

Observation geometry for uncertainty-aware Hamiltonian inference and experimental design in quantum magnets

Roy Liu,^{1,2} Venugopal Ranganathan,^{1,3} David Dahlbom,⁴ Shizhou Xu,^{5,6} Tianyu Zhang,^{1,7} Yuan Ni,^{6,8} Daniel M. Pajerowski,⁴ Garrett Granroth,⁴ Thomas Strohmmer,⁵ Matthew B. Stone,⁴ Andrew F. May,^{4,9} Mark D. Lumsden,⁴ Joshua J. Turner,^{6,8} Yongqiang Cheng,^{4,*} and Zhantao Chen^{1,3,7,†}

¹*Group for AI in Materials Modeling and Analysis,
The University of Texas at Austin, Austin, TX, USA.*

²*Department of Computer Science, The University of Texas at Austin, Austin, TX, USA.*

³*Walker Department of Mechanical Engineering, The University of Texas at Austin, Austin, TX, USA.*

⁴*Oak Ridge National Laboratory, Oak Ridge, TN, USA*

⁵*Department of Mathematics, University of California Davis, Davis, CA, USA.*

⁶*Stanford Institute for Materials and Energy Sciences, Stanford University, Stanford, CA 94305, USA*

⁷*Texas Materials Institute, The University of Texas at Austin, Austin, TX, USA.*

⁸*Linac Coherent Light Source, SLAC National Accelerator Laboratory, Menlo Park, CA, USA*

⁹*Materials Science and Technology Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA*

Determining microscopic interactions from spectroscopic and scattering measurements is central to understanding quantum materials, yet it often remains unclear which interactions can be reliably revealed by the available experimental data and how additional experimental modalities should be designed to resolve the remaining ambiguities. Here we present an artificial intelligence-enabled framework for uncertainty-aware Hamiltonian inference and adaptive experimental design. By combining Hamiltonian-conditioned neural surrogates with Bayesian inference and observation geometry, the framework characterizes how measurements constrain Hamiltonian parameter space, quantifies the identifiability of microscopic interactions, and propagates posterior uncertainty directly in the physical Hamiltonian parameter space rather than an abstract learned representation. Using multimodal powder and single-crystal inelastic neutron scattering measurements of the quantum magnet NiPS₃, we demonstrate physically interpretable Hamiltonian inference, modality-aware uncertainty quantification, and adaptive experimental design. The framework provides a general strategy for uncertainty-aware microscopic characterization and multimodal experimental design across quantum materials.

INTRODUCTION

Microscopic models of elementary excitations provide a fundamental link between experimental observations and the underlying physics of condensed matter and materials. For quantum magnets, this connection can be modeled through the spin Hamiltonian, whose parameters encode exchange couplings, anisotropies, and other microscopic interactions governing the excitation spectrum¹⁻⁴. By fitting model parameters to measured spectra, one can infer these parameters, thereby determining the associated collective excitations as well as the static and dynamical properties of materials. In practice, however, this inverse problem is often ill-posed: distinct Hamiltonians can produce nearly indistinguishable spectra within experimental resolution, making it difficult to determine whether subtle microscopic features are genuinely constrained by the available measurements⁵⁻¹¹. Consequently, conventional refinement procedures that report only a single best-fit Hamiltonian may obscure substantial parameter uncertainty, correlations, and alternative microscopic descriptions that remain consistent with the data^{12,13}. The central challenge is therefore not

only to identify a Hamiltonian that reproduces the observed spectra, but also to determine which interactions are uniquely identifiable, which remain ambiguous, and how much microscopic information a given measurement can truly reveal.

Inelastic neutron scattering (INS) is a particularly powerful probe of magnetic and lattice excitations because it directly and quantitatively measures the momentum- and energy-resolved dynamical structure factor, $S(\mathbf{Q}, \omega)$, for magnons and phonons with high resolution^{2-4,14-18}. Interpreting INS data typically relies on physically rigorous forward methods, including linear spin wave theory, exact diagonalization, density-matrix renormalization group calculations, and, more recently, neural quantum states, to generate theoretical spectra for comparison with experiment¹⁹⁻²⁵. Although their computational scaling differs substantially, repeatedly evaluating these forward models across a high-dimensional Hamiltonian parameter space can remain prohibitively expensive. Exhaustive grid-based searches scale exponentially with the dimensionality of the parameter space, whereas coarse parameter sweeps can overlook narrow regions containing plausible solutions. Consequently, Hamiltonian refinement often relies on selected spectral cuts, expert-guided parameter tuning, or simplified Hamiltonian parameterizations, making it difficult to systematically quantify parameter uncertainty and de-

* chengy@ornl.gov

† zhantao@austin.utexas.edu

termine which microscopic interactions are uniquely constrained by the experimental data.

Additional challenges arise from the increasingly multimodal nature of experimental materials characterization^{26–30}. In INS, both powder and single-crystal measurements are widely employed. Single-crystal INS preserves directional momentum information and can resolve fine features of the dynamical structure factor, $S(\mathbf{Q}, \omega)^{2–4}$. However, each sample orientation probes only a limited region of reciprocal space, requiring multiple measurements for comprehensive coverage, and sufficiently large, high-quality single crystals are often difficult to grow. By contrast, powder samples are generally easier to prepare in the quantities required for neutron-scattering experiments. Powder INS measures the orientationally averaged response, $S(|\mathbf{Q}|, \omega)$, sacrificing directional information in exchange for greater experimental accessibility and broader sampling of reciprocal space^{31–33}. These complementary measurement modalities therefore impose distinct observational constraints on the same underlying Hamiltonian. Moreover, multimodal experiments often differ in their coordinate representations, spatial and energy resolutions, selection rules, and noise characteristics, making it challenging to integrate heterogeneous datasets in a manner that is both mathematically rigorous and physically interpretable.

Artificial intelligence (AI) and machine learning (ML) have been playing increasingly important roles in bridging experimental observations and theories and computations^{24,34–38}. Recent advances in simulation-based inference, neural posterior estimation, and multimodal machine learning^{39–42} have provided powerful new tools for accelerating inverse problems and integrating heterogeneous datasets. Many of these approaches, however, rely on learned latent representations or gridded data parameterizations to relate observations across different modalities. In practice, discrepancies between the simulated data used for training and experimental measurements, such as instrumental effects, sample imperfections, and other phenomena not captured by the forward model, can degrade the robustness of learned latent representations and complicate their physical interpretation, particularly in high-dimensional settings. For the Hamiltonian-inference problem considered here, however, a natural shared representation is provided by the parameter space of the microscopic Hamiltonian itself. An effective framework for multimodal spectroscopy should therefore connect complementary measurements through this physically meaningful parameter space, propagate uncertainty consistently across measurement modalities, and guide experimental design by identifying the measurements that most effectively resolve the remaining ambiguities.

In this work, we develop a framework that asks not only which Hamiltonian best reproduces an INS spectrum, but also what microscopic information the measurements actually constrain. Our central idea is to treat each spectroscopic modality as an observation mapping

from the Hamiltonian parameter space to the experimentally accessible response function. This mapping induces an observation geometry on the Hamiltonian parameter space, where stiff directions correspond to parameter combinations that strongly influence the measured spectra, while sloppy directions produce nearly indistinguishable responses and are thus only weakly constrained^{5–7}.

To make this observation geometry computationally tractable, we construct Hamiltonian-conditioned, differentiable AI surrogate models that accurately emulate multimodal neutron-scattering spectra, leveraging recent advances in implicit neural representations (INRs) for spectroscopy^{43–45}. Combined with Bayesian inference, these surrogates enable rapid evaluation of spectral-discrepancy landscapes, quantification of modality-dependent parameter-wise identifiability, and propagation of posterior uncertainty directly within the physical Hamiltonian parameter space. Rather than fusing powder and single-crystal INS through an abstract learned embedding, our framework connects them through the shared microscopic Hamiltonian that generates both measurements, thereby keeping inference within this physically meaningful parameter space. As a result, posterior information inferred from one modality can naturally inform inference in another, enabling sequential Bayesian inference, joint parameter estimation, and uncertainty-aware experimental design^{46–50}.

We demonstrate this framework using simulated and experimental INS measurements of the quantum magnet NiPS₃. The observation geometry reveals modality-dependent stiff and sloppy Hamiltonian directions, powder-informed posterior initialization reshapes the subsequent single-crystal acquisition strategy, and joint inference on experimental powder and single-crystal spectra yields a posterior distribution of Hamiltonians that reproduce both modalities with high spectral agreement. This work establishes a general framework that uses AI surrogates not only to accelerate Hamiltonian inference, but also to efficiently uncover which microscopic interactions are resolved by each experimental modality, quantify the remaining ambiguities, and guide adaptive experimental design toward measurements with higher expected information gain.

RESULTS

Multimodal AI surrogate

In an INS experiment, illustrated schematically in Fig. 1(a), an incident neutron beam scatters from the sample, and the scattered neutrons are recorded by an array of detectors as a function of momentum and energy transfer. For a single-crystal sample, the crystal orientation is controlled by a goniometer with rotation angle ψ . For a fixed sample orientation ψ , the detector array samples the measured intensity $I(\mathbf{Q}, \omega) \propto S(\mathbf{Q}, \omega)$ over a region of the four-dimensional scattering space determined

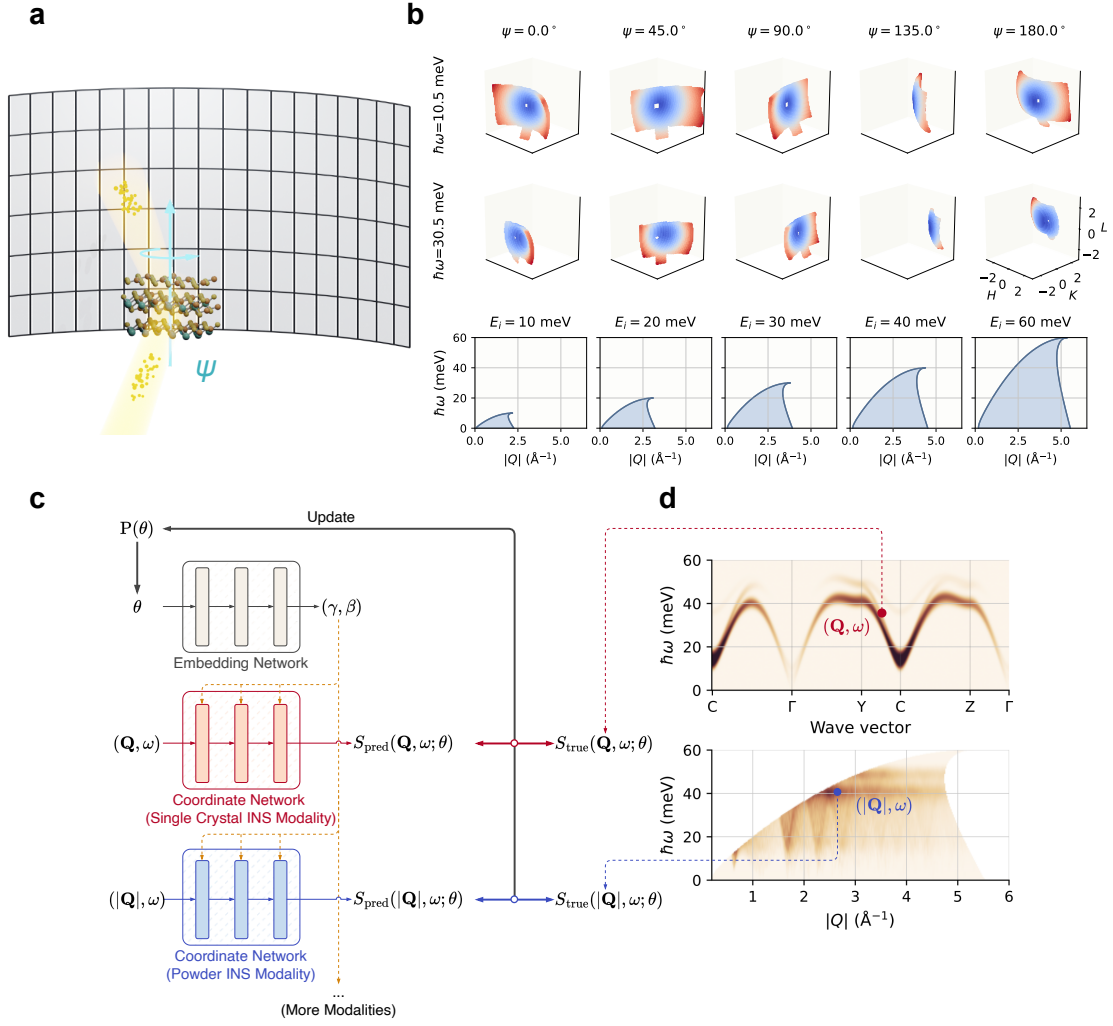


FIG. 1. **Multimodal inelastic neutron scattering (INS) measurements and Hamiltonian-conditioned AI surrogate model.** (a) Incident neutrons scatter from the sample and are recorded by a detector array, sampling a subset of the four-dimensional scattering space $(\mathbf{Q}, \omega)$ for single-crystal measurements or a bounded region in $(|\mathbf{Q}|, \omega)$ for powder measurements. (b) Experimental control variables and the corresponding measurement coverage. The upper panel illustrates the accessible $(\mathbf{Q}, \omega)$ region for different single-crystal sample orientations, while the lower panel shows the accessible $(|\mathbf{Q}|, \omega)$ region for different incident neutron energies in powder measurements. (c) AI surrogate model based on Hamiltonian-conditioned implicit neural representations (INRs) for multimodal INS. A shared embedding network encodes the Hamiltonian parameters into a common latent representation that conditions modality-specific coordinate networks, which predict scattering intensities at arbitrary $(\mathbf{Q}, \omega)$ coordinates for single-crystal measurements and $(|\mathbf{Q}|, \omega)$ coordinates for powder measurements. The trained surrogate enables efficient Bayesian inference of Hamiltonian parameters from multimodal INS measurements. (d) Representative single-crystal (top) and powder (bottom) INS spectra illustrating the complementary reciprocal-space coverage of the two measurement modalities.

by the detector coverage and the kinematically accessible combinations of momentum transfer $\mathbf{Q} = [H, K, L]$ and energy transfer $\hbar\omega$. Representative measurement coverages for different rotation angles are shown in the upper panels of Fig. 1(b). For a powder sample, the random orientation of crystallites reduces the measurement to the two-dimensional response $S(Q, \omega)$, where $Q = |\mathbf{Q}|$ is the magnitude of the momentum transfer. The corresponding detector coverage for different incident neutron energies is shown in the lower panels of Fig. 1(b).

We choose the van der Waals (vdW) quantum magnet

NiPS₃ as the representative material system throughout this work. Its spin Hamiltonian is given by³

$$H = \sum_{i,j} J_{ij} \mathbf{S}_i \cdot \mathbf{S}_j + \sum_i [A_x (S_i^x)^2 + A_z (S_i^z)^2], \quad (1)$$

where J_{ij} denotes the Heisenberg exchange interactions extending from the first- to fourth-nearest neighbors, parameterized by $[J_{1a}, J_{1b}, J_{2a}, J_{2b}, J_{3a}, J_{3b}, J_4]$, while A_x and A_z are the single-ion anisotropy parameters. Ref. 3 distinguished J_{1a} and J_{1b} to account for the crystallographic distortion while imposing $J_{2a} = J_{2b}$

and $J_{3a} = J_{3b}$. Because the same symmetry reduction formally permits the corresponding second- and third-neighbor exchanges to be distinct, we relax these constraints and treat J_{2a} , J_{2b} , J_{3a} , and J_{3b} as independent parameters. These seven exchange and two anisotropy parameters together define a nine-dimensional (9D) Hamiltonian parameter space, which is denoted as $\theta = [A_x, A_z, J_{1a}, J_{1b}, J_{2a}, J_{2b}, J_{3a}, J_{3b}, J_4]$.

To enable rapid, differentiable forward evaluations, we construct an AI surrogate model that maps Hamiltonian parameters to INS spectra. To accommodate multiple INS measurement modalities, the model consists of a shared Hamiltonian parameter embedding network coupled with modality-specific INRs⁴³, which serve as coordinate networks. The embedding network conditions the modality-specific INRs on the Hamiltonian parameters, allowing different parameter sets to generate distinct dynamical structure factors, $S(\mathbf{Q}, \omega)$. For example, the coordinate network for single-crystal INS maps the momentum and energy transfer coordinates, $(\mathbf{Q}, \omega)$, to the scattering intensity. Combined with the parameter embedding network, the surrogate realizes the conditional mapping $(\mathbf{Q}, \omega; \theta) \mapsto S(\mathbf{Q}, \omega)$, where θ denotes the Hamiltonian parameters. By using a shared parameter embedding network to condition multiple observables, the intrinsically multimodal design provides a physics-based, extensible, and flexible framework. The overall architecture is illustrated in Fig. 1(c), with additional details on the datasets and model architecture provided in Methods and Supplementary Note 4.

The differentiability and computational efficiency of the surrogate model enable seamless integration with Bayesian inference and Bayesian experimental design^{46–50}. In particular, the surrogate-predicted spectra, S_{pred} , serve as the forward model in the likelihood function, where likelihoods are evaluated by direct comparison with the measured spectra, S_{targ} , as illustrated in Fig. 1(d).

Observation-induced geometry of Hamiltonian parameter space

We define the empirical spectral distance between two Hamiltonian parameter sets using the normalized L^2 distance between their corresponding scalar scattering intensities,

$$d_N(\theta, \theta') = \left[\frac{1}{N} \sum_{i=1}^N |S(x_i; \theta) - S(x_i; \theta')|^2 \right]^{1/2}, \quad (2)$$

where $x_i = (\mathbf{Q}_i, \omega_i)$ are the sampled momentum–energy coordinates. This is a metric on the sampled spectral vectors and induces a pseudometric on the Hamiltonian parameter space; it becomes a metric there if the forward map is injective. To characterize its local geometry, we define the smooth squared-distance landscape $f(\theta) = d_N^2(\theta, \theta_{\text{ref}})$, where the reference Hamiltonian θ_{ref} is taken from Ref. 3, whose parameter values are listed in Supplementary Note 5.

An example of the local spectral-discrepancy landscape, $f(\theta)$, for single-crystal INS is shown in Fig. 2(a). Several notable features emerge. First, the landscape is relatively flat along the A_x and A_z directions, indicating that these parameters weakly influence $S(\mathbf{Q}, \omega)$ near the reference Hamiltonian. Second, correlated parameter combinations involving (J_{1a}, J_{1b}) and (J_{2a}, J_{2b}) produce similar spectra, leading to elongated valleys in the observation landscape. In contrast, anti-correlated variations involving J_{3a} and J_{3b} produce only small spectral changes. These observations illustrate the central principle: spectral distinguishability is governed by the geometry induced by the forward mapping, $\theta \mapsto S(\mathbf{Q}, \omega; \theta)$, rather than by Euclidean distance in the Hamiltonian parameter space.

To characterize the observation-induced geometry more generally, let $x \in \Omega$ denote a coordinate point in the observation space, where Ω is the experimentally accessible measurement domain. The spectroscopic forward map $S(x; \theta)$ induces a local metric on the Hamiltonian parameter space through the normalized $L^2(\Omega)$ inner product,

$$G_{mn}(\theta) = \langle \partial_{\theta_m} S(x; \theta), \partial_{\theta_n} S(x; \theta) \rangle_{\Omega} = \frac{1}{V_{\Omega}} \int_{\Omega} \partial_{\theta_m} S(x; \theta) \partial_{\theta_n} S(x; \theta) dx, \quad (3)$$

where V_{Ω} denotes the measure of the accessible momentum–energy domain and θ_m denotes the m th component of θ . For INS, $x = (\mathbf{Q}, \omega)$, and the normalized inner product reads

$$G_{mn}(\theta) = \frac{1}{V_{\Omega}} \int_{\Omega} \frac{\partial S(\mathbf{Q}, \omega; \theta)}{\partial \theta_m} \frac{\partial S(\mathbf{Q}, \omega; \theta)}{\partial \theta_n} d^3 \mathbf{Q} d\omega, \quad (4)$$

where $V_{\Omega} = \int_{\Omega} d^3 \mathbf{Q} d\omega$ is the accessible measurement

volume. In essence, this defines the intensity-induced L^2 pullback metric. In practice, the integral is approximated using Monte Carlo sampling,

$$G_{mn}(\theta) \approx \frac{1}{N} \sum_{i=1}^N \frac{\partial S(x_i; \theta)}{\partial \theta_m} \frac{\partial S(x_i; \theta)}{\partial \theta_n}, \quad (5)$$

where $x_i = (\mathbf{Q}_i, \omega_i)$ are sampled uniformly over Ω with

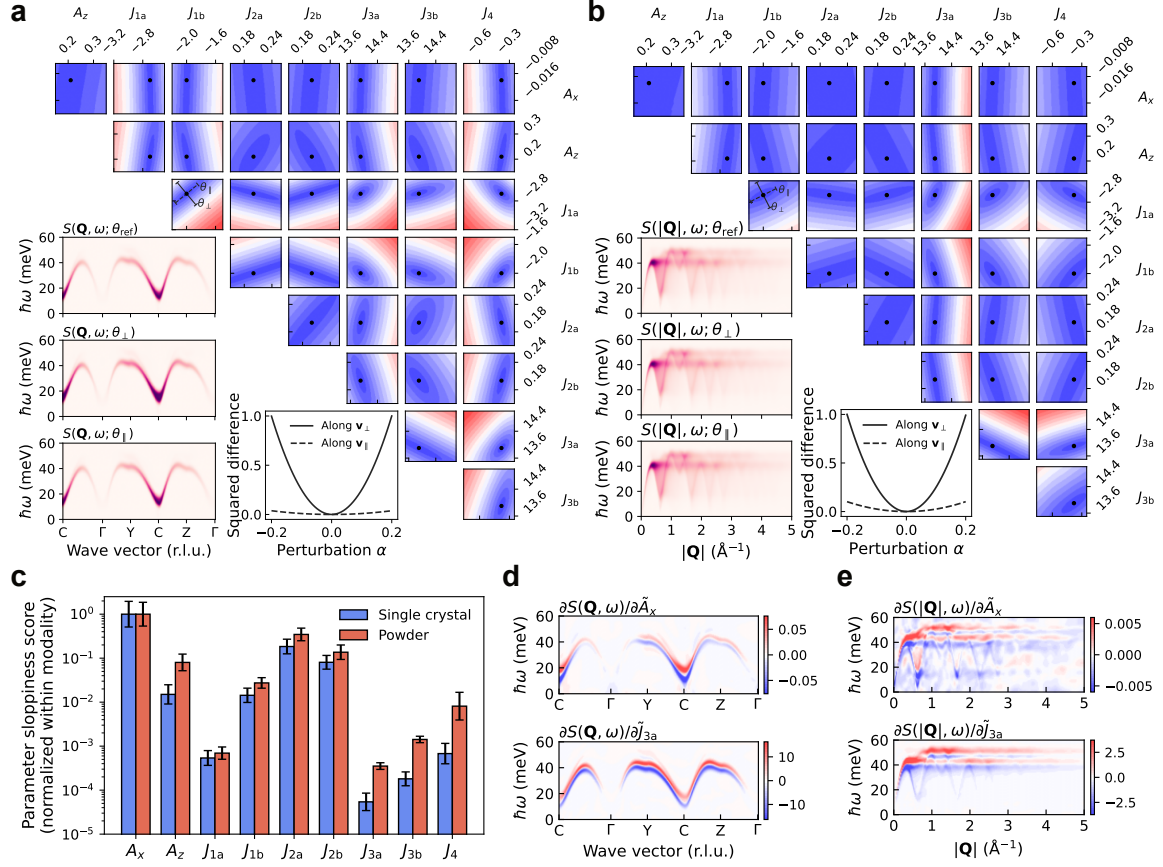


FIG. 2. Observation landscapes and local parameter identifiability across INS modalities. **a**, Local observation landscape for the single-crystal INS modality, centered at the reference Hamiltonian θ_{ref} . The corner plot shows two-parameter slices of the spectral-discrepancy landscape between the reference spectrum, $S(\mathbf{Q}, \omega; \theta_{\text{ref}})$, and spectra generated by perturbed Hamiltonians, $S(\mathbf{Q}, \omega; \theta)$. Circular markers indicate θ_{ref} . In the (J_{1a}, J_{1b}) subspace, the solid and dashed lines denote perturbations along the eigenvectors, $\mathbf{v}_\perp$ and $\mathbf{v}_\parallel$, of the local observation metric tensor $\mathbf{G}$, with eigenvalues satisfying $\lambda_\perp \geq \lambda_\parallel$. These directions correspond to the local stiff and sloppy directions in Hamiltonian parameter space, respectively. Representative spectra and the corresponding normalized squared differences demonstrate that perturbations of comparable magnitude along $\mathbf{v}_\perp$ produce substantially larger spectral changes than perturbations along $\mathbf{v}_\parallel$. **b**, Corresponding local observation landscape for the powder INS modality, illustrating the distinct observation geometry induced by orientational averaging. **c**, Parameter-wise sloppiness scores for single-crystal and powder INS, computed from the diagonal elements of the inverse local metric tensor, $\tilde{\mathbf{G}}^{-1}$, and normalized independently within each modality such that the maximum sloppiness score equals to 1. Larger values indicate Hamiltonian parameters that are less well constrained by the corresponding measurement modality. **d**, Representative partial derivatives of the dynamical structure factor with respect to selected normalized Hamiltonian parameters for the single-crystal modality, $\partial S(\mathbf{Q}, \omega) / \partial \tilde{A}_x$ and $\partial S(\mathbf{Q}, \omega) / \partial \tilde{J}_{3a}$. **e**, Representative partial derivatives of the dynamical structure factor with respect to selected normalized Hamiltonian parameters for the powder modality, $\partial S(|\mathbf{Q}|, \omega) / \partial \tilde{A}_x$ and $\partial S(|\mathbf{Q}|, \omega) / \partial \tilde{J}_{3a}$.

respect to the normalized momentum-energy volume measure. Unless otherwise specified, we use $N = 25,000$ samples to estimate the observation metric.

The eigenstructure of the pullback Gram matrix characterizes the local spectral sensitivity in the chosen Hamiltonian parameterization. Eigenvectors corresponding to large eigenvalues define *stiff* directions, along which small parameter perturbations produce substantial changes in the predicted spectra. Eigenvectors with small eigenvalues define *sloppy* directions, along which local parameter perturbations produce comparatively small spectral changes. These directions characterize sensitivity un-

der the selected L^2 measure and do not by themselves incorporate counting noise, acquisition time, or prior information.

An example within the (J_{1a}, J_{1b}) subspace is shown in the corner plot of Fig. 2(a). The perturbed Hamiltonians $\theta_\perp$ and $\theta_\parallel$ are displaced by an equal distance of 0.2 meV from the reference Hamiltonian θ_{ref} along the stiff and sloppy eigen-directions of the local 2D subspace metric tensor $\mathbf{G}(\theta_{\text{ref}})$, respectively. The corresponding spectra, $S(\mathbf{Q}, \omega; \theta_\perp)$ and $S(\mathbf{Q}, \omega; \theta_\parallel)$, are shown in the lower-left panels of Fig. 2, while the corresponding 9D eigenvectors are visualized in Supplementary Note 7. Pertur-

bations along the stiff direction produce substantially larger spectral changes relative to $S(\mathbf{Q}, \omega; \boldsymbol{\theta}_{\text{ref}})$ than perturbations along the sloppy direction. The normalized squared spectral differences, shown in the lower-right panels of Fig. 2, further highlight the different sensitivities along these two directions. These results demonstrate that parameter-wise resolving power is fundamentally governed by the observation geometry induced by the forward map. Rather than merely identifying which Hamiltonian best reproduces the measurements, the observation geometry characterizes what information the measurements can and cannot resolve.

In Fig. 2(b), we present the corresponding local observation landscape for the powder INS modality. Relative to single-crystal INS, powder averaging produces broader low-discrepancy regions in Hamiltonian parameter space, reflecting the loss of directional information. As a result, orientationally averaged spectra are less effective at distinguishing correlated parameter variations. The representative powder spectra and their corresponding squared-difference maps again reveal stiff and sloppy parameter combinations, although the spectral response is generally less discriminative than in single-crystal INS. Despite this reduced parameter-wise resolving power, the observation geometry demonstrates that powder INS remains informative for Hamiltonian inference. In particular, it provides meaningful constraints on the Hamiltonian parameter space, making it a practical alternative when sufficiently large single crystals are unavailable or when rapid measurements with limited beam time are preferred. The complete 9D eigen-structures of the observation metric for both single-crystal and powder modalities are provided in Supplementary Note 7.

To compare the local identifiability of Hamiltonian parameters with different scales, we evaluate the observation metric in locally normalized coordinates. For a metric evaluated at a reference parameter point $\hat{\boldsymbol{\theta}}$, we define the dimensionless coordinates

$$\tilde{\theta}_m = \frac{\theta_m - \hat{\theta}_m}{|\hat{\theta}_m|}. \quad (6)$$

The corresponding scaled observation metric is

$$\tilde{\mathbf{G}}(\hat{\boldsymbol{\theta}}) = \mathbf{D}(\hat{\boldsymbol{\theta}})\mathbf{G}(\hat{\boldsymbol{\theta}})\mathbf{D}(\hat{\boldsymbol{\theta}}), \quad \mathbf{D}(\hat{\boldsymbol{\theta}}) = \text{diag}\left(|\hat{\theta}_1|, \dots, |\hat{\theta}_M|\right), \quad (7)$$

where M is the number of Hamiltonian parameters. This scaled metric removes the dependence of the local geometry on the absolute magnitudes of the Hamiltonian parameters, enabling more intuitive comparisons of parameter identifiability across different Hamiltonian parameters. We quantify the sloppiness of parameter θ_m using the corresponding diagonal element of the inverse scaled metric,

$$\mathcal{S}_m(\hat{\boldsymbol{\theta}}) = \left[\left(\tilde{\mathbf{G}}(\hat{\boldsymbol{\theta}}) + \epsilon \mathbf{I} \right)^{-1} \right]_{mm}, \quad (8)$$

where ϵ is a small numerical regularizer ($\epsilon = 10^{-12}$ throughout this work). Larger values of $\mathcal{S}_m$ indicate that

parameter θ_m is less locally identifiable from the measured spectra.

To obtain a representative measure of parameter identifiability across the Hamiltonian parameter space, we evaluate the sloppiness scores at 100 randomly sampled Hamiltonians $\hat{\boldsymbol{\theta}}$ and report their averages in Fig. 2(c). For visualization, the sloppiness scores are normalized within each measurement modality by their largest value. For both powder and single-crystal INS, the transverse anisotropy A_x is identified as the least constrained parameter and therefore sets the normalization scale. Because the scores are normalized independently, this comparison concerns relative parameter-wise sloppiness within each modality rather than the absolute magnitude of the metric between modalities. Relative to this common modality-wise reference, all remaining Hamiltonian parameters exhibit systematically larger normalized sloppiness scores under powder INS than under single-crystal INS (Fig. 2(c)). This indicates that the parameter-wise resolving power is less differentiated in the powder modality, with a broader set of Hamiltonian parameters remaining relatively sloppy compared with the least constrained direction. This trend reflects the orientational averaging inherent in powder measurements, which suppresses directional spectral features and expands the set of Hamiltonian perturbations that produce nearly indistinguishable spectra.

The partial derivatives shown in Fig. 2(d,e) provide an intuitive visualization of the local observation geometry by revealing how infinitesimal perturbations of individual Hamiltonian parameters modify the measured spectra. Consistent with the sloppiness analysis, parameters with weaker spectral sensitivity generally exhibit smaller derivative magnitudes over the measured momentum-energy domain. Because these derivatives are obtained directly by automatic differentiation of the AI surrogate, they can be evaluated efficiently, in contrast to conventional forward solvers where derivative information is typically unavailable or substantially more expensive to compute numerically.

More generally, the parameter-wise sloppiness profile provides a quantitative description of how effectively a given measurement modality constrains different directions in the Hamiltonian parameter space. This observation suggests a practical principle for multimodal experimental design: complementary measurements are most informative when they exhibit complementary sloppy directions, allowing one modality to constrain parameter combinations that remain weakly identifiable in another. The powder-single-crystal combination considered here serves primarily as a proof of concept, since powder INS is effectively an orientationally averaged projection of single-crystal scattering. Even greater gains can be expected when combining experimental probes with genuinely complementary observation operators and correspondingly distinct observation geometries and sloppiness profiles.

Posterior-guided multimodal Hamiltonian inference and adaptive experimental design

The proposed framework advances our understanding of how Hamiltonian parameters are jointly constrained by multimodal experimental observables. By learning a differentiable and computationally efficient mapping between the Hamiltonian parameter space and INS spectra, the AI surrogate provides the foundation for Bayesian experimental design, enabling posterior-guided Hamiltonian inference that integrates heterogeneous measurements and adaptively designs subsequent experiments.

In INS, single-crystal measurements preserve directional momentum information and can resolve fine features of the dynamical structure factor, $S(\mathbf{Q}, \omega)$. However, each crystal orientation probes only a subset of reciprocal space, making comprehensive reciprocal-space coverage experimentally demanding. In contrast, powder INS rapidly provides an orientationally averaged overview of the excitation spectrum, although the averaging process obscures directional information. These complementary measurements therefore induce distinct observation geometries for the same underlying Hamiltonian: powder measurements provide an efficient global characterization, while single-crystal measurements subsequently resolve the remaining microscopic ambiguities.

To exploit this complementarity, we perform sequential Bayesian inference in which information inferred from powder INS is used to initialize the subsequent single-crystal analysis. We first infer the powder posterior $P_{\text{pd}}(\boldsymbol{\theta})$, represented computationally by the particle approximation $\mathcal{P}_{\text{pd}}$. The initial particle distribution for the single-crystal stage is then constructed from $\mathcal{P}_{\text{pd}}$ using the partial posterior initialization procedure described in Methods. After acquiring the $(t+1)$ th single-crystal measurement at orientation $\psi^{(t+1)}$, the posterior is updated according to

$$P_{\text{sc}}^{(t+1)}(\boldsymbol{\theta}) \propto P\left(S_{\text{sc}}^{(t+1)} \mid \boldsymbol{\theta}, \psi^{(t+1)}\right) P_{\text{sc}}^{(t)}(\boldsymbol{\theta}), \quad (9)$$

where $S_{\text{sc}}^{(t+1)}$ denotes the measured single-crystal spectrum acquired at iteration $t+1$, and $\psi^{(t+1)}$ is the corresponding sample orientation.

Unlike many multimodal learning approaches^{41,42}, which fuse heterogeneous observations through learned latent representations, our framework connects powder and single-crystal measurements directly through the shared Hamiltonian parameter space $\boldsymbol{\theta}$. The modality-specific surrogate branches map the same Hamiltonian parameters to powder and single-crystal response functions in their respective observation spaces. Consequently, posterior uncertainty inferred from powder INS can be propagated directly through the shared Hamiltonian parameterization to predict single-crystal spectra, enabling principled experimental planning in the subsequent measurement stage.

The next single-crystal measurement is selected using Bayesian experimental design. Here, we demonstrate the

experimental design capability using the crystal orientation ψ while this can be generalized to other experimental settings. We approximate the expected information gain using the predictive spectral variance^{45,48,49,51}, leading to the utility function

$$U^{(t+1)}(\psi) = \frac{1}{N_{\mathbf{Q}, \omega}^{(\psi)}} \sum_{(\mathbf{Q}, \omega) \in \Omega_{\psi}} \text{Std}_{\boldsymbol{\theta} \sim P_{\text{sc}}^{(t)}(\boldsymbol{\theta})} [S_{\text{sc}}(\mathbf{Q}, \omega; \boldsymbol{\theta})], \quad (10)$$

where Ω_{ψ} denotes the momentum–energy region accessible at experimental setting ψ , and $N_{\mathbf{Q}, \omega}^{(\psi)}$ is the number of sampled $(\mathbf{Q}, \omega)$ points within this region. In practice, the utility is further adjusted to account for experimental preferences, including penalties for repeated measurements and large changes in sample orientation (see Methods and Supplementary Information). The next experimental setting (i.e., orientation $\psi^{(t+1)}$) is then selected by maximizing $U^{(t+1)}(\psi)$ over the candidate orientations.

We demonstrate the proposed workflow using simulated INS experiments (Fig. 3). For each ground-truth Hamiltonian used in the simulated experiments, synthetic single-crystal spectra are generated from $S_{\text{sc}}(\mathbf{Q}, \omega; \boldsymbol{\theta})$ with Poisson counting noise and treated as experimental observations. Figures 3(a) and (b) illustrate the adaptive single-crystal acquisition strategies obtained without and with powder-informed initialization, respectively. In each panel, the heatmap shows the average predictive-variance utility, $U^{(t)}(\psi)$, computed from 10 simulated experiments whose ground-truth Hamiltonians are randomly sampled from the initial uniform prior $P^{(0)}(\boldsymbol{\theta})$ defined in Methods.

The powder-informed initialization reshapes the posterior uncertainty before the single-crystal stage begins, altering the predictive utility landscape over the candidate crystal orientations. Consequently, the preferred measurement sequence adapts to the remaining uncertainty, leading to different orientation trajectories than those obtained from an uninformative prior. The overlaid trajectory shows the sequence of crystal orientations selected for a representative simulated experiment generated from the same reference Hamiltonian, $\boldsymbol{\theta}_{\text{ref}}$, used in Fig. 2. Figure 3(c) compares the posterior evolution for the single-crystal-only and powder-initialized protocols. The marginal distributions along the top and left illustrate the progressive reduction of parameter uncertainty with successive measurements, while the lower-right corner plot shows the final posterior distribution after ten single-crystal acquisitions.

Figure 3(d) compares the Hamiltonian parameter estimation errors obtained using powder-only, single-crystal-only, and staged multimodal acquisition protocols. Consistent with the parameter-wise sloppiness analysis in Fig. 2(c), powder-only inference produces the largest errors for most parameters because orientational averaging removes directional information that is retained in single-crystal measurements. Consequently, single-crystal INS generally achieves the lowest point-estimation errors, reflecting its ability to resolve the full directional structure

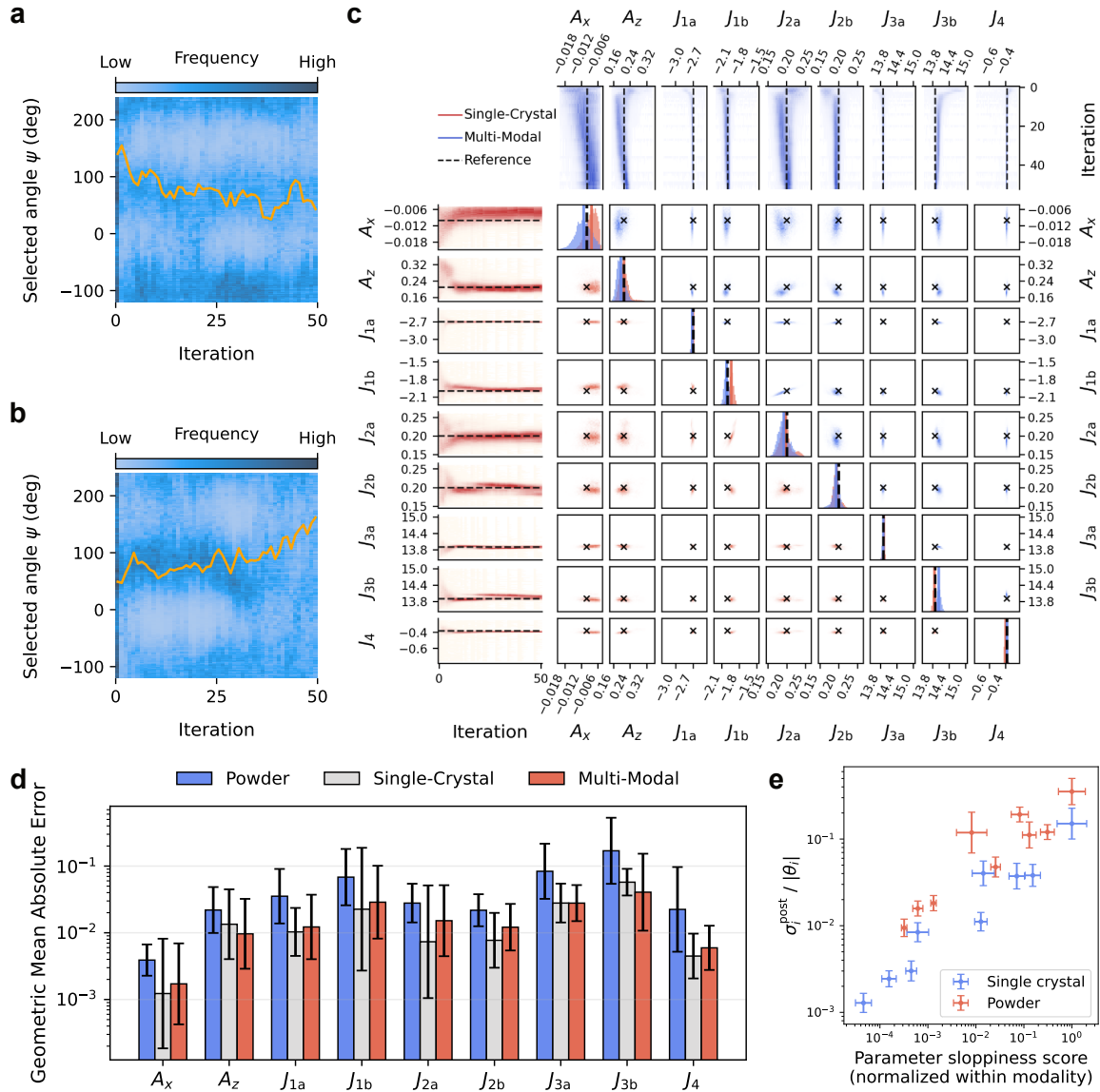


FIG. 3. Bayesian experimental design and multimodal Hamiltonian inference enabled by AI surrogates. **a**, Adaptive experimental design on orientation ψ for single-crystal INS without powder-derived prior information. The orange curve shows the sequence of selected sample orientations ψ for a representative simulated experiment generated from $S(\mathbf{Q}, \omega; \theta_{\text{ref}})$ with Poisson noise. The heatmap shows the average predictive-variance utility over 10 simulated experiments with ground-truth Hamiltonians sampled from the initial prior. **b**, Adaptive experimental design on orientation ψ for single-crystal INS initialized with a posterior obtained from powder INS measurements. Compared with panel (a), the suggested orientations shift toward different regions, reflecting the additional information contributed by the powder modality. **c**, Posterior distributions of Hamiltonian parameters. The top and left panels show the marginal posterior evolution as a function of measurement steps, while the corner plot illustrates the posterior distribution after ten single-crystal measurements. **d**, Parameter estimation error for powder-only, single-crystal-only, and multimodal inference. Bars show the geometric mean of absolute error across benchmark realizations over sampled θ , computed as $y = \exp[(\log \hat{\theta} - \theta^*)]$ and displayed on a logarithmic scale. Error bars indicate one log-standard-deviation interval, where $\sigma_{\log} = \text{std}[\log \hat{\theta} - \theta^*]$, corresponding to multiplicative intervals $[y \exp(-\sigma_{\log}), y \exp(\sigma_{\log})]$. **e**, Relationship between the parameter sloppiness score and the residual posterior uncertainty after sequential single-crystal inference. Each point corresponds to one Hamiltonian parameter. The positive correlation demonstrates that the local observation geometry provides a useful predictor of the uncertainty remaining after Bayesian inference.

of $S(\mathbf{Q}, \omega)$.

Although powder-initialized measurement does not uniformly reduce the point-estimation error relative to the single-crystal-only measurement, it produces narrower posterior distributions for sloppy Hamiltonian parameters including A_x , J_{2a} , and J_{2b} (Fig. 3(c)) compared with single-crystal-only measurements. This result indicates that powder measurements effectively concentrate the posterior distribution before single-crystal measurements are acquired, allowing the subsequent measurements to focus on the remaining parameter ambiguities. Rather than providing fundamentally independent information beyond an ideal, fully sampled single-crystal experiment, powder INS serves as an experimentally efficient initialization that enables the subsequent single-crystal measurement budget to be allocated more effectively.

To assess whether the local observation geometry predicts the practical identifiability of Hamiltonian parameters, Fig. 3(e) compares the parameter sloppiness scores with the normalized posterior uncertainties obtained after Bayesian inference. The horizontal axis shows the sloppiness score computed from the inverse scaled observation metric, while the vertical axis shows the posterior standard deviation normalized by the magnitude of the corresponding reference parameter, $\sigma_i^{\text{post}}/|\theta_i^{\text{ref}}|$. The positive correlation demonstrates that the local observation geometry provides an effective *a priori* predictor of the parameter uncertainty remaining after inference. This result suggests that the observation geometry can serve as a diagnostic of the parameter-resolving capability of a measurement configuration solely from the forward model, before experimental data are acquired.

The powder–single-crystal example presented here serves primarily as a proof of concept for the proposed multimodal framework. More generally, the same methodology applies to any combination of experimental probes whose observations can be linked through a shared microscopic Hamiltonian.

Validation with experimental multimodal INS data

We now apply our multimodal AI surrogate to infer Hamiltonian parameters from experimental INS data collected from powder and single-crystal NiPS₃ samples. Details about experimental data preparation are presented in Methods. As in the staged workflow introduced in the previous section, the two measurement modalities are coupled through the shared Hamiltonian parameter space. The objective is therefore not simply to fit each spectrum independently, but to infer a posterior distribution over Hamiltonian parameters that is simultaneously consistent with both powder and single-crystal observations.

Figure 4(a) shows the posterior distributions obtained from the two-stage Bayesian inference using the multimodal surrogate. The blue posterior corresponds to

the unconstrained nine-parameter Hamiltonian, whereas the red posterior incorporates the symmetry constraints $J_{2a} = J_{2b}$ and $J_{3a} = J_{3b}$ following Ref. 3, thereby reducing the parameter space from nine to seven dimensions. The constrained posterior is centered near the literature-reported Hamiltonian, while the unconstrained model reveals the associated uncertainty and parameter correlations when these symmetry constraints are relaxed.

Despite the additional degrees of freedom in the unconstrained model, the marginal posteriors of the remaining parameters are broadly consistent between the two formulations. The posterior widths and correlations further indicate that some exchange parameters remain more weakly constrained than others, such as A_x , A_z , J_{2a} , and J_{2b} , consistent with the sloppy directions identified by the local observation geometry.

This extension from a seven-dimensional to a nine-dimensional Hamiltonian is computationally nontrivial for conventional grid-based forward-model evaluation. For a fixed resolution of N points along each parameter dimension, the number of required forward calculations increases by a factor of N^2 . To put this into perspective, using the same computational budget as the surrogate training dataset used in this work (18,000 forward-model evaluations; see Methods), a direct grid search over the nine-dimensional Hamiltonian space would provide only about three grid points along each parameter dimension, making brute-force Hamiltonian fitting impractical. Instead, these forward-model evaluations are used to train a differentiable neural surrogate that can subsequently be evaluated rapidly throughout the Hamiltonian parameter space, enabling efficient Bayesian inference and experimental design.

Having obtained the unconstrained posterior, we next assess how well its Hamiltonian samples jointly reproduce the experimental powder and single-crystal measurements. We draw 2,048 samples from the posterior and use independent Sunny calculations at the corresponding Hamiltonian parameter sets to generate the associated spectra for comparison with experiment. The scatter plot in Fig. 4(b) compares the Pearson correlation coefficients (PCCs) between the experimental data and the corresponding Sunny-generated LSWT spectra for each sampled Hamiltonian. The horizontal and vertical axes denote PCC_{sc} (PCC against the single-crystal experimental data) and PCC_{pd} (PCC against the powder experimental data), respectively, so that each point represents a single Hamiltonian evaluated simultaneously against both experimental modalities. Larger PCC values therefore indicate better joint agreement with the measured spectra. A pixel-wise decomposition illustrating the contributions to the PCC metric is provided in Supplementary Note 8.

The scatter plot is overlaid with the reference Hamiltonian θ_{ref} reported in Ref. 3, which was obtained from single-crystal measurements. Although individual posterior samples can fit one modality better than the other, the joint posterior-predictive comparison identi-

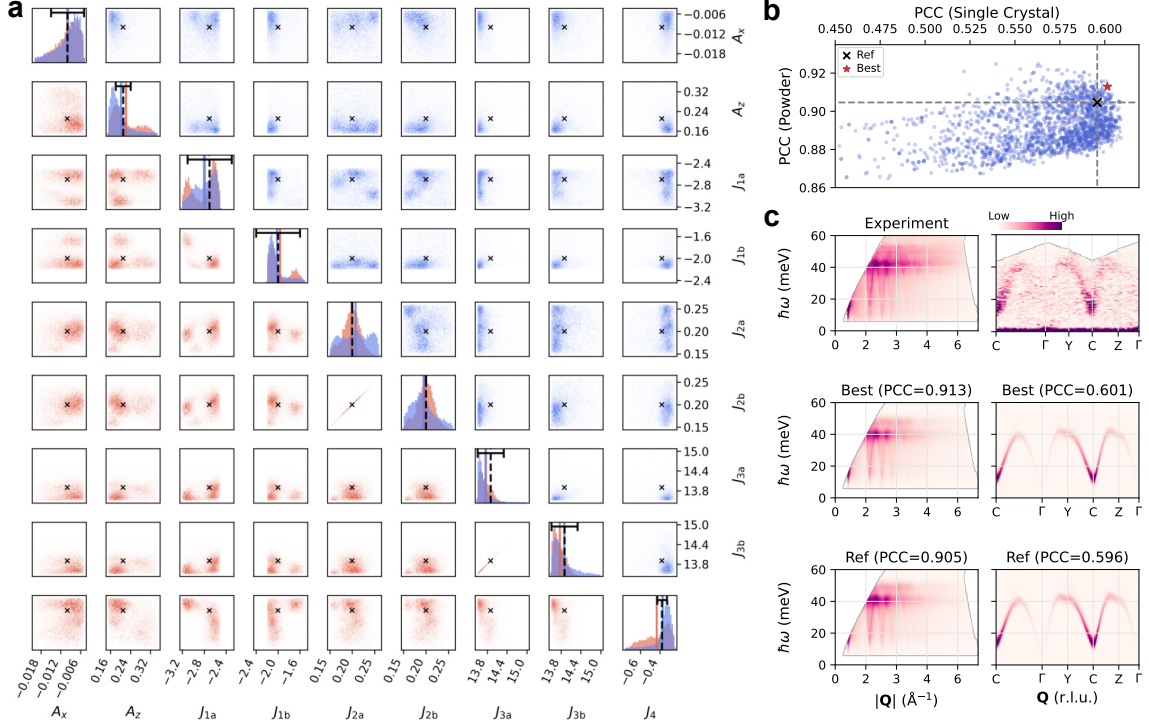


FIG. 4. **Validation with experimental multimodal INS data.** **a**, Posterior distributions inferred from multimodal INS measurements with (red) and without (blue) the exchange-parameter constraints $J_{2a} = J_{2b}$ and $J_{3a} = J_{3b}$. The corner plot shows the final posterior distributions and parameter correlations, while the diagonal panels show the corresponding marginal posterior distributions. Black crosses and dashed vertical lines denote the reference Hamiltonian parameters θ_{ref} , and horizontal error bars indicate the corresponding one-standard-deviation uncertainties reported in Ref. 3. **b**, Pearson correlation coefficients (PCCs) for powder and single-crystal INS evaluated using 2,048 samples drawn from the unconstrained posterior. The selected “best” Hamiltonian θ_{best} is the posterior sample with the smallest normalized distance to the ideal agreement point, as defined in Eq. (12). **c**, Comparison of powder (left) and single-crystal (right) INS spectra. Within each column, from top to bottom: experimental spectrum, prediction from the selected posterior sample θ_{best} , and prediction from the reference Hamiltonian θ_{ref} .

fies Hamiltonians that achieve balanced agreement across both measurements.

In particular, let $\mathcal{E}$ denote the set of posterior samples whose single-crystal and powder PCCs are no smaller than the corresponding reference values:

$$\mathcal{E} = \{\theta^{(k)} : \text{PCC}_{\text{sc}}(\theta^{(k)}) \geq \text{PCC}_{\text{sc}}(\theta_{\text{ref}}), \text{PCC}_{\text{pd}}(\theta^{(k)}) \geq \text{PCC}_{\text{pd}}(\theta_{\text{ref}})\}.$$

Assuming that $\mathcal{E}$ is nonempty and that the PCC range is nonzero for each modality, define

$$\widetilde{\text{PCC}}_{\alpha}(\theta) = \frac{\text{PCC}_{\alpha}(\theta) - \min_{\vartheta \in \mathcal{E}} \text{PCC}_{\alpha}(\vartheta)}{\max_{\vartheta \in \mathcal{E}} \text{PCC}_{\alpha}(\vartheta) - \min_{\vartheta \in \mathcal{E}} \text{PCC}_{\alpha}(\vartheta)}, \quad (11)$$

where $\alpha \in \{\text{sc}, \text{pd}\}$. We select the Hamiltonian closest to the component-wise ideal point (1, 1) in the normalized PCC plane:

$$\theta_{\text{best}} \in \arg \min_{\theta \in \mathcal{E}} \sqrt{[1 - \widetilde{\text{PCC}}_{\text{sc}}(\theta)]^2 + [1 - \widetilde{\text{PCC}}_{\text{pd}}(\theta)]^2}. \quad (12)$$

The corresponding Hamiltonian parameter vectors for both the reference model, θ_{ref} , and the selected posterior sample, θ_{best} , are listed in Supplementary Note 5.

Figures 4(b) and (c) compare the measured and predicted powder and single-crystal spectra, respectively. For the powder INS, the left panels show the experimental $S(|\mathbf{Q}|, \omega)$ together with the spectrum generated by the “best” posterior sample, $S(|\mathbf{Q}|, \omega; \theta_{\text{best}})$, and that generated by the literature reference Hamiltonian, $S(|\mathbf{Q}|, \omega; \theta_{\text{ref}})$. The right panels show the corresponding spectra for the single-crystal modality along the high-symmetry path.

We emphasize here that the purpose of this comparison is not to revise the Hamiltonian parameters reported in Ref. 3, but to demonstrate how the present framework identifies alternative parameter sets that are consistent with the processed experimental spectra and characterizes the associated parameter ambiguities.

Both Hamiltonian estimates reproduce the dominant spectral features in both modalities, but differ in subtle intensity distributions and dispersive details that remain important for resolving the underlying microscopic interactions. In particular, θ_{best} yields higher PCC values

for both powder and single-crystal measurements than θ_{ref} , indicating that the joint multimodal posterior identifies Hamiltonians with improved simultaneous agreement with the processed experimental spectra considered here. Interestingly, the posterior samples span a much narrower range of powder agreement than single-crystal agreement. This behavior reflects the weaker constraints imposed by powder INS: many distinct Hamiltonians produce similarly good powder spectra, whereas the directional information preserved by single-crystal INS more effectively differentiates among competing microscopic models.

Beyond returning a representative “best” Hamiltonian, the inference yields a full posterior distribution over Hamiltonian parameters, enabling direct quantification of the uncertainty associated with each microscopic interaction. Rather than collapsing the inference to a single point estimate, the posterior distributions and parameter correlations reveal which interaction combinations remain poorly resolved by the available measurements, consistent with the locally sloppy directions predicted by the observation geometry.

This experimental validation demonstrates that the proposed framework extends beyond simulated datasets to real multimodal INS measurements, while preserving physical interpretability through the shared Hamiltonian parameter space. The framework also provides a principled approach for determining not only which microscopic models are consistent with heterogeneous experimental observations, but also which interactions remain fundamentally unresolved by the available measurements.

DISCUSSION

In this work, we present an AI-enabled framework that not only accelerates Hamiltonian inference from experimentally measured spectra, but also provides a probabilistic and physically interpretable approach to Hamiltonian inference. By combining Hamiltonian-conditioned neural surrogates with Bayesian inference, local observation geometry, and Bayesian experimental design, the framework quantifies uncertainty, integrates complementary measurement modalities, and predicts the parameter-resolving capability of a given measurement configuration. Rather than simply producing a best-fit Hamiltonian, the framework characterizes the information content of spectroscopy measurements and the remaining ambiguities in the inferred microscopic model.

We demonstrate these capabilities using NiPS₃ as a representative quantum magnet with both powder and single-crystal INS measurements. The learned surrogate enables rapid, differentiable evaluation of neutron-scattering spectra throughout the Hamiltonian parameter space, making it practical to construct observation landscapes that reveal stiff and sloppy parameter directions. These landscapes provide intuitive and quantita-

tive insight into why certain interactions are well constrained by INS whereas others remain only weakly identifiable. Importantly, the observation geometry itself is independent of the neural surrogate and can, in principle, be defined for any differentiable forward model. However, repeated evaluation of conventional forward solvers rapidly becomes computationally prohibitive in high-dimensional Hamiltonian spaces. For example, a uniform nine-dimensional grid with the same computational budget as the training dataset used in this work would provide only about three sampling points along each parameter dimension, far too coarse to accurately resolve the posterior distribution. The surrogate therefore serves as an enabling computational accelerator that makes repeated evaluation of the observation geometry, Bayesian inference, and adaptive experimental design practical for realistic high-dimensional Hamiltonian spaces.

A key strength of the framework is its physically grounded treatment of multimodal inference. Rather than coupling heterogeneous measurements through an abstract learned latent representation, powder and single-crystal INS are linked through their shared microscopic Hamiltonian parameter space. This enables posterior information inferred from one modality to be propagated naturally to another, supporting sequential Bayesian inference, joint parameter estimation, and adaptive experimental design. In the experimental analysis of NiPS₃, this strategy yields posterior distributions that simultaneously reproduce both powder and single-crystal spectra while explicitly quantifying the remaining uncertainty. Notably, starting from an established Hamiltonian obtained through conventional refinement, the framework identifies alternative parameter combinations consistent with the observations, providing a complementary characterization of parameter ambiguities beyond the single best-fit solution.

The present framework also has several limitations. First, the surrogate model is trained on spectra generated from an assumed Hamiltonian family and forward solver, so its predictive accuracy is ultimately limited by the fidelity of the underlying physical model, the quality of the approximation method used to calculate the spectra (here spin wave theory), and the coverage of the training dataset. Second, although neural surrogates provide highly efficient forward evaluations after training, generating sufficiently informative training data remains computationally demanding for high-dimensional Hamiltonians. Third, the inferred posterior quantifies uncertainty only within the assumed model class. Missing interactions, disorder, domain structure, instrumental effects, or other sources of model mismatch may therefore manifest as broadened posterior distributions or systematic residuals rather than being explicitly represented. Future work could address these limitations by incorporating surrogate-model uncertainty, for example through Bayesian neural networks or ensemble methods, and by using cross-modal residuals as diagnostics of model inad-

equacy.

Beyond Hamiltonian inference, the inferred posterior may itself encode useful physical information. Real materials often exhibit spatial inhomogeneity, strain, stacking faults, domains, or local disorder, giving rise to distributions of effective microscopic parameters rather than a single uniform Hamiltonian. Although the present work interprets the posterior primarily as epistemic uncertainty, future hierarchical Bayesian models could distinguish uncertainty arising from limited measurements from intrinsic variations within the sample. Such extensions would enable multimodal spectroscopy to probe not only the most probable Hamiltonian but also the distribution of microscopic environments contributing to the measured response.

Our framework uses AI not as an end-to-end inverse model, but as a reusable surrogate for computationally expensive physics calculations while keeping the inference grounded in microscopic Hamiltonian parameters and experimentally defined observables. This separation between the physical model, the learned surrogate, and the Bayesian inference engine preserves physical interpretability while enabling efficient uncertainty quantification and experimental design. The central requirement of this methodology, namely a computationally efficient, differentiable forward model, is shared by many spectroscopic and scattering techniques, providing a basis for applying the framework beyond INS to domains such as resonant inelastic X-ray scattering, Raman spectroscopy, X-ray absorption spectroscopy, optical spectroscopy, and transport measurements. By revealing how measurement modalities constrain microscopic interactions and identifying experiments that efficiently resolve remaining ambiguities, the framework establishes a general strategy for uncertainty-aware Hamiltonian inference and adaptive multimodal experimental design in quantum materials. Because the observation geometry is defined entirely by the forward model, the same framework naturally extends to optimizing additional experimental control variables, such as sample orientation, incident energy, magnetic field, pressure, temperature, or other external perturbations, as well as coordinating complementary measurements across different experimental modalities. This statistical decision-making layer, combined with emerging agentic AI-driven laboratory automation^{52–56}, could enable closed-loop autonomous scientific experiments that iteratively identify the most informative measurements, acquire new data, and refine microscopic models. By unifying physically interpretable AI surrogates, Bayesian inference, observation geometry, and adaptive experimental design, this work provides a foundation for autonomous scientific discovery in quantum materials.

METHODS

Bayesian inference and experimental design

Unless otherwise specified, we adopt independent uniform priors over the nine Hamiltonian parameters,

$$P^{(0)}(\boldsymbol{\theta}) = \prod_{m=1}^9 \mathcal{U}(\theta_m; \theta_{m,\text{lb}}, \theta_{m,\text{ub}}), \quad (13)$$

where the lower and upper bounds are

$$\begin{aligned} \boldsymbol{\theta}_{\text{lb}} = & (-0.020, 0.15, -3.2, -2.2, 0.15, \\ & 0.15, 13.5, 13.5, -0.76) \text{ and} \\ \boldsymbol{\theta}_{\text{ub}} = & (-0.005, 0.35, -2.5, -1.5, 0.26, \\ & 0.26, 15.0, 15.0, -0.25), \end{aligned} \quad (14)$$

in the parameter order

$$\boldsymbol{\theta} = (A_x, A_z, J_{1a}, J_{1b}, J_{2a}, J_{2b}, J_{3a}, J_{3b}, J_4). \quad (15)$$

The prior ranges were chosen to encompass physically plausible Hamiltonian values around the reference parameter set while allowing reasonable variation along each parameter direction.

The posterior distribution was approximated using a sequential Monte Carlo particle approximation,

$$P^{(t)}(\boldsymbol{\theta}) \approx \sum_{n=1}^N w_n^{(t)} \delta(\boldsymbol{\theta} - \boldsymbol{\theta}_n^{(t)}), \quad (16)$$

where $N = 10,000$ particles were initialized by independent sampling from the prior in Eq. (13), with initial weights $w_n^{(0)} = 1/N$. At design iteration t , the selected experimental setting produces a set of measured coordinates and intensities $\mathcal{D}_t = \{(\mathbf{x}_j, y_j)\}_{j=1}^M$. Here $\mathbf{x}_j = (H_j, K_j, L_j, \omega_j)$ for single-crystal measurements.

For each particle, the surrogate forward model produced an intensity prediction $S_{\text{raw pred}}(\mathbf{Q}_j, \omega_j; \boldsymbol{\theta}_n^{(t)})$, which was scaled and clipped for the following particle-score calculation via

$$S_{\text{pred}}(\mathbf{Q}_j, \omega_j; \boldsymbol{\theta}_n^{(t)}) = \alpha_t \max\{S_{\text{raw pred}}(\mathbf{Q}_j, \omega_j; \boldsymbol{\theta}_n^{(t)}), 0\} \quad (17)$$

The global scale factor α_t was adaptively estimated from the ratio between the mean measured intensity and the mean particle-averaged predicted intensity, with exponential smoothing in log space. Particle weights were subsequently updated using a Poisson-derived robust score $\ell_n^{(t)}$,

$$w_n^{(t+1)} = \frac{w_n^{(t)} \exp(\ell_n^{(t)})}{\sum_{m=1}^N w_m^{(t)} \exp(\ell_m^{(t)})}. \quad (18)$$

Because $\ell_n^{(t)}$ is a robust effective score rather than a conventional log-likelihood, the resulting particle distribution is interpreted as a generalized posterior. More

details about the adaptive intensity scale factor α_t are provided in Supplementary Note 1, while the Poisson-derived robust score $\ell_n^{(t)}$ is described in Supplementary Note 2.

To monitor particle degeneracy, we used the standard effective sample size

$$N_{\text{eff}}^{(t)} = \frac{1}{\sum_{n=1}^N \left(w_n^{(t)}\right)^2}. \quad (19)$$

When $N_{\text{eff}}^{(t)} < N/2$, particles were rejuvenated using Liu–West resampling⁵⁷. Parent particles were sampled with replacement according to the current weights, and each child particle was shrunk toward the current weighted mean and perturbed by a Gaussian kernel,

$$\begin{aligned} \theta'_n &= a\theta_{I_n} + (1-a)\bar{\theta} + \epsilon_n, \\ \epsilon_n &\sim \mathcal{N}(\mathbf{0}, (1-a^2)\Sigma + 10^{-6}\mathbf{I}), \end{aligned} \quad (20)$$

where $I_n \sim \text{Categorical}(\{w_m\}_{m=1}^N)$, $\bar{\theta}$ and Σ are the weighted particle mean and covariance, and $a = 0.98$. After resampling, all weights were reset to $1/N$. Particles falling outside the prior bounds were replaced by new samples drawn uniformly within the allowed parameter ranges.

Because the masked powder measurements provide only weak constraints on several Hamiltonian parameters, directly propagating the entire powder particle set could result in overconfident initialization for the subsequent single-crystal inference. We therefore adopted a partial posterior initialization strategy. Specifically, particles were first resampled from the powder posterior according to their importance weights, after which the parameters A_x , J_{1a} , J_{1b} , J_{2a} , J_{2b} , and J_4 were reinitialized by independently drawing from their original uniform prior distributions, while the remaining parameters retained the powder-informed posterior values. The particle weights were subsequently reset to a uniform distribution before beginning the single-crystal inference.

Experimental settings were selected sequentially from the candidate setting space. For a candidate orientation ψ , we sampled coordinates $\mathbf{x}$ within the corresponding instrumental coverage region and drew parameter samples from the current particle approximation. The acquisition utility defined in Eq. (10) was multiplied by two scale factors to incorporate practical experimental considerations,

$$\tilde{U}^{(t+1)}(\psi) = s_{\text{rep}}^{(t+1)}(\psi) s_{\text{dist}}^{(t+1)}(\psi) U^{(t+1)}(\psi), \quad (21)$$

where s_{rep} penalizes repeated experimental settings and s_{dist} penalizes large changes relative to the previously selected single-crystal orientation. The specific forms of the two scale factors are provided in Supplementary Note 3. The next measurement was selected by maximizing the adjusted utility:

$$\psi_{t+1} = \arg \max_{\psi} \tilde{U}^{(t+1)}(\psi). \quad (22)$$

For the default Bayesian optimal experimental design (BOED) runs, utilities were estimated using 1000 sampled coordinates and 500 parameter samples at each design step.

Neural surrogate model

We used a multimodal neural surrogate to approximate the dynamical structure factor for both single-crystal and powder neutron-scattering geometries. The surrogate was based on a combined feature-wise linear modulation (FiLM) sinusoidal representation network (SIREN) architecture^{43,58}, in which the nine-dimensional Hamiltonian parameter vector $\theta = [A_x, A_z, J_{1a}, J_{1b}, J_{2a}, J_{2b}, J_{3a}, J_{3b}, J_4]$ was passed through a shared mapping network to generate layer-wise scale and shift modulations. The mapping network was a fully connected ReLU network with three hidden layers of 256 neurons, producing the FiLM parameters for the coordinate networks.

Separate coordinate networks were used for the single-crystal and powder modalities. The single-crystal network took (H, K, L, ω) as input, while the powder network took $(|\mathbf{Q}|, \omega)$ as input. Each coordinate network consisted of four FiLM-modulated sine layers with 256 hidden units and SIREN frequency parameter $\omega_0 = 30$, followed by a linear output layer producing a scalar intensity. The Hamiltonian-parameter mapping network was shared between modalities, while the coordinate networks were modality-specific.

The surrogate was trained jointly on the single-crystal and powder datasets, which together comprise 20,000 Hamiltonian samples (see “Training data generation” below). The dataset was randomly split into training, validation, and testing sets containing 90%, 5%, and 5% of the samples, respectively. We optimized a weighted mean-squared error loss with equal weights for the two modalities, using the Adam optimizer with a learning rate 10^{-4} and mini-batch size 4. Training was run for up to 300 epochs, and the model with the lowest validation loss (at epoch 239) was used for all BOED calculations. The detailed network architecture and training procedure are provided in Supplementary Note 4.

Experiment and data preparation

Polycrystalline NiPS₃ was prepared by heating a mixture of high-purity elements in a silica tube sealed under vacuum after argon purging. The reaction occurred by slowly heating (10 °C/h) to 400 °C followed by a 24 h dwell, then heating (10 °C/h) to 650 °C for 24 h. The furnace power was turned off and the ampoule cooled in the furnace. This produced a mixture of crystals and polycrystalline masses, which were ground in air and passed through a 0.5 mm sieve to reduce particle size. The resulting powder was annealed for 16 h at 600 °C

in an evacuated silica tube. The resulting material was lightly ground and passed through a 0.5 mm sieve before characterization by x-ray diffraction and magnetization measurements to verify the desired phase and expected Neel temperature. Prior to use, forming gas was used to reduce nickel powder at 450 °C for 12 h.

The powder sample of NiPS₃ was loaded under a helium atmosphere into aluminum sample cans. The sample along with an identical empty can were mounted to the cold-finger of the SEQUOIA instruments closed-cycle refrigerator sample changer⁵⁹. The third position of the sample changer had no sample installed for independent measurements of the sample environment background contribution. Inelastic neutron scattering measurements were performed at $T = 5$ K for incident energy $E_i = 210$ (6,3,3 C), 104.6 (15,12,2.9 C), 62.7 (12.3,9,1.5 C), and 29.3 (8.9,9,1.5 C) meV with the instruments high resolution Fermi chopper spinning at 600, 540, 420, and 240 Hz respectively. Values in parentheses correspond to accumulated proton charge in Coulombs on the spallation target for measurements of the sample, the empty can, and the empty sample environment respectively. Details about extraction of powder magnon spectrum from the total experimental INS data are presented in Supplementary Note 10.

Powder data. The powder INS data were reduced from the raw SEQUOIA event data using Mantid⁶⁰. For the powder spectrum used in Fig. 4(c), the nominal $E_i = 100$ meV dataset was used. Individual sample runs were merged prior to reduction, with the corresponding neutron-monitor data merged separately. The incident energy and time-zero correction were determined from the monitor data, and the scattering intensity was normalized to proton charge and corrected for detector efficiency using a processed vanadium standard. The reduced detector data were transformed to powder coordinates ($|\mathbf{Q}|, \hbar\omega$) using `ConvertToMD` and histogrammed using `BinMD` over $0 \leq |\mathbf{Q}| \leq 7 \text{ \AA}^{-1}$ and $0 \leq \hbar\omega \leq 80$ meV, with bin widths of $0.035 \text{ \AA}^{-1}$ and 0.4 meV, respectively. Measurements of the NiPS₃ sample together with the aluminum sample can and heat shield were reduced independently from corresponding empty-can measurements containing the can and heat shield, using the same reduction and histogramming procedure. Rather than directly subtracting the empty-can spectrum, the sample and empty-can spectra were retained as separate inputs to the vision-transformer-based source-separation workflow described in Supplementary Note 10. This workflow separates the magnetic scattering from sample-phonon and sample-environment contributions, after which the extracted magnon spectrum was processed by the feature-enhancement model and used for Hamiltonian inference.

Single-crystal data. The experimental spectrum along the high-symmetry path shown in Fig. 4(c) was generated from the original Mantid MDEventWorkspace (MDE) files⁶⁰ used to produce the figures in Ref. 3. The dataset comprises three merged MDE files containing complete sample-rotation measurements collected at

incident energies $E_i = 28, 60$, and 100 meV. For each incident energy, the data were histogrammed along the high-symmetry path $C - \Gamma, \Gamma - Y, Y - C, C - Z$, and $Z - \Gamma$ using the MDNorm algorithm as implemented in SHIVER^{61,62}. The histogram specifications followed Ref. 3, with integration widths of ± 0.05 reciprocal lattice units (r.l.u.) in the in-plane direction transverse to the path and ± 0.25 r.l.u. in the out-of-plane direction. An energy-bin width of 0.5 meV was used over $0 \leq \hbar\omega \leq 60$ meV. The individual path segments were concatenated for each incident energy and then combined into a single spectrum, using the $E_i = 28$ meV data for $0-20$ meV, the $E_i = 60$ meV data for $20-32$ meV, and the $E_i = 100$ meV data above 32 meV. The effect of the finite reciprocal-space integration used to construct these cuts on the Hamiltonian inference is examined further in Supplementary Note 9.

Training data generation

Training data were generated using linear spin wave calculations implemented in `Sunny.jl`²³. We randomly sampled 20,000 Hamiltonians by independently varying each Hamiltonian parameter over a range between zero and twice its value reported in Ref. 3, while preserving its sign; i.e., $\theta_i \sim \mathcal{U}(0, 2\theta_{\text{ref},i})$ for $\theta_{\text{ref},i} > 0$ and $\theta_i \sim \mathcal{U}(2\theta_{\text{ref},i}, 0)$ for $\theta_{\text{ref},i} < 0$. For each parameter set, the magnetic ground state was first obtained through classical energy minimization before constructing the corresponding spin wave model. The magnetic ground state search was always conducted on a single crystallographic unit cell. Parameter sets that failed to converge or yielded unstable magnon spectra were discarded and replaced with new samples.

Single-crystal training data were generated by evaluating the dynamical structure factor at 2,000 randomly sampled momentum-transfer points uniformly distributed within $[-3, 3]^3$ reciprocal lattice units (r.l.u.) and 100 uniformly spaced energy transfers between 0 and 100 meV. Powder spectra were generated using orientational powder averaging with 300 randomly sampled $|\mathbf{Q}|$ values and 3,000 crystal orientations. Both single-crystal and powder spectra employed Gaussian energy broadening with a full width at half maximum (FWHM) of 4 meV. Three rotational magnetic domains were averaged for the single-crystal calculations, and the Ni²⁺ magnetic form factor was included throughout. The Hamiltonian parameters and scattering coordinates were linearly normalized to the interval $[-1, 1]$ before training.

DATA AVAILABILITY

The experimental data supporting the findings of this study are available from the corresponding author upon reasonable request. The simulation datasets can be re-

generated using the code provided and will be made publicly available upon acceptance of the manuscript.

CODE AVAILABILITY

The source code required to reproduce the results presented in this work will be made publicly available through a GitHub repository with a permanent Zenodo archive upon acceptance of the manuscript.

ACKNOWLEDGMENTS

This work is primarily supported by the U.S. Department of Energy, Office of Science, Basic Energy Sciences (BES) under the Genesis Mission BES AI Pathfinder Program, MAIQMag: Multimodal AI for 2D Quantum Magnets. Sample synthesis and verification (A.F.M.) were supported by the U. S. Department of Energy, Office of Science, Basic Energy Sciences, Materials Sci-

ences and Engineering Division. This research used computational resources of the National Energy Research Scientific Computing Center (NERSC), a US Department of Energy Office of Science User Facility located at Lawrence Berkeley National Laboratory operated under contract DE-AC02-05CH11231, using NERSC award ERCAP0038195. The authors acknowledge the Texas Advanced Computing Center (TACC) at The University of Texas at Austin for providing computational resources under project code DMR25014 that have contributed to the research results reported within this paper. We thank Allen Scheie for providing the single-crystal INS data originally published in Ref. 3. A portion of this research used resources at the Spallation Neutron Source, a DOE Office of Science User Facility operated by the Oak Ridge National Laboratory. The beam time was allocated to SEQUOIA on proposal number IPTS-37598. During the preparation of this work, the authors used the large language model ChatGPT by OpenAI in order to refine the language and enhance the readability of this paper. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for all content in this publication.

-
- [1] N. Ntallis, V. Borisov, Y. O. Kvashnin, D. Thonig, E. Sjöqvist, A. Bergman, A. Delin, O. Eriksson, and M. Pereiro, Connection between magnetic interactions and the spin-wave gap of the insulating phase of NaOsO_3 , *Physical Review B* **104**, 134433 (2021).
 - [2] A. Petsch, N. Headings, D. Prabhakaran, A. Kolesnikov, C. Frost, A. Boothroyd, R. Coldea, and S. Hayden, High-energy spin waves in the spin-1 square-lattice antiferromagnet La_2NiO_4 , *Physical Review Research* **5**, 033113 (2023).
 - [3] A. Scheie, P. Park, J. W. Villanova, *et al.*, Spin wave hamiltonian and anomalous scattering in NiPS_3 , *Phys. Rev. B* **108**, 104402 (2023).
 - [4] M. Hase, R. Tamura, K. Hukushima, S. Asai, T. Masuda, S. Itoh, and A. Dönni, Inelastic neutron scattering studies on the eight-spin zigzag-chain compound $\text{KCu}_4\text{P}_3\text{O}_{12}$: Confirmation of the validity of a data-driven technique based on machine learning, *Physical Review B* **109**, 094434 (2024).
 - [5] R. N. Gutenkunst, J. J. Waterfall, F. P. Casey, K. S. Brown, C. R. Myers, and J. P. Sethna, Universally sloppy parameter sensitivities in systems biology models, *PLoS computational biology* **3**, e189 (2007).
 - [6] M. K. Transtrum, B. B. Machta, K. S. Brown, B. C. Daniels, C. R. Myers, and J. P. Sethna, Perspective: Sloppiness and emergent theories in physics, biology, and beyond, *The Journal of chemical physics* **143** (2015).
 - [7] S.-i. Amari, *Information geometry and its applications* (Springer, 2016).
 - [8] A. M. Samarakoon, K. Barros, Y. W. Li, M. Eisenbach, Q. Zhang, F. Ye, V. Sharma, Z. Dun, H. Zhou, S. A. Grigera, *et al.*, Machine-learning-assisted insight into spin ice $\text{Dy}_2\text{Ti}_2\text{O}_7$, *Nature communications* **11**, 892 (2020).
 - [9] Z. Chen, X. Shen, N. Andrejevic, T. Liu, D. Luo, T. Nguyen, N. C. Drucker, M. E. Kozina, Q. Song, C. Hua, *et al.*, Panoramic mapping of phonon transport from ultrafast electron diffraction and scientific machine learning, *Advanced Materials* **35**, 2206997 (2023).
 - [10] A. M. Samarakoon and D. Alan Tennant, Machine learning for magnetic phase diagrams and inverse scattering problems, *Journal of Physics: Condensed Matter* **34**, 044002 (2022).
 - [11] T. Misawa, R. Tamura, K. Yoshimi, and Y. Yamaji, Revisiting spin hamiltonian parameters in a kitaev material via bayesian optimization of magnetization curves, *arXiv preprint arXiv:2605.24857* (2026).
 - [12] A. Tarantola, *Inverse problem theory and methods for model parameter estimation* (SIAM, 2005).
 - [13] A. M. Stuart, Inverse problems: a bayesian perspective, *Acta numerica* **19**, 451 (2010).
 - [14] G. L. Squires, *Introduction to the theory of thermal neutron scattering* (Courier Corporation, 1996).
 - [15] S. W. Lovesey, *Theory of Neutron Scattering from Condensed Matter, Volume 2: Polarization Effects and Magnetic Scattering* (Clarendon Press, Oxford, 1984).
 - [16] J. R. Copley and T. J. Udovic, Neutron time-of-flight spectroscopy, *Journal of research of the National Institute of Standards and Technology* **98**, 71 (1993).
 - [17] A. Ivanov and P. Alekseev, Neutron spectroscopy: Principles and equipment, *Crystallography Reports* **67**, 18 (2022).
 - [18] T. Chatterji, *Neutron scattering from magnetic materials* (Elsevier, 2005).
 - [19] J. Jaklič and P. Prelovšek, Lanczos method for the calculation of finite-temperature quantities in correlated systems, *Physical Review B* **49**, 5065 (1994).

- [20] U. Schollwöck, The density-matrix renormalization group, *Rev. Mod. Phys.* **77**, 259 (2005).
- [21] S. Toth and B. Lake, Linear spin wave theory for single-q incommensurate magnetic structures, *Journal of Physics: Condensed Matter* **27**, 166002 (2015).
- [22] D. Blosser, N. Kestlin, K. Y. Povarov, R. Bewley, E. Coira, T. Giamarchi, and A. Zheludev, Finite-temperature correlations in a quantum spin chain near saturation, *Physical Review B* **96**, 134406 (2017).
- [23] D. Dahlbom, H. Zhang, C. Miles, S. Quinn, A. Niraula, B. Thipe, M. Wilson, S. Matin, H. Mankad, S. Hahn, *et al.*, Sunny.jl: A Julia package for spin dynamics, *arXiv preprint arXiv:2501.13095* (2025).
- [24] G. Carleo and M. Troyer, Solving the quantum many-body problem with artificial neural networks, *Science* **355**, 602 (2017).
- [25] T. Mendes-Santos, M. Schmitt, and M. Heyl, Highly resolved spectral functions of two-dimensional systems with neural quantum states, *Physical Review Letters* **131**, 046501 (2023).
- [26] S. Chillal, Y. Iqbal, H. O. Jeschke, J. A. Rodriguez-Rivera, R. Bewley, P. Manuel, D. Khalyavin, P. Steffens, R. Thomale, A. N. Islam, *et al.*, Evidence for a three-dimensional quantum spin liquid in $\text{PbCuTe}_2\text{O}_6$, *Nature communications* **11**, 2348 (2020).
- [27] X. Huang, H. Miao, H. Kim, A. Townsend, K. Champley, J. Tringe, V. Pascucci, and P.-T. Bremer, Bimodal visualization of industrial x-ray and neutron computed tomography data, *IEEE Transactions on Visualization and Computer Graphics* **31**, 2196 (2024).
- [28] P. Vestin, E. Schlautmann, O. Sans-Planell, N. Kardjilov, R. Woracek, A. Tengattini, W. G. Zeier, and S. Hall, 4D multimodal neutron and x-ray tomography of lithium transport in all-solid-state batteries using Li-7 contrast enhancement, *Cell Reports Physical Science* **7** (2026).
- [29] J.-G. Park, K.-X. Zhang, H. Cheong, J. H. Kim, C. A. Belvin, D. Hsieh, H. Ning, and N. Gedik, 2D van der waals magnets: from fundamental physics to applications, *Reviews of Modern Physics* **98**, 025003 (2026).
- [30] I. A. Moses, C. Chen, J. M. Redwing, and W. F. Reinhardt, Cross-modal characterization of thin-film MoS_2 using generative models, *Advanced Intelligent Systems* **8**, 2500613 (2026).
- [31] S. M. Avdoshenko, A. A. Kulbakov, E. Häußler, P. Schlender, T. Doert, J. Ollivier, and D. S. Inosov, Spin-wave dynamics in the KCeS_2 delafossite: A theoretical description of powder inelastic neutron-scattering data, *Physical Review B* **106**, 214431 (2022).
- [32] Y. Su and C. Li, Uncovering obscured phonon dynamics from powder inelastic neutron scattering using machine learning, *Machine Learning: Science and Technology* **5**, 035080 (2024).
- [33] Y. Cheng, M. B. Stone, and A. J. Ramirez-Cuesta, A database of synthetic inelastic neutron scattering spectra from molecules and crystals, *Scientific Data* **10**, 54 (2023).
- [34] Z. Chen, N. Andrejevic, N. C. Drucker, T. Nguyen, R. P. Xian, T. Smidt, Y. Wang, R. Ernstorfer, D. A. Tennant, M. Chan, *et al.*, Machine learning on neutron and x-ray scattering and spectroscopies, *Chemical Physics Reviews* **2** (2021).
- [35] R. Plumley, S. Chitturi, C. Peng, T. Assefa, N. Burdet, L. Shen, Z. Chen, A. Reid, G. Dakovski, M. Seaberg, *et al.*, On ultrafast x-ray scattering methods for magnetism, *Advances in Physics: X* **9**, 2423935 (2024).
- [36] J. P. Horwath, X.-M. Lin, H. He, Q. Zhang, E. M. Dufresne, M. Chu, S. K. Sankaranarayanan, W. Chen, S. Narayanan, and M. J. Cherukara, Ai-nerd: Elucidation of relaxation dynamics beyond equilibrium through ai-informed x-ray photon correlation spectroscopy, *Nature Communications* **15**, 5945 (2024).
- [37] Q. Li, R. Jiao, L. Wu, T. Zhu, W. Huang, S. Jin, Y. Liu, H. Weng, and X. Chen, Powder diffraction crystal structure determination using generative models, *Nature Communications* **16**, 7428 (2025).
- [38] K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev, and A. Walsh, Machine learning for molecular and materials science, *Nature* **559**, 547 (2018).
- [39] G. Papamakarios and I. Murray, Fast ε -free inference of simulation models with bayesian conditional density estimation, *Advances in neural information processing systems* **29** (2016).
- [40] J.-M. Lueckmann, J. Boelts, D. Greenberg, P. Goncalves, and J. Macke, Benchmarking simulation-based inference, in *International conference on artificial intelligence and statistics* (PMLR, 2021) pp. 343–351.
- [41] T. Baltrusaitis, C. Ahuja, and L.-P. Morency, Multimodal machine learning: A survey and taxonomy, *IEEE transactions on pattern analysis and machine intelligence* **41**, 423 (2019).
- [42] P. P. Liang, A. Zadeh, and L.-P. Morency, Foundations & trends in multimodal machine learning: Principles, challenges, and open questions, *ACM computing surveys* **56**, 1 (2024).
- [43] V. Sitzmann, J. Martel, A. Bergman, D. Lindell, and G. Wetzstein, Implicit neural representations with periodic activation functions, *Advances in neural information processing systems* **33**, 7462 (2020).
- [44] S. Chitturi, Z. Ji, A. Petsch, C. Peng, Z. Chen, R. Plumley, M. Dunne, S. Mardanya, S. Chowdhury, H. Chen, *et al.*, Capturing dynamical correlations using implicit neural representations, *arXiv preprint arXiv:2304.03949* (2023).
- [45] Z. Chen, A. N. Petsch, Z. Ji, S. R. Chitturi, C. Peng, C. Jia, A. I. Kolesnikov, J. B. Thayer, and J. J. Turner, Implicit neural representations for experimental steering of advanced experiments, *Cell Reports Physical Science* **6** (2025).
- [46] K. Chaloner and I. Verdinelli, Bayesian experimental design: A review, *Statistical science* , 273 (1995).
- [47] C. E. Granade, C. Ferrie, N. Wiebe, and D. G. Cory, Robust online hamiltonian learning, *New Journal of Physics* **14**, 103013 (2012).
- [48] X. Huan and Y. M. Marzouk, Simulation-based optimal bayesian experimental design for nonlinear systems, *Journal of Computational Physics* **232**, 288 (2013).
- [49] R. D. McMichael and S. M. Blakley, Simplified algorithms for adaptive experiment design in parameter estimation, *Physical review applied* **18**, 10 (2022).
- [50] X. Huan, J. Jagalur, and Y. Marzouk, Optimal experimental design: Formulations and computations, *Acta Numerica* **33**, 715 (2024).
- [51] Z. Chen, C. Peng, A. N. Petsch, S. R. Chitturi, A. Okullo, S. Chowdhury, C. H. Yoon, and J. J. Turner, Bayesian experimental design and parameter estimation for ultrafast spin dynamics, *Machine Learning: Science and Technology* **4**, 045056 (2023).

- [52] D. A. Boiko, R. MacKnight, B. Kline, and G. Gomes, Autonomous chemical research with large language models, *Nature* **624**, 570 (2023).
- [53] Z. Xiao, G. Zhang, Z. Morgan, V. Reshniak, and X. Wang, CrystalPilot: A machine-learning based software platform for single crystal neutron diffraction experiments, *Structural Dynamics* **12**, A125 (2025).
- [54] T. Hellert, D. Bertwistle, S. C. Leemann, A. Sulc, and M. Venturini, Agentic artificial intelligence for multistage physics experiments at a large-scale user facility particle accelerator, *Physical Review Research* **8**, L012017 (2026).
- [55] Z. Chen, A. N. Petsch, A. J. Israelski, R. Plumley, L. Shen, C. Wang, C. Peng, Y. Ni, A. Bansil, S. Chowdhury, *et al.*, An agentic artificially intelligent x-ray scientist, *Nature Machine Intelligence*, 1 (2026).
- [56] S. Qiu, P. D. Suh, N. H. Tran, X. Wang, H. Park, K. Akin, K. K. Multani, S. Jung, W. Cai, X. Liu, *et al.*, AIMS: An uncertainty-aware ai experimentalist for quantum matter, arXiv preprint arXiv:2607.16544 (2026).
- [57] J. Liu and M. West, Combined parameter and state estimation in simulation-based filtering, in *Sequential Monte Carlo methods in practice* (Springer, 2001) pp. 197–223.
- [58] E. Perez, F. Strub, H. De Vries, V. Dumoulin, and A. Courville, FiLM: Visual reasoning with a general conditioning layer, in *Proceedings of the AAAI conference on artificial intelligence*, Vol. 32 (2018).
- [59] M. B. Stone, G. E. Granroth, D. M. Pajerowski, D. L. Abernathy, D. L. Conner, L. DeBeer-Schmitt, V. R. Fanelli, R. Goyette, A. I. Kolesnikov, R. Mills, M. Odom, A. Podlesnyak, C. Schmitt, T. E. Sherline, L. Solomon, and J. F. Wenzel, Sample changers for direct geometry neutron chopper spectrometers, *Scientific Reports* **15**, 31936 (2025).
- [60] O. Arnold, J. C. Bilheux, J. M. Borreguero, A. Buts, S. I. Campbell, L. Chapon, *et al.*, Mantid—data analysis and visualization package for neutron scattering and μ SR experiments, *Nuclear Instruments and Methods in Physics Research Section A* **764**, 156 (2014).
- [61] A. T. Savici, M. A. Gigg, O. Arnold, R. Tolchenov, R. E. Whitfield, S. E. Hahn, W. Zhou, and I. A. Zaliznyak, Efficient data reduction for time-of-flight neutron scattering experiments on single crystals, *Journal of Applied Crystallography* **55**, 1514 (2022).
- [62] A. T. Savici, SHIVER: Spectroscopy histogram visualizer for event reduction (2025), Oak Ridge National Laboratory.

Supplementary Note 1. Global scale factor

For each particle, the surrogate forward model produced an intensity prediction $S_{\text{pred}}(\mathbf{Q}_j, \omega_j; \boldsymbol{\theta}_n^{(t)})$. Since the surrogate intensity scale need not match the measured count scale, we introduced an adaptive global scale factor, where M is the number of measured data points,

$$\hat{\alpha}_t = \frac{M^{-1} \sum_{j=1}^M \max(S_{\text{targ}}(\mathbf{Q}_j, \omega_j), 0)}{M^{-1} \sum_{j=1}^M N^{-1} \sum_{n=1}^N \max\{S_{\text{raw pred}}(\mathbf{Q}_j, \omega_j; \boldsymbol{\theta}_n^{(t)}), 0\} + \varepsilon}, \quad \varepsilon = 10^{-12}. \quad (\text{S1})$$

The scale used in the particle score, α_t , was obtained by smoothing $\hat{\alpha}_t$ with an exponential moving average in log space,

$$\log \alpha_t = \rho \log \alpha_{t-1} + (1 - \rho) \log \hat{\alpha}_t, \quad (\text{S2})$$

with $\rho = 0.95$, except for the first update where $\alpha_t = \hat{\alpha}_t$. The scaled surrogate prediction becomes

$$S_{\text{pred}}(\mathbf{Q}_j, \omega_j; \boldsymbol{\theta}_n^{(t)}) = \alpha_t \max \left[S_{\text{raw pred}}(\mathbf{Q}_j, \omega_j; \boldsymbol{\theta}_n^{(t)}), 0 \right]. \quad (\text{S3})$$

Supplementary Note 2. Poisson-derived robust score

We use a Poisson-derived robust particle score to define a generalized Bayesian update rather than directly evaluating the raw Poisson likelihood, which was found to be numerically too sharp for the sparse, high-intensity spectra considered here. We first computed

$$r_{nj}^{(t)} = \log \text{Pois}\left(S_{\text{targ}}(\mathbf{Q}_j, \omega_j); S_{\text{pred}}(\mathbf{Q}_j, \omega_j; \boldsymbol{\theta}_n^{(t)})\right). \quad (\text{S4})$$

For numerical stabilization, these scores were normalized over the measured data points for each particle,

$$p_{nj}^{(t)} = \frac{\exp(r_{nj}^{(t)})}{\sum_{k=1}^M \exp(r_{nk}^{(t)})}, \quad (\text{S5})$$

and then rescaled using min–max normalization across particles at each measured point,

$$\tilde{p}_{nj}^{(t)} = \frac{p_{nj}^{(t)} - \min_m p_{mj}^{(t)}}{\max_m p_{mj}^{(t)} - \min_m p_{mj}^{(t)} + \varepsilon}, \quad \varepsilon = 10^{-12}. \quad (\text{S6})$$

The score used for the particle update was

$$\ell_n^{(t)} = \frac{1}{M} \sum_{j=1}^M \log \tilde{p}_{nj}^{(t)}. \quad (\text{S7})$$

The resulting score is not a conventional log-likelihood because of the normalization and rescaling operations above. We therefore use it as an effective log-score in a generalized Bayesian particle update,

$$w_n^{(t+1)} = \frac{w_n^{(t)} \exp(\ell_n^{(t)})}{\sum_{m=1}^N w_m^{(t)} \exp(\ell_m^{(t)})}. \quad (\text{S8})$$

Accordingly, the resulting particle distribution should be interpreted as an effective (generalized) posterior induced by this robust score rather than as the posterior associated with the raw Poisson likelihood.

Supplementary Note 3. History-dependent setting penalties

The BOED acquisition utility was multiplied by two history-dependent penalty factors,

$$\tilde{U}^{(t+1)}(\psi) = s_{\text{rep}}^{(t+1)}(\psi) s_{\text{dist}}^{(t+1)}(\psi) U^{(t+1)}(\psi), \quad (\text{S9})$$

where s_{rep} discourages repeated measurements at the same setting and s_{dist} discourages abrupt jumps in the single-crystal rotation angle. These factors were introduced as pragmatic scheduling regularizers: they do not modify the posterior update, but only bias the selection of the next experimental setting among candidates with similar acquisition utilities.

Let $\mathcal{H}^{(t)} = \{\psi^{(1)}, \dots, \psi^{(t)}\}$ denote the history of selected angles before step $t + 1$. The repetition penalty was defined as

$$s_{\text{rep}}^{(t+1)}(\psi) = \exp \left[-\gamma_{\text{rep}} C^{(t)}(\psi) \right], \quad (\text{S10})$$

where

$$C^{(t)}(\psi) = \sum_{k=1}^t \delta(\psi - \psi^{(k)}) \quad (\text{S11})$$

counts the number of previous selections of angle ψ , within a small numerical tolerance ϵ_ψ . In the calculations reported here, $\gamma_{\text{rep}} = 0.5$. Thus, each repeated use of the same angle reduces its acquisition score by a factor of $\exp(-0.5)$.

For single-crystal BOED, we also applied a distance penalty relative to the most recently selected angle $\psi^{(t)}$,

$$s_{\text{dist}}^{(t+1)}(\psi) = \exp \left[-\gamma_{\text{dist}} \left(\frac{d_{\text{per}}(\psi, \psi^{(t)})}{\ell_\psi} \right)^2 \right]. \quad (\text{S12})$$

Here d_{per} is the periodic angular distance,

$$d_{\text{per}}(\psi, \psi^{(t)}) = \min \left(|\psi - \psi^{(t)}|, 360^\circ - |\psi - \psi^{(t)}| \right), \quad (\text{S13})$$

to account for the 360° periodicity of the sample rotation. We used $\gamma_{\text{dist}} = 0.1$ and $\ell_\psi = 15^\circ$. At the first design step, before any setting has been selected, $s_{\text{dist}} = 1$ for all candidate angles.

The final multiplicative penalty factor was clipped from below,

$$s_{\text{rep}}^{(t+1)}(\psi) s_{\text{dist}}^{(t+1)}(\psi) \leftarrow \max \left[s_{\text{rep}}^{(t+1)}(\psi) s_{\text{dist}}^{(t+1)}(\psi), 10^{-6} \right], \quad (\text{S14})$$

to avoid assigning exactly zero acquisition score to any candidate setting. The next setting was then chosen by maximizing the penalized acquisition utility,

$$\psi^{t+1} = \arg \max_{\psi} \tilde{U}^{(t+1)}(\psi). \quad (\text{S15})$$

Supplementary Note 4. AI surrogate architecture and training

The multimodal surrogate consists of a shared Hamiltonian-parameter mapping network and separate coordinate networks for the single-crystal and powder modalities. The mapping network takes the nine-dimensional Hamiltonian parameter vector θ as input to produce the FiLM modulation parameters used by the modality-specific coordinate networks. The single-crystal branch takes four-dimensional coordinates (H, K, L, ω) as input, whereas the powder branch takes two-dimensional coordinates $(|\mathbf{Q}|, \omega)$ as input. Both branches use four modulated sine layers with 256 hidden units, followed by a linear scalar-output layer. The complete model architecture, as reported by the implementation, is shown below.

```
MultiModalModulatedSiren(
  (coordinate_networks): ModuleDict(
    (single_crystal): Sequential(
      (0): ModulatedSineLayer(
        (linear): Linear(in_features=4, out_features=256, bias=True)
      )
      (1): ModulatedSineLayer(
        (linear): Linear(in_features=256, out_features=256, bias=True)
      )
      (2): ModulatedSineLayer(
        (linear): Linear(in_features=256, out_features=256, bias=True)
      )
      (3): ModulatedSineLayer(
        (linear): Linear(in_features=256, out_features=256, bias=True)
      )
      (4): Linear(in_features=256, out_features=1, bias=True)
    )
    (powder): Sequential(
      (0): ModulatedSineLayer(
        (linear): Linear(in_features=2, out_features=256, bias=True)
      )
      (1): ModulatedSineLayer(
        (linear): Linear(in_features=256, out_features=256, bias=True)
      )
      (2): ModulatedSineLayer(
        (linear): Linear(in_features=256, out_features=256, bias=True)
      )
      (3): ModulatedSineLayer(
        (linear): Linear(in_features=256, out_features=256, bias=True)
      )
      (4): Linear(in_features=256, out_features=1, bias=True)
    )
  )
  (mapping_net): MappingNetwork(
    (net): Sequential(
      (0): Linear(in_features=9, out_features=256, bias=True)
      (1): ReLU(inplace=True)
      (2): Linear(in_features=256, out_features=256, bias=True)
      (3): ReLU(inplace=True)
      (4): Linear(in_features=256, out_features=256, bias=True)
      (5): ReLU(inplace=True)
      (6): Linear(in_features=256, out_features=2048, bias=True)
    )
  )
)
```

The surrogate was trained for 300 epochs. Figure S1 shows the training and validation loss histories on a logarithmic scale. Both losses decrease rapidly during the initial stage of training and continue to improve more gradually at later epochs. The validation loss exhibits larger epoch-to-epoch fluctuations than the training loss, as expected, but does not show sustained divergence from the training curve. The checkpoint with the lowest validation loss was obtained at epoch 239 and was used in the subsequent inference and experimental-design calculations.

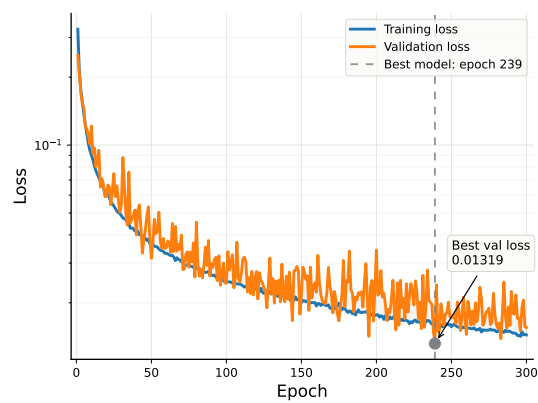


FIG. S1. **Training and validation loss histories of the multimodal neural surrogate.** Training and validation losses are shown as functions of training epoch on a logarithmic scale. The model was trained for 300 epochs. The dashed vertical line marks the selected checkpoint at epoch 239, which achieved the minimum validation loss of 0.01319. This checkpoint was used for the subsequent Hamiltonian inference and Bayesian experimental-design calculations.

Supplementary Note 5. Hamiltonian parameter values

Table S1 lists the Hamiltonian parameters of the literature reference model, $\boldsymbol{\theta}_{\text{ref}}$, and the selected posterior sample, $\boldsymbol{\theta}_{\text{best}}$, reported in the main text.

TABLE S1. Reference and inferred Hamiltonian parameters (meV). The reference Hamiltonian is taken from Ref. 3, while $\boldsymbol{\theta}_{\text{best}}$ denotes the posterior sample selected according to Eq. (12).

	A_x	A_z	J_{1a}	J_{1b}	J_{2a}	J_{2b}	J_{3a}	J_{3b}	J_4
$\boldsymbol{\theta}_{\text{ref}}$	-0.005	0.21	-2.70	-2.00	0.20	0.20	13.90	13.90	-0.38
$\boldsymbol{\theta}_{\text{best}}$	-0.00595	0.20	-2.87	-2.13	0.19	0.21	13.61	13.52	-0.39

Supplementary Note 6. Convergence rate comparison: sequential, random, and BOED

We compare the convergence behavior of three data-acquisition strategies: sequential acquisition, random acquisition, and BOED. For each strategy, we track the mean absolute error (MAE) of the inferred Hamiltonian parameters as a function of the acquisition step. The benchmark includes ten trials in total: nine Hamiltonian parameter sets randomly sampled from the prior distribution $P_0(\theta)$, together with the reference parameter set reported in Ref. 3.

As shown in Fig. S2, BOED substantially accelerates convergence compared to the sequential strategy, reaching a low-error regime within the first few measurements. In contrast, the sequential strategy decreases the error more gradually, indicating that a fixed acquisition order can spend measurements on configurations that provide limited additional information.

Random acquisition achieves a mean performance that is relatively close to BOED in this benchmark, but with noticeably greater variability across trials. This behavior is expected because random acquisition may occasionally sample informative configurations, but does not consistently target the measurements that are most informative under the current posterior uncertainty. We also note that fully random acquisition is rarely used as a practical experimental protocol, where measurement choices are typically constrained by time, prior knowledge, and instrument operation. BOED therefore provides a more systematic and reproducible strategy for rapidly reducing Hamiltonian uncertainty under a finite measurement budget.

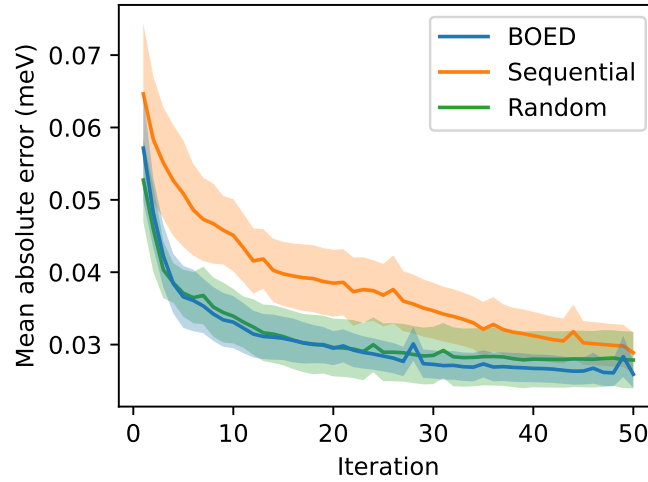


FIG. S2. Comparison of parameter-estimation error histories for sequential, random, and BOED acquisition strategies. Solid curves show the mean absolute error (MAE) averaged over ten benchmark trials, consisting of nine randomly sampled Hamiltonian parameter sets from the prior $P_0(\theta)$ and one reference parameter set reported in Ref. 3. Shaded regions indicate ± 1 standard deviation across trials. BOED converges substantially faster than the sequential strategy and achieves comparable or lower final error with reduced variability relative to random acquisition.

Supplementary Note 7. High-dimensional eigen-structure

To further illustrate the geometric interpretation of the metric tensor, Fig. S3 shows pairwise spectral discrepancy landscapes for the single-crystal modality, overlaid with the projections of the global stiffest (red, largest eigenvalue) and sloppiest (yellow, smallest eigenvalue) eigenvectors obtained from the full nine-dimensional metric tensor. Similar results for the powder modality are presented in Fig. S5.

Several observations are worth noting. First, the projected eigenvectors should be interpreted with caution because each panel represents only a two-dimensional projection of a nine-dimensional eigen-direction. Furthermore, the parameters span substantially different numerical ranges, producing unequal axis scales that can visually distort the projected directions. Nevertheless, the projected stiffest direction remains approximately perpendicular to the elongated discrepancy valleys for the most sensitive parameter pairs, such as the J_{1a} – J_{1b} and J_{3a} – J_{3b} subspaces, consistent with these parameter pairs contributing strongly to the largest-curvature directions of the global metric.

To directly validate the eigen-decomposition, Fig. S4 compares the spectral discrepancy along the global nine-dimensional stiff and sloppy eigenvectors with the corresponding principal directions obtained from the J_{1a} – J_{1b} two-dimensional subspace. The discrepancy increases most rapidly along the global stiff direction and most slowly along the global sloppy direction, while the two-dimensional principal directions exhibit intermediate behavior. This demonstrates that the full nine-dimensional eigen-directions provide upper and lower bounds on the local sensitivity observed within individual parameter subspaces. A similar trend is observed for the powder modality (Fig. S6), which is expected because powder INS provides an orientationally averaged measurement of the single-crystal response.

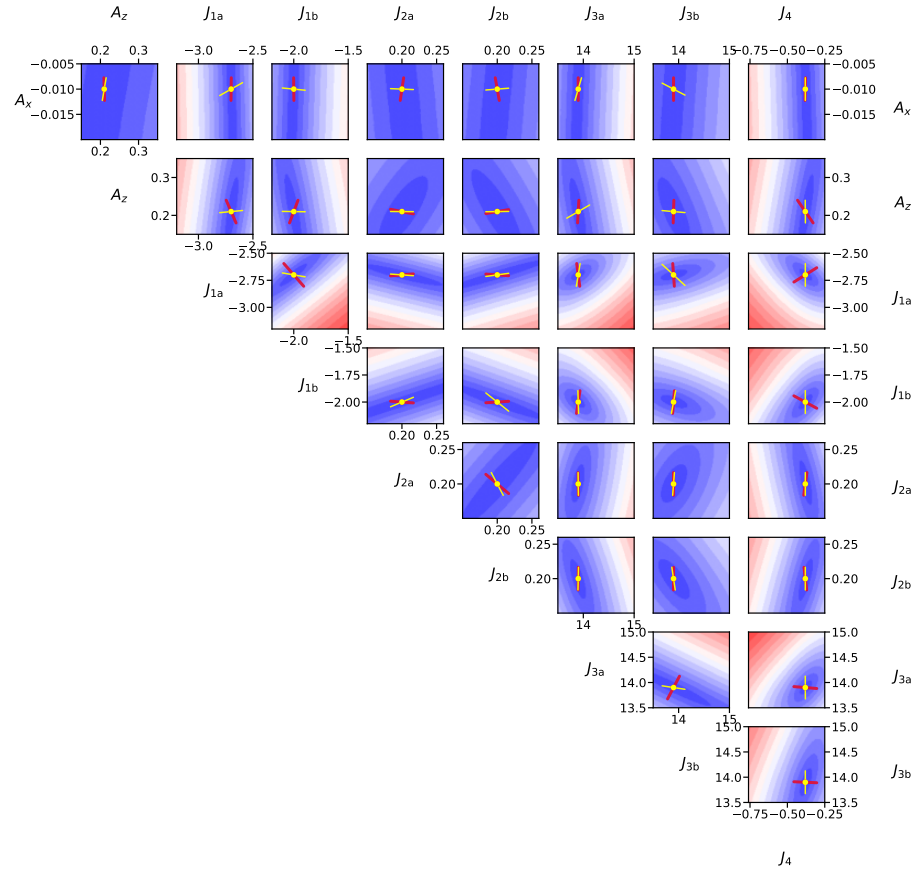


FIG. S3. Pairwise spectral discrepancy landscapes for the single-crystal modality with projections of the global nine-dimensional metric tensor eigen-directions. Red and yellow line segments denote the projected stiffest (largest eigenvalue) and sloppiest (smallest eigenvalue) eigenvectors, respectively. Since each panel represents a two-dimensional projection of the full nine-dimensional parameter space, the projected eigen-directions are not generally expected to align with the principal axes of the corresponding two-dimensional discrepancy landscapes.

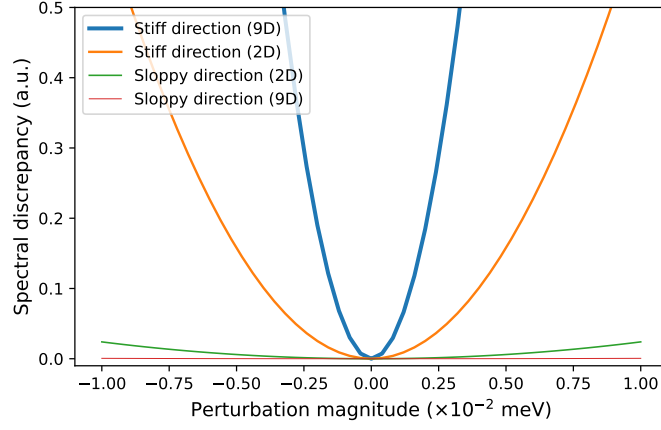


FIG. S4. Comparison of spectral discrepancy along the global nine-dimensional and local two-dimensional principal directions for the single-crystal modality. The blue and red curves correspond to perturbations along the stiffest and sloppiest eigenvectors of the full nine-dimensional metric tensor, respectively, while the orange and green curves correspond to the principal directions obtained from the J_{1a} – J_{1b} two-dimensional subspace. The global stiff and sloppy directions exhibit the fastest and slowest discrepancy growth, respectively, demonstrating that the full metric tensor eigen-directions bound the local sensitivity observed within lower-dimensional parameter subspaces.

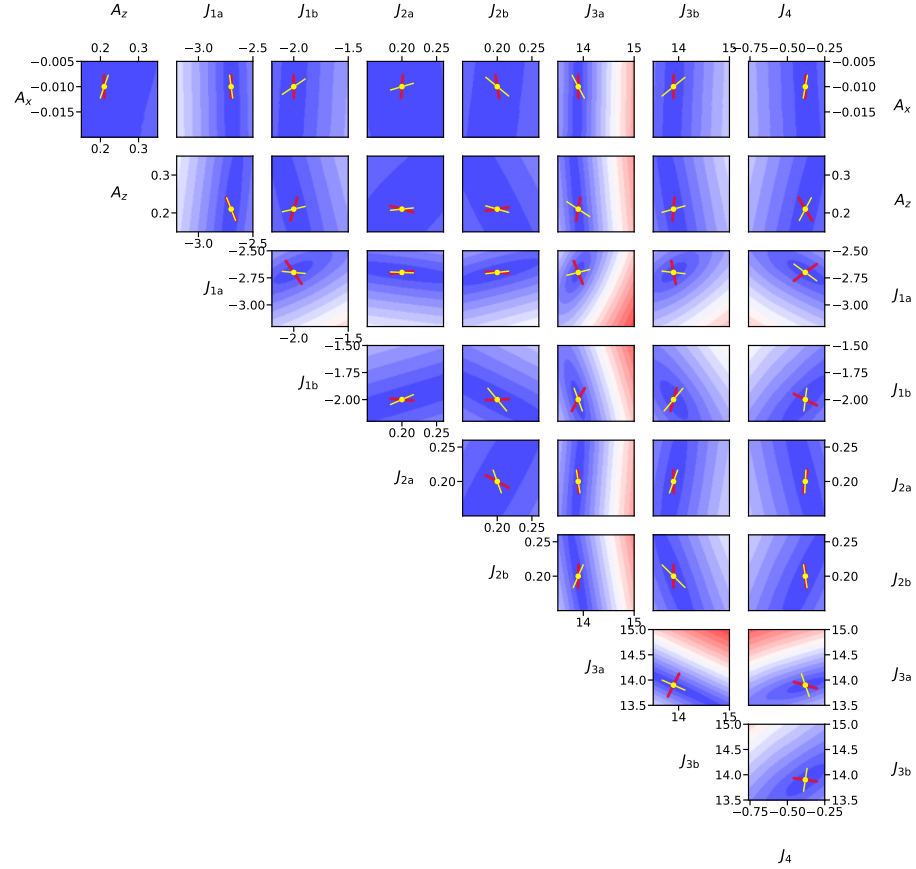


FIG. S5. Pairwise spectral discrepancy landscapes for the powder modality with projections of the global nine-dimensional metric tensor eigen-directions. Red and yellow line segments denote the projected stiffest (largest eigenvalue) and sloppiest (smallest eigenvalue) eigenvectors, respectively. Similar to the single-crystal case, the projected global eigen-directions are not generally expected to align with the principal axes of the two-dimensional discrepancy landscapes because they represent projections of directions defined in the full nine-dimensional parameter space.

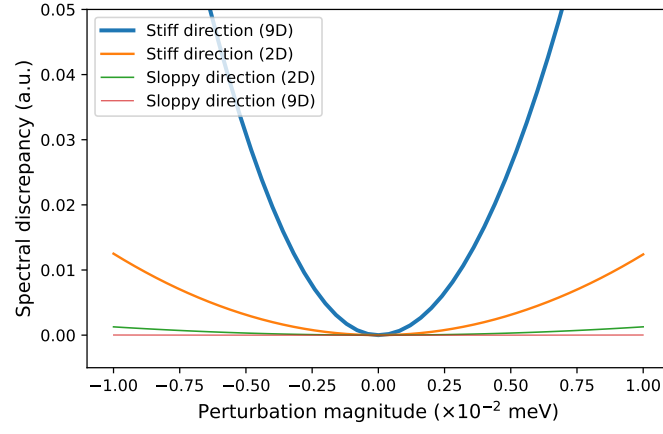


FIG. S6. Comparison of spectral discrepancy along the global nine-dimensional and local two-dimensional principal directions for the powder modality. The global stiff and sloppy eigen-directions exhibit the fastest and slowest discrepancy growth, respectively, while the corresponding $J_{1a}-J_{1b}$ two-dimensional principal directions show intermediate behavior, consistent with the observations for the single-crystal modality.

Supplementary Note 8. Pixel-wise decomposition of spectral correlation

To identify the spectral regions responsible for the observed agreement, we decompose the Pearson correlation coefficient into additive pixel-wise contributions. For either modality $\alpha \in \{\text{pd}, \text{sc}\}$, let $x_i^{(\alpha)}$ and $y_i^{(\alpha)}$ denote the predicted and experimental intensities, respectively, at valid pixel i . The corresponding Pearson correlation coefficient can be written as

$$\text{PCC}_\alpha = \frac{\sum_i \left(x_i^{(\alpha)} - \bar{x}^{(\alpha)} \right) \left(y_i^{(\alpha)} - \bar{y}^{(\alpha)} \right)}{\sqrt{\sum_i \left(x_i^{(\alpha)} - \bar{x}^{(\alpha)} \right)^2} \sqrt{\sum_i \left(y_i^{(\alpha)} - \bar{y}^{(\alpha)} \right)^2}} = \sum_i c_i^{(\alpha)}, \quad (\text{S16})$$

where

$$c_i^{(\alpha)} = \frac{\left(x_i^{(\alpha)} - \bar{x}^{(\alpha)} \right) \left(y_i^{(\alpha)} - \bar{y}^{(\alpha)} \right)}{\sqrt{\sum_j \left(x_j^{(\alpha)} - \bar{x}^{(\alpha)} \right)^2} \sqrt{\sum_j \left(y_j^{(\alpha)} - \bar{y}^{(\alpha)} \right)^2}} \quad (\text{S17})$$

is the contribution of the i th pixel to the global correlation PCC_α . Figures S7 and S8 show the pixel-wise contribution maps for the powder and single-crystal spectra, respectively, comparing the selected Hamiltonian θ_{best} and the reference Hamiltonian θ_{ref} with the experimental data. Larger positive contributions identify spectral regions that support the global correlation, whereas negative contributions identify regions that reduce it. These maps reveal which spectral regions account for the differences in global agreement reported in Fig. 4.

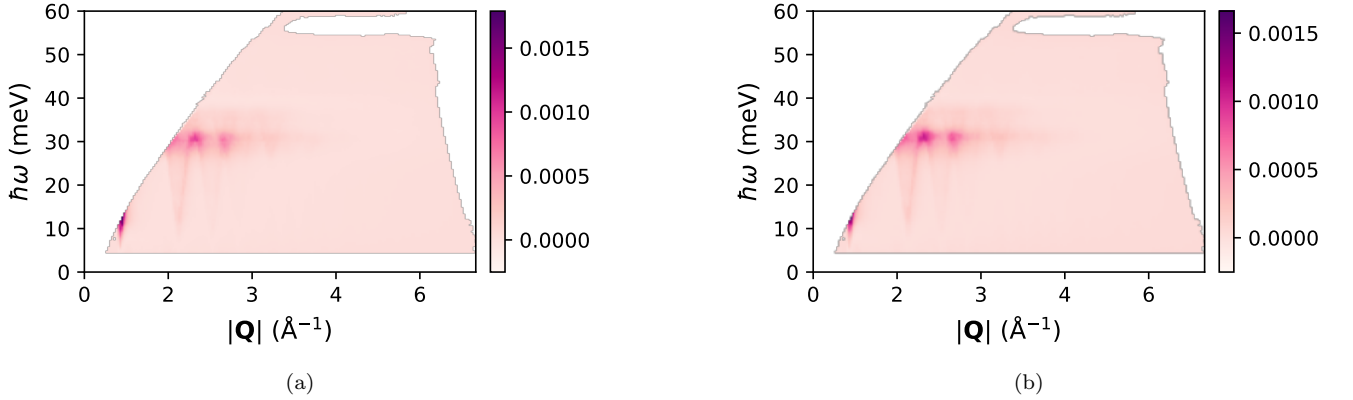


FIG. S7. Pixel-wise decomposition of the powder spectra Pearson correlation coefficient, PCC_{pd} , for (a) the selected Hamiltonian θ_{best} and (b) the literature reference Hamiltonian θ_{ref} .

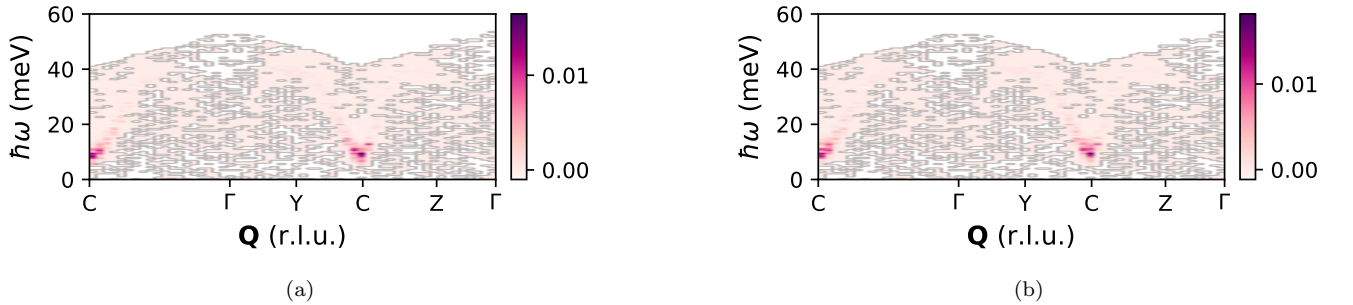


FIG. S8. Pixel-wise decomposition of the single-crystal spectra Pearson correlation coefficient, PCC_{sc} , for (a) the selected Hamiltonian θ_{best} and (b) the literature reference Hamiltonian θ_{ref} .

Supplementary Note 9. Effect of finite momentum-space binning

The single-crystal experimental spectra along the high-symmetry path were obtained by integrating the measured intensity over finite regions of reciprocal space. Directly incorporating this binning procedure during random-coordinate training data generation would substantially increase the computational cost. We therefore use the same sharp-coordinate surrogate and subsequently average its predictions over the corresponding experimental momentum-bin volumes. This separation also preserves the generality of the surrogate, allowing different momentum-space resolutions and integration widths to be applied at inference time without retraining. We do not apply additional energy binning because the surrogate is trained on simulated spectra with an energy broadening of $\text{FWHM} = 4 \text{ meV}$.

Figure S9 presents the resulting finite-bin-averaged spectral comparison, while Figs. S10 and S11 show the corresponding pixel-wise PCC contribution maps. The similar overall conclusions obtained with and without finite-bin averaging indicate that the reported multimodal spectral comparison is robust to this difference in representation. Table S2 lists the inferred Hamiltonian parameter values.

TABLE S2. Inferred Hamiltonian parameters obtained with finite $\mathbf{Q}$ -space averaging (meV).

	A_x	A_z	J_{1a}	J_{1b}	J_{2a}	J_{2b}	J_{3a}	J_{3b}	J_4
$\theta_{\text{best,binned}}$	-0.0082	0.21	-2.55	-2.10	0.21	0.17	13.78	14.86	-0.49

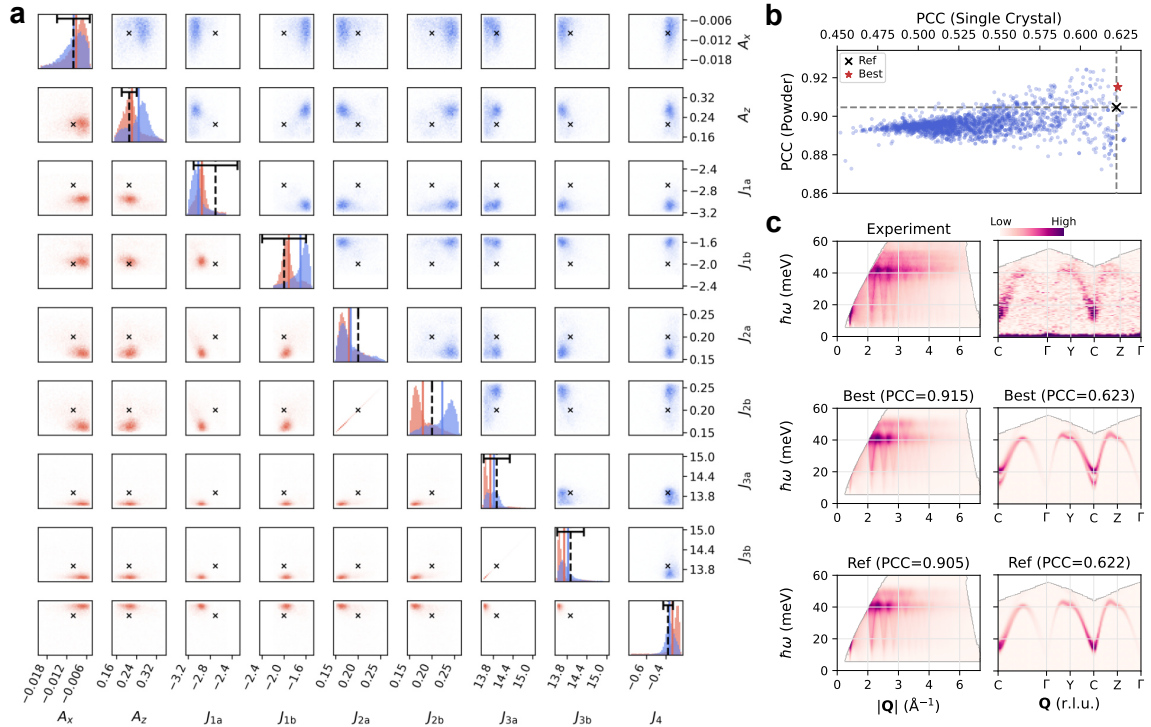


FIG. S9. Finite $\mathbf{Q}$ -space binning comparison corresponding to Fig. 4. The single-crystal surrogate spectra are averaged over the experimental momentum-bin volumes during Bayesian inference to obtain the reported posterior distribution. All panel definitions, symbols, and plotting conventions follow Fig. 4; see its caption for details.

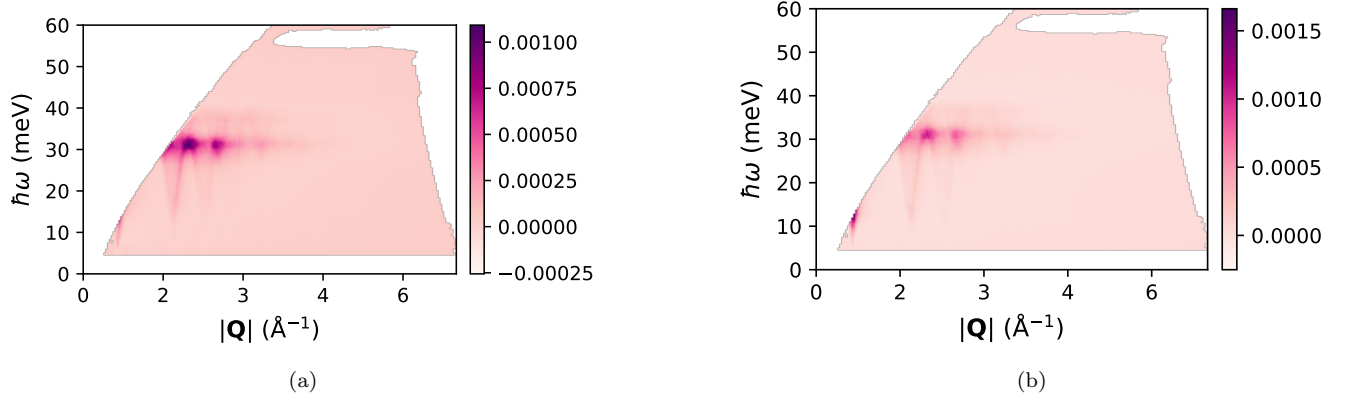


FIG. S10. Pixel-wise decomposition of the powder spectra Pearson correlation coefficient, PCC_{pd} , for **(a)** the selected Hamiltonian θ_{best} obtained with finite $\mathbf{Q}$ -space binning and **(b)** the literature reference Hamiltonian θ_{ref} .

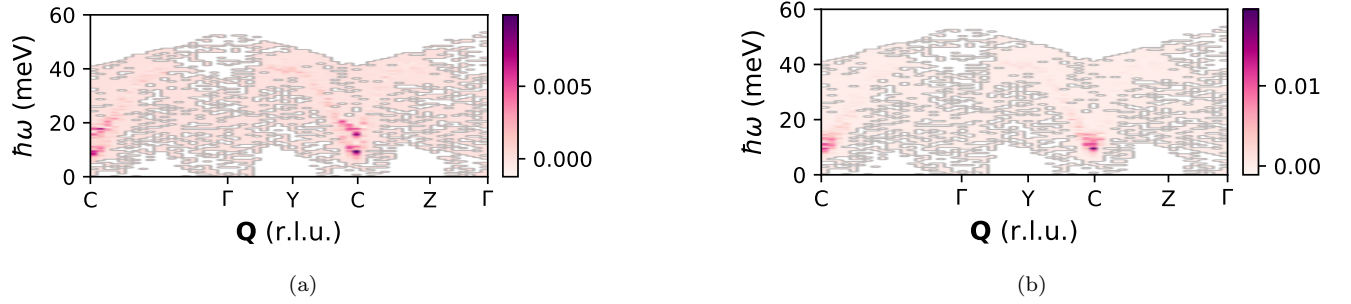


FIG. S11. Pixel-wise decomposition of the single-crystal spectra Pearson correlation coefficient, PCC_{sc} , for **(a)** the selected Hamiltonian θ_{best} obtained with finite $\mathbf{Q}$ -space binning and **(b)** the literature reference Hamiltonian θ_{ref} .

Supplementary Note 10. Extraction of powder magnon spectrum from the experimental INS data

INS measures not only magnetic excitations (magnons) but also lattice excitations (phonons) from the sample. In addition, the nonmagnetic aluminum sample holder contributes phonon scattering to the measured spectrum. Magnon scattering is typically strongest at low momentum transfer ($|\mathbf{Q}|$) and decreases with increasing $|\mathbf{Q}|$ due to the magnetic form factor. In contrast, phonon scattering generally becomes more prominent at higher $|\mathbf{Q}|$. In orientationally averaged powder spectra, the magnon and phonon contributions partially overlap. In this case, although most magnon intensity is concentrated at $|\mathbf{Q}| < 3 \text{ \AA}^{-1}$ and most phonon intensity appears at $|\mathbf{Q}| > 3 \text{ \AA}^{-1}$, the two contributions still overlap, and their separation remains essential because the AI surrogate model predicts only magnon spectra. To isolate the magnetic signal, we trained a source-separation vision transformer using synthetic datasets that include a variety of NiPS_3 magnon spectra, NiPS_3 phonon spectra, and aluminum phonon spectra. The extracted magnon spectrum was subsequently denoised using a feature-enhancement vision transformer to improve signal quality prior to inference. The spectra that go through the data processing workflow are illustrated in Fig. S12.

As shown in Fig. S12, the final NiPS_3 magnon spectrum preserves the key features of the measured magnetic signal. The source separation effectively isolates the magnon contribution from the phonon background, while the feature-enhancement model suppresses noise and sharpens spectral features within the constraints imposed by the underlying physical model. A detailed description of the source separation and spectral processing workflow using vision transformers will be presented in a separate publication, as its methodology is beyond the scope of the present work.

Because the source-separation and feature-enhancement procedures introduce additional model-dependent processing beyond the raw experimental measurement, the resulting powder spectrum should be interpreted as an effective magnetic-spectrum estimate rather than a model-independent experimental observable. The experimental powder analysis presented here is intended primarily to demonstrate the multimodal inference workflow; quantitative differences in the inferred Hamiltonian may depend on the specific preprocessing procedure. A systematic comparison of alternative powder-data processing strategies is left for future work.

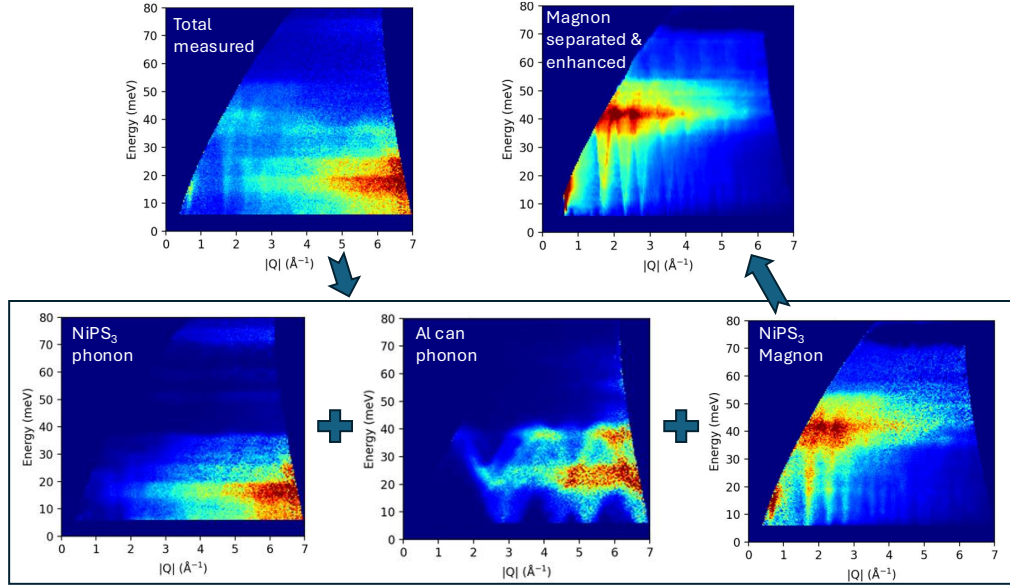


FIG. S12. Source separation of the measured INS spectrum into three components: sample magnons, sample phonons, and phonons originating from the aluminum sample holder. Each panel uses an independent color scale to better visualize the spectral features of the corresponding component. On a common intensity scale, the measured total spectrum is the sum of these three contributions. The isolated magnon spectrum is subsequently processed with a feature-enhancement model to suppress noise and sharpen physically consistent spectral features. The elastic peak and its low-energy tail near $E = 0$ have been masked.