



Cite this: DOI: 10.1039/d6dd00008h

# Machine learning inversion of interatomic force constants from single-crystal inelastic neutron scattering

Aiden Sable,<sup>a</sup> Bander Linjawi,<sup>a</sup> Kyle Bradbury,<sup>b</sup> Jordan Malof<sup>c</sup>  
and Olivier Delaire<sup>d</sup>\*

Atomic vibrations govern many macroscopic properties of materials, but experiments to comprehensively probe them remain challenging. Inelastic neutron scattering (INS) is a powerful technique to map phonon dispersions in crystals, especially when leveraging modern time-of-flight (ToF) spectrometers with large detectors. However, efficiently and robustly extracting interatomic force constants (FCs) parameterizing phonon dynamics from experimental spectra remains a bottleneck due to the complexity and high dimensionality of ToF INS datasets. Here, we present a machine learning approach for the direct inversion of FCs from single-crystal INS measurements. The framework leverages synthetic training data generated using universal machine-learned force fields and an efficient physics-based forward model. We benchmark two neural architectures—one emphasizing structured latent representation learning and the other direct, supervised spectral regression—across simulated datasets for two materials under idealized and noisy conditions. The latent-representation model is subsequently applied to experimental single-crystal INS data on germanium. The model is shown to reproduce FCs derived from both first-principles simulations and from iterative optimization, and furthermore achieves reliable inference even from sparse, single-orientation measurements representing short data acquisitions. Analysis of the learned latent space reveals semantically continuous and physically interpretable encodings that support strong cross-domain generalization. By bridging theoretical and experimental domains, we establish a path toward rapid inversion of experimental spectra and data-driven interpretation of temperature-dependent lattice dynamics.

Received 9th January 2026

Accepted 16th March 2026

DOI: 10.1039/d6dd00008h

rsc.li/digitaldiscovery

## Introduction

Atomic structure and dynamics play a pivotal role in determining many key material properties, such as thermodynamics,<sup>1–3</sup> thermal conductivity,<sup>4,5</sup> superconductivity,<sup>6,7</sup> and structural phase transitions.<sup>8,9</sup> As a result, gaining a more comprehensive understanding of atomic vibrations through both experimental and theoretical approaches has become a central goal in materials research.<sup>10–12</sup> A key link between vibrational models and experimental spectra lies in the interatomic force constants (FCs), which parameterize the Taylor expansion of the ionic potential energy surface about a reference crystal structure and allow for a compact encoding

of phonon behavior.<sup>13,14</sup> While higher-order FCs control anharmonic phenomena, including phonon lifetimes and therefore thermal conductivity, or soft-mode phase transitions, our focus in this work is restricted to the harmonic FCs that primarily govern the phonon dispersion. Since the mid-20th century, harmonic interatomic FCs have been determined by applying lattice-dynamical and Born–von Kármán-type models to phonon spectra obtained from fitted peak positions measured using triple-axis neutron spectrometry,<sup>15–17</sup> providing early foundations for quantitative structure–dynamics relationships grounded in scattering experiments. With modern instrumentation, single-crystal inelastic neutron scattering (INS) using ToF direct-geometry spectrometers with large pixelated detector banks and powerful spallation neutron sources<sup>18–21</sup> is now enabling comprehensive mapping of the four-dimensional momentum ( $\mathbf{Q}$ ) and energy ( $E$ )-dependent dynamical structure factor  $S(\mathbf{Q}, E)$  of materials.<sup>22</sup> Parallel advances in theoretical modeling, such as density functional theory, finite-temperature lattice-dynamics methods,<sup>23–25</sup> and emerging universal machine-learning force fields (uMLFFs),<sup>26,27</sup> have yielded increasingly accurate simulations that complement INS and have deepened understanding of atomistic

<sup>a</sup>Department of Mechanical Engineering and Materials Science, Duke University, Durham, NC, USA. E-mail: olivier.delaware@duke.edu

<sup>b</sup>Department of Electrical and Computer Engineering, Duke University, Durham, NC, USA

<sup>c</sup>Department of Electrical Engineering and Computer Science, University of Missouri–Columbia, Columbia, MO, USA

<sup>d</sup>Department of Chemistry, Duke University, Durham, NC, USA

\*Department of Physics, Duke University, Durham, NC, USA



behavior, including anharmonic effects. Together, these developments now make it possible to map atomic vibrations with unprecedented detail.

However, recovering interatomic FCs directly from experimental spectra remains a significant challenge, especially given the information-rich, high-dimensional, and anisotropic nature of modern single-crystal measurements. Because solving this inversion problem provides a powerful route to enable studies of temperature-dependent lattice dynamics, several approaches have been proposed to estimate interatomic FCs using physics-based or probabilistic frameworks. Among these, Bayesian inference schemes based on replica-exchange Markov chain Monte Carlo have been used to recover FCs from synthetic phonon dispersion data.<sup>28,29</sup> Although these methods achieved good performance on a simplified phonon dispersion model, they were computationally intensive, required significant statistical expertise, and lacked validation on experimental data. Beyond probabilistic inference, in prior work,<sup>30</sup> a physics-informed hierarchical optimization approach was developed to extract FCs from single-crystal INS spectra by iteratively minimizing the error between measured and simulated  $S(\mathbf{Q}, E)$  by optimizing a reduced set of symmetry-unique FC parameters. This method successfully reproduced experimental phonon dispersions in silicon, but its overall scalability and automation were limited: it required manual data selection and processing, full forward  $S(\mathbf{Q}, E)$  simulations at each optimization step, and *a priori* DFT calculations that may not be readily available in experimental settings or feasible for certain material systems. More broadly, current practices for extracting interatomic FCs from large experimental INS datasets are hampered by the lack of robust workflows to connect the measured  $S(\mathbf{Q}, E)$  intensities in large four-dimensional  $(\mathbf{Q}, E)$  domains with a real-space model of lattice dynamics. This need motivates the development of quantitative approaches that can efficiently and robustly extract deeper physical insights from complex experiments, and that could be leveraged to better optimize measurements.

Machine learning (ML) methods have been increasingly adopted across condensed matter physics and the scattering sciences, offering new pathways for automation, acceleration, and physical model extraction.<sup>31–34</sup> Further, recent studies have demonstrated that neural networks can accurately predict phonon spectra and densities of states from crystal structures,<sup>35–38</sup> underscoring the maturity of ML-based forward modeling. In contrast, ML-based approaches for inverting scattering data to recover physically meaningful parameters have received comparatively less attention. Chang *et al.*<sup>39</sup> demonstrated that neural networks can extract effective interactions from small-angle neutron scattering data, supporting inverse modeling in soft matter systems. In single-crystal diffuse scattering, autoencoders have been employed to infer magnetic Hamiltonian parameters in spin-ice systems by learning compressed representations of three-dimensional scattering patterns.<sup>40,41</sup> Within phonon dynamics, Su and Li<sup>42</sup> trained a variational autoencoder (VAE) exclusively on DFT-based powder INS simulations to recover interatomic FCs in aluminum, demonstrating the promise of self-supervised ML

for inverse phonon problems. Despite this progress, existing efforts remain confined to diffraction or powder data, which are inherently lower-dimensional and lack energy-resolved information and directional momentum resolution, respectively. By contrast, single-crystal INS comes with substantial experimental and computational challenges: data reduction and processing are complex, information content and signal quality vary strongly across the data volume, and analogous forward modeling is computationally demanding. These factors have so far precluded the development of ML-based inversion methods capable of capitalizing on modern experimental single-crystal INS data.

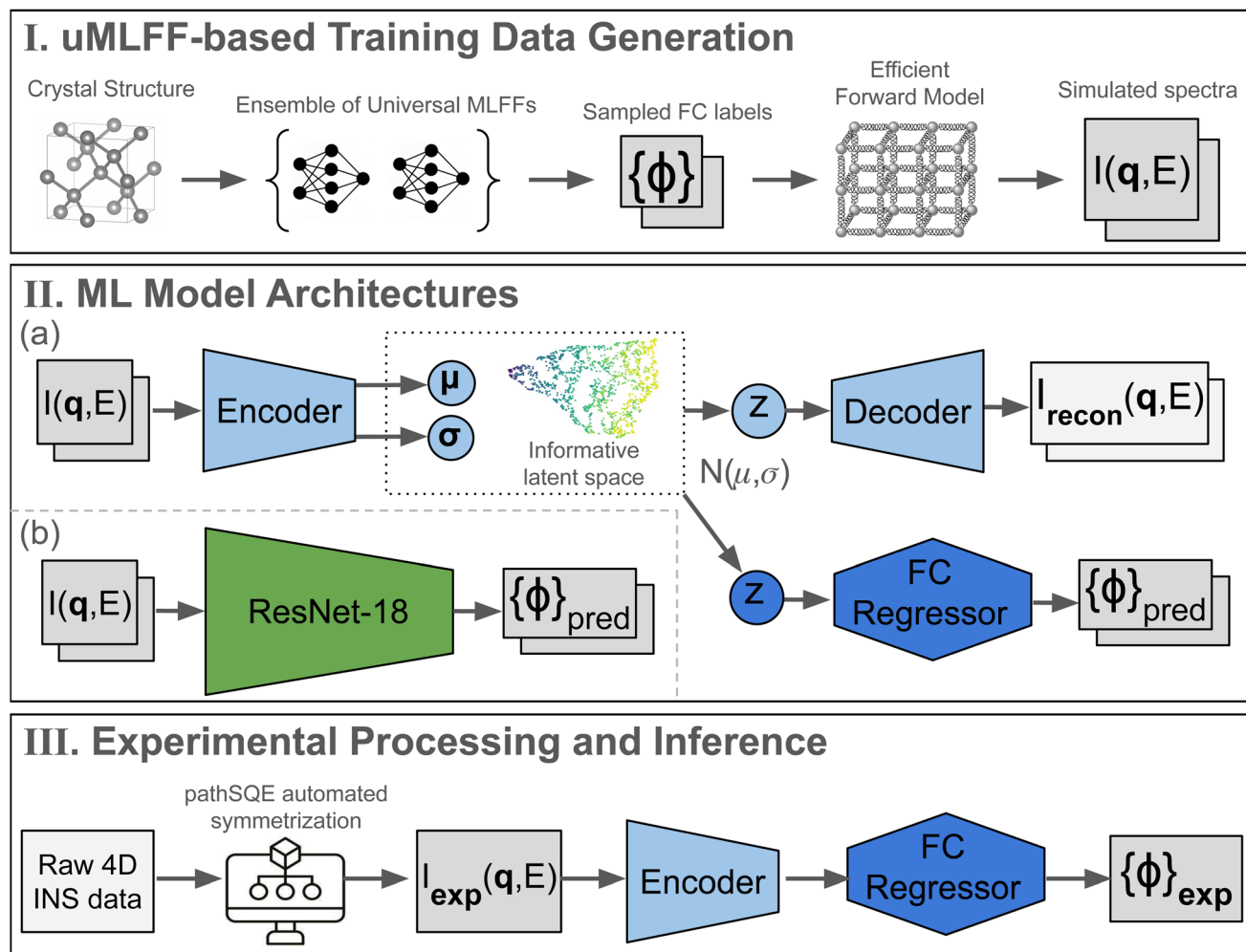
In this work, we present an ML approach for the inversion of interatomic FCs from single-crystal INS data. Our dataset generation procedure relies on uMLFFs to define the bounds of the FC domain, and is made computationally inexpensive due to an efficient physics-based forward model. For the inversion task, we evaluate both a dual model architecture, consisting of a VAE and a separate FC regressor, and a transfer-learned ResNet-18 architecture, benchmarking their performance on two material systems, Ge and Nb. During inference with realistic spectral noise augmentation, we find that the VAE-based approach exhibits superior robustness and domain adaptability. We validate our framework on experimental single-crystal INS measurements of Ge, demonstrating close agreement with FCs obtained from both DFT and traditional iterative optimization. We further extend the model to sparse inputs from single-orientation scans representing short acquisitions, where it maintains high predictive accuracy. Finally, a detailed analysis of the VAE latent space behavior reveals semantically continuous, physically aligned encodings that underpin its generalizability across both simulated and experimental domains. Once trained, the VAE-based model enables rapid inversion of experimental spectra with no additional forward simulations. Overall, this framework provides a potential path toward real-time inversion and experimental steering in neutron scattering workflows, and provides a promising route to directly extract physically meaningful information about a material's potential energy surface from temperature-dependent measurements of its collective vibrational dynamics.

## Results and discussion

### Framework for ML-driven force constant inversion from single-crystal inelastic neutron scattering spectra

The approach developed for inverting interatomic FCs from INS spectra consists of three primary stages, as outlined in Fig. 1. In Stage I, an efficient, uMLFF-based procedure is employed to generate a large training dataset of paired  $[\phi, I(\mathbf{q}, E)]$  values. A simple crystal structure file is used as input to an ensemble of uMLFFs, specifically 14 submodels from the SevenNet,<sup>43</sup> MACE,<sup>44</sup> M3GNet,<sup>45</sup> and CHGNet<sup>46</sup> families, used here to sample a broad and physically plausible region of FC space rather than to privilege any single potential as ground truth. Each uMLFF model predicts a potential energy surface, which is then used to generate interatomic FC tensors using the finite-difference method, as implemented in Phonopy.<sup>47,48</sup> Along





**Fig. 1** Overall framework for force constant inversion from single-crystal INS spectra. (I) Training FC labels are generated from an ensemble of uMLFFs, sampling the phase space of potential FCs. An efficient, physics-based forward model simulates unpolarized INS spectra from symmetry-reduced FCs. (II) For the dual-model approach in (a), a VAE learns low-dimensional latent representations of spectra; a separate regressor is trained to predict FCs from the latent encodings. In the direct inversion approach in (b), ResNet-18 is transfer-learned to predict FCs from input spectra. (III) Experimental 4D data volumes are symmetrized and passed through the frozen, pretrained encoder. Encoded latent distributions are sampled and regressed to infer experimental FC predictions and associated uncertainties.

a standardized high-symmetry  $\mathbf{q}$ -path through the Brillouin zone, based on the conventions of<sup>49,50</sup> and illustrated in SI Fig. 1 and 6, phonon dispersion calculations are performed for each uMLFF (SI Fig. 2 and 7). FC sets yielding imaginary phonon frequencies (*i.e.*, dynamically unstable configurations) are discarded. A sensitivity analysis is subsequently performed to determine how many FCs are needed to sufficiently capture variations in the phonon dispersion, as shown in SI Fig. 4 and 9. On this basis, the largest eight FCs are retained for constructing the final training datasets. This reduced parameterization captures the dominant subspace governing spectral variation along the sampled  $\mathbf{q}$ -path, while the remaining FCs are left unmodified at their uMLFF-derived reference values in the forward model. The retained FCs are then jointly sampled over an expanded yet physically bounded range (uniformly sampled from  $[-25\%, +150\%]$  of their observed uMLFF bounds) to define the training domains shown in SI Fig. 3 and 8, and the resulting

sampled FC labels are used to generate simulated INS spectra spanning a broad range of phonon behaviors.

The forward model is grounded in Born–von Kármán lattice dynamics and is designed for computational efficiency. Simulations use symmetry-reduced FCs—obtained by consolidating tensor elements invariant under the crystal's space group operations—thus significantly reducing the number of parameters. Additionally, the model omits computationally intensive steps such as phonon eigenvector evaluations, Debye–Waller (DW) factors, and polarization effects, based on the assumption that these contributions are effectively averaged through Brillouin zone folding of  $S(\mathbf{Q}, E)$  data (see SI Fig. 11). In the present work, model validation and experimental inversion were performed at low temperature, where DW attenuation and related thermal broadening effects are minimal. Under such conditions, omission of explicit DW factors introduces negligible distortion in the folded  $I(\mathbf{q}, E)$  spectra used for training and



inference. This approximation enables faster simulations of unpolarized  $I(\mathbf{q}, E)$  spectra while maintaining sufficient fidelity for ML model training. For applications at elevated temperatures, these effects can be incorporated by performing full  $S(\mathbf{Q}, E)$  simulations over the experimentally sampled  $\mathbf{Q}, E$  volume—including temperature-dependent DW factors and polarization terms—prior to folding into  $I(\mathbf{q}, E)$ . Such calculations are readily implemented within existing scattering simulation frameworks (e.g., pathSQE<sup>51</sup>), and would allow extension of the present data generation procedure to more detailed temperature-dependent effects. From these simulations using perturbed FC values, only dynamically stable spectra are retained, yielding a curated training dataset of 10 000  $[\phi, I(\mathbf{q}, E)]$  pairs. This full uMLFF-based procedure was applied to both Ge and Nb to generate the corresponding training datasets (see SI Section 1).

Herein, we investigate two different model architectures for the FC inversion task. As shown in Stage II(a) of Fig. 1, in the dual-model approach, a VAE and a separate feedforward FC regressor are trained sequentially. The VAE compresses each high-dimensional  $I(\mathbf{q}, E)$  spectrum into a low-dimensional latent representation and reconstructs it *via* a decoder. The model is trained by minimizing the evidence lower bound, a dual-objective loss that combines reconstruction error with a scaled Kullback–Leibler divergence. This formulation encourages the latent space to be smooth and well-structured, such that nearby points correspond to similar spectra and samples drawn from the prior distribution decode into physically plausible reconstructions.<sup>52</sup> In a VAE, the encoder outputs a mean and variance for each latent dimension, defining the latent activations ( $\mu, \sigma^2$ ) that characterize both the compressed representation and its associated uncertainty. The variational bottleneck is particularly well suited for this task, as the input spectra are sparse and high-dimensional ( $\sim 20\,000$  pixels), while the underlying physical degrees of freedom—symmetry-reduced FCs—are comparatively few. After training, the encoder is frozen, and the FC regressor, a multi-layer feedforward neural network, is trained to predict FC values from the latent encodings. In the direct inversion approach shown in Stage II(b) of Fig. 1, we employ ResNet-18 (ref. 53) and apply transfer learning to train the network to predict FCs directly from input spectra. Unlike the VAE-based model, which provides probabilistic latent representations, the ResNet-18 functions as a deterministic regressor without explicit uncertainty estimation. Owing to its much deeper architecture and residual connections, ResNet-18 has substantially greater representational capacity than the dual-model approach. Both the VAE-based and ResNet-based models were trained on the same simulated datasets of 10 000  $I(\mathbf{q}, E)$  spectra, as described in detail in SI Section 2. While both models were trained on the same datasets, note we reserve experimental inference for the VAE-based approach, as ResNet-18 proved more sensitive to spectral noise and underperformed in noisy or domain-shifted settings (see benchmarking results).

The final stage of the framework involves experimental INS data processing and FC inference, as illustrated in Stage III of Fig. 1. Recall the measured INS signal constitutes a four-dimensional volume of scattering intensity as a function of momentum and energy transfer. After standard event-mode

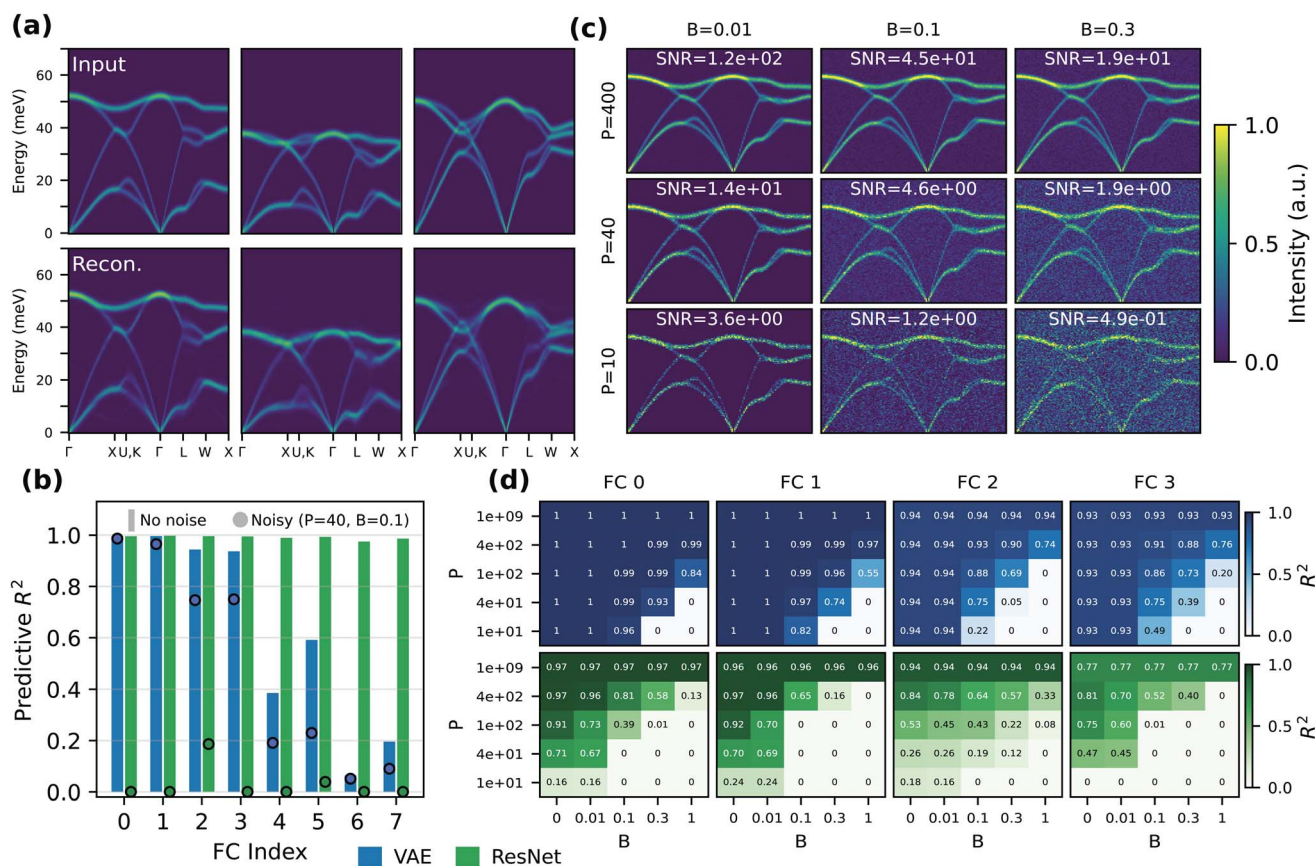
data reduction and transformation into  $S(\mathbf{Q}, E)$ ,<sup>54,55</sup> the data are processed using the pathSQE software to perform automated Brillouin zone folding.<sup>51</sup> This produces a lower-dimensional, symmetrized representation  $\sim I(\mathbf{q}, E)$  and improves statistical quality by aggregating signal from symmetrically equivalent regions of reciprocal space. Following Min–Max scaling to match the intensity range of the training data, the folded experimental spectrum is passed through the frozen, pre-trained encoder to obtain its latent distribution. Multiple latent samples are drawn and propagated through the frozen FC regressor to generate predicted FCs. This procedure yields both the mean predicted FC values and associated uncertainties based on sample variation. Notably, because the first two stages—training data generation and model training—can be completed in advance, Stage III could support fast, fully automated inference on experimental datasets. In practice, this enables near real-time application of the framework during an experiment, offering the potential for adaptive data collection strategies, on-the-fly assessment of measurement sufficiency, and future integration with experimental steering and optimization workflows.

### Benchmarking predictive performance and robustness to spectral noise on simulated data for Ge and Nb

To benchmark the predictive performance of the ML framework, we evaluated both model architectures—the VAE-based dual-model and the direct-inversion ResNet-18—on simulated test data for Ge and Nb. These two materials serve distinct purposes: Ge provides a benchmark system where the physics is well-understood and experimental validation is available and discussed herein, while Nb offers an opportunity to assess model generalization to a chemically and structurally distinct system, where longer range interactions also become important. For each material and model, we assess spectral reconstruction accuracy, FC predictions on clean (*i.e.*, no noise) test spectra, and robustness to realistic noise. Results for Ge and Nb are summarized in Fig. 2 and 3, respectively.

We begin by evaluating model performance on clean simulated spectra. To visualize reconstruction fidelity across different regions of the FC domains, test spectra with diverse FC configurations were selected. As shown in Fig. 2a and 3a, the VAE produces high-fidelity reconstructions that preserve overall dispersion structure and finer spectral details, indicating that the VAE effectively captures and compresses the complex spectral information across the FC phase space. We next evaluated the accuracy of the FC predictions generated by each of the inversion approaches. The test datasets of 1500 simulated spectra were passed through each of the frozen models to predict their corresponding symmetry-reduced FC values. SI Fig. 16 and 18 show parity plots for the all of the FC parameters across both models, along with their associated  $R^2$  and mean absolute error metrics. As summarized in the bar plots in Fig. 2b and 3b, the ResNet model, which maps directly from spectra to FCs without intermediate reconstruction, achieves very high predictive accuracy on the clean test sets for both Ge and Nb, with  $R^2$  scores near unity across all FCs. The VAE-based





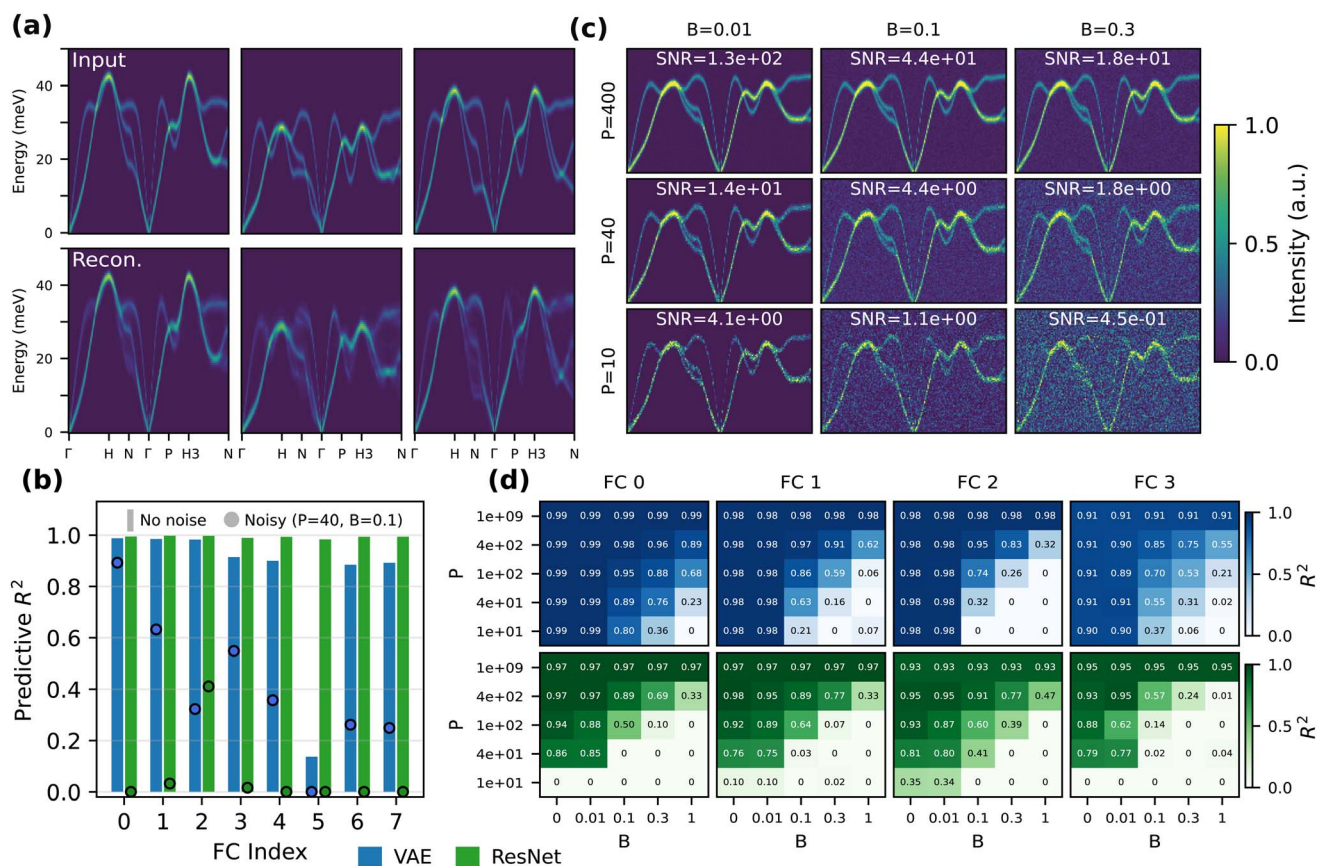
**Fig. 2** Predictive performance and spectral noise robustness on simulated test data for the case of Ge. (a) Reconstructions of simulated  $I(q, E)$  spectra demonstrate accurate preservation of spectral features. (b) Bar plots show predictive  $R^2$  scores for each FC using clean test data, with circles indicating inference performance for a representative noisy case. Blue corresponds to the VAE model and green to the ResNet baseline. (c) Example spectra perturbed with increasing constant background (B) and Poisson counting noise (P); signal-to-noise ratios (SNRs) are indicated in each input. (d) Heatmaps of  $R^2$  scores from VAE-based and ResNet-based predictions of FC indices 0–3 across all noise conditions. Values are clipped to [0, 1] for clarity, with entries  $\leq 0$  shown as 0.

approach, in contrast, exhibits lower, more selective accuracy: it reliably recovers the most influential FCs (e.g., FC0–FC3 for Ge and all except FC5 for Nb) with  $R^2 \geq 0.9$ , while its performance declines for weaker terms that are known to have a more limited impact on the spectral features. This distinction highlights a key difference in model behavior. ResNet appears to rely more heavily on fine-grained spectral features to fit all FCs equally well under ideal conditions, while the VAE tends to recover the FCs with the strongest influence on the spectra, since the compressed latent representations preferentially retain high-impact spectra variations. These results are consistent with expectations from the FC sensitivity analysis and underscore the interpretability of the VAE-based approach.

To further assess the generalizability of the models, we evaluated their robustness to realistic spectral noise augmentation applied to the simulated test datasets, for which ground-truth labels remain available. The noise model was designed to reflect the physical and statistical processes of INS, incorporating a tunable constant background component (B) and scalable Poisson counting statistics (P) (see SI Section 4). Fig. 2c and 3c show example test spectra under varied noise levels, along with the corresponding SNRs. To quantify predictive

stability, we applied 25 combinations of background and Poisson noise (spanning a  $5 \times 5$  parameter grid, shown in SI Fig. 21–24) across the test datasets and performed FC inference. For each configuration, we computed  $R^2$  scores for the predicted FC values relative to ground truth, tracking the degradation in accuracy with increasing noise. The results for FCs 0–3 are shown in Fig. 2d and 3d, while less impactful FCs 4–7 are displayed in SI Fig. 17 and 19. From this analysis, we find the VAE model maintains good performance over a wide range of signal-to-noise ratios, with graceful degradation that more heavily affects the less sensitive FCs. By contrast, ResNet exhibits a sharper decline in accuracy, with noticeable  $R^2$  drops for several FCs even under mild noise levels. This suggests that ResNet may overfit to subtle features of the clean forward model spectra, resulting in poor generalization when spectral quality deteriorates. For both materials, VAE-based predictions show greater resilience, consistent with the ability of the latent encoding to filter noise and focus on the dominant phonon features. Importantly, although the VAE model was trained exclusively on clean simulated data, these results indicate it generalizes well to realistic noisy inputs. This behavior supports the domain robustness of the model and, by leveraging known





**Fig. 3** Predictive performance and spectral noise robustness on simulated test data for the case of Nb. (a) Reconstructions of simulated  $I(\mathbf{q}, E)$  spectra demonstrate accurate preservation of spectral features. (b) Bar plots show predictive  $R^2$  scores for each FC using clean test data, with circles indicating inference performance for a representative noisy case. Blue corresponds to the VAE model and green to the ResNet baseline. (c) Example spectra perturbed with increasing constant background ( $B$ ) and Poisson counting noise ( $P$ ); SNRs are indicated in each input. (d) Heatmaps of  $R^2$  scores from VAE-based and ResNet-based predictions of FC indices 0–3 across all noise conditions. Values are clipped to  $[0, 1]$  for clarity, with entries  $\leq 0$  shown as 0.

ground-truth labels, offers quantitative validation of its potential for experimental application.

These findings underscore a key distinction: while both models achieve high accuracy under ideal conditions, the VAE-based architecture exhibits superior robustness and domain adaptability, attributes essential for real-world deployment. Its consistent performance across Ge and Nb highlights material generality and the ability to learn transferable representations of vibrational dynamics for different material systems. Given this demonstrated stability, we focus exclusively on the VAE-based model for subsequent experimental inference. Together, these benchmarking results establish a solid quantitative foundation for further applying the framework to noisy, sparse, and domain-shifted experimental data.

### Experimental force constant inference under full-dataset and sparse single-scan conditions

Having established strong reconstruction and predictive performance on both clean and noisy simulated data, we next apply the trained VAE-based models to real INS measurements in order to extract experimental FCs parameterizing the phonon

behavior of a physical sample. The experimental dataset was collected on a large ( $m = 25$  g), high-quality Ge single crystal using the ARCS spectrometer<sup>56</sup> at the Spallation Neutron Source, with measurements performed at 5 K and acquired in  $1^\circ$  rotational increments spanning a total angular range of  $120^\circ$ . We consider two experimental inference scenarios: (1) the full dataset, acquired over approximately three hours, and (2) single-orientation scans representing one-minute measurements. These two cases are illustrated in the 2D elastic  $\mathbf{Q}$ -space map shown in SI Fig 25, where the colored dashed lines indicate the curved  $S(\mathbf{Q}, E)$  shells sampled at fixed orientations. While the full dataset case reflects post-experimental data from a more typical acquisition approach, the sparse single-scan case enables us to probe the limits of the model and assess how much physically meaningful information can be extracted from severely limited measurements. All data were preprocessed using pathSQE to perform Brillouin zone folding along the same  $\mathbf{q}$  path in SI Fig. 1, and the resulting folded  $I(\mathbf{q}, E)$  spectra—shown in the left-most column of Fig. 4a–d—serve as the input to the trained VAE-based model for experimental FC inference.

We first evaluate reconstruction and FC inversion performance using the full experimental dataset, as shown in Fig. 4a.



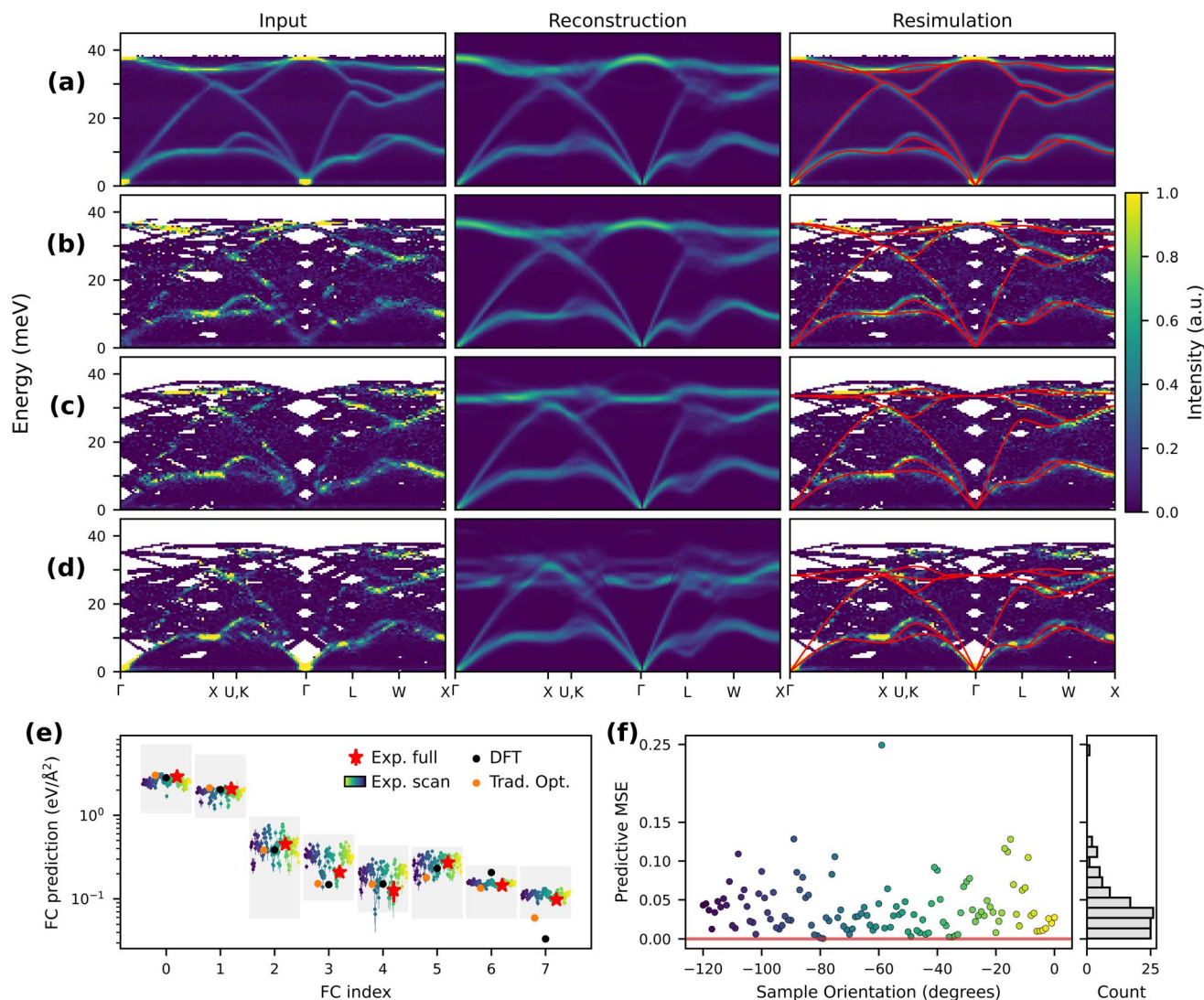


Fig. 4 Experimental force constant inference under full-dataset and sparse single-scan conditions. (a–d) Input spectra, reconstructions, and forward-simulated dispersions based on predicted FCs for (a) the full dataset and for fixed-orientation scans at (b)  $\omega = 79^\circ$ , (c)  $64^\circ$ , and (d)  $59^\circ$ , corresponding to the best, median, and worst single-scan FC inversion performance relative to the full-dataset result. For visual clarity, the spectra are cropped at 45 meV to better match the measured energy range; however, all model inputs and reconstructions are defined over the full 0–70 meV range used during training. (e) Predicted experimental FCs for all measurements, with DFT and traditionally optimized (Trad. Opt.) values shown for reference. Colored points correspond to single-scan inferences, with color indicating scan angle. The gray shaded boxes indicate the range of FC values sampled in the training dataset for each FC index and are shown for reference. (f) FC prediction mean squared error (MSE) across all single-scan measurements as a function of sample orientation.

The VAE-generated reconstruction closely replicates the overall phonon dispersion structure and mode energies present in the measured INS spectrum. Some fine spectral features—such as small energy splittings between closely spaced modes, particularly in the transverse acoustic branches between  $K$ – $\Gamma$  and  $L$ – $W$ —are partially blurred in the reconstruction, consistent with the smoothing behavior expected from a VAE. To assess the accuracy of the inferred FCs, we perform a forward phonon dispersion calculation using the ML-predicted FCs and overlay the simulated dispersion on the experimental spectrum, as shown in the right-most column of Fig. 4a. The close visual agreement across the Brillouin zone suggests that the inverted FCs successfully capture the underlying lattice dynamics measured

in the experiment. For a more quantitative comparison, we examine the predicted FC values directly against two independent references: those obtained from DFT and those derived *via* a traditional iterative optimization procedure (see SI Section 6). As shown in Fig. 4e, the ML-inferred FCs are in close agreement with both references, with minor deviations for the least influential terms. Taken together, these results demonstrate that the model provides very accurate experimental inference under full dataset conditions and establishes a high-fidelity baseline for evaluating performance in more data-limited regimes.

We now turn to inference on the more challenging and unconventional single-scan data, which represent one-minute measurements at fixed sample orientations. Because the full



dataset is constructed by aggregating many such scans in a set of step-wise sample rotations, we can treat each orientation as a separate test input, yielding approximately 120 independent opportunities to evaluate inference performance. For clarity and balance, we focus first on three representative cases: those with the best, median, and worst FC prediction accuracy relative to the full-dataset result, as shown in Fig. 4b–d. Despite the limited  $q$ - $E$  coverage and presence of dispersion discontinuities, the model effectively reconstructs signals present in the sparse data. The reconstruction in Fig. 4b (best-performing scan) exhibits striking similarity to the full-dataset reconstruction in Fig. 4a, demonstrating that the model can recover meaningful structure from minimal input. In contrast, the reconstruction in Fig. 4d (worst-performing scan) appears blurrier, particularly in the optical region where the input lacks usable signal, reflecting the model's uncertainty in the absence of informative features. When evaluating FC inversion, the forward-simulated dispersions based on predicted FCs align well with the experimental data across all three cases. For both the best and median orientations, the calculated dispersions show close agreement with the observed spectrum, with only minor discrepancies in regions of incomplete coverage. Even in the worst case, the model still captures key features in the acoustic region, despite the lack of intensity in the optical branches, which limits inversion accuracy. These results demonstrate that the combined VAE-based approach retains strong inference capabilities even under highly sparse measurement conditions.

Finally, we evaluate the FC predictions across all 120 individual single-scan measurements. A summary of these results is shown in Fig. 4e, where the predicted FCs are plotted as scatter points and colored by sample orientation. For the most influential parameters, FC0 and FC1, the predictions are well-constrained within a narrow subregion of the training domain and generally align with those obtained from the full dataset, DFT, and traditional iterative optimization methods. FCs 2–4, which are less dominant but still influential, show greater variation across orientations—suggesting that additional spectral detail may be required for their accurate recovery, which is not always present in sparse single-scan inputs. In contrast, predictions for the least influential terms, FC6 and FC7, remain narrowly distributed but do not reflect improved accuracy. Instead, this behavior mirrors what was observed during testing on clean simulated data, where the FC regressor failed to capture meaningful trends in the latent encodings for these less impactful parameters, yielding near-constant outputs and  $R^2$  scores near zero (see SI Fig. 16). To quantify prediction quality across orientations, we compute the MSE of each scan's predicted FC vector relative to the full-dataset prediction and plot the results as a function of scan angle in Fig. 4f. This analysis shows that the model produces near-perfect predictions for about 10 scan orientations. It also reveals that the worst case (Fig. 4d) is a substantial outlier, exhibiting nearly double the MSE of the next-worst prediction. Interestingly, the MSE distribution shows some angular clustering, suggesting that certain sample orientations consistently capture more informative phonon features and highlighting the directional

dependence of information content in INS experiments. Together, these results underscore the model's high accuracy under full-dataset conditions and its good performance even from a single one-minute scan. These findings point toward the potential for real-time experimental integration, where on-the-fly FC inversion could assist data acquisition or help prioritize measurements toward more informative orientations. In practice, preliminary spectra could be rapidly analyzed to compare inversion consistency across orientations, suggesting a pathway toward adaptive measurement workflows. This type of approach would allow experiments to dynamically allocate measurement time to the most informative regions of reciprocal space, thereby improving overall efficiency and the scientific return of limited beamtimes. Realizing a fully closed-loop optimization scheme would require additional strategies for quantifying orientation-dependent information content, which we leave for future work.

### Structure, interpretability, and generalization of the learned latent space

We conclude this study with a detailed investigation of the learned latent space to better understand the mechanisms underlying the model's strong FC inversion performance, particularly focusing on its interpretability and generalization across data domains. To this end, we analyzed the latent embeddings produced by the frozen, pre-trained encoder for both simulated and experimental input spectra. For each of the 1500 simulated test spectra, the encoder outputs a 20-dimensional latent distribution parameterized by a mean vector ( $\mu$ ) and log-scale variance vector ( $\log(\sigma^2)$ ), reflecting the variational nature of the model. To evaluate how the encoder utilizes each latent dimension, we visualized the aggregated  $\mu$  and  $\log(\sigma^2)$  values across all test spectra in Fig. 5a and b, respectively. In each plot, the  $x$ -axis denotes the latent dimension index, and the  $y$ -axis shows the values for all 1500 spectra. The black points and error bars represent the mean and standard deviation of the values within each dimension. As shown in Fig. 5a, approximately six dimensions exhibit tightly clustered  $\mu$  values near zero, while the corresponding  $\log(\sigma^2)$  values in Fig. 5b are also near zero. This indicates that these dimensions remain close to the standard normal prior and are effectively unused by the model. In contrast, the remaining 14 dimensions show well-spread  $\mu$  values ranging roughly from  $-3$  to  $3$ , suggesting active utilization for encoding meaningful spectral variation. The associated  $\log(\sigma^2)$  values for these active dimensions are generally negative, implying high confidence in the latent representations. Crucially, the  $\log(\sigma^2)$  values are not overly small, confirming that the encoder retains stochasticity and avoids posterior collapse into a deterministic autoencoder. Given the excellent FC inference performance achieved with this latent structure, these results indicate that the model effectively compresses the information from the  $\sim 20\,000$ -dimensional INS spectra into just 14 active and well-behaved latent dimensions.

Although the dimensionality of the spectra has already been significantly reduced through latent encoding, applying



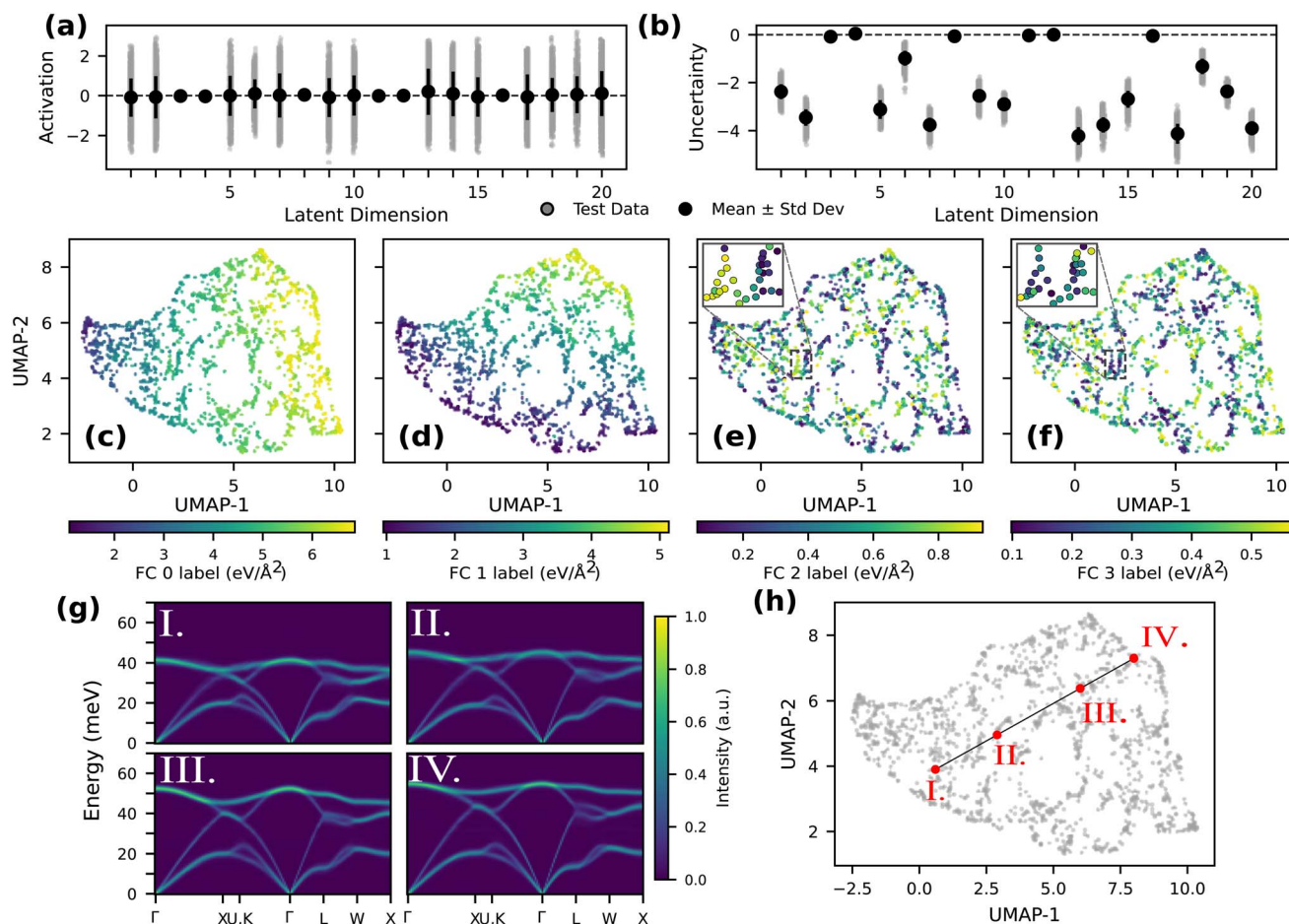
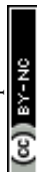


Fig. 5 Structure and interpretability of the learned latent space. (a and b) Simulated test spectra latent (a) activations  $\mu$  and (b) uncertainties  $\log(\sigma^2)$ . The black scatter points with error bars and dashed lines correspond to the mean  $\pm$  standard deviation and the initialized standard normal prior distributions, respectively. (c–f) Two-dimensional UMAP projections of the latent encodings colored by individual FC values (FC0–FC3) show a semantically and physically organized latent space. (g and h) Reconstructed spectra (g) from points sampled along a latent-space ray (h) further illustrate the physical alignment of the latent space and highlight semantic consistency in the decoder.

additional nonlinear dimensionality reduction can yield further insight into the structure and organization of the learned latent space. To this end, we apply Uniform Manifold Approximation and Projection (UMAP),<sup>57</sup> a widely used unsupervised ML method known for preserving both local and global structure in high-dimensional data. Specifically, we extract the encoded latent  $\mu$  from the simulated test set (as shown in Fig. 5a) and pass them to UMAP to produce two-dimensional embeddings that retain the meaningful relationships from the original latent space. The resulting 2D representations are plotted in Fig. 5c–f, with each point corresponding to a single spectrum and colored by one of the symmetry-reduced FC value labels. This analysis reveals a key result regarding the interpretability of the latent space: not only is it semantically organized, as encouraged by the variational loss, but its structure also reflects the underlying physical interactions encoded by the FCs. Most notably, distinct global gradients are observed in the UMAP embeddings for the dominant FCs (FC0 and FC1), with clear directional trends that span the entire latent manifold. Even for less influential FCs (FC2 and FC3), meaningful clustering and local variation persist as shown in the insets, despite their

reduced sensitivity and the further compression to only two UMAP dimensions. Similar behavior is also observed in the Nb case, as shown in SI Fig. 20, further highlighting the generality across material systems. To further probe the semantic consistency of the latent space, we linearly interpolated along a latent ray in the UMAP-projected space, shown in Fig. 5h, and decoded the resulting sampled latent vectors. As shown in Fig. 5g, the corresponding reconstructions vary smoothly and capture physically plausible spectral transitions, suggesting that the latent space supports a coherent and continuous spectral manifold. Together, these results demonstrate that the latent space not only encodes semantically relevant structure, but does so in a manner that is coherent, interpretable, and physically aligned.

Finally, as shown in Fig. 6, we examine the domain adaptability of the model—an essential factor underpinning its generalization across data sources. In this work, inference was performed across four distinct domains: clean simulated spectra, noise-augmented simulated spectra, full experimental spectra, and sparse single-scan spectra. To investigate how the encoder treats these varied inputs, we passed the experimental



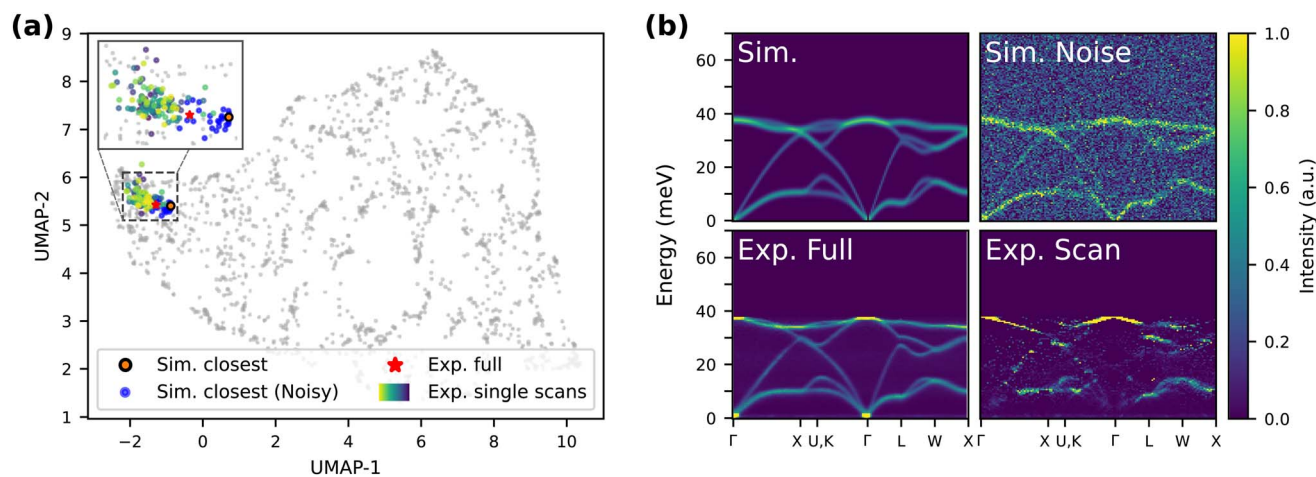


Fig. 6 Cross-domain generalization of the learned latent space. (a) Latent embeddings for clean and noisy simulations, full experimental data, and single-orientation experimental scans cluster near one another, indicating robustness and cross-domain generalization. (b) Input spectra across data domains corresponding to the most closely embedded points in (a), confirming consistent domain adaptation between simulated and experimental data.

spectra through the frozen encoder to obtain latent embeddings, which were then projected into two dimensions using the pre-fitted UMAP model. For reference and context, we also include the latent embedding of the simulated test spectrum most similar (in MSE) to the full experimental input, along with its noise-augmented variants across the full range of noise settings. The resulting projections are shown in Fig. 6a, where grey points denote the full simulated test set (as in Fig. 5h). Notably, all four data domains—despite differences in noise level, spectral smoothness, and  $Q$ - $E$  coverage—are mapped to a compact, shared region of the latent space. This tight clustering suggests that the model effectively identifies the underlying dynamical structure common to these spectra and encodes them consistently, facilitating robust FC inversion across domains. To further illustrate this behavior, Fig. 6b displays input spectra from the four representative cases: the full experimental spectrum, its most similar clean simulation, the most closely embedded noise-augmented spectrum, and the most similar single-scan measurement. Despite substantial visual differences between the spectra, their proximity in the latent space highlights the ability of the model to selectively extract physically meaningful features while ignoring irrelevant variations. This domain-consistent encoding is a key contributor to the model's strong real-world performance and suggests that variational latent spaces can provide a powerful foundation for generalizable, physics-informed spectral inference.

## Conclusions and outlook

The present study evaluates the performance of two ML-driven approaches for extracting interatomic FCs from single-crystal INS spectra. Through a systematic comparison of a VAE-based architecture and a transfer-learned ResNet model, we find that the VAE offers superior robustness to spectral noise, missing data, and domain shift, while simultaneously yielding latent representations that encode physically interpretable

relationships. In contrast, the ResNet model, when trained with target scaling, exhibits nearly perfect raw regression performance within the broad training domain but reduced generalizability under realistic experimental perturbations. These complementary behaviors underscore a central trade-off between precision and adaptability in data-driven inversion schemes, particularly in domains such as INS for which experimental datasets are resource-limited and models must typically be trained on synthetic data. Both models were trained exclusively on clean simulated spectra in order to isolate architectural effects. The observed robustness of the VAE under realistic noise perturbations therefore suggests that architectural bottlenecks and variational regularization contribute meaningfully to improved generalization. At the same time, augmentation of the training domain to include more realistic experimental effects—such as instrumental artifacts, noise, or limited ( $Q$ ,  $E$ ) coverage—may further enhance inversion accuracy, and explicit noise augmentation during training may improve the robustness of direct regression models such as ResNet. More broadly, alternative generative and sequence-based architectures, including diffusion models and transformer-based encoders, represent promising directions for extending ML-driven inversion frameworks beyond the present comparison. Collectively, these results demonstrate a practical basis for robust, interpretable, and computationally efficient inversion of physical models from experimental observables.

Building on this foundation, we next consider the assumptions and scope of the present framework, as well as avenues for its generalization and adaptive extension to more complex material systems. The framework proceeds under the assumption that the crystal structure is known, constraining the inversion task to the determination of interatomic FCs and excluding the problem of structural identification. This assumption reflects the experimental context of single-crystal INS, where crystallographic descriptors are typically determined independently prior to measurement, and enables the



trained regressors to focus solely on mapping spectral features to interatomic interactions. However, highly automated methods exist to determine or refine a crystal structure based on reciprocal space mapping and could be incorporated as a preliminary step. With these structural assumptions in place, we next consider the simulation strategy used to generate training data. An ensemble of uMLFFs is employed to generate data without direct reliance on DFT, thereby reducing computational cost and avoiding dependence on a single DFT functional, which may exhibit known limitations for certain material classes. The present demonstrations on Ge and Nb-systems for which DFT calculations remain tractable—serve as controlled benchmarks to validate model performance and to verify that synthetic data derived from ML surrogate (*i.e.*, non-first-principles) models can yield regressors that accurately capture the relevant physics observed in an experiment. Although the approach is conceptually general, full transferability across atomic species and structural motifs remains non-trivial. In particular, the efficiency of Brillouin zone folding and FC symmetry reduction as dimensionality-lowering strategies depends on crystal symmetry, and lower-symmetry systems may offer reduced redundancy, increasing the effective dimensionality of the inversion problem. Phonon spectra encode system-specific fingerprints that depend nonlinearly on bonding geometry, atomic masses, and crystal symmetry, making universal mapping inherently difficult. Furthermore, extending the ML architectures to incorporate chemical descriptors or symmetry-aware representations, and employing strategies such as active learning or uncertainty-aware retraining, represent promising routes toward improved transferability and adaptive domain coverage. Such mechanisms would allow the model to accommodate experimental spectra that lie outside the nominal training manifold while progressively reducing the need for extensive pre-training.

Beyond improvements to the data generation and ML architectures, physical considerations also point toward several extensions beyond the idealized harmonic case. Although disordered systems pose challenges for conventional Born–von Kármán approaches, virtual-crystal approximations could provide a pathway for mapping effective FCs. The direct inversion of measured intensities ( $I(\mathbf{q}, E)$  or  $S(\mathbf{Q}, E)$ ) avoids the need for explicit dispersion fitting and extraction, which is particularly advantageous for such materials where band-tracking is intractable. Employing temperature-dependent effective harmonic approaches (TDEP) and incorporating resolution effects, as well as developing deconvolution schemes to recover higher-order anharmonic interactions from spectral linewidths, represent promising next steps toward a more complete experimental inversion toolkit. These considerations delineate the present scope and identify clear directions for extending ML-based inversion approaches to increasingly complex material systems. The framework ultimately points toward real-time, uMLFF-driven characterization of phonon behavior, wherein learned models adaptively refine themselves from experimental data to reveal microscopic lattice dynamics with minimal prior input and human intervention.

## Author contributions

AS implemented the machine learning software, performed the study of its behavior and performance, and wrote the manuscript. BL implemented the initial ResNet-18 model for the study of phonons in niobium, with help from KB and JM. OD supervised the research project and edited the manuscript.

## Conflicts of interest

There are no conflicts to declare.

## Data availability

All numerical, experimental, and computed data supporting the findings of this study are openly available through the Duke Research Data Repository at <https://doi.org/10.7924/r4tm7gw38>. The codes developed and used in this work are included in the same Duke RDR and also available on GitHub at <https://github.com/delaire-lab-duke>.

Supplementary information (SI): additional figures and text descriptions of training data generation, machine learning model training, inference using trained models, spectral noise implementation, and iterative force constant optimization. See DOI: <https://doi.org/10.1039/d6dd00008h>.

## Acknowledgements

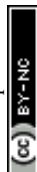
AS was partially supported by the National Science Foundation Graduate Research Fellowship under grant DGE-2139754 and by the National Science Foundation under grant DGE-2022040. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. Neutron scattering work by AS and OD was supported by the U.S. Department of Energy, Office of Science, Basic Energy Sciences, Materials Sciences and Engineering Division, under Award No. DE-SC0019978. BAL acknowledges support from the Alfred P. Sloan foundation through an Energy Data Analytics Fellowship at Duke University. This research used resources at the Spallation Neutron Source, a DOE Office of Science User Facilities operated by the Oak Ridge National Laboratory. The authors thank Francesco Luzi for help with initial training of the ResNet-18 model for Nb on GPU. We thank Douglas Abernathy, Matthew Stone, Jennifer Niedziela, and Dipanshu Bansal for help with collecting INS data on Ge using ARCS (IPTS-13861). We thank David Carlson, Guannan Zhang, Volker Blum, and Greg Herschlag for helpful discussions.

## References

- 1 D. C. Wallace and H. Callen, *Am. J. Phys.*, 1972, **40**, 1718–1719.
- 2 G. Grimvall, *Thermophysical Properties of Materials*, Elsevier, 1999.



- 3 D. C. Wallace, *Statistical Physics Of Crystals And Liquids: A Guide To Highly Accurate Equations Of State*, World Scientific, 2002.
- 4 G. A. Slack, *J. Phys. Chem. Solids*, 1973, **34**, 321–335.
- 5 D. G. Cahill and R. O. Pohl, *Annu. Rev. Phys. Chem.*, 1988, **39**, 93–121.
- 6 J. Bardeen, L. N. Cooper and J. R. Schrieffer, *Phys. Rev.*, 1957, **108**, 1175–1204.
- 7 P. B. Allen and R. Dynes, *Phys. Rev. B*, 1975, **12**, 905.
- 8 R. A. Cowley, *Phys. Rev. B*, 1976, **13**, 4877–4885.
- 9 M. T. Dove, *Am. Mineral.*, 1997, **82**, 213–244.
- 10 C. Mao, X. He, H.-M. Lin, M. K. Gupta, P. Postec, T. Lanigan-Atkins, M. Krogstad, D. M. Pajeroski, T. Hong, T. J. Williams, J. R. Stewart, D. Y. Chung, M. G. Kanatzidis, S. Rosenkranz, R. Osborn and O. Delaire, *Phys. Rev. Mater.*, 2025, **9**, 065401.
- 11 J. Ding, M. K. Gupta, C. Rosenbach, H.-M. Lin, N. C. Osti, D. L. Abernathy, W. G. Zeier and O. Delaire, *Nat. Phys.*, 2025, **21**, 118–125.
- 12 X. He, M. K. Gupta, D. L. Abernathy, G. E. Granroth, F. Ye, B. L. Winn, L. Boatner and O. Delaire, *Proc. Natl. Acad. Sci. U. S. A.*, 2025, **122**, e2419159122.
- 13 M. Born and T. von Kármán, *Phys. Z.*, 1912, **13**, 297–309.
- 14 M. Born and K. Huang, *Dynamical theory of crystal lattices*, Oxford university press, 1996.
- 15 B. t. Brockhouse and P. Iyengar, *Phys. Rev.*, 1958, **111**, 747.
- 16 B. Brockhouse, T. Arase, G. Caglioti, K. Rao and A. Woods, *Phys. Rev.*, 1962, **128**, 1099.
- 17 A. D. Zdetsis and C. S. Wang, *Phys. Rev. B*, 1979, **19**, 2999–3003.
- 18 M. B. Stone, J. L. Niedziela, D. L. Abernathy, L. DeBeer-Schmitt, G. Ehlers, O. Garlea, G. E. Granroth, M. Graves-Brook, A. I. Kolesnikov, A. Podlesnyak, *et al.*, *Rev. Sci. Instrum.*, 2014, **85**, 045113.
- 19 R. Bewley, J. Taylor and S. Bennington, *Nucl. Instrum. Methods Phys. Res., Sect. A*, 2011, **637**, 128–134.
- 20 R. Bewley, R. Eccleston, K. McEwen, S. Hayden, M. Dove, S. Bennington, J. Treadgold and R. Coleman, *Phys. B*, 2006, **385–386**, 1029–1031.
- 21 R. Kajimoto, T. Yokoo, M. Nakamura, Y. Kawakita, M. Matsuura, H. Endo, H. Seto, S. Itoh, K. Nakajima and S. Ohira-Kawamura, *Phys. B*, 2019, **562**, 148–154.
- 22 G. L. Squires, *Introduction to the theory of thermal neutron scattering*, Cambridge University Press, New York, NY, 1978.
- 23 O. Hellman, P. Steneteg, I. A. Abrikosov and S. I. Simak, *Phys. Rev. B:Condens. Matter Mater. Phys.*, 2013, **87**, 104111.
- 24 L. Monacelli, R. Bianco, M. Cherubini, M. Calandra, I. Errea and F. Mauri, *J. Phys.: Condens. Matter*, 2021, **33**, 363001.
- 25 T. Tadano, Y. Gohda and S. Tsuneyuki, *J. Phys.: Condens. Matter*, 2014, **26**, 225402.
- 26 T. Mueller, A. Hernandez and C. Wang, *J. Chem. Phys.*, 2020, **152**, 050902.
- 27 R. Jacobs, D. Morgan, S. Attarian, J. Meng, C. Shen, Z. Wu, C. Y. Xie, J. H. Yang, N. Artrith, B. Blaiszik, G. Ceder, K. Choudhary, G. Csanyi, E. D. Cubuk, B. Deng, R. Drautz, X. Fu, J. Godwin, V. Honavar, O. Isayev, A. Johansson, B. Kozinsky, S. Martiniani, S. P. Ong, I. Poltavsky, K. Schmidt, S. Takamoto, A. P. Thompson, J. Westermayr and B. M. Wood, *Curr. Opin. Solid State Mater. Sci.*, 2025, **35**, 101214.
- 28 H. Sakamoto, S. Katakami, K. Muto, K. Nagata, T.-h. Arima and M. Okada, *J. Phys. Soc. Jpn.*, 2020, **89**, 124002.
- 29 S. Katakami, H. Sakamoto, K. Nagata, T.-h. Arima and M. Okada, *Phys. Rev. E*, 2022, **105**, 065301.
- 30 F. Bao, R. Archibald, J. Niedziela, D. Bansal and O. Delaire, *Nanotechnology*, 2016, **27**, 484002.
- 31 Z. Chen, N. Andrejevic, N. C. Drucker, T. Nguyen, R. P. Xian, T. Smidt, Y. Wang, R. Ernstorfer, D. A. Tennant, M. Chan and M. Li, *Chem. Phys. Rev.*, 2021, **2**, 031301.
- 32 N. C. Drucker, T. Liu, Z. Chen, R. Okabe, A. Chotrattanapituk, T. Nguyen, Y. Wang and M. Li, *Synchrotron Radiat. News*, 2022, **35**, 16–20.
- 33 A. S. Anker, K. T. Butler, R. Selvan and K. M. Å. Jensen, *Chem. Sci.*, 2023, **14**, 14003–14019.
- 34 B. Han, R. Okabe, A. Chotrattanapituk, M. Cheng, M. Li and Y. Cheng, *Digital Discovery*, 2025, **4**, 584–624.
- 35 Y. Cheng, G. Wu, D. M. Pajeroski, M. B. Stone, A. T. Savici, M. Li and A. J. Ramirez-Cuesta, *Mach. Learn.*, 2023, **4**, 015010.
- 36 Z. Chen, N. Andrejevic, T. Smidt, Z. Ding, Q. Xu, Y. Chi, Q. T. Nguyen, A. Alatas, J. Kong and M. Li, *Adv. Sci.*, 2021, **8**, 2004214.
- 37 R. Okabe, A. Chotrattanapituk, A. Boonkird, N. Andrejevic, X. Fu, T. S. Jaakkola, Q. Song, T. Nguyen, N. Drucker, S. Mu, Y. Wang, B. Liao, Y. Cheng and M. Li, *Nat. Comput. Sci.*, 2024, **4**, 522–531.
- 38 B. Han, A. T. Savici, M. Li and Y. Cheng, *Comput. Phys. Commun.*, 2024, **304**, 109288.
- 39 M.-C. Chang, C.-H. Tung, S.-Y. Chang, J. M. Carrillo, Y. Wang, B. G. Sumpter, G.-R. Huang, C. Do and W.-R. Chen, *Commun. Phys.*, 2022, **5**, 46.
- 40 A. M. Samarakoon, K. Barros, Y. W. Li, M. Eisenbach, Q. Zhang, F. Ye, V. Sharma, Z. L. Dun, H. Zhou, S. A. Grigera, C. D. Batista and D. A. Tennant, *Nat. Commun.*, 2020, **11**, 892.
- 41 A. Samarakoon, D. A. Tennant, F. Ye, Q. Zhang and S. A. Grigera, *Commun. Mater.*, 2022, **3**, 84.
- 42 Y. Su and C. Li, *Mach. Learn.*, 2024, **5**, 035080.
- 43 Y. Park, J. Kim, S. Hwang and S. Han, *J. Chem. Theory Comput.*, 2024, **20**, 4857–4868.
- 44 I. Batatia, D. P. Kovács, G. N. C. Simm, C. Ortner and G. Csányi, *Proceedings of the 36th International Conference on Neural Information Processing Systems*, Red Hook, NY, USA, 2022.
- 45 C. Chen and S. P. Ong, *Nat. Comput. Sci.*, 2022, **2**, 718–728.
- 46 B. Deng, P. Zhong, K. Jun, J. Riebesell, K. Han, C. J. Bartel and G. Ceder, *Nat. Mach. Intell.*, 2023, **5**, 1031–1041.
- 47 A. Togo, *J. Phys. Soc. Jpn.*, 2023, **92**, 012001.
- 48 A. Togo, L. Chaput, T. Tadano and I. Tanaka, *J. Phys.: Condens. Matter*, 2023, **35**, 353001.
- 49 Y. Hinuma, G. Pizzi, Y. Kumagai, F. Oba and I. Tanaka, *Comput. Mater. Sci.*, 2017, **128**, 140–184.
- 50 A. Togo, K. Shinohara and I. Tanaka, *Sci. Technol. Adv. Mater.:Methods*, 2024, **4**, 2384822.



- 51 A. Sable, A. T. Savici, B. Linjawi and O. Delaire, *J. Appl. Crystallogr.*, 2026, **59**, 248–262.
- 52 D. P. Kingma and M. Welling, *arXiv*, 2013, preprint, arXiv: 1312.6114, DOI: [10.48550/arXiv.1312.6114](https://doi.org/10.48550/arXiv.1312.6114).
- 53 K. He, X. Zhang, S. Ren and J. Sun, *Proceedings Of The IEEE Conference On Computer Vision And Pattern Recognition*, 2016, pp. 770–778.
- 54 O. Arnold, J. Bilheux, J. Borreguero, A. Buts, S. Campbell, L. Chapon, M. Doucet, N. Draper, R. Ferraz Leal, M. Gigg, V. Lynch, A. Markvardsen, D. Mikkelsen, R. Mikkelsen, R. Miller, K. Palmen, P. Parker, G. Passos, T. Perring, P. Peterson, S. Ren, M. Reuter, A. Savici, J. Taylor, R. Taylor, R. Tolchenov, W. Zhou and J. Zikovsky, *Nucl. Instrum. Methods Phys. Res., Sect. A*, 2014, **764**, 156–166.
- 55 A. T. Savici, M. A. Gigg, O. Arnold, R. Tolchenov, R. E. Whitfield, S. E. Hahn, W. Zhou and I. A. Zaliznyak, *J. Appl. Crystallogr.*, 2022, **55**, 1514–1527.
- 56 D. L. Abernathy, M. B. Stone, M. J. Loguillo, M. S. Lucas, O. Delaire, X. Tang, J. Y. Y. Lin and B. Fultz, *Rev. Sci. Instrum.*, 2012, **83**, 015114.
- 57 L. McInnes, J. Healy, N. Saul and L. GroÅyberger, *J. Open Source Softw.*, 2018, **3**, 861.

