Highlights

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

Erdem Caliskan, Anik Das Anto, Stephanie TerMaath, Reza Abedi, Massimiliano Lupo Pasini

- First neural-network prediction of the per-SVE maximum FIP, instead of grain averages
- One graph model spans the FIP hierarchy: within-grain shape, grain peak, SVE maximum
- Closed-form Dirac and Gumbel EVT bridge with a measured active-grain fraction is derived
- EVT bridge recovers within-size variation of the extreme that direct regression misses
- Four FIP cases and five SVE sizes with calibrated grain- and SVE-level UQ intervals

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

Erdem Caliskan[a], Anik Das Anto[a], Stephanie TerMaath[a], Reza Abedi[a], Massimiliano Lupo Pasini[b,*]

[a] *Department of Mechanical, Aerospace and Biomedical Engineering, University of Tennessee Knoxville, 1506 Middle Drive, Knoxville, 37916, TN, USA*

[b] *Oak Ridge National Laboratory, Computational Sciences and Engineering Division, 1 Bethel Valley Rd, Oak Ridge, 37831, TN, USA*

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 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 interest. 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 inactive Dirac component at zero with an active-grain Gumbel component. The graph-level head predicts the per-SVE maximum directly and is the practical estimator across all four cases. Evaluated as a reference with the true within-SVE moments, the closed-form bridge attains the lowest error among the tested estimators, exceeding the trained head on three of four cases and matching it on the fourth; this reference sets the attainable ceiling of the extreme-value map. With predicted moments instead, the bridge remains accurate on the high-cycle cases, where the per-grain head preserves the active-grain upper tail, but degrades substantially on the low-cycle cases. A size-controlled analysis shows that direct prediction captures the size scaling of the maximum but not its within-size variation, which the reference bridge recovers from the lower-level statistics.

Keywords: crystal plasticity; fatigue; graph neural network; extreme-value statistics; fatigue indicator parameter; statistical volume element

Nomenclature

Statistical operators act over the population named in the conditioning: $M(\cdot \mid \cdot)$ maximum, $A(\cdot \mid \cdot)$ average, and $\sigma(\cdot \mid \cdot)$ standard deviation. Subscripts carry the conditioning: $v \mid g$ voxel-in-grain, $g \mid S$ grain-in-SVE with subsets $ga \mid S$ grain-active and $gi \mid S$ grain-inactive.

Symbol	Meaning
f_v	voxel-level FIP within a grain
$M_g = M(f_v \mid g)$	per-grain peak (maximum over voxels)
$A_g = A(f_v \mid g)$	within-grain mean
$\mu_g^{\ln}$, $\sigma_g^{\ln}$	within-grain lognormal location and scale of $\ln f_v$
$M_S = M(M_g \mid S)$	per-SVE maximum (the engineering extreme)
$\mu_{g\mid S}$, $\sigma_{g\mid S}$	mean and standard deviation of $\{M_g\}_{g\in S}$
$\phi_{gi\mid S}$, $\phi_{ga\mid S}$	grain-inactive and grain-active fractions in S
$\mathcal{A}(S)$	active-grain subset $\{g : M_g > \mu_{g\mid S} + \sigma_{g\mid S}\}$
$N_{g\mid S}$, $N_{ga\mid S}$	number of grains, number of active grains in S
$\alpha_{ga\mid S}$, $\beta_{ga\mid S}$	active-component Gumbel location and scale
$\mu_{ga\mid S}$, $\sigma_{ga\mid S}$	mean and standard deviation on the active subset $\mathcal{A}(S)$
γ_E	Euler–Mascheroni constant
$\widehat{M}_g$, $\widehat{M}_S$	predicted per-grain peak and per-SVE maximum
D	dedicated per-SVE-maximum graph head
$\mathrm{EVTB_t}$, $\mathrm{EVTB_p}$	EVT bridge from truth / predicted per-grain moments
$\hat{\sigma}_{\mathrm{a},g}$, $\hat{\sigma}_{\mathrm{a},S}$	aleatoric predictive scale, per grain / per SVE
$\hat{\sigma}_e$	epistemic predictive scale (Monte-Carlo dropout)

1. Introduction

Cyclic loading drives fatigue failure in metallic alloys through rare localization events rather than through the average macroscopic response. Classical Basquin [1] and Coffin–Manson [2, 3] 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 [4–8]. In polycrystals, crack

*This manuscript has been authored in part by UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the US Department of Energy (DOE). The US government retains and the publisher, by accepting the article for publication, acknowledges that the US government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for US government purposes. DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan).

*Corresponding author
Email address: lupopasinim@ornl.gov (Massimiliano Lupo Pasini)

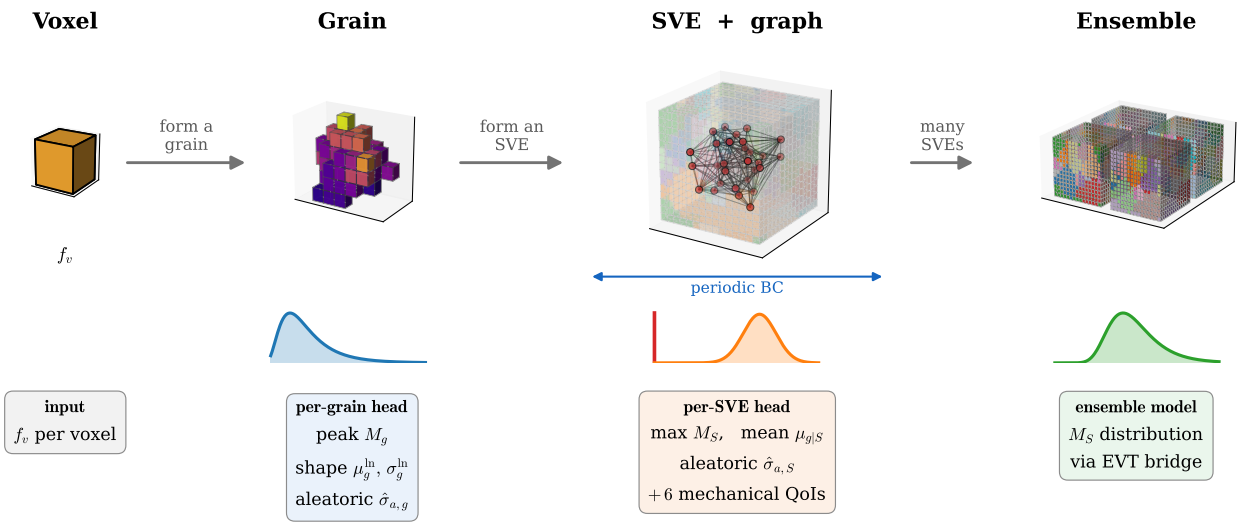
nucleation is often governed by extreme cyclic-plastic activity at the grain and sub-grain scale. The relevant quantity is therefore not a macroscopic average, but the upper tail of a heterogeneous local field.

Crystal-plasticity finite-element modeling (CPFEM) is a physics-based method for resolving this localization. It captures cyclic slip, hardening, and the resulting fatigue-driving fields at the grain and sub-grain scales [9–16]. Modern CPFEM simulations can resolve hundreds to thousands of grains within a *Representative Volume Element* (RVE) or *Statistical Volume Element* (SVE) using slip-system hardening laws calibrated against macroscopic experiments and single-crystal data [17–19]. Prior SVE studies show that the size required for statistical convergence of homogenized responses depends on the response quantity: elastic properties generally converge at smaller sizes than plastic properties, while monotonic and cyclic responses can exhibit different size dependence [20, 21]. Such high-fidelity simulations are well suited to studying microstructure-sensitive fatigue localization. However, their computational cost limits ensemble-scale analyses spanning many microstructural realizations, SVE sizes, loading regimes, and sources of uncertainty.

Fatigue indicator parameters (FIPs) are commonly used to summarize CPFEM cyclic-plasticity fields through scalar quantities associated with crack initiation. The Fatemi–Socie indicator ($\mathrm{FIP}_{\mathrm{FS}}$) [22, 23] 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 [24–29]. The plastic-work-based indicator ($\mathrm{FIP}_{\mathrm{WP}}$) [30, 31] 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 [31–35]. 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 and exhibit pronounced SVE-size dependence [21].

This observation connects microstructure-sensitive fatigue prediction to extreme-value theory (EVT). 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 (Figure 1). 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 [36–40]. This framing is consistent with recent microstructure-sensitive fatigue studies [21, 41–46]. 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 [39, 40]. A surrogate for fatigue criticality should therefore preserve the statistical hierarchy of the CPFEM field, rather than reduce the response to a single bulk-level homogenized scalar value. 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 Gaus-



FIP-head targets are produced for each of the four cases (HCF/LCF × FS/WP).

Figure 1: Hierarchy-aware FIP surrogate (case HCF/FS) across length scales: voxel → grain → SVE → ensemble (smallest size, $L_{\mathrm{SVE}} = 125\,\mu$m, 30 grains). Grains are graph nodes (edges are face adjacencies) with periodic boundaries on all faces. Boxes list the prediction targets at each scale; the small plots show the statistical model at each level (within-grain f_v lognormal, within-SVE peaks M_g mixture, cross-SVE M_S Gumbel).

sian negative log-likelihood (NLL) provides a way to learn feature-dependent predictive scales [47–49], while Monte Carlo dropout (MCD) approximates epistemic uncertainty through stochastic inference [50–52]. These approaches have been used in materials property prediction and crystal-plasticity contexts [53–55]. 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 [56–65]. 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 [42, 64, 66–68] using either ML methods or statistical methods. A multitask grain-graph surrogate has also been developed to predict SVE-level homogenized elastoplastic quantities of interest and stress–strain responses in dual-phase polycrystals [69]. A surrogate framework that predicts tail-relevant statistics across the voxel, grain, and SVE levels, tests whether an extreme-value scale bridge can reconstruct the per-SVE maximum from lower-level statistics, and evaluates uncertainty coverage on the same hierarchy has not yet been reported.

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 (computed on the test split) is reported at both the grain and SVE levels. Second, we derive a closed-form Gumbel-mean scale bridge from a two-component mixture working model on the within-SVE distribution of per-grain peaks: an effectively inactive component at zero combined with an active-grain Gumbel component. The active-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 practical quality of per-grain prediction.

Three aspects of this work are new. First, to our knowledge the maximum over the domain has not previously been predicted with neural networks; the surrogate introduced here is extreme-value- and hot-spot-aware. Second, four fatigue cases are treated and contrasted within one model. Third, taking the maximum of the per-grain predictions to estimate the domain maximum is shown to be inaccurate for fatigue, and a theoretical scale bridge that estimates the maximum from the per-grain distribution is introduced in its place. This comparison has no prior counterpart in microstructure-sensitive fatigue surrogate work.

Section 2 describes the CPFEM dataset, the FIP definitions, the multilevel statistics, and the Gumbel-mean bridge. Section 3 presents the graph construction, input features, multi-task architecture, training objective, and interpretation protocol. Section 4 reports the four-case results, including the direct-versus-bridge comparison, prediction-interval coverage, and mechanical auxiliary predictions. Section 5 summarizes the implications and limitations.

2. Dataset and CPFEM modeling

2.1. Crystal-plasticity simulation and fatigue indicator parameters

The dataset originates from the CPFEM simulations of periodic austenitic stainless steel SVEs reported by [21] and is used here without modification. The polycrystals are single-phase, face-centered-cubic (FCC), with random crystallographic orientations, statistically isotropic grain morphologies, and a mean equivalent-spherical grain diameter of $50\,\mu$m. The microstructures are generated in DREAM.3D [70]. The complete crystal-plasticity formulation, material parameters, grain-size distribution, finite-element discretization, and

periodic boundary conditions are given in [21] and are not repeated here. General treatments of crystal-plasticity models are given in [10, 11]. This subsection summarizes the loading conditions and the two voxel-level FIPs used as surrogate targets. The SVE ensemble and data partitions are given in Section 2.2.

The applied loading is fully reversed strain control at macroscopic strain amplitudes $\varepsilon_{\text{amp}} = 0.001$ and 0.0055, approximately 0.5 and 2.75 times the macroscopic yield strain, respectively. The case at $\varepsilon_{\text{amp}} = 0.001$ is referred to as the high-cycle-fatigue (HCF) regime and is predominantly elastic with sparse plastic activity. The case at $\varepsilon_{\text{amp}} = 0.0055$ is referred to as the low-cycle-fatigue (LCF) regime and has multiple active slip systems with stabilized plastic hysteresis. The voxel-level FIPs are evaluated from the stabilized cyclic response.

Two voxel-level indicators are evaluated for complementary fatigue-localization mechanisms. The plastic-work-based indicator $\text{FIP}_{\text{WP}}(\mathbf{x})$ is the plastic energy density accumulated over one stabilized cycle at voxel $\mathbf{x}$ [25, 30],

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

where τ^{α} is the resolved shear stress and $\dot{\gamma}^{\alpha}$ is the slip rate on slip system α. The sum over the twelve FCC slip systems represents the accumulated cyclic plastic energy at the voxel.

The crystallographic Fatemi–Socie indicator $\text{FIP}_{\text{FS}}(\mathbf{x})$ combines the cyclic shear-strain amplitude on a slip system with the maximum tensile normal stress on the same slip plane [22, 24],

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

where $\Delta\gamma^{\alpha}$ is the cyclic shear-strain range on slip system α, σ_n^{α} is the tensile normal stress on the same slip plane, k is the Fatemi–Socie material parameter, and σ_y is the macroscopic yield stress. The maximum over slip systems selects the most damaging crystallographic plane at each voxel, making FIP_{FS} a critical-plane proxy for slip-band crack initiation. Combining the two FIP definitions with the two loading regimes gives four cases: $\mathcal{C} = \{\text{HCF/FS, HCF/WP, LCF/FS, LCF/WP}\}$, that index the analysis in the remainder of the paper. Each case $c \in \mathcal{C}$ defines its own voxel-level FIP field f_v. The four fields are analyzed in parallel and form the voxel level of the statistical hierarchy used in this work.

2.2. SVE ensemble and four-case dataset

The ensemble contains microstructure variability and SVE size effects simultaneously, so the surrogate is exposed to statistical heterogeneity both within one size and across sizes. 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 [21]. Five SVE edge lengths are sampled, $L_{\text{SVE}} \in \{125, 177, 250, 354, 500\}\,\mu$m, spanning mean grain counts $\bar{N}_{g|S}$ from 30 to 1883 across the size range (Table 1). Figure 2 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 1: Four-case SVE dataset partitioned by edge length L_{SVE} and SVE index.

			N_S			
size bucket	L_{SVE} (μm)	$\bar{N}_{g\|S}$	train	validation	test	total
b0	125	30	180	10	10	200
b1	177	83	179	10	10	199
b2	250	233	180	10	10	200
b3	354	664	180	10	10	200
b4	500	1883	127	7	7	141
		total	846	47	47	940

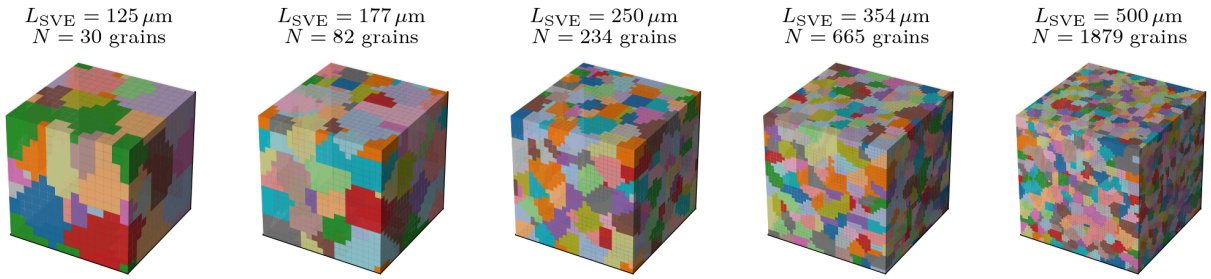


Figure 2: Voxelated SVE microstructures at the five edge lengths $L_{\mathrm{SVE}} \in \{125, 177, 250, 354, 500\}\,\mu$m for one representative realization per bucket. Grains are colored by integer index. The annotated grain counts ($N = 30, 82, 234, 665, 1879$) correspond to these specific realizations and lie close to the ensemble means.

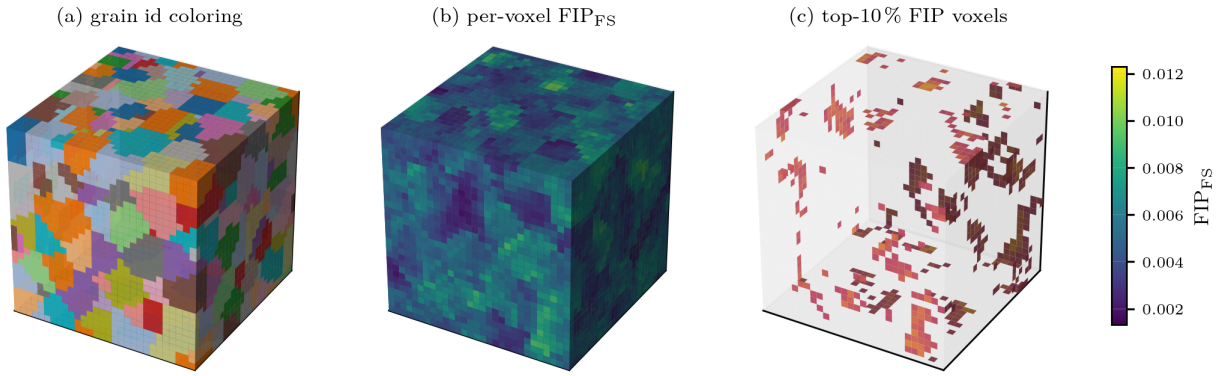


Figure 3: One SVE from the $L_{\mathrm{SVE}} = 250\,\mu$m size bucket ($N = 234$ grains, 29^3 voxels): (a) grain-id coloring, (b) per-voxel $\mathrm{FIP}_{\mathrm{FS}}$ field on the same volume (LCF), and (c) the top-10% voxels of $\mathrm{FIP}_{\mathrm{FS}}$ that identify hot-spot localization (LCF).

Figure 3 maps a single SVE from the $L_{\mathrm{SVE}} = 250\,\mu$m bucket through the three views that motivate the statistical hierarchy of the problem: panel (a) shows the grain-id coloring, panel (b) shows the corre-

sponding per-voxel $\mathrm{FIP}_{\mathrm{FS}}$ field, and panel (c) highlights the top-10% voxels by $\mathrm{FIP}_{\mathrm{FS}}$ that identify hot-spot localization. The highest FIP values occupy a small subset of voxels within a small subset of grains, so the per-volume extreme cannot be represented by a volume average. The SVE ensemble is generated with random crystallographic texture, so no preferred orientation is imposed.

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 1. The dataset corresponds to one alloy and one voxel resolution; cross-alloy and cross-resolution transfer is not evaluated here.

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. We test four candidate parametric shapes (lognormal, normal, Gumbel, and Weibull) using the Kolmogorov–Smirnov (KS) statistic on the non-zero within-grain population. Across the four FIP cases and five SVE size buckets, the within-grain population is well represented by a lognormal working model,

$$\ln f_v \,\big|\, g \;\sim\; \mathcal{N}\big(\mu_g^{\ln},\, (\sigma_g^{\ln})^2\big)\,, \qquad v = 1, \ldots, N_g, \tag{3}$$

where $\mu_g^{\ln}$ and $\sigma_g^{\ln}$ 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).

Median KS p-values per case-by-bucket combination are reported in Table A.1, 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 supports a compact empirical description of the within-grain bulk, that is, the main body of the distribution rather than its upper tail; it is not proof of an exact distributional law for that tail.

The fitted log-scale $\sigma_g^{\ln}$ 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 follow. First, a scalar summary of the within-grain field, such as the arithmetic voxel mean A_g, discards the distributional information encoded in $\sigma_g^{\ln}$; the per-grain peak M_g 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.

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 the per-grain peak FIP

$$M_g \;:=\; \max_{v\in g} f_v \;=\; M(f_v \mid g)\,, \tag{4}$$

one per grain. 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 $N_{g|S}$, we therefore adopt a two-component mixture working model on the within-SVE distribution of per-grain peaks

$$M_g \,\big|\, S \;\sim\; \phi_{gi|S}\,\delta(0) \;+\; \big(1-\phi_{gi|S}\big)\,\mathrm{Gumbel}\big(\alpha_{ga|S},\,\beta_{ga|S}\big), \qquad g\in S, \tag{5}$$

where $\phi_{gi|S}$ is the inactive-grain fraction and $1-\phi_{gi|S}$ the active-grain fraction in the SVE. The active-component $\mathrm{Gumbel}(\alpha_{ga|S},\beta_{ga|S})$ has location $\alpha_{ga|S}$ and scale $\beta_{ga|S}$ parameters. This choice is consistent with the within-grain working model of Section 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 [39], the per-grain peak distribution for active grains is approximately Gumbel at the voxel counts N_g of the present dataset. Inactive grains therefore have per-grain peaks near zero and do not enter the upper tail of the within-SVE distribution.

The active-grain fraction $1-\phi_{gi|S}$ is identified per SVE from the empirical $\{M_g\}_{g\in S}$ distribution as $1-\phi_{gi|S} = \Pr\big(M_g > \mu_{g|S}+\sigma_{g|S} \mid S\big)$, where $\mu_{g|S}$ and $\sigma_{g|S}$ are the within-SVE mean and standard deviation of the full $\{M_g\}_{g\in S}$ sample. The Gumbel parameters of the active component are then estimated by the method of moments on the active subset $\mathcal{A}(S) = \{M_g : M_g > \mu_{g|S}+\sigma_{g|S}\}$,

$$\beta_{ga|S} \;=\; \sigma_{ga|S}\,\frac{\sqrt{6}}{\pi}, \qquad \alpha_{ga|S} \;=\; \mu_{ga|S} - \gamma_E\,\beta_{ga|S}, \tag{6}$$

where $\mu_{ga|S}$ and $\sigma_{ga|S}$ are the sample mean and standard deviation of $M_g \in \mathcal{A}(S)$, $\alpha_{ga|S}$ is the Gumbel location parameter, and γ_E is the Euler-Mascheroni constant. The full within-SVE moments $(\mu_{g|S},\sigma_{g|S})$ enter through the active-subset threshold; the bridge formula in Section 2.5 consumes the active-component parameters $(\alpha_{ga|S},\beta_{ga|S})$ together with the grain count $N_{g|S}$.

Empirically, the active-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 active-grain fraction across all SVEs is 0.21 for HCF/FS, 0.19 for HCF/WP, 0.16 for LCF/FS, and 0.15 for LCF/WP, with case-to-bucket variation under 0.04 per case. Top row of Figure 4 shows the pooled within-SVE density of M_g for

each case together with the fitted mixture, decomposed into the inactive Dirac component near zero and the active-grain Gumbel component.

Equation (3) is checked further through the standardized within-grain location of the per-grain peak, $z_g := (\ln M_g - \mu_g^{\ln})/\sigma_g^{\ln}$. Because M_g is the largest of the N_g voxel values in the grain rather than a typical value, z_g is not centered on zero. For N_g independent draws from Equation (3) the expected standardized maximum is $\mathbb{E}[z_{\max}] \approx \Phi^{-1}\left(\frac{N_g - \frac{3}{8}}{N_g + \frac{1}{4}}\right)$, which gives 2.5 at the mean voxel count $N_g \approx 105$ of this dataset and 2.47 when averaged over the measured grain-size distribution. The measured mean is $\bar{z}_g = 2.30$–2.48 across the four FIP cases and five size buckets, with standard deviation approximately 0.47 and no systematic dependence on L_{SVE} between $N_{g|S} \approx 30$ and $N_{g|S} \approx 1900$; it corresponds to the 99th percentile of the fitted within-grain lognormal. These values are slightly below the independent-sample reference, consistent with spatial correlation between neighboring voxels reducing the effective sample size below N_g. This within-grain size-invariance is the voxel-level counterpart of the bucket-invariance of the active-grain fraction reported above.

2.5. Cross-SVE extremes and the EVT bridge

The engineering-relevant target is the per-SVE maximum FIP, the maximum of the per-grain peaks of Equation (4) over the grains of one SVE,

$$M_S := \max_{g \in S} M_g = M(M_g \mid S), \tag{7}$$

one per SVE. It is predicted by the graph-level head of Section 3.3 and by the closed-form bridge derived below. Across an ensemble of SVEs at fixed L_{SVE}, M_S is itself a random variable, and earlier work on microstructure-sensitive fatigue [21, 41, 42] 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 1): 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.[1] Bottom row of Figure 4 visualizes the cross-SVE distribution of M_S across the four FIP cases at $L_{\mathrm{SVE}} = 250\,\mu$m, as both density and cumulative distribution, together with the maximum-likelihood Gumbel fit per case.

[1] Because the distribution parameters are estimated from the same samples used for the test, these plug-in p-values are approximate similar to within-grain lognormal distribution in Section 2.3. We therefore also perform a parametric bootstrap with $M = 200$ resamples per combination. Under this calibration the Gumbel hypothesis is not rejected at $\alpha = 0.05$ in 15 of 20 combinations, against 10 of 20 for a lognormal and 7 of 20 for a normal fit to the same per-SVE maxima; the bootstrap therefore separates the Gumbel form from the alternatives more clearly than the plug-in test, which is anti-conservative for all three families. Three of the five Gumbel rejections fall at the smallest size bucket, where each SVE contains too few grains for the block-maximum asymptotics to hold. We therefore treat the Gumbel form as the operative cross-SVE working model, best supported at the larger SVEs.

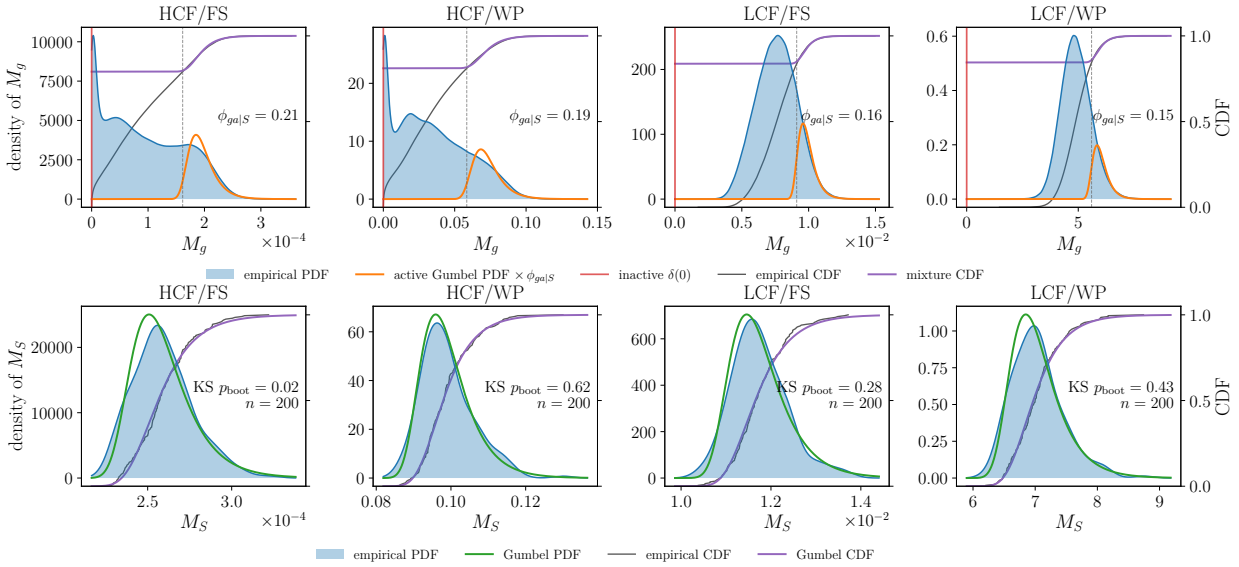


Figure 4: Two-level statistical cascade across the four FIP cases; probability density function (PDF) on the left axis, cumulative distribution function (CDF) on the right. Top row: pooled per-grain peaks M_g with the fitted Dirac-plus-Gumbel mixture (active fraction $\phi_{ga|S}$ annotated, threshold $\mu_{g|S} + \sigma_{g|S}$ dashed). Bottom row: cross-SVE M_S at $L_{\text{SVE}} = 250\,\mu\text{m}$ ($n_{\text{SVE}} = 200$) with the maximum-likelihood Gumbel PDF and CDF and the parametric-bootstrap Kolmogorov–Smirnov p-value.

The empirical cross-SVE Gumbel form is consistent with the within-SVE mixture working model of Section 2.4. Within an SVE, only the active-grain subpopulation contributes to the maximum, since the inactive-grain Dirac component places M_g near zero. Let $\mathcal{A}(S)$ denote the active subset of an SVE and $N_{ga|S} := |\mathcal{A}(S)|$ its integer cardinality. Treating the active-grain peaks as independent and identically distributed (iid) draws from Gumbel$(\alpha_{ga|S}, \beta_{ga|S})$ and conditioning on $N_{ga|S}$, the max-stability of the Gumbel family, a consequence of the Fisher–Tippett–Gnedenko theorem [36, 38, 39], gives

$$M_S \,\big|\, S,\ N_{ga|S} \ \sim\ \text{Gumbel}\Big(\alpha_{ga|S} + \beta_{ga|S} \ln N_{ga|S},\ \beta_{ga|S}\Big). \tag{8}$$

The bridge method of this work uses the effective-count approximation $N_{ga|S} \approx N_{g|S}(1 - \phi_{gi|S})$, replacing the integer active-grain cardinality by its expected value under the mixture. The corresponding analytical mean of M_S is

$$\mathbb{E}\big[M_S \,\big|\, \alpha_{ga|S}, \beta_{ga|S}, N_{g|S}, \phi_{gi|S}\big] \ \approx\ \alpha_{ga|S} \ +\ \beta_{ga|S}\big(\ln\big(N_{g|S}(1 - \phi_{gi|S})\big) + \gamma_E\big)\,. \tag{9}$$

Equation (8) specifies a distribution for M_S, and the bridge uses its mean, Equation (9), as the point estimate. The mean is used because the estimators are scored by squared-error criteria (R^2 and RMSE in Section 4.2), and the mean of the fitted law minimizes the expected squared error under that law. The median and the mode of the same law, lower than the mean by $\beta_{ga|S}(\gamma_E + \ln\ln 2) \approx 0.21\,\beta_{ga|S}$ and $\beta_{ga|S}\gamma_E \approx 0.58\,\beta_{ga|S}$, would be used under absolute-error and modal criteria. Equation (9) is exact under the iid approximation on $\mathcal{A}(S)$ when the integer cardinality $N_{ga|S}$ is used; the effective-count form used in this work carries the

additional approximation $\ln N_{ga|S} \approx \ln N_{g|S}(1-\phi_{gi|S})$, which is accurate to within a few percent at the active-grain fractions of the present dataset.

The expected per-SVE maximum grows linearly in $\ln\big(N_{g|S}(1-\phi_{gi|S})\big)$, that is, in the logarithm of the number of active grains rather than of all grains, at a rate set by the active-component Gumbel scale $\beta_{ga|S}$. This is consistent with the linear-in-$\ln N_{g|S}$ length-scale scaling reported in earlier work on the same ensemble [21].

Two limiting cases are worth noting. When $\phi_{gi|S} \to 0$ all grains are active, the inactive Dirac component vanishes, and Equation (9) reduces to the pure-Gumbel form with $N_{g|S}(1-\phi_{gi|S}) = N_{g|S}$. Replacing the within-SVE Gumbel component by a Gaussian on the full $\{M_g\}_{g\in S}$ population yields the Gaussian-parent extreme-value formula [39, 71], in which the leading-order extreme-location increment scales with $\sigma_{g|S}\sqrt{2\ln N_{g|S}}$ together with an additive Gaussian-maximum correction. Neither case applies to the present dataset, where the active-grain fraction $1-\phi_{gi|S}$ lies between 0.15 and 0.21 across the four cases and where the within-SVE population 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 (5)–(9) define the closed-form bridge used in this work: Given the within-SVE active-component Gumbel parameters $(\alpha_{ga|S}, \beta_{ga|S})$, the active-grain fraction $1-\phi_{gi|S}$, and the grain count $N_{g|S}$, 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 4.2. The bridge assumes an effective sample of $N_{g|S}(1-\phi_{gi|S})$ independent draws from the active-grain Gumbel parent. Spatial correlation among neighboring grains and SVE-level mechanical constraints can reduce the effective sample size below this nominal value. Figure 5 summarizes the three-step recipe: applied to the held-out true per-grain peaks it gives the empirical-moment bridge $\mathrm{EVTB_t}$, and applied to the per-grain head's predictions it gives the predicted-moment bridge $\mathrm{EVTB_p}$. The empirical sufficiency of the working model is tested directly in Section 4.2.

2.6. Cross-criterion and cross-regime correlation structure

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 6 reports the Spearman rank correlation of the voxel-level FIPs across the four cases, pooled over all voxels in the dataset. Three 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 $\mathrm{FIP_{FS}}$ is closely aligned with the plastic-work accumulation that drives $\mathrm{FIP_{WP}}$, so the two indicators induce nearly the same voxel-level ordering. At high

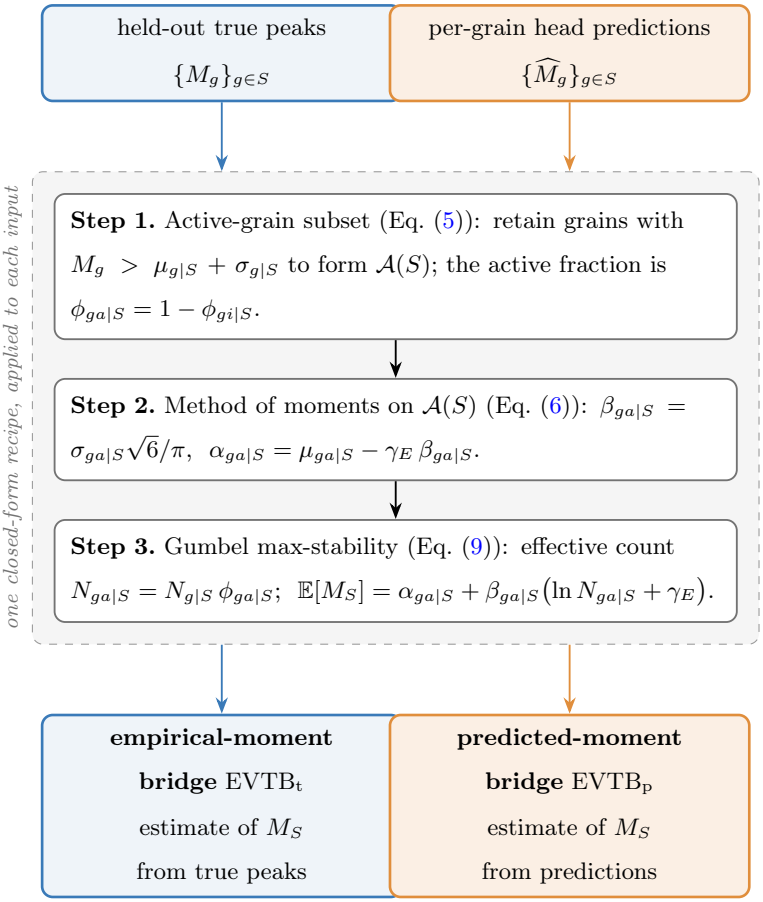


Figure 5: Closed-form extreme-value bridge from per-grain peaks to the per-SVE maximum M_S. One three-step recipe (active-grain subset, method of moments, Gumbel max-stability) is applied to two inputs: true peaks $\{M_g\}$ give the empirical-moment bridge $\mathrm{EVTB_t}$, and predicted peaks $\{\widehat{M}_g\}$ the predicted-moment bridge $\mathrm{EVTB_p}$.

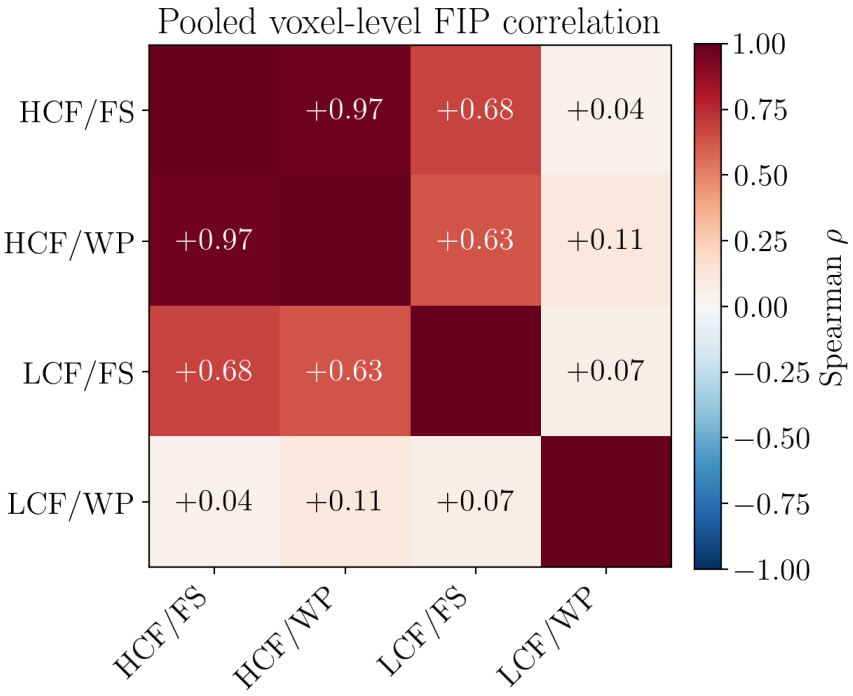


Figure 6: Pooled voxel-level Spearman rank correlation between the four FIP cases. FS and WP are nearly identical under HCF ($\rho_S = 0.97$) but decoupled under LCF ($\rho_S = 0.07$); FS is partially regime-transferable ($\rho_S = 0.68$ between HCF/FS and LCF/FS); LCF/WP correlates weakly with the other three.

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.

This correlation structure is used to interpret the surrogate results; it is not 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 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. Surrogate architecture and training

3.1. Graph construction

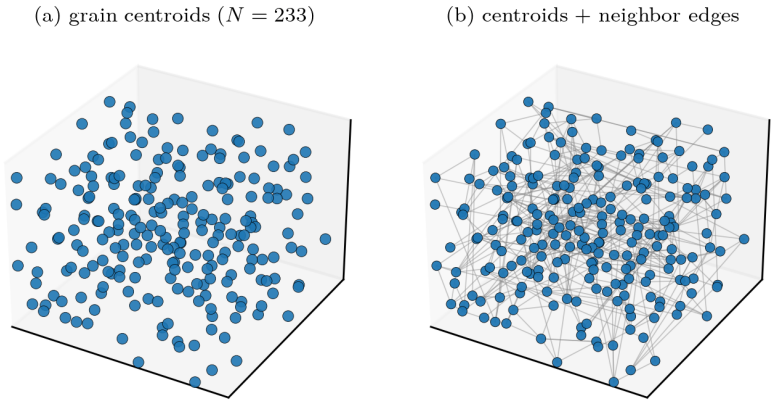


Figure 7: Graph construction for one $L_{\mathrm{SVE}} = 250\,\mu$m SVE ($N_{g|S} = 233$ grains, 1684 undirected edges): (a) grain centroids treated as graph nodes, and (b) the same centroids with neighbor edges drawn between voxel-adjacent grains, including periodic-face neighbors (not shown in the figure).

Each SVE is converted to an undirected grain-adjacency graph $G = (\mathcal{V}, \mathcal{E})$, where each node in the vertex set, $g \in \mathcal{V}$, represents one grain and carries the input feature vector $\mathbf{x}_g$ described in Section 3.2. An edge in the edge set, $(g, g') \in \mathcal{E}$, 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 3.4 consumes the graph adjacency and node features only; edge attributes are not used in the selected configuration. Figure 7 illustrates the construction for one $L_{\mathrm{SVE}} = 250\,\mu$m SVE.

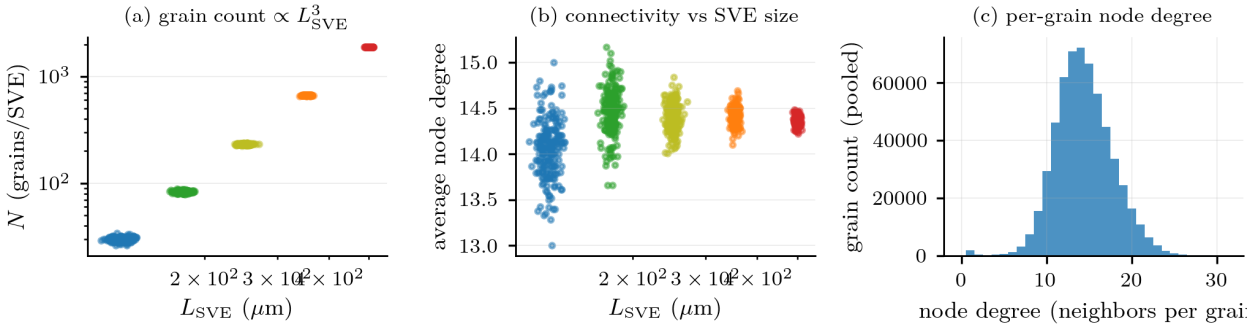


Figure 8: Graph-level statistics across the dataset: (a) grain count $N_{g|S}$ per SVE versus edge length L_{SVE}, exhibiting the expected cubic scaling; (b) per-SVE average node degree, approximately size-invariant near $\bar{k} \approx 14.4$; and (c) pooled per-grain node-degree histogram, with a mode near 14 neighbors and a long tail to about 30.

Two properties of the graph ensemble matter for generalization across SVE sizes (Figure 8). First, the grain count $N_{g|S}$ scales approximately as L_{SVE}^3 (Figure 8a), 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 8b), with a slightly lower $\bar{k}$ for the smallest SVE size, which has a relatively higher population of grains near the SVE boundaries. Moreover the pooled per-grain degree distribution spans roughly 5 to 30 neighbors with a mode near 14 (Figure 8c). The empirical graph diameter increases from a median of 2 hops on the smallest SVEs ($L_{\mathrm{SVE}} = 125\,\mu$m, $N_{g|S} \approx 30$, range 2–4) to a median of 9 hops on the largest SVEs ($L_{\mathrm{SVE}} = 500\,\mu$m, $N_{g|S} \approx 1880$, 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.

3.2. 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-grain feature vector is given in Appendix B.

The crystallographic block (16 scalars) encodes the grain's orientation 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 3.5); the per-grain maximum Schmid factor S_1, defined as the leading entry of the sorted Schmid vector, has an empirical distribution sharply peaked between 0.4 and 0.5 (Figure 9a). The per-grain morphology block (9 scalars) encodes grain vol-

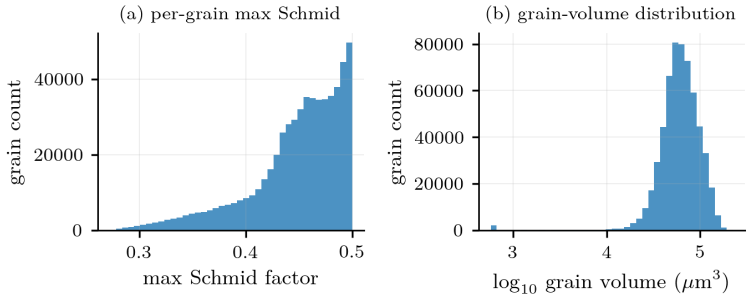


Figure 9: Pooled per-grain input-feature distributions: (a) maximum Schmid factor S_1 (peaked at 0.4–0.5, as expected for random-texture FCC); (b) grain volume ($\log_{10}$, approximately log-normal).

ume, 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 9b). 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 8c). 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. The SVE size is not available from $\mathbf{x}_g$: the per-grain features describe a grain and its immediate neighborhood, and a fixed-depth message-passing encoder covers a decreasing fraction of the graph as the SVE grows (Section 3.1), while the per-SVE maximum increases with the grain count through Equation (9). Earlier ablation studies on this dataset (not shown here for brevity) show that exposing L_{SVE} explicitly to the encoder removes a small size-dependent bias in the per-SVE maximum prediction and leaves the per-grain prediction quality unchanged, as expected if the missing information is at the graph level only. 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 to prevent leakage.

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 A_g, and the within-grain lognormal parameters $\mu_g^{\ln}$ and $\sigma_g^{\ln}$ from Equation (3). 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 targets per SVE: the four per-SVE maxima $M_S = M(M_g \mid S)$ of Equation (7), the four within-SVE means $\mu_{g|S} = A(M_g \mid S)$ of the per-grain peak population $\{M_g\}_{g\in S}$ defined in Section 2.4, and six volume-averaged mechanical auxiliaries. These mechanical auxiliaries are the yield strength σ_y, hardening modulus H_p, 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 4.2, where it is compared against the post-hoc maximum and the analytical EVT bridge of Equation (9).

Aleatoric uncertainty is attached only to the M_g and M_S channels. Each per-case M_g output is co-trained with a heteroscedastic Gaussian NLL variance output, yielding a learned per-grain aleatoric uncertainty $\hat{\sigma}_{\text{a},g}(\mathbf{x}_g)$. Each per-case M_S output is co-trained analogously, yielding a learned per-graph aleatoric uncertainty $\hat{\sigma}_{\text{a},S}(\mathbf{X}_G)$, where $\mathbf{X}_G$ is the full graph-level input. In the ideal population-risk setting for a correctly specified Gaussian NLL and a sufficiently flexible model, these variance heads estimate conditional residual variances,[2]

$$\hat{\sigma}^2_{\text{a},g}(\mathbf{x}_g) \to \text{Var}[M_g \mid \mathbf{x}_g], \qquad \hat{\sigma}^2_{\text{a},S}(\mathbf{X}_G) \to \text{Var}[M_S \mid \mathbf{X}_G]. \tag{10}$$

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:

$$L(\theta) \;=\; \sum_{c\in\mathcal{C}}\Big[w_m\, L_m^{(c)} + w_M\, L_M^{(c)} + w_a \sum_{s\in\mathcal{S}} L_s^{(c)}\Big] \;+\; w_{\text{aux}} \sum_{q\in\mathcal{Q}} L_q, \tag{11}$$

where $\mathcal{C}$ indexes the four FIP cases, $\mathcal{S} = \{A_g, \mu_g^{\ln}, \sigma_g^{\ln}, \mu_{g|S}\}$ enumerates the distributional auxiliary targets, and $\mathcal{Q}$ enumerates the six volume-averaged mechanical auxiliaries (σ_y, H_p, 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 (the top 10% of the target up-weighted by 50 times relative to the bulk) and a heteroscedastic Gaussian NLL on $(M_g, \hat{\sigma}_{\text{a},g})$. The per-SVE maximum loss $L_M^{(c)}$ is the sum of a Huber loss on M_S and a heteroscedastic Gaussian NLL on $(M_S, \hat{\sigma}_{\text{a},S})$. The four remaining per-case losses $L_s^{(c)}$, $s \in \mathcal{S}$, are those of the distributional targets: the within-grain mean A_g, the within-grain lognormal parameters $\mu_g^{\ln}$ and $\sigma_g^{\ln}$, and the within-SVE mean $\mu_{g|S}$ of the per-grain peaks. Each of these, and each mechanical auxiliary loss L_q, is a standard mean-squared error on a Yeo–Johnson- and z-score-normalized target, with predictions inverted to raw FIP or mechanical units

[2]The present finite-sample multi-task setting departs from this idealization, so we interpret $\hat{\sigma}_{\text{a},g}$ and $\hat{\sigma}_{\text{a},S}$ as aleatoric uncertainties and assess their reliability through held-out interval coverage in Section 4.3.

at evaluation. The four loss-weight constants $\{w_m, w_M, w_a, w_{\text{aux}}\}$ are fixed at unity, so the tail emphasis enters through the top-decile Huber term rather than the task weights. The encoder-and-head architecture and data flow are summarized in Figure 10.

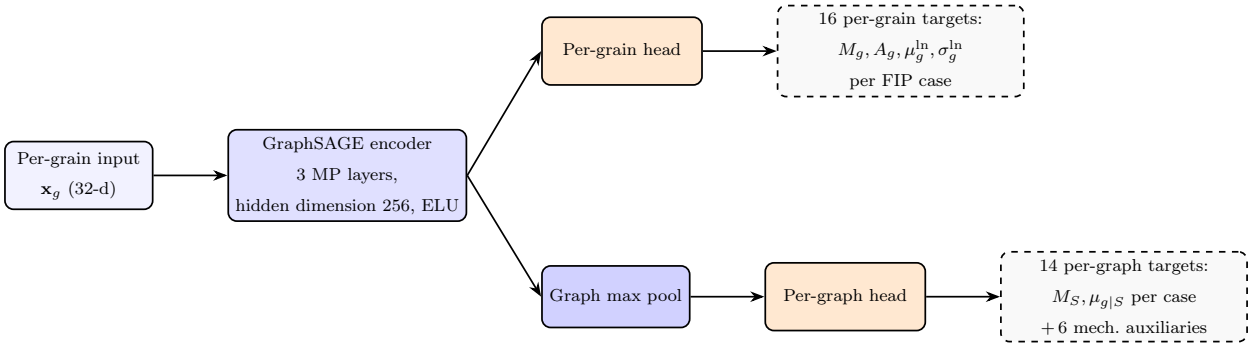


Figure 10: Hierarchy-aware multi-task architecture. Per-grain features $\mathbf{x}_g$ enter a shared 3-layer GraphSAGE encoder; per-grain states feed the per-grain head, while graph-level max pooling feeds the per-graph head. Outputs: the 16 within-grain summaries (per-grain) and, per graph, the four M_S, four $\mu_{g|S}$, and six mechanical auxiliaries.

3.4. Training and hyper-parameter selection

The encoder is a stack of GraphSAGE convolutions [56] 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 and regularization hyper-parameters (hidden dimension, number of message-passing layers, per-grain and per-graph head widths, dropout rate, activation function, graph-attribute conditioning mode, and learning rate) were selected by a multi-objective NSGA-II search; graph pooling, batch size, weight decay, and the loss recipe were held fixed. The two search objectives are the per-grain tail R^2 and a proper scoring rule on the predictive distribution, subject to five admissibility floors (on per-grain peak accuracy, top-1 hit rate, interval coverage, mechanical-auxiliary R^2, and σ-residual rank correlation). The selected configuration uses hidden dimension 256, three message-passing layers, ELU activation, dropout rate of 0.10, and graph-level max pooling. Full hyper-parameter optimization (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 C.

The total training loss is the case-symmetric multi-task objective of Equation (11). 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 results in Section 4. The GraphSAGE backbone was the HPO-selected operating point for the joint per-grain and per-graph objective; it naturally accommodates irregular grain counts and en-

codes microstructural anisotropy through the input features (*i.e.,* 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.[3]

3.5. Model interpretation: permutation importance and ablation

Once trained, the model is analyzed to identify which input features it 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 measures how much of the trained model's prediction quality depends on 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 3.4 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 11. The twelve-dimensional slip-system Schmid block, which projects the macroscopic loading axis onto each of the twelve octahedral slip systems, has by far the largest effect. Permuting this block produces $\Delta R^2 = 1.50 \pm 0.02$ relative to an unpermuted per-grain baseline of $R^2 = 0.504 \pm 0.005$ (mean $\pm$ standard deviation over the ten split seeds), driving the per-grain peak R^2 below zero. The permuted-model value is $0.504 - 1.50 \approx -1.0$: with the Schmid factors scrambled, the surrogate predicts grain-peak magnitude worse than the marginal mean of the target. The per-grain morphology and topology blocks each contribute at most $\Delta R^2 \approx 0.011$ (grain volume, surface area, and node degree are the largest of these), 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 bears on two results reported later. First, the weak LCF/WP performance in Section 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,

[3] An edge-attribute ablation matched to the selected configuration, identical except for the message-passing operator and seven standardized grain-boundary edge features (misorientation angle, edge length, the three relative-position components, the slip-transfer factor m', and the elastic mismatch), replaced the GraphSAGE backbone with a PNA backbone. At the matched split seed it left the high-cycle per-grain R^2 essentially unchanged (HCF/FS 0.949 versus 0.949 for GraphSAGE) and changed the low-cycle per-grain R^2 only marginally, while lowering the per-SVE R^2 on all four cases (HCF/FS 0.544 versus 0.658). The added edge physics therefore did not improve the per-SVE target, and the simpler GraphSAGE backbone was retained.

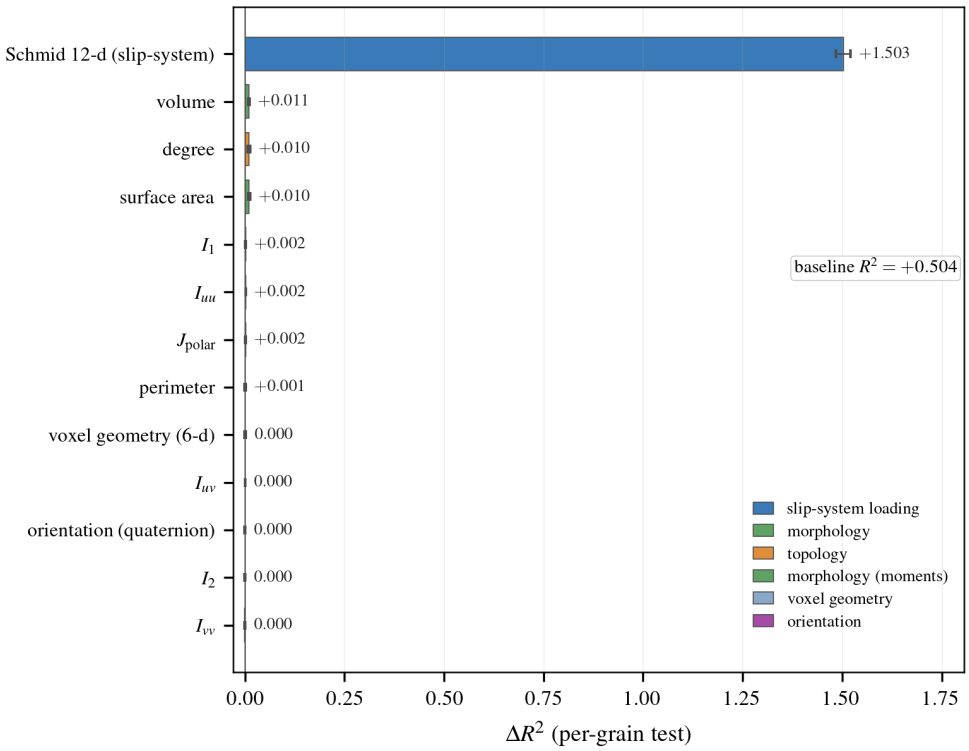


Figure 11: Per-feature permutation importance for per-grain peak prediction (held-out test split, selected configuration, averaged over ten replicates), grouped by mechanics block, with the mean ΔR^2 printed beside each bar. The twelve-dimensional Schmid block carries $\Delta R^2 = 1.50 \pm 0.02$ (baseline $R^2 = 0.504 \pm 0.005$); the remaining blocks contribute $\Delta R^2 \leq 0.011$ and are not resolvable on the common axis scale.

the predicted-moment extreme-value bridge of Section 4.2 stays accurate on the high-cycle cases, where the model relies most strongly on the well-resolved slip-system block, and degrades on the low-cycle cases where that signal weakens.

4. Results and discussion

4.1. Grain level predictions and learnability across cases

The four FIP cases are predicted in parallel by the multi-task encoder of Section 3.4. We first examine whether the surrogate recovers the distributional quantities that support the hierarchy of Sections 2.3–2.5, and which cases are learnable from the present grain-level descriptors, before turning to the per-SVE maximum M_S.

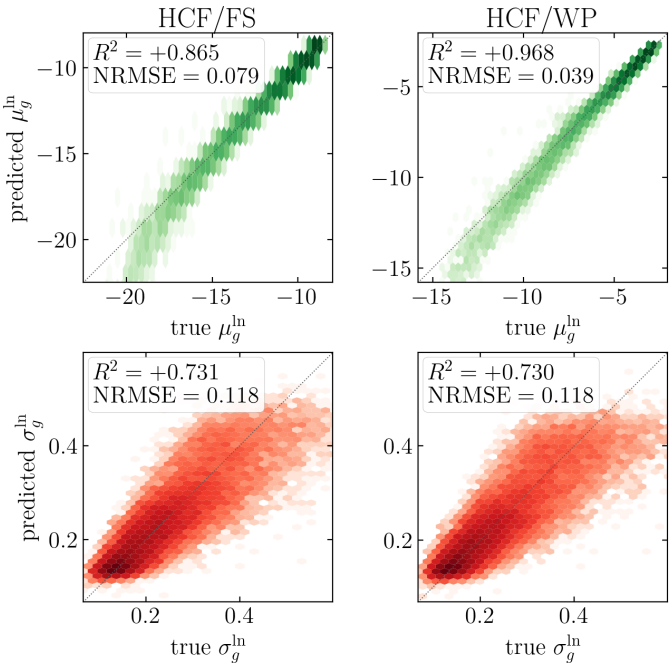


Figure 12: Within-grain log-shape parity for the HCF cases, one column per case: hexbin density of predicted versus true $\mu_g^{\ln}$ (top row) and $\sigma_g^{\ln}$ (bottom row) over the held-out test grains, with R^2 and normalized RMSE annotated per panel. The multi-task encoder recovers the within-grain log-mean $\mu_g^{\ln}$ at $R^2 = +0.97$ for HCF/WP and +0.87 for HCF/FS (top row), and the within-grain log-scale $\sigma_g^{\ln}$ at $R^2 = +0.73$ for both HCF cases (bottom row).

The per-grain head predicts the within-grain lognormal parameters $\mu_g^{\ln}$ and $\sigma_g^{\ln}$ of Equation (3), while the per-graph head predicts the within-SVE mean $\mu_{g|S}$ of the per-grain peak population $\{M_g\}_{g\in S}$, that is, $\mu_{g|S} = A(M_g \mid S)$. The within-SVE scale $\sigma_{g|S}$ 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 12 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 $R^2 = +0.73$

for both HCF cases. The corresponding LCF results are weaker, and are not shown for brevity. Their weaker performance follows the same case ordering as the per-grain peak prediction of Figure 14.

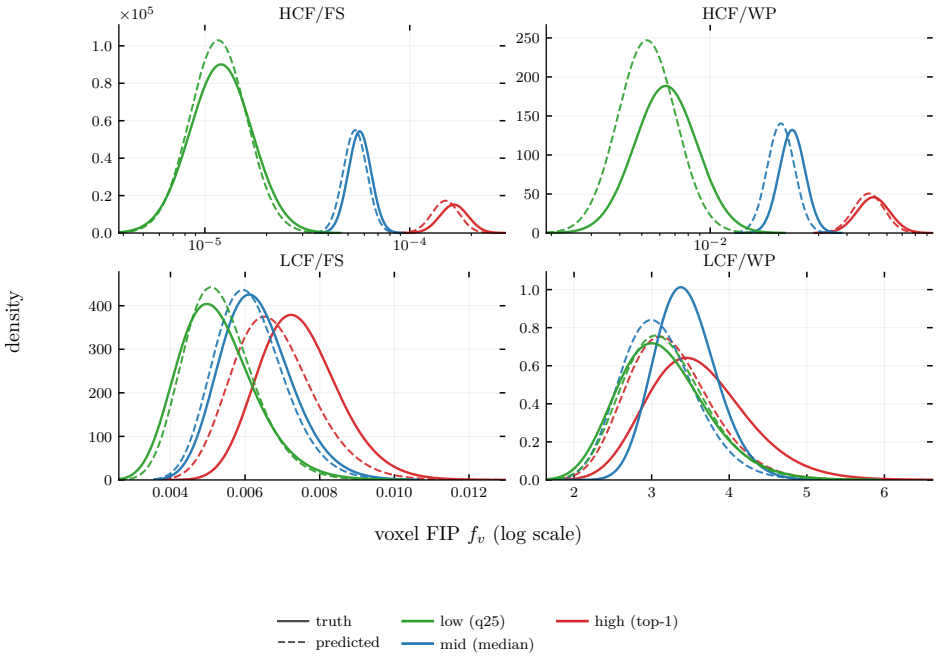


Figure 13: Predicted versus true within-grain voxel-FIP density for three representative grains per case (low, median, and high quantile by per-grain peak M_g). Solid lines: true within-grain density; dashed lines: surrogate-predicted density obtained by inverting $(\hat{\mu}_g^{\ln}, \hat{\sigma}_g^{\ln})$ to a lognormal density.

Figure 13 shows directly what the log-shape predictions result in. For representative grains selected at low, median, and high quantiles of the per-grain peak M_g, the predicted lognormal density is reconstructed from $(\hat{\mu}_g^{\ln}, \hat{\sigma}_g^{\ln})$ and compared with the empirical within-grain voxel-FIP density. The within-grain distributions of FIP values are closely reproduced for the HCF cases. The LCF cases show larger deviations, including a modest mode shift toward lower predicted values on the highest-peak LCF/WP grains, consistent with the weak cross-case correlation of LCF/WP reported in Section 2.6 and with its per-grain peak result in Figure 14.

Figure 14 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 R^2 of the displayed replicate annotated per panel. The replicate is one of the $n_{\text{rep}} = 10$ independent test-split seeds; the multi-seed values below are averages over all ten. 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. The LCF/FS case is moderately learnable ($R^2 = +0.50 \pm 0.01$), whereas the LCF/WP case is not learnable from the present descriptors ($R^2 = -0.47 \pm 0.02$), worse than predicting the marginal mean.

The case-by-case ranking is consistent with the data-side correlation structure of Section 2.6. The two

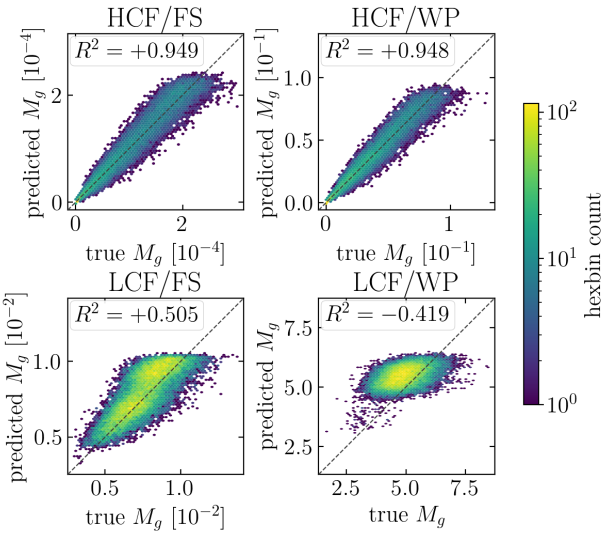


Figure 14: Per-grain peak FIP parity on the held-out test split, one panel per case. Hexbin density of predicted M_g versus true M_g, with the per-case (selected)R^2 annotated.

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 Schmid signal identified by the permutation-importance analysis of Section 3.5. Both HCF cases therefore reach large R^2. 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 shared structure that supports the other three heads is not available to the multi-task encoder for this case.

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.

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 15 reports the recall curves for $K \in \{1, 5, 10, 20, 50\}$ percent, averaged across test SVEs and multi-seed replicates. The HCF cases are well 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. Per-grain R^2 characterizes the average learnability of grain-level hot spots, but it is not sufficient to assess fatigue-critical response: the

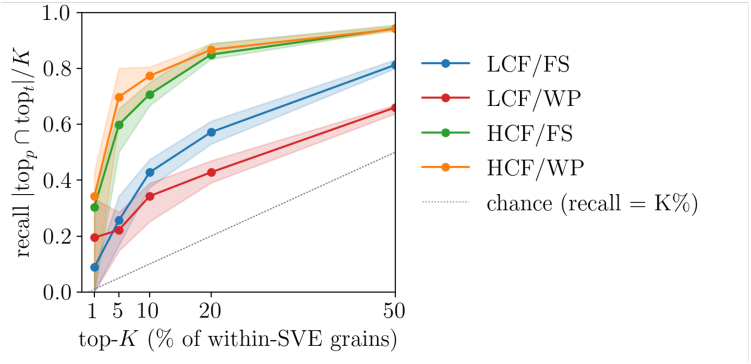


Figure 15: Within-SVE top-K recall of the per-grain head across the four FIP cases (averaged over test SVEs and replicates): the fraction of the truly top-K% grains that the predicted ranking also places in its top K. The HCF cases are well above the chance diagonal; the LCF cases are weaker.

engineering target is the per-SVE maximum M_S.

The limitation matters when the per-grain predictions are used to compute the SVE maximum FIP. 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 (as discussed later in Figure 18, 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 (as can be seen in Figure 14), 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.

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 predictions of the previous subsection are 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 post-hoc estimator $\max_g \widehat{M}_g$ obtained by taking the maximum over the per-grain predictions. The third is the empirical-moment EVT bridge, EVTB_t, with subscript t for truth moments, which evaluates the mixture-Gumbel mean of Equation (9) on the held-out true within-SVE active-grain moments of $\{M_g\}_{g\in S}$. Because EVTB_t uses held-out true per-grain peaks, which are the quantities the surrogate is trained to predict, it is a diagnostic of the EVT map and not an estimator usable in practice; we report its R^2 as an upper bound on what the closed-form bridge could achieve if the per-grain distribution were known exactly. The fourth is the predicted-moment EVT bridge, EVTB_p, which evaluates the same formula using moments estimated from the predicted $\{\widehat{M}_g\}_{g\in S}$; unlike EVTB_t, EVTB_p uses only predicted quantities and can be applied in practice. The dedicated head D tests direct learning of the volume extreme, EVTB_t tests the extreme-value

map when the lower-level (grain-level) moments of FIP are known, and $\mathrm{EVTB_p}$ tests whether that map can be reached from the per-grain head alone.

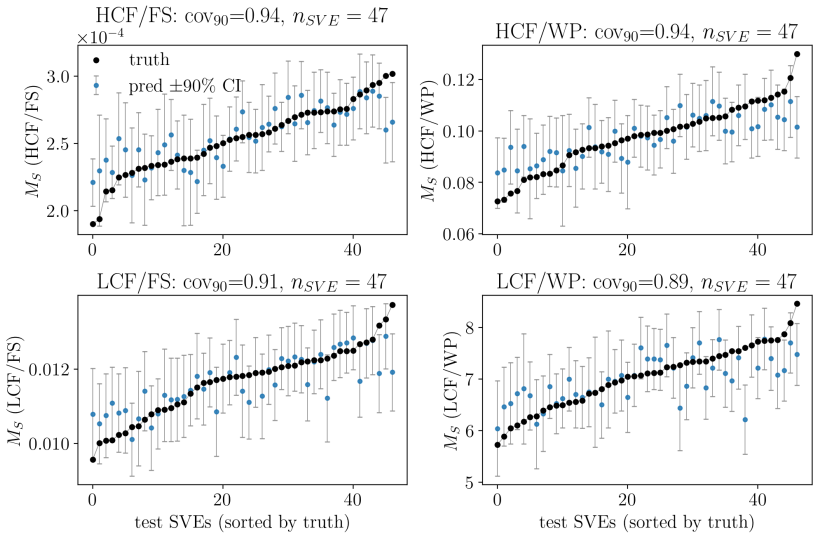


Figure 16: Ordered 90% intervals on the per-SVE maximum from the dedicated graph head, single replicate. Test SVEs are sorted by truth ($n_{\mathrm{SVE}} = 47$); black points show the truth, blue points show the predicted mean with the symmetric interval. Empirical 90% coverages for this replicate are annotated per panel.

The dedicated graph head trained directly on M_S provides informative per-SVE predictions on all four cases. Figure 16 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 aleatoric uncertainty from Section 3.3 and $\hat{\sigma}_e$ is the MCD standard deviation across $K = 30$ stochastic forward passes, which is used to quantify the epistemic uncertainty. The mean empirical 90% coverages across the ten replicates are 0.86 ± 0.05 on HCF/FS, 0.84 ± 0.09 on HCF/WP, 0.83 ± 0.10 on LCF/FS, and 0.88 ± 0.05 on LCF/WP, slightly below the nominal 0.90. 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 18, blue bars).

The EVT bridge gives a complementary route to the same target. The empirical-moment bridge $\mathrm{EVTB_t}$ isolates the adequacy of the closed-form mixture-Gumbel map, while the predicted-moment bridge $\mathrm{EVTB_p}$ adds the error induced by predicting the within-SVE moments from the per-grain head. Figure 17 shows the four estimators on a single replicate, and Figure 18 summarizes their multi-seed R^2 distributions. The two variants separate two questions. When $\mathrm{EVTB_t}$ succeeds but $\mathrm{EVTB_p}$ fails, the bridge itself is informative but the predicted within-SVE moments are not accurate enough to use in practice; when both succeed, the bridge can substitute for or complement the dedicated head.

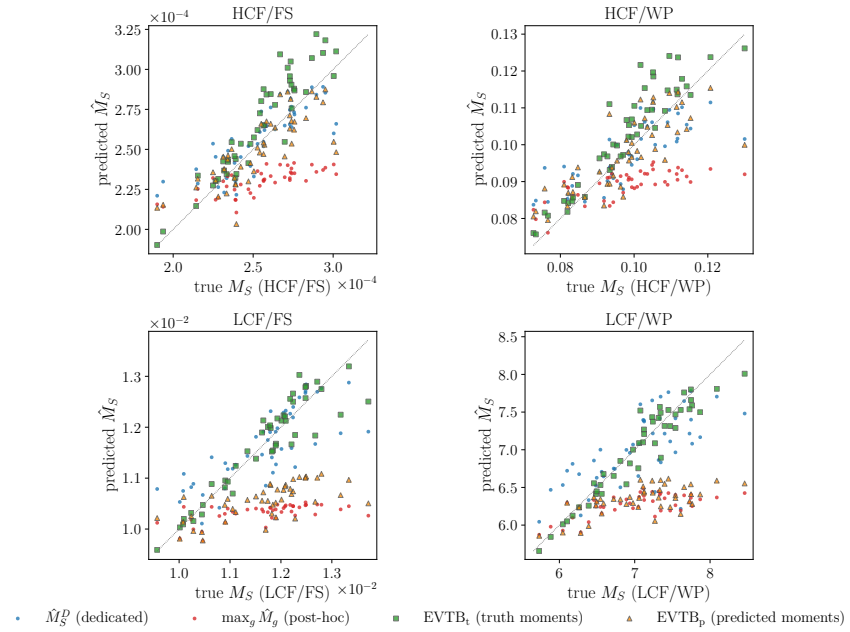


Figure 17: Four estimators of the per-SVE maximum M_S, one panel per case: the dedicated head $\widehat{M}_S^D$, the post-hoc maximum $\max_g \widehat{M}_g$, and the mixture-Gumbel bridge evaluated with truth ($\mathrm{EVTB_t}$) and predicted ($\mathrm{EVTB_p}$) moments. Estimator labels are those of Table 3.

Used as a reference (since the true distribution of $\{M_g\}_{g\in S}$ is not known in practice), 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. These values exceed the dedicated head on three of four cases and match it on the fourth, which marks the attainable ceiling of the closed-form map rather than a practical improvement. The reference bridge attains higher R^2 than D in 9 of 10 replicates on HCF/FS, and in 10 of 10 replicates on both LCF/FS and LCF/WP, while on HCF/WP it leads in 7 of 10 replicates and the two estimators are statistically indistinguishable. The corresponding sign-test and Wilcoxon p-values are 0.022 and 0.006 on HCF/FS, 0.002 and 0.002 on both LCF cases, and 0.34 on HCF/WP.[4] The reference bridge is also more stable across replicates on LCF/FS than the dedicated head, with $\sigma_{R^2}(\mathrm{EVTB_t}) = 0.04$ compared with $\sigma_{R^2}(D) = 0.27$, where σ_{R^2} denotes the across-replicate standard deviation of R^2.

The predicted-moment bridge $\mathrm{EVTB_p}$ behaves differently on the HCF and LCF cases. On both HCF cases, $\mathrm{EVTB_p}$ 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 ($\mathrm{EVTB_p} - \mathrm{EVTB_t} \approx -0.16$ on HCF/FS, -0.05 on HCF/WP). On these cases the per-grain head reproduces the within-SVE active-grain moments accurately

[4]The ten replicates are re-splits of the same finite SVE pool rather than independent datasets, so these paired-test p-values do not account for the re-use of data and should be read as indicative rather than as strict significance levels; the HCF/FS margin in particular is one replicate from non-significance.

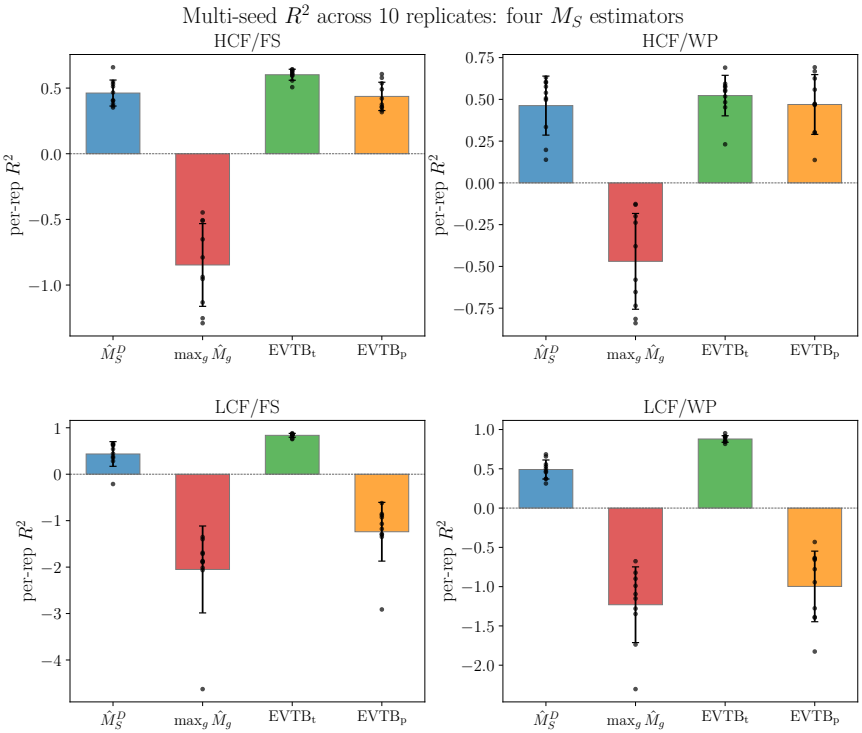


Figure 18: Per-replicate R^2 of the four M_S estimators across ten multi-seed replicates, one panel per case (bars: mean ± std; points: individual replicates). Estimator labels are those of Table 3.

enough that the bridge applied to the predicted moments retains positive R^2. On both LCF cases, EVTB$_p$ 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 (EVTB$_p$ − EVTB$_t$ ≈ −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 active-grain subset, and the predicted-moment bridge therefore fails on those cases even though the empirical-moment bridge provides accurate results.

Because M_S grows systematically with SVE size, the pooled R^2 values above partly reflect this size dependence. Recomputing each estimator's R^2 within size buckets, pooling residuals across the ten replicates and centering the total sum of squares on the per-bucket mean, isolates the skill that is not explained by size. The dedicated head's within-bucket R^2 is approximately zero on all four cases (−0.08 on HCF/FS, −0.09 on HCF/WP, −0.06 on LCF/FS, −0.10 on LCF/WP): it captures the size scaling of the extreme but does not resolve SVE-to-SVE variation in M_S at fixed size. The empirical-moment reference bridge instead retains positive within-bucket R^2 (+0.19, +0.02, +0.65, +0.74 in the same order), so the within-SVE moment structure it consumes carries size-independent information about the extreme that direct regression on M_S does not recover; the predicted-moment bridge loses this within-bucket skill, consistent with its moment-prediction errors. Within-size resolution of M_S is therefore available in principle from the lower-level statistics but is not yet reachable from the per-grain head alone.

These patterns separate the four cases by whether the predicted-moment bridge is usable, 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 is slightly higher than the head on HCF/FS and indistinguishable from it on HCF/WP. On both LCF cases the empirical-moment bridge has the highest R^2, but the predicted-moment bridge is not usable in practice, so the dedicated head remains the practical estimator until the per-grain prediction quality on the active-grain upper tail is improved.

Table 2: 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.

case	observed pattern	dominant interpretation	implied intervention
HCF/FS	grain head strong; empirical-moment bridge higher than the dedicated head in 9 of 10 replicates	empirical-moment bridge is informative; predicted-moment bridge stays positive	no targeted correction
HCF/WP	grain head strong; dedicated head and both bridge variants statistically indistinguishable	balanced reference case; predicted moments accurate enough to match the head	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

Table 2 summarizes the qualitative pattern quantified in Figure 18. Under the mixture working model of Section 2.4, the empirical-moment bridge $\mathrm{EVTB_t}$ returns positive R^2 on all four cases. The cases then separate by predicted-moment-bridge behavior. On both HCF cases, the predicted-moment bridge $\mathrm{EVTB_p}$ also returns positive R^2, because the per-grain head reproduces the active-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 practical estimator on those cases, with $\mathrm{EVTB_t}$ providing the structural reference. The empirical-moment reference bridge attains higher R^2 than the dedicated head on LCF/FS, LCF/WP, and HCF/FS, marking the structural ceiling of the map (indicative paired tests at $n_{\mathrm{rep}} = 10$). On HCF/WP neither bridge variant separates from the dedicated head across the ten replicates (sign-test $p = 0.34$ for $\mathrm{EVTB_t}$ and $p = 1.0$ for $\mathrm{EVTB_p}$ against D), and only the post-hoc maximum fails. The following subsection quantifies these estimator comparisons with R^2 and raw error metrics.

The mixture working model and the bridge method also transfer across alloys. Re-fitting the active-grain fraction and re-evaluating the empirical-moment bridge on the publicly available PRISMS-Fatigue dataset

for Al 7075-T6 under high-cycle loading [42], the bridge returns a positive pooled R^2 under the same per-SVE fitting rule, with an alloy-specific active-grain fraction. Only the HCF/FS case is testable, since PRISMS-Fatigue does not release $\mathrm{FIP}_{\mathrm{WP}}$ or LCF data, but this indicates that the closed-form extreme-value map is not specific to the austenitic-steel dataset studied here.

Table 3 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 fit, with the caveat below; the predicted-moment bridge matches the dedicated head on the HCF cases and degrades strongly on the LCF cases.

Table 3: Extended error statistics for the four M_S estimators across the four FIP cases under the mixture working model, multi-seed ($n_{\mathrm{rep}} = 10$) test-set values (mean ± std). RMSE and MAE are in raw FIP units; MAPE is in percent. The best R^2 per case is in bold. The same estimator symbols are used in Figures 17 and 18.

Case	Estimator	R^2	RMSE	MAE	MAPE (%)
HCF/FS	$\widehat{M}_S^D$ (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
	$\max_g \widehat{M}_g$ (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
	$\mathrm{EVTB_t}$ (truth moments)	$\mathbf{+0.60 \pm 0.04}$	$(1.63 \pm 0.13) \times 10^{-5}$	$(1.29 \pm 0.13) \times 10^{-5}$	4.9 ± 0.45
	$\mathrm{EVTB_p}$ (predicted moments)	$+0.44 \pm 0.11$	$(1.94 \pm 0.22) \times 10^{-5}$	$(1.46 \pm 0.20) \times 10^{-5}$	5.6 ± 0.74
HCF/WP	$\widehat{M}_S^D$ (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
	$\max_g \widehat{M}_g$ (post-hoc)	-0.47 ± 0.29	0.0126 ± 0.00073	0.0107 ± 0.0006	10.4 ± 0.54
	$\mathrm{EVTB_t}$ (truth moments)	$\mathbf{+0.52 \pm 0.12}$	$(7.13 \pm 0.86) \times 10^{-3}$	$(5.59 \pm 0.56) \times 10^{-3}$	5.6 ± 0.68
	$\mathrm{EVTB_p}$ (predicted moments)	$+0.47 \pm 0.18$	$(7.39 \pm 1.28) \times 10^{-3}$	$(5.87 \pm 0.94) \times 10^{-3}$	6.0 ± 1.05
LCF/FS	$\widehat{M}_S^D$ (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
	$\max_g \widehat{M}_g$ (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
	$\mathrm{EVTB_t}$ (truth moments)	$\mathbf{+0.84 \pm 0.04}$	$(3.66 \pm 0.42) \times 10^{-4}$	$(2.67 \pm 0.31) \times 10^{-4}$	2.2 ± 0.26
	$\mathrm{EVTB_p}$ (predicted moments)	-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	$\widehat{M}_S^D$ (dedicated)	$+0.49 \pm 0.12$	0.432 ± 0.067	0.333 ± 0.045	4.8 ± 0.62
	$\max_g \widehat{M}_g$ (post-hoc)	-1.23 ± 0.48	0.899 ± 0.041	0.757 ± 0.045	10.3 ± 0.56
	$\mathrm{EVTB_t}$ (truth moments)	$\mathbf{+0.88 \pm 0.04}$	0.206 ± 0.025	0.145 ± 0.018	2.0 ± 0.24
	$\mathrm{EVTB_p}$ (predicted moments)	-1.00 ± 0.45	0.835 ± 0.143	0.704 ± 0.124	9.6 ± 1.85

4.3. Distribution-aware prediction and interval coverage

This subsection evaluates whether the predictive scales attached to the peak and maximum channels give reliable held-out interval coverage. The surrogate is not only a regressor for M_g and M_S; it is also intended to provide uncertainty intervals.

We assess the heteroscedastic outputs that characterize the aleatoric uncertainty $\hat{\sigma}_{\mathrm{a},g}(\mathbf{x}_g)$ and $\hat{\sigma}_{\mathrm{a},S}(\mathbf{X}_G)$ 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 19 reports this comparison at both hierarchy levels for one representative replicate.

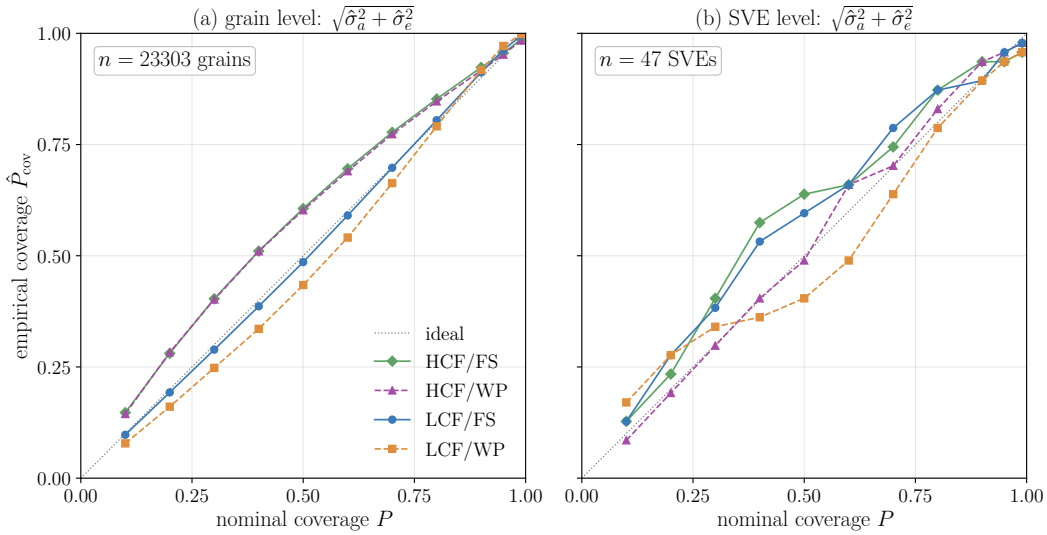


Figure 19: Calibration of the combined aleatoric-plus-epistemic predictive intervals, $\sqrt{\hat{\sigma}_a^2 + \hat{\sigma}_e^2}$. $\hat{\sigma}_a$ is the aleatoric uncertainty from heteroscedastic NLL and $\hat{\sigma}_e$ is the epistemic uncertainty from MCD standard deviation. (a) Grain-level coverage $\hat{P}_{\text{cov}}(P)$ for $n = 23{,}303$ grains; (b) SVE-level coverage for $n = 47$ SVEs. Dotted diagonal is perfect calibration.

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.

The aleatoric/epistemic decomposition of the per-grain uncertainty is shown in Figure 20. The aleatoric term dominates the epistemic term by at least an order of magnitude across all four cases. The per-grain epistemic variance fraction $\hat{\sigma}_e^2/(\hat{\sigma}_{\text{a},g}^2 + \hat{\sigma}_e^2)$ concentrates near zero for the LCF cases, with means of approximately 0.01 for LCF/WP and 0.02 for LCF/FS, whereas values for the HCF cases extend to several tens of percent, with means of approximately 0.11 and 0.12. This pattern does not mean that the LCF predictions are more certain in absolute terms. 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. At the per-SVE level the epistemic fraction is similarly small ($\approx$ 2–3% across cases), so the combined M_S intervals of Section 4.2 are likewise dominated by the learned

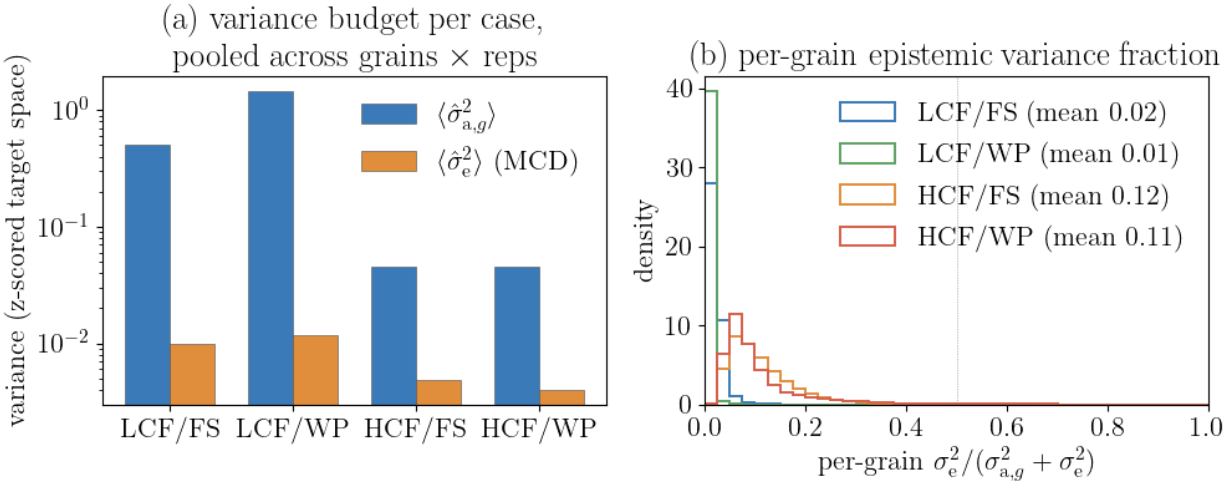


Figure 20: Aleatoric/epistemic variance decomposition of the per-grain predictive variance, pooled across grains and multi-seed replicates. (a) Variance budget per case on a logarithmic axis, (b) Histograms of per-grain epistemic variance fraction.

aleatoric scale.

4.4. Mechanical quantities of interest

The FIP heads test whether the surrogate recovers fatigue-localization statistics. The same CPFEM simulations also produce volume-averaged mechanical quantities that summarize the elastoplastic 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 σ_y, hardening modulus H_p, 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 21 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. The inclusion of these six auxiliary targets does not degrade FIP-head performance (Sections 4.1 and 4.2). These auxiliary R^2 values indicate that the shared representation carries microstructural information relevant to the cyclic constitutive response in general, not only to the FIP-specific signal.

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.

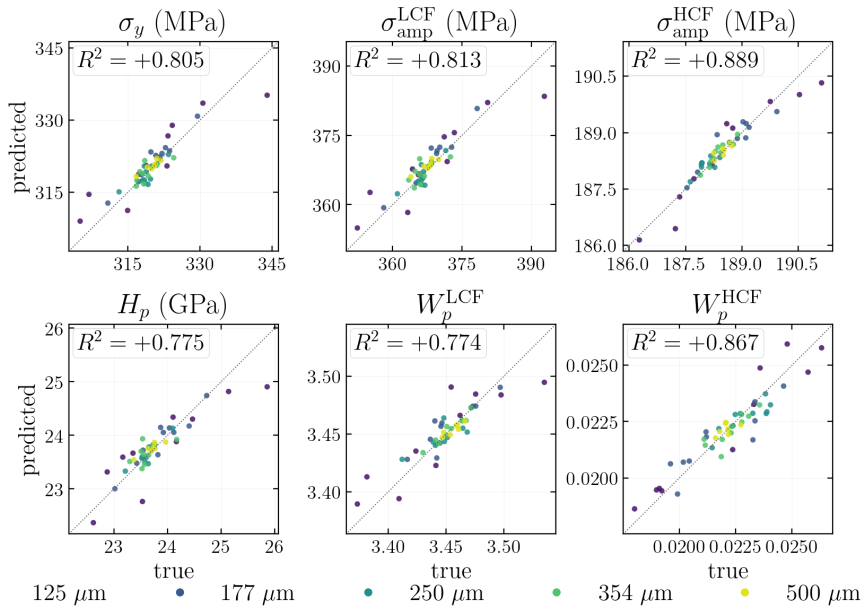


Figure 21: Parity of the six volume-averaged mechanical auxiliary outputs on the held-out test split, colored by SVE edge length L_{SVE}.

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 $(\mu_g^{\ln}, \sigma_g^{\ln})$, per-grain peaks (M_g), within-SVE moments of those peaks $(\mu_{g|S})$, and per-SVE maxima (M_S) 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. The inclusion of these six auxiliary quantities does not degrade (or result in no observable trade-off) the prediction performance of the FIP heads.

Second, the extreme-value bridge is structurally informative when the within-SVE statistics of the active-grain population are known. Under the two-component mixture working model, which combines an effectively inactive Dirac component at zero with an active-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. These values exceed the directly trained graph-level maximum head on three of the four FIP cases and match it on the fourth (indicative paired tests); a size-controlled analysis further shows that the bridge resolves within-size variation in the maximum that the direct head, which captures mainly its size scaling, does not. The bridge therefore identifies the attainable ceiling of the closed-form map rather than a practical replacement for the head: when the active-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 active-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 in the extreme-value bridge itself, but in the under-prediction of the active-grain upper tail by the present per-grain surrogate in the LCF regimes.

The main limitation is the degradation of the predicted-moment bridge in the LCF cases, where the learned per-grain peaks do not preserve the active-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.

Appendix A. Within grain goodness of fit

Within-grain Median Kolmogorov–Smirnov p-values are provided in Table A.1.

Appendix B. Per-grain input feature inventory

Table B.1 enumerates the 32 scalar components of the per-grain input vector $\mathbf{x}_g$ defined in Section 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.

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.

Table A.1: Within-grain shape goodness-of-fit. Median Kolmogorov–Smirnov p-value across the within-grain voxel populations $\{f_v\}_{v \in g}$ 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	lognormal	normal	Gumbel	Weibull
HCF/FS	b0	**0.54**	**0.54**	0.45	0.22
HCF/FS	b1	0.55	0.38	**0.57**	0.31
HCF/FS	b2	**0.70**	0.59	0.47	0.27
HCF/FS	b3	0.44	**0.56**	0.38	0.30
HCF/FS	b4	**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

Table B.1: Per-grain input feature vector $\mathbf{x}_g$, total dimension 32. Block widths sum to $16 + 9 + 6 + 1 = 32$. Quaternion components are normalized; 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

Appendix C. Hyper-parameter optimization

The selected configuration of Section 3.4 was chosen by a multi-objective NSGA-II search (population size 12, 50 trials). Every trial trained the full multi-task model on all four FIP cases; the selection criteria were evaluated on the high-cycle cases, where the per-grain signal leaves the most room for improvement. Two objectives were maximized: the mean per-grain top-decile tail R^2 and the negative mean Gaussian continuous ranked probability score (CRPS), a proper scoring rule that jointly rewards predictive coverage and sharpness. The search space spanned eight knobs: hidden dimension, number of message-passing layers, per-grain and per-graph head widths, dropout rate, activation function, graph-attribute conditioning mode, and learning rate (log-uniform in $[5 \times 10^{-5}, 5 \times 10^{-4}]$). Graph pooling (max), batch size, weight decay, and the multi-task loss recipe were held fixed rather than searched. Five admissibility floors, evaluated on the validation split of the high-cycle cases, pruned degenerate trials: per-grain peak $R^2 \geq 0.85$, heteroscedastic coverage $P_{\mathrm{cov}}(z{=}1.65) \geq 0.85$, mean per-grain top-1 hit rate ≥ 0.10, mean mechanical-auxiliary $R^2 \geq 0.50$, and mean σ-residual Spearman correlation ≥ 0.50; the last two screen out trials with strong point accuracy but degenerate uncertainty. The Pareto front and per-trial metrics are committed alongside the code.

Table C.1 summarizes the selected configuration used for all 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 4.

Table C.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
message-passing operator	SAGE (GraphSAGE)
hidden dimension	256
number of message-passing 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 (11) are set to unity, so the per-grain and per-graph head blocks enter at equal weight (1:1) and the tail emphasis is carried inside the Huber terms rather than by the task weights: in both L_m and L_M the top decile (top 10%) of the target is up-weighted by a factor of 50 relative to the bulk, and the per-case heteroscedastic NLL terms enter at unit weight.

These constants are fixed by the multi-task loss recipe rather than searched in the HPO above (the full configuration JSON is 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.

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. **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. **Massimiliano Lupo Pasini:** Writing - review and editing, Conceptualization, Funding acquisition, Methodology.

Code availability

The open-source HydraGNN library [72], used to generate the numerical results described in this article, can be accessed at the following GitHub repository: https://github.com/ORNL/HydraGNN.

Data availability

The dataset will be made publicly available upon acceptance/publication of this manuscript.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

AI writing assistance (Anthropic Claude) was used for language editing, LaTeX table formatting, and TikZ figure generation (Figs. 5 and 10). After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

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.

References

[1] O. Basquin, The exponential law of endurance tests, in: Proceedings of the American Society for Testing and Materials, Vol. 10, 1910, pp. 625–630.

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

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

[4] S. Suresh, Fatigue of materials, Cambridge University Press, 1998.

[5] D. McDowell, F. Dunne, Microstructure-sensitive computational modeling of fatigue crack formation, International Journal of Fatigue 32 (9) (2010) 1521–1542.

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

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

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

[9] R. J. Asaro, Micromechanics of crystals and polycrystals, Advances in Applied Mechanics 23 (1983) 1–115.

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

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

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

[13] J. Segurado, R. A. Lebensohn, J. LLorca, Computational homogenization of polycrystals, Advances in Applied Mechanics 51 (2018) 1–114.

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

[15] M. Knezevic, B. Drach, M. Ardeljan, I. J. Beyerlein, Three dimensional predictions of grain scale plasticity and grain boundaries using crystal plasticity finite element models, Computer Methods in Applied Mechanics and Engineering 277 (2014) 239–259.

[16] T. Zhang, J. Jiang, B. Britton, B. Shollock, F. Dunne, 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) (2016).

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

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

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

[20] A. D. Anto, R. Fleishel, S. TerMaath, R. Abedi, Size dependency of elastic and plastic properties of metallic polycrystals using statistical volume elements, Applied Sciences 14 (18) (2024) 8207.

[21] A. D. Anto, S. TerMaath, R. Abedi, Size dependence of homogenized monotonic and cyclic responses and extreme-value fatigue indicators in polycrystalline microstructures, Mechanics of Materials (2026) 105698.

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

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

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

[25] A. Manonukul, F. Dunne, 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) (2004) 1881–1903.

[26] J. Hochhalter, D. Littlewood, R. Christ Jr, M. Veilleux, J. Bozek, A. Ingraffea, A. Maniatty, 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) (2010) 045004.

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

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

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

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

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

[32] A. Cerrone, C. Stein, R. Pokharel, C. Hefferan, J. Lind, H. Tucker, R. Suter, A. Rollett, A. Ingraffea, 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) (2015) 035006.

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

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

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

[36] R. A. Fisher, L. H. C. Tippett, Limiting forms of the frequency distribution of the largest or smallest member of a sample, in: Mathematical proceedings of the Cambridge philosophical society, Vol. 24, Cambridge University Press, 1928, pp. 180–190.

[37] E. J. Gumbel, Statistics of Extremes, Columbia University Press, 1958.

[38] S. Coles, J. Bawa, L. Trenner, P. Dorazio, An introduction to statistical modeling of extreme values, Vol. 208, Springer, 2001.

[39] P. Embrechts, C. Klüppelberg, T. Mikosch, Modelling extremal events: for insurance and finance, Vol. 33, Springer Science & Business Media, 2013.

[40] M. R. Leadbetter, G. Lindgren, H. Rootzén, Extremes and related properties of random sequences and processes, Springer Science & Business Media, 2012.

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

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

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

[44] M. D. Sangid, The physics of fatigue crack initiation, International Journal of Fatigue 57 (2013) 58–72.

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

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

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

[48] D. A. Nix, A. S. Weigend, Estimating the mean and variance of the target probability distribution, in: Proceedings of 1994 IEEE International Conference on Neural Networks (ICNN'94), Vol. 1, IEEE, 1994, pp. 55–60.

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

[50] Y. Gal, Z. Ghahramani, Dropout as a bayesian approximation: Representing model uncertainty in deep learning, in: International Conference on Machine Learning, PMLR, 2016, pp. 1050–1059.

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

[52] Y. Gal, J. Hron, A. Kendall, Concrete dropout, Advances in Neural Information Processing Systems 30 (2017).

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

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

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

[56] W. Hamilton, Z. Ying, J. Leskovec, Inductive representation learning on large graphs, Advances in Neural Information Processing Systems 30 (2017).

[57] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, G. E. Dahl, Neural message passing for quantum chemistry, in: International Conference on Machine Learning, Pmlr, 2017, pp. 1263–1272.

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

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

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

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

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

[63] T. Xie, J. C. Grossman, Hierarchical visualization of materials space with graph convolutional neural networks, The Journal of chemical physics 149 (17) (2018).

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

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

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

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

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

[69] E. Caliskan, A. D. Anto, M. Lupo Pasini, S. TerMaath, R. Abedi, Multitask graph neural networks for elastoplastic response prediction in dual-phase polycrystals, Journal of Materials Science: Materials Theory 10 (1) (2026) 8.

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

[71] P. Hall, On the rate of convergence of normal extremes, Journal of Applied Probability 16 (2) (1979) 433–439.

[72] M. L. Pasini, J. Y. Choi, K. Mehta, P. Zhang, R. Weaver, R. Messerly, A. Chowdhury, A. Raman, A. M. Aji, User manual-hydragnn v5. 0: Distributed implementation of multi-tasking graph neural networks (2026).