stiffness tensors for from the GNN ($\mathbf{C}^{\text{GNN},*}$). Table 1.2 provides the mean absolute error (MAE), mean absolute relative error (MARE), maximum absolute relative error for GNN and MFM (Max ARE), and R^2 values for the $\mathbf{C}^{\text{GNN},*}$ component averages for each r_E. The table also provides the CoV values for C_{11}^{*} obtained from $\mathbf{C}^{\text{GNN},*}$ and true CoV values of isotropic $C_{11}^{*,\text{Iso}}$ from FEM (shown in Figure 1.4a earlier). Across all contrast ratios, the GNN sharply reduces both the average and worst-case relative errors. For instance, at the extreme stiff-inclusion case $r_E = 1000$, the MFM incurs a 53% mean and 131% max error, whereas the GNN limits these to 4.5% and 28%, respectively. The advantage persists in the compliant-inclusion regime ($r_E \ll 1$). Even at $r_E = 0.001$ the GNN reduces the MFM's mean error by more than half while keeping the maximum error below one-third of MFM's.

The data show that the mean absolute relative error at $r_E = 0.001$ is the highest for GNN. In addition, the mean relative error is higher overall for $r_E \ll 1$, indicating that it is more challenging to predict. In this regime, the inclusions may create percolation-type paths through the matrix, which amplify small geometric perturbations. The table shows that in addition to the stiffness tensor values, CoV is captured with near-perfect accuracy for all contrast ratios. In contrast, the MARE is higher for $r_E \ll 1$ which might be caused by orders of magnitude higher absolute values in those r_E values, and in line with the observation made from the $\mathbf{C}^{\text{norm}}$ parity plots in Figure 1.9a.

Parity plots for $\mathbf{C}^{\text{norm}}$ colored with v^I are plotted in Figure 1.9b for each component. Since the low v^I generally requires less correction, they tend to be clustered around zero and exhibit high prediction accuracy. However, the lowest

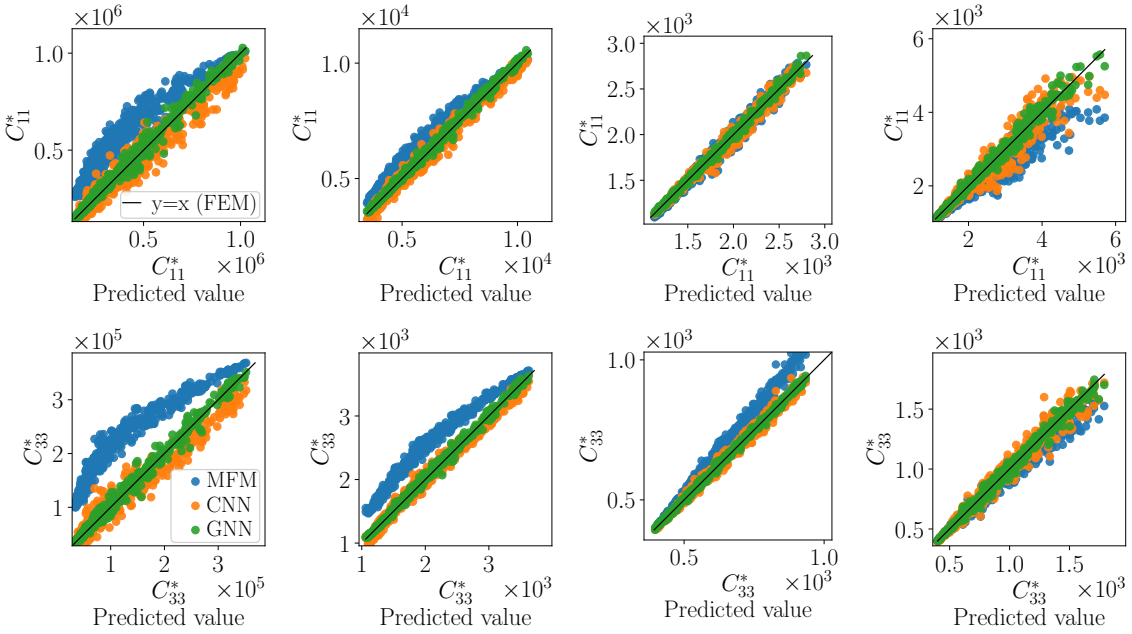
Table 1.2: Error statistics for $\mathbf{C}^{\mathrm{GNN},*}$, mean and maximum relative error for $\mathbf{C}^{\mathrm{MFM},*}$, and component averages and CoV of C_{11}^{*} for each contrast ratio.

	MAE	MARE (%)		Max ARE (%)		C_{11}^{*} statistics		
r_E	GNN	GNN	MFM	GNN	MFM	R^2	CoV	True CoV
0.001	1.2e+04	4.53	53.18	28.33	130.53	0.989	0.552	0.559
0.01	1.1e+03	3.78	40.13	22.29	96.13	0.991	0.507	0.511
0.1	6.1e+01	1.53	12.64	8.71	27.45	0.996	0.324	0.323
10	1.1e+01	0.94	5.53	4.63	14.84	0.997	0.249	0.255
100	3.0e+01	1.97	7.09	9.06	24.62	0.992	0.368	0.381
1000	4.5e+01	2.62	9.19	12.96	31.38	0.988	0.411	0.428

accuracy is not necessarily observed for high v^I. Instead, the worst predictions by GNN seem to be related to the anisotropy. Figures B.1 and B.2 (in Appendix B) illustrate the microstructures with the highest relative error for the diagonal components of the $\mathbf{C}^{\mathrm{GNN},*}$. Note that the mean relative error provided in Table 1.2 is calculated using absolute values, while the maximum values are calculated by averaging the tensor components. In contrast, the values in Figures B.1 and B.2 are represented as both positive and negative, indicating overfitting and underfitting, respectively.

As we discussed in Section 1.2.5, high anisotropy results in high values of $\mathbf{C}^{\mathrm{norm}}$ for a given v^I. The samples exhibit a clear bias in their distribution, with the inclusions generally aligned diagonally for $r_E = 0.001$, and horizontally, vertically, and diagonally aligned for $r_E = 1000$. Considering that the average A_{Gao} values corresponding to $r_E = 0.001$ and $r_E = 1000$ are 0.23 and 0.15, respectively, most of the samples are outliers. However, some microstructures with high relative error do not exhibit high anisotropy but have clustered inclusions, which may explain the high relative error.

We plotted some examples from the test set values against their ground truth (FEM) in the parity plots shown in Figure 1.10. The parity plots for the whole dataset are included in the Figure B.3. In these plots, we can see that MFM and ML methods provide fairly accurate results for $r_E \gg 1$ cases (Figures 1.10c and 1.10d). The



(a) $r_E = 0.001$ **(b)** $r_E = 0.1$ **(c)** $r_E = 10$ **(d)** $r_E = 1000$

Figure 1.10: Example parity plots for the denormalized values of the stiffness tensor. Plots are selected from low and high contrast ratios for the stiff inclusion with compliant matrix an high contrast stiff matrix, compliant inclusion cases arranged from left to right.

accuracy of the effective stiffness from MFM clearly decreases for the high contrast ratios, which corresponds to higher v^I since the inclusion is stiffer. By training the ML model to correct the difference by accounting for geometry, the effective stiffness of the high v^I samples can be improved significantly. On the other hand, MFM yields very poor predictions for $r_E \ll 1$ cases, with an R^2 value of around zero. However, when used in conjunction with HydraGNN as normalization to homogenize VEs, the R^2 value increases to over 99%.

Now, we will compare the results of HydraGNN with those from a CNN architecture, which consists of two convolutional layers followed by max-pooling operations and two fully connected layers. The first convolutional layer applies 4 filters of size 3×3 with stride 1 and padding 1 to preserve spatial dimensions, followed by a 2×2 max-pooling layer that downsamples the feature maps. The second convolutional layer increases the depth to 16 feature maps using the same kernel size, stride, and padding, and is again followed by a 2×2 max-pooling layer. The output is then flattened and passed through a fully connected layer with 128 neurons and finally through an output layer with 4 neurons, corresponding to the number of output classes or features. This model has 8 million learnable parameters compared to around 50,000 for the HydraGNN model. Figure 1.11 illustrates the R^2 values for all contrast ratios and four stiffness tensor components. First, as we explained earlier in Section 1.2, we start with a good estimate of the absolute values of the stiffness tensor by using MFM normalization, specifically when the constitutive phases are similar to each other and for low v^I. Thus, there is a small window for improvement in these cases. That said, the plot illustrates that every material pair case for all linearly independent components of the stiffness tensor can be predicted with higher accuracy by using a GNN. This improvement is especially impressive when considering the reduction of learnable parameters from 8 million with the CNN model to 50,000 for the GNN model, which significantly decreases the computational cost of training and inference.

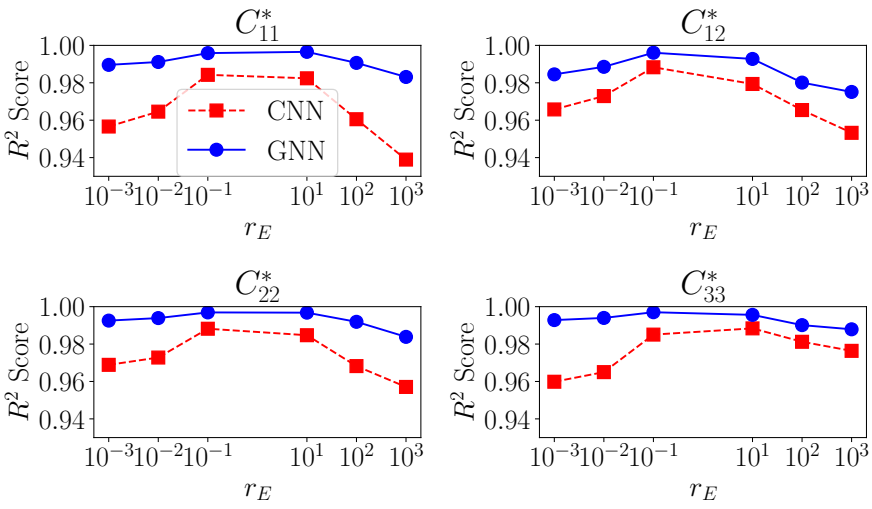


Figure 1.11: R^2 values for the GNN model compared with the CNN model.

1.4.3 Predicting peak strength under bulk damage

In this section, we will discuss the predictions of the GNN model by using the dataset explained in Section 1.3.1. The nonlinearity introduced by the phase-field damage model used in the tensile simulations of these structures makes it challenging to predict the peak strength a VE can carry using surrogate models, as noted in the original research paper Gao et al. (2023). In that work, researchers compare two approaches. First, they used UNET as image-to-value and observed poor predictions (illustrated in Figure 1.13); then, they introduced a two-step model that first maps the microstructure image to the damage field image, which is then fed into a different ML model to predict the peak strength. Although the latter approach improves the predictions compared to the first method, as it learns the peak strength from the damage field image rather than the microstructure image as we consider here, the predictions have a clear bias, resulting in lower values for lower peak strengths and higher values for higher peak strengths, similar to the first model.

For the bulk damage dataset, we considered the 0.5 v^I case labeled as dataset A, composed of 1000 samples in Gao et al. (2023). This dataset includes the peak strength values for fracture and doesn't include the homogenized stiffness tensor; thus,

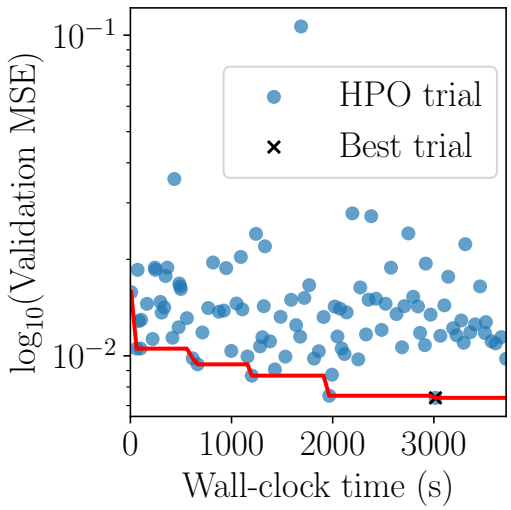


Figure 1.12: Evolution of the validation loss during the hyperparameter optimization.

we only predict a scalar value. We used PNA with edge length and local Cartesian as edge features. We observed significant improvements in predictions using the inverse of edge length (1/edge length) compared to using the edge length directly, which is attributed to the correlation between strength and the minimum distance between inclusions.

Having a much smaller dataset compared to Micro2D provided an opportunity for a parametric hyperparameter optimization (HPO) study, as training is inexpensive and HPO can be conducted without extensive resource usage. We utilized the Optuna library Akiba et al. (2019) for HPO. The initial model had an architecture with six graph convolutional layers and a hidden dimensionality of 5. The model includes a graph-level output head composed of two shared layers of dimension 5, followed by two specialized head layers with 50 and 25 units, respectively. Training is performed for 50 epochs using the AdamW optimizer with an adaptive maximum learning rate of 5×10^{-6} to 10^{-4}, a batch size of 64, the ReLU activation function, and MSE as the loss function. The model is trained on 80% of the available data, 10% is used for validation, and 10% is reserved for testing. We observed significant improvements in

predictions using the inverse of edge length (1/edge length) compared to using the edge length directly, which is attributed to the correlation between strength and the minimum distance between inclusions. The hyperparameters used for this dataset are listed in Table B.6.

For the HPO, the dimensionality of the graph convolutional layers is selected from a range of 8 to 256, and the number of layers varies from 1 to 5 to control the model's depth. The architecture of the shared layers and output heads are also optimized, with the number of layers ranging from 1 to 5 and the dimensionality of each layer independently sampled between 8 and 256 for both shared and head layers. To balance training efficiency and generalization, the batch size is selected from a categorical set of values: [1, 4, 16, 32]. The evolution of the validation loss during the HPO with 100 trials is shown in Figure 1.12. The resulting HydraGNN model comprises four graph-convolutional layers with a dimension of 196, two shared layers with a dimension of 5, and four head layers with dimensions of 78, 91, 165, and 163 neurons, with a batch size of 16. This model was trained in around 5 minutes on an NVIDIA RTX 3060 with 12 GB of memory.

Predictions from the optimized GNN model and the UNET model from Gao et al. (2023) are provided in the parity plot, along with the R^2 values shown in Figure 1.13. Our GNN model addressed this bias in the predictions, as evident from the plot. The GNN model achieves an R^2 value of 0.64, utilizing approximately 2.5 million learnable parameters, compared to UNET, which has an R^2 value of 0.29 with approximately 31 million learnable parameters (excluding the fully connected layers at the end for regression). GNN doesn't require the use of two neural networks for regression on the peak strength from the damage field to achieve a better result, as done in the previous research. Furthermore, efficiency is significantly increased since the input dimensions have decreased from 512×512 to only 25 (the number of fibers), thus reducing the training cost.

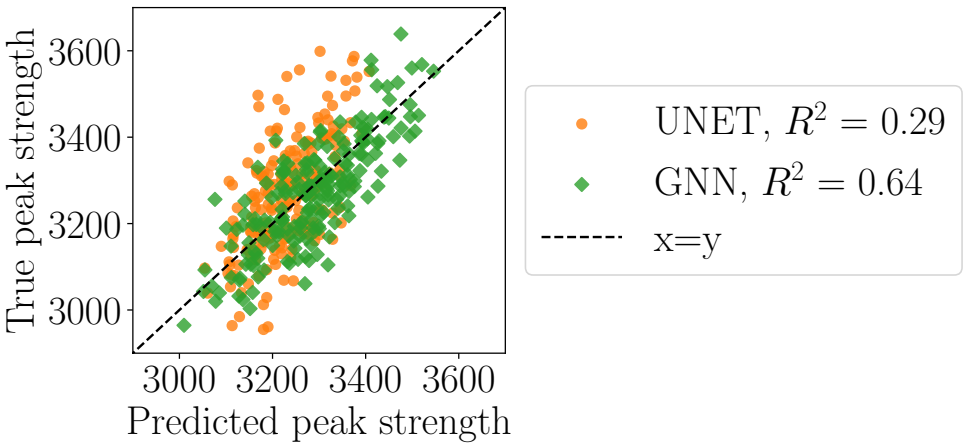


Figure 1.13: Parity plots for the peak strength predictions. UNET results are from Gao et al. (2023).

1.4.4 Predicting initiation strength under interface damage

This section shows GNN's ability to predict another important failure condition: interface damage in composites. To illustrate this, we will use the dataset explained in Section 1.3.1. In this dataset, the VEs are created from Voronoi cells, which makes them unsuitable for regular-grid methods such as CNNs since the boundaries are not rectangular, as shown in Figure 1.5c. This dataset presents the most challenging case for an ML model, as the number of samples is limited to 400.

The samples with varying sizes are shuffled before splitting. 360 samples are allocated to the training set, while the test and validation sets each have 20 samples. The GNN model for this case consists of three graph-convolutional layers and two shared layers, with dimension sizes of 256 and 64, respectively, followed by two head layers with 256 and 126 neurons. All layers used the ReLU activation function. The model has been trained using a batch size of 20 for 60 epochs with a Huber loss function. We used the AdamW optimizer with an adaptive learning rate ranging from 10^{-7} to 10^{-5}. The same model is used to compare the effect of input labels. The hyperparameters used for this dataset are listed in Table B.7.

Using only node positions as input features, as done in the previous two sections, does not provide accurate predictions of the quantities of interest, as shown in the

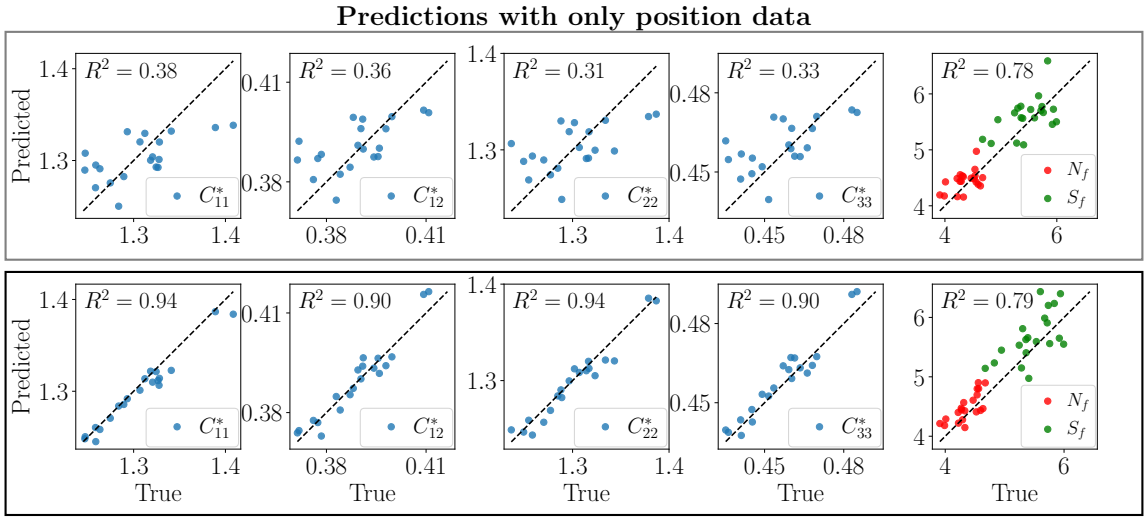


Figure 1.14: Parity plots for stiffness tensor components: comparison of the GNN with only position as node-level input with results from the GNN with Voronoi labels as node-level input.

top row of Figure 1.14, regardless of model size. To address the low accuracy issue for small datasets, we introduce additional features that leverage the properties of polygons generated by Voronoi tessellation, as explained in Section 1.3.2. Adding the Voronoi features significantly improves the predictions, as shown in the bottom row of Figure 1.14. The neural network cannot differentiate between different VEs without the Voronoi data due to the statistical similarities of the microstructures. Adding these additional features helps distinguish between otherwise similar VEs.

Figure 1.14 presents the parity plots for the stiffness tensor components (first four columns) and the initiation strengths in normal and shear directions (last column), denoted as $\mathrm{N_f}$ and $\mathrm{S_f}$, respectively. Although there is a nearly threefold increase in the R^2 values of the stiffness tensor components, the prediction accuracy for the strength does not improve significantly. Additionally, the accuracy of the strength is significantly higher than that of the stiffness tensor predictions in the position-only case, whereas it is lower for the Voronoi data-added case. Thus, we can conclude that the stiffness tensor predictions benefit from the additional node-level information

provided by the Voronoi cells, while the strength-related properties (similar to the peak strength predictions in Section 1.4.3) are more affected by the edge features.

To determine how much predictive value each of the 16 Voronoi-derived descriptors adds, we carried out a three-stage ablation analysis. First, we considered each feature by themselves. This single-feature tests showed that no individual variable can match the baseline performance (Table B.1). Then, we conducted group-wise experiments, which is either isolating (only-one-group, Table B.2) or excluding (leave-one-group-out, Table B.3) semantic bundles. We identified the shape metrics (area, perimeter, number of edges) as the most informative group. Keeping only this these three, or similarly dropping just the positional coordinates, the loss decreases from the baseline. Finally, we conducted a sequential-forward selection (Table B.4) which confirmed that the three shape metrics are chosen first and that adding further features yields diminishing returns. However, because the improvement is marginal relative to cross-validation and different datasets may benefit from the full set of descriptors, we retain all 16 features in the final model while highlighting the dominant role of the shape metrics.

1.5 Summary and conclusions

In this work, we illustrate the effectiveness of GNNs in homogenizing 2D fiber composites. Our general observations can be summarized as follows:

- **Inhomogeneity and anisotropy of VEs:** We argue that for effective ML homogenization, VEs must show significant statistical variability in homogenized properties. Using popular elasticity anisotropy indices, we show that small-size VEs (SVEs) exhibit considerable anisotropy in their stiffness tensor. Therefore, instead of predicting bulk and shear moduli alone, the entire tensor should be predicted.

- **MFM-based normalization:** We laid out an MFM-based normalization method for the effective stiffness tensor, enabling the generalization of the ML model to extreme material pairs while accounting for anisotropy.

- **Graph representation for input efficiency:** By utilizing graph representations derived from microstructure topology, we achieved a substantial reduction in input dimensionality. This approach effectively captures essential features, such as low proximity between inclusions, enabling GNNs to outperform CNNs in both accuracy and computational efficiency.

- **Performance comparison of GNN algorithms:** We evaluated two widely used message-passing GNN algorithms: GIN and PNA. Our results show that PNA outperforms GIN in mechanics-related problems, thanks to its advanced aggregators and ability to incorporate edge features, resulting in superior accuracy and robustness in property predictions.

The general insights should guide future surrogate design regardless of microstructural system, while the dataset-specific findings highlight where the present approach may need adaptation (*e.g.,* in low-contrast or limited-sample regimes). Our dataset-specific observations are as follows:

- **Contrast-ratio dependence:** For the Micro2D dataset, mean relative error is highest at $r_E = 0.001$, even though R^2 remains higher than 0.99, indicating larger absolute values dominate the fit.

- **Extension to nonlinear properties:** The proposed method and conclusions are not limited to linear elastic homogenization. We illustrated the applicability of GNNs in predicting nonlinear properties, such as peak strength under bulk damage (PDF dataset) and initiation strength under interface damage (Voronoi dataset). These properties are critical for understanding the two primary failure modes of composite structures.

- **Voronoi cell features improve small-dataset accuracy:** In the 400-sample Voronoi dataset, adding polygon-shape descriptors boosts R^2 on stiffness components from around 0.4 to around 0.9, whereas strength predictions benefit more from edge-length features. This technique does not improve accuracy for very high volume fraction volume elements (such as in PFD), as the Voronoi cells are ill-defined. In addition, when the number of samples is high and different contrast ratios are considered (such as in Micro2D), the method either provides marginal improvements at best (*e.g.,* $r_E = 0.001$, 0.989 to 0.993). In some cases, it may decrease the accuracy (*e.g.,* $r_E = 1000$, 0.992 to 0.99) since the contrast ratio is the most important feature in the input.

Building on the present framework, future work can focus on enriching the node-level descriptor set to encode anisotropic or polydisperse shapes (e.g., ellipsoids with variable aspect ratio). Furthermore, model generalization or extrapolation performance can be studied. Domain adaptation and transfer learning techniques can be used to port the trained GNN to entirely new morphology classes (e.g., connected or graded inclusions) with minimal additional data, and incorporating lightweight physics-based constraints during training to further stabilize extrapolation beyond the originally sampled parameter bounds.

Even though we used only 2-phase composites, the same methodology would apply to n-phase composites by adding the phase label to the corresponding nodes. Adding global attention read-out (GATv2) after the final layer may help with capturing long-range fiber chains, specifically for highly anisotropic microstructures. In addition, introducing a graph-level spectral positional encoding that injects coarse anisotropy cues (such as the anisotropy indices) may also help with these cases. Furthermore, an SO(3) equivariant GNN architecture can be applied to predict the effective stiffness and nonlinear properties, leveraging the rotational symmetry.

CRediT authorship contribution statement

Erdem Caliskan: Writing – original draft, Visualization, Investigation, Validation, Methodology, Software, Data curation, Formal analysis. **Reza Abedi:** Writing – review and editing, Data curation, Conceptualization, Formal analysis, Methodology. **Massimiliano Lupo Pasini:** Writing – review and editing, Supervision, Conceptualization, Project administration, Funding acquisition, Methodology.

Code availability

The open-source HydraGNN library Lupo Pasini et al. (2021), used to generate the numerical results described in this article, can be accessed at the following GitHub repository: https://github.com/ORNL/HydraGNN/tree/JMADE-2025. The scripts used in this work are located under examples/fiber_composite folder.

Data availability

The Micro2D dataset from Robertson *et al.*, the PFD dataset from Gao *et al.*, and the Voronoi dataset from Acton *et al.* are referenced as reference numbers Robertson et al. (2024), Gao et al. (2023), and Acton and Baxter (2018), respectively, in the bibliography of the manuscript.

Acknowledgements

Massimiliano Lupo Pasini thanks Dr. Vladimir Protopopescu for his valuable feedback in the preparation of the manuscript. Erdem Caliskan would like to thank Professor Katherine Acton for providing the Voronoi dataset. Erdem Caliskan also appreciates the help from Professor Ravindra Duddu for providing the phase-field damage dataset.

This research is sponsored by the Artificial Intelligence Initiative as part of the Laboratory Directed Research and Development (LDRD) Program of Oak Ridge National Laboratory, managed by UT-Battelle, LLC, for the US Department of Energy under contract DE-AC05-00OR22725. This work used resources of the Oak Ridge Leadership Computing Facility, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725, under INCITE award CPH161. Reza Abedi acknowledges the partial support of his time from Air Force Office of Scientific Research (AFOSR) under grant number FA9550-22-1-0359. The authors would also like to thank the University of Tennessee Infrastructure for Scientific Applications and Advanced Computing (ISAAC). A portion of the computation for this work was performed with ISAAC computational resources.

Chapter 2

Multitask Graph Neural Networks for Elastoplastic Response Prediction in Dual-Phase Polycrystals

A version of this chapter is under review as: Erdem Caliskan, Anik Das Anto, Massimiliano Lupo Pasini, Stephanie TerMaath, and Reza Abedi, "Multitask Graph Neural Networks for Elastoplastic Response Prediction in Dual-Phase Polycrystals," Journal of Materials Science: Materials Theory, under review.

Abstract

Microstructure-sensitive prediction of elastoplastic response remains a recurring bottleneck in multiscale damage and fatigue modeling, where large ensembles of statistically distinct polycrystals are required to quantify variability and extreme-value behavior. In this work, we develop a multitask graph neural network (GNN) surrogate that maps dual-phase ferrite–martensite polycrystal microstructures to Statistical Volume Element (SVE)-level elastoplastic quantities of interest. Each SVE is represented as a grain-adjacency graph, with node features encoding phase, geometry, and crystallographic orientation, and edge features encoding relative misorientation. A message-passing graph convolution generates node embeddings, which are pooled into a graph representation and passed to a multitask regression head that jointly predicts 10 scalar QoIs and vector-valued stress–strain responses in orthogonal loading directions across multiple martensite volume fractions and SVE sizes. Results show high accuracy for scalar QoIs and strong agreement for full stress–strain trajectories, with population envelopes reproducing both median behavior and finite-SVE variability across compositions and partition scales. A unified model trained on pooled volume-fraction data preserves most within-regime accuracy relative to regime-specific models while also capturing the broader cross-regime variation reflected in the pooled test set. Distributional comparisons further demonstrate that the surrogate preserves heterogeneity under SVE partitioning, enabling statistically consistent block-wise random-field construction for mesoscale analyses. Overall, the proposed grain-graph surrogate provides a practical pathway

to accelerate ensemble-based studies of constitutive variability relevant to crack nucleation and early-stage damage.

2.1 Introduction

Polycrystalline metals exhibit a strong microstructure-sensitive elastoplastic response where grain morphology, crystallographic texture, and phase topology jointly shape the onset of yielding, the evolution of hardening, and the directional character of the apparent stiffness. This sensitivity is central to multiscale modeling workflows where macroscopic predictions must reflect not only mean constitutive behavior but also variability, tails of distributions, and spatially localized hot spots that drive early-stage damage. In practice, these requirements lead to ensemble computations over many statistically distinct microstructures, and the cost of repeated high-fidelity simulations becomes the dominant bottleneck.

Computational homogenization formalizes this scale-bridging task through the use of *Representative Volume Element* (RVE), which provides an effective representation of the material's behavior once scale separation is achieved Sab (1992); Ostoja-Starzewski et al. (2007); Nemat-Nasser and Hori (2013); Stapleton et al. (2016). However, in many applications, the relevant domain is not sufficiently large to reach an RVE limit, the RVE size can be prohibitively large (particularly for nonlinear properties), or the RVE may not exist in a strict sense for certain responses such as softening and post-peak behavior Ostoja-Starzewski (1998); Kanit et al. (2003); Gitman et al. (2007); Pélissou et al. (2009); Abedi et al. (2023). These limitations motivate *Statistical Volume Elements* (SVEs), which provide a continuum description via *apparent* properties that retain finite-size variability and directional dependence Baxter and Graham (2000); Huyse and Maes (2001); Ostoja-Starzewski (2006); Tregger et al. (2006); Salmi et al. (2012). From an engineering standpoint, SVEs offer a controllable compromise: one can trade computational cost against the degree of heterogeneity preserved, while still maintaining a consistent mapping from

microstructure to homogenized quantities across the sizes at which variability remains non-negligible Garrard and Abedi (2020a,b); Anto et al. (2024). Most importantly, by resulting in an inhomogeneous and random coarse-grained representation of the material microstructure, they can systematically relate material aleatory uncertainties to the distribution of macroscopic Quantities of Interest (QoIs). This propagation of statistics across scales cannot be achieved by using RVEs.

A common route to SVE-based mesoscale modeling is the moving-window method, where local apparent properties are obtained by partitioning a larger microstructural domain into a hierarchy of windows Graham and Baxter (2001); Graham-Brady et al. (2003); Acton and Graham-Brady (2009a,b, 2010). This construction naturally supports random-field descriptions of constitutive parameters, which are essential when spatial variability governs macroscopic failure patterns. For example, the ability to resolve intermediate-scale variability (below the RVE limit but above the grain scale) can strongly influence predicted crack paths and failure statistics Acton et al. (2018); Garrard and Abedi (2020a); Abedi et al. (2023). For polycrystalline metals, the same logic applies to constitutive variability relevant to crack nucleation, where local yielding and hardening fluctuations can matter even when global averages appear well converged.

Despite their modeling value, SVEs impose a substantial computational burden when evaluated using full-field crystal plasticity. Crystal Plasticity Finite Element Method (CPFEM) simulations remain a standard tool for resolving grain-scale anisotropy and heterogeneous slip, but they are expensive in multi-phase settings and in regimes where large ensembles are required Asgharzadeh et al. (2023); Fuhg et al. (2022). Moreover, the SVE size needed for plastic descriptors can be substantially larger than that needed for elastic descriptors, implying that the most informative statistics are often the most costly to obtain Anto et al. (2024). This creates a practical gap for multiscale studies that require (i) many SVEs per microstructure for block-wise random-field construction, (ii) many microstructures per material state for uncertainty quantification, and (iii) nonlinear response.

Surrogate modeling has therefore become an active strategy to accelerate microstructure-property linkages. Classical reduced-order and data-driven homogenization frameworks have demonstrated that nonlinear effective response can be learned from micromechanical simulations when suitable microstructural descriptors are available Latypov and Kalidindi (2017); Latypov et al. (2019). Machine learning has also been used to boost nonlinear homogenization by coupling data-driven components with established homogenization formalisms Tannous et al. (2025). More recently, deep learning has been used to approximate crystal plasticity response and related field mappings, including approaches that target CPFEM-based predictions or embed learned constitutive components into finite element workflows Heidenreich et al. (2023); He et al. (2024); Pasparakis et al. (2025); Mao et al. (2025). Neural operator formulations further extend this line of work by learning mappings between function spaces for stress or field prediction, with recent efforts introducing physics-aware constraints, such as preserving equilibrium Khorrami et al. (2024). While these developments are promising, microstructure representation remains a recurring challenge: image/voxel-based encodings can be data-intensive and resolution-tied, and they struggle to accommodate variable numbers of grains across SVE sizes without careful architectural choices.

Graph representations provide a principled alternative because they match the discrete topology of polycrystals. In geometric deep learning, graphs provide a native representation for irregular domains and their relational structure Bronstein et al. (2017); Wu et al. (2020); Sanchez-Lengeling et al. (2021). For polycrystalline microstructures, grains can be treated as nodes and grain-boundary neighborhoods as edges, enabling the encoding of crystallographic state at the node level and misorientation-driven interactions at the edge level. Graph-based GNN surrogates have also demonstrated accurate homogenization in composite microstructures using unstructured topology-derived graphs, including prediction of tensorial stiffness and strength measures across diverse configurations Caliskan et al. (2025). This representation supports size variability (different grain counts per SVE), preserves physical

interpretability, and allows message passing to model short-range mechanisms that dominate heterogeneous response. Recent mechanics-focused studies have further demonstrated the utility of GNNs for material structure-property prediction across diverse settings, including microstructure-informed elasticity and polycrystal-related tasks Zhao et al. (2024); Dai et al. (2021, 2023); Qin et al. (2024). Graph-based learning has also been explored in fatigue response predictions, where localized values are of primary interest, thereby reinforcing the suitability of grain graphs for capturing rare yet consequential response modes Sim et al. (2025). In addition, open software ecosystems for graph learning and scientific GNN training reduce barriers to scalable surrogate deployment Fey and Lenssen (2019); Lupo Pasini et al. (2025, 2023).

In this work, we develop a deterministic multitask graph neural network surrogate for predicting SVE-level elastoplastic responses in dual-phase polycrystals. Each SVE is represented as a grain-adjacency graph with node features encoding phase, geometry, and crystallographic orientation, and with edge features encoding relative misorientation. The surrogate learns a shared microstructure embedding via Principal Neighbourhood Aggregation (PNA) Corso et al. (2020) message passing and predicts both scalar QoIs, including stiffness components, isotropic moduli, yield strengths, and hardening descriptors, as well as vector-valued stress–strain responses in orthogonal loading directions. Beyond pointwise accuracy, we evaluate whether the surrogate preserves finite-SVE variability across sizes and compositions in a manner consistent with SVE-based mesoscale modeling.

The main contributions of this study are:

- A grain-graph representation and multitask GNN surrogate for mapping polycrystal SVEs to elastic and plastic homogenized quantities across multiple SVE sizes.

- Joint prediction of scalar properties and full stress–strain trajectories within a shared embedding framework, supporting nonlinear response characterization beyond scalars.

- Validation at the population level (distributions and variability) to assess suitability for SVE-based block-wise random-field construction and ensemble-driven multiscale analyses.

2.2 Dataset Description

The dataset employed in this study is derived from the high-fidelity CPFEM simulations reported in our prior work Anto et al. (2024). In that work, the size dependency of homogenized elastic and plastic properties of dual-phase polycrystalline materials was systematically investigated using the SVE framework. The present study uses that validated computational database to construct a structured microstructure-property dataset suitable for deterministic surrogate modeling.

2.2.1 Material System and Microstructure Generation

The material system consists of dual-phase ferrite-martensite polycrystals. Three microstructural configurations were generated corresponding to martensite volume fractions (v^I) of 10% (M1), 45% (M2), and 90% (M3). Each parent microstructural domain has dimensions $400\,\mu\text{m} \times 400\,\mu\text{m} \times 1\,\mu\text{m}$ with an average grain equivalent diameter of $\bar{d} = 5\,\mu\text{m}$. The grains are modeled as equiaxed, and crystallographic orientations are assigned randomly from a uniform orientation distribution function, resulting in an effectively untextured microstructure. Phase assignment is performed at the grain level according to the prescribed martensite volume fraction, producing statistically representative dual-phase microstructures with heterogeneous yet controlled spatial distributions. Phase-dependent elastic constants and crystal plasticity parameters were assigned at the grain level. Microstructures were generated using DREAM.3D Groeber and Jackson (2014) and discretized for full-field CPFEM simulations.

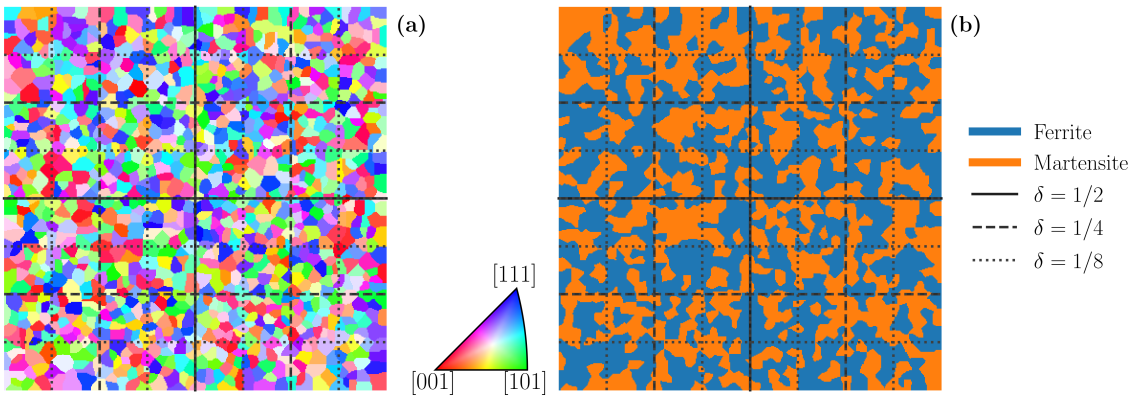


Figure 2.1: Example 400 μm $\times$ 400 μm parent microstructure showing (a) grain orientations colored according to the inverse pole figure (IPF) key and (b) phase distribution for ferrite and martensite. Grid lines indicate the SVE partitioning levels corresponding to $\delta = 1/2$, $1/4$, and $1/8$.

This simplified microstructure is used to isolate the effects of phase fraction and SVE size, without introducing additional variability from texture or anisotropic grain morphology. Other features such as grain elongation, crystallographic texture, and phase clustering can influence the elastoplastic response, but are not considered here. These effects are left for future work.

2.2.2 SVE Extraction

To capture spatial variability and size effects, the SVEs were extracted using a space-filling moving-window partitioning approach. Figure 2.1 illustrates an example microstructure from $v^I = 0.45$, where the grain orientations are colored using inverse pole figure (IPF) and the partitions used for SVEs are plotted over the domain. The considered SVE sizes are $L_{\text{SVE}} \in \{50, 100, 200, 400\}\,\mu$m, which tile 30 different 400 μm $\times$ 400 μm parent domains into 8×8, 4×4, 2×2, and 1×1 grids, respectively, without overlap or gaps. These correspond to size $\delta = L_{\text{SVE}}/400\ \mu\text{m} \in \{1/8, 1/4, 1/2, 1\}$. The parameter δ denotes the subdivision ratio relative to the full domain, so decreasing L_{SVE} (larger subdivision) increases the number of SVEs obtained from each parent microstructure and, consequently, the share of samples at

that size. In total, there are 30, 120, 480, and 1920 samples per v^I for $\delta = 1$, 1/2, 1/4, and 1/8, respectively. All SVEs are non-overlapping and collectively cover the entire domain.

This construction ensures that the dataset explicitly captures composition dependence through the martensite v^I, size dependence through the SVE length scale, and intrinsic statistical variability across different microstructural realizations. As a result, the database systematically encodes the coupled effects of phase contrast, observation window size, and microstructural randomness on the homogenized mechanical response.

2.2.3 CPFEM Simulation and Homogenization

Each SVE was analyzed using a full-field crystal plasticity formulation under mixed boundary conditions (MBC). For the elastic response, three independent linear simulations were performed, corresponding to normal loading in the x direction, normal loading in the y direction, and in-plane shear. These loading cases were used to determine the components of the homogenized in-plane stiffness tensor. For the elastoplastic response, two nonlinear simulations were carried out under normal loading in the x and y directions. The resulting stress–strain responses were used to compute the plastic QoIs, including directional yield strengths and hardening behavior.

Under the MBC, displacements were prescribed in the loading direction, while the remaining directions were not fully constrained. Traction-free conditions were applied in the transverse directions to allow lateral deformation. On the top and bottom surfaces, the out-of-plane displacement was constrained ($u_z = 0$) and the out-of-plane shear tractions were set to zero ($\sigma_{zx} = \sigma_{zy} = 0$). The homogenized stress and strain were obtained by volume averaging of the microscopic fields.

2.2.4 Output Quantities of Interest

For each SVE, both elastic and plastic homogenized quantities were reported.

Elastic Properties

The elastic quantities were obtained from the linear simulations of each SVE. The homogenized stress and strain were expressed in Voigt notation as

$$\boldsymbol{\sigma} = \begin{bmatrix} \sigma_{xx} \\ \sigma_{yy} \\ \sigma_{xy} \end{bmatrix}, \qquad \boldsymbol{\gamma} = \begin{bmatrix} \varepsilon_{xx} \\ \varepsilon_{yy} \\ 2\varepsilon_{xy} \end{bmatrix}, \tag{2.1}$$

and the homogenized elastic constitutive relation was written as

$$\boldsymbol{\sigma} = \mathcal{C}\,\boldsymbol{\gamma}, \tag{2.2}$$

where the in-plane homogenized elastic stiffness tensor is given by

$$\mathcal{C} = \begin{bmatrix} \mathcal{C}_{11} & \mathcal{C}_{12} & \mathcal{C}_{13} \\ \mathcal{C}_{12} & \mathcal{C}_{22} & \mathcal{C}_{23} \\ \mathcal{C}_{13} & \mathcal{C}_{23} & \mathcal{C}_{33} \end{bmatrix}. \tag{2.3}$$

Here, $\mathcal{C}_{33}$ denotes the in-plane shear stiffness.

The elastic outputs include the in-plane stiffness components $\mathcal{C}_{11}, \mathcal{C}_{12}, \mathcal{C}_{22}, \mathcal{C}_{33}$, and bulk modulus κ together with isotropic shear modulus μ obtained through Reuss angular averaging of $\mathcal{C}$, *cf.* Ostoja-Starzewski (1998). Although the full in-plane stiffness tensor in Eq. 2.3 is computed from the three CPFEM loading cases for every SVE, the present dataset is dominated by the orthotropic components $\{\mathcal{C}_{11}, \mathcal{C}_{12}, \mathcal{C}_{22}, \mathcal{C}_{33}\}$. Because the coupling terms $\mathcal{C}_{13}$ and $\mathcal{C}_{23}$ remain consistently close to zero across the dataset, only these orthotropic-dominant components are retained for learning and reporting.

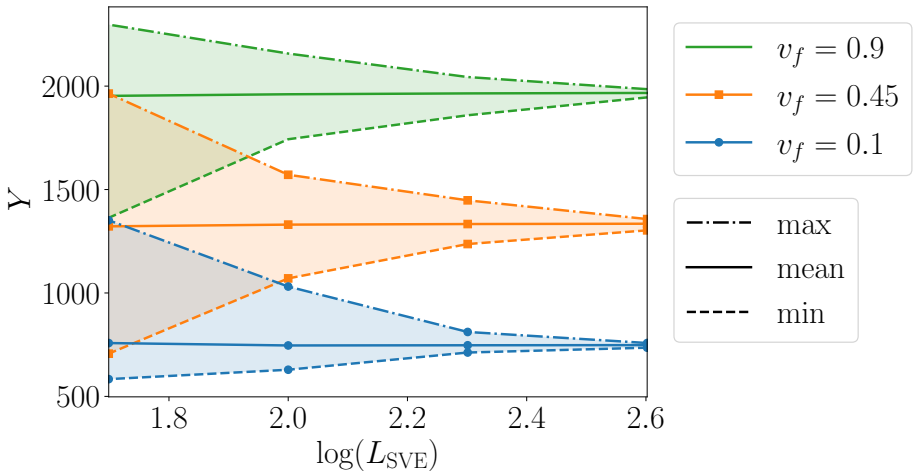


Figure 2.2: Variation of the average yield strength Y (MPa) with SVE size for three martensite volume fractions plotted against $\log(L_{\mathrm{SVE}})$. Solid lines denote the mean response across SVE realizations, while dashed and dash-dotted lines represent the minimum and maximum values, respectively.

Plastic Properties

Plastic quantities were extracted from the homogenized stress–strain curves. The yield strength Y was determined using the 0.2% offset criterion. For the loading case with normal loading in the x direction, the response σ_{xx}-ε_{xx} was used to compute Y_x, while for the loading case with normal loading in the y direction, the response σ_{yy}-ε_{yy} was used to compute Y_y. The reported yield strength was defined as $Y = (Y_x + Y_y)/2$.

The directional hardening moduli H_x and H_y were obtained from the corresponding σ_{xx}–ε_{xx} and σ_{yy}–ε_{yy} responses for loading in the x and y directions, respectively, using linear regression over the final six increments in the stabilized hardening region. The tail hardening modulus was defined as $H = (H_x + H_y)/2$. The homogenized stress–strain responses $\sigma_{xx}(\varepsilon)$ and $\sigma_{yy}(\varepsilon)$ were sampled at 24 uniformly spaced logarithmic strain increments over the loading range up to $\varepsilon = 0.095$, corresponding to approximately 10% engineering strain.

2.2.5 Statistical Characteristics

The dataset exhibits several key characteristics. Property variance decreases with increasing SVE size, indicating the convergence toward the representative-volume-element behavior. An example convergence is given in Figure 2.2. Elastic and plastic properties exhibit distinct scaling trends, with plastic quantities, particularly the hardening modulus H, showing larger variability and slower convergence. Intermediate martensite fractions exhibit increased anisotropy and statistical dispersion. Furthermore, property distributions at small SVE sizes exhibit the highest level of anisotropy and inhomogeneity. Overall, the dataset spans multiple compositions, length scales, and mechanical regimes, providing a comprehensive microstructure-property database suitable for surrogate modeling of SVE-level homogenized response.

2.3 Graph Neural Network Framework

Figure 2.3 summarizes the grain-graph surrogate used to predict SVE-level elastoplastic QoIs from dual-phase polycrystal microstructures. The framework maps each SVE to a graph, performs message passing to obtain node embeddings that encode local neighborhood interactions, and then aggregates these embeddings into a compact SVE representation that drives a multitask regression head.

Each SVE is represented as a graph $G = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the set of microstructural entities (nodes) and $\mathcal{E}$ is the set of neighborhood relations (edges). We use a grain-based VE representation because it significantly reduces input dimensionality while preserving microstructure. In grain-based graphs, each node corresponds to a grain. This choice provides a physically interpretable representation in which nodes align with crystallographic state variables (orientation, phase) and geometric descriptors (grain size).

Edges encode short-range microstructural interactions. In the grain-graph setting, we connect nodes using adjacency across grain boundaries: two grains i and j are

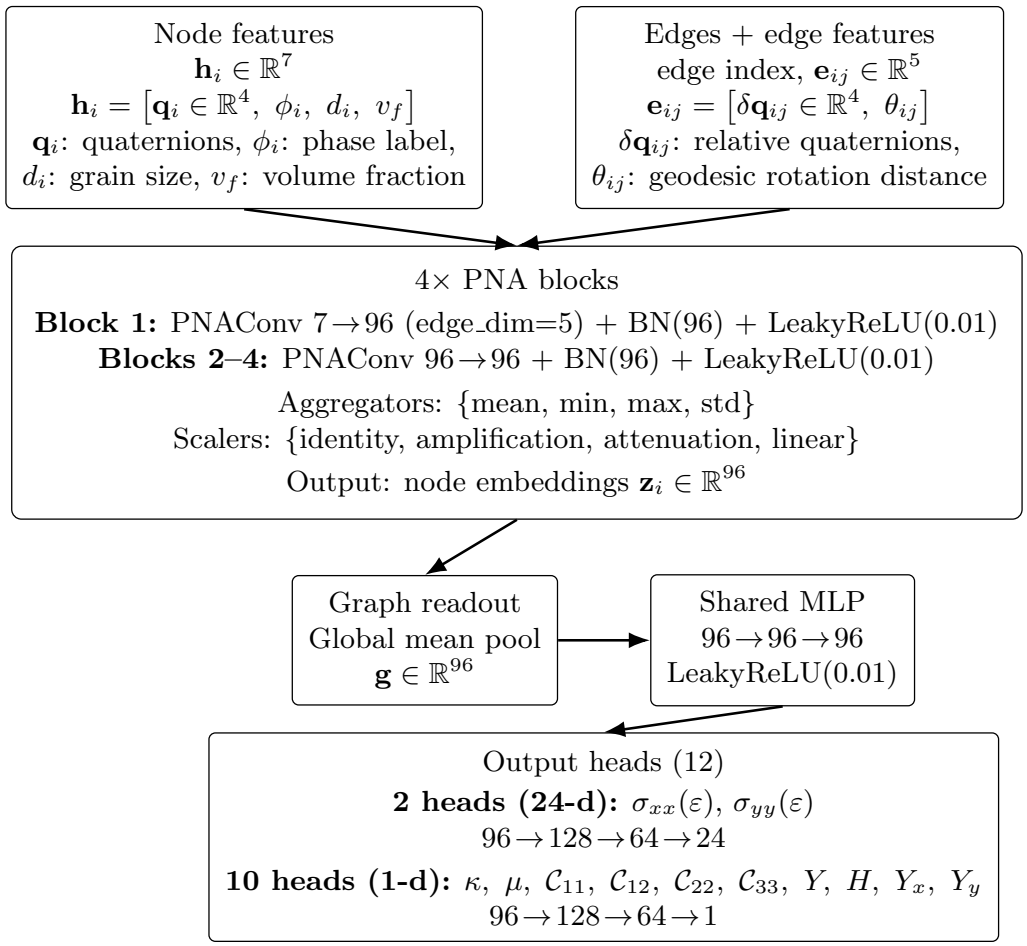


Figure 2.3: Simplified PNA-based GNN: node features $\mathbf{h}_i = [\mathbf{q}_i, \phi_i, d_i, v^I]$, edge features $\mathbf{e}_{ij} = [\delta\mathbf{q}_{ij}, \theta_{ij}]$, 4 PNA blocks using the shown aggregators/scalers, global mean pooling, shared Multilayer Perceptron (MLP), and 12 prediction heads.

neighbors if they share an interface in the tessellation/segmentation. We treat the graph as undirected in the physical sense, but implement it with two directed edges $(i \to j)$ and $(j \to i)$ for message passing. This construction produces sparse graphs with locality aligned to the mechanisms that drive effective response: Phase connectivity, crystallographic misorientation, and neighboring grains.

Each node $i \in \mathcal{V}$ is assigned a feature vector $\mathbf{h}_i \in \mathbb{R}^7$ (Fig. 2.3)

$$\mathbf{h}_i = \left[\mathbf{q}_i \in \mathbb{R}^4,\ \phi_i,\ d_i,\ v^I\right], \tag{2.4}$$

where $\mathbf{q}_i$ is the grain orientation encoded as a unit quaternion,* ϕ_i is a phase label (binary encoding for ferrite/martensite), d_i is a grain size descriptor (e.g., number of elements), and v^I is the SVE-level volume fraction. The scalar v^I is replicated to each node to provide global compositional context during message passing.

Each directed edge $(i, j) \in \mathcal{E}$ carries an edge feature vector $\mathbf{e}_{ij} \in \mathbb{R}^5$

$$\mathbf{e}_{ij} = \left[\delta\mathbf{q}_{ij} \in \mathbb{R}^4,\ \theta_{ij}\right], \tag{2.5}$$

where $\delta\mathbf{q}_{ij}$ is the relative quaternion between grains i and j, and θ_{ij} is the geodesic rotation distance, providing a scalar misorientation measure. This feature design separates node-level descriptors (state and geometry) from edge-level descriptors (pairwise crystallographic relationships), which is particularly relevant for plasticity-sensitive QoIs.

The surrogate is a graph neural network with four message-passing (PNA) layers. In each layer, every node (grain) updates its internal feature vector by combining its own information with information coming from its neighboring grains through the grain-boundary edges. We start from 7 input features per node and 5 features

*Grain orientations are represented with unit quaternions rather than Euler angles because quaternions provide a smooth, singularity-free description of 3D rotations. In contrast, Euler angles can suffer from gimbal-lock-type singularities and discontinuities, and their representation is more redundant under crystal symmetry. For learning on grain graphs, quaternions therefore provide a more stable basis for encoding orientations and computing relative misorientations between neighboring grains.

per edge; the first PNA layer maps the node description from $\mathbb{R}^7$ to a 96-component hidden representation, and the next three layers keep the same 96-dimensional size while progressively refining it as shown in Figure 2.3. The output of these four layers is a learned 96-dimensional vector $\mathbf{z}_i \in \mathbb{R}^{96}$ for each node, which is called a node embedding, providing a compact numerical summary of the local microstructural context around grain i.

PNA is designed to perform well when different nodes have very different numbers of neighbors. To achieve this, it summarizes each node's neighborhood in several complementary ways using the aggregators $\{\text{mean}, \text{min}, \text{max}, \text{std}\}$, and then adjusts these summaries with degree-dependent scalers $\{\text{identity}, \text{amplification}, \text{attenuation}, \text{linear}\}$. Using multiple summaries, together with degree-aware rescaling, makes the layer more stable and robust across SVEs whose grain-adjacency graphs can vary in connectivity.

Node embeddings are reduced to an SVE representation via global mean pooling

$$\mathbf{g} = \frac{1}{|\mathcal{V}|} \sum_{i \in \mathcal{V}} \mathbf{z}_i, \qquad \mathbf{g} \in \mathbb{R}^{96}. \tag{2.6}$$

Mean pooling provides a stable summary when the number of grains varies across SVEs and partition scales, a significant advantage over CNNs. A shared Multilayer Perceptron (MLP) refines the global mean pooling by adding three more layers and using LeakyReLU (0.01) as the activation function. The shared latent is then mapped to 12 task-specific output heads:

- Two 24-dimensional heads for stress–strain responses, $\sigma_{xx}(\varepsilon)$ and $\sigma_{yy}(\varepsilon)$, each parameterized by $96 \to 128 \to 64 \to 24$.
- Ten scalar heads for $\{\kappa, \mu, \mathcal{C}_{11}, \mathcal{C}_{12}, \mathcal{C}_{22}, \mathcal{C}_{33}, Y, H, Y_x, Y_y\}$, each parameterized by $96 \to 128 \to 64 \to 1$.

This structure enforces a shared microstructure embedding while allowing the final mappings to specialize by the QoI type (scalar versus curve).

Training minimizes a multi-output mean squared error (MSE) aggregated over all heads. Let the set of tasks be $\mathcal{T} = \{1, \ldots, 12\}$, and let task t have output dimension d_t (with $d_t = 1$ for scalar QoIs and $d_t = 24$ for stress–strain heads). For a graph G, predictions are $\hat{\mathbf{y}}^{(t)}(G)$ and targets are $\mathbf{y}^{(t)}(G)$. The combined loss function is

$$\mathcal{L} = \frac{1}{|\mathcal{T}|} \sum_{t \in \mathcal{T}} \frac{1}{d_t} \left\| \mathbf{y}^{(t)}(G) - \hat{\mathbf{y}}^{(t)}(G) \right\|_2^2 . \tag{2.7}$$

The per-task normalization by d_t prevents the 24-dimensional curve heads from dominating the optimization purely by dimensionality. In practice, targets are standardized using statistics computed on the training set to reduce scale imbalance across elastic constants and plastic metrics; predictions are then mapped back to physical units for reporting.

To prevent information leakage arising from correlated SVEs extracted from the same parent microstructure, we perform an image-level split. Parent microstructure images are first divided into train/validation/test sets. Then, all SVEs derived from a given parent image are assigned to the same subset, including the full hierarchy of partitions (e.g., the $\delta = 1/1$ SVE and its corresponding $\delta = 1/2$, $\delta = 1/4$, and $\delta = 1/8$ sub-SVEs). We target equal representation for each martensite v^I regime within each subset (stratified by v^I), and we preserve representation across SVE sizes through hierarchical assignment. In the representative split used for the main experiments, selection is performed at the image level (e.g., 26 parent images for training and 4 parent images for testing), after which all corresponding partition-derived SVEs are included automatically in the same subset.

Table 2.1 summarizes the SVE sampling strategy and the resulting dataset composition. For each martensite $v_f \in \{0.10, 0.45, 0.90\}$, we consider 30 parent microstructures over a 400 μm $\times$ 400 μm domain and extract non-overlapping SVEs of L_{SVE} and δ as explained in Section 2.2. The table also reports the average grain count per SVE, emphasizing that smaller observation windows contain fewer grains and therefore represent more localized microstructural environments. As a representative

Table 2.1: Summary of the SVE dataset (derived from Anto *et al.* Anto et al. (2024) for three martensite volume fractions: $v_f = \{0.10, 0.45, 0.90\}$, with 30 parent microstructures per v_f). Coefficient of variance (CoV) of Y is given as an example.

L_{SVE} [μm]	δ	Avg. grain count	Samples per v_f	Share [%]	CoV(Y)
400	1/1	1180	30	1.2	0.006
200	1/2	315	120	4.7	0.023
100	1/4	89	480	18.8	0.051
50	1/8	27	1920	75.3	0.113

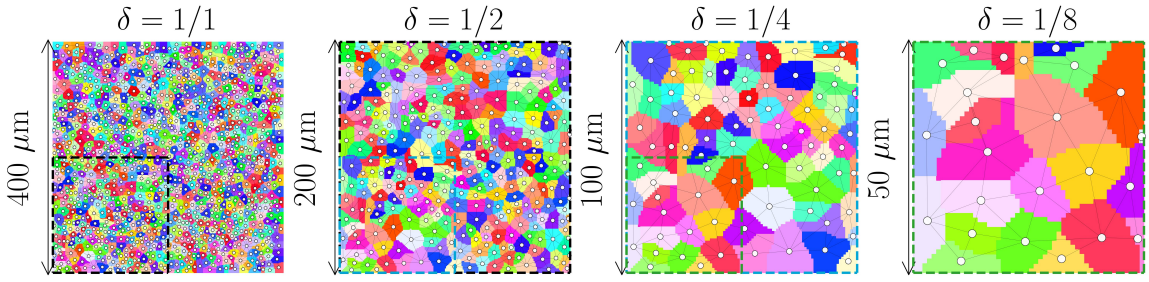


Figure 2.4: Example microstructure domains at four SVE sizes together with the corresponding grain-graph representations used as inputs to the GNN model.

measure of size-dependent scatter, the coefficient of variation (CoV) of the yield strength Y is listed; the monotonic increase in CoV with decreasing L_{SVE} reflects stronger microstructural variability at smaller length scales. Figure 2.4 illustrates an example domain and corresponding graph from each size referenced in Table 2.1.

2.4 Results

This section evaluates the predictive performance and physical consistency of the proposed multitask grain-graph surrogate. We first examine whether a single model trained on multiple martensite volume fractions can maintain accuracy across individual composition regimes (Section 2.4.1). This analysis establishes that pooling data across volume fractions enables a unified surrogate while preserving most predictive fidelity. We then quantify the accuracy of scalar elastic and plastic quantities of interest using parity plots, highlighting regions where the model achieves

Table 2.2: Effect of pooling volume fractions during training. $R^2_{\text{spec}}(v^I)$ denotes within-regime R^2 for a model trained only on regime v^I and evaluated on test SVEs from the same regime. $R^2_{\text{unif}|v^I}$ denotes within-regime R^2 for the unified model $\mathcal{M}_{\text{all}}$ evaluated on the test subset restricted to regime v^I. The pooled scores compare (i) concatenated predictions from specialized models (each test sample predicted by its regime-specific model; columns 2 to 7) and (ii) a single unified model evaluated on the concatenated test set (columns 8 and 9).

QoI	R^2_{spec}			$R^2_{\text{unif}\|v^I}$			$\cup v^I$	
	v^I=0.1	v^I=0.45	v^I=0.9	v^I=0.1	v^I=0.45	v^I=0.9	$R^2_{\text{spec}}{}^{\text{pool}}$	$R^2_{\text{unif}}{}^{\text{pool}}$
$\mathcal{C}_{11}$	0.881	0.905	0.761	0.865	0.863	0.751	0.993	0.990
$\mathcal{C}_{12}$	0.734	0.762	0.392	0.704	0.714	0.422	0.969	0.969
$\mathcal{C}_{22}$	0.894	0.910	0.768	0.869	0.880	0.738	0.993	0.991
$\mathcal{C}_{33}$	0.143	0.209	0.075	0.152	0.209	0.103	0.807	0.828
κ	0.966	0.962	0.919	0.931	0.940	0.908	0.998	0.996
μ	0.504	0.564	0.247	0.550	0.565	0.335	0.947	0.950
Y	0.792	0.853	0.586	0.771	0.812	0.611	0.987	0.985
H	0.594	0.745	0.323	0.571	0.648	0.403	0.964	0.966
Y_x	0.779	0.843	0.577	0.738	0.799	0.605	0.986	0.985
Y_y	0.734	0.848	0.561	0.748	0.812	0.591	0.986	0.984

strong agreement with CPFEM results (Section 2.4.2). Next, we evaluate vector-valued stress–strain predictions in the x and y loading directions using both curve-level metrics and population envelopes to assess fidelity in nonlinear response and variability (Section 2.4.3). Finally, we examine whether the surrogate preserves microstructural heterogeneity across SVE partitions by comparing predicted and reference distributions and spatial field patterns (Sections 2.4.4).

2.4.1 Volume-fraction generalization

To assess whether a single model can generalize across martensite volume fractions without sacrificing accuracy, we compare two training strategies: (i) v^I*-specific* models trained on a single regime, and (ii) a *unified* model trained on the pooled dataset spanning all v^I. Let $\mathcal{D}_{v^I}$ denote the subset of SVEs with volume fraction $v^I \in \{0.10, 0.45, 0.90\}$. We train specialized models $\mathcal{M}_{v^I}$ on $\mathcal{D}_{v^I}$ and a unified

model $\mathcal{M}_{\text{all}}$ on $\bigcup_{v^I} \mathcal{D}_{v^I}$. For evaluation, we report two complementary coefficients of determination: (a) *within-regime* (conditional) scores, $R^2(\mathcal{M}, \mathcal{D}^{\text{test}}_{v^I})$, computed on test SVEs restricted to a fixed v^I, and (b) *pooled* scores computed after concatenating all test SVEs across regimes. For the pooled score using specialized models, each test sample is predicted by its corresponding $\mathcal{M}_{v^I}$ (a post-hoc regime-aware reference), whereas the unified pooled score uses $\mathcal{M}_{\text{all}}$ for all samples.

Before comparing regime-specific and unified training strategies, it is useful to note that direct cross-volume-fraction transfer is generally poor when a model is trained on a single composition regime and evaluated on another. As an illustrative example, Appendix Figure C.1 shows parity plots for a model trained only on $vf = 0.10$ and tested on $v_f = 0.45$, where several scalar QoIs exhibit substantial degradation. This behavior motivates the volume-fraction generalization study below and supports the use of a unified model trained on pooled v^I data.

Table 2.2 shows that unified training does not degrade performance in a systematic way: the conditional R^2 values of $\mathcal{M}_{\text{all}}$ remain close to those of the specialized $\mathcal{M}_{v^I}$ models across most outputs, while eliminating the need to maintain separate models for each composition regime. Moreover, the pooled R^2 from the unified model is comparable to that obtained by combining specialized models, indicating that a single set of learned microstructure features can support accurate predictions across volume fractions. The main exception is the direction-dependent shear component $\mathcal{C}_{33}$, which remains challenging for both strategies; however, even for this case, the unified pooled score is competitive, consistent with the broader observation that shear-dominated response is particularly sensitive to finite-SVE anisotropy effects, which is explained in detail in the next section.

Table 2.2 should therefore be interpreted in two complementary ways. The unified model remains competitive with the regime-specific models within individual v^I regimes, but the strongest gains are observed in the pooled evaluation, where the model benefits from learning the broader variation in responses across compositions. Accordingly, the main result is not that pooled training is uniformly superior across

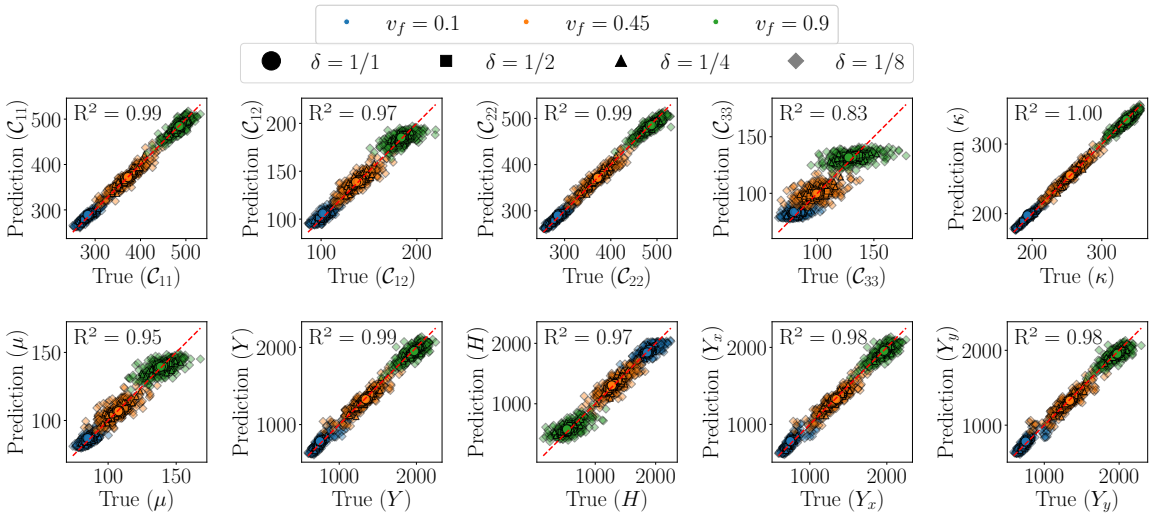


Figure 2.5: Parity plots for the linear and nonlinear scalar QoIs.

all regimes, but that a single unified surrogate preserves most within-regime fidelity while also capturing the dominant cross-regime trends. Here, the pooled R^2 primarily measures fidelity across the broader response range obtained by combining all v^I regimes, whereas the conditional R^2 is the stricter measure of within-regime sensitivity at fixed v^I.

2.4.2 Elastic and plastic QoIs

Figure 2.5 reports parity plots for the ten scalar QoIs. All stiffness and strength units in the manuscript are MPa. Overall, the surrogate achieves very high accuracy for both elastic and plastic properties, indicating that the learned SVE embedding is sufficiently expressive to resolve microstructure-induced trends beyond simple compositional averaging. The best-performing elastic targets are $\mathcal{C}_{11}$ and $\mathcal{C}_{22}$, followed by $\mathcal{C}_{12}$. The derived isotropic bulk modulus κ exhibits the highest agreement among the reported scalars with an R^2 higher than 0.995, while the isotropic shear modulus μ remains strongly correlated ($R^2 = 0.95$). These results suggest that the model captures both the principal stiffness response and derived moduli consistently within a single multitask representation.

Plastic scalars show a similarly strong performance. The yield strength Y is predicted with $R^2 = 0.99$ and the hardening descriptor H with $R^2 = 0.97$. Directional yield strengths are predicted with nearly identical accuracy in both loading directions (Y_x: $R^2 = 0.98$, Y_y: $R^2 = 0.98$); this is particularly important for maintaining directional consistency when SVEs exhibit apparent anisotropy due to finite sampling and phase topology. The parity scatter plots for these plastic quantities remain concentrated near the identity line, indicating that the surrogate preserves relative ranking across SVEs and does not systematically bias toward the mean.

Among the elastic scalars, shear modulus $\mathcal{C}_{33}$ exhibits the largest dispersion ($R^2 = 0.83$), despite the high accuracy obtained for isotropic shear modulus μ ($R^2 = 0.95$). This contrast suggests that the dominant challenge is not capturing the overall elastic stiffness level, but resolving the direction-dependent shear response that emerges at finite SVE sizes. In particular, $\mathcal{C}_{33}$ is more strongly influenced by the microstructure topology, including phase connectivity, interface networks, and local constraint pathways, than the angle-averaged response encoded in μ. At small-to-intermediate SVE sizes, these topology-driven effects can also manifest as apparent anisotropy, even when the parent microstructure is statistically isotropic, thereby increasing sample-to-sample variability and reducing the predictability of a single shear component. The relatively higher dispersion of $\mathcal{C}_{33}$ is also influenced by the MBC used in the crystal plasticity simulations. In the adopted formulation, roller-type constraints are applied on certain faces of the SVE. These constraints work well for the normal loading cases that provide $\mathcal{C}_{11}$ and $\mathcal{C}_{22}$. However, $\mathcal{C}_{33}$ is obtained from the shear loading case, which appears to be more sensitive to this boundary treatment. As a result, the shear response obtained from the SVE simulations may contain additional variability compared with the normal stiffness components.[†] This

[†]Other types of MBC Huet (1990); Hazanov and Huet (1994); Hazanov and Amieur (1995) may provide a more accurate shear response. The corresponding $\mathcal{C}_{33}$ values could correlate more strongly with the microstructure and therefore show lower dispersion in the surrogate GNN model. However, the original CP dataset used in this work did not include results obtained with such alternative boundary conditions.

interpretation is also consistent with our prior CPFEM study of the same SVE database, where the shear modulus exhibited smaller convergence rates and larger RVE sizes than the bulk modulus, indicating stronger sensitivity of shear-dominated quantities to finite-SVE variability and anisotropy Anto et al. (2024).

Finally, the evaluation is designed to measure generalization at the parent-image level. The train and test split is performed on microstructure images with stratification across v^I regimes, and all partition-derived SVEs from a given parent image are kept in the same split as explained in Section 2.3. The high parity agreement across scalars, therefore, reflects transfer to unseen microstructures rather than interpolation among correlated subwindows.

2.4.3 Stress–strain response

We next evaluate the vector-valued stress–strain targets, $\sigma_{xx}(\varepsilon)$ and $\sigma_{yy}(\varepsilon)$, each represented as a 24-dimensional output. Figure 2.6 summarizes population envelopes (median and 5–95% bands) across the considered volume-fraction regimes and loading directions. The surrogate reproduces the median curve shape across the full strain range, including the initial elastic regime, the yield transition, and subsequent hardening. This indicates that the shared latent embedding captures the coupled features needed to predict both elastic slope and post-yield evolution.

A key outcome is the agreement in response variability. The predicted 5–95% envelopes track the observed spread, suggesting that the surrogate preserves heterogeneity across SVEs rather than collapsing to a narrow mean response. This matters for downstream reliability and damage analyses, where variability and extremes (not only averages) govern initiation and uncertainty. Visually, the agreement is maintained across all volume-fraction regimes, supporting the model's ability to represent both stiffness/strength shifts and changes in scatter with microstructure.

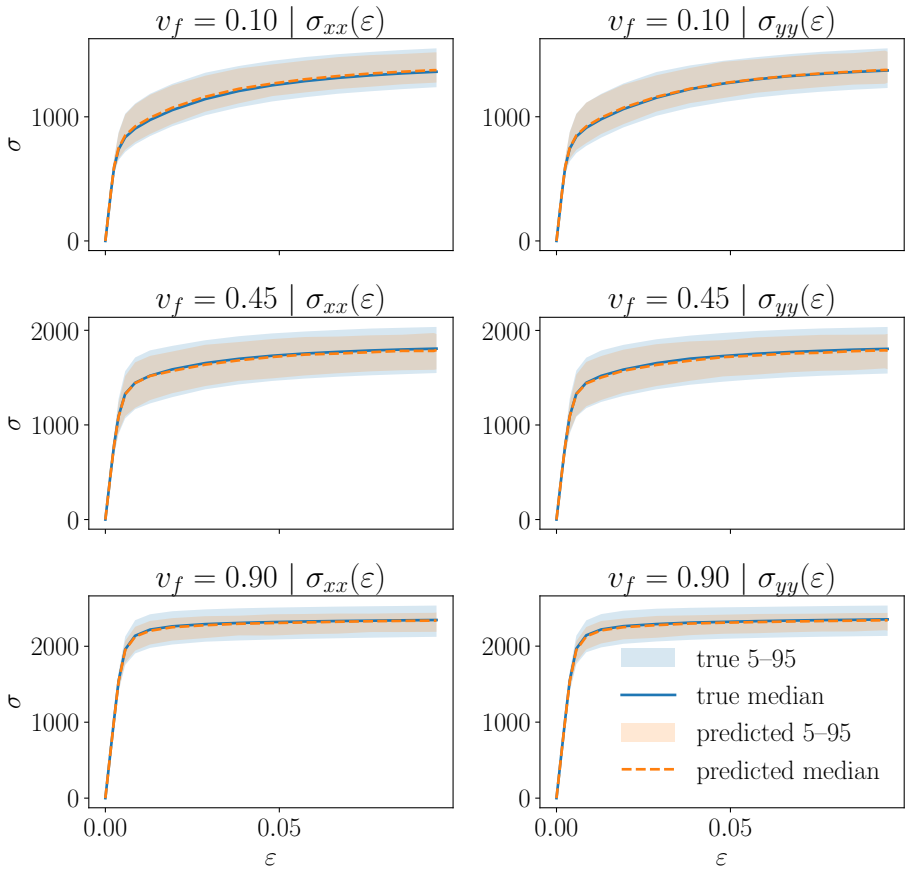


Figure 2.6: Population envelopes for predicted and reference stress–strain responses under uniaxial loading in the x and y directions. For each martensite volume-fraction regime ($v^I = 0.10$, 0.45, and 0.90), the solid curves denote the reference median response and the shaded region indicates the 5–95% interval across SVE realizations.

Table 2.3: Curve-level performance metrics for predicted stress–strain responses in the x and y loading directions. Metrics are computed over the discretized strain grid for each response and then aggregated over the evaluation set.

Output	R^2	MAE (MPa)	RMSE (MPa)	$\mathrm{NRMSE}_{\mathrm{peak}}$ (%)	L_∞ (MPa)
$\sigma_{xx}(\varepsilon)$	0.9561	154.33	194.05	7.17	307.10
$\sigma_{yy}(\varepsilon)$	0.9558	159.94	195.28	7.29	284.60

Quantitative curve metrics are summarized in Table 2.3. We report R^2 to assess the explained variance of the full discretized response, and MAE/RMSE to quantify the average and quadratic errors in stress magnitude, respectively, across the sampled strain points. To enable comparison across regimes with different stress scales, we also report normalized values relative to the peak reference stress ($\mathrm{NRMSE}_{\mathrm{peak}}$). Finally, the L_∞ error captures the worst-case pointwise deviation across the strain grid, highlighting localized mismatches near the yield surface or in regions of high curvature. Across both loading directions, these metrics indicate comparable accuracy, supporting the conclusion that the surrogate does not over-specialize to a single loading axis and can represent directional stress–strain trajectories within the same multitask model.

2.4.4 Distributions and spatial consistency

Beyond pointwise accuracy, we evaluate whether the surrogate model preserves pointwise distribution and spatial consistency induced by SVE partitioning. Figure 2.7 compares predicted and reference probability density functions (PDFs) for representative scalars (κ, μ, Y, and H) across multiple partition scales. Across regimes, the predicted PDFs closely follow the reference PDFs in both the value range and the local maxima, indicating that the model reproduces not only mean shifts but also the variability structure associated with finite SVEs. This is a nontrivial requirement for microstructure surrogates, since distributional fidelity is typically more stringent than the minimization of average error.

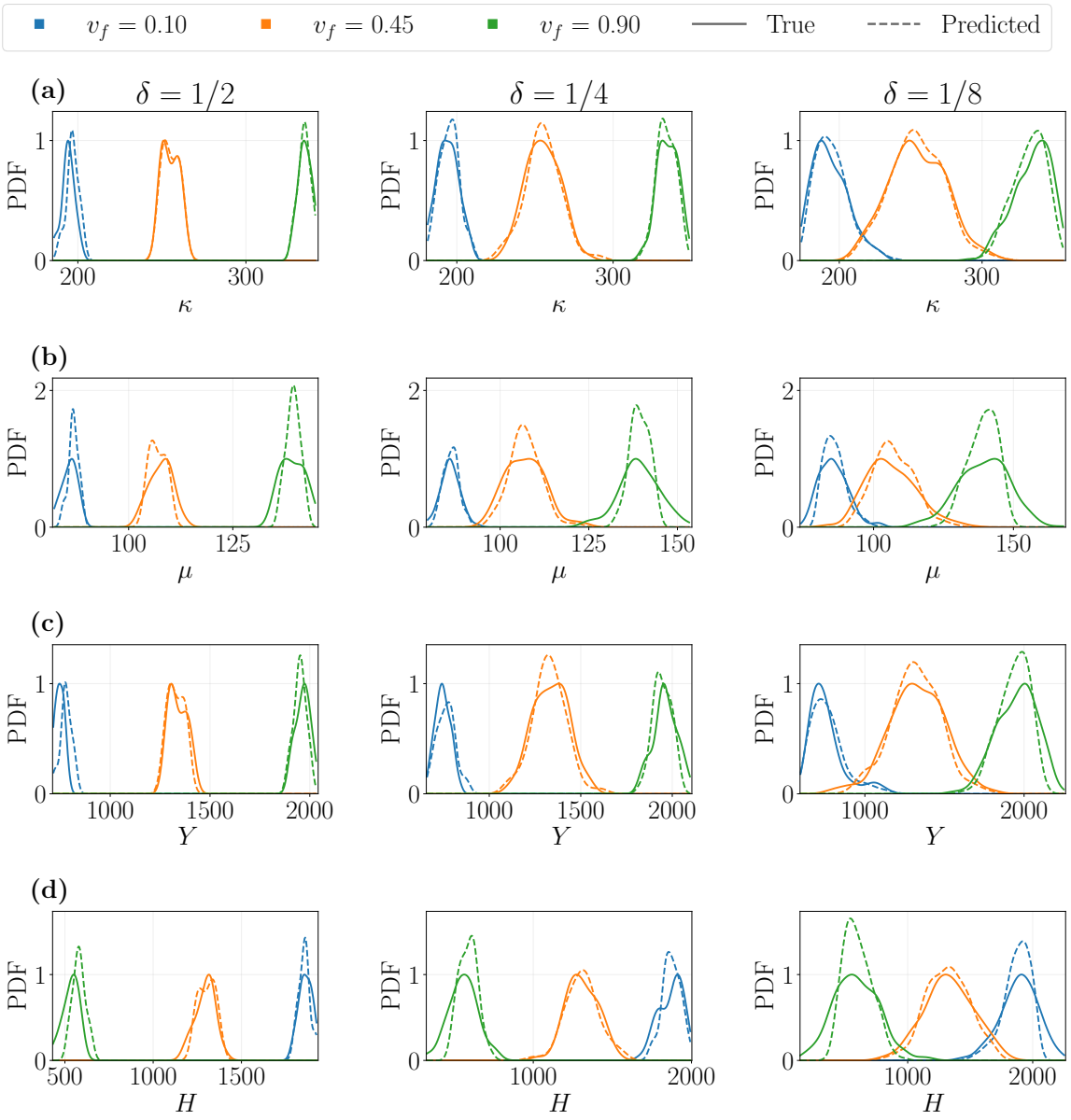


Figure 2.7: Comparison of predicted and reference probability density functions (PDFs) for representative scalar QoIs across SVE partition scales and martensite volume fractions. Solid lines denote reference PDFs and dashed lines denote surrogate-predicted PDFs for (a) κ, (b) μ, (c) Y, and (d) H at partition scales $\delta \in \{1/2, 1/4, 1/8\}$.

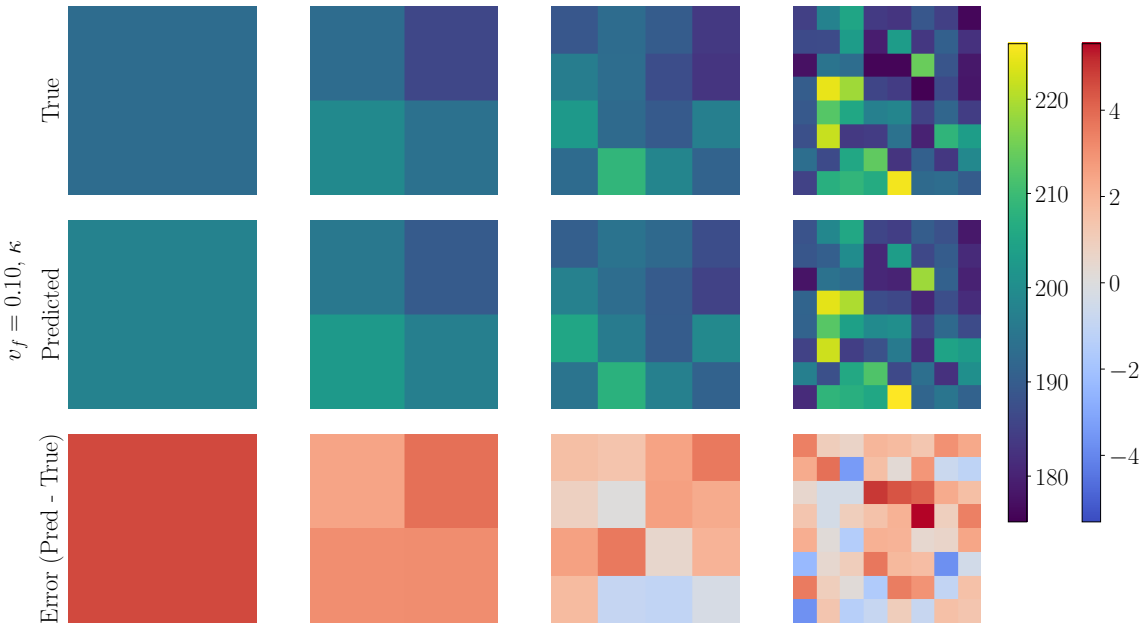


Figure 2.8: Spatial consistency of surrogate predictions across SVE partitions. For a representative case, subdomain maps are shown for the reference field, the surrogate-predicted field, and the corresponding error field for the selected QoI (κ) at progressively finer partition resolutions.

We also assess the spatial consistency of homogenized fields using subdomain maps shown in Figure 2.8. The predicted fields reproduce the dominant spatial patterns across subwindows, and the corresponding error maps remain localized rather than exhibiting global bias. This behavior suggests that the model learns microstructure-dependent heterogeneity and spatial consistency that are consistent across partition levels. This is particularly relevant when partitioned predictions are later assembled into larger domains within a multiscale method, as in Bahmani et al. (2019).

Taken together, the pointwise distribution and field-level comparisons indicate that the surrogate model accurately represents the underlying material heterogeneity, which is important for downstream damage and fatigue modeling. In such settings, localized response patterns and the tails of property distributions can dominate the risk of crack nucleation, even when global averages appear well predicted. Moreover, retaining distributional fidelity enables the construction of statistically consistent random fields for mesoscale simulations, in which one seeks to populate larger

structural domains with spatially varying elastic and fracture constitutive parameters while matching the target n-point statistics. Such random-field representations are particularly useful for uncertainty propagation, reliability analysis, and large-area fracture/fatigue assessments, where explicitly resolving microstructure everywhere is infeasible with direct numerical simulations (DNS), yet preserving the correct variability and spatial heterogeneity remains essential and feasible, with the proposed SVE-based upscaling approach.

2.5 Conclusions

This work developed and evaluated a multitask grain-graph surrogate for predicting SVE-level elastoplastic response from dual-phase polycrystal microstructures. The main conclusions are as follows:

1. **Accurate multitask prediction of coupled elastic and plastic scalars.** Parity analyses demonstrate strong performance across most scalar QoIs, including stiffness components, isotropic moduli, and plastic descriptors (yield strength and hardening). The consistent accuracy across both elastic and plastic targets indicates that a shared microstructure embedding can capture coupled trends without training independent models for each QoI.

2. **High-fidelity stress–strain trajectory prediction with directional consistency.** The surrogate predicts discretized stress–strain responses in both x and y loading directions with strong agreement in median behavior and variability envelopes. Curve-level metrics (Table 2.3) indicate that the overall error and explained variance are comparable across directions, supporting stable directional generalization within a single multitask architecture.

3. **Distributional and spatial agreement under SVE partitioning.** Predicted PDFs closely match the reference PDFs across partition scales and volume-fraction regimes. Moreover, partition-level field maps reproduce the

dominant spatial patterns and exhibit localized residual errors. These results show that the model preserves microstructure-driven heterogeneity rather than collapsing to a mean response, which is essential when downstream analyses depend on pointwise distribution and spatial consistency.

4. **Implications for damage and fatigue modeling at larger scales.** The surrogate provides rapid microstructure-conditioned constitutive predictions that can support high-throughput screening, uncertainty propagation, and statistical characterization of mesoscale property fields. In particular, the demonstrated distributional fidelity enables construction of statistically consistent random fields for larger-domain simulations, where spatially varying constitutive parameters or strength proxies are required but explicit microstructure resolution everywhere is impractical. The ability to preserve tail behavior and spatial heterogeneity is particularly relevant to crack nucleation and early damage processes driven by localized extremes.

Overall, the proposed grain-graph surrogate demonstrates that physics-informed microstructure descriptors combined with message passing and multitask learning can accurately reproduce both scalar homogenized properties and full nonlinear stress–strain trajectories, while maintaining the variability structure induced by finite SVEs. These capabilities align with the needs of modern fracture and fatigue modeling workflows that increasingly rely on AI-assisted coarse-graining and large-scale computational studies.

Author contribution

Erdem Caliskan: Writing - original draft, Visualization, Investigation, Validation, Methodology, Software, Data curation, Formal analysis. **Anik Das Anto:** Writing - review and editing, Data curation, Investigation, Validation, Methodology, Software. **Massimiliano Lupo Pasini:** Writing - review and editing, Conceptualization,

Funding acquisition, Methodology. **Stephanie TerMaath:** Writing - review and editing, Conceptualization, Project administration, Funding acquisition, Methodology. **Reza Abedi:** Writing - review and editing, Supervision, Conceptualization, Formal analysis, Project administration, Funding acquisition, Methodology.

Code availability

The open-source HydraGNN library Lupo Pasini et al. (2025), used to generate the numerical results described in this article, can be accessed at the following GitHub repository: https://github.com/ORNL/HydraGNN/tree/cp2d_paper. The scripts used in this work are located under examples/cp2d.

Data availability

The datasets generated and analyzed during this study have been deposited in Zenodo under the reserved DOI 10.5281/zenodo.19041349. The repository will be made publicly available upon acceptance/publication of this manuscript.

Ethics declaration

Not applicable.

Declaration of Competing Interest

Erdem Caliskan reports that financial support was provided by Oak Ridge Institute for Science and Education (ORISE). The other authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Massimiliano Lupo Pasini and Erdem Caliskan thank Dr. Vladimir Protopopescu for his valuable feedback in the preparation of the manuscript.

This research is sponsored by the Artificial Intelligence Initiative as part of the Laboratory Directed Research and Development (LDRD) Program of Oak Ridge National Laboratory, managed by UT-Battelle, LLC, for the US Department of Energy under contract DE-AC05-00OR22725. This work used resources of the Oak Ridge Leadership Computing Facility, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725, under ASCR Leadership Computing Challenge (ALCC) award LRN070.

This research was supported in part by the Air Force Office of Scientific Research (AFOSR) under grant number FA9550-22-1-0359. Erdem Caliskan acknowledges partial support by an appointment to the Oak Ridge National Laboratory GRO Program, sponsored by the U.S. Department of Energy and administered by the Oak Ridge Institute for Science and Education (ORISE).

The authors would also like to thank the University of Tennessee Infrastructure for Scientific Applications and Advanced Computing (ISAAC). A portion of the computation for this work was performed with ISAAC computational resources.

Chapter 3

Extreme-value-aware graph surrogates for fatigue localization in austenitic stainless steel

Author list: Erdem Caliskan, Anik Das Anto, Stephanie TerMaath, Reza Abedi, and Massimiliano Lupo Pasini

To be submitted.

Abstract

Fatigue-critical response in polycrystalline metals is governed by extreme statistics of heterogeneous cyclic-plasticity fields, where grain-scale hot spots control crack initiation. Direct crystal-plasticity finite-element (CPFEM) simulation resolves these fields, but its cost limits ensemble-scale fatigue assessment. We present a hierarchy-aware multi-task graph neural network surrogate that maps austenitic stainless steel statistical volume elements (SVEs) to per-grain and per-SVE statistics of voxel-level fatigue indicator parameter (FIP) fields. The predicted hierarchy includes within-grain FIP distribution parameters, per-grain peak FIPs, within-SVE statistics of grain-level peaks, per-SVE maximum FIPs, and volume-averaged mechanical quantities of interests. The framework is evaluated across four cases obtained by crossing Fatemi–Socie and plastic-work-based FIPs with high- and low-cycle fatigue loading over five SVE edge lengths. Beyond direct prediction of the per-SVE maximum by a graph-level head, we derive a closed-form extreme-value bridge from a mixture model for within-SVE per-grain peaks, combining an effectively elastic Dirac component at zero with a plastic-grain Gumbel component. When supplied with empirical within-SVE moments, this reference bridge achieves the lowest error among the tested per-SVE maximum estimators across all four cases and outperforms the trained graph head on three, identifying a structurally informative extreme-value map. Comparing empirical- and predicted-moment bridges separates the sources of error: the predicted-moment bridge remains accurate on the high-cycle cases, where the per-grain head preserves the plastic-grain upper tail, but degrades substantially on the low-cycle cases, where direct graph-level prediction remains the practical estimator.

3.1 Introduction

Cyclic loading drives fatigue failure in metallic alloys through rare localization events rather than through the average macroscopic response. Classical Basquin (Basquin, 1910) and Coffin–Manson (Coffin Jr, 1954; Manson, 1965) laws capture population-mean stress-life or strain-life trends, but they cannot by themselves resolve the microstructure-sensitive variability that controls the lower tail of fatigue life in engineering-critical regimes such as high-cycle fatigue (HCF) in turbine alloys and low-cycle fatigue (LCF) in landing-gear steels (Suresh, 1998; McDowell and Dunne, 2010; Przybyla and McDowell, 2010; Przybyla et al., 2013; Kuang et al., 2026). In polycrystals, crack nucleation is often governed by extreme cyclic-plastic activity at the grain and sub-grain scale. The relevant quantity is therefore not only a macroscopic average, but the upper tail of a heterogeneous local field.

Crystal-plasticity finite-element modeling (CPFEM) provides a physics-based route for resolving this localization. CPFEM captures cyclic slip, hardening, and the resulting fatigue-driving fields at the grain and sub-grain scale (Asaro, 1983; Kalidindi et al., 1992; Roters et al., 2010; Bargmann et al., 2018; Segurado et al., 2018; Ghosh et al., 2008; Knezevic et al., 2014; Zhang et al., 2016). Modern CPFEM studies can resolve hundreds to thousands of grains per representative volume element or statistical volume element (SVE) using slip-system hardening laws calibrated against macroscopic experiments and single-crystal data (Musinski and McDowell, 2012; Castelluccio and McDowell, 2016; Ashraf and Castelluccio, 2023). This resolution makes CPFEM suitable for studying microstructure-sensitive fatigue localization, but its computational cost limits ensemble-scale analyses over many microstructural realizations, SVE sizes, loading regimes, and uncertainty sources.

Fatigue indicator parameters (FIPs) are commonly used to summarize CPFEM cyclic-plasticity fields through scalar quantities associated with crack initiation. The Fatemi–Socie indicator ($\mathrm{FIP_{FS}}$) (Fatemi and Socie, 1988; Fatemi and Kurath, 1988) combines cyclic shear-strain amplitude on a slip plane with a tensile normal-stress

factor on the same plane, and is commonly used as a critical-plane proxy for stage-I slip-band crack initiation under HCF conditions (Castelluccio and McDowell, 2012; Manonukul and Dunne, 2004; Hochhalter et al., 2010; Hennessey et al., 2017; Muth et al., 2021; Ashraf et al., 2022). The plastic-work-based indicator (FIP_{WP}) (Korsunsky et al., 2007; Wan et al., 2014) accumulates the per-cycle plastic-energy density over slip systems, and acts as an energy-density proxy for damage accumulation under LCF conditions where multiple slip systems are active (Cerrone et al., 2015; Wan et al., 2014; Yuan et al., 2018; Song et al., 2022; Wang et al., 2023). These indicators are local fields, not single deterministic material constants. Their grain-scale hot spots and per-volume maxima therefore inherit the heterogeneity of the underlying microstructure.

This observation connects microstructure-sensitive fatigue prediction to extreme-value statistics. Voxel-level FIP fields define within-grain distributions; within-grain extremes define per-grain peak FIPs; the collection of per-grain peaks within an SVE defines a within-SVE distribution; and the maximum over that distribution defines the per-SVE extreme. When the underlying distribution belongs to the Gumbel maximum domain of attraction, the cross-volume maximum approaches a Gumbel form under appropriate centering and scaling (Fisher and Tippett, 1928; Gumbel, 1958; Coles et al., 2001; Embrechts et al., 2013; Leadbetter et al., 2012). This framing is consistent with recent microstructure-sensitive fatigue studies (Anto et al., 2026; Gu et al., 2020; Yaghoobi et al., 2023, 2021; Sangid, 2013; Castelluccio and McDowell, 2015; Stopka et al., 2023). At the finite grain counts encountered in engineering SVEs, however, second-order corrections, finite-size effects, and working-model assumptions can strongly influence the predicted upper tail (Embrechts et al., 2013; Leadbetter et al., 2012; Hall, 1979). A surrogate for fatigue criticality should therefore preserve the statistical hierarchy of the CPFEM field, rather than reduce the response to a single bulk scalar. Uncertainty quantification (UQ) is also required because both the grain-level hot spots and the per-SVE maximum are tail-sensitive quantities.

Materials machine learning has increasingly been used to accelerate structure-property prediction, but fatigue-localization surrogates require both tail sensitivity and uncertainty awareness. Heteroscedastic Gaussian negative log-likelihood (NLL) provides a way to learn feature-dependent predictive scales (Kendall and Gal, 2017; Nix and Weigend, 1994; Seitzer et al., 2022), while Monte Carlo dropout (MCD) approximates epistemic uncertainty through stochastic inference (Gal and Ghahramani, 2016; Kingma et al., 2015; Gal et al., 2017). These approaches have been used in materials property prediction and crystal-plasticity contexts (Tran et al., 2020; Frankel et al., 2020; Olivier et al., 2021). For fatigue-critical quantities, however, uncertainty must be assessed at the same levels as the target itself: the grain-level hot spots and the per-SVE maximum.

Graph neural networks (GNNs) provide a natural surrogate architecture for polycrystalline microstructures because grains can be represented as nodes and grain–grain contacts as edges (Hamilton et al., 2017; Gilmer et al., 2017; Battaglia et al., 2018; Bronstein et al., 2017; Xie and Grossman, 2018a; Park and Wolverton, 2020; Reiser et al., 2022; Xie and Grossman, 2018b; Sim et al., 2025; Hu et al., 2024). Message passing aggregates crystallographic, morphological, and neighborhood information into per-grain embeddings, while graph pooling produces per-SVE predictions. Prior CPFEM-surrogate studies have predicted per-grain mean FIPs, bulk-volume FIPs, or related SVE-level responses (Sim et al., 2025; Frankel et al., 2019; Yaghoobi et al., 2023; Mangal and Holm, 2018; Hansen et al., 2024). What remains less developed is a surrogate framework that predicts tail-relevant statistics across the voxel, grain, and SVE levels; tests whether an extreme-value bridge can reconstruct the per-SVE maximum from lower-level statistics; and evaluates uncertainty coverage on the same hierarchy.

In this work, we address this gap using a hierarchy-aware graph surrogate for CPFEM-derived FIP fields in austenitic stainless steel. Both $\mathrm{FIP}_{\mathrm{FS}}$ and $\mathrm{FIP}_{\mathrm{WP}}$ are evaluated under HCF and LCF loading regimes, producing four FIP cases per SVE

that share the same microstructure but differ in the post-processing of the cyclic-plasticity response. Across all four cases, the engineering target is the per-SVE maximum FIP, because fatigue life is governed by the most damaging local event. The surrogate predicts within-grain FIP distribution parameters, per-grain peak FIPs, within-SVE statistics of grain-level peaks, per-SVE maximum FIPs, and volume-averaged mechanical auxiliary quantities produced by the same CPFEM simulations.

The contributions of this work are threefold. First, we develop a multi-task grain-graph surrogate that predicts four nested FIP statistics: within-grain distribution parameters, per-grain peaks, within-SVE moments of those peaks, and per-SVE maxima jointly across the four cases obtained by crossing $\mathrm{FIP}_{\mathrm{FS}}$ and $\mathrm{FIP}_{\mathrm{WP}}$ with HCF and LCF loading. Six volume-averaged mechanical auxiliary quantities are co-trained on the same encoder, and held-out prediction-interval coverage is reported at both the grain and SVE levels. Second, we derive a closed-form Gumbel-mean bridge from a two-component mixture working model on the within-SVE distribution of per-grain peaks: an effectively elastic component at zero combined with a plastic-grain Gumbel component. The plastic-grain fraction takes alloy- and regime-specific values on this dataset and is treated as a measured property of the cyclic-plasticity response rather than a free hyperparameter. Third, we use the comparison between the trained graph head, the bridge with empirical within-SVE moments, and the bridge with predicted moments as a diagnostic that separates the structural extreme-value map from the deployment-side quality of per-grain prediction. To our knowledge, this two-route comparison has no prior counterpart in microstructure-sensitive fatigue surrogate work.

Section 3.2.1 describes the CPFEM dataset, the FIP definitions, the multilevel statistics, and the Gumbel-mean bridge. Section 3.3 presents the graph construction, input features, multi-task architecture, training objective, and interpretation protocol. Section 3.4 reports the four-case results, including the direct-versus-bridge comparison, prediction-interval coverage, and mechanical auxiliary predictions. Section 3.5 summarizes the implications and limitations.

3.2 Dataset and CPFEM modelling

3.2.1 Crystal-plasticity simulation and fatigue indicator parameters

The dataset originates from CPFEM simulations on austenitic stainless steel SVEs reported by (Anto et al., 2026). This subsection summarizes the physical model and the two voxel-level FIPs used as surrogate targets. The ensemble-level dataset statistics, including SVE sizes, realization counts, and train/validation/test partitions, are given in Section 3.2.2.

The polycrystals are single-phase, face-centered-cubic, with random crystallographic orientations and equiaxed equivalent-spherical grain diameters drawn from a truncated lognormal distribution centered at $50\,\mu$m, consistent with observations for solution-annealed stainless steel (Anto et al., 2026; Asaro, 1983). Each SVE is a periodic cube voxelated in DREAM.3D (Groeber and Jackson, 2014). The deformation gradient is split multiplicatively as $\mathbf{F} = \mathbf{F}^e\mathbf{F}^p$, with plastic velocity gradient

$$\mathbf{L}^p = \dot{\mathbf{F}}^p\mathbf{F}^{p,-1} = \sum_{\alpha=1}^{N_{\text{slip}}} \dot{\gamma}^\alpha\, \mathbf{s}^\alpha \otimes \mathbf{m}^\alpha, \tag{3.1}$$

where the sum is restricted to the twelve octahedral slip systems of the face-centered-cubic lattice, with slip directions $\mathbf{s}^\alpha$ and slip-plane normals $\mathbf{m}^\alpha$. Slip evolves according to a viscoplastic flow rule with backstress offset and saturation, and the constitutive response combines isotropic and kinematic hardening at the slip-system level. The parameter set follows the solution-annealed steel calibration of (Anto et al., 2026) and the crystal-plasticity formulation of (Kalidindi et al., 1992; Roters et al., 2010). The applied loading is fully reversed strain control with macroscopic strain amplitude ε_{amp}, and FIPs are evaluated on the final stabilized cycle, typically the third cycle.

Two voxel-level indicators are evaluated, corresponding to complementary fatigue-localization mechanisms. The plastic-work-based indicator $\text{FIP}_{\text{WP}}(\mathbf{x})$ is the per-cycle

plastic energy density at voxel $\mathbf{x}$ (Korsunsky et al., 2007; Manonukul and Dunne, 2004),

$$\mathrm{FIP}_{\mathrm{WP}}(\mathbf{x}) = \sum_{\alpha=1}^{N_{\mathrm{slip}}} \int_{\mathrm{cycle}} \tau^{\alpha}(\mathbf{x})\, \dot{\gamma}^{\alpha}(\mathbf{x})\, dt, \tag{3.2}$$

where τ^{α} is the resolved shear stress and $\dot{\gamma}^{\alpha}$ is the slip rate on slip system α. This quantity acts as an energy-density proxy for damage accumulation, particularly when multiple slip systems contribute to cyclic plastic dissipation.

The crystallographic Fatemi–Socie indicator $\mathrm{FIP}_{\mathrm{FS}}(\mathbf{x})$ combines the cyclic shear-strain amplitude on a slip plane with the tensile normal stress acting on that plane (Fatemi and Socie, 1988; Castelluccio and McDowell, 2012),

$$\mathrm{FIP}_{\mathrm{FS}}(\mathbf{x}) = \max_{\alpha} \frac{\Delta\gamma^{\alpha}(\mathbf{x})}{2} \left(1 + k \frac{\sigma_n^{\alpha}(\mathbf{x})}{Y}\right), \tag{3.3}$$

where $\Delta\gamma^{\alpha}$ is the peak-to-peak cyclic shear-strain range on slip system α, σ_n^{α} is the peak tensile normal stress on the same slip plane, k is the Fatemi–Socie material parameter, and Y is the macroscopic yield stress. The maximum over slip systems selects the most damaging crystallographic plane at each voxel, making $\mathrm{FIP}_{\mathrm{FS}}$ a critical-plane proxy for stage-I slip-band crack initiation.

The two FIP definitions are evaluated under two cyclic-loading regimes: an HCF regime at $\varepsilon_{\mathrm{amp}} = 0.001 \approx 0.5\,\varepsilon_y$, which is predominantly elastic with sparse plastic activity, and an LCF regime at $\varepsilon_{\mathrm{amp}} = 0.0055 \approx 2.75\,\varepsilon_y$, in which multiple slip systems are active and stabilized plastic hysteresis develops. Crossing the two FIP definitions with the two loading regimes produces four cases: HCF/FS, HCF/WP, LCF/FS, and LCF/WP. Each case defines its own voxel-level FIP field f_v, with magnitudes, size dependence, and learnability analyzed in parallel throughout the paper.

3.2.2 SVE ensemble and four-case dataset

The ensemble samples microstructure variability and SVE size effects simultaneously, so that the surrogate is exposed to the statistical heterogeneity that drives engineering

scatter. Periodic cubic SVEs are voxelated at a fixed voxel edge length of 8.5 μm, yielding approximately 105 Hex8 finite elements per grain on average (Anto et al., 2026). Five SVE edge lengths are sampled, $L_{\mathrm{SVE}} \in \{125, 177, 250, 354, 500\}\,\mu$m, spanning mean grain counts from order 30 to nearly 1900 across the size range (Table 3.1). Figure 3.1 shows one representative SVE from each size bucket, with the grain count for that specific realization annotated. Between approximately 140 and 200 independent microstructure realizations are generated per size bucket, capturing variability from the random grain morphology and crystallographic orientation distributions.

Table 3.1: Four-case SVE dataset partitioned by edge length L_{SVE} and SVE index. The same realizations carry all four FIP cases, so the partition counts apply uniformly to (FS, WP) $\times$ (HCF, LCF). Per-bucket grain counts vary across realizations and are visualized in Figure 3.1.

L_{SVE} (μm)	train	validation	test	total
125	180	10	10	200
177	179	10	10	199
250	180	10	10	200
354	180	10	10	200
500	127	7	7	141
total	846	47	47	940

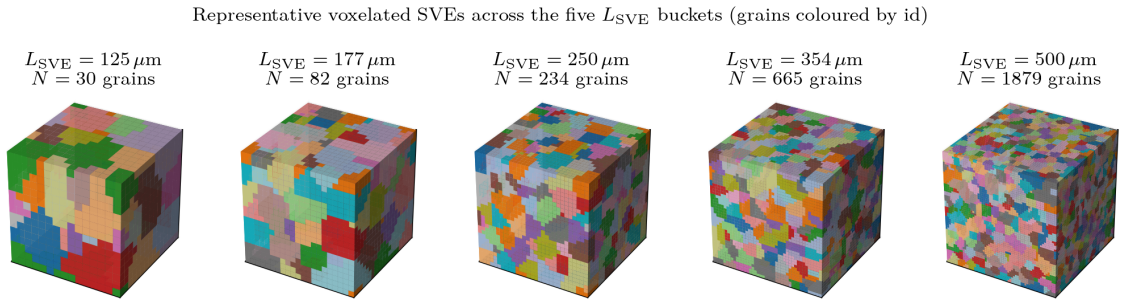


Figure 3.1: Voxelated SVE microstructures at the five edge lengths $L_{\mathrm{SVE}} \in \{125, 177, 250, 354, 500\}\,\mu$m for one representative realization per bucket. Grains are colored by integer index for visual contrast. The annotated grain counts ($N = 30, 82, 234, 665, 1879$) correspond to these specific realizations and lie close to the ensemble means.

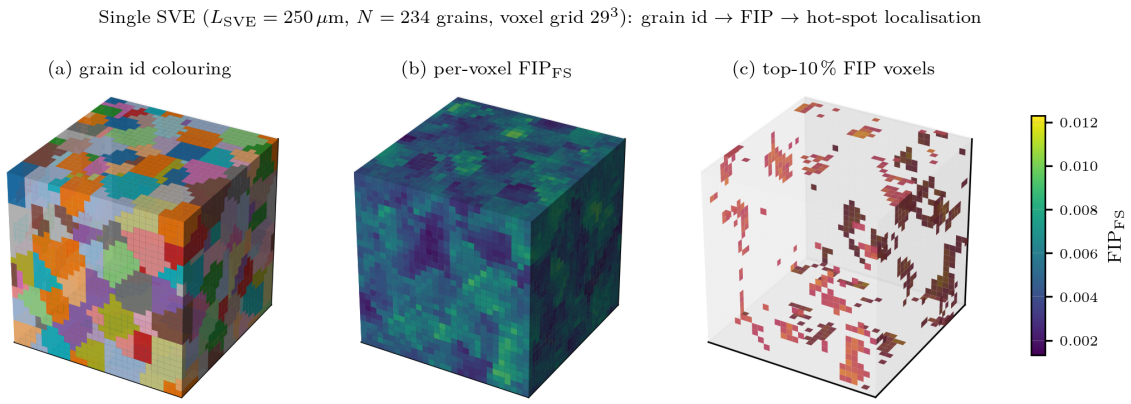


Figure 3.2: One SVE from the $L_{SVE} = 250\,\mu$m bucket ($N = 234$ grains, 29^3 voxels): (a) grain-id coloring, (b) per-voxel FIP_{FS} field on the same volume (LCF), and (c) the top-10% voxels of FIP_{FS} that identify hot-spot localization (LCF). The within-grain field is heterogeneous, and the highest values are concentrated in a small subset of grains.

Figure 3.2 maps a single SVE from the $L_{SVE} = 250\,\mu$m bucket through the three views that motivate the statistical hierarchy of Sections 3.2.3–3.2.5: panel (a) shows the grain-id coloring, panel (b) shows the corresponding per-voxel FIP_{FS} field, and panel (c) highlights the top-10% voxels by FIP_{FS} that identify hot-spot localization. These views illustrate why the per-volume extreme cannot be represented by a simple volume average: the highest FIP values occupy a small subset of voxels within a small subset of grains. The SVE ensemble is generated with random crystallographic texture, so no preferred orientation is imposed.

The same microstructure realizations are evaluated under both FIP definitions (FS, WP) and both loading regimes (HCF, LCF), producing four FIP cases per SVE that share the underlying microstructure and differ only in the post-processing of the cyclic-plasticity response. The ensemble is partitioned at the SVE level into train, validation, and test subsets that are disjoint by SVE index, and the same partition is used for all four cases. This prevents the same microstructure from appearing in different splits through another FIP case, while preserving the case-symmetric structure used by the multi-task surrogate. Counts per size bucket are summarized in Table 3.1. The per-grain peak distribution across the four cases, together with

the mixture working model, is shown in Figure 3.3 of Section 3.2.5. The dataset corresponds to one alloy and one voxel resolution; cross-alloy and cross-resolution transfer is not evaluated here. The next subsection moves from this ensemble-level description to the within-grain statistics that summarize the voxel-level FIP fields.

3.2.3 Within-grain voxel-level distributions

The first level of the statistical hierarchy is the population of voxel-level FIP values $\{f_v\}_{v\in g}$ inside each grain. Voxels with zero FIP correspond to elastic regions and carry no logarithmic information, so they are removed before fitting. On the remaining within-grain population, we test four candidate parametric shapes (lognormal, normal, Gumbel, and Weibull) using the Kolmogorov–Smirnov (KS) statistic. Across the four FIP cases and five SVE size buckets, the within-grain population is well represented by a lognormal working model,

$$\ln f_v \,\big|\, g \;\sim\; \mathcal{N}\big(\mu_g,\, \sigma_g^2\big)\,, \qquad v = 1, \ldots, N_g, \tag{3.4}$$

where μ_g and σ_g are the within-grain log-location and log-scale. These two quantities are predicted directly by the surrogate's per-grain head (Section 3.3.3).

Median KS p-values per case-by-bucket combination are reported in Table 3.2, computed on a stratified sample of approximately 20 grains per combination. The median plug-in KS p-value for the lognormal fit exceeds 0.05 in all 20 case-by-bucket combinations, and the lognormal is the largest-p shape in 11 of 20 combinations, with the alternatives within 0.10 in the remaining nine. Because the distribution parameters are estimated from the same samples used in the KS test, these plug-in p-values are approximate rather than distribution-free. We therefore also perform a parametric bootstrap with $M = 200$ resamples per grain, drawing from the fitted lognormal model on $\ln f_v$ and re-fitting the parameters on each resample. The bootstrap median p-value remains above 0.05 in all 20 case-by-bucket combinations, supporting the use of the lognormal as a working shape model for the within-grain

FIP population. This evidence should be read as support for a compact empirical description of the within-grain bulk, not as proof of an exact distributional law for the upper tail.

Table 3.2: Within-grain shape goodness-of-fit. Median Kolmogorov–Smirnov p-value across the within-grain populations of approximately 20 grains per case-by-bucket combination, by FIP case and SVE size bucket. The largest-p shape per row is bolded. The median bootstrap p-value for the lognormal working model remains above $\alpha = 0.05$ in all 20 case-by-bucket combinations.

case	bucket (L_{SVE}, $\bar{N}_S$)	lognormal	normal	Gumbel	Weibull
HCF/FS	b0 (125 μm, ∼32)	**0.54**	**0.54**	0.45	0.22
HCF/FS	b1 (177 μm, ∼88)	0.55	0.38	**0.57**	0.31
HCF/FS	b2 (250 μm, ∼232)	**0.70**	0.59	0.47	0.27
HCF/FS	b3 (354 μm, ∼656)	0.44	**0.56**	0.38	0.30
HCF/FS	b4 (500 μm, ∼1858)	**0.58**	0.51	0.35	0.44
HCF/WP	b0	**0.49**	0.24	0.39	0.15
HCF/WP	b1	**0.87**	0.62	0.53	0.28
HCF/WP	b2	**0.54**	0.48	0.39	0.42
HCF/WP	b3	0.66	0.20	**0.75**	0.24
HCF/WP	b4	**0.57**	0.34	**0.57**	0.16
LCF/FS	b0	0.66	**0.77**	0.39	0.41
LCF/FS	b1	**0.60**	0.49	0.53	0.21
LCF/FS	b2	0.48	**0.74**	0.39	0.42
LCF/FS	b3	0.51	**0.67**	0.39	0.49
LCF/FS	b4	**0.60**	0.50	0.41	0.24
LCF/WP	b0	**0.75**	0.50	0.55	0.36
LCF/WP	b1	0.54	**0.62**	0.47	0.46
LCF/WP	b2	**0.59**	0.42	0.53	0.25
LCF/WP	b3	**0.71**	0.55	0.53	0.30
LCF/WP	b4	0.72	**0.78**	0.43	0.40

The fitted log-scale σ_g varies across grains by more than an order of magnitude even at fixed L_{SVE}, and is therefore retained as a per-grain target rather than collapsed into a single representative shape parameter. Two consequences frame the rest of the statistical hierarchy. First, a scalar summary of the within-grain field, such as the arithmetic voxel mean $\bar{f}_g$, discards the distributional information encoded in σ_g; the per-grain peak $m_g = \max_{v \in g} f_v$ is therefore not recoverable from the mean alone.

Second, the lognormal working model on f_v provides a compact description of the voxel-to-grain level of the hierarchy, which the next subsection connects to the within-SVE distribution of per-grain peaks.

3.2.4 Within-SVE per-grain maximum distributions

The next level of the statistical hierarchy is the population of per-grain peak FIPs within a single SVE. Define

$$m_g \;:=\; \max_{1\le v\le N_g} f_v, \qquad M_S \;:=\; \max_{g\in S} m_g, \tag{3.5}$$

where m_g is the per-grain peak FIP and M_S is the per-SVE maximum, the engineering-relevant extreme. Under cyclic loading at amplitudes near the macroscopic yield, only a subset of the grains in a given SVE accumulate appreciable plastic activity; the remainder contribute peak FIPs near zero. Conditional on a single SVE S with grain count N_S, we therefore adopt a two-component mixture working model on the within-SVE distribution of per-grain peaks,

$$m_g \,\big|\, S \;\sim\; q(S)\,\delta(0) \;+\; \big(1-q(S)\big)\,\mathrm{Gumbel}\big(\xi_P(S),\,\beta_P(S)\big), \qquad g\in S, \tag{3.6}$$

where $q(S)$ is the elastic-grain fraction and $1-q(S)$ the plastic-grain fraction in the SVE. The plastic-component $\mathrm{Gumbel}(\xi_P,\beta_P)$ has location $\xi_P(S)$ and scale $\beta_P(S)$. This choice is consistent with the within-grain working model of Section 3.2.3: since m_g is the maximum of N_g lognormal voxel FIPs and the lognormal parent lies in the Gumbel maximum-domain of attraction (Embrechts et al., 2013), the per-grain peak distribution for plastically-active grains is approximately Gumbel at the voxel counts N_g of the present dataset. Effectively elastic grains contribute m_g near zero and do not affect M_S.

The plastic-grain fraction $1-q(S)$ is identified per SVE from the empirical $\{m_g\}_{g\in S}$ distribution as $1-q(S)=\Pr\big(m_g>\mu_M(S)+\sigma_M(S)\mid S\big)$, where $\mu_M(S)$ and $\sigma_M(S)$ are

the within-SVE mean and standard deviation of the full $\{m_g\}_{g\in S}$ sample. The Gumbel parameters of the plastic component are then estimated by method of moments on the plastic subset $\mathcal{P}(S) = \{m_g : m_g > \mu_M(S) + \sigma_M(S)\}$,

$$\beta_P(S) \;=\; s_P(S)\,\frac{\sqrt{6}}{\pi}, \qquad \xi_P(S) \;=\; \bar{m}_P(S) - \gamma_E\,\beta_P(S), \tag{3.7}$$

where $\bar{m}_P(S)$ and $s_P(S)$ are the sample mean and standard deviation of $\mathcal{P}(S)$, $\xi_P(S)$ is the Gumbel location parameter, and γ_E is the Euler-Mascheroni constant. The full within-SVE moments $(\mu_M(S), \sigma_M(S))$ enter through the plastic-subset threshold; the bridge formula in Section 3.2.5 consumes the plastic-component parameters $(\xi_P(S), \beta_P(S))$ together with the grain count N_S.

Empirically, the plastic-grain fraction is approximately invariant across the five SVE size buckets within each FIP case and varies between cases. Across the four FIP cases, the median plastic-grain fraction across all SVEs is 0.21 on HCF/FS, 0.19 on HCF/WP, 0.16 on LCF/FS, and 0.15 on LCF/WP, with case-to-bucket variation under 0.04 per case. Row (a) of Figure 3.3 shows the pooled within-SVE density of m_g for each case together with the fitted mixture, decomposed into the elastic Dirac component near zero and the plastic-grain Gumbel component.

The per-graph head of Section 3.3.3 predicts the within-SVE mean $\mu_M(S)$ of the full $\{m_g\}_{g\in S}$ population directly. The plastic-grain Gumbel parameters $(\xi_P(S), \beta_P(S))$ and the plastic-grain fraction $1 - q(S)$ are recovered post-hoc from the per-grain head's predicted peaks $\{\widehat{m}_g\}_{g\in S}$ for the deployment bridge, or from the empirical $\{m_g\}_{g\in S}$ for the empirical-moment reference. The two bridge variants compared in Section 3.4.2 differ only in this choice of input sample.

As a further consistency check on Equation (3.4), the standardized within-grain peak location $z_g := (\ln m_g - \mu_g)/\sigma_g$ has empirical mean in the narrow range $\bar{z}_g = 2.37$–2.49 across the four FIP cases and five size buckets, with standard deviation approximately 0.45 and no systematic dependence on L_{SVE} between $N_S \approx 30$ and

$N_S \approx 1900$. This within-grain size-invariance is the voxel-level counterpart of the bucket-invariance of the plastic-grain fraction reported above.

3.2.5 Cross-SVE extremes and the EVT bridge

The engineering-relevant target is M_S, the per-SVE maximum FIP defined in Equation (3.5). Across an ensemble of SVEs at fixed L_{SVE}, M_S is itself a random variable, and earlier work on microstructure-sensitive fatigue (Anto et al., 2026; Gu et al., 2020; Yaghoobi et al., 2023) reports an approximately Gumbel cross-SVE distribution at sizes beyond the smallest size buckets. We observe the same empirical pattern across the four FIP cases and five size buckets (20 case-by-bucket combinations on 141 to 200 SVEs per combination, per Table 3.1): the Gumbel hypothesis is not rejected by Kolmogorov–Smirnov at $\alpha = 0.05$ in 19 of 20 combinations and is selected by the Akaike information criterion (AIC) in 14 of 20, with non-Gumbel-best combinations concentrated at the smallest size bucket where finite-size effects are strongest. Because distribution parameters are estimated from the same samples used for the test, the plug-in p-values are approximate rather than distribution-free; parametric-bootstrap p-values would provide a more formal calibration of the goodness-of-fit statistic. We therefore treat these results as empirical support for the Gumbel form rather than as a formal distributional decision. Row (b) of Figure 3.3 visualizes the cross-SVE density of M_S across the four FIP cases at $L_{\mathrm{SVE}} = 250\,\mu$m together with the maximum-likelihood Gumbel fit per case.

The empirical cross-SVE Gumbel form is consistent with the within-SVE mixture working model of Section 3.2.4. Within an SVE, only the plastic-grain subpopulation contributes to the maximum, since the elastic-grain Dirac component places m_g near zero. Let $\mathcal{P}(S)$ denote the plastic subset of an SVE and $N_{\mathcal{P}}(S) := |\mathcal{P}(S)|$ its integer cardinality. Treating the plastic-grain peaks as iid draws from Gumbel$(\xi_P(S), \beta_P(S))$ and conditioning on $N_{\mathcal{P}}(S)$, the max-stability of the Gumbel family, a consequence of the Fisher–Tippett–Gnedenko theorem (Fisher and Tippett, 1928; Coles et al., 2001;

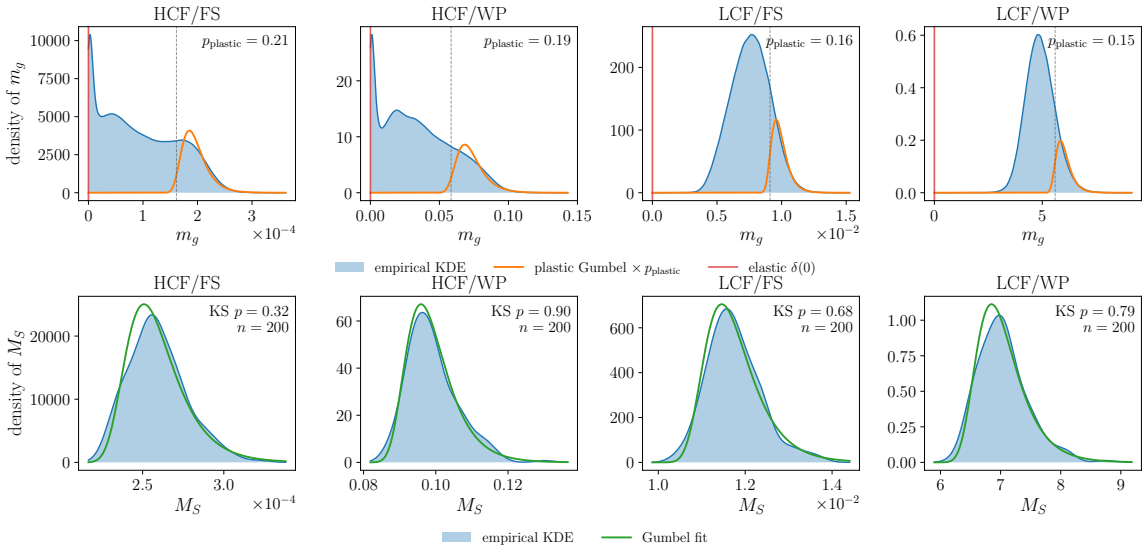


Figure 3.3: Two-level statistical cascade across the four FIP cases. Row (a): pooled per-grain peak density of m_g (filled, empirical KDE) with the fitted Dirac-plus-Gumbel mixture overlaid as the plastic-grain Gumbel component scaled by the plastic-grain fraction (orange) and the elastic Dirac component at $m_g = 0$ (red vertical); the threshold $\mu_M(S)+\sigma_M(S)$ is marked (grey dashed). The plastic-grain fraction $1-q(S)$ is annotated per panel (values rounded to two decimals from the per-SVE medians reported in Section 3.2.4). Row (b): cross-SVE density of M_S at $L_{\text{SVE}} = 250\,\mu$m ($n = 200$ SVEs per case) with the maximum-likelihood Gumbel fit (green) and the Kolmogorov–Smirnov p-value annotated. The two rows visualize the two working-model statements in Equations (3.6) and (3.8).

Embrechts et al., 2013), gives

$$M_S \,\big|\, S,\, N_{\mathcal{P}}(S) \;\sim\; \text{Gumbel}\Big(\xi_P(S) + \beta_P(S)\ln N_{\mathcal{P}}(S),\; \beta_P(S)\Big). \tag{3.8}$$

The bridge of this work uses the effective-count approximation $N_{\mathcal{P}}(S) \approx N_S(1-q(S))$, replacing the integer plastic cardinality by its expected value under the mixture. The corresponding analytical mean of M_S is

$$\mathbb{E}\big[M_S \,\big|\, \xi_P(S), \beta_P(S), N_S, q(S)\big] \;\approx\; \xi_P(S) \;+\; \beta_P(S)\big(\ln\big(N_S(1-q(S))\big) + \gamma_E\big)\,, \tag{3.9}$$

where γ_E is the Euler-Mascheroni constant. Equation (3.9) is exact under the iid approximation on $\mathcal{P}(S)$ when the integer cardinality $N_{\mathcal{P}}(S)$ is used in place of the effective count $N_S(1-q(S))$; the effective-count form used in this work carries the additional approximation $\ln N_{\mathcal{P}}(S) \approx \ln N_S(1-q(S))$, which is accurate to within a few percent at the plastic-grain fractions of the present dataset. The leading-order growth in N_S is linear in $\ln\big(N_S(1-q(S))\big)$, consistent with the linear-in-$\ln N_S$ length-scale scaling reported in earlier work (Anto et al., 2026). Two degenerate cases are worth noting. When $q(S) \to 0$, all grains are plastically active, the elastic Dirac component vanishes, and Equation (3.9) reduces to the pure-Gumbel form with $N_S(1-q) = N_S$. An alternative working model that replaces the within-SVE Gumbel by a Gaussian on $\{m_g\}_{g\in S}$ yields the Gaussian-parent extreme-value formula (Embrechts et al., 2013; Hall, 1979), in which the leading-order extreme-location increment scales with $\sigma_M\sqrt{2\ln N_S}$ together with an additive Gaussian-maximum correction. Neither case applies to the present dataset, where the plastic-grain fraction $1-q(S)$ lies between 0.15 and 0.21 across the four cases and where the within-SVE mixture is not well approximated by a single Gaussian. The mixture form is therefore adopted as the operative bridge in the remainder of this work.

Equations (3.6)–(3.9) define the closed-form bridge used in this work: given the within-SVE plastic-component Gumbel parameters $(\xi_P(S), \beta_P(S))$, the plastic-grain

fraction $1-q(S)$, and the grain count N_S, the bridge returns a Gumbel-mean estimate of M_S. It is the analytical counterpart of the dedicated graph-level maximum head $\widehat{M}_S$ examined in Section 3.4.2. The bridge assumes an effective sample of $N_S(1-q(S))$ independent draws from the plastic-grain Gumbel parent. Spatial correlation among neighboring grains and SVE-level mechanical constraints can reduce the effective sample size below this nominal value. The empirical sufficiency of the working model is tested directly in Section 3.4.2.

3.2.6 Cross-criterion and cross-regime correlation structure

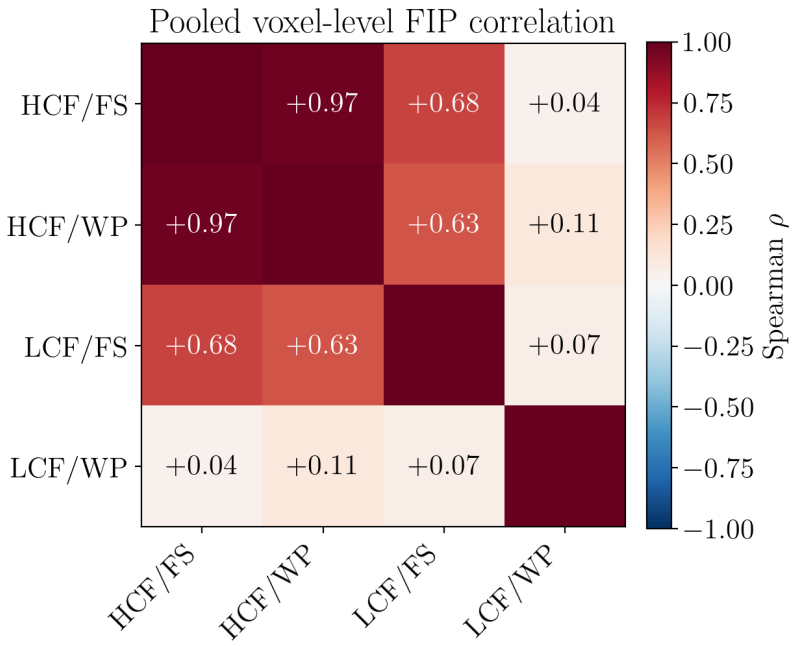


Figure 3.4: Pooled voxel-level Spearman rank correlation between the four FIP cases, computed across all voxels in the dataset. The FS and WP indicators are nearly indistinguishable in the HCF regime ($\rho_S = 0.97$) but only weakly correlated in the LCF regime ($\rho_S = 0.07$); the FS criterion is partially transferable across regimes ($\rho_S = 0.68$ between HCF/FS and LCF/FS); and LCF/WP correlates only weakly with the other three cases.

The four FIP cases (HCF/FS, HCF/WP, LCF/FS, LCF/WP) share the same microstructure ensemble and differ only in the post-processing rule that maps the

resolved cyclic-plasticity fields to voxel-level FIP values. The cross-case correlation of the voxel-level fields is therefore a data property, not a surrogate property, and it provides useful context for the symmetric multi-task structure used in the results section. Figure 3.4 reports the Spearman rank correlation of the voxel-level FIPs across the four cases, pooled over all voxels in the dataset. Three structural features are visible.

First, the FS and WP indicators rank voxels almost identically in the HCF regime ($\rho_S = 0.97$) but are only weakly correlated in the LCF regime ($\rho_S = 0.07$). At low macroscopic strain, plastic activity is sparse and the slip-system shear-strain amplitude that drives $\mathrm{FIP}_{\mathrm{FS}}$ is closely aligned with the plastic-work accumulation that drives $\mathrm{FIP}_{\mathrm{WP}}$, so the two indicators induce nearly the same voxel-level ordering. At high macroscopic strain, multiple slip systems activate within a voxel, and the maximum over slip systems used in $\mathrm{FIP}_{\mathrm{FS}}$ decouples from the slip-summed plastic energy entering $\mathrm{FIP}_{\mathrm{WP}}$. Second, the FS criterion is partially transferable across loading regimes: $\mathrm{FIP}_{\mathrm{FS}}$ ranks voxels with $\rho_S \approx 0.68$ between HCF and LCF, indicating that the critical-plane proxy retains some common ordering of slip localization across strain amplitudes. Third, the LCF/WP case correlates only weakly with the other three cases (off-diagonal $\rho_S \leq 0.11$), so the voxel-level information shared by the other cases is largely absent in LCF/WP.

We use this correlation structure as an interpretive constraint on the surrogate results rather than as a model input. In particular, the weak cross-case correlation of LCF/WP supports the view that the present grain-level descriptors do not resolve the dominant localization structure of that case; this point is revisited in Section 3.4.1. The reported correlations are pooled and voxel-level. Per-grain peak and per-SVE maximum correlations inherit the same qualitative structure but are more sensitive to finite-sample effects in the upper tail.

3.3 Surrogate architecture and training

3.3.1 Graph construction

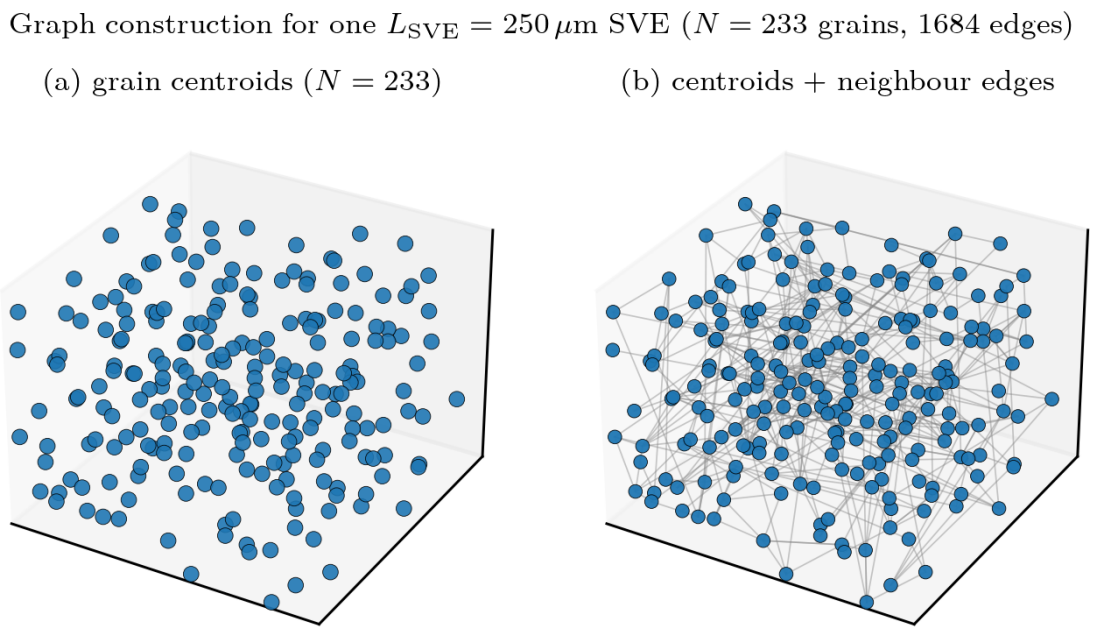


Figure 3.5: Graph construction for one $L_{\mathrm{SVE}} = 250\,\mu$m SVE ($N_S = 233$ grains, 1684 undirected edges): (a) grain centroids treated as graph nodes, and (b) the same centroids with neighbor edges drawn between voxel-adjacent grains, including periodic-face neighbors.

Each SVE is converted to an undirected grain-adjacency graph $G = (\mathcal{V}, \mathcal{E})$, where each node $g \in \mathcal{V}$ represents one grain and carries the input feature vector $\mathbf{x}_g$ described in Section 3.3.2. An edge $(g, g') \in \mathcal{E}$ is placed between two grains when at least one pair of voxels across the $\langle 100 \rangle$ voxel grid shares a grain–grain interface. Periodic boundary conditions on the SVE are inherited by the graph, so grains touching opposite faces of the periodic cube are connected when they are periodic neighbors. The GraphSAGE encoder of Section 3.3.4 consumes the graph adjacency and node features only; edge attributes are not used in the selected configuration. Figure 3.5 illustrates the construction for one $L_{\mathrm{SVE}} = 250\,\mu$m SVE.

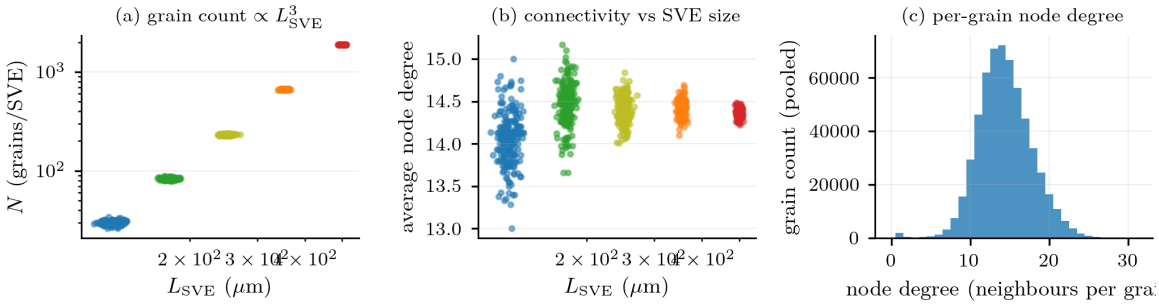


Figure 3.6: Graph-level statistics across the dataset: (a) grain count N_S per SVE versus edge length L_{SVE}, exhibiting the expected cubic scaling; (b) per-SVE average node degree, approximately size-invariant near $\bar{k} \approx 14$; and (c) pooled per-grain node-degree histogram, with a mode near 14 neighbors and a long tail to about 30.

Two structural properties of the graph ensemble are important for generalization across SVE sizes. First, the grain count N_S scales approximately as L^3_{SVE} (Figure 3.6a), so the receptive field of a fixed-depth message-passing encoder spans a decreasing fraction of the full graph as SVE size increases. Second, the local connectivity is approximately size-invariant: the per-SVE mean node degree clusters near $\bar{k} \approx 14.4$ across the five size buckets (Figure 3.6b), and the pooled per-grain degree distribution spans roughly 5 to 30 neighbors with a mode near 14 (Figure 3.6c). The empirical graph diameter increases from a median of 2 hops on the smallest SVEs ($L_{\mathrm{SVE}} = 125\,\mu$m, $N_S \approx 30$, range 2–4) to a median of 9 hops on the largest SVEs ($L_{\mathrm{SVE}} = 500\,\mu$m, $N_S \approx 1880$, range 9–10). A three-layer encoder can therefore propagate information across nearly the full smallest SVEs, but only across a local neighborhood on the largest SVEs. Global size information is therefore introduced through graph-level pooling and explicit size conditioning, rather than through message passing across the entire largest graph.

3.3.2 Node and graph features

Each grain node carries a 32-dimensional input feature vector $\mathbf{x}_g$ assembled from four families of microstructural descriptors that summarize the information available

to the surrogate without leaking any FIP-derived quantity: crystallography, per-grain morphology, voxel-geometry moments, and graph topology. The full per-block enumeration is given in Appendix D.2.

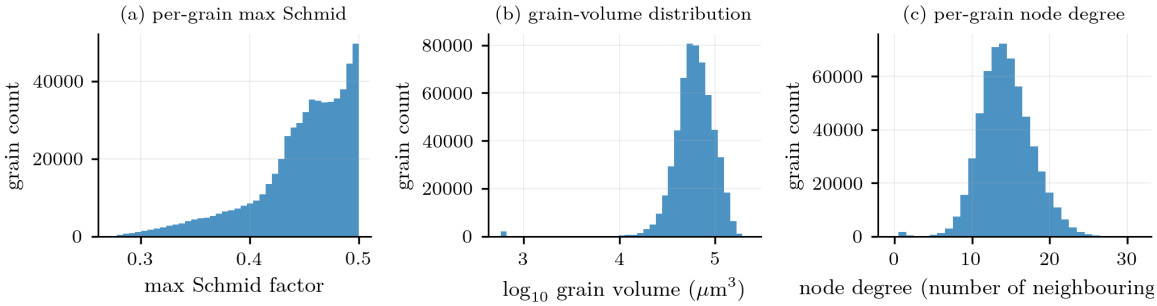


Figure 3.7: Pooled per-grain input-feature distributions across the dataset: (a) the per-grain maximum Schmid factor S_1, sharply peaked between 0.4 and 0.5 as expected for a random-texture FCC ensemble; (b) the grain volume on a $\log_{10}$ scale, approximately log-normal; (c) the per-grain node degree, with a mode near 14 neighbors and a long tail to about 30. These three quantities span the crystallography, morphology and topology blocks of the per-grain feature vector $\mathbf{x}_g$.

The crystallographic block (16 scalars) encodes the grain's orientation in the fundamental zone using a unit-quaternion representation, augmented by the twelve-dimensional vector of sorted-descending Schmid factors that projects the macroscopic loading axis onto each of the twelve octahedral FCC slip systems. The Schmid sub-block identifies the slip systems most favorably oriented for shear and is the strongest mechanics-side feature used by the trained model for $\mathrm{FIP}_{\mathrm{FS}}$ in the present dataset (Section 3.3.5); the per-grain maximum Schmid factor S_1, defined as the leading entry of the sorted vector, has an empirical distribution sharply peaked between 0.4 and 0.5 (Figure 3.7a). The per-grain morphology block (9 scalars) encodes grain volume, surface area, and perimeter together with six 2D-projected inertia descriptors of the grain shape; the volume distribution is approximately log-normal across the dataset (Figure 3.7b). The voxel-geometry block (6 scalars) summarizes the voxel-level boundary geometry of each grain (interior-versus-surface voxel fractions, distance-to-boundary moments, and grain voxel count); these features arise from voxel-level adjacency counts and do not use any cyclic-plasticity field. The topological block

(1 scalar) exposes the per-grain node degree, that is, the number of voxel-adjacent grain neighbors (Figure 3.7c). Together these four blocks form the 32-dimensional per-grain input vector $\mathbf{x}_g$ used by the encoder.

A small number of optional graph-level scalar conditioners are available alongside $\mathbf{x}_g$, namely the SVE edge length L_{SVE} and its size-bucket index. Earlier ablation studies on this dataset show that exposing L_{SVE} explicitly to the encoder removes a small size-dependent bias in the per-SVE maximum prediction without changing the per-grain prediction quality; we adopt this conditioning by default. The feature set excludes sub-grain descriptors (intra-grain voxel statistics beyond a per-grain summary) and any quantity computed from the cyclic-plasticity fields; richer sub-grain descriptors are a clear avenue for the LCF/WP case whose data-side correlation structure is independent of the other three cases (Section 3.2.6).

3.3.3 Multi-task output heads

The encoder produces both per-grain and per-graph (per-SVE) outputs. Targets are organized symmetrically across the four FIP cases, with no case privileged at training time, and complemented by mechanical auxiliary quantities predicted from the same graph representation.

The per-grain head predicts four within-grain summaries for each FIP case: the per-grain peak m_g, the within-grain arithmetic mean $\bar{f}_g$, and the within-grain lognormal parameters μ_g and σ_g from Equation (3.4). Across four FIP cases, this gives 16 supervised per-grain targets. The peak target m_g is trained with a top-decile-weighted Huber loss to emphasize the upper tail of the within-grain distribution; the remaining three target families use standard mean-squared error on Yeo–Johnson- and z-score-normalized targets. Each per-grain head is wrapped with a learnable linear residual skip connection from the encoder hidden state, which provides a small gain in upper-tail accuracy without changing the architecture footprint.

The per-graph head predicts 14 supervised mean targets per SVE: the four per-SVE maxima $M_S = \max_{g \in S} m_g$, the four within-SVE means $\mu_M(S)$ of the per-grain peak population $\{m_g\}_{g \in S}$, and six volume-averaged mechanical auxiliaries motivated in Section 3.4.4. These mechanical auxiliaries are the yield strength Y, hardening modulus H, plastic-work density W_p at HCF and LCF, and stress amplitude σ_{amp} at HCF and LCF. The four M_S targets are trained as graph-level outputs directly, rather than recovered by post-processing the per-grain predictions. The rationale for this direct maximum head is evaluated in Section 3.4.2, where it is compared against the post-hoc maximum and the analytical EVT bridge of Equation (3.9).

Predictive uncertainty is attached only to the peak and maximum channels. Each per-case m_g output is co-trained with a heteroscedastic Gaussian negative-log-likelihood (NLL) variance output, yielding a learned per-grain predictive scale $\hat{\sigma}_{\text{pn}}(\mathbf{x}_g)$. Each per-case M_S output is co-trained analogously, yielding a learned per-graph predictive scale $\hat{\sigma}_{\text{pg}}(\mathbf{X}_G)$. In the ideal population-risk setting for a correctly specified Gaussian NLL and a sufficiently flexible model, these variance heads estimate conditional residual variances,

$$\hat{\sigma}^2_{\text{pn}}(\mathbf{x}_g) \to \text{Var}[m_g \mid \mathbf{x}_g], \qquad \hat{\sigma}^2_{\text{pg}}(\mathbf{X}_G) \to \text{Var}[M_S \mid \mathbf{X}_G]. \tag{3.10}$$

The present finite-sample multi-task setting departs from this idealization, so we interpret $\hat{\sigma}_{\text{pn}}$ and $\hat{\sigma}_{\text{pg}}$ as learned predictive scales and assess their reliability through held-out interval coverage in Section 3.4.3. The mechanical auxiliary outputs share the per-graph head architecture but are trained as standard MSE regressions without NLL variance outputs; the multi-task layout is therefore symmetric across the four FIP cases but asymmetric across FIP statistics and mechanical auxiliaries.

The full training loss combines the head-specific losses into a single case-symmetric objective:

$$L(\theta) \;=\; \sum_{c \in \mathcal{C}} \Big[w_m \, L_m^{(c)} + w_M \, L_M^{(c)} + w_a \sum_{s \in \mathcal{S}} L_s^{(c)} \Big] \;+\; w_{\text{aux}} \sum_{q \in \mathcal{Q}} L_q, \tag{3.11}$$

where $\mathcal{C} = \{\text{HCF/FS}, \text{HCF/WP}, \text{LCF/FS}, \text{LCF/WP}\}$ indexes the four FIP cases, $\mathcal{S} = \{\bar{f}_g, \mu_g, \sigma_g, \mu_M(S)\}$ enumerates the distributional auxiliary targets, and $\mathcal{Q}$ enumerates the six volume-averaged mechanical auxiliaries (Y, H, W_p and σ_{amp} at HCF and LCF). The per-grain peak loss $L_m^{(c)}$ is the sum of a top-decile-weighted Huber loss on m_g and a heteroscedastic Gaussian NLL on $(m_g, \hat{\sigma}_{\text{pn}})$. The per-SVE maximum loss $L_M^{(c)}$ is the sum of a Huber loss on M_S and a heteroscedastic Gaussian NLL on $(M_S, \hat{\sigma}_{\text{pg}})$. Each $L_s^{(c)}$ and L_q is a standard MSE on a Yeo–Johnson- and z-score-normalized target, with predictions inverted to raw FIP or mechanical units at evaluation. The four loss-weight constants $\{w_m, w_M, w_a, w_{\text{aux}}\}$ are HPO-tuned; selected values are reported in Appendix D.1. The encoder-and-head architecture and data flow are summarized in Figure 3.8.

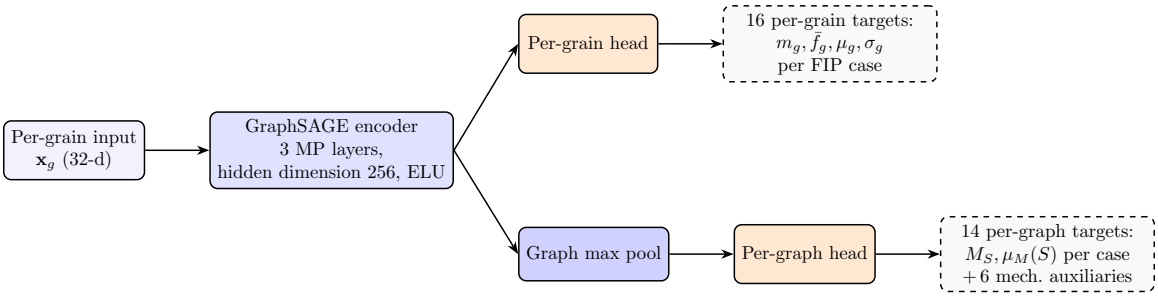


Figure 3.8: Hierarchy-aware multi-task architecture. Per-grain features $\mathbf{x}_g$ enter a shared 3-layer GraphSAGE encoder; per-grain hidden states feed the per-grain head directly, while graph-level max pooling aggregates them into a graph-level embedding that feeds the per-graph head. Per-grain outputs are the 16 within-grain summaries across the four FIP cases; per-graph outputs are the four per-SVE maxima M_S, the four within-SVE means $\mu_M(S)$, and the six volume-averaged mechanical auxiliaries.

3.3.4 Training and hyper-parameter selection

The encoder is a stack of GraphSAGE convolutions (Hamilton et al., 2017) terminated by graph-level max pooling. The message-passing layers produce per-grain hidden states that feed the per-grain head directly, while max pooling aggregates the same hidden states into a graph-level embedding for the per-graph head. The architecture hyper-parameters that control model capacity and regularization (hidden dimension,

number of message-passing layers, dropout rate, activation function, graph pooling, and loss-weight coefficients) were selected by a multi-objective HPO search across the four FIP cases. The target metrics balance tail R^2, held-out interval coverage, and mechanical-auxiliary R^2. The selected configuration uses hidden dimension 256, three message-passing layers, ELU activation, dropout rate 0.10, and graph-level max pooling. Full HPO details, including the search space, NSGA-II algorithm, convergence behavior, configuration JSON path, dataset hash, and per-claim seed protocol, are reported in Appendix D.1.

The total training loss is the case-symmetric multi-task objective of Equation (3.11), summed across the four FIP cases and the six mechanical auxiliaries. Targets are Yeo–Johnson- and z-score-normalized inside the loss and inverted to raw FIP or mechanical units only at evaluation. Early stopping uses the validation R^2 of the per-grain peak prediction. Multi-seed training uses ten independent split seeds with a fixed initialization seed, producing ten replicates of the selected configuration for the headline results in Section 3.4. The GraphSAGE backbone was the HPO-selected operating point for the joint per-grain and per-graph objective; it accommodates irregular grain counts naturally and encodes microstructural anisotropy through the input features (L_{SVE}, the Schmid block, and quaternion orientation) rather than through the message-passing operator. Architecture variants beyond this selected GraphSAGE family, including attention-based and equivariant alternatives, were not benchmarked here; the selected configuration is therefore held fixed across the four FIP cases.*

*An edge-attribute ablation using a PNA backbone with interface area, misorientation angle, and slip-direction inner product raised the HCF/FS per-grain R^2 marginally (0.579 versus 0.554 for GraphSAGE) but lowered the per-graph R^2 in the multi-task setup, and was therefore not retained.

3.3.5 Model interpretation: permutation importance and ablation

Once trained, the surrogate is interrogated to identify which input features the selected model relies on. We use feature-block permutation importance: at evaluation time, a feature or feature block in $\mathbf{x}_g$ is permuted across grains within the held-out test SVEs, and the resulting drop in per-grain peak R^2 is recorded. The ΔR^2 of a feature relative to the unpermuted baseline quantifies how much of the trained model's prediction quality flows through that feature. A feature that the model ignores produces a ΔR^2 near zero, whereas a feature on which the model strongly relies produces a large ΔR^2. The procedure is run on the selected configuration of Section 3.3.4 and averaged across multi-seed replicates to suppress seed-level noise. We complement permutation importance with leave-one-block ablation, in which the encoder is retrained from scratch with one mechanics block removed from $\mathbf{x}_g$; both diagnostics give consistent block rankings, and we display the permutation-importance result here.

The result is summarized in Figure 3.9. The twelve-dimensional slip-system Schmid block, which projects the macroscopic loading axis onto each of the twelve octahedral slip systems, carries by far the largest model reliance. Permuting this block produces $\Delta R^2 \approx 0.51$ relative to an unpermuted per-grain baseline of $R^2 = 0.535$, reducing the trained model's per-grain peak accuracy to near zero. Two scalar diagnostics derived from the same block, the maximum factor S_1 (the leading entry of the sorted vector) and the asymmetry between the two largest factors, are also permuted as additional probes of the schmid12 contribution and each produces $\Delta R^2 \approx 0.06$. The per-grain morphology and topology blocks contribute below $\Delta R^2 \approx 0.01$, and the orientation quaternion and voxel-geometry moments are effectively unused by the selected model.

This ranking is consistent across multi-seed replicates and across the HCF and LCF cases. It provides context for two results developed later. First, the weak

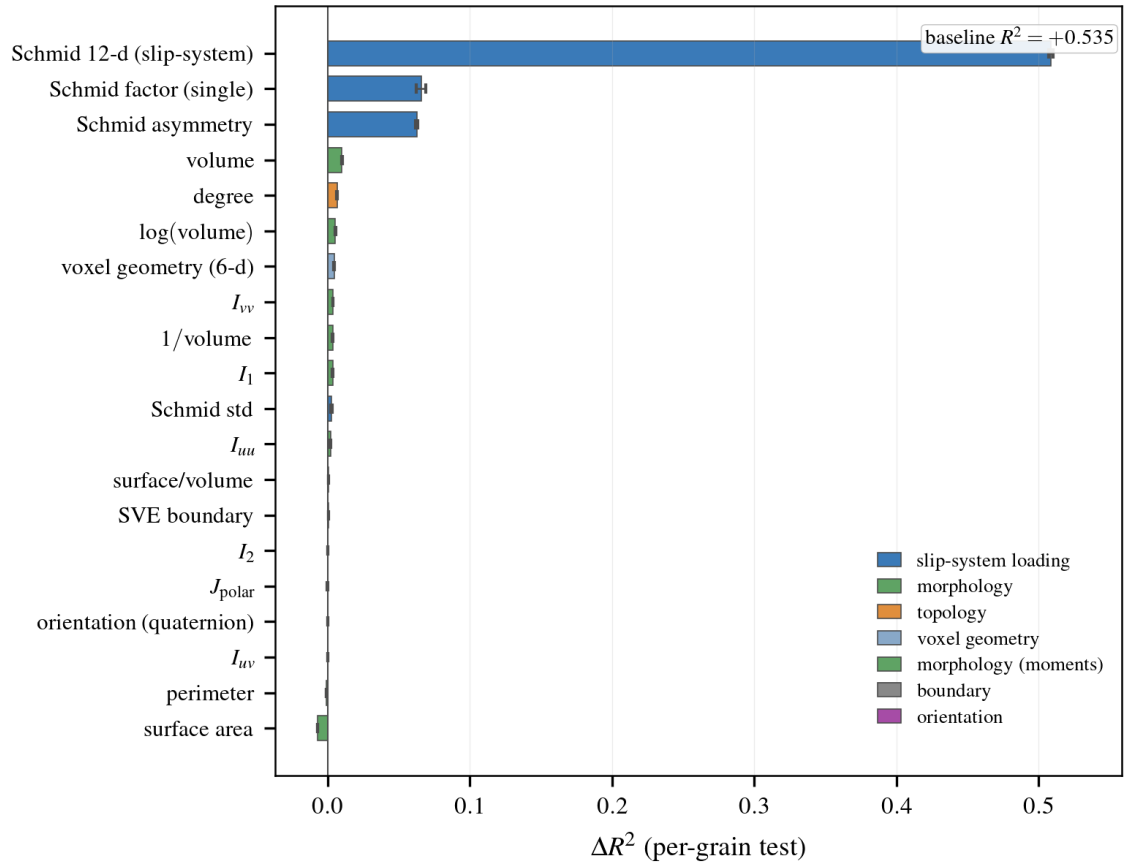


Figure 3.9: Per-feature permutation importance for per-grain peak prediction, evaluated on the held-out test split with the selected configuration. Bars are grouped by mechanics block. The twelve-dimensional slip-system Schmid block carries $\Delta R^2 \approx 0.51$ relative to a baseline per-grain $R^2 = 0.535$, while the remaining feature blocks contribute $\Delta R^2 \leq 0.07$ each.

LCF/WP performance in Section 3.4.1 occurs in a case whose dominant predictable signal is still the slip-system Schmid block, while the additional microstructural information needed for low-cycle plastic-work localization is not resolved by the present descriptors. Second, the HCF/FS bridge pathology discussed in Section 3.4.2 occurs in the regime where the selected model relies most strongly on the slip-system block. These connections should be read as model-interpretation evidence rather than as causal statements about the underlying CPFEM fields. Permutation importance ranks features by how much the trained model uses them; it is conditional on the selected architecture, feature set, and training objective.

3.4 Results and discussion

3.4.1 Learnability map across fatigue criteria and loading regimes

The four FIP cases are predicted in parallel by the multi-task encoder of Section 3.3.4, with no case privileged at training time. We first ask which cases are learnable from the present grain-level descriptors before turning to the per-SVE maximum M_S. Figure 3.10 reports the per-grain peak parity on the held-out test split for the selected configuration: hexbin density of predicted m_g versus true m_g, one panel per case, with the per-case single-replicate R^2 annotated. The HCF cases are highly learnable at the grain level, with multi-seed $R^2 = +0.95 \pm 0.00$ on both HCF/FS and HCF/WP across the full grain population ($n_{\text{rep}} = 10$). The LCF/FS case is moderately learnable ($R^2 = +0.50 \pm 0.01$, single-replicate $+0.51$), whereas the LCF/WP case is not learnable from the present descriptors ($R^2 = -0.47 \pm 0.02$, single-replicate -0.42), worse than predicting the marginal mean.

The case-by-case ranking is consistent with the data-side correlation structure of Section 3.2.6. The two HCF cases have nearly identical voxel-level rank ordering ($\rho_S = 0.97$), and the per-grain ranking they induce is dominated by the same slip-system

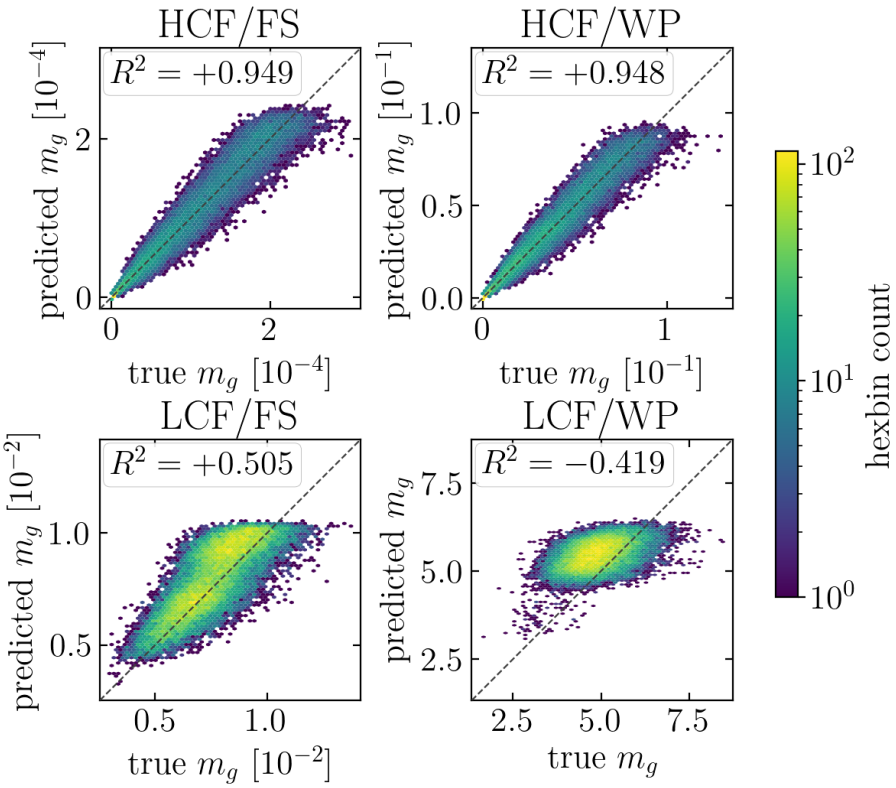


Figure 3.10: Per-grain peak FIP parity on the held-out test split, one panel per case. Hexbin density of predicted m_g versus true m_g, with the per-case R^2 annotated. The HCF cases are highly learnable ($R^2 = +0.95$ on both); LCF/FS is moderately learnable ($R^2 = +0.51$ on the displayed replicate); and LCF/WP is feature-limited ($R^2 = -0.42$ on the displayed replicate).

Schmid signal identified by the permutation-importance analysis of Section 3.3.5. Both HCF cases therefore occupy the large-R^2 regime. The LCF/FS case shares partial transferability with HCF/FS through the same critical-plane criterion ($\rho_S = 0.68$), but loses correlation with the WP-based cases, placing its per-grain R^2 between the HCF cases and LCF/WP. The LCF/WP case has weak off-diagonal voxel-level correlations with all other cases ($\rho_S \leq 0.11$), so the multi-task encoder cannot borrow the shared structure that supports the other three heads.

The negative LCF/WP R^2 reflects a feature-side limitation under the current representation rather than a generic training failure. The present descriptors resolve the sparse, Schmid-dominated localization structure of the HCF cases and partially resolve LCF/FS, but they do not capture the microstructural information controlling plastic-work localization in the LCF/WP case, and this limitation persists across the explored GraphSAGE/HPO search space. All four cases are reported symmetrically. The two dominant failure modes are distinguished in Section 3.4.2, and the LCF/WP feature limitation is returned to in the future-work discussion.

Table 3.3: Qualitative summary across the four FIP cases. Each row records the observed multi-seed pattern, the dominant interpretation, and the implied intervention. Quantitative errors are reported in Table 3.4.

case	observed pattern	dominant interpretation	implied intervention
HCF/FS	grain head strong; bridge slightly edges the dedicated head	empirical-moment bridge is informative; deployable bridge stays positive	no targeted correction
HCF/WP	grain head strong; all estimators viable	balanced reference case	no targeted correction
LCF/FS	empirical-moment bridge beats trained head; predicted-moment bridge collapses	EVT map is informative when moments are known	improve moment prediction
LCF/WP	grain head weak; predicted-moment bridge collapses	grain-level feature limitation	richer per-grain descriptors

Per-grain R^2 characterizes the average learnability of grain-level hot spots, but it is not sufficient to assess fatigue-critical response: the engineering target is the per-SVE maximum M_S. Table 3.3 previews the qualitative pattern that is quantified in Section 3.4.2. Under the mixture working model of Section 3.2.4, the empirical-moment bridge A_t returns positive R^2 on all four cases. The cases then separate by deployment-bridge behavior. On both HCF cases, the predicted-moment bridge A_p also returns positive R^2, because the per-grain head reproduces the plastic-grain upper tail of $\{m_g\}_{g\in S}$ closely enough that the bridge applied to the predicted moments preserves the cross-SVE ranking. On both LCF cases, the predicted-moment bridge collapses to negative R^2; the dedicated head D is the deployable estimator on those cases, with A_t providing the structural reference. The empirical-moment bridge advantage previously isolated to LCF/FS now extends to LCF/WP and HCF/FS, with paired-test support at $n_{\text{rep}} = 10$. HCF/WP is the balanced case in which all three non-baseline estimators are viable. The following subsection quantifies these estimator comparisons with R^2 and raw error metrics.

3.4.2 Predicting the per-SVE maximum

This subsection evaluates prediction of the per-SVE maximum M_S, the engineering target of the hierarchy. The per-grain learnability map of Section 3.4.1 is necessary but not sufficient for this evaluation, because M_S is not determined by average per-grain accuracy alone. We compare four estimators. The first is the dedicated graph head, denoted D, trained directly on M_S. The second is the naive order-statistic estimator $\max_g \hat{m}_g$ obtained by taking the maximum over the per-grain predictions. The third is the empirical-moment EVT bridge, A_{t}, which evaluates the mixture-Gumbel mean of Equation (3.9) using the held-out true within-SVE plastic-grain moments of the true $\{m_g\}_{g\in S}$. Because A_{t} consumes held-out true per-grain peaks (the very quantities the surrogate is trained to predict), it is a structural diagnostic of the EVT map and not a deployable surrogate; we report its R^2 as an upper bound on

what the closed-form bridge could achieve if per-grain prediction were perfect. The fourth is the predicted-moment EVT bridge, A_{p}, which evaluates the same formula using moments estimated from the predicted $\{\widehat{m}_g\}_{g\in S}$; unlike A_{t}, A_{p} is deployable. The dedicated head D tests direct learning of the volume extreme, A_{t} tests the extreme-value map when the lower-level moments are known, and A_{p} tests whether that map can be reached from the per-grain head alone.

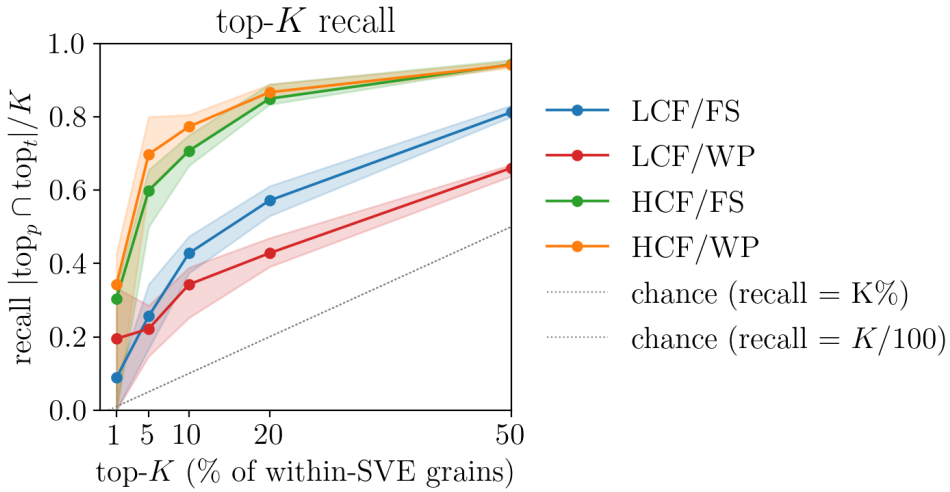


Figure 3.11: Within-SVE top-K recall of the per-grain head across the four FIP cases, averaged over test SVEs and multi-seed replicates. The recall is the fraction of the truly highest K percent of grains within an SVE that the predicted ranking also places in its top K. The HCF cases lift sharply above the chance diagonal recall = $K/100$; the LCF cases are weaker but still informative.

The within-SVE ranking quality of the per-grain head is summarized by top-K recall: the fraction of the truly highest K percent of grains within an SVE that the predicted ranking also places in its top K. Figure 3.11 reports the recall curves for $K \in \{1, 5, 10, 20, 50\}$ percent, averaged across test SVEs and multi-seed replicates. The HCF cases lift sharply above the chance diagonal, recall = $K/100$: top-10% recall reaches 0.77 on HCF/WP and 0.70 on HCF/FS, while top-1% recall reaches 0.34 and 0.30, respectively. The LCF cases are weaker but still informative, with top-10% recall of 0.43 on LCF/FS and 0.34 on LCF/WP. The per-grain head therefore

provides useful within-SVE ranking information, especially on the HCF cases, but reliable identification of the single most damaging grain remains difficult.

This limitation becomes decisive when the per-grain predictions are used as an order statistic. Taking $\max_g \widehat{m}_g$ to estimate M_S fails on all four cases: across the multi-seed grid, the resulting per-SVE R^2 values are -0.85 ± 0.32 on HCF/FS, -0.47 ± 0.29 on HCF/WP, -2.05 ± 0.94 on LCF/FS, and -1.23 ± 0.48 on LCF/WP (Figure 3.14, red bars). The failure is the expected order-statistic consequence of regression-to-the-mean shrinkage: even when the per-grain head has high population-level R^2, it under-predicts the far upper tail, so the maximum of the predictions is a biased estimator of the maximum of the truth. The per-SVE maximum must therefore be treated as a separate statistical target rather than as a post-processing operation on per-grain predictions.

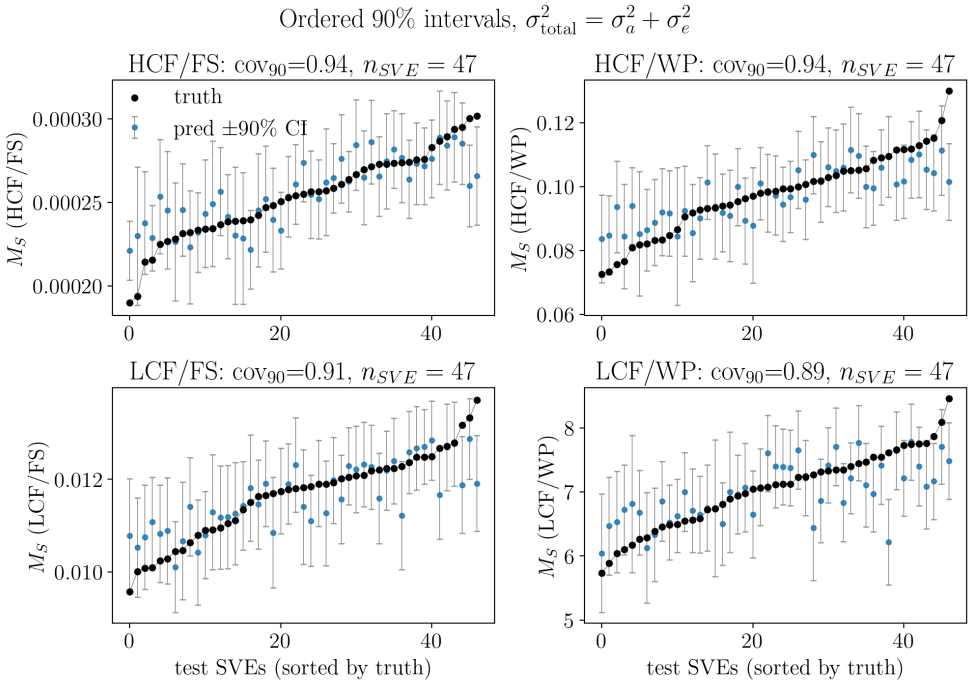


Figure 3.12: Ordered 90% intervals on the per-SVE maximum from the dedicated graph head, single rep. Test SVEs are sorted by truth ($n_{\text{SVE}} = 47$); black points show the truth, blue points show the predicted mean with the symmetric interval $\pm 1.645\sqrt{\hat{\sigma}^2_a + \hat{\sigma}^2_e}$. Empirical 90% coverages, annotated per panel, lie within 0.04 of nominal on all four cases.

The dedicated graph head trained directly on M_S provides informative per-SVE predictions on all four cases. Figure 3.12 shows the ordered 90% interval representation for one representative replicate: within each case, the $n_{\mathrm{SVE}} = 47$ test SVEs are sorted by the true M_S, and the predicted mean is plotted with the symmetric interval

$$\hat{M}_S \pm 1.645\sqrt{\hat{\sigma}_a^2 + \hat{\sigma}_e^2},$$

where $\hat{\sigma}_a$ is the heteroscedastic NLL scale from Section 3.3.3 and $\hat{\sigma}_e$ is the MCD standard deviation across $K = 30$ stochastic forward passes. The empirical 90% coverages on this replicate are 0.94 on HCF/FS and HCF/WP, 0.91 on LCF/FS, and 0.89 on LCF/WP. The multi-seed dedicated-head R^2 values are $+0.46 \pm 0.10$ on HCF/FS, $+0.46 \pm 0.18$ on HCF/WP, $+0.44 \pm 0.27$ on LCF/FS, and $+0.49 \pm 0.12$ on LCF/WP (Figure 3.14, blue bars).

The EVT bridge gives a complementary route to the same target. The empirical-moment bridge A_{t} isolates the adequacy of the closed-form mixture-Gumbel map, while the predicted-moment bridge A_{p} adds the error induced by predicting the within-SVE moments from the per-grain head. Figure 3.13 shows the four estimators on a single replicate, and Figure 3.14 summarizes their multi-seed R^2 distributions. The contrast between A_{t} and A_{p} is central: when A_{t} succeeds but A_{p} fails, the bridge itself is informative but the moment estimates are not deployable; when both succeed, the bridge can substitute for or complement the dedicated head.

The empirical-moment bridge returns positive multi-seed R^2 on every case: $+0.60 \pm 0.04$ on HCF/FS, $+0.52 \pm 0.12$ on HCF/WP, $+0.84 \pm 0.04$ on LCF/FS, and $+0.88 \pm 0.04$ on LCF/WP. Against the dedicated head D, the bridge wins in 9 of 10 replicates on HCF/FS (sign-test $p = 0.022$, Wilcoxon $p = 0.006$), in 10 of 10 replicates on LCF/FS (sign-test $p = 0.002$, Wilcoxon $p = 0.002$), and in 10 of 10 replicates on LCF/WP (sign-test $p = 0.002$, Wilcoxon $p = 0.002$). On HCF/WP the bridge wins in 7 of 10 replicates without paired-test significance at $n_{\mathrm{rep}} = 10$ (sign-test $p = 0.34$), so on that case the dedicated head and the empirical-moment

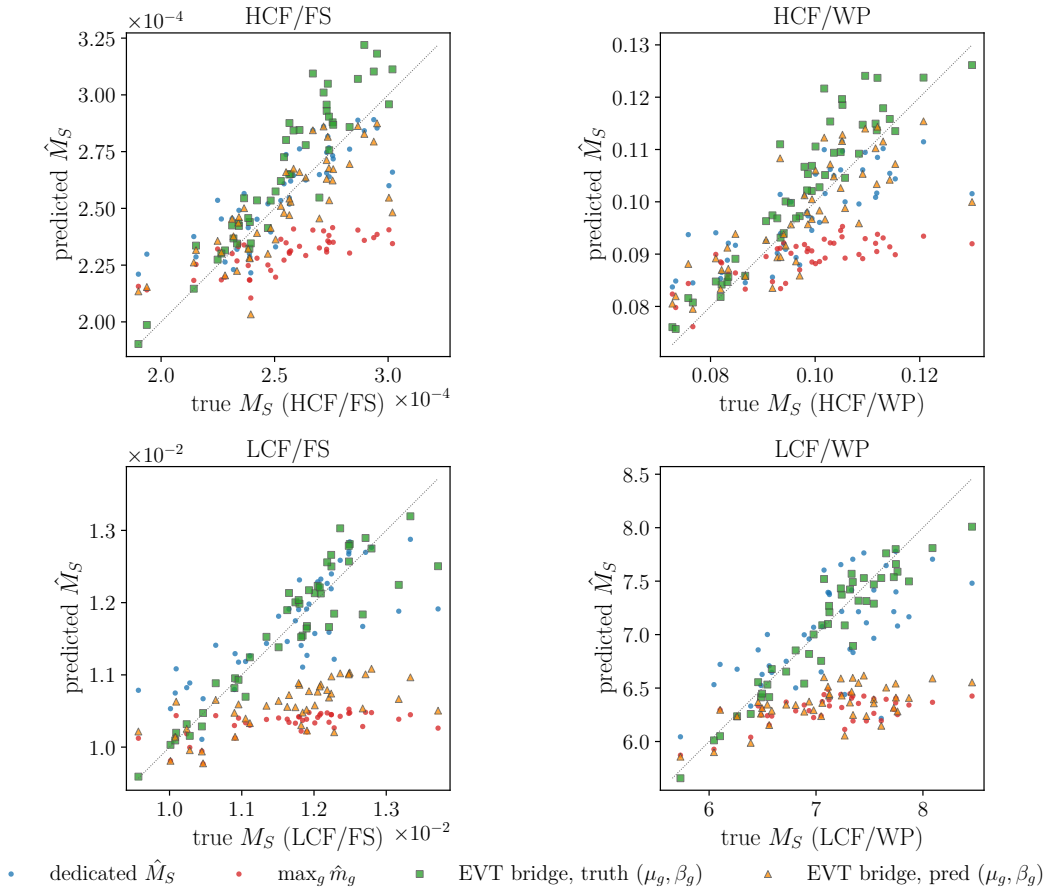


Figure 3.13: Four estimators of the per-SVE maximum M_S on a single replicate, one panel per case: dedicated graph head $\hat{M}_S$ (blue), post-hoc $\max_g \hat{m}_g$ (red), mixture-Gumbel bridge with empirical moments of the plastic-grain subset (green), and mixture-Gumbel bridge with predicted moments (orange). The empirical-moment bridge tracks the diagonal most cleanly across all four cases; the predicted-moment bridge tracks the diagonal on the HCF cases but collapses on the LCF cases, where the per-grain head does not preserve the plastic-grain upper tail.

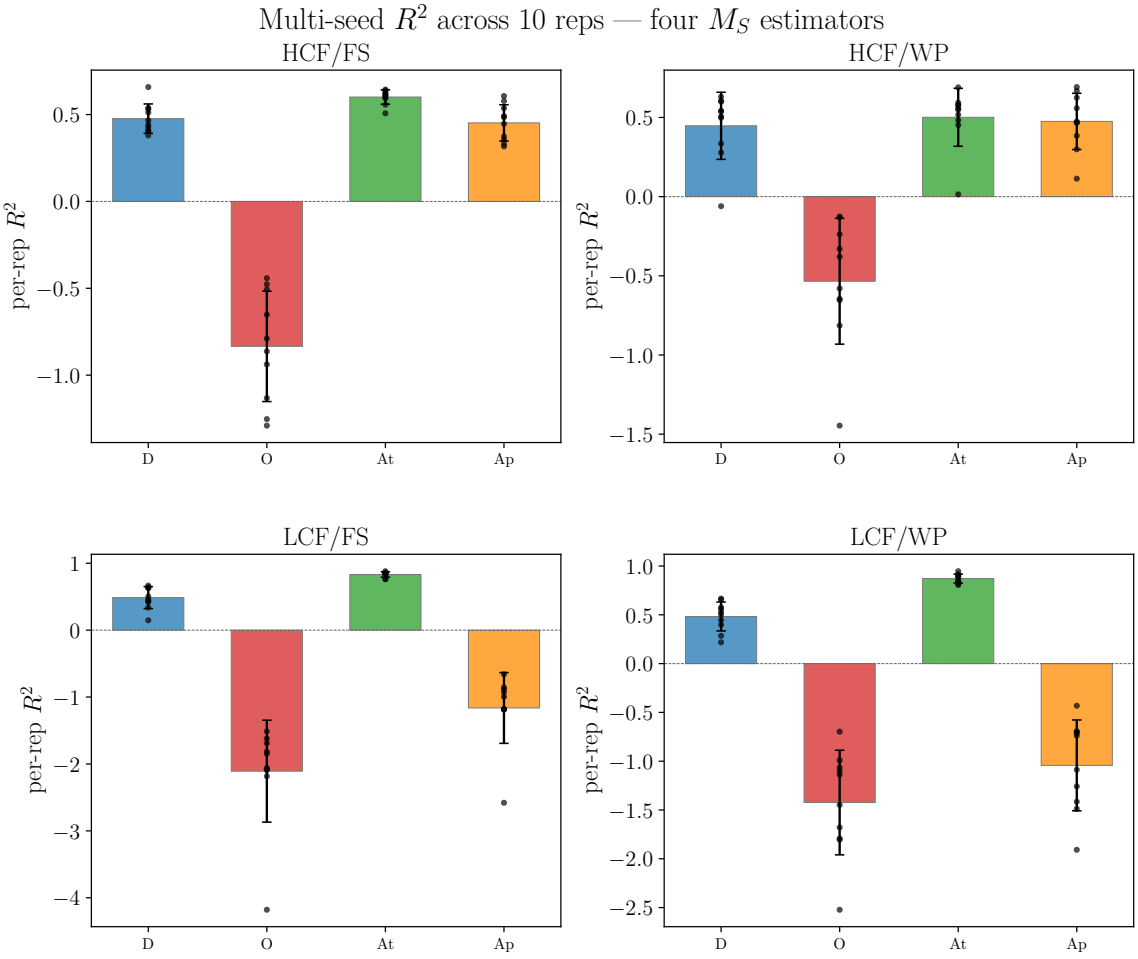


Figure 3.14: Per-rep R^2 of the four estimators across ten multi-seed replicates ($n_{\text{rep}} = 10$), one panel per case. Bars show the mean across reps with ± 1 standard deviation; black points show individual reps. The empirical-moment mixture bridge A_{t} attains the highest structural R^2 on every case and beats the dedicated head on three of four (HCF/FS: 9/10 reps, $p = 0.022$; LCF/FS: 10/10, $p = 0.002$; LCF/WP: 10/10, $p = 0.002$). The predicted-moment bridge A_{p} remains positive on the HCF cases but collapses on the LCF cases, where the per-grain head does not preserve the plastic-grain upper tail. The post-hoc max baseline $\max_g \hat{m}_g$ returns negative R^2 on every case. Companion error magnitudes (RMSE, MAE, MAPE in raw FIP units) appear in Table 3.4.

bridge are statistically indistinguishable. The bridge advantage previously identified on LCF/FS therefore extends to LCF/WP and HCF/FS under the mixture working model. The bridge is also more stable across replicates on LCF/FS than the dedicated head, with $\sigma_r(A_\mathrm{t}) = 0.04$ compared with $\sigma_r(D) = 0.27$.

The predicted-moment bridge A_p behaves differently on the HCF and LCF cases. On both HCF cases, A_p returns positive R^2 ($+0.44 \pm 0.11$ on HCF/FS, $+0.47 \pm 0.18$ on HCF/WP), and the loss relative to the empirical-moment bridge is modest ($A_\mathrm{p} - A_\mathrm{t} \approx -0.16$ on HCF/FS, -0.05 on HCF/WP). On these cases the per-grain head reproduces the within-SVE plastic-grain moments accurately enough that the bridge applied to the predicted moments retains positive R^2. On both LCF cases, A_p collapses to strongly negative R^2 (-1.24 ± 0.63 on LCF/FS, -1.00 ± 0.45 on LCF/WP), and the gap to the empirical-moment bridge is large ($A_\mathrm{p} - A_\mathrm{t} \approx -2.08$ on LCF/FS, -1.88 on LCF/WP). The per-grain head's bulk-level performance on the LCF cases does not translate into accurate within-SVE moments of the plastic-grain subset, and the deployment bridge therefore fails on those cases even though the empirical-moment bridge is the strongest estimator.

These patterns separate the four cases by deployment behavior rather than by a within-SVE working-model failure. On both HCF cases the dedicated head, the empirical-moment bridge, and the predicted-moment bridge are all viable; the bridge edges the head on HCF/FS and is indistinguishable on HCF/WP. On both LCF cases the empirical-moment bridge has the highest structural R^2 with paired-test support, but the predicted-moment bridge is not deployable, so the dedicated head remains the practical estimator until the per-grain prediction quality on the plastic-grain upper tail is improved.[†] The four-estimator comparison is reported for the selected

[†]As a cross-alloy check on the mixture working model, we re-fit the plastic-grain fraction and re-evaluated the empirical-moment bridge on the publicly available PRISMS-Fatigue dataset for Al 7075-T6 under high-cycle loading (Yaghoobi et al., 2023). The bridge returns positive pooled R^2 on that dataset under the same per-SVE fitting rule, with an alloy-specific plastic-grain fraction; only the HCF/FS case is testable since PRISMS-Fatigue does not release FIP_WP or LCF data.

architecture and feature set on this dataset at $n_{\text{rep}} = 10$; architecture variants are not benchmarked here.

Table 3.4 reports the same estimator comparison using RMSE, MAE, and MAPE in raw FIP units alongside R^2. The case-dependent ordering is consistent across the metrics. The empirical-moment bridge gives the lowest error on every case as a structural fit, with the caveat below; the predicted-moment bridge matches the dedicated head on the HCF cases and degrades strongly on the LCF cases. Thus the diagnostic patterns in Figure 3.14 are not artifacts of unit-free R^2; they remain visible in engineering-unit error statistics. The next subsection turns from per-SVE point prediction to distributional recovery and held-out interval coverage.

3.4.3 Distribution-aware prediction and interval coverage

Section 3.4.2 evaluated point prediction of the per-SVE maximum M_S. We now ask whether the surrogate also recovers the distributional quantities that support the hierarchy of Sections 3.2.3–3.2.5, and whether the predictive scales attached to the peak and maximum channels produce reliable held-out interval coverage. This distinction matters because the surrogate is not only a regressor for m_g and M_S; it is also intended to represent the lower-level statistical structure from which the extreme-value bridge is constructed.

The per-grain head predicts the within-grain lognormal parameters μ_g and σ_g of Equation (3.4), while the per-graph head predicts the within-SVE mean $\mu_M(S)$ of the per-grain peak population. The within-SVE scale $\sigma_M(S)$ is not a separate supervised output; it is recovered post-hoc as the empirical standard deviation of the predicted per-grain peaks $\{\widehat{m}_g\}_{g \in S}$. Figure 3.15 reports the within-grain log-shape parity on the HCF cases, where distributional recovery is strongest. The log-location is recovered with $R^2 = +0.97$ on HCF/WP and $R^2 = +0.87$ on HCF/FS, and the log-scale with $R^2 = +0.73$ on both HCF cases. The corresponding LCF results are weaker, following the same case ordering as the per-grain peak prediction in Section 3.4.1.

Table 3.4: Extended error statistics for the four M_S estimators across the four FIP cases under the mixture working model, multi-seed ($n_{\text{rep}} = 10$) test-set values (mean $\pm$ std). The bridge A_t is the closed-form Gumbel-mean evaluated at the empirical within-SVE moments of the plastic-grain subset, and A_p the same formula evaluated at the per-grain head's predicted moments. RMSE and MAE are in raw FIP units; MAPE is in percent. The empirical-moment bridge A_t attains the highest structural R^2 on every case but is not deployable because it consumes held-out true per-grain peaks; it is reported as an upper bound on what the closed-form bridge could achieve if per-grain prediction were perfect. The deployable predicted-moment bridge A_p remains positive on the HCF cases but collapses on the LCF cases.

Case	Estimator	R^2	RMSE	MAE	MAPE (%)
HCF/FS	$\hat{M}_S^D$ (dedicated)	$+0.46 \pm 0.10$	$(1.90 \pm 0.23) \times 10^{-5}$	$(1.48 \pm 0.17) \times 10^{-5}$	5.8 ± 0.61
	$\max_g \hat{m}_g$ (post-hoc)	-0.85 ± 0.32	$(3.50 \pm 0.21) \times 10^{-5}$	$(2.95 \pm 0.24) \times 10^{-5}$	11.0 ± 0.89
	$\hat{M}_S^{A_t}$ (mixture-truth)	$+0.60 \pm 0.04$	$(1.63 \pm 0.13) \times 10^{-5}$	$(1.29 \pm 0.13) \times 10^{-5}$	4.9 ± 0.45
	$\hat{M}_S^{A_p}$ (mixture-pred)	$+0.44 \pm 0.11$	$(1.94 \pm 0.22) \times 10^{-5}$	$(1.46 \pm 0.20) \times 10^{-5}$	5.6 ± 0.74
HCF/WP	$\hat{M}_S^D$ (dedicated)	$+0.46 \pm 0.18$	$(7.57 \pm 0.85) \times 10^{-3}$	$(5.96 \pm 0.50) \times 10^{-3}$	6.2 ± 0.51
	$\max_g \hat{m}_g$ (post-hoc)	-0.47 ± 0.29	0.0126 ± 0.00073	0.0107 ± 0.0006	10.4 ± 0.54
	$\hat{M}_S^{A_t}$ (mixture-truth)	$+0.52 \pm 0.12$	$(7.13 \pm 0.86) \times 10^{-3}$	$(5.59 \pm 0.56) \times 10^{-3}$	5.6 ± 0.68
	$\hat{M}_S^{A_p}$ (mixture-pred)	$+0.47 \pm 0.18$	$(7.39 \pm 1.28) \times 10^{-3}$	$(5.87 \pm 0.94) \times 10^{-3}$	6.0 ± 1.05
LCF/FS	$\hat{M}_S^D$ (dedicated)	$+0.44 \pm 0.27$	$(6.62 \pm 0.86) \times 10^{-4}$	$(5.16 \pm 0.74) \times 10^{-4}$	4.4 ± 0.61
	$\max_g \hat{m}_g$ (post-hoc)	-2.05 ± 0.94	$(1.57 \pm 0.09) \times 10^{-3}$	$(1.36 \pm 0.09) \times 10^{-3}$	11.2 ± 0.72
	$\hat{M}_S^{A_t}$ (mixture-truth)	$+0.84 \pm 0.04$	$(3.66 \pm 0.42) \times 10^{-4}$	$(2.67 \pm 0.31) \times 10^{-4}$	2.2 ± 0.26
	$\hat{M}_S^{A_p}$ (mixture-pred)	-1.24 ± 0.63	$(1.33 \pm 0.21) \times 10^{-3}$	$(1.14 \pm 0.19) \times 10^{-3}$	9.4 ± 1.59
LCF/WP	$\hat{M}_S^D$ (dedicated)	$+0.49 \pm 0.12$	0.432 ± 0.067	0.333 ± 0.045	4.8 ± 0.62
	$\max_g \hat{m}_g$ (post-hoc)	-1.23 ± 0.48	0.899 ± 0.041	0.757 ± 0.045	10.3 ± 0.56
	$\hat{M}_S^{A_t}$ (mixture-truth)	$+0.88 \pm 0.04$	0.206 ± 0.025	0.145 ± 0.018	2.0 ± 0.24
	$\hat{M}_S^{A_p}$ (mixture-pred)	-1.00 ± 0.45	0.835 ± 0.143	0.704 ± 0.124	9.6 ± 1.85

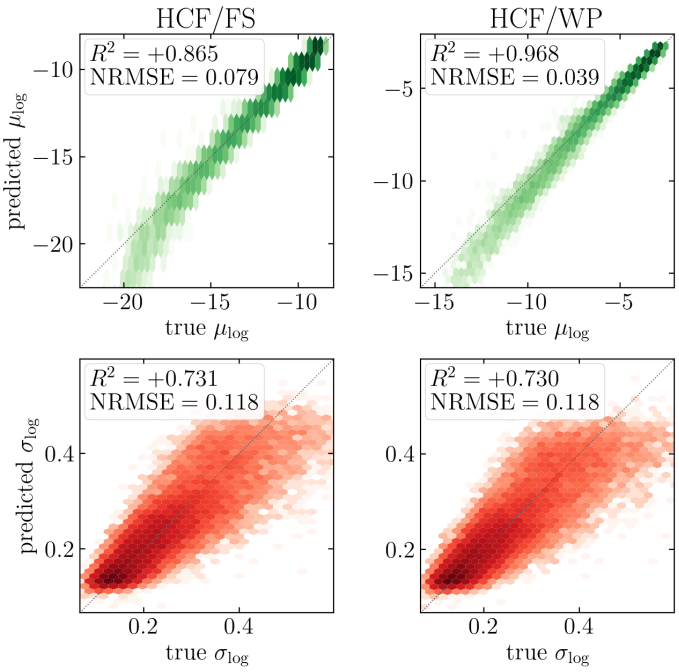


Figure 3.15: Within-grain log-shape parity on the HCF cases. The multi-task encoder recovers the within-grain log-mean μ_g at $R^2 = +0.97$ on HCF/WP and $+0.87$ on HCF/FS (top row), and the within-grain log-scale σ_g at $R^2 = +0.73$ on both HCF cases (bottom row). Hexbin density of predicted versus true on the held-out test split.

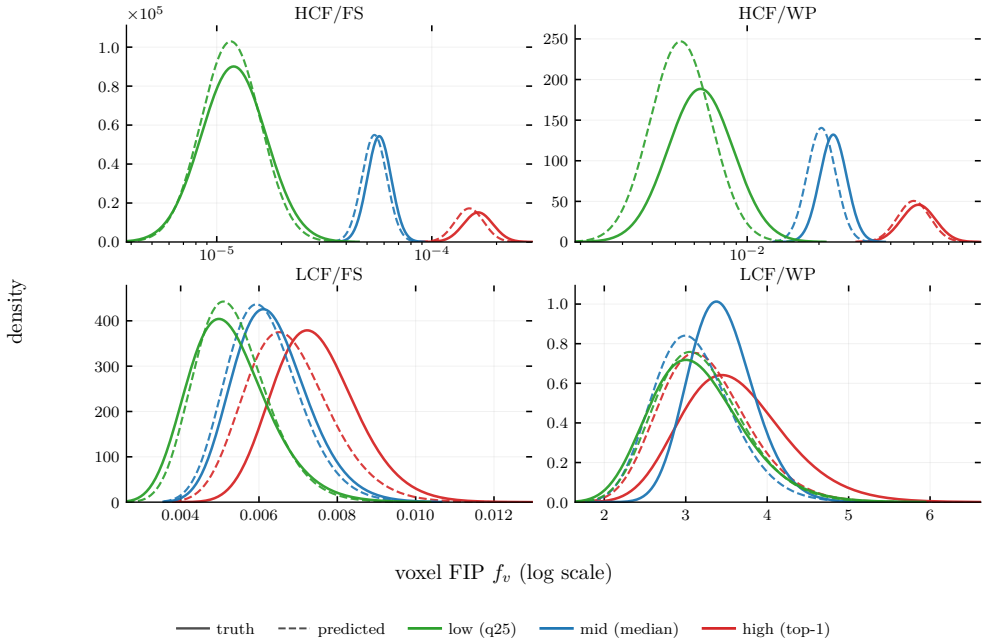


Figure 3.16: Predicted versus true within-grain voxel-FIP density for three representative grains per case (low, median, and high quantile by per-grain peak m_g). Solid lines: true within-grain density; dashed lines: surrogate-predicted density obtained by inverting $(\hat{\mu}_g, \hat{\sigma}_g)$ to a lognormal density. The HCF cases match closely across the within-grain bulk; the LCF cases exhibit a modest mode shift on the hardest grains.

Figure 3.16 provides a more direct view of what the log-shape predictions imply. For representative grains selected at low, median, and high quantiles of the per-grain peak m_g, the predicted lognormal density is reconstructed from $(\hat{\mu}_g, \hat{\sigma}_g)$ and compared with the empirical within-grain voxel-FIP density. The HCF cases are reproduced closely across the bulk of the within-grain distribution. The LCF cases show larger deviations, including a modest mode shift on the most difficult LCF/WP grains, consistent with the feature limitation identified in Section 3.4.1.

We next evaluate the predictive scales attached to the peak and maximum channels. The heteroscedastic outputs $\hat{\sigma}_{\mathrm{pn}}(\mathbf{x}_g)$ and $\hat{\sigma}_{\mathrm{pg}}(\mathbf{X}_G)$ are interpreted as learned predictive scales rather than guaranteed calibrated standard deviations. We therefore assess them empirically through coverage curves: for each nominal coverage level P, we compute the fraction $\widehat{P}_{\mathrm{cov}}(P)$ of held-out targets falling inside the corresponding symmetric predictive interval. Figure 3.17 reports this comparison at both hierarchy levels for one representative replicate.

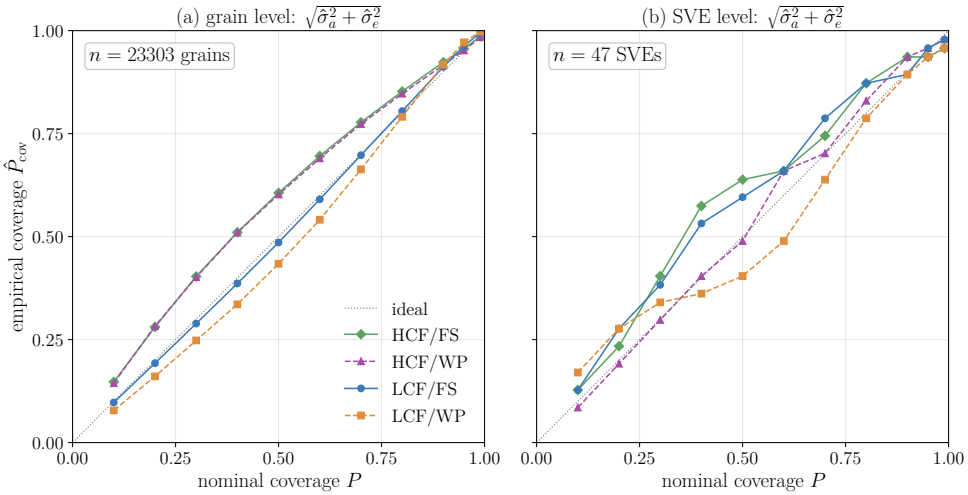


Figure 3.17: Calibration of the combined aleatoric-plus-epistemic predictive intervals, $\sqrt{\hat{\sigma}_a^2 + \hat{\sigma}_e^2}$. $\hat{\sigma}_a$ is the heteroscedastic NLL scale and $\hat{\sigma}_e$ is the Monte-Carlo Dropout standard deviation. (a) Grain-level coverage $\hat{P}_{\mathrm{cov}}(P)$ on $n = 23{,}303$ grains; (b) SVE-level coverage on $n = 47$ SVEs. Dotted diagonal is perfect calibration. Grain level tracks within 0.05 of nominal across all four cases; SVE-level scatter is wider due to the smaller sample, with LCF/WP under-covering across the nominal range.

At the grain level ($n_{\text{grain}} = 23{,}303$), the empirical coverage curves remain within approximately 0.05 of the diagonal over the nominal range. The HCF cases are slightly conservative, whereas the LCF cases either track the diagonal closely or under-cover mildly. At the SVE level, the test count is much smaller ($n_{\text{SVE}} = 47$ per case), so the apparent case-to-case ordering must be interpreted with finite-sample uncertainty: a single empirical coverage estimate has a binomial 95% uncertainty of about ± 0.14 at $P = 0.5$ and about ± 0.09 at $P = 0.9$. Within that limitation, HCF/FS and LCF/FS are slightly conservative at moderate nominal coverage, while LCF/WP under-covers across much of the range. This under-coverage is consistent with the LCF/WP feature limitation and suggests that the same missing descriptors that degrade the per-grain head may also contribute to under-estimated predictive uncertainty.

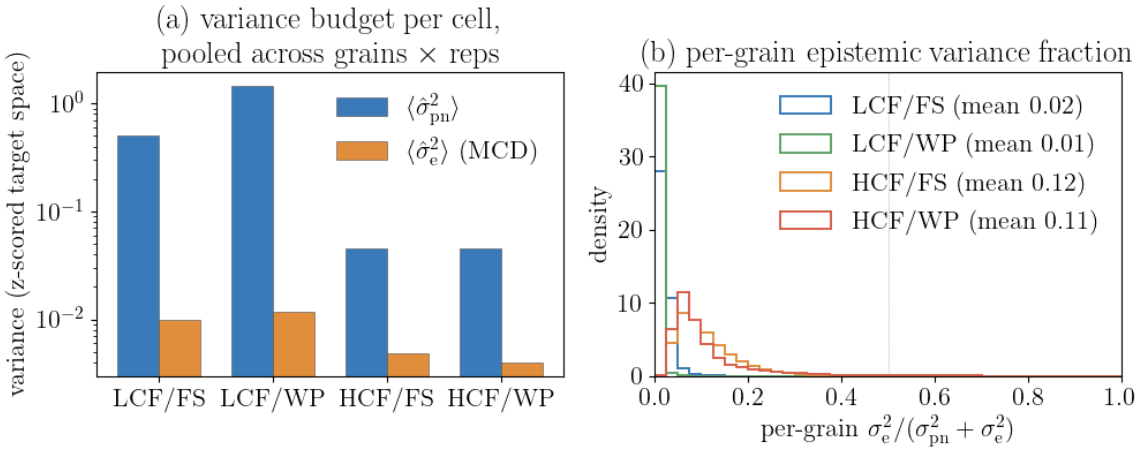


Figure 3.18: Aleatoric/epistemic variance decomposition of the per-grain predictive variance, pooled across grains and multi-seed reps. (a) Variance budget per case on a logarithmic axis: the NLL-trained $\langle\hat{\sigma}^2_{\text{pn}}\rangle$ dominates the Monte-Carlo Dropout $\langle\hat{\sigma}^2_{\text{e}}\rangle$ by an order of magnitude or more. (b) Per-grain epistemic variance fraction $\hat{\sigma}^2_{\text{e}}/(\hat{\sigma}^2_{\text{pn}} + \hat{\sigma}^2_{\text{e}})$, a histogram per case; LCF cases concentrate near zero (mean 0.01 and 0.02), HCF cases extend to several tens of percent (mean 0.11 and 0.12).

The aleatoric/epistemic decomposition of the per-grain predictive variance is shown in Figure 3.18. The NLL-trained aleatoric term dominates the MCD epistemic term by at least an order of magnitude across all four cases. The per-grain epistemic variance fraction $\hat{\sigma}^2_e/(\hat{\sigma}^2_{\text{pn}} + \hat{\sigma}^2_e)$ concentrates near zero on the LCF cases, with means

of approximately 0.01 on LCF/WP and 0.02 on LCF/FS, whereas the HCF cases extend to several tens of percent, with means of approximately 0.11 and 0.12. This pattern should not be read as evidence that the LCF predictions are more certain in an absolute sense. Rather, together with the negative LCF/WP per-grain R^2, it indicates that the dominant LCF limitation is data-side under the present feature representation; additional training capacity alone is unlikely to close this gap without richer descriptors. On the HCF cases, the non-negligible epistemic fraction suggests that multi-seed averaging or additional model capacity may still reduce part of the residual uncertainty.

The mechanical auxiliary heads reported next do not carry NLL variance outputs and therefore sit outside the interval-coverage analysis above. They instead test whether the same encoder representation also captures volume-averaged constitutive response quantities produced by the CPFEM simulations.

3.4.4 Mechanical quantites of interests

The FIP heads test whether the surrogate can recover fatigue-localization statistics, but the same CPFEM simulations also produce volume-averaged mechanical quantities that summarize the cyclic constitutive response. We use these quantities as auxiliary graph-level targets to test whether the shared encoder captures mechanically relevant information beyond the FIP-specific signal. The six auxiliary targets are the yield strength Y, hardening modulus H, plastic-work density W_p at HCF and LCF, and stress amplitude σ_{amp} at HCF and LCF. Because these quantities are volume-averaged responses of the same SVEs, they provide a macroscopic consistency check on the learned graph representation.

Figure 3.19 reports the parity of all six auxiliary outputs on the held-out test split, colored by SVE size bucket. The shared encoder recovers each channel with per-channel R^2 between $+0.77$ and $+0.89$ (mean $R^2 \approx 0.82$), approximately size-invariant across the five size buckets and without observable trade-off against the FIP

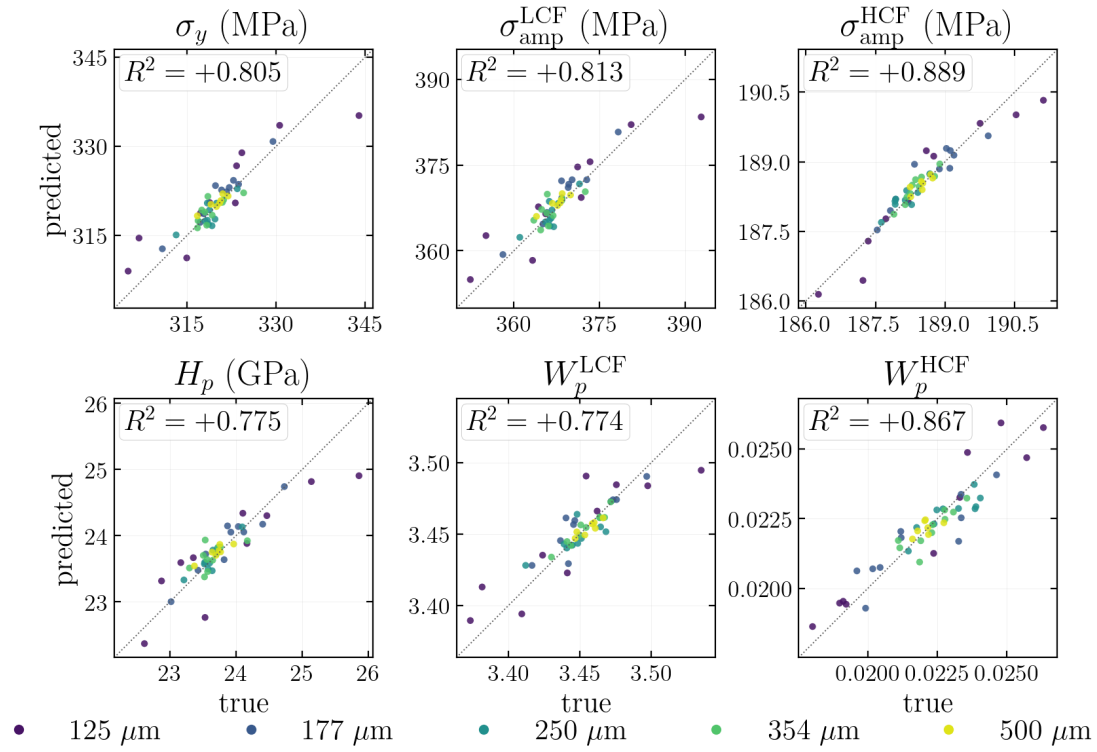


Figure 3.19: Parity of the six volume-averaged mechanical auxiliary outputs on the held-out test split, colored by SVE edge length L_{SVE}.

heads of Sections 3.4.1 and 3.4.2. We read these auxiliary R^2 values as supporting evidence that the encoder is mechanically grounded: the shared representation contains microstructural information that drives the cyclic constitutive response broadly, not only the FIP-specific signal. The auxiliary outputs are not validated against experimental constitutive measurements in the present work; whether they are accurate enough for downstream constitutive-calibration or fatigue-life-mapping use is a question for external-data validation we leave to future work.

3.5 Conclusion

This work developed a hierarchy-aware graph surrogate for microstructure-sensitive fatigue localization in austenitic stainless steel. The central objective was not only to predict the per-SVE maximum FIP, but also to determine whether that extreme can be explained from lower-level statistics of the CPFEM field. Three conclusions follow.

First, the multi-task graph surrogate provides a unified representation of the FIP hierarchy across the four cases obtained by crossing $\mathrm{FIP}_{\mathrm{FS}}$ and $\mathrm{FIP}_{\mathrm{WP}}$ with HCF and LCF loading. The model predicts within-grain distribution parameters, per-grain peaks, within-SVE moments of those peaks, and per-SVE maxima in a single architecture, while also providing heteroscedastic prediction intervals with approximately calibrated held-out coverage at both the grain and SVE levels. The same encoder also supports six volume-averaged mechanical auxiliary quantities, which achieve per-channel R^2 between $+0.77$ and $+0.89$ (mean ≈ 0.82) on held-out SVEs. These auxiliary predictions do not introduce an observable trade-off with the FIP heads.

Second, the extreme-value bridge is structurally informative when the within-SVE statistics of the plastic-grain population are known. Under the two-component mixture working model, which combines an effectively elastic Dirac component at zero with a plastic-grain Gumbel component, the closed-form Gumbel-mean bridge

achieves multi-seed R^2 values between $+0.52$ and $+0.88$ (mean ≈ 0.71, $n_{\mathrm{rep}} = 10$) when supplied with empirical within-SVE moments. It also outperforms the directly trained graph-level maximum head on three of the four FIP cases with paired-test support. This result shows that, when the plastic-grain subset is known, the two-component Gumbel working model is a useful finite-SVE approximation to the extreme statistics of the per-grain peak population.

Third, the comparison between the trained maximum head, the bridge with empirical moments, and the bridge with predicted moments separates two different sources of error. The empirical-moment bridge tests the adequacy of the extreme-value map, whereas the predicted-moment bridge tests whether the learned per-grain description preserves the plastic-grain upper tail well enough to use that map. The predicted-moment bridge remains useful on the two HCF cases, with $R^2 \approx +0.45$, but fails on the two LCF cases, where it gives strongly negative R^2 values (-1.24 on LCF/FS and -1.00 on LCF/WP). The failure is therefore not primarily a failure of the extreme-value bridge itself; rather, it reflects under-prediction of the plastic-grain upper tail by the present per-grain surrogate in the LCF regimes.

The main limitation is the degradation of the deployment bridge in the LCF cases, where the learned per-grain peaks do not preserve the plastic-grain upper tail. This behavior suggests that the localization mechanisms active under LCF loading, especially for the plastic-work indicator, are not fully represented by the present pairwise grain-adjacency graph. Future work should therefore explore architectures that can represent higher-order microstructural interactions and mechanism-dependent neighborhoods, such as graph-attention models. These extensions may be particularly useful for cases where fatigue criticality is controlled by collective slip activity, interface-mediated constraint, or crack-driving measures more closely related to plastic work than to a single critical-plane hot spot.

Bibliography

Abedi, R., Garrard, J., and Acton, K. (2023). Elastic and strength properties of statistical volume elements: Determination of isotropic and homogeneous size limits. *Computers & structures*, 277:106959. 3, 5, 10, 17, 18, 19, 51, 52, 154

Abedi, R., Garrard, J., Yang, M., Acton, K., and Soghrati, S. (2022). Effect of boundary condition and statistical volume element size on inhomogeneity and anisotropy of apparent properties. *Mechanics of materials*, 173:104408. 5, 10, 17, 18, 19

Acton, K. and Baxter, S. (2018). Characterization of Random Composite Properties Based on Statistical Volume Element Partitioning. *ASCE Journal of Engineering Mechanics*, 144:04017168 1–9. 20, 23, 47

Acton, K. and Graham-Brady, L. (2009a). *Fitting an Anisotropic Yield Surface Using the Generalized Method of Cells*, pages 27–41. Springer Netherlands, Dordrecht. 3, 52

Acton, K. and Graham-Brady, L. (2009b). Meso-scale modeling of plasticity in composites. *Computer methods in applied mechanics and engineering*, 198(9-12):920–932. 3, 52

Acton, K. and Graham-Brady, L. (2010). Elastoplastic mesoscale homogenization of composite materials. *Journal of engineering mechanics*, 136(5):613–624. 3, 52

Acton, K., Sherod, C., Bahmani, B., and Abedi, R. (2019). Effect of Volume Element Geometry on Convergence to a Representative Volume. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B: Mechanical Engineering*, 5(3). 3

Acton, K. A., Baxter, S. C., Bahmani, B., Clarke, P. L., and Abedi, R. (2018). Voronoi tessellation based statistical volume element characterization for use in fracture modeling. *Computer Methods in Applied Mechanics and Engineering*, 336:135–155. 3, 20, 23, 52

Ahmad, R., Liu, M., Ortiz, M., Mukerji, T., and Cai, W. (2023). Homogenizing elastic properties of large digital rock images by combining CNN with hierarchical homogenization method. arXiv:2305.06519 [physics]. 6, 10, 11

Akiba, T., Sano, S., Yanase, T., Ohta, T., and Koyama, M. (2019). Optuna: A Next-generation Hyperparameter Optimization Framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '19, pages 2623–2631, New York, NY, USA. Association for Computing Machinery. 40

Alnaes, M. S., Blechta, J., Hake, J., Johansson, A., Kehlet, B., Logg, A., Richardson, C. N., Ring, J., Rognes, M. E., and Wells, G. N. (2015). The FEniCS Project Version 1.5. *Archive of Numerical Software*, 3. 22

Anto, A. D., Fleishel, R., TerMaath, S., and Abedi, R. (2024). Size dependency of elastic and plastic properties of metallic polycrystals using statistical volume elements. *Applied Sciences*, 14(18):8207. xii, 52, 55, 65, 70, 167

Anto, A. D., TerMaath, S., and Abedi, R. (2026). Size dependence of homogenized monotonic and cyclic responses and extreme-value fatigue indicators in polycrystalline microstructures. *Mechanics of Materials*, page 105698. 82, 85, 87, 93, 95

Asaro, R. J. (1983). Micromechanics of crystals and polycrystals. *Advances in Applied Mechanics*, 23:1–115. 81, 85

Asgharzadeh, A., Park, T., Nazari Tiji, S., and Pourboghrat, F. (2023). A computationally efficient multiscale, multi-phase modeling approach based on cpfem to assess the effect of second phase particles on mechanical properties. *Crystals*, 13(8):1199. 52

Ashraf, F. and Castelluccio, G. M. (2023). History and temperature dependent cyclic crystal plasticity model with material-invariant parameters. *International Journal of Plasticity*, 161:103494. 81

Ashraf, F., Cini, A., and Castelluccio, G. M. (2022). Analytical fatigue life formulation for notches informed by crystal plasticity. *International Journal of Fatigue*, 163:107072. 82

Bahmani, B., Yang, M., Nagarajan, A., Clarke, P. L., Soghrati, S., and Abedi, R. (2019). Automated homogenization-based fracture analysis: Effects of sve size and boundary condition. *Computer Methods in Applied Mechanics and Engineering*, 345:701–727. 74

Bargmann, S., Klusemann, B., Markmann, J., Schnabel, J. E., Schneider, K., Soyarslan, C., and Wilmers, J. (2018). Generation of 3d representative volume elements for heterogeneous materials: A review. *Progress in Materials Science*, 96:322–384. 81

Basquin, O. (1910). The exponential law of endurance tests. In *Proceedings of the American Society for Testing and Materials*, volume 10, pages 625–630. 81

Battaglia, P. W., Hamrick, J. B., Bapst, V., Sanchez-Gonzalez, A., Zambaldi, V., Malinowski, M., Tacchetti, A., Raposo, D., Santoro, A., Faulkner, R., et al. (2018). Relational inductive biases, deep learning, and graph networks. *arXiv:1806.01261*. 83

Baxter, S. C. and Graham, L. L. (2000). Characterization of random composites using moving-window technique. *Journal of Engineering Mechanics*, 126(4):389–397. 3, 51

Berryman, J. G. (1980). Long-wavelength propagation in composite elastic media I. Spherical inclusions. *The Journal of the Acoustical Society of America*, 68(6):1809–1819. 8, 152

Böhm, H. J. (2004). A short introduction to continuum micromechanics. In *Mechanics of microstructured materials*, pages 1–40. Springer. 8

Bradski, G. (2000). The OpenCV Library. *Dr. Dobb's Journal of Software Tools*. 25

Bronstein, M. M., Bruna, J., LeCun, Y., Szlam, A., and Vandergheynst, P. (2017). Geometric deep learning: going beyond euclidean data. *IEEE Signal Processing Magazine*, 34(4):18–42. 5, 53, 83

Cabello-Solorzano, K., Ortigosa de Araujo, I., Peña, M., Correia, L., and J. Tallón-Ballesteros, A. (2023). The impact of data normalization on the accuracy of machine learning algorithms: a comparative analysis. In *International conference on soft computing models in industrial and environmental applications*, pages 344–353. Springer. 6, 10

Caliskan, E., Abedi, R., and Pasini, M. L. (2025). Graph neural networks for mechanical property prediction of 2d fiber composites. *Materials & Design*, page 114500. 53

Castelluccio, G. M. and McDowell, D. L. (2012). Assessment of small fatigue crack growth driving forces in single crystals with and without slip bands. *International Journal of Fracture*, 176(1):49–64. 82, 86

Castelluccio, G. M. and McDowell, D. L. (2015). Microstructure and mesh sensitivities of mesoscale surrogate driving force measures for transgranular fatigue cracks in polycrystals. *Materials Science and Engineering: A*, 639:626–639. 82

Castelluccio, G. M. and McDowell, D. L. (2016). Microstructure-sensitive small fatigue crack growth assessment: Effect of strain ratio, multiaxial strain state, and geometric discontinuities. *International Journal of Fatigue*, 82:521–529. 81

Cerrone, A., Stein, C., Pokharel, R., Hefferan, C., Lind, J., Tucker, H., Suter, R., Rollett, A., and Ingraffea, A. (2015). Implementation and verification of a microstructure-based capability for modeling microcrack nucleation in lshr at room temperature. *Modelling and Simulation in Materials Science and Engineering*, 23(3):035006. 82

Chung, D. and Buessem, W. (1967). The elastic anisotropy of crystals. *Journal of Applied Physics*, 38(5):2010–2012. 155

Chung, J., Ahmad, R., Sun, W., Cai, W., and Mukerji, T. (2024). Prediction of effective elastic moduli of rocks using Graph Neural Networks. *Computer Methods in Applied Mechanics and Engineering*, 421:116780. 5, 6, 10, 24, 28

Coffin Jr, L. F. (1954). A study of the effects of cyclic thermal stresses on a ductile metal. *Transactions of the American Society of Mechanical engineers*, 76(6):931–949. 81

Coles, S., Bawa, J., Trenner, L., and Dorazio, P. (2001). *An introduction to statistical modeling of extreme values*, volume 208. Springer. 82, 93

Corso, G., Cavalleri, L., Beaini, D., Liò, P., and Veličković, P. (2020). Principal neighbourhood aggregation for graph nets. *Advances in neural information processing systems*, 33:13260–13271. 31, 32, 54

Dai, M., Demirel, M. F., Liang, Y., and Hu, J.-M. (2021). Graph neural networks for an accurate and interpretable prediction of the properties of polycrystalline materials. *npj Computational Materials*, 7(1):103. 6, 54

Dai, M., Demirel, M. F., Liu, X., Liang, Y., and Hu, J.-M. (2023). Graph neural network for predicting the effective properties of polycrystalline materials: A comprehensive analysis. *Computational Materials Science*, 230:112461. 54

Eidel, B. (2023). Deep CNNs as universal predictors of elasticity tensors in homogenization. *Computer Methods in Applied Mechanics and Engineering*, 403:115741. 4

Embrechts, P., Klüppelberg, C., and Mikosch, T. (2013). *Modelling extremal events: for insurance and finance*, volume 33. Springer Science & Business Media. 82, 91, 95

Fang, Y., Wang, Y., Imtiaz, H., Liu, B., and Gao, H. (2019). Energy-ratio-based measure of elastic anisotropy. *Physical review letters*, 122(4):045502. 155

Fatemi, A. and Kurath, P. (1988). Multiaxial fatigue life predictions under the influence of mean-stresses. *Journal of Engineering Materials and Technology*, 110(4):380–388. 81

Fatemi, A. and Socie, D. F. (1988). A critical plane approach to multiaxial fatigue damage including out-of-phase loading. *Fatigue & Fracture of Engineering materials & structures*, 11(3):149–165. 81, 86

Feng, H., Subramaniyan, S. P., Tewani, H., and Prabhakar, P. (2024). Physics-Constrained Neural Network for design and feature-based optimization of weave architectures. *Composites Part A: Applied Science and Manufacturing*, 187:108465. 4

Fey, M. and Lenssen, J. E. (2019). Fast graph representation learning with pytorch geometric. *arXiv preprint arXiv:1903.02428*. 26, 54

Fisher, R. A. and Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. In *Mathematical*

proceedings of the Cambridge philosophical society, volume 24, pages 180–190. Cambridge University Press. 82, 93

Frankel, A. L., Jones, R. E., Alleman, C., and Templeton, J. A. (2019). Predicting the mechanical response of oligocrystals with deep learning. *Computational Materials Science*, 169:109099. 83

Frankel, A. L., Jones, R. E., and Swiler, L. P. (2020). Tensor basis gaussian process models of hyperelastic materials. *Journal of Machine Learning for Modeling and Computing*, 1(1). 83

Fuhg, J. N., van Wees, L., Obstalecki, M., Shade, P., Bouklas, N., and Kasemer, M. (2022). Machine-learning convex and texture-dependent macroscopic yield from crystal plasticity simulations. *Materialia*, 23:101446. 52

Gal, Y. and Ghahramani, Z. (2016). Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *International Conference on Machine Learning*, pages 1050–1059. PMLR. 83

Gal, Y., Hron, J., and Kendall, A. (2017). Concrete dropout. *Advances in Neural Information Processing Systems*, 30. 83

Gao, Y., Berger, M., and Duddu, R. (2023). CNN-Based Surrogate for the Phase Field Damage Model: Generalization across Microstructure Parameters for Composite Materials. *Journal of Engineering Mechanics*, 149(6):04023025. Publisher: American Society of Civil Engineers. xvi, 5, 20, 21, 39, 41, 42, 47

Garrard, J. M. and Abedi, R. (2020a). Statistical volume element averaging scheme for fracture of quasi-brittle materials. *Computers and Geotechnics*, 117:103229. 3, 52

Garrard, J. M. and Abedi, R. (2020b). Statistical volume elements for the characterization of angle-dependent fracture strengths in anisotropic microcracked

materials. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B: Mechanical Engineering*, 6(2):021008. 3, 52

Ghosh, S., Bhandari, Y., and Groeber, M. (2008). Cad-based reconstruction of 3d polycrystalline alloy microstructures from fib generated serial sections. *Computer-Aided Design*, 40(3):293–310. 81

Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O., and Dahl, G. E. (2017). Neural Message Passing for Quantum Chemistry. In *Proceedings of the 34th International Conference on Machine Learning*, pages 1263–1272. PMLR. ISSN: 2640-3498. 32, 83

Gitman, I., Askes, H., and Sluys, L. (2007). Representative volume: Existence and size determination. *Engineering fracture mechanics*, 74(16):2518–2534. 51

Gomez-Flores, A., Cho, H., Hong, G., Nam, H., Kim, H., and Chung, Y. (2024). A critical review on machine learning applications in fiber composites and nanocomposites: Towards a control loop in the chain of processes in industries. *Materials & Design*, 245:113247. 4

Graham, L. and Baxter, S. (2001). Simulation of local material properties based on moving-window gmc. *Probabilistic Engineering Mechanics*, 16(4):295–305. 3, 52

Graham-Brady, L., Siragy, E., and Baxter, S. (2003). Analysis of heterogeneous composites based on moving-window techniques. *Journal of engineering mechanics*, 129(9):1054–1064. 3, 52

Groeber, M. A. and Jackson, M. A. (2014). Dream. 3d: a digital representation environment for the analysis of microstructure in 3d. *Integrating materials and manufacturing innovation*, 3(1):56–72. 55, 85

Gu, T., Stopka, K. S., Xu, C., and McDowell, D. L. (2020). Prediction of maximum fatigue indicator parameters for duplex ti–6al–4v using extreme value theory. *Acta Materialia*, 188:504–516. 82, 93

Gumbel, E. J. (1958). *Statistics of Extremes*. Columbia University Press. 82

Hall, P. (1979). On the rate of convergence of normal extremes. *Journal of Applied Probability*, 16(2):433–439. 82, 95

Hamilton, W., Ying, Z., and Leskovec, J. (2017). Inductive representation learning on large graphs. *Advances in Neural Information Processing Systems*, 30. 83, 103

Hamilton, W. L. (2020). Graph representation learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 14(3):1–159. 28, 30

Hansen, C. K., Whelan, G. F., and Hochhalter, J. D. (2024). Interpretable machine learning for microstructure-dependent models of fatigue indicator parameters. *International Journal of Fatigue*, 178:108019. 83

Hashin, Z. and Shtrikman, S. (1963). A variational approach to the theory of the elastic behaviour of multiphase materials. *Journal of the Mechanics and Physics of Solids*, 11(2):127–140. 152

Hazanov, S. and Amieur, M. (1995). On overall properties of elastic heterogeneous bodies smaller than the representative volume. *International Journal of Engineering Science*, 33(9):1289–1301. 9, 69

Hazanov, S. and Huet, C. (1994). Order relationships for boundary conditions effect in heterogeneous bodies smaller than the representative volume. *Journal of the Mechanics and Physics of Solids*, 42(12):1995–2011. 9, 11, 69

Haşegan, D., Geniesse, C., Chowdhury, S., and Saggar, M. (2023). Deconstructing the Mapper algorithm to extract richer topological and temporal features from functional neuroimaging data. *bioRxiv*, page 2023.10.13.562304. 24

He, Y., Heider, Y., and Markert, B. (2024). Embedding an ann-based crystal plasticity model into the finite element framework using an abaqus user-material subroutine. *arXiv preprint arXiv:2410.08214*. 53

Heidenreich, J. N., Gorji, M. B., and Mohr, D. (2023). Modeling structure-property relationships with convolutional neural networks: Yield surface prediction based on microstructure images. *International Journal of Plasticity*, 163:103506. 53

Hennessey, C., Castelluccio, G. M., and McDowell, D. L. (2017). Sensitivity of polycrystal plasticity to slip system kinematic hardening laws for al 7075-t6. *Materials Science and Engineering: A*, 687:241–248. 82

Hill, R. (1963). Elastic properties of reinforced solids: some theoretical principles. *Journal of the Mechanics and Physics of Solids*, 11(5):357–372. Publisher: Elsevier. 3

Hill, R. (1965). A self-consistent mechanics of composite materials. *Journal of the Mechanics and Physics of Solids*, 13(4):213–222. 8, 152

Hill, R. (1967). The essential structure of constitutive laws for metal composites and polycrystals. *Journal of the Mechanics and Physics of Solids*, 15(2):79–95. 9

Hill, R. (1972). On constitutive macro-variables for heterogeneous solids at finite strain. *Proceedings of the Royal Society A: Mathematical, Physical, and Engineering SCiences*, 326:131–147. 3

Hochhalter, J., Littlewood, D., Christ Jr, R., Veilleux, M., Bozek, J., Ingraffea, A., and Maniatty, A. (2010). A geometric approach to modeling microstructurally small fatigue crack formation: Ii. physically based modeling of microstructure-dependent slip localization and actuation of the crack nucleation mechanism in aa 7075-t651. *Modelling and Simulation in Materials Science and Engineering*, 18(4):045004. 82

Hu, Y., Zhou, G., Lee, M.-G., Wu, P., and Li, D. (2024). A temporal graph neural network for cross-scale modelling of polycrystals considering microstructure interaction. *International Journal of Plasticity*, 179:104017. 83

Huet, C. (1990). Application of variational concepts to size effects in elastic heterogeneous bodies. *Journal of the Mechanics and Physics of Solids*, 38(6):813–841. 8, 11, 69

Huyse, L. and Maes, M. A. (2001). Random field modeling of elastic properties using homogenization. *Journal of Engineering Mechanics*, 127(1):27–36. 3, 51

Kalidindi, S. R., Bronkhorst, C. A., and Anand, L. (1992). Crystallographic texture evolution in bulk deformation processing of fcc metals. *Journal of the Mechanics and Physics of Solids*, 40(3):537–569. 81, 85

Kanit, T., Forest, S., Galliet, I., Mounoury, V., and Jeulin, D. (2003). Determination of the size of the representative volume element for random composites: statistical and numerical approach. *International Journal of solids and structures*, 40(13-14):3647–3679. 17, 51

Kendall, A. and Gal, Y. (2017). What uncertainties do we need in bayesian deep learning for computer vision? *Advances in Neural Information Processing Systems*, 30. 83

Khorrami, M. S., Goyal, P. K., Mianroodi, J. R., Svendsen, B., Benner, P., and Raabe, D. (2024). Divergence-free neural operators for stress field modeling in polycrystalline materials. *arXiv preprint arXiv:2408.15408*. 53

Kingma, D. P., Salimans, T., and Welling, M. (2015). Variational dropout and the local reparameterization trick. *Advances in Neural Information Processing Systems*, 28. 83

Kipf, T. N. and Welling, M. (2016). Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*. 32

Knezevic, M., Drach, B., Ardeljan, M., and Beyerlein, I. J. (2014). Three dimensional predictions of grain scale plasticity and grain boundaries using crystal plasticity

finite element models. *Computer Methods in Applied Mechanics and Engineering*, 277:239–259. 81

Korsunsky, A. M., Dini, D., Dunne, F. P., and Walsh, M. J. (2007). Comparative assessment of dissipated energy and other fatigue criteria. *International Journal of Fatigue*, 29(9-11):1990–1995. 82, 86

Kramer, O. (2013). K-nearest neighbors. In *Dimensionality reduction with unsupervised nearest neighbors*, pages 13–23. Springer. 26

Kuang, X.-C., Jiang, W.-G., Mao, L.-H., Wu, Z.-K., Liu, F.-C., Zhou, X., Ling, P.-H., Zhang, Y.-C., and Yi, M. (2026). Microstructure-induced fatigue scatter of additively manufactured inconel 718: Insight from multilevel simulations and dislocation-based strain gradient crystal plasticity. *International Journal of Plasticity*, page 104632. 81

Latypov, M. I. and Kalidindi, S. R. (2017). Data-driven reduced order models for effective yield strength and partitioning of strain in multiphase materials. *Journal of Computational Physics*, 346:242–261. 53

Latypov, M. I., Toth, L. S., and Kalidindi, S. R. (2019). Materials knowledge system for nonlinear composites. *Computer Methods in Applied Mechanics and Engineering*, 346:180–196. 53

Leadbetter, M. R., Lindgren, G., and Rootzén, H. (2012). *Extremes and related properties of random sequences and processes*. Springer Science & Business Media. 82

Ledbetter, H. and Migliori, A. (2006). A general elastic-anisotropy measure. *Journal of Applied Physics*, 100(6):063516. 155

Leman, A. and Weisfeiler, B. (1968). A reduction of a graph to a canonical form and an algebra arising during this reduction. *Nauchno-Technicheskaya Informatsiya*, 2(9):12–16. 31

Li, Q., Wang, Z., Li, L., Hao, H., Chen, W., and Shao, Y. (2023). Machine learning prediction of structural dynamic responses using graph neural networks. *Computers & Structures*, 289:107188. 5

Li, R., Shao, Q., Gao, E., and Liu, Z. (2020). Elastic anisotropy measure for two-dimensional crystals. *Extreme Mechanics Letters*, 34:100615. 154

Loh, J. Y. Y., Yeoh, K. M., Raju, K., Pham, V. N. H., Tan, V. B. C., and Tay, T. E. (2024). A Review of Machine Learning for Progressive Damage Modelling of Fiber-Reinforced Composites. *Applied Composite Materials*, 31(6):1795–1832. 4

Loshchilov, I. and Hutter, F. (2019). Decoupled Weight Decay Regularization. arXiv:1711.05101 [cs]. 33

Lupo Pasini, M., Choi, J. Y., Mehta, K., Zhang, P., Weaver, R., Chowdhury, A., Li, C., Ye, Z., and Baker, J. (2025). Hydragnn v4. 0. Technical report, Oak Ridge National Laboratory (ORNL), Oak Ridge, TN (United States). 54, 77

Lupo Pasini, M., Jung, G. S., and Irle, S. (2023). Graph neural networks predict energetic and mechanical properties for models of solid solution metal alloy phases. *Computational Materials Science*, 224:112141. 31, 54

Lupo Pasini, M., Reeve, S. T., Zhang, P., and Choi, J. Y. (2021). Hydragnn. [Computer Software] https://doi.org/10.11578/dc.20211019.2. 30, 47

Mandel, J. (1966). Contribution theorique aletude de l'ecrouissage et des lois de l'ecoulement plastique. *Applied Mechanics*, pages 502–509. 3

Mangal, A. and Holm, E. A. (2018). Applied machine learning to predict stress hotspots i: Face centered cubic materials. *International Journal of Plasticity*, 111:122–134. 83

Manonukul, A. and Dunne, F. (2004). High–and low–cycle fatigue crack initiation using polycrystal plasticity. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 460(2047):1881–1903. 82, 86

Manson, S. S. (1965). Fatigue: A complex subject—some simple approximations. *Experimental Mechanics*, 5(4):193–226. 81

Mao, Y., Keshavarz, S., Kilic, M. N. T., Wang, K., Li, Y., Reid, A. C., Liao, W.-k., Choudhary, A., and Agrawal, A. (2025). A deep learning-based crystal plasticity finite element model. *Scripta Materialia*, 254:116315. 53

McDowell, D. and Dunne, F. (2010). Microstructure-sensitive computational modeling of fatigue crack formation. *International Journal of Fatigue*, 32(9):1521–1542. 81

Meyer, P. P., Bonatti, C., Tancogne-Dejean, T., and Mohr, D. (2022). Graph-based metamaterials: Deep learning of structure-property relations. *Materials & Design*, 223:111175. 5

Miehe, C. and Bayreuther, C. G. (2007). On multiscale FE analyses of heterogeneous structures: from homogenization to multigrid solvers. *International Journal for Numerical Methods in Engineering*, 71(10):1135–1180. 3

Miehe, C., Hofacker, M., and Welschinger, F. (2010). A phase field model for rate-independent crack propagation: Robust algorithmic implementation based on operator splits. *Computer Methods in Applied Mechanics and Engineering*, 199(45):2765–2778. 4

Miles, R. E. (1970). On the homogeneous planar Poisson point process. *Mathematical Biosciences*, 6:85–127. 27

Mirkhalaf, M. and Rocha, I. (2024). Micromechanics-based deep-learning for composites: Challenges and future perspectives. *European Journal of Mechanics-A/Solids*, 105:105242. 4

Mori, T. and Tanaka, K. (1973). Average stress in matrix and average elastic energy of materials with misfitting inclusions. *Acta Metallurgica*, 21(5):571–574. 8, 152

Musinski, W. D. and McDowell, D. L. (2012). Microstructure-sensitive probabilistic modeling of hcf crack initiation and early crack growth in ni-base superalloy in100 notched components. *International Journal of Fatigue*, 37:41–53. 81

Muth, A., John, R., Pilchak, A., Kalidindi, S. R., and McDowell, D. L. (2021). Analysis of fatigue indicator parameters for ti-6al-4v microstructures using extreme value statistics in the hcf regime. *International Journal of Fatigue*, 145:106096. 82

Nemat-Nasser, S. and Hori, M. (2013). *Micromechanics: overall properties of heterogeneous materials*, volume 37. Elsevier, Amsterdam, The Netherlands. 51

Nix, D. A. and Weigend, A. S. (1994). Estimating the mean and variance of the target probability distribution. In *Proceedings of 1994 IEEE International Conference on Neural Networks (ICNN'94)*, volume 1, pages 55–60. IEEE. 83

Nye, J. (1985). *Physical properties of crystals: their representation by tensors and matrices*. Oxford university press. 154

Ogden, R. W. (1974). On the overall moduli of non-linear elastic composite materials. *Journal of the Mechanics and Physics of Solids*, 22(6):541–553. 3

Olivier, A., Shields, M. D., and Graham-Brady, L. (2021). Bayesian neural networks for uncertainty quantification in data-driven materials modeling. *Computer Methods in Applied Mechanics and Engineering*, 386:114079. 83

Orifici, A. C., Herszberg, I., and Thomson, R. S. (2008). Review of methodologies for composite material modelling incorporating failure. *Composite Structures*, 86(1):194–210. 4

Ostoja-Starzewski, M. (1998). Random field models of heterogeneous materials. *International Journal of Solids and Structures*, 35(19):2429–2455. 3, 8, 11, 51, 58

Ostoja-Starzewski, M. (2006). Material spatial randomness: From statistical to representative volume element. *Probabilistic engineering mechanics*, 21(2):112–132. 3, 51

Ostoja-Starzewski, M., Du, X., Khisaeva, Z., and Li, W. (2007). Comparisons of the size of the representative volume element in elastic, plastic, thermoelastic, and permeable random microstructures. *International Journal for Multiscale Computational Engineering*, 5(2). 3, 51

Park, C. W. and Wolverton, C. (2020). Developing an improved crystal graph convolutional neural network framework for accelerated materials discovery. *Physical Review Materials*, 4(6):063801. 83

Pasparakis, G. D., Graham-Brady, L., and Shields, M. D. (2025). Bayesian neural networks for predicting uncertainty in full-field material response. *Computer Methods in Applied Mechanics and Engineering*, 433:117486. 53

Pathan, M. V., Ponnusami, S. A., Pathan, J., Pitisongsawat, R., Erice, B., Petrinic, N., and Tagarielli, V. L. (2019). Predictions of the mechanical properties of unidirectional fibre composites by supervised machine learning. *Scientific Reports*, 9(1):13964. 4

Pélissou, C., Baccou, J., Monerie, Y., and Perales, F. (2009). Determination of the size of the representative volume element for random quasi-brittle composites. *International Journal of Solids and Structures*, 46(14-15):2842–2855. 51

Przybyla, C. P. and McDowell, D. L. (2010). Microstructure-sensitive extreme value probabilities for high cycle fatigue of ni-base superalloy in100. *International Journal of Plasticity*, 26(3):372–394. 81

Przybyla, C. P., Musinski, W. D., Castelluccio, G. M., and McDowell, D. L. (2013). Microstructure-sensitive hcf and vhcf simulations. *International Journal of Fatigue*, 57:9–27. 81

Qin, Y., DeWitt, S., Radhakrishnan, B., and Biros, G. (2024). Graingnn: A dynamic graph neural network for predicting 3d grain microstructure. *Journal of Computational Physics*, 510:113061. 54

Ranganathan, S. and Ostoja-Starzewski, M. (2008). Universal elastic anisotropy index. *Physical Review Letters*, 101(5):055504. 154, 156

Reiser, P., Neubert, M., Eberhard, A., Torresi, L., Zhou, C., Shao, C., Metni, H., van Hoesel, C., Schopmans, H., Sommer, T., et al. (2022). Graph neural networks for materials science and chemistry. *Communications Materials*, 3(1):93. 83

Robertson, A. E., Generale, A. P., Kelly, C., Buzzy, M. O., and Kalidindi, S. R. (2024). MICRO2D: A Large, Statistically Diverse, Heterogeneous Microstructure Dataset. *Integrating Materials and Manufacturing Innovation*, 13(1):120–154. 12, 20, 21, 47

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer. 4

Roters, F., Eisenlohr, P., Hantcherli, L., Tjahjanto, D. D., Bieler, T. R., and Raabe, D. (2010). Overview of constitutive laws, kinematics, homogenization and multiscale methods in crystal plasticity finite-element modeling: Theory, experiments, applications. *Acta Materialia*, 58(4):1152–1211. 81, 85

Rychlewski, J. and Zhang, J. (1989). Anisotropy degree of elastic materials. *Archives of Mechanics*, 41(5):697–715. 155

Sab, K. (1992). On the homogenization and the simulation of random materials. *Eur. J. Mech,. A/Solids*, 11(5):585–607. 51

Salmi, M., Auslender, F., Bornert, M., and Fogli, M. (2012). Apparent and effective mechanical properties of linear matrix-inclusion random composites: Improved bounds for the effective behavior. *International Journal of Solids and Structures*, 49(10):1195–1211. 3, 51

Sanchez-Lengeling, B., Reif, E., Pearce, A., and Wiltschko, A. B. (2021). A gentle introduction to graph neural networks. *Distill*, 6(9):e33. 53

Sangid, M. D. (2013). The physics of fatigue crack initiation. *International Journal of Fatigue*, 57:58–72. 82

Segurado, J., Lebensohn, R. A., and LLorca, J. (2018). Computational homogenization of polycrystals. *Advances in Applied Mechanics*, 51:1–114. 81

Seitzer, M., Tavakoli, A., Antic, D., and Martius, G. (2022). On the pitfalls of heteroscedastic uncertainty estimation with probabilistic neural networks. *arXiv:2203.09168*. 83

Sim, G.-J., Lee, M.-G., and Latypov, M. I. (2025). Fip-gnn: Graph neural networks for scalable prediction of grain-level fatigue indicator parameters. *Scripta Materialia*, 255:116407. 54, 83

Singh, A. and Pal, S. (2021). Multi-phase field modeling for various fracture mechanisms in composites. *Engineering Fracture Mechanics*, 241:107348. 4

Singh, D. and Singh, B. (2020). Investigating the impact of data normalization on classification performance. *Applied Soft Computing*, 97:105524. 6, 10

Singh, G., Memoli, F., and Carlsson, G. (2007). *Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition*. The Eurographics Association. ISSN: 1811-7813. 5, 24

Smith, S. L., Kindermans, P.-J., Ying, C., and Le, Q. V. (2017). Don't decay the learning rate, increase the batch size. *arXiv preprint arXiv:1711.00489*. 5

Song, K., Wang, K., Zhang, L., Zhao, L., Xu, L., Han, Y., and Hao, K. (2022). Insights on low cycle fatigue crack formation and propagation mechanism: A microstructurally-sensitive modeling. *International Journal of Plasticity*, 154:103295. 82

Stapleton, S., Appel, L., Simon, J.-W., and Reese, S. (2016). Representative volume element for parallel fiber bundles: Model and size convergence. *Composites Part A: Applied Science and Manufacturing*, 87:170–185. 3, 51

Stopka, K. S., Yaghoobi, M., Allison, J. E., and McDowell, D. L. (2023). Microstructure-sensitive modeling of surface roughness and notch effects on extreme value fatigue response. *International Journal of Fatigue*, 166:107295. 82

Suresh, S. (1998). *Fatigue of materials*. Cambridge University Press. 81

Tannous, M., Ghnatios, C., Castelnau, O., Castañeda, P. P., and Chinesta, F. (2025). Machine learning-boosted nonlinear homogenization. *Mechanics of Materials*, 201:105229. 53

Tao, Y. and Ge, S. (2025). A distribution-guided mapper algorithm. *BMC bioinformatics*, 26(1):73. 24

Tasdemir, B., Pellegrino, A., Su, X., and Tagarielli, V. L. (2025). Learning the non-proportional multiaxial elastic–plastic response of an aluminium alloy with neural networks. *Materials & Design*, 253:113956. 4

Torquato, S. (1991). Random heterogeneous media: Microstructure and improved bounds on effective properties. *Applied Mechanics Reviews*, 44(2):37–76. 15, 152

Torquato, S. (1997). Effective stiffness tensor of composite media—I. exact series expansions. *Journal of the Mechanics and Physics of Solids*, 45(9):1421–1448. 9

Torquato, S. (1998). Effective stiffness tensor of composite media: II. applications to isotropic dispersions. *Journal of the Mechanics and Physics of Solids*, 46(8):1411–1440. 9

Tran, K., Neiswanger, W., Yoon, J., Zhang, Q., Xing, E., and Ulissi, Z. W. (2020). Methods for comparing uncertainty quantifications for material property predictions. *Machine Learning: Science and Technology*, 1(2):025006. 83

Tregger, N., Corr, D., Graham-Brady, L., and Shah, S. (2006). Modeling the effect of mesoscale randomness on concrete fracture. *Probabilistic Engineering Mechanics*, 21(3):217–225. 3, 51

Walpole, L. (1966). On bounds for the overall elastic moduli of inhomogeneous systems—i. *Journal of the Mechanics and Physics of Solids*, 14(3):151–162. 8

Walpole, L. J. (1969). On the overall elastic moduli of composite materials. *Journal of the Mechanics and Physics of Solids*, 17(4):235–251. 150

Wan, V., MacLachlan, D., and Dunne, F. (2014). A stored energy criterion for fatigue crack nucleation in polycrystals. *International Journal of Fatigue*, 68:90–102. 82

Wang, H., Tan, P. F. M., Foo, Y. Z., Owyong, S., Lu, Z., Chen, H. T., Nguyen, T. H., and Lin, A. (2025). Integrating graph neural networks with physics-informed loss function for mechanical response prediction of hollow concrete structures with morphed honeycomb configurations. *Materials & Design*, 251:113659. 5

Wang, K., Song, K., Xin, R., Zhao, L., and Xu, L. (2023). Cyclic microstructure analysis, crack propagation and life prediction of inconel 750h considering the slip fracture energy. *International Journal of Plasticity*, 167:103660. 82

Wang, L.-F. and Zheng, Q.-S. (2007). Extreme anisotropy of graphite and single-walled carbon nanotube bundles. *Applied physics letters*, 90(15):153113. 154

Willis, J. R. (1981). Variational and related methods for the overall properties of composites. *Advances in applied mechanics*, 21:1–78. 8

Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., and Yu, P. S. (2020). A comprehensive survey on graph neural networks. *IEEE transactions on neural networks and learning systems*, 32(1):4–24. 53

Xie, T. and Grossman, J. C. (2018a). Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Physical Review Letters*, 120(14):145301. 83

Xie, T. and Grossman, J. C. (2018b). Hierarchical visualization of materials space with graph convolutional neural networks. *The Journal of chemical physics*, 149(17). 83

Xu, K., Hu, W., Leskovec, J., and Jegelka, S. (2019). How Powerful are Graph Neural Networks? arXiv:1810.00826 [cs, stat]. 31, 33

Xu, K., Li, C., Tian, Y., Sonobe, T., Kawarabayashi, K.-i., and Jegelka, S. (2018). Representation Learning on Graphs with Jumping Knowledge Networks. arXiv:1806.03536 [cs]. 31

Yaghoobi, M., Stopka, K. S., Lakshmanan, A., Sundararaghavan, V., Allison, J. E., and McDowell, D. L. (2021). Prisms-fatigue computational framework for fatigue analysis in polycrystalline metals and alloys. *npj Computational Materials*, 7(1):38. 82

Yaghoobi, M., Stopka, K. S., McDowell, D. L., Graham-Brady, L., and Teferra, K. (2023). Effect of sample size on the maximum value distribution of fatigue driving forces in metals and alloys. *International Journal of Fatigue*, 176:107853. 82, 83, 93, 116

Yang, M., Garrard, J., Abedi, R., and Soghrati, S. (2021). Effect of microstructural variations on the failure response of a nano-enhanced polymer: a homogenization-based statistical analysis. *Computational Mechanics*, 67(1):315–340. 3

Yuan, M., Paradiso, S., Meredig, B., and Niezgoda, S. R. (2018). Machine learning–based reduce order crystal plasticity modeling for icme applications. *Integrating Materials and Manufacturing Innovation*, 7(4):214–230. 82

Zhang, T., Jiang, J., Britton, B., Shollock, B., and Dunne, F. (2016). Crack nucleation using combined crystal plasticity modelling, high-resolution digital image correlation and high-resolution electron backscatter diffraction in a superalloy containing non-metallic inclusions under fatigue. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 472(2189). 81

Zhao, Y., Li, H., Zhou, H., Attar, H. R., Pfaff, T., and Li, N. (2024). A review of graph neural network applications in mechanics-related domains. *Artificial Intelligence Review*, 57(11):315. 6, 54

Zhou, Y. and Hubert, P. (2024). Characterization and property prediction of fibre structures within discontinuous-fibre reinforced polymer matrix composites using 3D fibre cells assisted by contrastive learning. *Composites Part A: Applied Science and Manufacturing*, 187:108506. 4

Appendices

Appendix A

Appendix: Chapter 1

A.1 Mean-Field Method (MFM)

A general form of Hashin–Shtrikman (HS) formula for the stiffness tensor of a composite is Walpole (1969),

$$\mathbf{C}^* = \left\langle \mathbf{C}\mathbf{A}^0_{\text{dil}} \right\rangle \left\langle \mathbf{A}^0_{\text{dil}} \right\rangle^{-1}, \tag{A.1}$$

where $\mathbf{C}^*$ is the RVE-limit homogenized stiffness tensor, $\langle f \rangle = 1/|V| \int_V f(\mathbf{x})\ \mathrm{d}\mathbf{x}$ is the volumetric average of a quantity f over the VE V with volume $|V|$, $\mathbf{C}(\mathbf{x})$ is the position-dependent stiffness tensor, *cf.* SM Eq. 1, and $\mathbf{A}^0_{\text{dil}}$ is the dilute approximation strain concentration tensor relative to a reference material 0 with stiffness $\mathbf{C}^0$. The role of the reference material will be further discussed below. For an n-phase material with volume fractions v^i, $i \in \{1, \cdots, n\}$, $\langle \mathbf{C}\mathbf{A}^0_{\text{dil}} \rangle = \sum_{i=1}^n v^i \mathbf{C}^i \mathbf{A}^{i,0}_{\text{dil}}$ and $\langle \mathbf{A}^0_{\text{dil}} \rangle = \sum_{i=1}^n v^i \mathbf{C}^i \mathbf{A}^{i,0}_{\text{dil}}$. When the phase i is formed or approximated by elliptical (or ellipsoidal in 3D) inclusions, the dilute average to phase i strain map $\mathbf{A}^{i,0}_{\text{dil}}$ (corresponding to material 0 as the matrix) can be analytically computed using the Eshelby tensor.

To simplify the formulas for the RVE-limit isotropic two-phase composites in this paper, the pair $\mathbf{p} := [p_1, p_2] = [\kappa, \mu]$ is used to contain the in-plane bulk modulus κ

and shear modulus μ of a material. The superscripts $I, M, *$ and 0 refer to quantities associated with inclusion, matrix, homogenized RVE, and the reference material, respectively. The subscripts $j = 1$ and $j = 2$ refer to bulk and shear-related quantities. For a given 2D isotropic material, the stiffness tensor $\mathbf{C}$ and the bulk and shear moduli are related by,

$$\mathbf{C} = \begin{bmatrix} \kappa + \mu & \kappa - \mu & 0 \\ \kappa - \mu & \kappa + \mu & 0 \\ 0 & 0 & \mu \end{bmatrix} \quad \Leftrightarrow \quad \begin{cases} \kappa = C_{11} - C_{33} \\ \mu = C_{33} \end{cases} \tag{A.2}$$

The process for obtaining homogenized $\mathbf{C}^*$ for the two-phase composite considered is discussed next.

First, if not already provided for M and I phases, compute their bulk and shear moduli. This can be from the isotropic $\mathbf{C}$ in Eq. (A.2). Alternatively, if elastic modulus E and Poisson ratio ν are provided $\mu = E/2(1-\nu)$. Moreover, $\kappa = E/2(1-2\nu)$ and $\kappa = E/2(1-\nu)$ for plane-strain and plane-stress modes, respectively.

Second, given the choice of reference material (discussed next), compute the homogenized properties,

$$p_j^* = p_j^M \left[1 + \frac{v^I \left(\frac{p_j^I}{p_j^M} - 1 \right) \left(1 + s_j^0 \left(\frac{p_j^M}{p_j^0} - 1 \right) \right)}{v^M \left(1 + s_j^0 \left(\frac{p_j^I}{p_j^0} - 1 \right) \right) + v^I \left(1 + s_j^0 \left(\frac{p_j^M}{p_j^0} - 1 \right) \right)} \right], \tag{A.3}$$

for $j = 1$ (bulk modulus κ^*) and $j = 2$ (shear modulus μ^*). In 2D, the bulk and shear isotropic Eshelby tensor components for a circle in the reference material 0 are given by,

$$s_1^0 := s_\kappa^0 = \frac{p_1^0}{p_1^0 + p_2^0} = \frac{\kappa^0}{\kappa^0 + \mu^0} \tag{A.4a}$$

$$s_2^0 := s_\mu^0 = \frac{p_1^0 + 2p_2^0}{2(p_1^0 + p_2^0)} = \frac{\kappa^0 + 2\mu^0}{2(\kappa^0 + \mu^0)} \tag{A.4b}$$

For all choices of the MFM, except the DEM method, the homogenized bulk and shear moduli are computed from Eq. (A.3). For the DEM method, a differential form of Eq. (A.3) will be used to determine the homogenized solutions, as discussed later.

Third, having κ^* and μ^*, the homogenized stiffness tensor $\mathbf{C}^*$ is obtained from Eq. (A.2).

Finally, different MFM bounds and estimates are obtained by different choices of the reference material,

- Voigt and Reuss bounds are obtained by letting $\mathbf{p}^0 = [\infty, \infty]$ and $[0, 0]$, respectively. They result in the upper Voigt $p_j^* = v^M p_j^M + v^I p_j^I$ and lower Reuss $1/p_j^* = v^M/p_j^M + v^I/p_j^I$ bounds, respectively.

- Hashin—Shtrikman (HS) Hashin and Shtrikman (1963) upper (HS+) and lower (HS-) bounds correspond to $\mathbf{p}^0 = [\max(\kappa^M, \kappa^I), \max(\mu^M, \mu^I)]$ and $\mathbf{p}^0 = [\min(\kappa^M, \kappa^I), \min(\mu^M, \mu^I)]$, respectively. These bounds are tighter than Voigt-Reuss bounds.

- Mori-Tanaka (MT) Mori and Tanaka (1973) estimate corresponds to $\mathbf{p}^0 = [\kappa^M, \mu^M]$. If the material is well-ordered (that is, one material has the highest bulk and shear moduli), the MT estimate corresponds to HS- or HS+ when the inclusion or matrix is the stiffer phase. All composites considered in this paper are well-ordered.

- Self-consistent (SC) estimate Hill (1965) corresponds to $\mathbf{p}^0 = [\kappa^*, \mu^*]$. The resulting nonlinear equation Eq. (A.4) is often solved by the fixed-point iterative method. One problem with the classical SC method is its breakdown for porous media ($\kappa^I = 0, \mu^I = 0$) with circular voids for $v^I \geq 1/3$ (1/2 limit for spherical voids in 3D) due to the percolation of voids Torquato (1991).

- Differential Effective Medium (DEM) Berryman (1980): In this method, Eq. (A.1) is solved in the differential form, without the term $\langle \mathbf{A}_{\text{dil}}^0 \rangle^{-1}$. The

instantaneous volume fraction of inclusions c^I, increases from 0 to its terminal value v^I. Throughout the process, the reference material 0 corresponds to the current homogenized material *. This results in the following *Ordinary Differential Equation* (ODE),

$$\frac{\mathrm{d}\mathbf{C}^*(c^I)}{\mathrm{d}c^I} = \frac{1}{1-c^I}\left(\mathbf{C}^I - \mathbf{C}^*(c^I)\right)\mathbf{A}^{I,*}_{\mathrm{dil}} \tag{A.5}$$

integrated for $c^I = 0$ to v^I with the initial condition $\mathbf{C}^*(c^I = 0) = \mathbf{C}^M$. For circular inclusions and isotropic phases, following a similar process as above, the ODE for $\mathbf{p}^*$ is,

$$\frac{\mathrm{d}p_j^*(c^I)}{\mathrm{d}c^I} = \frac{1}{1-c^I}\frac{p_j^I - p_j^*(c^I)}{1 + s_j^*(c^I)\left(\frac{p_j^I}{p_j^*(c^I)} - 1\right)}, \tag{A.6}$$

for $c^I = 0$ to v^I and initial condition $p_j^*(c^I = 0) = p_j^M$ for $j = 1$ and 2. s_j^* is computed from Eq. (A.4) using the instantaneous homogenized material $(*)$ as the reference (0) material.

A.2 Anisotropy measures

An isotropic material can be represented by only two elastic constants. For example, for an isotropic material, everything is presented in terms of longitudinal C_{11} and shear $\mu = C_{33}$ moduli: $C_{22} = C_{11}$, $C_{12} = C_{21} = C_{11} - 2C_{33}$. Alternatively, an isotropic elasticity response can be represented in terms of the in-plane bulk modulus $\kappa = C_{11} - C_{33}$ and shear modulus μ. Finally, either of the pairs above can be represented in terms of the elastic modulus E and Poisson ratio ν, as described in Section A.1.

One can quantify whether the VEs of a given size have reached such homogeneous and isotropic limits. Independent of the isotropy of $\mathbf{C}$, the in-plane bulk modulus is

defined as Abedi et al. (2023).

$$\kappa = \frac{1}{D_{11} + D_{22} + 2D_{12}}. \tag{A.7}$$

Let us consider a population of VEs of the same size, composed of the same material constituents. We can represent the homogeneity of $\mathbf{C}$ matrices, through the *Coefficient of Variation* (CoV) of the bulk modulus,

$$c_\kappa = \frac{\varsigma_\kappa}{\bar{\kappa}}, \tag{A.8}$$

where ς_κ and $\bar{\kappa}$ are the nominal mean and standard deviation of the bulk moduli of the VEs.

The anisotropy of the stiffness tensor can be measured using a different index. One useful concept is the derivation of the isotropic stiffness tensor $\mathbf{C}^{\text{V,Iso}}$ from $\mathbf{C}$. Let $\mathbf{C}(\theta)$ denote the stiffness of the material rotated by angle θ in (x_1, x_2) plane. $\mathbf{C}^{\text{V,Iso}}$ is defined by Voigt averaging of $\mathbf{C}(\theta)$ over angle θ Ranganathan and Ostoja-Starzewski (2008),

$$\mathbf{C}^{\text{V,Iso}} = \frac{1}{2\pi}\int_0^{2\pi} \mathbf{C}(\theta)\mathrm{d}\theta. \tag{A.9}$$

$\mathbf{C}^{\text{V,Iso}}$ has several uses. For example, unlike the bulk modulus, the shear modulus C_{33} is angle-dependent. Even is $\mathbf{C}$ is anisotropic, one can define the isotropic shear modulus Li et al. (2020),

$$\mu^{\text{V,Iso}} := C_{33}^{\text{V,Iso}} = \frac{1}{8}(C_{11} + C_{22}) - \frac{1}{4}C_{12} + \frac{C_{33}}{2}. \tag{A.10}$$

It is also used to define a few anisotropy indices, such as *Nye, Zheng* anisotropy index $A^{||\Delta\mathbf{C}||} = ||\Delta\mathbf{C}||/||\mathbf{C}||$ Nye (1985); Wang and Zheng (2007) and *Ranganathan, Ostoja-Starzewski* anisotropy index A_{RO} Ranganathan and Ostoja-Starzewski (2008); $||.||$ refers to an objective fourth-order tensor norm and $\Delta\mathbf{C} := \mathbf{C} - \mathbf{C}^{\text{V,Iso}}$ is the difference of the stiffness tensor and its Voigt isotropic value. The *universal* anisotropy

index A^U measures the difference (through their ratios) between the Voigt Eq. (A.9) and Reuss isotropic tensor bulk and shear moduli, based on the averaging of the compliance tensor. Some other general anisotropy indices are proposed in Chung and Buessem (1967); Ledbetter and Migliori (2006); Rychlewski and Zhang (1989); Fang et al. (2019). Given that these indices provide a relatively wide range of values, two indices from these references will also be considered in Section 1.2.5. First is an index based on the anisotropy of wave speed proposed by Ledbetter and Migliori (2006). Considering a planar wave propagating in θ direction in the (x, y) 2D plane, $c_{\max}(\theta)$ and $c_{\min}(\theta)$ correspond to the maximum and minimum wave speeds that are obtained from Christoffel equation. The anisotropy index is defined as $A^{c_{\max}} = [\max_\theta(c_{\max}(\theta))/\min_\theta(c_{\max}(\theta))]^2 - 1$.* That is, it measures the anisotropy of wave speed. Second, Fang et al. (2019) proposes a general energy-based anisotropic index that finds a strain loading for which the largest ratio of strain energy is observed across all angles (in-plane θ angle in 2D). We refer to this index by $A^{\epsilon}_{\text{energy}}$.

A.3 Orthotropic versus isotropic stiffness tensor prediction

Numerical results are used to show why predicting the orthotropic stiffness tensor $\mathbf{C}^*$ is preferred over predicting only isotropic properties. The former is the only viable choice for a macroscopically anisotropic material. As discussed in Sections 1.1 and 1.2.2, predicting a mechanical property from images is more successful and meaningful if SVEs are (much) smaller than the RVE limit. Accordingly, individual SVEs can exhibit notable anisotropies, justifying the prediction of the entire $\mathbf{C}^*$. The anisotropy of the Micro2D dataset in Section 1.2.5 and Figure 1.4, warrants the prediction of $\mathbf{C}^*$ from the discussion above. To strengthen this argument, we will demonstrate how much the predictions would suffer if only isotropic properties were considered.

*The index in Ledbetter and Migliori (2006) does not include the subtraction by one. The subtraction ensures that, similar to other indices, A is equal to zero for an isotropic material.

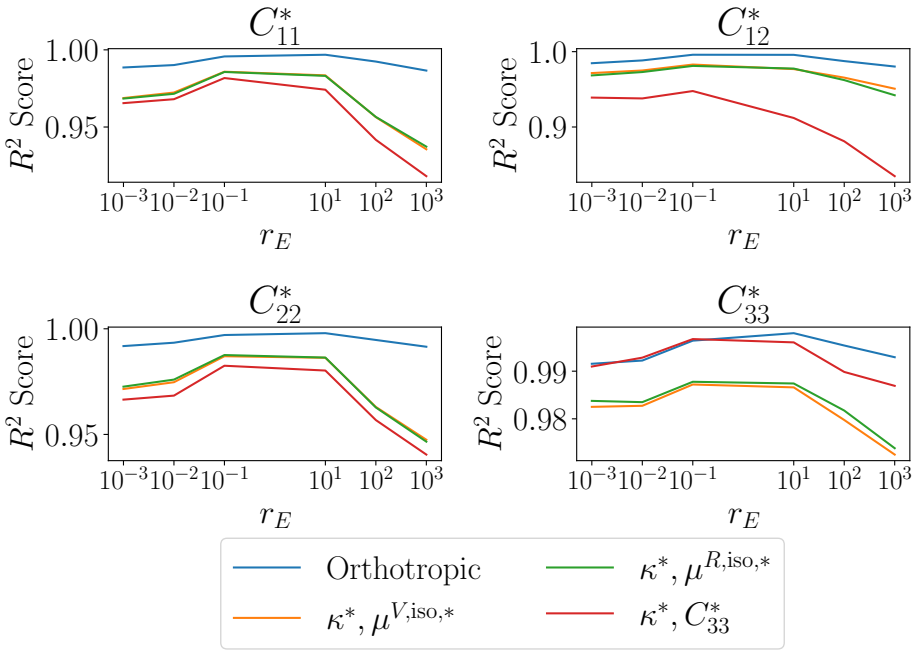


Figure A.1: Comparison of R^2 values for the orthotropic and isotropic predictions. Blue line shows the results from the predictions of the four effective stiffness tensor components. Orange, green and red lines shows the results from isotropic (two independent components) predictions with the effective bulk modulus, κ^* and different choices for the effective shear modulus, $\mu^{\text{V,Iso},*}$, $\mu^{\text{R,Iso},*}$, and C^*_{33}, respectively.

An isotropic elastic material is best represented by its bulk and shear moduli. The 2D bulk modulus is unambiguously defined even for an anisotropic material from (A.7). In contrast, the shear modulus changes by the angle θ. Accordingly, three common approaches of extracting μ^* from $\mathbf{C}^*$ are considered: 1) $\mu^{\text{V,Iso}}$ from (A.10) uses the Voigt-averaged, isotropic $\mathbf{C}^{\text{V,Iso},*}$; 2) $\mu^{\text{R,Iso}}$ uses the Reuss isotropic angle-averaging of the material. This provides $\mu^{\text{R,Iso},*} = 2/(D^*_{11} + D^*_{22} - 2D^*_{12} + D^*_{33})$;[†] 3) C_{33}. This is the least justifiable and easiest choice and simply takes the shear modulus of the (potentially anisotropic) SVE at $\theta = 0$.

Figure A.1, compares the performance of our proposed approach, *i.e.,* predicting the entire $\mathbf{C}^*$, versus the approaches that use κ^* and μ^* to form $\mathbf{C}^*$ from Eq. (A.2). The orthotropic label corresponds to the first approach as Micro2D reports $\mathbf{C}^*$ as

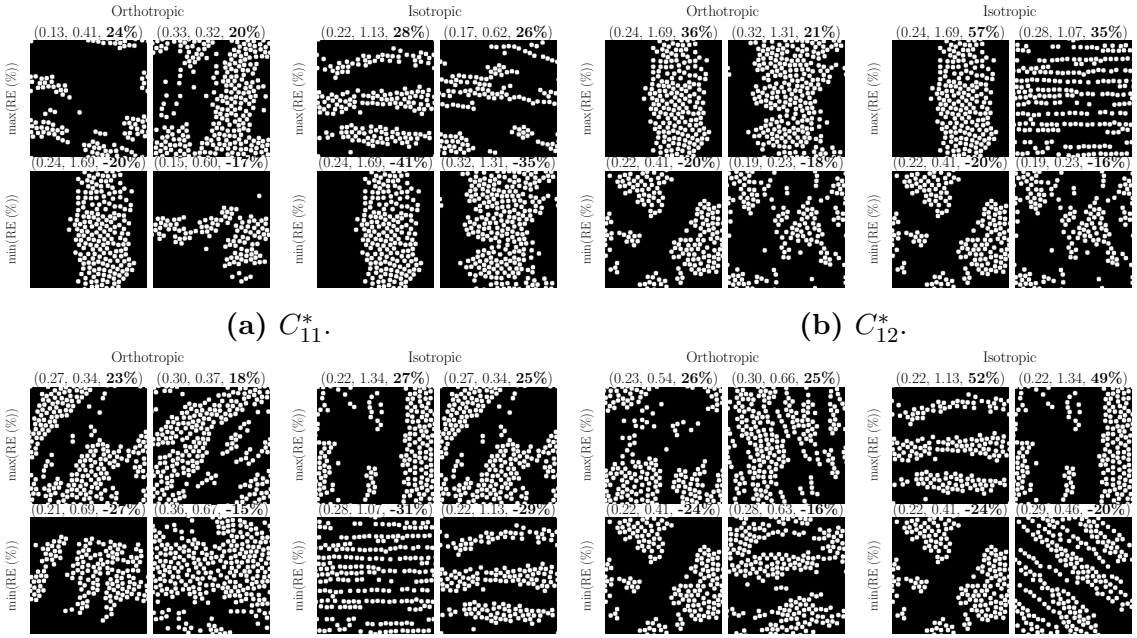
[†]In fact, $\mu^{\text{V,Iso},*} \geq \mu^{\text{R,Iso},*}$ and the ratio of the two is used to define A^U in Ranganathan and Ostoja-Starzewski (2008).

orthotropic. The three comparison results based on κ^* and μ^* differ by the choice of μ^* discussed above. The R^2 of our approach is well above 0.98 for all components and r_E. The R^2 based on (κ^*, μ^*) methods is notably lower, with the worse R^2 of 0.85 to 0.97 for $r_E = 10^3$. An interesting case is the κ^*, C^*_{33} approach; since C^*_{33} directly takes the homogenized C_{33} of the SVE, this approach is very accurate for C_{33}; in fact, its R^2 is comparable to our approach. However, this comes at a cost of severely affecting R^2 of other components, especially C_{12}.

This example demonstrates the advantages of predicting the (entire) stiffness tensor, even if the material is macroscopically isotropic. Figures SM1 and SM2 strengthens this argument by showing that the worst performing SVEs of the (κ^*, $\mu^{\mathrm{V,Iso,}*}$) approach can be about twice as bad for certain components of $\mathbf{C}^*$ and r_E, based on the relative error. Finally, if only isotropic properties are predicted, as is common in literature, one first needs to turn the homogenized $\mathbf{C}^*$ isotropic (Voigt and Reuss approaches showed the same accuracy), rather than directly extracting μ^* from $\mathbf{C}^*$.

Appendix B

Supplemental Materials: Chapter 1



(a) C^*_{11}. **(b)** C^*_{12}.

(c) C^*_{22}. **(d)** C^*_{33}.

Figure B.1: Microstructures with highest relative errors in $\mathbf{C}^{\mathrm{GNN},*}$ diagonal components for $r_E = 0.001$. Orthotropic effective stiffness predictions are compared with isotrpic predictions using κ and $\mu^{\mathrm{V,Iso},*}$. Values in parenthesis are: v^I, $A^{\epsilon}_{\mathrm{energy}}$, and relative error.

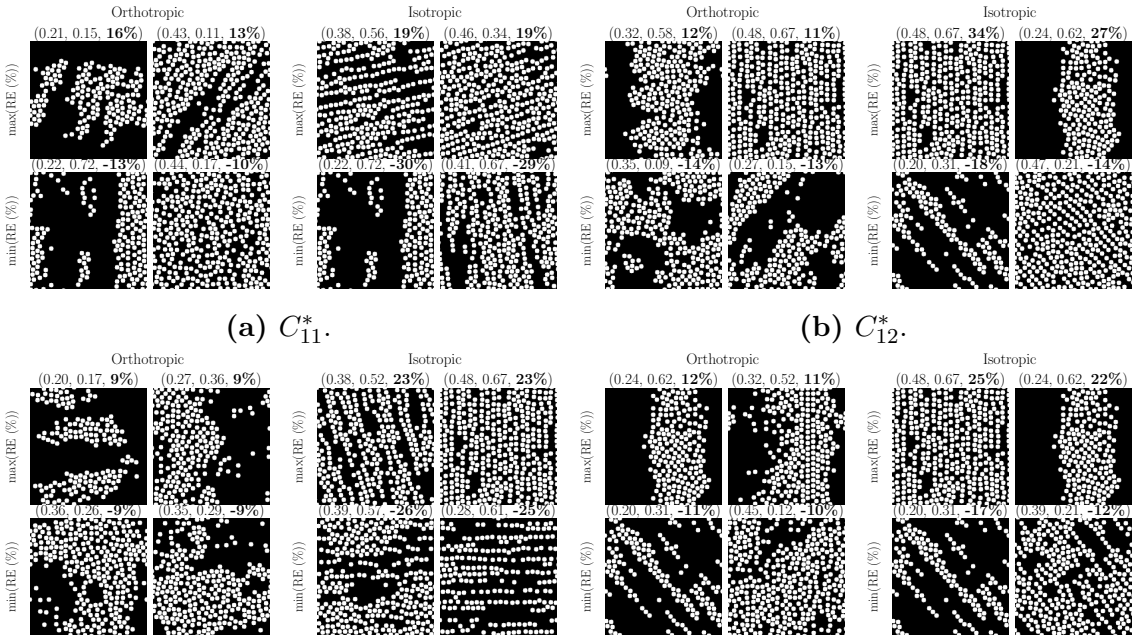


Figure B.2: Microstructures with highest relative errors in $\mathbf{C}^{\mathrm{GNN},*}$ diagonal components for $r_E = 1000$. Orthotropic effective stiffness predictions are compared with isotrpic predictions using κ and $\mu^{\mathrm{V,Iso},*}$. Values in parenthesis are: v^I, $A^{\epsilon}_{\mathrm{energy}}$, and relative error.

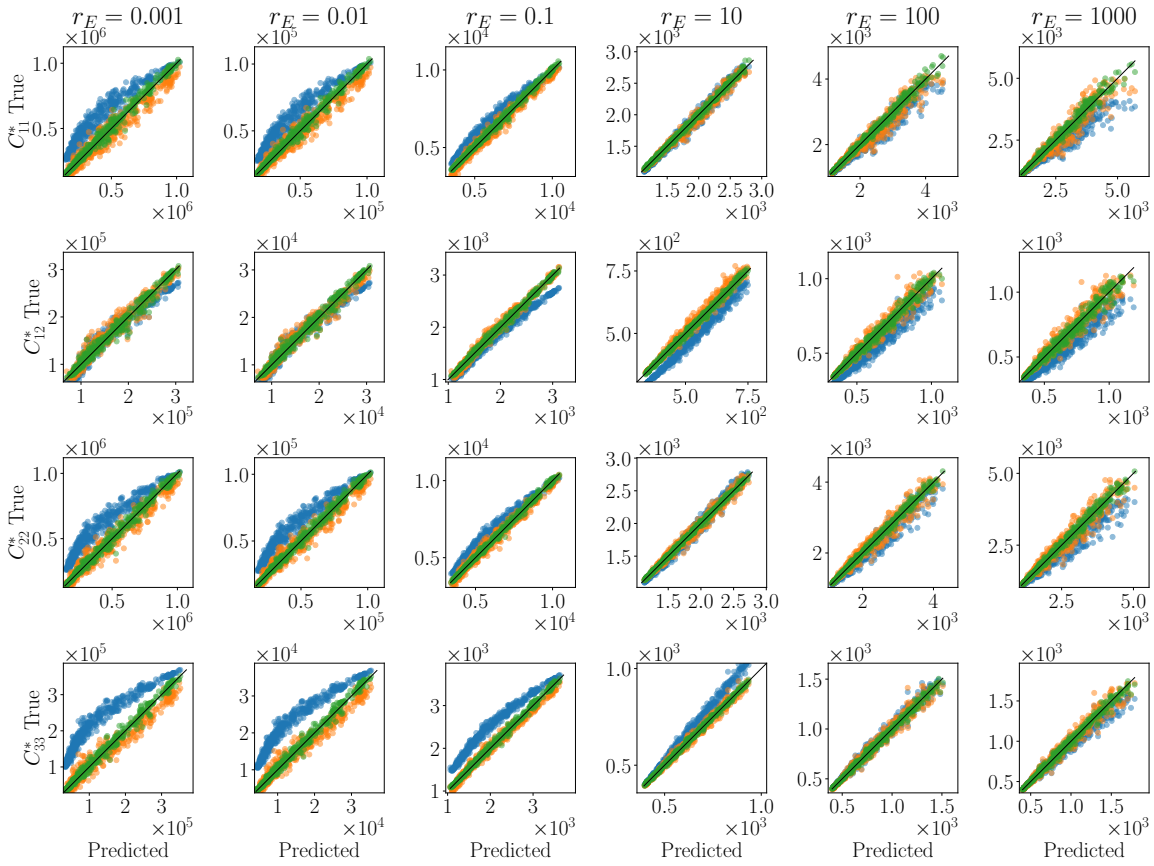
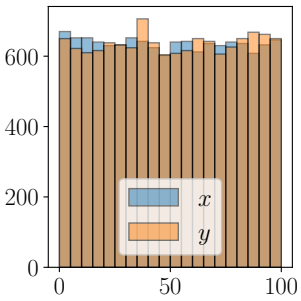
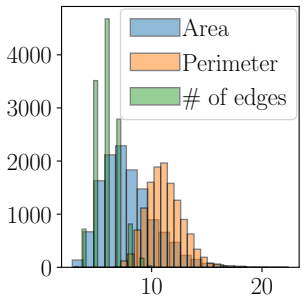


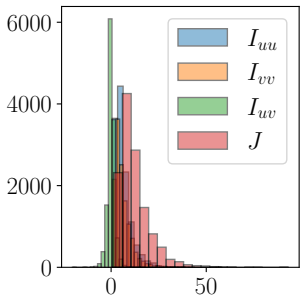
Figure B.3: Parity plots for the Micro2D dataset for DEM (blue), CNN (orange), and GNN (green). Rows correspond to stiffness tensor components, and columns have different contrast ratios.



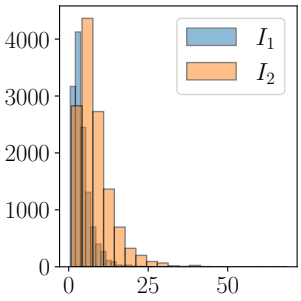
(a) Centroid locations.



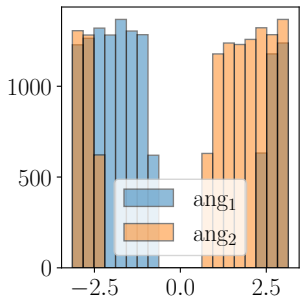
(b) Areas, perimeters, and number of edges.



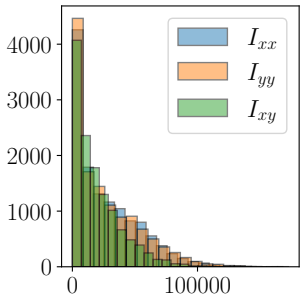
(c) Centroidal and polar moments of inertia.



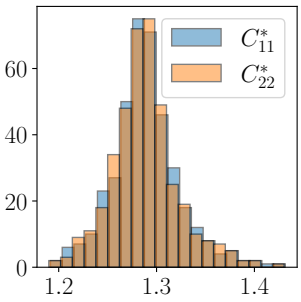
(d) Principal moments w.r.t. origin.



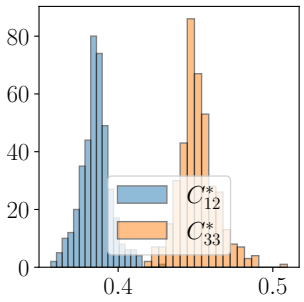
(e) Principle angles.



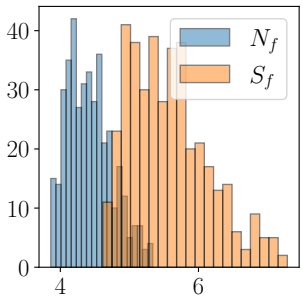
(f) Moments of inertia w.r.t. origin.



(g) Stiffness tensor components.



(h) Stiffness tensor components.



(i) Normal (N_f) and shear (S_f) strength.

Figure B.4: Histograms of the node level features (a to f) and graph level labels (g to i) used with Voronoi partitioning in the initiation strength dataset.

Table B.1: Comparison of all Voronoi features (baseline) with using a single feature at a time.

Feature	Total Loss	Δ Total Loss
All (Baseline)	0.002121	-
Area	0.002950	+0.000829
I_1	0.003265	+0.001143
J	0.003522	+0.001401
Perimeter	0.003852	+0.001731
I_{vv}	0.003874	+0.001752
I_{uu}	0.004083	+0.001962
I_2	0.004097	+0.001976
I_{uv}	0.006987	+0.004865
# of edges	0.007180	+0.005059
ang_2	0.007593	+0.005472
I_{xy}	0.007750	+0.005628
I_{yy}	0.007758	+0.005637
I_{xx}	0.007762	+0.005640
ang_1	0.007799	+0.005678
y	0.007903	+0.005782
x	0.008130	+0.006008

Table B.2: Only-one-group (kept a semantic group and removed the rest) comparison with baseline Voronoi features.

Group name	Input Features	Total Loss	Δ Total Loss
All	(Baseline)	0.002121	-
Shape metrics	Area, Perimeter, # of edges	0.001767	-0.000354
Principal moments	I_1, I_2	0.002262	+0.000141
Centroidal and polar moments of inertia	I_{uu}, I_{vv}, I_{uv}, J	0.002559	+0.000438
Moments of inertia	I_{xx}, I_{yy}, I_{xy}	0.006324	+0.004203
Principle angles	ang_1, ang_2	0.006632	+0.004511
Positions	x, y	0.007071	+0.004950

Table B.3: Leave-one-group-out (removed one semantic group at a time) comparison with baseline Voronoi features.

Group name	Excluded Features	Total Loss	Δ Total Loss
All	(Baseline)	0.002121	-
Positions	x, y	0.001729	-0.000393
Moments of inertia	I_{xx}, I_{yy}, I_{xy}	0.001895	-0.000227
Shape metrics	Area, Perimeter, # of edges	0.002203	+0.000082
Principal moments	I_1, I_2	0.002277	+0.000155
Principal angles	ang_1, ang_2	0.002524	+0.000403
Centroidal and polar moments of inertia	I_{uu}, I_{vv}, I_{uv}, J	0.002546	+0.000424

Table B.4: Informed Feature Selection (SFS) Summary vs. Baseline

Round	Full Set	Total Loss	Δ Total Loss
Baseline	All 16 Features	0.002121	-
1	[2]	0.002883	+0.000762
2	[2, 4]	0.001994	-0.000127
3	[2, 3, 4]	0.001726	-0.000395
4	[2, 3, 4, 11]	0.002034	-0.000088

Table B.5: Complete hyper-parameter specification for the PNA-based GNN used for Micro2D.

Graph generation	
Neighborhood size (k)	5
Architecture	
Model type	PNA
Hidden dimension (d_{hidden})	32
Number of convolution layers	3
Edge features	{`edge length, x, y`}
Activation function	`tanh`
Graph output head	
Shared layers (count / dim)	2 / 32
Head layers (count / dims)	4 / [32, 32, 32, 32]
Variables of interest	
Input node features	Positions x, y and contrast ratio r_E
Output dimension	4 (normalized coefficients for orthotropic stiffness tensor)
Training	
Epochs	100
Training split	90 %
Loss function	Mean-squared error (MSE)
Batch size	64
Optimizer	AdamW
Initial learning rate	1×10^{-4}
LR scheduler	`ReduceLROnPlateau` (*mode* = 'min', factor = 0.5, patience = 5, min(lr) = 10^{-6})

Table B.6: Complete hyper-parameter specification for the PNA-based GNN used for PDF.

Graph generation	
Neighborhood size (k)	5
Architecture	
Model type	PNA
Hidden dimension (d_{hidden})	196
Number of convolution layers	4
Edge features	{`edge length, x, y`}
Activation function	`ReLU`
Graph output head	
Shared layers (count / dim)	2 / 5
Head layers (count / dims)	4 / [78, 91, 165, 163]
Variables of interest	
Input node features	Peak load
Output dimension	1
Training	
Epochs	50
Training split	80 %
Loss function	Mean-squared error (MSE)
Batch size	16
Optimizer	AdamW
Initial learning rate	1×10^{-4}
LR scheduler	`ReduceLROnPlateau` (`mode = 'min'`, `factor` = 0.5, `patience` = 5, `min(lr)` = 10^{-5})

Table B.7: Complete hyper-parameter specification for the PNA-based GNN used for Voronoi dataset.

Graph generation	
Neighborhood size (k)	10
Architecture	
Model type	PNA
Hidden dimension (d_{hidden})	256
Number of convolution layers	3
Edge features	{`edge length`, `x`, `y`}
Activation function	`ReLU`
Graph output head	
Shared layers (count / dim)	2 / 64
Head layers (count / dims)	2 / [265, 128]
Variables of interest	
Input node features	Effective stiffness tensor components (orthotropic) and the initiation strength (normal and shear)
Output dimension	6
Training	
Epochs	60
Training split	90 %
Loss function	Huber
Batch size	20
Optimizer	AdamW
Initial learning rate	1×10^{-5}
LR scheduler	`ReduceLROnPlateau` (`mode` = `'min'`, `factor` = 0.5, `patience` = 5, `min(lr)` = 10^{-7})

Appendix C

Appendix: Chapter 2

C.1 CPFEM Model and Material Properties

C.1.1 Crystal Plasticity Model

A rate-dependent phenomenological crystal plasticity formulation was used to describe the constitutive behavior at the grain level. Plastic deformation occurs by crystallographic slip, with the slip rate on each system α governed by a power-law relation,

$$\dot{\gamma}^{\alpha} = \dot{\gamma}_0 \left(\frac{|\tau^{\alpha}|}{\tau_c^{\alpha}} \right)^n \text{sign}(\tau^{\alpha}), \tag{C.1}$$

where $\dot{\gamma}^{\alpha}$, τ^{α}, and τ_c^{α} denote the shear strain rate, resolved shear stress, and critical resolved shear stress, respectively. The evolution of slip resistance follows a Voce-type hardening law,

$$\Delta\tau_c^{\alpha} = \sum_{\beta} H_{\beta}^{\alpha} \, \Delta h^{\beta}, \tag{C.2}$$

$$\Delta h^{\beta} = h_0 \left(1 - \frac{\tau_c^{\beta}}{S_s} \right)^m |\dot{\gamma}^{\beta}| \Delta t, \tag{C.3}$$

where H_{β}^{α} is the hardening interaction matrix, and h_0, S_s, and m are the hardening parameters. Δt denotes the time increment. Additional details of the CPFEM formulation are available in Anto et al. (2024).

C.1.2 Material Properties

Phase-dependent elastic constants and plastic parameters were assigned at the grain level for ferrite and martensite phases. The material parameters used in the simulations are summarized in Table C.1

Table C.1: Material parameters for ferrite and martensite.

Phase	$\mathcal{C}_{11}$ (GPa)	$\mathcal{C}_{12}$ (GPa)	$\mathcal{C}_{33}$ (GPa)	$\dot{\gamma}_0$	n	h_0	S_s (MPa)	m
Martensite	417.40	242.40	211.10	0.001	20	40000	700	2.5
Ferrite	218.37	113.31	105.34	0.001	20	4500	370	1.3

C.2 Cross-volume-fraction transfer example

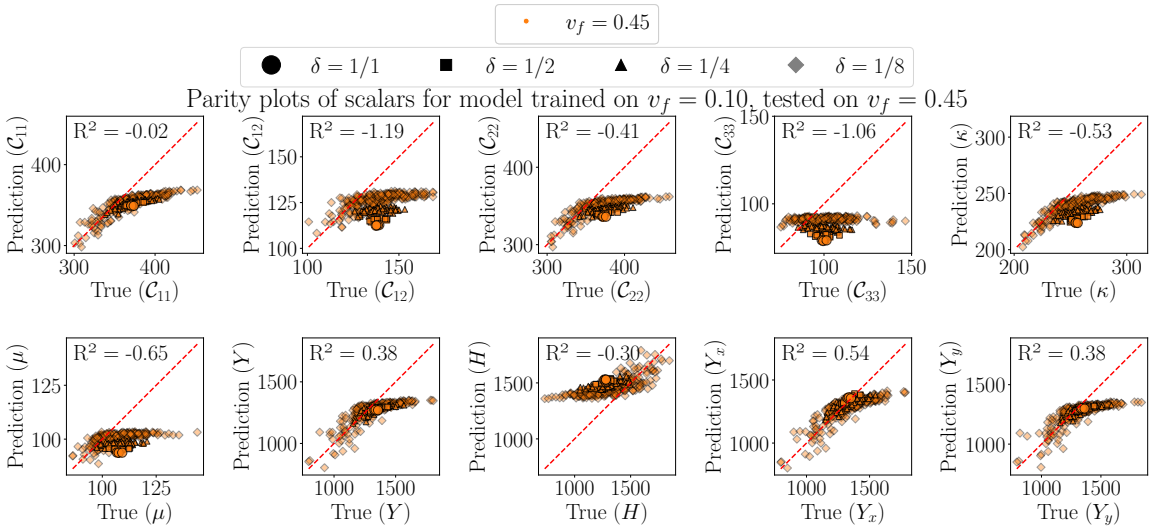


Figure C.1: Parity plots for scalar QoIs for a model trained on $v_f = 0.10$ and evaluated on $v_f = 0.45$. Marker shape denotes the SVE partition scale δ. The noticeable loss of accuracy across several outputs indicates limited direct transferability between composition regimes, motivating the pooled volume-fraction training strategy discussed in Section 2.4.1.

Appendix D

Appendix: Chapter 3

D.1 Hyper-parameter optimization

The selected configuration of Section 3.3.4 was selected by a multi-objective NSGA-II search across the four FIP cases with target metrics balancing tail R^2, calibrated coverage, and the mechanical-auxiliary R^2. The search budget was 200 trials over a 13-knob space spanning encoder capacity (hidden dimension, number of message-passing layers), regularization (dropout rate, activation function), graph pooling, target-loss weights, and the heteroscedastic-NLL weighting on the per-grain peak and per-volume MAX heads. Three admissibility floors guarded against degenerate solutions: bulk per-grain $R^2 \geq 0.55$ on HCF/FS, calibrated 90% coverage on the per-grain head within ± 0.10 of nominal on each case, and mean mechanical-auxiliary $R^2 \geq 0.7$. Pareto trials passing the floors were sorted by lexicographic preference (HCF tail R^2, then per-grain R^2, then within-grain σ-recovery R^2). Per-trial convergence and the full multi-objective Pareto front are committed alongside the code.

Table D.1 summarizes the selected configuration used for all body results. Architecture and training knobs are held fixed across the four FIP cases; only the train/validation/test split seed varies across the 10 replicates that produce the multi-seed mean and standard-deviation reports of Section 3.4. The full configuration

JSON, the dataset hash, and the per-claim seed protocol are committed to the project repository so that any reported body number can be regenerated from the corresponding seed and config.

Table D.1: Selected configuration of the multi-task graph surrogate ($n = 10$ multi-seed replicates). All replicates share the same architecture and training knobs; only the random split seed varies.

field	value
MPNN type	SAGE (GraphSAGE)
hidden dimension	256
number of MP layers	3
graph pooling	max
activation	ELU
dropout rate	0.10
graph-attr conditioning	on (L_{SVE}, size-bucket index)
input feature dimension	32
early-stopping metric	per-grain R^2 on validation
multi-seed replicates	$n = 10$ split seeds

The four loss-weight constants $\{w_m, w_M, w_a, w_{\mathrm{aux}}\}$ of Equation (3.11) and the per-case NLL-to-Huber weight on the heteroscedastic heads are HPO-tuned and reported in the configuration JSON committed alongside the code. A sensitivity sweep on the per-grain target weight w_a shows that it acts as an effective rebalancing knob between bulk and tail per-grain R^2 but does not affect the mechanical-auxiliary R^2, which is controlled by the encoder and head capacity rather than by loss weighting.

D.2 Per-grain input feature inventory

Table D.2 enumerates the 32 scalar components of the per-grain input vector $\mathbf{x}_g$ defined in Section 3.3.2. The four blocks are organized by physical origin and apply identically across the four FIP cases; no FIP-derived quantity enters $\mathbf{x}_g$. The voxel-geometry block is computed from per-grain voxel adjacency counts and distance transforms and does not use any cyclic-plasticity field.

Table D.2: Per-grain input feature vector $\mathbf{x}_g$, total dimension 32. Block widths sum to $16 + 9 + 6 + 1 = 32$. Quaternion components are taken in the fundamental zone; Schmid factors are sorted in descending order within each grain so that the leading entry is the maximum factor S_1. The 2D inertia descriptors are computed on the grain's projected cross-section.

block	feature	dim.
crystallography (16)	unit quaternion q_0, q_1, q_2, q_3	4
	sorted Schmid factors over 12 FCC slip sys.	12
per-grain morphology (9)	grain volume	1
	grain surface area	1
	grain perimeter	1
	2D inertia components I_{uu}, I_{vv}, I_{uv}	3
	polar moment of inertia J_{polar}	1
	principal 2D moments I_1, I_2	2
voxel-geometry moments (6)	surface-voxel fraction	1
	interior-voxel fraction	1
	mean distance to grain boundary	1
	maximum distance to grain boundary	1
	mean exposed-face count per voxel	1
	total voxel count	1
topology (1)	node degree (voxel-adjacent neighbors)	1

The unit quaternion encodes the grain's orientation in the fundamental zone. The Schmid block stores the twelve octahedral FCC slip-system Schmid factors sorted in descending order. Per-grain morphology covers the geometric size, surface, and 2D-projected moment-of-inertia descriptors of the grain shape. The voxel-geometry moments summarize the voxel-level boundary structure: the fraction of grain voxels that sit on the grain–grain interface versus the interior, distance-transform moments to the nearest boundary voxel, the mean number of exposed faces per voxel, and the total voxel count. Topology contributes a single scalar: the per-grain node degree, defined as the number of voxel-adjacent grain neighbors.

Vita

Erdem Caliskan received his Bachelor of Science degree in Mechanical Engineering from Istanbul Technical University in 2020 and his Master of Science degree in Mechanical Engineering from Istanbul Technical University in 2021. He began his doctoral studies in Mechanical Engineering at the University of Tennessee, Knoxville, in 2022 under the supervision of Prof. Reza Abedi.

His doctoral research focused on finite element modeling, multiscale mechanics, fatigue and failure of heterogeneous materials, graph neural networks, and data-driven modeling for materials science. After completing his Ph.D., he will continue his career as a Structural Analysis Engineer, applying computational mechanics and simulation-based design to advanced engineering systems.