Supplementary Information for

Machine Learning Inversion of Interatomic Force Constants from Single-Crystal Inelastic Neutron Scattering

Aiden Sable[a], Bander Linjawi[a], Kyle Bradbury[b], Jordan Malof[c], and Olivier Delaire[1] [a,d,e]

[a] *Department of Mechanical Engineering and Materials Science, Duke University, Durham, NC, USA*
[b] *Department of Electrical and Computer Engineering, Duke University, Durham, NC, USA*
[c] *Department of Electrical Engineering and Computer Science, University of Missouri–Columbia, Columbia, MO, USA*
[d] *Department of Chemistry, Duke University, Durham, NC, USA*
[e] *Department of Physics, Duke University, Durham, NC, USA*

[1]E-mail: olivier.delaire@duke.edu

1 uMLFF-based Training Data Generation

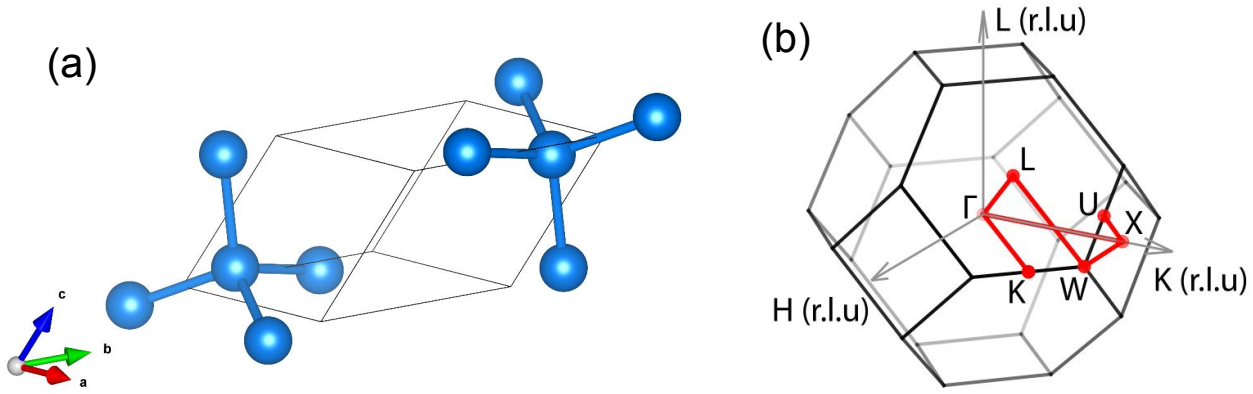

Figure S1: **Ge crystal structure and Brillouin zone.** (a) Primitive unit cell of diamond-cubic Ge. (b) First Brillouin zone with labeled high-symmetry points. The one-dimensional piecewise path connecting these points defines the momentum axis used in the phonon spectra presented in the main text.

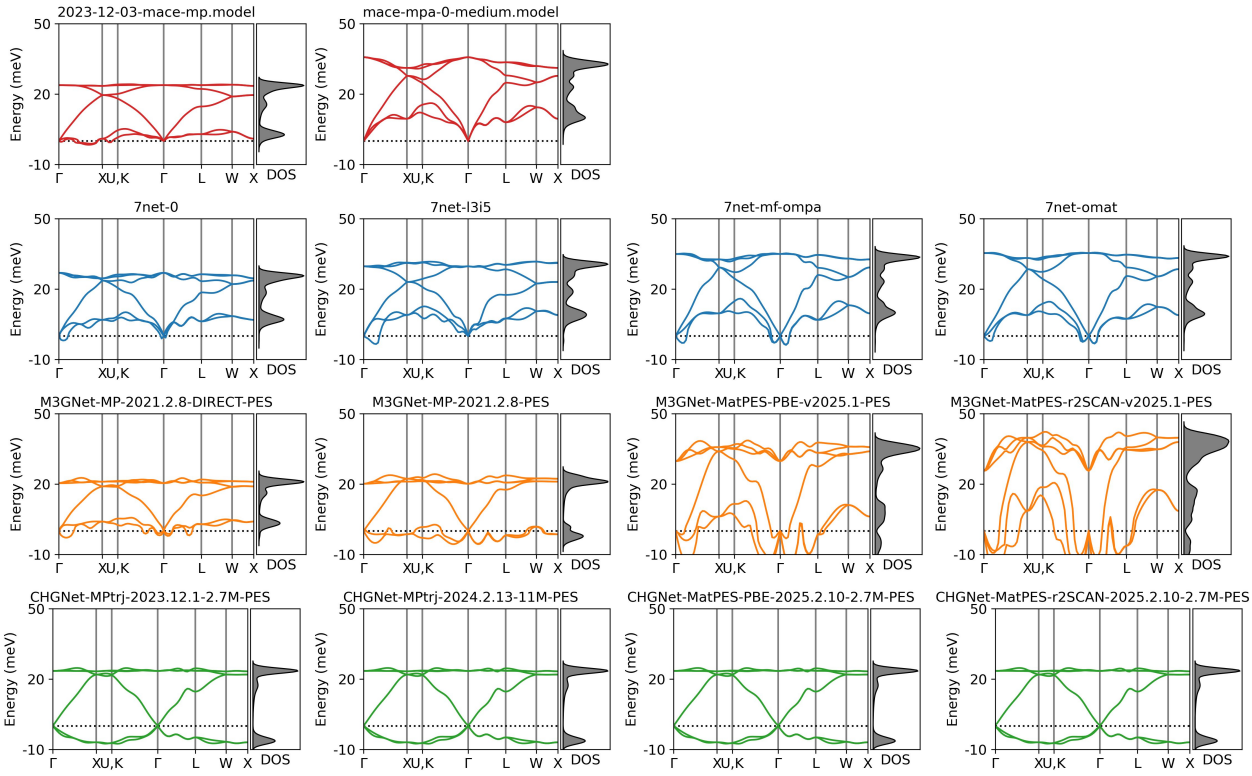

Figure S2: **Ge phonon dispersion and density-of-states (DOS) predictions from universal ML force fields (uMLFFs).** The same Ge crystal structure was supplied as input to 14 uMLFF models drawn from the MACE (red), SevenNet (blue), M3GNet (orange), and CHGNet (green) families. Specific model instances are indicated in the subplot titles. The dispersions are shown as curves on the energy axis along a wave-vector high-symmetry path, connecting symmetry points as labeled on the horizontal axis. The DOS is shown as a distribution of phonon states along the vertical energy axis.

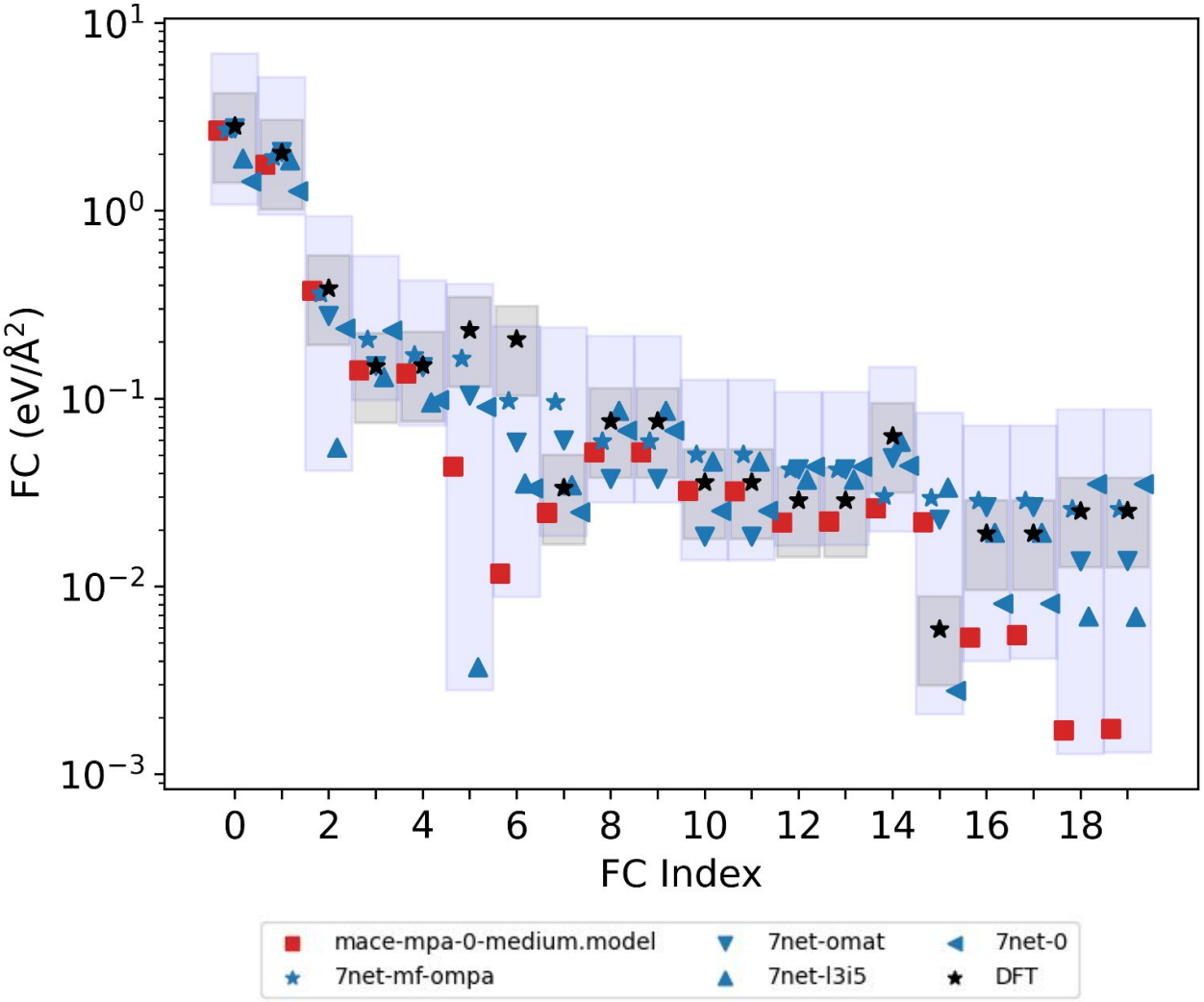

Figure S3: **Force constant (FC) training domain for Ge.** The five predicted FC tensors that yielded dynamically stable phonon dispersions were symmetry-reduced, and the resulting bounds were expanded to $[0.75 \times \min(FC_i),\ 2.5 \times \max(FC_i)]$ to define the training domain. This expanded range spans a broad set of physically plausible lattice dynamics and is intentionally asymmetric to facilitate identification of poorly performing models that collapse toward predicting the training-set mean. FC labels were then jointly and uniformly sampled from this domain and used in forward simulations to generate the corresponding inelastic neutron spectra. The DFT$\pm$50% range is shown in black for reference.

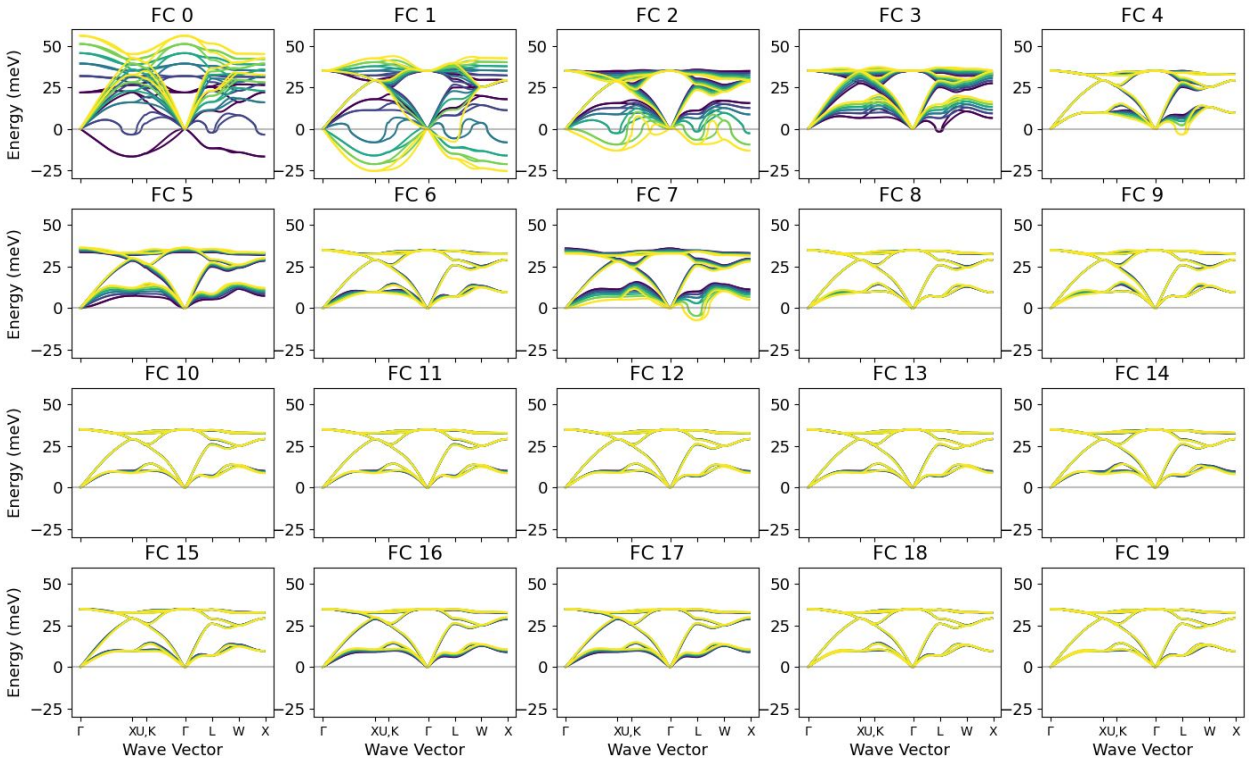

Figure S4: **Sensitivity of Ge phonon dispersions to the 20 largest symmetry-reduced FCs.** Force constants are ranked in descending order by magnitude. Line color denotes the value of the perturbed FC as a percentage of its sampled range, while all remaining FCs are held fixed. The first eight FCs produce observable changes in the dispersion, with the first two having the strongest influence.

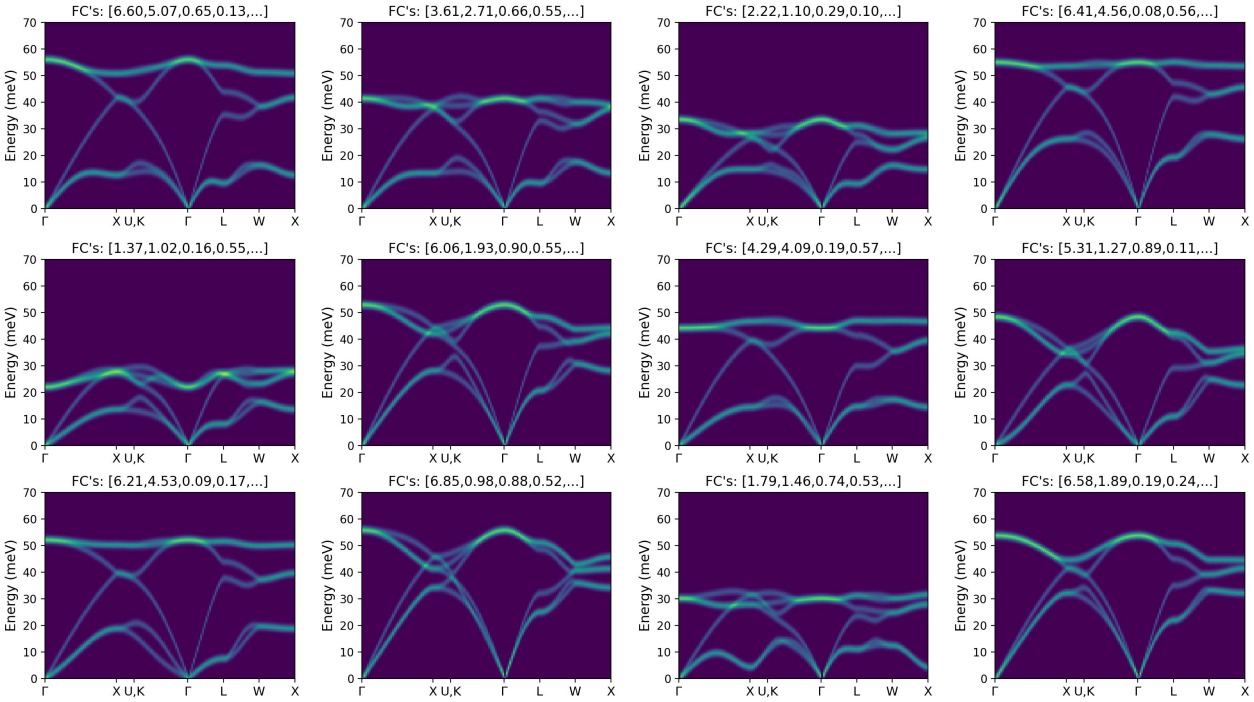

Figure S5: **Representative simulated $I(\mathbf{q}, E)$ spectra from the Ge training dataset.** Each subplot shows a randomly selected spectrum from the training dataset, together with its corresponding FC labels (shown in the title). The dataset spans a wide range of phonon behaviors and dynamical regimes.

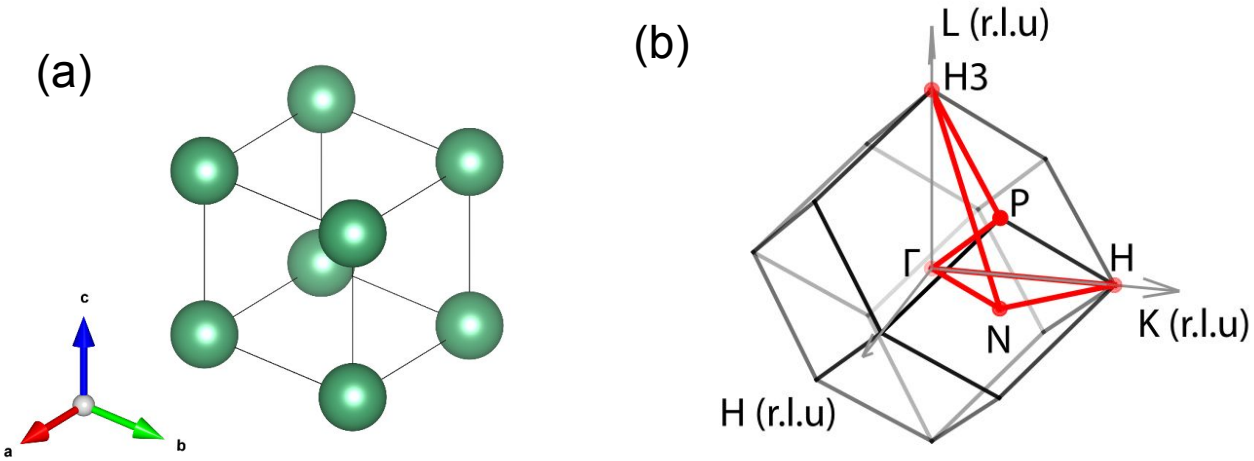

Figure S6: **Nb crystal structure and Brillouin zone.** (a) Primitive unit cell of body-centered-cubic Nb. (b) First Brillouin zone with labeled high-symmetry points. The one-dimensional piecewise path connecting these points defines the momentum axis used in the phonon spectra presented in the main text.

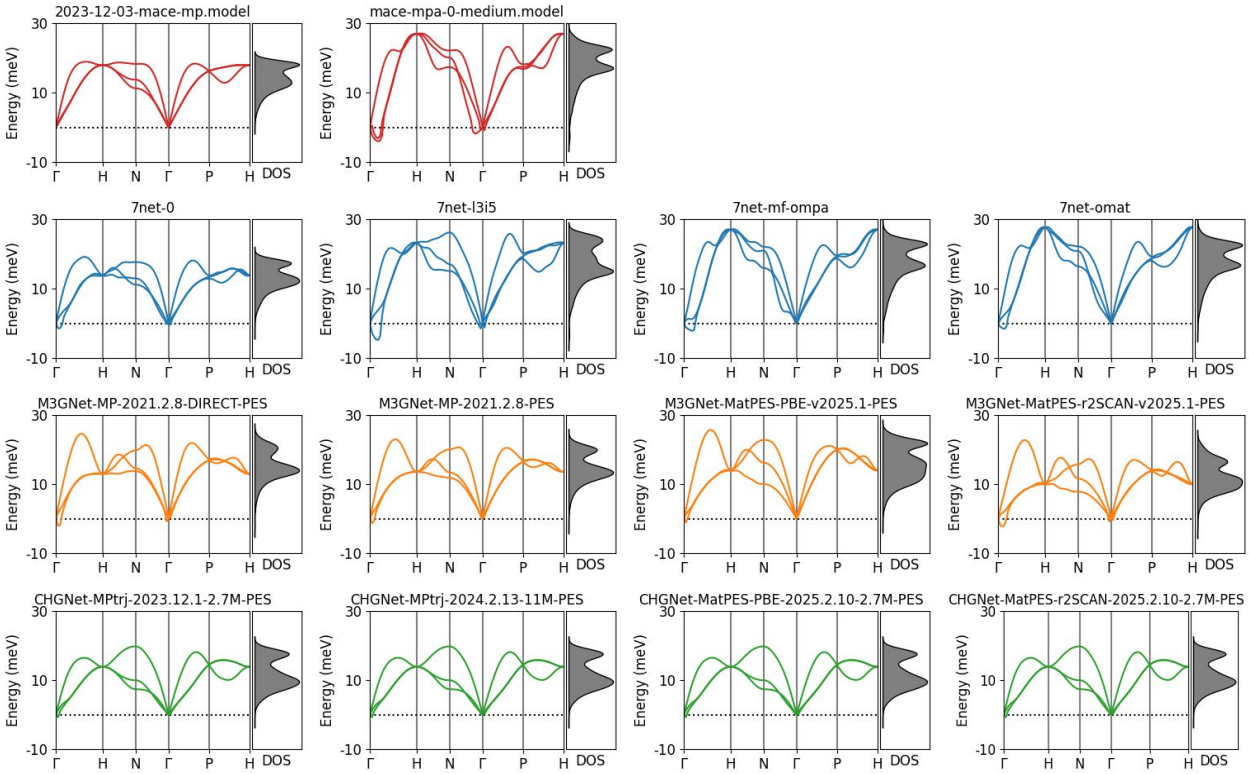

Figure S7: **Nb phonon dispersion and density-of-states (DOS) predictions from universal ML force fields (uMLFFs).** The same Nb crystal structure was supplied as input to 14 uMLFF models drawn from the MACE (red), SevenNet (blue), M3GNet (orange), and CHGNet (green) families. Specific model instances are indicated in the subplot titles. The dispersions are shown as curves on the energy axis along a wave-vector high-symmetry path, connecting symmetry points as labeled on the horizontal axis. The DOS is shown as a distribution of phonon states along the vertical energy axis.

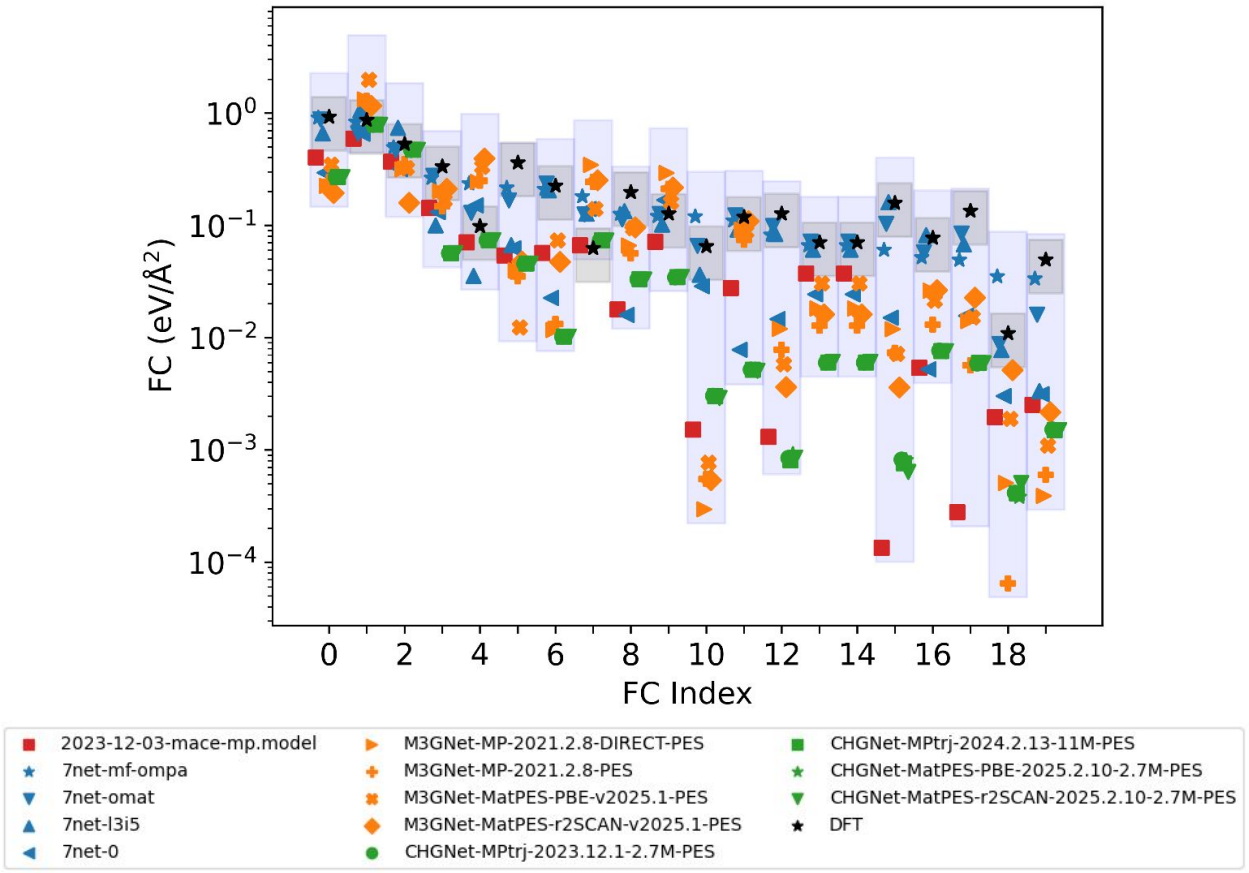

Figure S8: **Force constant (FC) training domain for Nb.** The 13 predicted FC tensors that yielded dynamically stable phonon dispersions were symmetry-reduced, and the resulting bounds were expanded to $[0.75 \times \min(FC_i),\ 2.5 \times \max(FC_i)]$ to define the training domain. This expanded range spans a broad set of physically plausible lattice dynamics and is intentionally asymmetric to facilitate identification of poorly performing models that collapse toward predicting the training-set mean. FC labels were then jointly and uniformly sampled from this domain and used in forward simulations to generate the corresponding inelastic neutron spectra. The DFT$\pm$50% range is shown in black for reference.

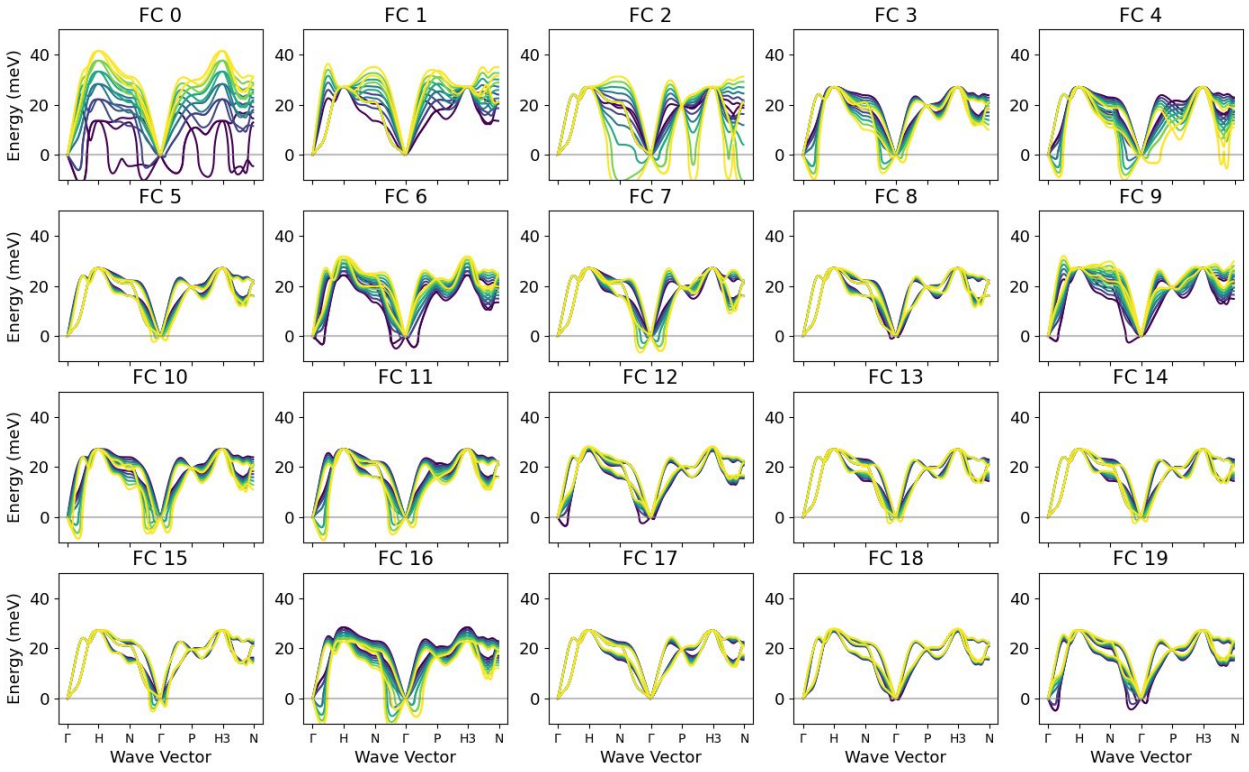

Figure S9: **Sensitivity of Nb phonon dispersions to the 20 largest symmetry-reduced FCs.** Force constants are ranked in descending order by magnitude. Line color denotes the perturbed FC value as a percentage of its sampled range, with all others held fixed. The noticeable response across multiple FCs indicates the presence of notable long-range interactions in Nb.

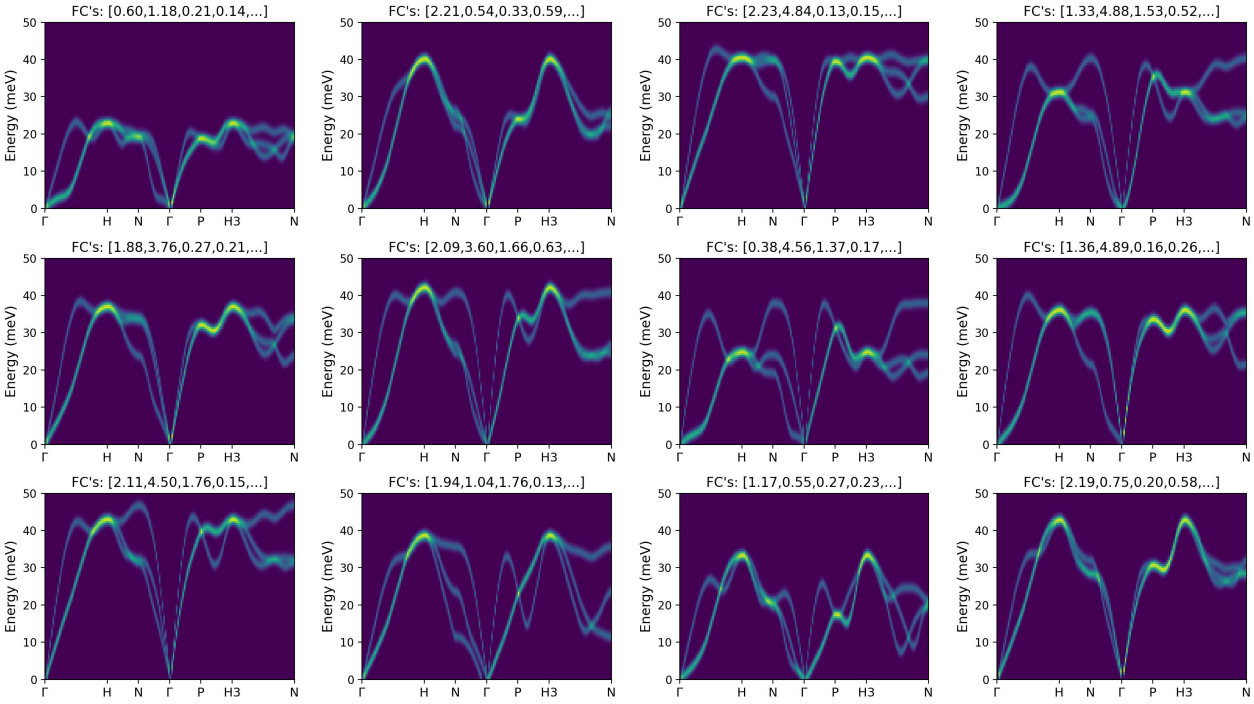

Figure S10: **Representative simulated $I(\mathbf{q}, E)$ spectra from the Nb training dataset.** Each subplot shows a randomly selected spectrum from the training dataset, together with its corresponding FC labels (shown in the title). The dataset spans a wide range of phonon behaviors and dynamical regimes.

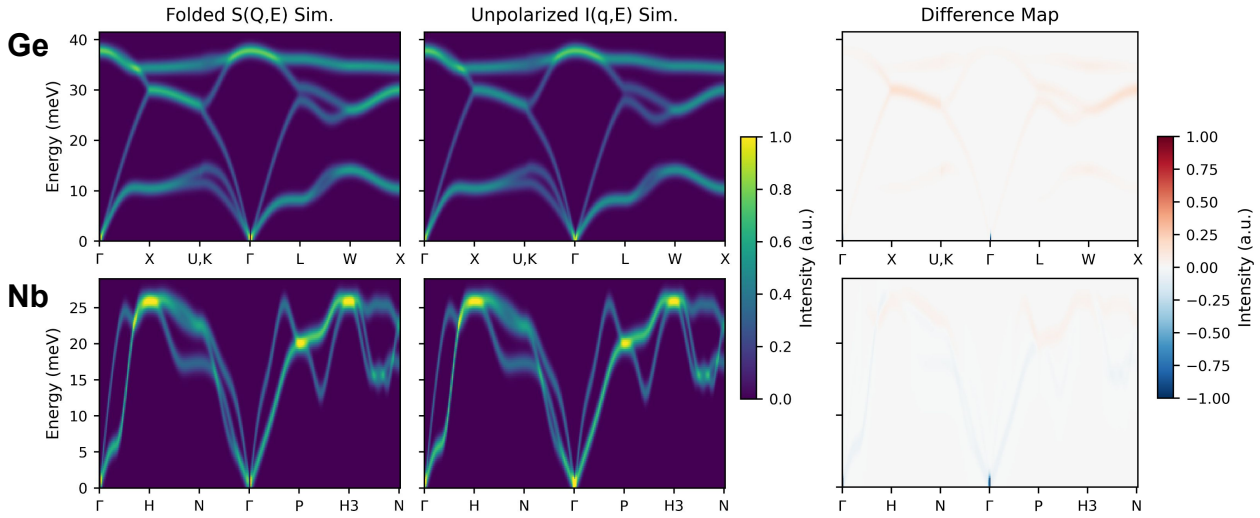

Figure S11: **Comparison of folded $\mathbf{S(Q,E)}$ and unpolarized forward-model spectra for Ge and Nb.** Top row: simulated Ge spectra showing (left) the fully computed $S(\mathbf{Q}, E)$ intensity summed over all symmetrically equivalent $\mathbf{Q}$-paths (*folded* $S(\mathbf{Q}, E)$), and (middle) the corresponding unpolarized $I(\mathbf{q}, E)$ spectrum obtained from the simplified forward model. The difference map from subtracting unpolarized from folded is also shown (right). Bottom row: analogous results for Nb. The close agreement between folded and unpolarized spectra demonstrates that Brillouin-zone folding effectively averages over polarization and other orientation-dependent effects, thereby justifying the use of the computationally efficient unpolarized model for large-scale dataset generation at low temperatures.

2 Model training details

Variational Autoencoder for spectral representation.
A fully connected variational autoencoder was implemented in PyTorch to learn compact latent representations of the simulated $I(\mathbf{q}, E)$ spectra prior to force-constant regression. The encoder consists of an input layer with dimension equal to the flattened spectrum length, followed by a 400-unit hidden layer with `tanh` activation. Two parallel linear heads produce the latent-space mean $\boldsymbol{\mu}$ and log-variance $\log \boldsymbol{\sigma}^2$ for a latent vector of dimension L. The decoder mirrors this structure symmetrically, mapping the latent variable $\mathbf{z} \in \mathrm{R}^L$ through a 400-unit hidden layer to reconstruct the normalized spectral intensity, using a `sigmoid` output activation to constrain values to $[0, 1]$.

The overall training objective combined a mean-squared reconstruction loss $\mathcal{L}_{\mathrm{rec}}$ and a KL divergence regularization term $\mathcal{L}_{\mathrm{KL}}$, weighted by a time-dependent factor β_t:

$$\mathcal{L}_{\mathrm{VAE}} = \mathcal{L}_{\mathrm{rec}} + \beta_t \, \mathcal{L}_{\mathrm{KL}}, \tag{S1}$$

with

$$\mathcal{L}_{\mathrm{rec}} = \frac{1}{N} \sum_i \|\hat{I}_i - I_i\|^2, \quad \mathcal{L}_{\mathrm{KL}} = -\frac{1}{2} \sum_j (1 + \log \sigma_j^2 - \mu_j^2 - \sigma_j^2). \tag{S2}$$

Here, β_t followed a cyclical annealing schedule over 50 epochs with a linear ramp during the first 40 epochs of each cycle, reaching a maximum value $\beta_{\max}$ tuned during hyperparameter search. This schedule mitigates posterior collapse by allowing the reconstruction term to dominate early training before gradually enforcing latent regularization. Models were trained for 200 epochs using the Adam optimizer (Kingma and Ba, 2017) (learning rate 10^{-3}, batch size 64) on a single compute node. Each epoch included both training and validation passes. Early stopping was disabled to ensure full-cycle Kullback–Leibler (KL) warm-up coverage. All runs were initialized with the same random seed to minimize stochastic variation.

Hyperparameters were tuned through systematic grid and sweep searches to ensure stable convergence and well-balanced latent representations. The search space included dataset size, latent-space dimension, and maximum KL-divergence weight ($\beta_{\max}$), each varied within representative physically and computationally feasible ranges. Because training and evaluation of the VAE are computationally demanding, the full parameter space was explored in targeted sub-grids focused on the most sensitive variables, as identified from preliminary experiments. Validation loss, reconstruction quality, and latent stability were used as selection criteria, and this strategy provided comprehensive coverage of the most influential hyperparameters while maintaining feasible training times.

VAE hyperparameter search space and final selections.

Parameter	Range Tested	Final Value
Dataset size	2,000–15,000	10,000
Latent dimension (L)	2–100	20 (Ge), 30 (Nb)
KL-weight $\beta_{\max}$	0–10^{-2}	10^{-6}

The total loss decreased smoothly during training over the first ∼150 epochs and reached equilibrium thereafter (see Fig. S12). The reconstruction loss $\mathcal{L}_{\text{rec}}$ decreased monotonically, while the KL term $\mathcal{L}_{\text{KL}}$ stabilized around a small positive value once the latent posterior distribution became well regularized. No evidence of posterior collapse (vanishing $\mathcal{L}_{\text{KL}}$) was observed, indicating that the warm-up schedule effectively balanced reconstruction fidelity and latent-space disentanglement.

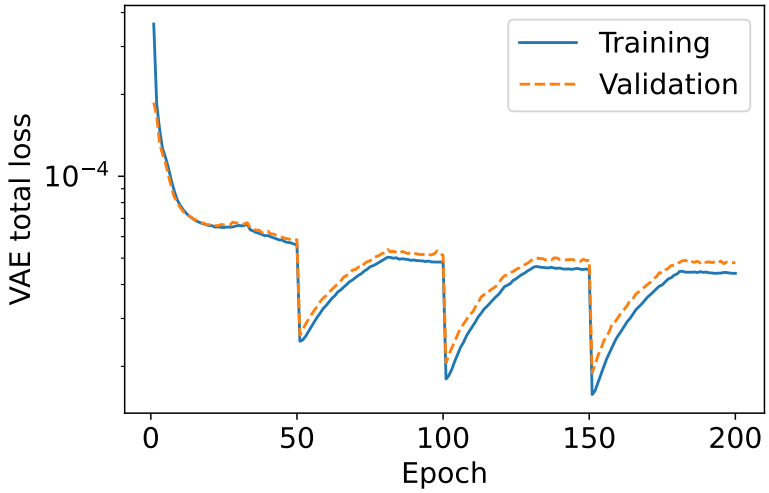

Figure S12: **VAE training dynamics.** Representative training and validation total loss curves for the VAE used to learn latent representations of $I(\mathbf{q}, E)$ spectra.

Latent-to-FC regressor.
Following VAE pretraining, a fully connected feed-forward network was trained to map the latent code $\mathbf{z}$ to the target set of symmetry-reduced FCs. The regression head comprised four hidden layers with dimensions of 100, 1000, 1000, and 100 neurons, each using `tanh` activation. A final linear output layer returned the vector of FCs corresponding to the training configuration.

Training used the Adam optimizer (Kingma and Ba, 2017) (learning rate 10^{-3}, batch size 64, 100 epochs) with mean-squared error loss and a `ReduceLROnPlateau` scheduler (factor 0.5, patience 10) to adaptively decrease the learning rate upon plateauing validation loss. All latent vectors were normalized, and inference used the encoder mean $\boldsymbol{\mu}$ rather than a sampled $\mathbf{z}$ to avoid stochasticity injection. Training and validation losses for the FC predictor converged smoothly, typically within approximately 75 epochs, and showed no evidence of overfitting (see Fig. S13). The final models used for inference were selected based on minimum validation MSE and good performance in the corresponding pre-trained VAE.

To motivate the dataset size used throughout this work, Fig. S14 shows the final training and validation losses as a function of dataset size for the VAE and the downstream FC predictor.

Transfer-learned ResNet-18 for direct FC regression.

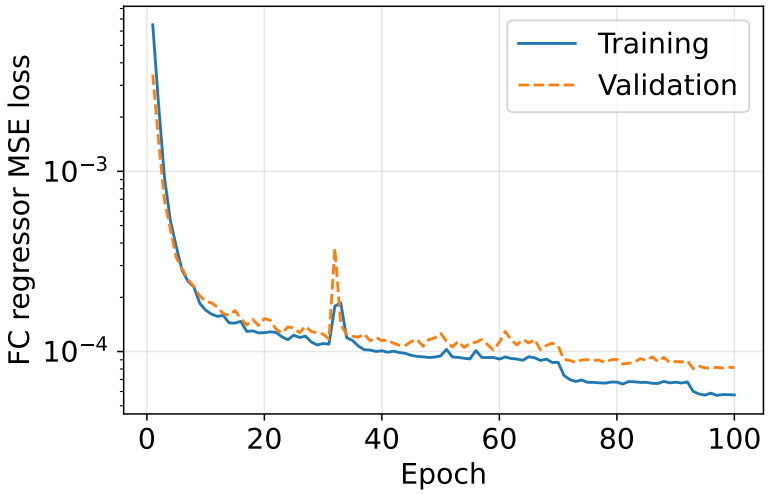

Figure S13: **FC regressor training dynamics.** Representative training and validation loss curves for the latent-to-FC predictor network trained on VAE latent representations.

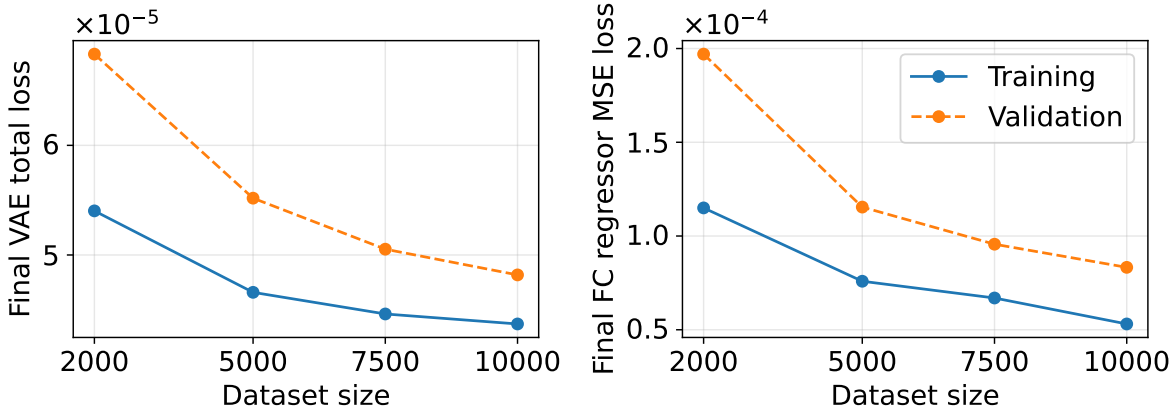

Figure S14: **Dataset size ablation.** Final (last-epoch) training and validation losses for the VAE (left) and the latent-to-FC regressor (right) as a function of the number of spectra comprising the dataset, using fixed hyperparameters.

For the direct inversion approach, a convolutional neural network based on the ResNet-18 architecture (He *et al.*, 2016) was employed to map the simulated $I(\mathbf{q}, E)$ spectra directly to the corresponding set of symmetry-reduced force constants (FCs). The model was initialized with weights pretrained on ImageNet and fine-tuned for regression using the spectral datasets. The final fully connected layer of ResNet-18 was replaced with a linear output head of dimension equal to the number of target FCs. All convolutional layers were retained to leverage pretrained feature extraction, while batch normalization and residual connections stabilized optimization on the relatively small spectral-domain dataset.

Each two-dimensional spectrum was expanded to three channels by duplication to match the input shape expected by ResNet-18 (`[B, 3, H, W]`). Force-constant labels were scaled prior to training using the reciprocal of their per-component median values across the training set:

$$\tilde{\mathbf{FC}} = \mathbf{FC} \odot \mathrm{median}(\mathbf{FC})^{-1}, \tag{S3}$$

where $\odot$ denotes elementwise multiplication. This scaling equalized the dynamic range across FC components, ensuring balanced gradient magnitudes during optimization and preventing larger-magnitude FCs from dominating the loss. During validation and inference, predicted FCs were rescaled back to their original units using the same median factors.

The regression objective minimized the mean-squared error (MSE) between predicted and true FC vectors, defined as

$$\mathcal{L}_{\mathrm{MSE}} = \frac{1}{N} \sum_i \|\hat{\mathbf{FC}}_i - \mathbf{FC}_i\|^2. \tag{S4}$$

The loss was computed in the scaled FC space for backpropagation but tracked in physical (unscaled) units for interpretability. Models were trained for 75 epochs using stochastic gradient descent (SGD) with momentum $m = 0.9$, weight decay $\lambda = 10^{-4}$, and learning rate $\eta = 10^{-3}$. A batch size of 5 was used for the final reported models. The `StepLR` scheduler progressively reduced the learning rate by a factor of 0.95 each epoch:

$$\eta_t = \eta_0 \, (0.95)^t, \tag{S5}$$

where t is the epoch index.

Hyperparameters were tuned through systematic grid and sweep searches to ensure stable convergence and generalization. The search covered dataset size, learning rate, momentum, weight decay, and batch size, each varied within representative, computationally feasible ranges. Because of the computational cost of training large convolutional networks, the full hyperparameter space was explored in sub-grids targeting the most sensitive parameters, as identified from preliminary experiments. Validation loss and convergence stability served as the selection criteria, and the configuration yielding the lowest validation loss with smooth convergence is reported below.

Training and validation losses decreased relatively monotonically over the first $\sim$50 epochs and converged thereafter (see Fig. S15). The learning-rate schedule effectively mitigated overfitting and improved validation consistency across random seeds. The final reported

ResNet-18 hyperparameter search space and final selections.

Parameter	Range Tested	Final Value
Dataset size	5,000–50,000	10,000
Learning rate (η)	10^{-3}–10^{-1}	10^{-3}
Momentum (m)	0.1–0.9	0.9
Weight decay (λ)	10^{-8}–10^{-2}	10^{-4}
Batch size	1–500	5

models correspond to the checkpoints at 50 epochs before any observable overfitting, and the pretrained convolutional filters successfully transferred to the spectral regression task without over-adaptation.

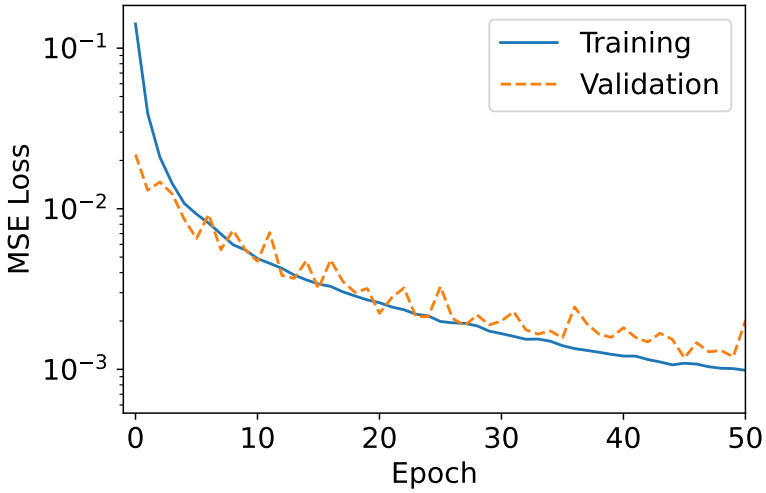

Figure S15: **ResNet-18 training dynamics.** Representative training and validation loss curves for the transfer-learned ResNet-18 regression model trained to map $I(\mathbf{q}, E)$ spectra directly to symmetry-reduced FCs.

3 Model performance, robustness, and interpretability

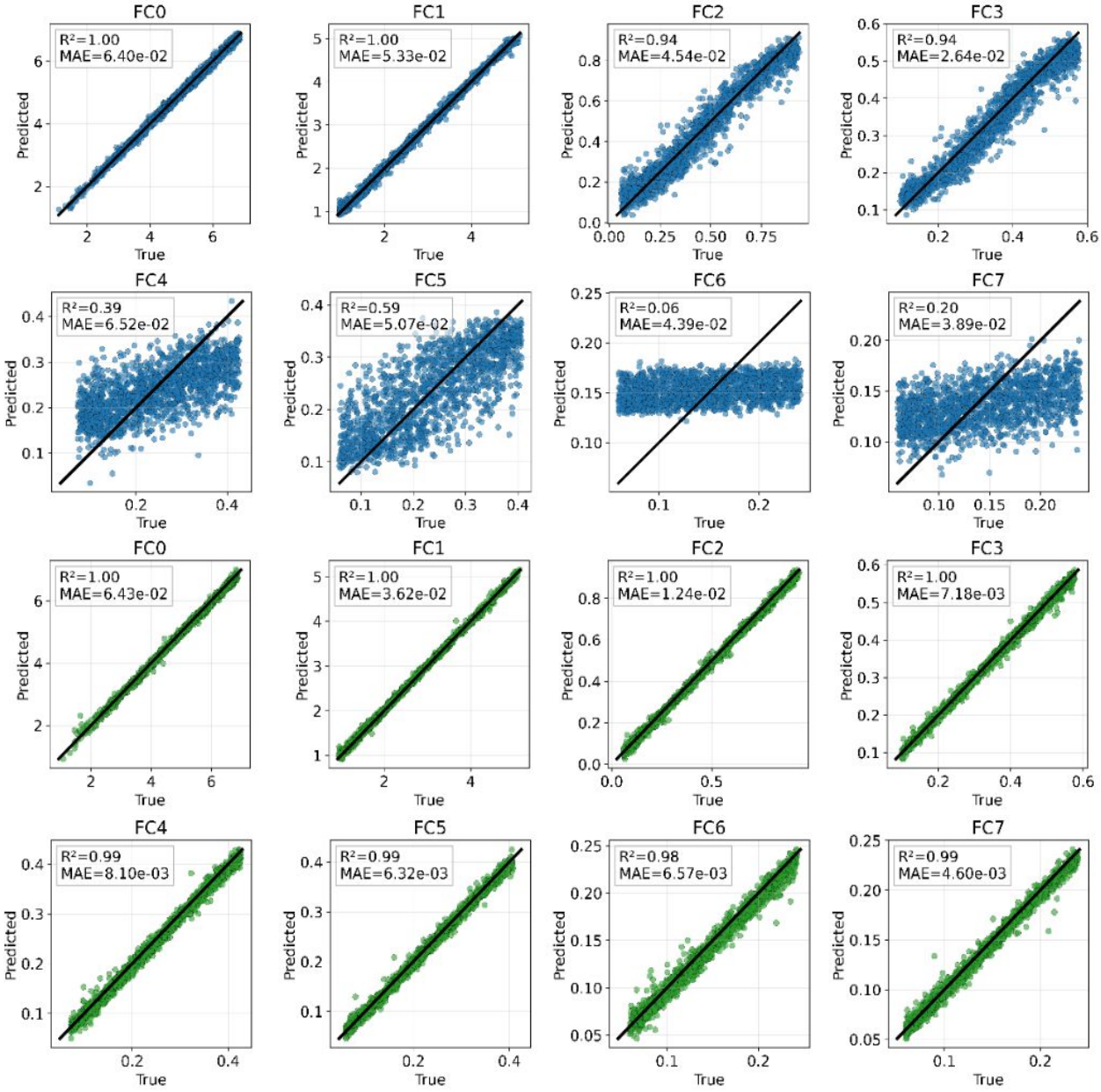

Figure S16: **Ge force-constant predictions on clean data.** Parity plots showing true values versus predicted force constants from the VAE-based model (top, blue) and the ResNet-18 model (bottom, green) using 1,500 simulated test spectra. Reported metrics include R^2 and mean absolute error (MAE). The black line indicates perfect agreement.

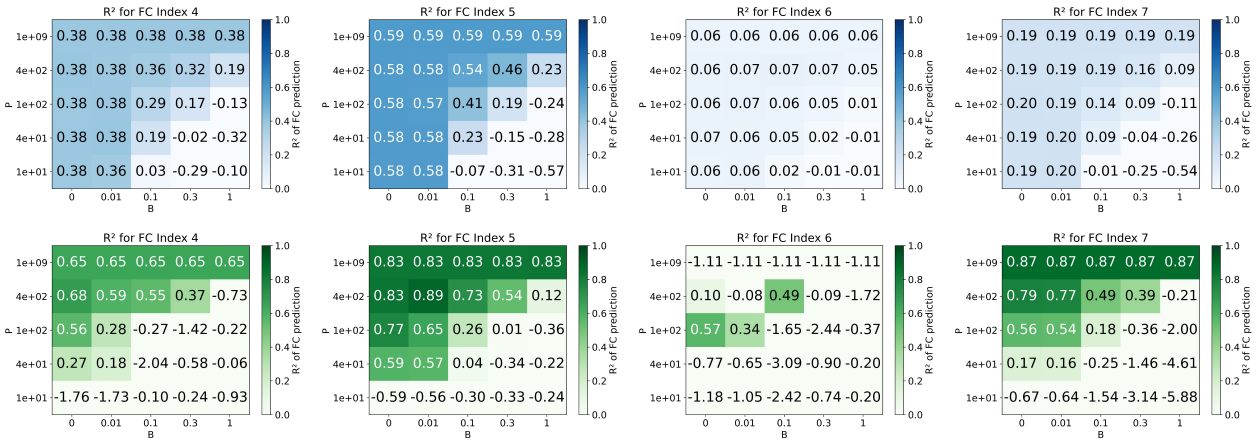

Figure S17: **Prediction robustness to noise for less influential Ge force constants.** Heatmaps of R^2 scores for VAE-based (top, blue) and ResNet-based (bottom, green) predictions of Ge FC indices 4–7 under all noise conditions (see Fig. S21). Indices 0–3, corresponding to the more influential FCs, are shown in the main text.

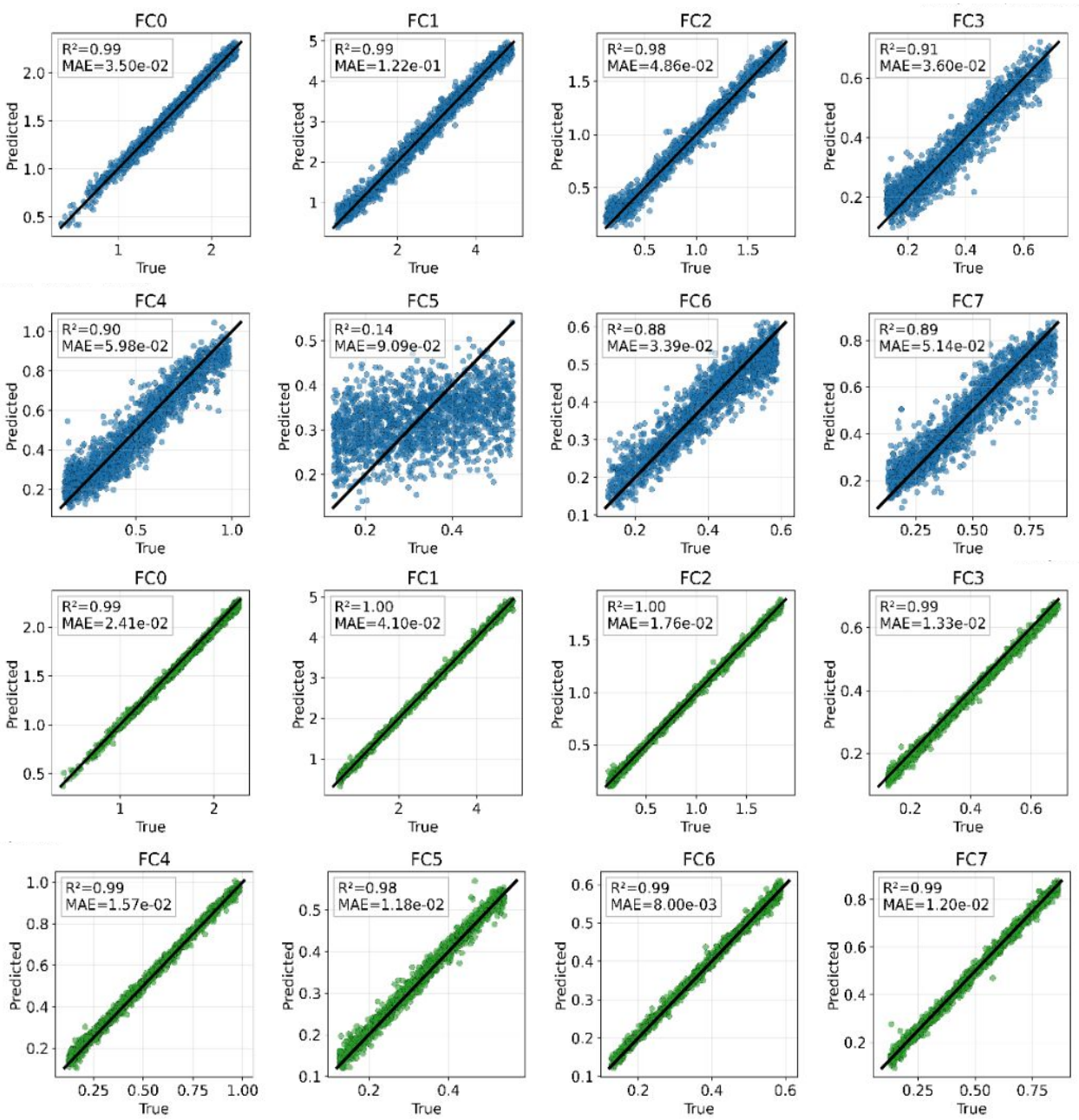

Figure S18: **Nb force-constant predictions on clean data.** Parity plots showing true values versus predicted force constants from the VAE-based model (top, blue) and the ResNet-18 model (bottom, green) using 1,500 simulated test spectra. Reported metrics include R^2 and mean absolute error (MAE). The black line indicates perfect agreement.

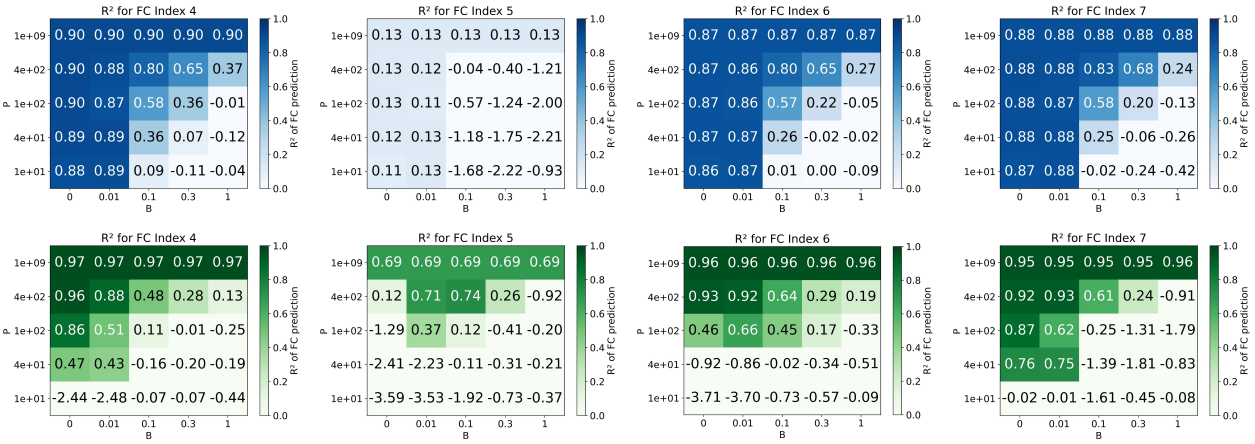

Figure S19: **Prediction robustness to noise for less influential Nb force constants.** Heatmaps of R^2 scores for VAE-based (top, blue) and ResNet-based (bottom, green) predictions of Nb FC indices 4–7 under all noise conditions (see Fig. S23). Indices 0–3, corresponding to the more influential FCs, are shown in the main text.

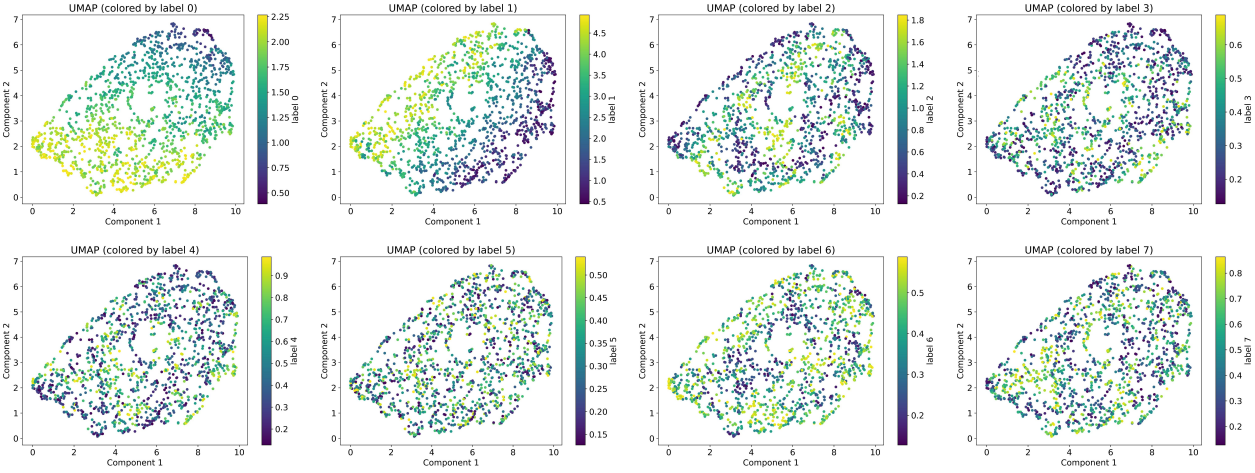

Figure S20: **Nb latent-space structure and interpretability.** Two-dimensional UMAP (McInnes *et al.*, 2018) projections of the latent encodings (μ) obtained by inputting simulated test spectra into the trained encoder. The encodings are colored by individual force-constant values, revealing a semantically and physically organized latent space, particularly for the most influential FCs.

4 Spectral noise implementation

To assess the robustness of models trained on clean (noise-free) data to domain shifts introduced by experimental noise, synthetic perturbations were applied directly to the simulated $I(\mathbf{q}, E)$ spectra used for testing (see Figs. S21–S24). The noise model was designed to approximate the dominant sources of variation in inelastic neutron scattering data, namely a constant background signal and Poisson-distributed counting statistics.

Noise injection was performed using a two-stage transformation implemented in PyTorch. Each spectrum was first normalized by a global scaling constant estimated from the training dataset to maintain approximate intensity range consistency prior to noise injection. A constant background offset B was then added, followed by sampling from a Poisson process with scale factor P that emulates the discrete counting statistics of neutron detectors. In code, the background-adjusted signal $I' = I + B$ was multiplied by P and converted to integer counts, from which pixel-wise Poisson samples were drawn. The result was divided back by P to restore the original scale. After sampling, the noisy spectrum was post-processed by subtracting its fifth intensity percentile (to remove residual baseline bias) and renormalized by its 99.5^{th} percentile value to confine intensities to approximately $[0, 1]$.

For diagnostic purposes, a per-spectrum signal-to-noise ratio was computed during noise generation to quantify the relative contributions of stochastic and deterministic perturbations. After noise injection, the constant background offset B was subtracted from the noisy spectrum I_{noisy}, and the SNR was computed as

$$\text{SNR} = \frac{\mathbb{E}[I^2]}{\mathbb{E}[(I_{\text{noisy}} - I)^2]}, \tag{S6}$$

where I is the clean, normalized input spectrum and I_{noisy} is the perturbed output following Poisson sampling and baseline removal. This formulation measures the ratio of the average signal power to the noise power, isolating the stochastic contribution of the Poisson process while excluding the deterministic background component, since the background offset was already incorporated into the Poisson sampling step. The computed SNR values were used solely for visualization and analysis and were not used during model training or evaluation.

Noise transforms were applied dynamically through a custom `Dataset` class. For each sample index i, an `IndexAwareNoise` object generated a unique random seed (`base_seed + i`) to ensure deterministic yet non-repeating noise realizations across the dataset splits (train/validation/test). This framework enabled consistent augmentation while preserving reproducibility and ground-truth FC label correspondence.

To systematically probe noise sensitivity, a 5×5 parameter grid was defined with $B \in [0, 1.0]$ and $P \in [0, 10^9]$. Each combination corresponds to a unique pairing of constant background level and Poisson count scaling, yielding 25 distinct noise conditions per spectrum. Representative examples of Ge and Nb spectra perturbed under the full 5×5 noise grid are shown in Figs. S21–S24, together with corresponding one-dimensional $I(E)$ cuts at various columns that illustrate the relative noise intensities and background effects.

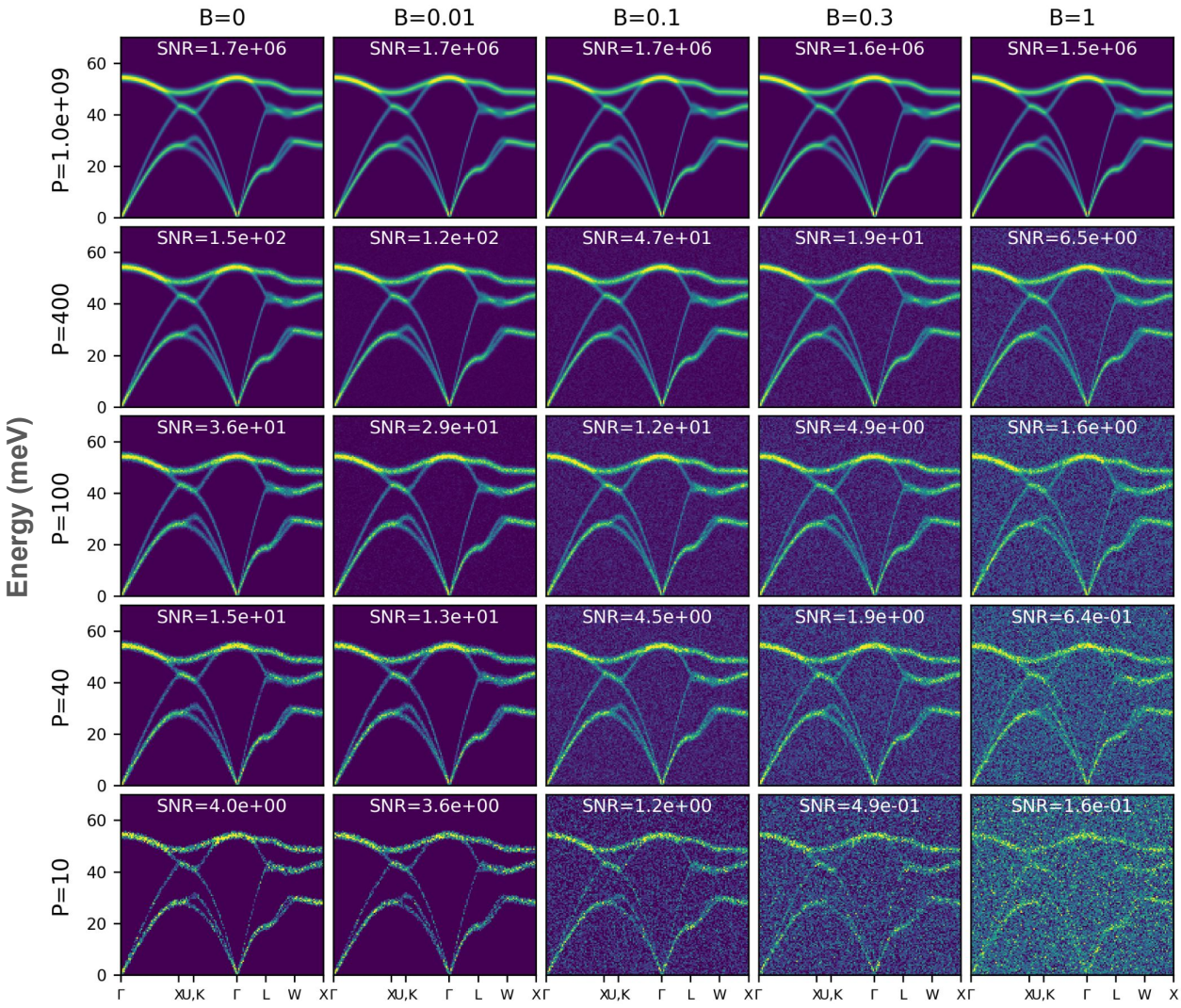

Figure S21: **Example Ge spectrum under the full 5×5 grid of noise conditions used for noise-robustness analysis.** Each panel corresponds to a unique combination of constant background (B) and Poisson counting noise (P). Signal-to-noise ratios (SNRs) are indicated for each input, illustrating the relative contribution of simulation and stochastic noise components.

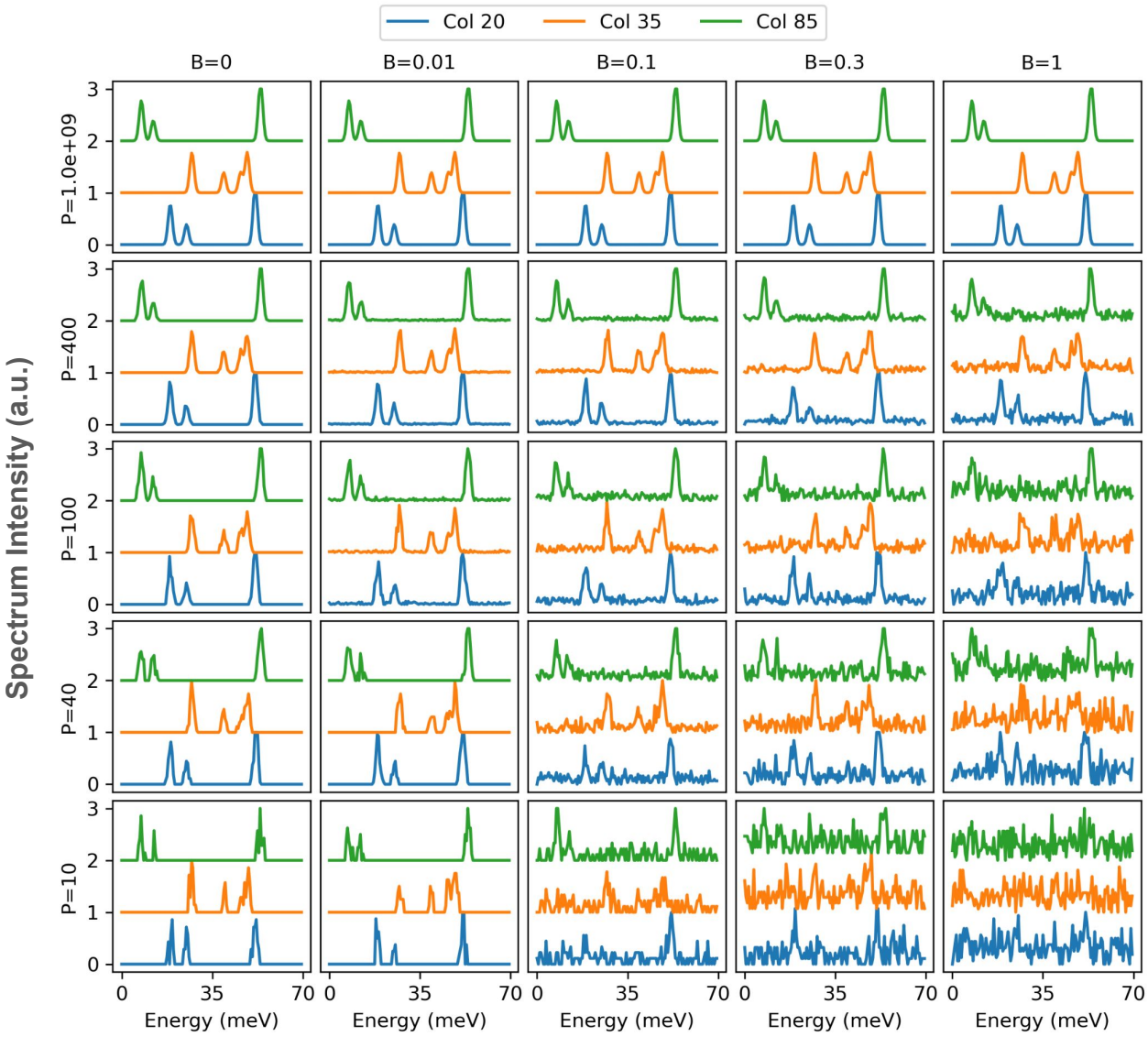

Figure S22: **One-dimensional $I(E)$ cuts of the example Ge spectrum under the full 5×5 grid of noise conditions.** The spectra shown in Fig. S21 are sliced along columns 20 (blue), 35 (orange), and 85 (green), and the resulting intensity profiles are overlaid to provide a quantitative view of the noise levels. Constant background (B) and Poisson counting noise (P) values are labeled consistently with Fig. S21.

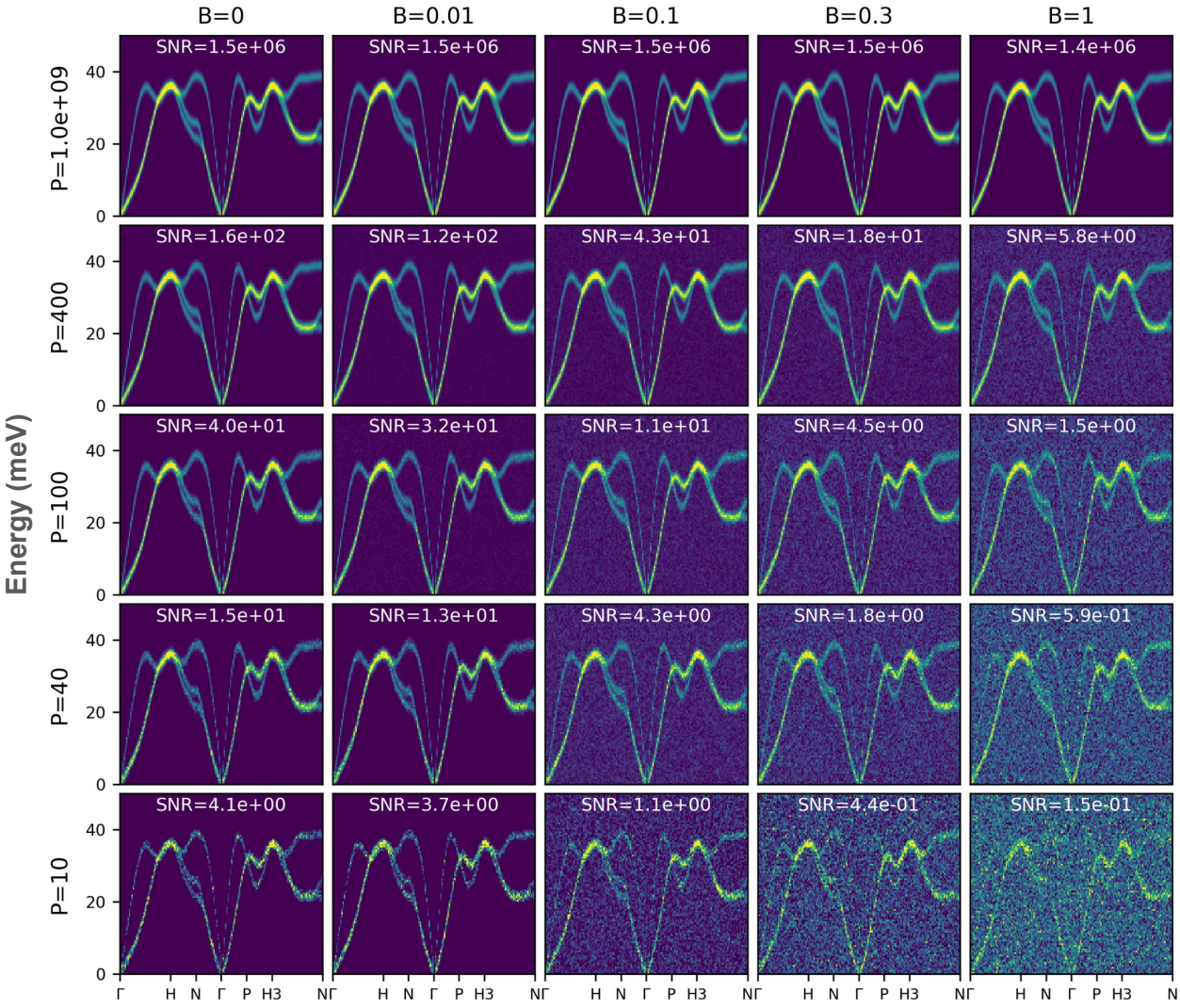

Figure S23: **Example Nb spectrum under the full 5×5 grid of noise conditions used for noise-robustness analysis.** Each panel corresponds to a distinct combination of constant background (B) and Poisson counting noise (P), with signal-to-noise ratios (SNRs) indicated in each case.

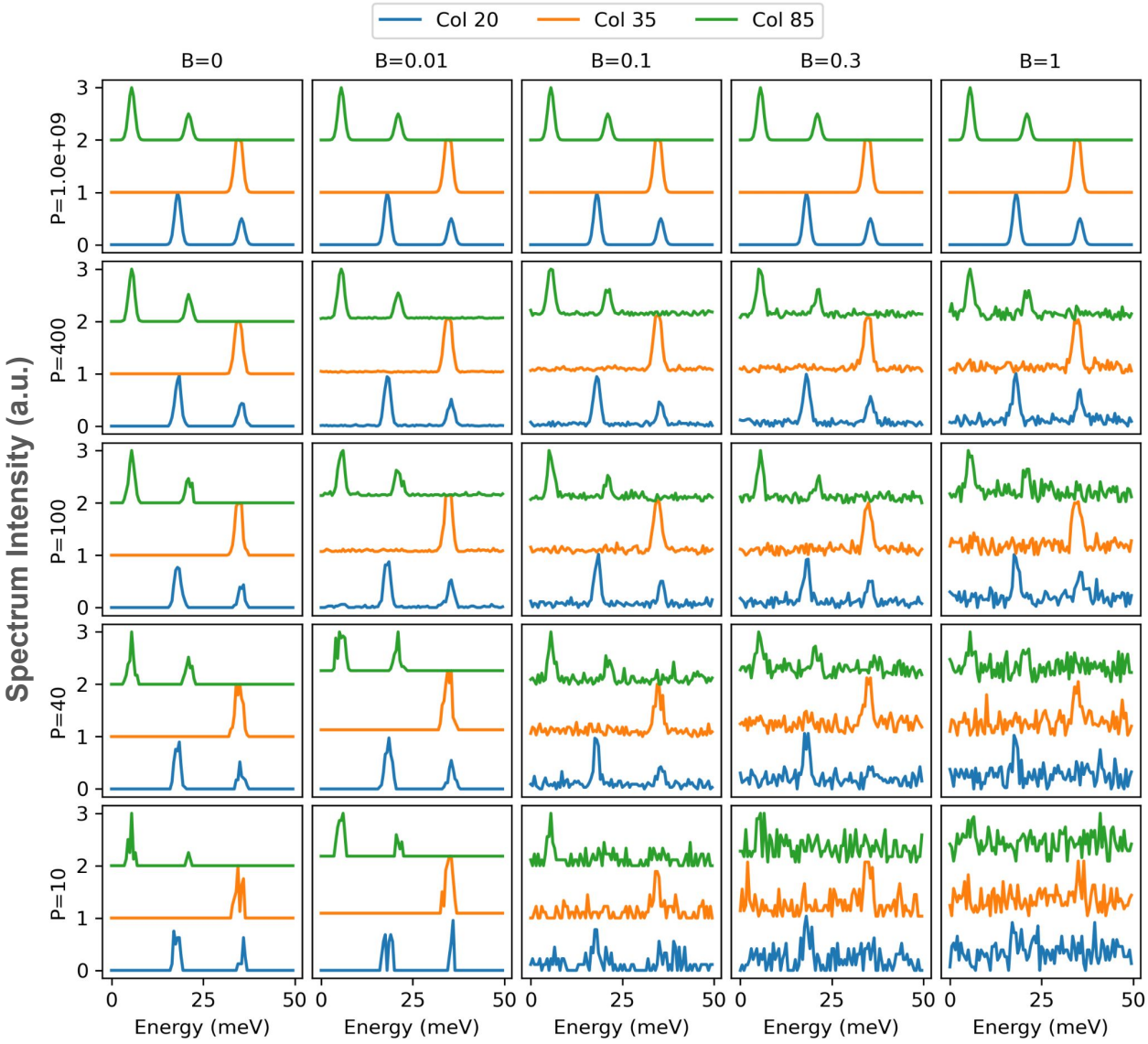

Figure S24: **One-dimensional $I(E)$ cuts of the example Nb spectrum under the full 5×5 grid of noise conditions.** The spectra in Fig. S23 are sliced along columns 20 (blue), 35 (orange), and 85 (green), and the overlaid signal profiles highlight how the relative levels of additive (B) and Poisson (P) noise affect peak definition.

5 Ge ARCS experimental elastic Q map

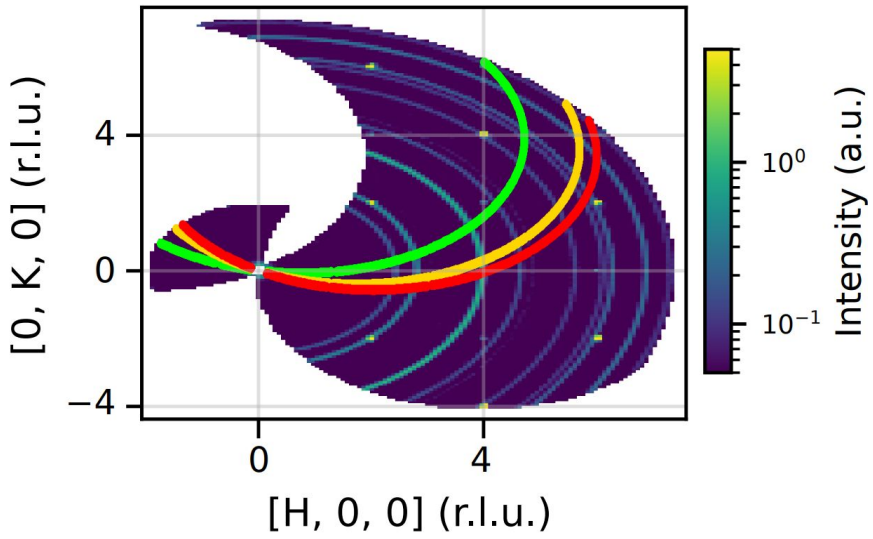

Figure S25: Elastic **Q**-space coverage for the full Ge dataset measured on ARCS at the SNS (Abernathy *et al.*, 2012). Three fixed sample orientation scans are shown at $\omega = -79°$ (green), $-64°$ (yellow), and $-59°$ (red), which correspond to the best, median, and worst single-scan performance based on FC prediction error relative to the full-dataset result.

6 Iterative force constant optimization

To provide a baseline for comparison with the learned inversion models, an iterative force-constant (FC) optimization was performed using a direct forward–model fitting approach. This traditional optimization establishes how well the underlying spectral model can recover FC values when guided only by experimental data, without machine-learning inversion.

The differential-evolution algorithm (Storn and Price, 1997) implemented in the `scipy.optimize` package was employed to minimize the least-squares discrepancy between forward-simulated and experimental $I(\mathbf{q}, E)$ spectra. Differential evolution is a population-based, gradient-free method that is well suited for this task because it efficiently explores high-dimensional, non-linear parameter spaces and avoids premature convergence to local minima. The algorithm was run with its default settings (population size proportional to the number of free parameters, mutation constant 0.5–1.0, crossover probability 0.7), which were found to perform robustly without additional tuning. Owing to its stochastic nature, ten independent trials were conducted to assess reproducibility and convergence variability.

The optimization objective was defined as the mean-squared error between the forward-simulated and experimental intensity maps,

$$\chi^2 = \frac{1}{N} \sum_i \left[I_{\mathrm{sim}}(\mathbf{q}_i, E_i) - I_{\mathrm{exp}}(\mathbf{q}_i, E_i) \right]^2 . \tag{S7}$$

For each candidate parameter vector proposed by the optimizer, the corresponding FC tensor was generated by scaling the symmetry-reduced FC components within the full tensor, reapplying acoustic sum rules to enforce translational invariance, and writing the result as a `FORCE_CONSTANTS` file compatible with `phonopy` (Togo, 2023; Togo *et al.*, 2023). The updated tensor was then passed to the forward simulation routine to compute the corresponding $I(\mathbf{q}, E)$, from which χ^2 was evaluated.

The search parameters consisted of the eight largest symmetry-reduced FC components extracted from the MLFF-derived tensors, which were identified in prior sensitivity analyses as having the dominant impact on phonon dispersion. The utilized search domain was matched exactly to the FC sampling range used for the machine-learning training data, ensuring a consistent parameter space across all inversion approaches. Initial parameter vectors were randomly sampled within these bounds to promote diverse starting populations among the ten independent trials.

The evolution of the objective function across all trials (Fig. S26) shows steady improvement during the early iterations, with intermittent discovery of new minima as indicated by red markers. The corresponding FC parameter trajectories (Fig. S27) reveal that the optimizer rapidly constrains the largest FCs to narrow subdomains of the search range, while smaller-magnitude FCs remain more weakly determined owing to their limited influence on the dispersion. To obtain a statistically robust estimate of the optimized FC values, parameter samples corresponding to the lowest 1^{st} percentile of objective-function evaluations were aggregated across all trials (Fig. S28). The resulting distributions highlight the degree of certainty for each FC and collectively define the best-fit harmonic FC tensor obtained from direct forward optimization.

This iterative differential-evolution procedure provides a physically interpretable reference for evaluating the learned inversion models. Although computationally expensive relative to the machine-learning approaches, it offers valuable insight into the structure of the optimization landscape and provides an avenue to benchmark the consistency between ML-driven and forward-fit FC predictions.

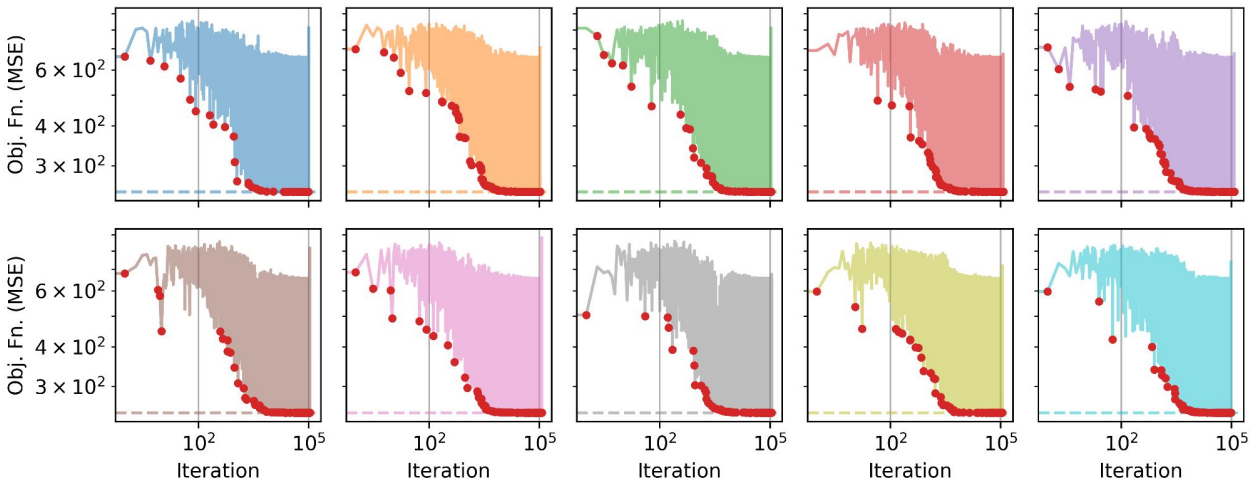

Figure S26: **Objective function convergence across iterative optimization trials.** Objective function values are shown as a function of iteration number for ten independent optimization trials. Red markers denote iterations at which a new minimum objective function value was achieved, illustrating the stochastic convergence behavior and relative progression of the optimization process.

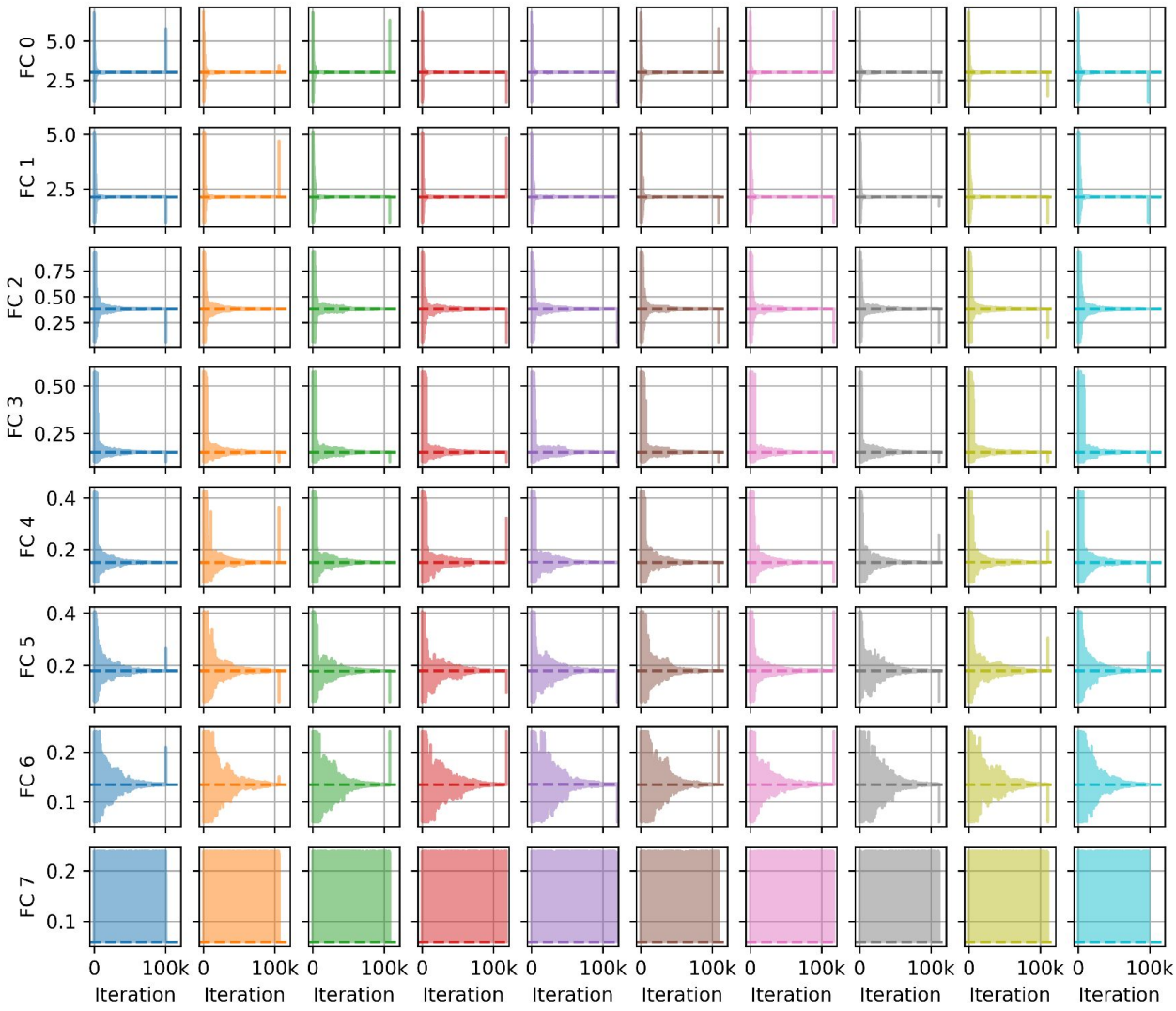

Figure S27: **Evolution of force-constant parameters during iterative optimization.** Each curve represents the progression of an individual symmetry-reduced FC parameter over the course of the optimization for all ten trials. The y-axis limits correspond to the total FC training domain, and horizontal dashed lines indicate the FC values associated with the minimum objective function achieved in each trial, providing a visual reference for convergence behavior and parameter stability.

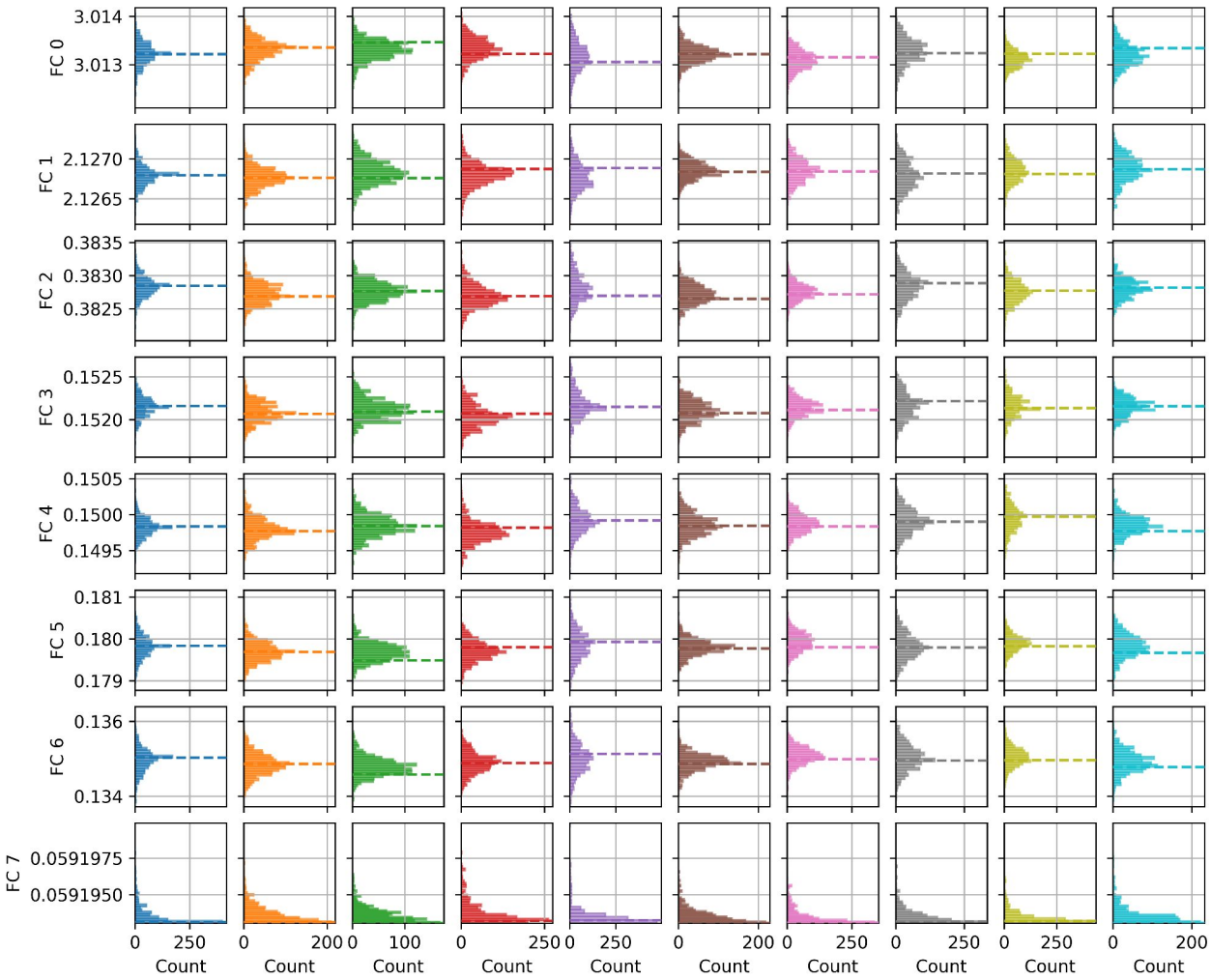

Figure S28: **Distribution of optimal force-constant parameters across trials.** Shown are the distributions of FC parameter guesses corresponding to the lowest 1st percentile of objective function evaluations aggregated over all ten trials. Horizontal dashed lines denote the FC values corresponding to the minimum objective function for each individual trial. These distributions highlight the slight variability in converged solutions and the sensitivity of the optimization landscape.

References

D. P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, 2017, `http://arxiv.org/abs/1412.6980`, arXiv:1412.6980 [cs].

K. He, X. Zhang, S. Ren and J. Sun, Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.

L. McInnes, J. Healy, N. Saul and L. Großberger, *Journal of Open Source Software*, 2018, **3**, 861.

D. L. Abernathy, M. B. Stone, M. J. Loguillo, M. S. Lucas, O. Delaire, X. Tang, J. Y. Y. Lin and B. Fultz, *Review of Scientific Instruments*, 2012, **83**, 015114.

R. Storn and K. Price, *Journal of global optimization*, 1997, **11**, 341–359.

A. Togo, *Journal of the Physical Society of Japan*, 2023, **92**, 012001.

A. Togo, L. Chaput, T. Tadano and I. Tanaka, *Journal of Physics: Condensed Matter*, 2023, **35**, 353001.