Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

MELT-3D: Zero-shot cross-material prediction of three-dimensional full-field crystal-plasticity evolution with a latent transformer

Reeju Pokharel

reeju@lanl.gov

Los Alamos National Laboratory

Alexander Scheinker

Los Alamos National Laboratory

Article

Keywords:

Posted Date: November 6th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11299724/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

MELT-3D: Zero-shot cross-material prediction of three-dimensional full-field crystal-plasticity evolution with a latent transformer

Reeju Pokharel[1,*] and Alexander Scheinker[1]

[1]Los Alamos National Laboratory, Los Alamos, NM, 87544, USA
[*]Corresponding author: reeju@lanl.gov

Abstract

Full-field crystal-plasticity surrogates require simulations from each target material. We demonstrate zero-shot prediction of the complete three-dimensional material state of an unseen material, including all stress, elastic-strain, and plastic-strain tensor components and crystallographic orientation. Prediction requires only the initial full-field state and constitutive descriptors, with no target-material trajectories for training, validation, fine-tuning, or normalization. MELT-3D (Microstructure Evolution via Latent Transformer) combines a variational autoencoder, a non-autoregressive conditioned transformer, and a hybrid mechanics–machine-learning model that predicts field statistics for normalization and reconstruction. Trained on Cu and CoCrFeMnSi and evaluated on CoCrFeMnNi, MELT-3D achieves accuracy comparable to held-out training-material microstructures. For example, its normalized voxelwise stress error is 0.196, compared with 0.184 and 0.192 for the held-out materials. Volume averaging the predicted axial-stress fields gives a 1.0% end-state macroscopic stress error. Error decomposition shows that encoding and decoding account for three-quarters of the stress-field error and most orientation error, while transformer-based latent-state prediction has smaller overall contribution.

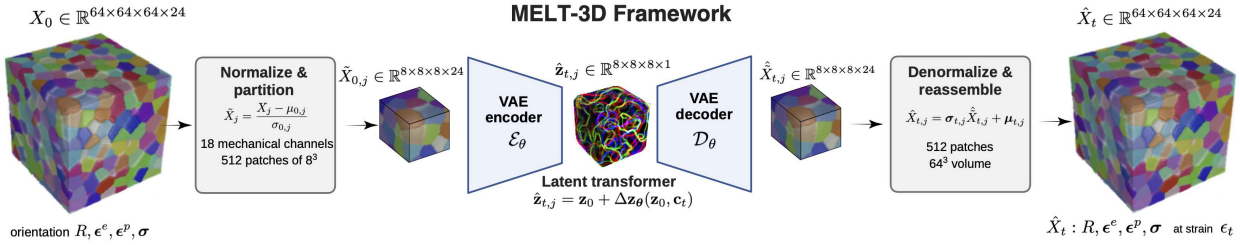


Figure 1: MELT-3D (Microstructure Evolution via Latent Transformer) framework for predicting the evolution of three-dimensional microstructures and their associated micromechanical fields during deformation. A variational autoencoder (VAE) compresses the voxel-resolved crystallographic orientation and micromechanical tensor fields into a spatially structured latent representation. A residual transformer then predicts the latent representation at a specified deformation state.

1 Introduction

The mechanical response of a polycrystalline material emerges from interactions among neighboring grains and depends on their crystallographic orientations, grain morphology, elastic anisotropy,

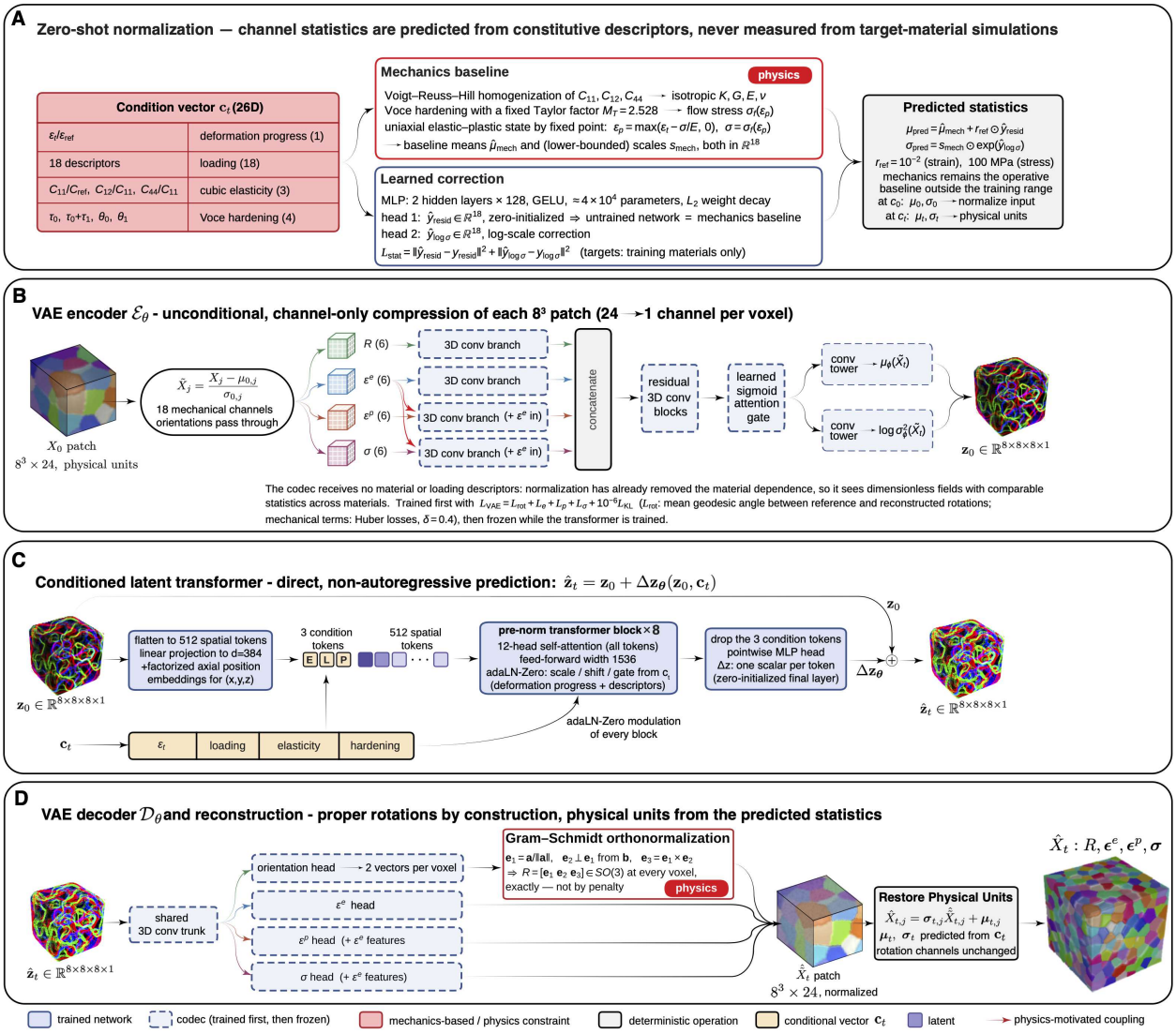


Figure 2: **Overall setup of the approach.** (A) General normalization. (B) VAE encoder. (C) Conditional latent transformer. (D) VAE decoder.

and activated deformation mechanisms such as crystallographic slip and/or twinning. These interactions generate heterogeneous stress, strain, and lattice-rotation fields even under macroscopically homogeneous loading. Understanding which deformation mechanisms are activated and how they evolve is important because they influence texture development, stress concentrations and strain localization, which govern yielding, damage nucleation, fatigue, and material failure. Full-field crystal-plasticity methods including crystal-plasticity finite-element and spectral fast Fourier transform formulations describe these processes by solving the equilibrium and constitutive equations throughout a three-dimensional microstructure [1, 2]. Their computational cost, however, limits their repeated application in uncertainty quantification, materials screening, inverse problems, experimental planning, and real-time analysis and feedback.

Data-driven surrogate models offer a complementary approach by learning the relationship among microstructure, loading, and mechanical response from high-fidelity simulations. Several approaches efficiently predict homogenized behavior [3–5], but they do not recover the coupled

voxel-resolved fields required to study localization and intragranular heterogeneity. Other work has focused on spatially resolved responses. The Materials Knowledge Systems framework established calibrated localization links to predict microscale fields under imposed macroscale loading [6–8]. Convolutional and operator-learning models have been used to predict stress fields in two-dimensional polycrystals undergoing elastic–plastic deformation [9], and three-dimensional surrogates have predicted spatially resolved crystallographic orientation and mechanical tensor fields during uniaxial deformation [10–12]. Adaptive 3D CNNs have also been developed for mapping X-ray diffraction intensities to the corresponding 3D electron densities to track them under dynamic time-varying conditions [13, 14]. Physics-encoded neural operators have shown that constraints such as stress equilibrium can be embedded directly into a field-prediction architecture [15]. Recent architectures have expanded the range of microstructures and loading conditions that can be addressed by a single model [16, 17]. However, in these studies, the constitutive behavior remains fixed. Applying a trained model to a material with different elastic constants or hardening parameters still requires new simulations and model adaptation for that material. This limitation was evident in our previous surrogate for the evolution of the full elastic-strain tensor in three-dimensional polycrystals [18]. A recurrent and recursive convolutional network trained on Cu did not transfer with sufficient accuracy to other cubic material systems, even when their microstructural statistics were comparable. Transfer learning reduced the amount of new training data required, but it still depended on target-material evolution trajectories and material-specific optimization.

In this work we examine whether the subsequent deformation of a target material can be predicted from its initial full-field state and constitutive properties without using any part of its evolution trajectory for model training, validation, fine-tuning, or empirical preprocessing. We call this task *zero-shot material evolution*. The model identifies a material through physical descriptors comprising its cubic elastic constants and Voce hardening parameters and receives one initial state but no material-evolution trajectory. This setup is analogous to direct numerical simulation, which begins from a specified initial state, material parameters, and boundary conditions.

We introduce MELT-3D (Microstructure Evolution via Latent Transformer), a surrogate model for predicting the coupled evolution of crystallographic orientation, elastic strain, plastic strain, and stress in voxel-resolved polycrystals. MELT-3D predicts the latent state at any target deformation step directly from the initial microstructure and continuous physical descriptors. A high level overview of the approach is shown in Fig. 1 and a more detailed schematic is shown in Fig. 2. This non-autoregressive approach avoids sequential error accumulation across deformation steps. A hybrid mechanics–machine-learning model predicts the field statistics required for normalization and reconstruction, eliminating the need for categorical material labels or target-material normalization data. The MELT-3D surrogate is trained on Cu and CoCrFeMnSi and evaluated on CoCrFeMnNi without material-specific fine-tuning or deformation trajectories. Its accuracy for CoCrFeMnNi is comparable to that obtained for held-out microstructures of the training materials. Although the macroscopic response of CoCrFeMnNi lies broadly between those of Cu and CoCrFeMnSi, several of its Voce hardening descriptors fall outside the training ranges. These results provide a proof-of-concept demonstration of zero-shot material evolution within the investigated family of FCC constitutive responses.

2 Results

The MELT-3D framework, shown schematically in Fig. 1, predicts the coupled three-dimensional state of a deforming polycrystal from its initial state and a set of constitutive and loading descriptors. The voxel-resolved state comprises crystallographic orientation, elastic strain, plastic strain,

and stress, represented by 24 field channels on a $64 \times 64 \times 64$ grid. At inference, MELT-3D takes the initial full-field state X_0 and a condition vector c_t describing the constitutive properties, applied loading, and target deformation state. The prediction proceeds through five stages.

1. An auxiliary statistics model predicts the channel means and standard deviations used to normalize the input state and transform the predicted target state back to physical units. The statistics for the initial and target states are obtained from their corresponding condition vectors using a mechanics-based baseline with a learned correction. No empirical field statistics from a target-material deformation trajectory are required.

2. The normalized initial state is partitioned into 8^3 patches. A variational autoencoder, referred to as the codec, encodes each patch into a spatially structured latent field z_0, reducing the representation at each voxel from 24 variables to one latent variable.

3. A transformer conditioned on c_t predicts the latent state at the specified deformation step as a residual relative to the initial latent state:

$$\hat{z}_t = z_0 + \Delta z_\theta(z_0, c_t).$$

   Each target state is predicted directly from z_0 and c_t, independently of the predictions at other deformation steps. The model therefore does not advance recursively through the deformation trajectory.

4. The frozen decoder reconstructs the normalized target field from $\hat{z}_t$. The predicted target-state statistics then transform the 18 micromechanical channels back to physical units.

5. The 512 predicted patches are returned to their original spatial locations and reassembled to reconstruct the complete 64^3 volume.

We evaluate MELT-3D on held-out Cu and CoCrFeMnSi (HEA-Si) microstructures and on CoCrFeMnNi (HEA-Ni). No HEA-Ni data were used to train or validate any learned component of the framework; HEA-Ni therefore serves as the zero-shot test material. Although its macroscopic stress–strain response lies between those of Cu and HEA-Si, its constitutive descriptors are not all within the corresponding training ranges. The three elastic descriptors and the initial slip resistance τ_0 lie within their componentwise training ranges, whereas $\tau_0 + \tau_1$, θ_0, and θ_1 lie outside them (Supplementary Table S2). HEA-Ni therefore represents an unseen point in the multidimensional constitutive-parameter space, with elastic descriptors that interpolate within the training ranges and several hardening descriptors that require extrapolation beyond them.

2.1 Macroscopic stress–strain response

MELT-3D predicts spatially resolved fields and macroscopic responses can be calculated by volume averaging the local predictions. Figure 3 compares the resulting stress–strain responses with the corresponding crystal-plasticity results. The volume-averaged axial stress obtained from the complete prediction pipeline closely follows the reference response through the elastic regime, the elastic–plastic transition, and subsequent strain hardening. At the final deformation state, the relative errors are 1.01% for Cu, 0.74% for HEA-Si, and 1.01% for the zero-shot HEA-Ni material. The mean relative errors over the complete loading trajectories are 0.88%, 0.88%, and 0.69%, respectively. The corresponding maximum relative errors are 5.14%, 7.49%, and 10.56%. These maxima occur during the early elastic regime, where the reference axial stress is small, so even a small absolute discrepancy produces a comparatively large relative error.

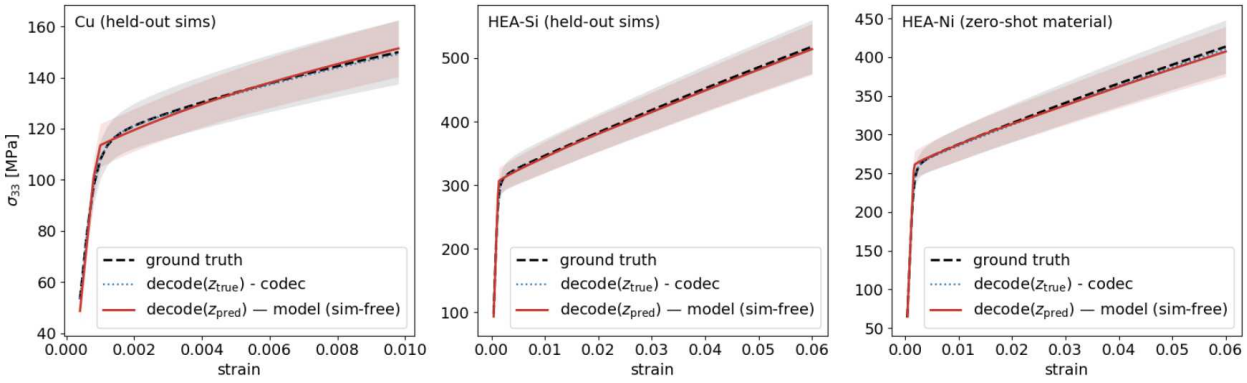


Figure 3: Macroscopic stress–strain responses for held-out Cu and HEA-Si microstructures and zero-shot HEA-Ni microstructures. The volume-averaged axial stress σ_{33} from the crystal-plasticity simulations (black dashed line) is compared with the codec reconstruction $D_\psi(z_t)$ (blue dotted line) and the complete model prediction $D_\psi(\hat{z}_t)$ (red solid line). $D_\psi(z_t)$ is obtained by encoding and decoding the reference crystal-plasticity state and therefore isolates the error introduced by latent compression. In contrast, $D_\psi(\hat{z}_t)$ includes errors from both latent compression and transformer prediction. The shaded regions indicate ± 1 standard deviation across the mean axial stresses of the 512 nonoverlapping 8^3-voxel patches and therefore represent patch-scale spatial heterogeneity rather than predictive uncertainty. Results are averaged over six Cu, six HEA-Si, and ten HEA-Ni microstructures.

Although no HEA-Ni deformation trajectories were used for training, MELT-3D predicts its macroscopic response with accuracy comparable to that obtained for the held-out Cu and HEA-Si microstructures. The model also reproduces the magnitude of the patch-scale stress heterogeneity. At the final deformation state, the standard deviations across the 512 patch-averaged σ_{33} values are 12.5 and 11.1 MPa for the reference and predicted Cu fields, respectively; 42.1 and 40.1 MPa for HEA-Si; and 34.5 and 32.2 MPa for HEA-Ni. Thus, the model reproduces both the mean axial stress and the magnitude of its spatial variation with comparable accuracy for the zero-shot material and the two materials represented during training.

The mechanical fields are transformed back to physical units using channel statistics predicted from the constitutive descriptors and deformation state (Methods). An oracle diagnostic isolated the error introduced by this procedure by replacing the predicted statistics with empirical statistics calculated from the target crystal-plasticity fields. These empirical statistics were used only for the diagnostic comparison. Relative to the oracle results, using the predicted statistics changed the normalized σ_{33} error by 0.0059 for the transformer-predicted latent states and by 0.0046 for the reference latent states. The small difference between these changes, 0.0013, indicates only a weak interaction between the latent-state prediction and statistics prediction errors. The zero-shot normalization therefore introduces little additional error while removing the need for empirical HEA-Ni field statistics during inference.

2.2 Voxel-resolved mechanical fields

Macroscopic agreement alone does not quantify the accuracy of the underlying voxel-resolved fields. Characterization of the local fields are needed to describe stress and strain localization, intragranular heterogeneity, and the conditions associated with damage initiation. We therefore, evaluate all six components of the stress, elastic-strain, and plastic-strain tensors. Figure 4 compares the

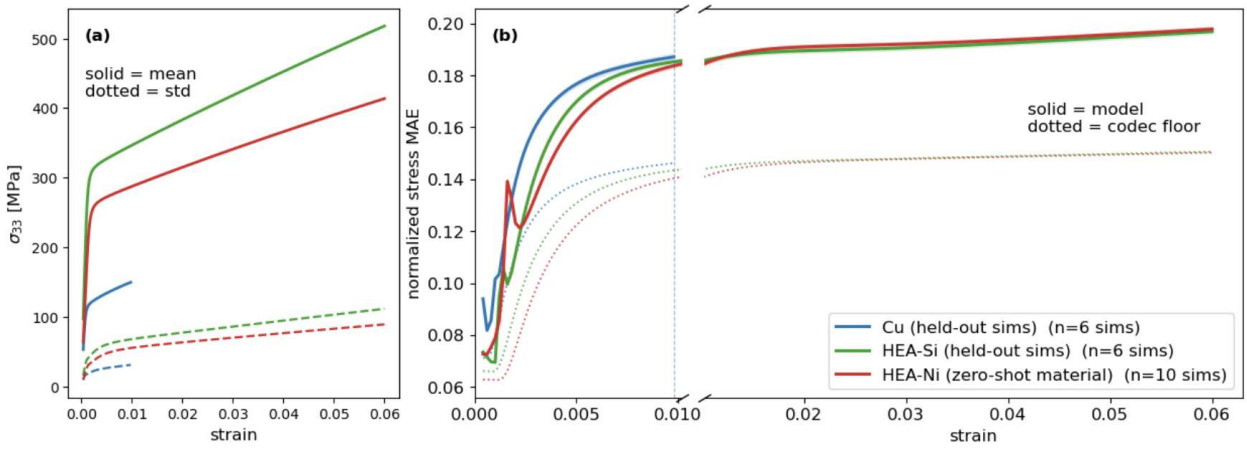


Figure 4: Evolution of the axial stress response and stress-field prediction error. (a) Volume-averaged axial stress σ_{33} (solid lines) and the standard deviation of the patch-averaged axial stresses across the 512 nonoverlapping 8^3-voxel patches (dashed lines). (b) Normalized mean absolute error of the predicted stress fields (solid lines) and the corresponding codec reconstruction error (dotted lines). The codec error is obtained by encoding and decoding the reference fields without transformer prediction and therefore represents the error introduced by latent compression. Curves show averages over six Cu, six HEA-Si, and ten HEA-Ni microstructures; shaded regions in panel (b) indicate ± 1 standard deviation across microstructures. The break in the horizontal axis accommodates the different deformation ranges of Cu (0–1% strain) and the two high-entropy alloys (0–6% strain).

normalized stress-field error with the corresponding macroscopic response. Normalizing the error by the spatial variation of each reference stress component enables comparison among materials with different absolute stress levels.

For all three materials, the normalized stress-field error increases through the elastic regime and the elastic–plastic transition before approaching a plateau of approximately 0.19–0.20. The HEA-Ni error remains within the range obtained for the two materials represented during training. The small variation among independent microstructures further shows that the zero-shot accuracy is consistent across the HEA-Ni test set and is not associated with a particular microstructure.

The normalized error reaches a plateau even as the mean stress and patch-to-patch stress variation continue to evolve [Fig. 4(a)]. This behavior indicates that the absolute prediction error scales with the characteristic spatial variation of the stress field rather than accumulating progressively with deformation. The dotted curves in Fig. 4(b) show the corresponding codec reconstruction error obtained without transformer prediction. This contribution is approximately 0.15 and is examined further in Section 2.4.

Table 1 extends this comparison to the elastic-strain and plastic-strain fields over the second half of each trajectory. The normalized errors are 0.184–0.196 for stress, 0.211–0.219 for elastic strain, and 0.243–0.251 for plastic strain. The ordering is consistent across all three materials: stress is reproduced most accurately, elastic strain less so, and plastic strain least accurately. Plastic strain is the most spatially heterogeneous of the three fields and develops the sharpest localized features, so this ordering is consistent with a representation whose fidelity decreases as the fine-scale content of the field increases. For every field group, the zero-shot HEA-Ni value lies within or immediately adjacent to the range spanned by held-out Cu and HEA-Si, with the largest relative gap occurring for the volume-averaged axial stress (1.29% against 0.50–0.76%).

Table 1: **Accuracy of the decoded fields**, averaged over the second half of each trajectory (ε beyond 50% of the sampled deformation range; $n = 5$ states per material). Uncertainties are one standard deviation over these states. The relative error in mean σ_{33} is based on the volume-averaged axial stress; it differs from the end-state values quoted above because it is averaged over the second half of the trajectory. The normalized mean absolute error (MAE) is the voxelwise MAE divided by the spatial standard deviation of the corresponding reference field and is dimensionless.

Quantity	Cu	HEA-Si	HEA-Ni
Relative error in mean σ_{33} (%)	0.4985 ± 0.3620	0.7628 ± 0.0106	1.2910 ± 0.2040
Normalized MAE, stress	0.1838 ± 0.0031	0.1924 ± 0.0021	0.1959 ± 0.0020
Normalized MAE, elastic strain	0.2107 ± 0.0035	0.2192 ± 0.0013	0.2180 ± 0.0013
Normalized MAE, plastic strain	0.2433 ± 0.0018	0.2505 ± 0.0012	0.2472 ± 0.0012
Normalized MAE, rotation representation	0.0954 ± 0.0002	0.1018 ± 0.0018	0.1017 ± 0.0016

Figure 5 compares representative spatial fields from the crystal-plasticity simulations with their codec reconstructions and complete model predictions. The model reproduces the principal spatial variations in the elastic-strain, plastic-strain, and stress fields, including the grain-scale stress partitioning and the banded plastic-strain features that develop between neighboring grains, although discrepancies remain visible in the difference maps. The two difference columns separate the information lost during compression and reconstruction from the effect of replacing the encoded target state z_t with the transformer-predicted state $\hat{z}_t$. These two contributions have visibly different spatial character: the codec error appears fine-grained and locally uncorrelated, whereas the error associated with latent-state prediction extends more coherently over neighboring voxels. Because two error fields can have similar mean absolute magnitudes while differing substantially in spatial organization, this distinction is quantified in Section 2.4.

2.3 Crystallographic orientation

Crystallographic orientation requires separate treatment because componentwise error measures on the stored rotation components do not correspond to angular differences between crystal orientations. For the complete predictions, the normalized componentwise MAE of the six-dimensional rotation representation was 0.095–0.102 (Table 1), lower than the normalized errors of the mechanical fields. This quantity cannot be interpreted directly as an angular error, and agreement between componentwise variances does not establish voxelwise crystallographic agreement. We therefore evaluate orientation using disorientation reduced over the 24 proper rotations of the cubic point group.

Figure 6 and Table 2 report the disorientation between the reference orientations and their encode–decode reconstructions. The median disorientation was 4.0–4.3°, so a typical reconstructed voxel lies within approximately 4° of its reference orientation. The 90th-percentile disorientation was 13.6–14.3°, and approximately 4.2% of the volume exceeded 30°, producing 99th-percentile disorientations of approximately 48°. The reconstruction therefore recovers most of the orientation field with small angular error while producing a pronounced high-angle tail. The reconstructed kernel average misorientation (KAM) was 1.31–1.44 times the reference value under the same neighborhood definition, indicating greater local orientation variation in the reconstructions than in the reference fields. This behavior is not apparent from the componentwise dispersion ratio, which was 1.0000 to the reported precision, and illustrates why crystallographic metrics are required: the reconstruction preserves the global orientation distribution while roughening it locally.

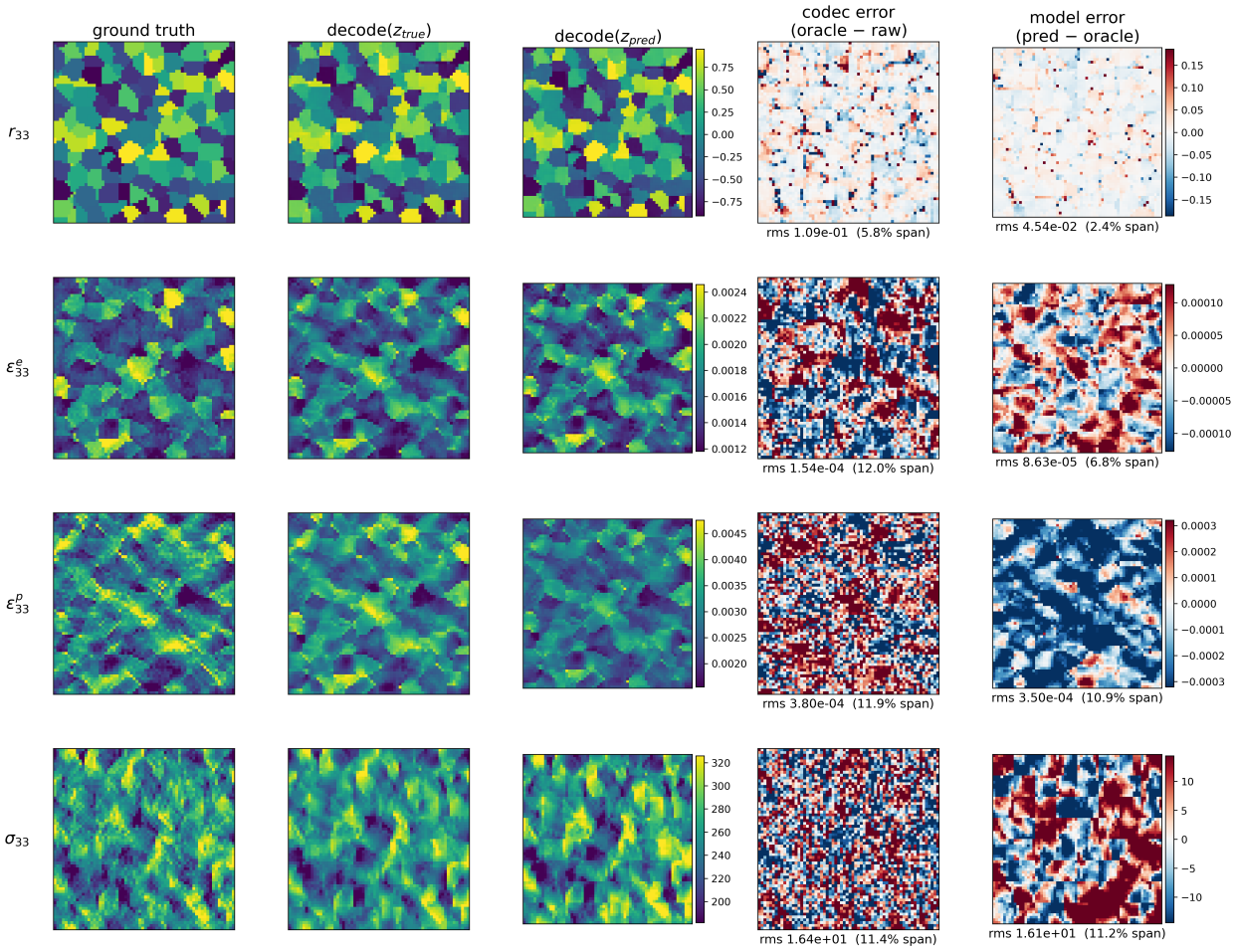


Figure 5: Representative mid-plane slices of the predicted and reference fields. Rows show one component from each field group: the rotation-matrix component R_{33}, elastic strain ε^e_{33}, plastic strain ε^p_{33}, and stress σ_{33}. Columns show the reference crystal-plasticity field, the codec reconstruction decode(z_{true}), the complete model prediction decode(z_{pred}), the codec error, and the additional error associated with latent-state prediction. For each field, the two difference maps use the same color range, corresponding to $\pm 10\%$ of the range of the reference field.

Median and upper-percentile statistics must be reported together. For HEA-Ni at $\varepsilon = 0.06$, the end-to-end distribution has a median of 5.10°, a mean of 7.82°, and a 90th percentile of 14.96°; the corresponding codec values are 4.38°, 7.19°, and 14.38°. The difference between the mean and median, together with the substantially larger 90th percentile, shows that most voxels have lower disorientation than the high-error tail. The median characterizes typical voxelwise performance, while the upper percentiles characterize a smaller high-disorientation population, which is analyzed in Supplementary Section S5.4.

2.4 Compression dominates the error and is nearly material-independent

The preceding sections show that the zero-shot material has errors comparable to those of held-out microstructures from the training materials. To identify the sources of these errors, we separate the contributions from latent compression and latent prediction. The reference latent z_t and the

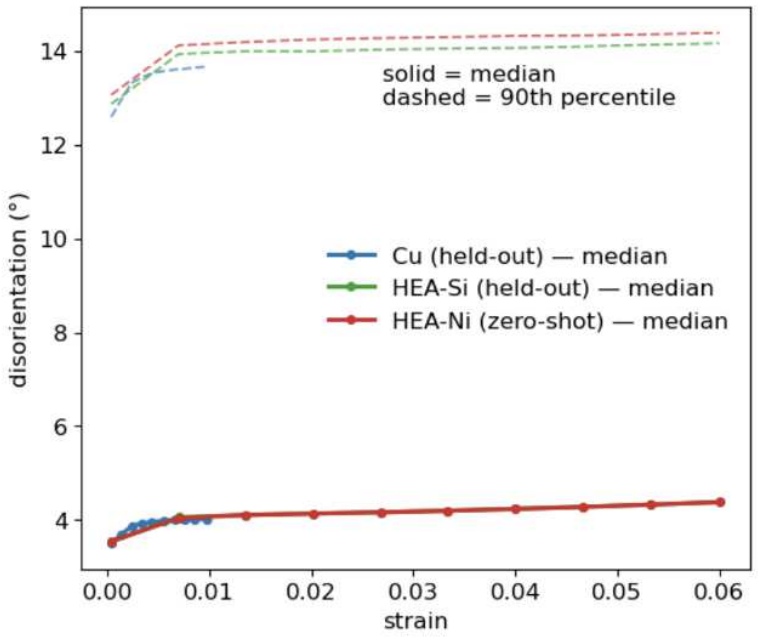


Figure 6: **Cubic-symmetry-reduced disorientation between the reference orientation field and its shared-codec reconstruction**, reported as the median (solid lines) and 90th percentile (dashed lines). Cu and HEA-Si are held-out test materials, whereas HEA-Ni is evaluated zero-shot.

predicted latent $\hat{z}_t$ are passed through the same frozen decoder, giving $X_t^{\text{codec}} = D_\psi(z_t)$ and $X_t^{\text{pred}} = D_\psi(\hat{z}_t)$. Comparing these outputs isolates the contribution of latent prediction under the same decoder, although the comparison still includes the nonlinear response of the decoder to differences in the latent states.

The transformer under predicts the magnitude of latent evolution. The predicted latent displacement follows the rapid increase and subsequent plateau observed in the encoded reference trajectories, but its magnitude is systematically smaller (Fig. 7). Over the second half of the deformation range, the ratio $\|z_{\text{pred}} - z_0\|/\|z_{\text{true}} - z_0\|$ is 0.805–0.813 across the three materials (Table 3). The transformer therefore reproduces approximately 81% of the reference displacement from the initial latent state. The nearly identical ratios for held-out Cu and HEA-Si and zero-shot HEA-Ni show that this contraction is not amplified for the unseen material.

The latent dispersion ratio is approximately 0.994 and the correlation between predicted and reference latent states is 0.991–0.992. These values do not contradict the drift result because they quantify different properties: the dispersion ratio measures the overall spread of the latent variables and the correlation their statistical association, whereas the drift ratio measures the magnitude of the change from the supplied initial state. The predicted latent states therefore retain a similar overall organization while underestimating the magnitude of their evolution away from the initial condition. The approximately 19% contraction does not correspond to a 19% error in the decoded physical response: the relative error in volume-averaged σ_{33} over the same interval is 0.50–1.29%, and the normalized voxelwise errors are 0.184–0.251 depending on field group (Table 1). The latent drift ratio and the decoded-field errors use different definitions and normalizations and should not be interpreted as proportional measures. Residual-direction controls (Supplementary Section S4.3)

Table 2: Crystallographic metrics for the shared-codec orientation reconstructions, averaged over the second half of each trajectory (ε beyond 50% of the sampled deformation range; $n = 5$ states per material). Uncertainties are one standard deviation over these states. Disorientation was minimized over the 24 proper rotations of the cubic point group. The KAM ratio is the reconstructed value divided by the corresponding reference value.

Quantity	Cu	HEA-Si	HEA-Ni
Dispersion ratio std(recon)/std(raw)	1.0000 ± 0.0000	1.0000 ± 0.0000	1.0000 ± 0.0000
Median disorientation (°)	4.0062 ± 0.0140	4.2854 ± 0.0719	4.2852 ± 0.0726
Mean disorientation (°)	6.8389 ± 0.0138	7.0865 ± 0.0596	7.1123 ± 0.0580
90th pct. disorientation (°)	13.6185 ± 0.0356	14.0969 ± 0.0469	14.3337 ± 0.0341
99th pct. disorientation (°)	48.2227 ± 0.0518	48.3334 ± 0.0247	48.1500 ± 0.0348
Voxels $> 10°$ (%)	14.8299 ± 0.0569	15.5578 ± 0.1400	15.8992 ± 0.1492
Voxels $> 30°$ (%)	4.1750 ± 0.0051	4.1866 ± 0.0135	4.1547 ± 0.0050
KAM ratio, recon/reference	1.4420 ± 0.0056	1.3240 ± 0.0322	1.3140 ± 0.0315

Table 3: **Latent-space metrics** averaged over the second half of each trajectory (ε beyond 50% of the sampled deformation range; $n = 5$ states per material). Uncertainties are one standard deviation over these states. The drift ratio compares the magnitudes of the predicted and reference displacements from the initial latent state and is not itself a field-error percentage.

Quantity	Cu	HEA-Si	HEA-Ni
Drift ratio $\|z_{\text{pred}} - z_0\|/\|z_{\text{true}} - z_0\|$	0.8061 ± 0.0016	0.8131 ± 0.0033	0.8049 ± 0.0070
Dispersion ratio $\text{std}(z_{\text{pred}})/\text{std}(z_{\text{true}})$	0.9942 ± 0.0001	0.9941 ± 0.0001	0.9944 ± 0.0001
Correlation $\text{corr}(z_{\text{pred}}, z_{\text{true}})$	0.9924 ± 0.0002	0.9915 ± 0.0001	0.9912 ± 0.0001

show that the observed latent residual produces a decoded-field perturbation comparable to those produced by norm-matched control residuals, providing no evidence that the residual is concentrated in an unusually decoder-insensitive latent direction.

Compression establishes most of the delivered error. For the stress field, the codec-only reconstruction error in Fig. 4(b) plateaus at approximately 0.15, against an end-to-end plateau of 0.19–0.20. Roughly three-quarters of the normalized stress error is therefore already present after encoding and decoding the reference field, before any prediction is introduced. The remaining difference between the codec-only and complete-pipeline curves is similar for Cu, HEA-Si, and HEA-Ni, so the additional error contributed by latent-state prediction is comparable for the zero-shot material and the materials represented during training.

The orientation results give the same picture (Fig. 8, Table 4). The codec has the larger and more stable error contribution: its median disorientation is approximately 3.5° at the beginning of loading and reaches around 4.0° by $\varepsilon \approx 0.007$, after which it increases more gradually to 4.38° at $\varepsilon = 0.06$ for both HEA-Si and HEA-Ni (Fig. 8a). At matched strain, the HEA-Si and HEA-Ni codec medians differ by no more than approximately 0.08°. The codec therefore introduces a nearly material-independent orientation-reconstruction error, with only a modest dependence on deformation state over the investigated range. The decoded discrepancy introduced by latent-state prediction (Fig. 8b) rises through the early elastic–plastic transition and changes more gradually beyond $\varepsilon \approx 0.007$, reaching median values of 3.01° for HEA-Si and 3.10° for HEA-Ni at $\varepsilon = 0.06$.

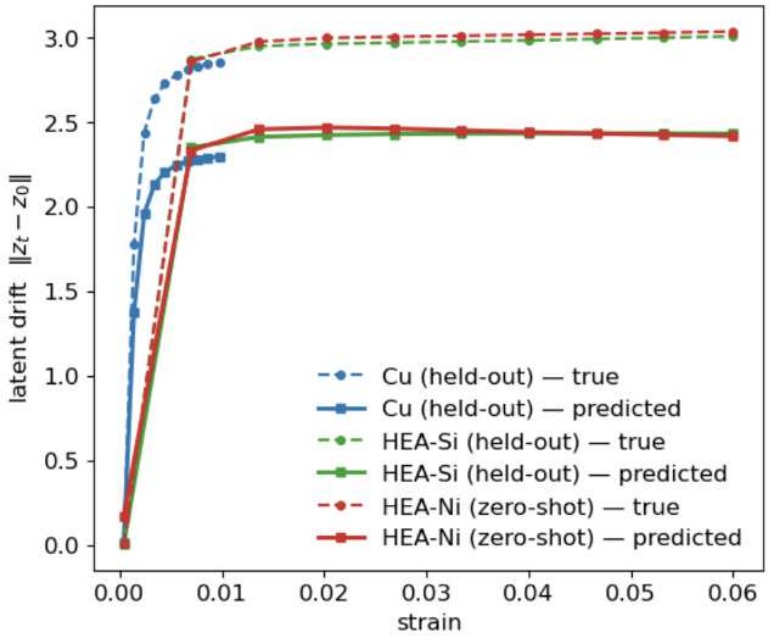


Figure 7: Displacement from the supplied initial latent state, $\|z - z_0\|$, for the encoded reference states (dashed lines) and transformer predictions (solid lines). The predicted trajectory reproduces the shape of the reference trajectory but contracts its magnitude by approximately 19%, and does so by the same factor for the zero-shot material as for the two training materials.

The codec and transformer disorientations cannot be added to obtain the end-to-end disorientation. At $\varepsilon = 0.06$ for HEA-Ni, the codec and transformer medians are 4.38° and 3.10°, respectively, whereas the end-to-end median is 5.10°, not 7.48°. At each voxel, the three orientations define geodesic distances on $SO(3)$ that satisfy the triangle inequality,

$$|\theta_{\text{codec}} - \theta_{\text{xf}}| \;\le\; \theta_{\text{e2e}} \;\le\; \theta_{\text{codec}} + \theta_{\text{xf}}, \tag{1}$$

and the reported medians are calculated independently over the voxels of each disorientation field. The relevant comparison is therefore between the codec and end-to-end medians. At the final evaluated state, the difference between them is 0.29° for Cu at $\varepsilon = 0.0098$, 0.71° for HEA-Si at $\varepsilon = 0.06$, and 0.72° for HEA-Ni at $\varepsilon = 0.06$. The codec-only median is thus much closer to the end-to-end median than the standalone transformer median would suggest, showing that latent-state prediction produces a relatively modest increase in the median error of the delivered orientation field. The decomposition is nearly identical for held-out HEA-Si and zero-shot HEA-Ni, indicating that the orientation error observed for HEA-Ni is not a distinct consequence of zero-shot material prediction.

The two error sources have distinct spatial signatures. The qualitative difference visible in the difference maps of Fig. 5 is quantified by the spatial autocorrelation of the two error contributions (Fig. 9). For each field, the correlation length is the first spatial lag at which the normalized autocorrelation falls below $1/e$. The codec errors in the mechanical fields have short correlation lengths of approximately one to two voxels, whereas the errors associated with latent-state prediction remain correlated over approximately three to four voxels, which is of the order of a grain diameter at this resolution. The rotation field is an exception, with both contributions

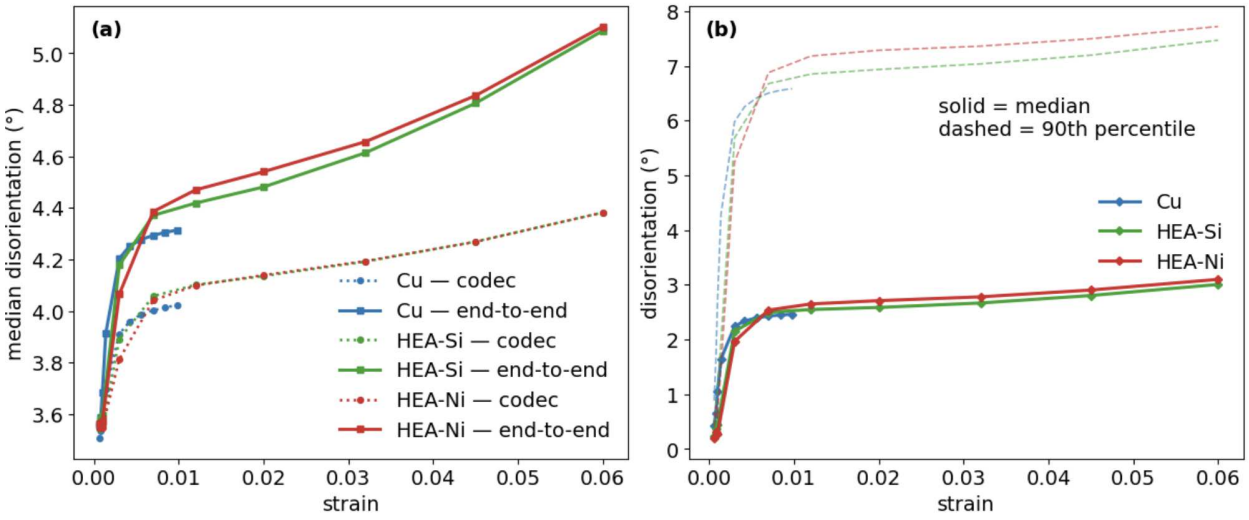


Figure 8: **Decomposition of orientation error into codec and transformer contributions.** (a) Median disorientation between the reference orientation field and its encode–decode reconstruction (codec, dotted) and between the reference field and the complete decoded prediction (end-to-end, solid), plotted against applied strain for Cu, HEA-Si, and zero-shot HEA-Ni. The dotted curves reproduce the medians of Fig. 6 for comparison with the end-to-end result. (b) Decoded discrepancy introduced by latent-state prediction, defined as the disorientation between the decoded reference latent and the decoded predicted latent, shown as the median (solid) and 90th percentile (dashed). The codec curves are nearly independent of material and establish the larger contribution to the median end-to-end orientation error. All angles are disorientations minimized over cubic crystal symmetry.

decaying within approximately one voxel. No systematic periodic peak is observed at multiples of the eight-voxel patch width.

The difference in correlation length indicates that the two stages introduce different forms of error. The codec primarily loses fine-scale spatial information during compression and reconstruction, whereas latent-state prediction introduces discrepancies organized over several neighboring voxels, consistent with features rendered at the wrong magnitude, position, or spatial extent. This distinction explains why error fields with comparable pointwise magnitudes can nevertheless have visibly different spatial structures.

To determine whether patchwise processing introduces localized boundary effects, we compare the error at patch faces with that in patch interiors. The boundary-to-interior ratio is the mean absolute error at voxels on an 8^3-patch face divided by the mean absolute error at interior voxels; a value near unity indicates no detectable boundary dependence. Because the codec reconstruction can itself introduce field-dependent boundary effects, the boundary dependence associated with latent-state prediction is assessed relative to the codec baseline. For the elastic-strain and stress fields, the latent-prediction error is moderately elevated at patch faces relative to that baseline. The plastic-strain and rotation fields show no comparable additional boundary enhancement. The behavior observed for elastic strain and stress is consistent with the patchwise formulation of the model: each 8^3 patch is predicted from its own initial latent state and condition vector without access to adjacent patches, so spatial features extending across a patch boundary must be reconstructed independently on either side. Correlation lengths and boundary-to-interior ratios averaged over five sampled deformation states are reported in Supplementary Table S7.

Table 4: **Disorientation statistics for the codec, transformer, and end-to-end error measures**, at representative deformation states. Codec denotes the disorientation between the reference orientation field and decode(z_{true}). Transformer denotes the decoded discrepancy between decode(z_{true}) and decode(z_{pred}); using the same decoder for both latent states isolates the effect of latent-state prediction after decoding but does not remove decoder sensitivity. End-to-end denotes the disorientation between the reference field and decode(z_{pred}). Statistics are calculated over all 262,144 voxels in each volume, and all angles are minimized over cubic crystal symmetry. The complete state-by-state table is given in Supplementary Table S9.

		Codec			Transformer			End-to-end	
Material	ε	Mean	Med.	p90	Mean	Med.	p90	Med.	p90
Cu	0.0010	6.358	3.582	12.796	1.831	1.050	2.738	3.683	12.785
Cu	0.0030	6.738	3.911	13.396	3.674	2.242	5.978	4.204	13.722
Cu	0.0098	6.855	4.023	13.661	3.994	2.464	6.583	4.314	13.926
HEA-Si	0.0030	6.732	3.891	13.553	3.518	2.151	5.686	4.179	13.967
HEA-Si	0.0200	6.969	4.136	13.994	4.165	2.586	6.934	4.481	14.317
HEA-Si	0.0600	7.166	4.382	14.161	4.584	3.006	7.468	5.087	14.627
HEA-Ni†	0.0030	6.669	3.813	13.676	3.256	1.963	5.238	4.066	13.953
HEA-Ni†	0.0200	6.996	4.139	14.243	4.350	2.710	7.284	4.540	14.738
HEA-Ni†	0.0600	7.190	4.382	14.382	4.688	3.099	7.720	5.104	14.964

†Zero-shot: HEA-Ni was not seen during training.

The same qualitative spatial-error signatures are observed for Cu, HEA-Si, and the zero-shot HEA-Ni material. The zero-shot result therefore extends beyond agreement in aggregate error magnitude: the spatial organization of the error is also consistent with that observed for the materials represented during training.

3 Discussion

The coupled field evolution of HEA-Ni was predicted without including any HEA-Ni deformation trajectories in training, validation, fine-tuning, or empirical field normalization. Across all quantities examined including macroscopic stress, voxelwise stress, elastic strain, plastic strain, crystallographic orientation, spatial error correlation, patch-boundary sensitivity, and latent-trajectory contraction, the errors for zero-shot material are comparable to those obtained for held-out microstructures of the training materials. Within the material family considered here, introducing an unseen material therefore does not produce a distinct or substantially larger source of error.

The error decomposition in Section 2.4 helps explain this result. Compression accounts for the larger fraction of the total error, approximately three-quarters of the normalized stress error is already present after the reference fields are encoded and decoded, before latent-state prediction is introduced. The codec also accounts for most of the median orientation error. Importantly, this compression error varies little among the three materials. The codec is unconditional and receives neither material properties nor loading descriptors. In addition, normalization by the predicted channel statistics reduces material-dependent differences in the magnitudes and scales of the mechanical fields before encoding. The encoder therefore operates on dimensionless fields with broadly comparable statistical distributions. Because the three systems also have similar grain morphology and crystallographic texture, they exhibit common spatial features, including

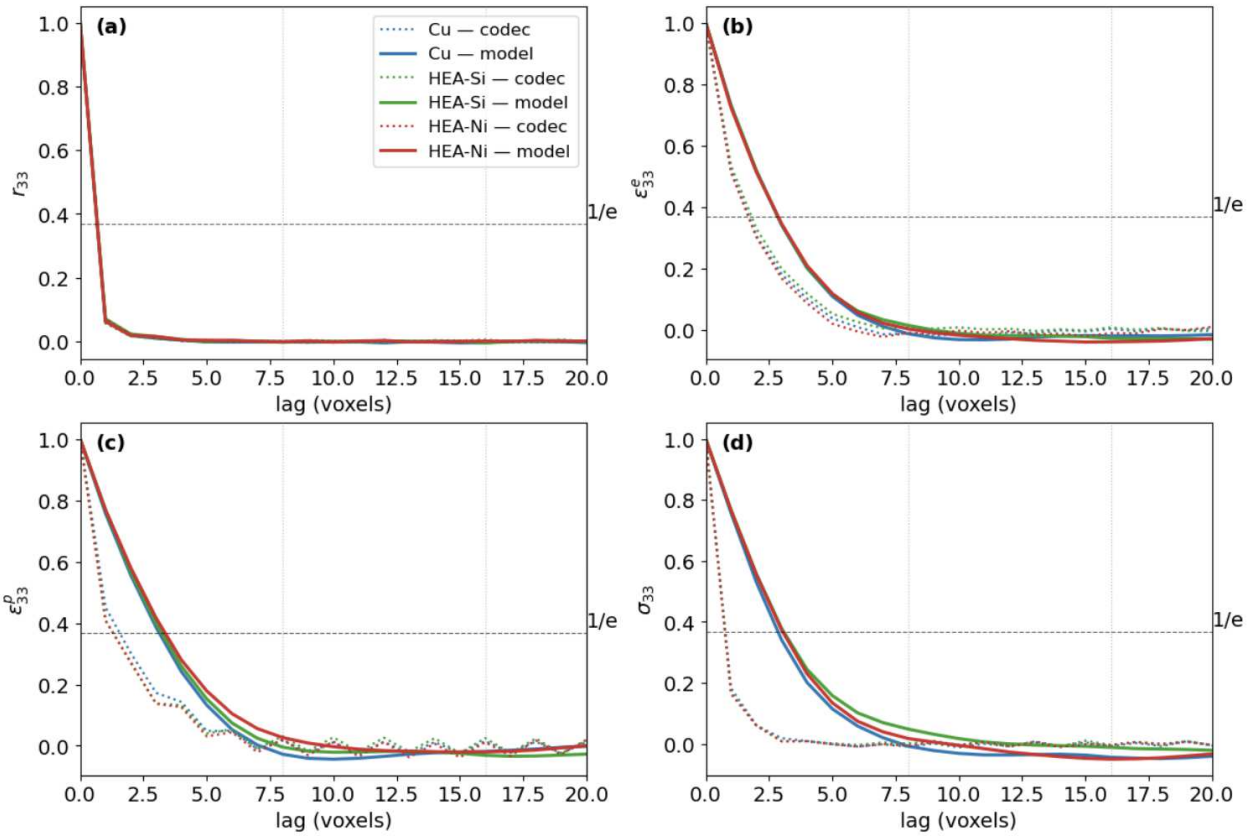


Figure 9: Spatial autocorrelation of the codec and latent-prediction error fields for R_{33}, ε^e_{33}, ε^p_{33}, and σ_{33} at a representative deformation state. Dotted curves denote the codec error, and solid curves denote the additional error associated with latent-state prediction. The horizontal line at $1/e$ defines the threshold used to determine the correlation length. Vertical lines at lags of 8 and 16 voxels indicate multiples of the patch width and are included to identify possible periodic signatures associated with patchwise processing.

grain-scale field partitioning, boundary structure, and intragranular gradients. The codec can consequently represent the normalized fields of HEA-Ni with approximately the same accuracy as those of Cu and HEA-Si.

Although the codec has no explicit access to the material identity, it can still learn spatial features specific to the distribution of fields present in the training data. The results show that the normalized HEA-Ni fields remain within the representational range learned from Cu and HEA-Si. This similarity is consistent with the fact that all three materials have initially equiaxed microstructures and are subjected to the same monotonic loading path. The three materials are consequently subject to a similar compression-imposed error floor. A material or microstructure that produces substantially different normalized field patterns may not exhibit the same behavior.

The decomposition also shows that the observed zero-shot transfer is determined partly by the location of the dominant error, rather than solely by extrapolation of the conditioned predictor in constitutive-parameter space. The material-dependent components, which are the latent transformer and the statistics predictor, account for a smaller fraction of the final error, and their contributions are similar for the zero-shot material and the training materials. Two aspects of the model design likely contribute to this result. First, material dependence is introduced through

continuous descriptors of cubic elasticity, Voce hardening, loading, and deformation progress. No categorical material label is used. This formulation allows the predicted response to vary continuously with the material parameters, although it does not by itself guarantee accurate extrapolation beyond the training domain. Second, the normalization statistics are predicted using a hybrid mechanics–machine-learning model. Voigt–Reuss–Hill homogenization and an approximate uniaxial Voce response provide the baseline, while a small neural network predicts a bounded correction. The mechanics-based component therefore constrains the statistics prediction when the learned correction is uncertain. Consistent with this interpretation, the oracle comparison shows that replacing the predicted statistics with empirical values changes the normalized stress error by only 0.006, corresponding to a few percent of the total error. The errors associated with latent-state prediction and statistics prediction therefore do not amplify one another appreciably.

The direct, non-autoregressive prediction strategy also affects how errors evolve with deformation. Each requested state is predicted independently from the initial latent state and the corresponding condition vector; a predicted state is never used as the input for a subsequent prediction. The normalized field errors increase through the elastic–plastic transition and then remain approximately constant, even as the stress magnitude and patch-to-patch variation continue to increase. This behavior is consistent with the absence of sequential error accumulation. It does not establish stability beyond the deformation range examined here, but it confirms that the reported errors do not arise from autoregressive propagation. The same formulation also permits any deformation state within the modeled range to be queried directly, without evaluating the intervening states.

The scope of the demonstrated zero-shot generalization remains limited. HEA-Ni was excluded entirely from training and fine-tuning, and three of its four hardening descriptors lie outside the componentwise ranges represented by the training materials. Its macroscopic stress–strain response, however, is broadly bounded by the responses of Cu and HEA-Si. Moreover, all three materials are face-centered-cubic polycrystals with comparable grain morphology and crystallographic texture. The present evaluation therefore demonstrates transfer across constitutive behavior within a restricted family of microstructures; it does not test simultaneous extrapolation across crystal structure, grain morphology, texture, loading mode, and loading history. The results should consequently be interpreted as evidence that continuous conditioning can support zero-shot prediction within the investigated material domain, rather than as evidence of generalization to arbitrary materials. The transfer observed here relies on the normalized fields remaining compatible with the representation learned by the codec. Substantial changes in microstructural geometry or deformation mechanisms may alter those fields more strongly than the constitutive-parameter variations considered here and therefore represent an important test for future work.

Future directions

The error decomposition identifies compression as the primary target for further model development. A preliminary experiment indicates that the orientation representation can be substantially improved. We trained a separate codec using only the six rotation components, with patches extracted from undeformed synthetic polycrystals. This codec was not coupled to the latent transformer and does not affect any of the results reported above. Across the evaluated strain range its median disorientation ranges from 1.03° to 1.74°, against 3.55° to 4.38° for the shared codec (Supplementary Table S10), and at $\varepsilon = 0.030$ the reconstructed-to-reference KAM ratio falls from 1.37 to 1.11 (Supplementary Fig. S1). The two codecs also differ in training resolution and in exposure to intragranular orientation gradients, so the comparison does not isolate channel specialization alone. Both differences favor the shared codec, however, and when voxels are grouped by

the reference local orientation gradient to place the two codecs at matched local difficulty, the ratio of shared to dedicated median error remains between 2.5 and 4.3 in every bin, at every strain, for all three materials (Supplementary Fig. S2). The reported factors are therefore lower bounds on what a dedicated orientation representation can achieve. Additionally, these results show that the orientation error of the shared codec is not an inherent consequence of representing rotations by the first two columns of the rotation matrix. Instead, they suggest that a representation specialized for orientation is a promising direction. Whether the improvement is retained after coupling the specialized codec to the transformer remains to be determined. More generally, the same approach may benefit the other field groups considered here, all of which exhibit errors dominated by compression.

One possible extension is therefore to replace the shared 24-channel codec with field-specific codecs and concatenate their latent representations. For example, separate latent channels could be assigned to orientation and to the coupled mechanical fields, producing an $8 \times 8 \times 8 \times 2$ latent representation instead of the present $8 \times 8 \times 8 \times 1$ representation. This design would increase latent capacity while allowing each codec to learn features appropriate to its fields. The additional cost to the transformer would be modest because the number of spatial tokens would remain 512; only the dimension of each token and the corresponding input projection would change. However, increased latent capacity would not resolve all observed errors. The residual population of voxels with large disorientation is nearly independent of strain and is concentrated in regions with large orientation gradients (see Supplementary Section S5.4). This behavior is more consistent with the spatial resolution of the 64^3 microstructures, in which some grains span only a few voxels, than with insufficient latent capacity. Similarly, the contraction of the predicted latent trajectory originates in the conditioned predictor and would persist even with an improved codec. This contraction further suggests that a Euclidean latent-space loss may not assign sufficient weight to all physically relevant changes. Possible alternatives include supplementing the latent loss with decoded-field losses, directly penalizing errors in predicted increments, or introducing multiscale objectives. The benefits and trade-offs of these choices will need to be systematically evaluated.

Three additional limitations warrant further study. First, treating the microstructure as independent 8^3 patches prevents direct exchange of information between neighboring regions. The elevated boundary-to-interior error ratios for stress and elastic strain indicate that the absence of this context affects the predictions. Overlapping patches, halo regions, neighboring-patch tokens, or hierarchical representations could provide the missing information. Their performance should be assessed using spatially resolved diagnostics, including boundary-to-interior errors, rather than aggregate error alone. Second, the present architecture does not explicitly enforce mechanical equilibrium, strain compatibility, or constitutive consistency. This limitation is important if the predicted fields are to be used to study strain localization or damage initiation, because a field may reproduce the reference statistics while remaining mechanically inadmissible. Future work should quantify equilibrium and compatibility residuals directly, examine whether these residuals differ between held-out and zero-shot materials, and evaluate physics-encoded formulations such as neural operators [15]. Third, the limits of generalization should be established by independently varying constitutive response, crystallographic texture, grain-size distribution, grain morphology, crystal structure, loading direction, and loading history. Cyclic, nonproportional, and multiaxial loading are particularly important tests. The current condition vector specifies the requested deformation state but does not represent an arbitrary preceding loading path; extension to such histories may therefore require explicit path-dependent internal variables.

Within the range examined here, the results indicate that a full-field crystal-plasticity surrogate need not be retrained separately for every material. Instead, the field evolution can be conditioned on constitutive descriptors and an initial microstructural state. In the present numerical evaluation,

this initial state is obtained from the first crystal-plasticity increment. In a future experimental implementation, the initial crystallographic orientation and elastic-strain fields could be measured using high-energy diffraction microscopy, and the initial stress could be inferred from the elastic strain and known elastic constants [19]. The plastic-strain field could reasonably be initialized to zero for an undeformed specimen, whereas application to a pre-deformed specimen would require an independent estimate of its initial plastic state. Avoiding material-specific retraining would substantially improve the feasibility of using this type of surrogate for uncertainty quantification, alloy screening, and inverse problems, where repeated evaluations across many candidate materials are required.

4 Methods

Crystal-plasticity simulations and data preparation

Spectral crystal-plasticity simulations were performed using the MASSIF code [20]. Each simulation was initialized using a three-dimensional synthetic polycrystalline microstructure generated with DREAM.3D [21] and discretized on a $64 \times 64 \times 64$ voxel grid. All simulations were subjected to strain-controlled uniaxial tension using a constant strain increment of $\Delta\varepsilon = 2 \times 10^{-4}$. The Cu simulations were loaded to approximately 1% strain over 49 increments, while the HEA-Si and HEA-Ni simulations were loaded to 6% strain over 300 increments. Elastic constants and Voce hardening parameters are provided in Supplementary Section S1. The synthetic microstructures were generated with comparable grain morphology and crystallographic-texture statistics across the three material systems, so the HEA-Ni evaluation examines transfer across constitutive behavior rather than simultaneous extrapolation across both material response and microstructural morphology.

The training, validation, and test data were obtained from previous work [10, 11, 18]; no new simulations were performed. Cu and HEA-Si were used for model development, with 150 simulations available per material: 135 training, 5 validation, and 10 held-out test. HEA-Ni was excluded from all model training, validation, and fine-tuning, and 40 HEA-Ni simulations formed an independent zero-shot test pool. The datasets were partitioned at the simulation level before individual deformation states and spatial patches were extracted, so no patch from a held-out or zero-shot simulation appeared in a training batch, and these simulations did not influence early stopping or hyperparameter selection. Supplementary Table S1 summarizes the material exposure of each learned component.

For each simulation, the recorded data were stored as an array with dimensions (N_t, N, N, N, C), where N_t is the number of recorded deformation states, $N = 64$ and $C = 24$. Each state contains six components representing the first two columns of the crystallographic rotation matrix R [22], six elastic-strain components, six plastic-strain components, and six stress components; the third column of R is recovered from the cross product of the first two. The spectral solver produces no zero-strain output, so the first recorded state corresponds to an applied strain of 2×10^{-4} rather than to the undeformed configuration. This state is supplied to the model as the initial condition, and every subsequent recorded state is a prediction target compared with the crystal-plasticity field at the same applied strain. The Cu simulations comprise 49 recorded states and the HEA-Si and HEA-Ni simulations 300, giving 48 and 299 prediction targets and final applied strains of 0.0098 and 0.0600, respectively. For computationally efficient training, each $64 \times 64 \times 64 \times 24$ state was partitioned into nonoverlapping patches of size $8 \times 8 \times 8 \times 24$, giving $8^3 = 512$ patches per deformation state. During inference the 512 patches are predicted independently and reassembled at their original spatial locations.

Field normalization and zero-shot prediction of channel statistics

For mechanical channel j at deformation state t, the normalized field is

$$\tilde{X}_{t,j}(\mathbf{r}) = \frac{X_{t,j}(\mathbf{r}) - \mu_{t,j}}{\sigma_{t,j}}, \tag{2}$$

where $\mathbf{r}$ denotes voxel position and $\mu_{t,j}$ and $\sigma_{t,j}$ are the corresponding spatial mean and standard deviation. The six rotation components are retained in their bounded six-dimensional representation and are not normalized using material- and timestep-dependent empirical statistics.

This normalization is the point at which a conventional surrogate would reacquire a dependence on the target material, because empirical values of $\mu_{t,j}$ and $\sigma_{t,j}$ would have to be computed from a simulated trajectory of every material encountered at inference. Genuine zero-shot operation therefore requires the normalization statistics to be inferred rather than measured. During training on Cu and HEA-Si, the empirical statistics are calculated using only simulations assigned to the training set; for inference, they are supplied by the auxiliary model described below.

Condition vector. The statistics predictor and latent transformer receive the same 26-dimensional condition vector c_t. Deformation progress is represented by the applied strain of the requested state divided by a fixed reference strain of $\varepsilon_{\mathrm{ref}} = 0.10$; the normalizer is fixed rather than set by the length of each trajectory, so a given value denotes the same absolute deformation for every material. The condition vector also contains 18 loading descriptors, three descriptors of cubic elasticity, C_{11}/C_{ref}, C_{12}/C_{11}, and C_{44}/C_{11}, and four Voce hardening descriptors, τ_0, $\tau_0+\tau_1$, θ_0, and θ_1. Definitions and scaling are given in Supplementary Section S2.

Mechanics baseline. The cubic single-crystal elastic constants are homogenized using Voigt–Reuss–Hill averaging [23] to obtain the effective isotropic bulk modulus K, shear modulus G, Young's modulus E, and Poisson's ratio ν. These are combined with the Voce hardening law [24] to construct an approximate uniaxial elastic–plastic response. A Taylor factor of $M_T = 2.528$, determined by least squares against the volume-averaged axial stresses of the Cu and HEA-Si training simulations and held fixed thereafter, relates the single-crystal slip resistance to the effective polycrystalline flow stress [25, 26]; it is not re-fit for the zero-shot material. A single fixed value is adequate because the mechanics model supplies only a baseline estimate, and systematic departures are absorbed by the learned correction. At each deformation state, the equivalent plastic strain is obtained from

$$\varepsilon_p = \max\!\left(\varepsilon_t - \frac{\sigma}{E},\, 0\right), \qquad \sigma = \sigma_f(\varepsilon_p), \tag{3}$$

solved iteratively by fixed-point consistency. The resulting uniaxial solution supplies mechanics-based estimates of the 18 channel means, $\hat{\mu}_{\mathrm{mech}}$, and positive channel-specific scales s_{mech}; lower bounds are imposed on these scales to maintain numerical stability during the early stages of loading, where several plastic-strain channels are identically zero.

Learned correction. The correction is a multilayer perceptron with two hidden layers of width 128 and GELU activations, containing approximately 4×10^4 trainable parameters, with an L_2 weight-decay coefficient of 10^{-4}. It has two output heads: the first predicts normalized residual corrections to the mechanics-based means, $\hat{y}_{\mathrm{resid}} \in \mathbb{R}^{18}$, and is zero-initialized so that the untrained network reproduces the mechanics baseline exactly; the second predicts logarithmic corrections to the mechanics-based scales, $\hat{y}_{\log\sigma} \in \mathbb{R}^{18}$. The final predicted statistics are

$$\mu_{\mathrm{pred}} = \hat{\mu}_{\mathrm{mech}} + r_{\mathrm{ref}} \odot \hat{y}_{\mathrm{resid}}, \qquad \sigma_{\mathrm{pred}} = s_{\mathrm{mech}} \odot \exp(\hat{y}_{\log\sigma}), \tag{4}$$

where $\odot$ denotes element-wise multiplication and r_{ref} contains 10^{-2} for the strain channels and 100 MPa for the stress channels. This parameterization retains the mechanics model as the operative baseline outside the training-material range. The predictor is trained by minimizing

$$\mathcal{L}_{\mathrm{stat}} = \|\hat{y}_{\mathrm{resid}} - y_{\mathrm{resid}}\|_2^2 + \|\hat{y}_{\log\sigma} - y_{\log\sigma}\|_2^2, \tag{5}$$

with targets defined in Supplementary Section S2, using Adam with an initial learning rate of 10^{-3}, learning-rate reduction on a validation-loss plateau, and early stopping. For zero-shot HEA-Ni inference, both the initial-state and target-state normalization statistics are obtained from this predictor; no empirical HEA-Ni means or standard deviations are supplied at any point.

Variational autoencoder for state compression

Each normalized material-state patch $\tilde{X}_t \in \mathbb{R}^{8\times8\times8\times24}$ is compressed by a variational autoencoder [27] into a spatially structured latent field $z_t \in \mathbb{R}^{8\times8\times8\times1}$. The compression is purely in the channel dimension; spatial resolution is preserved throughout the encoder and decoder. The autoencoder is unconditional: it receives no material or loading descriptors. Material dependence has already been removed at this stage by normalization, so what reaches the encoder is a dimensionless field whose statistics are comparable across materials. Material dependence enters the framework at normalization and at latent prediction, in both cases through the same condition vector, so conditioning the codec would be redundant; it would also be undesirable for the zero-shot objective, since an unconditional codec cannot specialize to the identity of the training materials.

The encoder processes orientation, elastic strain, plastic strain, and stress using separate three-dimensional convolutional branches. The plastic-strain branch additionally receives the elastic-strain channels, and the stress branch likewise, so that the coupling among the mechanical fields is retained rather than left to be recovered after fusion. Branch outputs are concatenated and processed using residual convolutional blocks and a learned sigmoid attention gate, and separate convolutional towers predict the latent mean $\mu_\phi(\tilde{X}_t)$ and log-variance $\log\sigma^2_\phi(\tilde{X}_t)$. During training the latent field is sampled using $z_t = \mu_\phi(\tilde{X}_t) + \sigma_\phi(\tilde{X}_t) \odot \epsilon$ with $\epsilon \sim \mathcal{N}(0, I)$; after training, the posterior mean is used as the deterministic latent representation. The decoder contains a shared convolutional trunk followed by separate heads for each field group, mirroring the encoder's coupling. The orientation head produces two three-dimensional vectors at each voxel, which a differentiable Gram–Schmidt operation orthonormalizes, so the decoded orientation is a proper rotation by construction rather than by penalty [22]. The VAE is trained using

$$\mathcal{L}_{\mathrm{VAE}} = \mathcal{L}_{\mathrm{rot}} + \mathcal{L}_e + \mathcal{L}_p + \mathcal{L}_\sigma + 10^{-6}\mathcal{L}_{\mathrm{KL}}, \tag{6}$$

where $\mathcal{L}_{\mathrm{rot}}$ is the mean geodesic angle between reference and reconstructed rotation matrices, evaluated without minimization over the cubic symmetry operators during training, and the mechanical terms are Huber losses with $\delta = 0.4$. Complete architectures are given in Supplementary Section S3.

Conditioned latent prediction

Let $z_0 = \mu_\phi(\tilde{X}_0)$ denote the deterministic latent representation of the initial state. The latent state at target deformation step t is predicted as

$$\hat{z}_t = z_0 + \Delta z_\theta(z_0, c_t), \tag{7}$$

where Δz_θ is the condition-dependent latent change predicted by the transformer [28]. Each training example consists of (z_0, c_t, z_t), and each target deformation state is treated as an independent

example. The model does not use $\hat{z}_{t-1}$ to predict $\hat{z}_t$, so errors from earlier predicted states cannot accumulate through autoregressive rollout, and any requested state can be evaluated without integrating through the intervening ones.

The $8\times8\times8$ initial latent field is flattened into 512 spatial tokens, and each scalar latent value is projected into a 384-dimensional embedding. Trainable factorized axial embeddings represent the three spatial coordinates. Elasticity, loading, and plasticity descriptors are separately embedded and prepended as three condition tokens, and deformation progress together with selected condition embeddings additionally modulate every transformer block through adaptive layer normalization with zero initialization [29]. The network contains eight pre-normalized transformer encoder blocks, each using 12-head self-attention and a feed-forward hidden dimension of 1536. Because the sequence dimension represents spatial positions rather than successive timesteps, no causal attention mask is applied; self-attention therefore permits interaction among all latent positions within an 8^3 patch, but not across patches. After the transformer blocks the condition tokens are removed and a pointwise multilayer-perceptron head maps each spatial token to a scalar latent increment. The final prediction layer and adaptive residual gates are zero-initialized, so the untrained network begins from the identity mapping $\hat{z}_t = z_0$.

The VAE parameters are frozen during transformer training. The transformer minimizes the mean Huber loss

$$\mathcal{L}_{\mathrm{Tr}} = \frac{1}{512}\sum_{i=1}^{512} H_{\delta_{\mathrm{Tr}}}(\hat{z}_{t,i} - z_{t,i}), \tag{8}$$

using AdamW, global gradient-norm clipping, linear learning-rate warm-up, and cosine learning-rate decay. The transformer contains approximately 22.0 million trainable parameters. The predicted latent field is decoded to obtain the normalized target field, $\hat{\tilde{X}}_t = D_\psi(\hat{z}_t)$, and the predicted target-state statistics transform the 18 mechanical channels back to physical units.

Evaluation metrics

Predictions are evaluated after decoding and transformation to physical units. The macroscopic axial stress $\sigma_{33}(t)$ is the volume average of the predicted voxelwise stress over the reconstructed 64^3 volume, and its relative error is calculated with respect to the crystal-plasticity result; relative errors at very early deformation states are interpreted cautiously because the reference stress is small.

Local mechanical-field errors are quantified using a normalized mean absolute error: for each component of a six-component field group, the voxelwise mean absolute error is divided by the spatial standard deviation of the corresponding reference component, and the six normalized values are averaged. Normalizing by field variation rather than by the mean avoids ill-conditioning for shear components, whose volume averages are near zero, and permits comparison among materials with different absolute stress levels. A normalized error of 0.19 therefore means that the voxelwise mean absolute error is 19% of the spatial standard deviation of the reference field, not 19% of its magnitude.

Crystallographic accuracy is evaluated using disorientation minimized over the 24 proper rotations of the cubic point group,

$$\theta = \min_{S\in G_{\mathrm{cubic}}} \cos^{-1}\left(\frac{\mathrm{tr}(SR\hat{R}^T)-1}{2}\right), \tag{9}$$

and kernel average misorientation is calculated from the symmetry-reduced disorientation between

each voxel and its local neighborhood. Symmetry reduction is essential because crystallographically equivalent orientations would otherwise be assigned artificially large angular differences.

Errors are separated into a compression contribution and a latent-prediction contribution by decoding the reference latent z_t and the predicted latent $\hat{z}_t$ through the same frozen decoder, giving $X_t^{\mathrm{codec}} = D_\psi(z_t)$ and $X_t^{\mathrm{pred}} = D_\psi(\hat{z}_t)$. Because the decoder acts on both latent fields, comparing them isolates the effect of latent prediction after passage through a common decoder but does not remove the decoder's nonlinear sensitivity; the standalone prediction discrepancy should therefore not be read as an additive component of the end-to-end error. For crystallographic orientation this non-additivity is exact rather than approximate, because disorientations are geodesic distances on $SO(3)$ subject to the triangle inequality. Correlation lengths, boundary-to-interior ratios, and latent-residual controls are defined in Supplementary Section S4.

Unless otherwise stated, reported curves and uncertainties represent the mean and standard deviation over independent simulations, with the number of simulations given in each figure or table caption.

Data availability

Data supporting the findings of this study are available within the article or from the corresponding authors upon reasonable request.

Code availability

Codes supporting the findings of this study are available within the article or from the corresponding authors upon reasonable request.

Acknowledgements

Los Alamos National Laboratory is operated by Triad National Security, LLC, for the National Nuclear Security Administration of the U.S. Department of Energy under Contract No. 89233218CNA000001. The framework developed in this work was initially conceived with the support from LANL's LDRD ER-Seedling Project No. 20240825ER, "Towards Integrated Scale Bridging with Embedded Microstructure Analysis." The detailed analysis and testing were supported by the Dynamic Materials Properties Campaign under DOE-NNSA.

Author contributions

R.P. and A.S. developed the MELT-3D framework and the algorithm, R.P. developed and trained the deep learning architecture and analyzed the data, R.P. wrote the original draft and R.P and A.S. contributed to reviewing and editing the manuscript.

Competing interests

The authors declare no competing interests.

References

[1] Lebensohn, R. A., Kanjarla, A. K. & Eisenlohr, P. An elasto-viscoplastic formulation based on fast Fourier transforms for the prediction of micromechanical fields in polycrystalline materials. *International Journal of Plasticity* **32–33**, 59–69 (2012).

[2] Roters, F. *et al.* DAMASK–the düsseldorf advanced material simulation kit for modeling multi-physics crystal plasticity, thermal, and damage phenomena from the single crystal up to the component scale. *Computational Materials Science* **158**, 420–478 (2019).

[3] Liu, Z., Wu, C. T. & Koishi, M. A deep material network for multiscale topology learning and accelerated nonlinear modeling of heterogeneous materials. *Computer Methods in Applied Mechanics and Engineering* **345**, 1138–1168 (2019).

[4] Bonatti, C., Berisha, B. & Mohr, D. From CP-FFT to CP-RNN: Recurrent neural network surrogate model of crystal plasticity. *International Journal of Plasticity* **158**, 103430 (2022).

[5] He, J. *et al.* Material-response-informed DeepONet and its application to polycrystal stress–strain prediction in crystal plasticity. *JOM* **76**, 5744–5754 (2024).

[6] Yabansu, Y. C., Patel, D. K. & Kalidindi, S. R. Calibrated localization relationships for elastic response of polycrystalline aggregates. *Acta Materialia* **81**, 151–160 (2014).

[7] Yabansu, Y. C. & Kalidindi, S. R. Representation and calibration of elastic localization kernels for a broad class of cubic polycrystals. *Acta Materialia* **94**, 26–35 (2015).

[8] Montes de Oca Zapiain, D. & Kalidindi, S. R. Localization models for the plastic response of polycrystalline materials using the material knowledge systems framework. *Modelling and Simulation in Materials Science and Engineering* **27**, 074008 (2019).

[9] Frankel, A. L., Tachida, K. & Jones, R. E. Prediction of the evolution of the stress field of polycrystals undergoing elastic–plastic deformation with a hybrid neural network model. *Machine Learning: Science and Technology* **1**, 035005 (2020).

[10] Pandey, A. & Pokharel, R. Machine learning based surrogate modeling approach for mapping crystal deformation in three dimensions. *Scripta Materialia* **193**, 1–5 (2021).

[11] Pokharel, R., Pandey, A. & Scheinker, A. Physics-informed data-driven surrogate modeling for full-field 3d microstructure and micromechanical field evolution of polycrystalline materials. *JOM* **73**, 3371–3382 (2021).

[12] Ibragimova, O. *et al.* A convolutional neural network based crystal plasticity finite element framework to predict localised deformation in metals. *International Journal of Plasticity* **157**, 103374 (2022).

[13] Scheinker, A. & Pokharel, R. Adaptive 3D convolutional neural network-based reconstruction method for 3D coherent diffraction imaging. *Journal of Applied Physics* **128** (2020).

[14] Scheinker, A. & Pokharel, R. Enabling dynamic 3D coherent diffraction imaging via adaptive latent space tuning of generative autoencoders. *npj Computational Materials* **10**, 283 (2024).

[15] Khorrami, M. S. *et al.* A physics-encoded Fourier neural operator approach for surrogate modeling of divergence-free stress fields in solids. *arXiv preprint arXiv:2408.15408* (2025).

[16] Yu, G. *et al.* A robust and scalable surrogate model for 3d crystal plasticity simulation under diverse loading conditions. *Acta Materialia* 121441 (2025).

[17] Atkinson, M. D., White, M. D., Plowman, A. J. & Shanthraj, P. Microstructure sensitive recurrent neural network surrogate model of crystal plasticity. *arXiv preprint arXiv:2510.06904* (2025).

[18] Lenau, A., Pokharel, R., Scheinker, A. & Niezgoda, S. Full-field elastic strain tensor evolution of 3d polycrystals with recurrent neural networks and transfer learning. *Mechanics of Materials* **208**, 105389 (2025).

[19] Pokharel, R. & Lebensohn, R. A. Instantiation of crystal plasticity simulations for micromechanical modelling with direct input from microstructural data collected at light sources. *Scripta Materialia* **132**, 73–77 (2017).

[20] Tari, V. *et al.* Validation of micro-mechanical fft-based simulations using high energy diffraction microscopy on ti-7al. *Acta materialia* **154**, 273–283 (2018).

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

[22] Zhou, Y., Barnes, C., Lu, J., Yang, J. & Li, H. On the continuity of rotation representations in neural networks. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 5738–5746 (IEEE, 2019).

[23] Hill, R. The elastic behaviour of a crystalline aggregate. *Proceedings of the Physical Society. Section A* **65**, 349–354 (1952).

[24] Voce, E. The relationship between stress and strain for homogeneous deformation. *Journal of the Institute of Metals* **74**, 537–562 (1948).

[25] Taylor, G. I. Plastic strain in metals. *J. Inst. Metals* **62**, 307–324 (1938).

[26] Kocks, U. The relation between polycrystal deformation and single-crystal deformation. *Metallurgical Transactions* **1**, 1121–1143 (1970).

[27] Kingma, D. P. & Welling, M. Auto-encoding variational Bayes. In *International Conference on Learning Representations* (2014). `1312.6114`.

[28] Vaswani, A. *et al.* Attention is all you need. In *Advances in Neural Information Processing Systems*, vol. 30, 5998–6008 (2017).

[29] Peebles, W. & Xie, S. Scalable diffusion models with transformers. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, 4172–4182 (IEEE, 2023).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- melt3dsupplementary.pdf