

Neural Posterior Estimation via Hierarchical Bayesian Distillation

Sina Negarandeh

Thesis submitted to the University of Ottawa
in partial Fulfillment of the requirements for the
Master of Computer Science
Concentration in Applied Artificial Intelligence

School of Electrical Engineering and Computer Science
Faculty of Engineering
University of Ottawa

© Sina Negarandeh, Ottawa, Canada, 2026

Abstract

Solving inverse problems for spectral data is fundamentally hindered by data scarcity, ill-posedness, and the strict requirement for accurate uncertainty quantification (UQ). Addressing these challenges in the context of X-ray diffraction (XRD), an essential technique for materials discovery and structural health monitoring, we propose an amortized neural posterior estimation framework. Our approach distills a computationally expensive, physics-informed Bayesian teacher into a highly efficient conditional normalizing flow (CNF) student. By enforcing crystallographic constraints within our base generative model, we construct a representative synthetic training corpus to optimize the CNF, effectively bypassing the bottleneck of limited real-world data. We extensively validate the neural posterior’s fidelity against the Bayesian teacher using posterior predictive checks and Wasserstein distances. Finally, we demonstrate the practical utility of this approach on a real-world structural damage detection task. Evaluated in a leave-one-group-out cross-validation (LOGOCV) setup, our amortized posteriors maintain the predictive classification performance and UQ calibration of the Bayesian baseline while accelerating inference by orders of magnitude.

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my supervisor, Dr. Verena Kantere, for her invaluable guidance, mentorship, and support throughout my Master's studies. I am equally thankful for the time I spent in the Modern Management of Data and Analytics (M²oDA) lab, surrounded by an inspiring group of people.

A special note of appreciation is owed to the Fatigue and Stress Analysis Laboratory (FATSLab) at the University of Waterloo. The core experimental data utilized in this thesis was acquired through their extensive efforts. I am particularly thankful for our close collaboration in designing and conducting the data generation pipeline.

This work was made possible by the generous financial support of the Natural Sciences and Engineering Research Council of Canada (NSERC) through the NSERC Alliance grant.

Finally, I wish to thank my family and friends. Their immense patience, understanding, and unwavering emotional support throughout the entirety of this academic journey are what allowed me to continue, and for that, I am beyond grateful.

Table of Contents

Abstract	ii
Acknowledgements	iii
List of Figures	x
List of Tables	xii
List of Abbreviations	xiii
1 Introduction	1
1.1 Motivating Application: Structural Health Monitoring	2
1.2 Problem Statement	3
1.3 Proposed Approach	4
1.4 Contributions	5
1.5 Thesis Organization	6
2 Background and Related Work	8
2.1 X-ray Diffraction Physics and Damage Characterization	8
2.1.1 Physical Foundations and Spectral Modeling	9
2.1.2 Diffraction Signatures of Fatigue Damage	9
2.1.3 Analytical Models for Fatigue Life Assessment	10

2.1.4	Machine Learning in XRD: Opportunities and Limitations	10
2.2	Bayesian Spectral Deconvolution and Hierarchical Modeling	11
2.2.1	Bayesian Spectral Deconvolution	12
2.2.2	Hierarchical Models and Partial Pooling	12
2.2.3	Probabilistic Inference: MCMC versus Variational Methods	13
2.3	Simulation-Based Inference and Normalizing Flows	14
2.3.1	Simulation-Based Inference	14
2.3.2	Neural Posterior Estimation	15
2.3.3	Normalizing Flows and Conditional Density Estimation	16
2.3.4	Masked Autoregressive Flow and MADE	17
2.4	Summary	18
3	Problem Formulation	19
3.1	Notation	19
3.2	Grouped Sample Structure	19
3.3	Spectral Observations	20
3.4	Inverse Problem	20
3.5	Problem Statement	22
3.6	Key Challenges	23
3.7	Summary	24
4	Data Acquisition and Preprocessing	25
4.1	Material Science Background	25
4.2	Data Generation	27
4.2.1	Specimen Acquisition	28
4.2.2	Damage Induction	28
4.2.3	XRD Characterization	29
4.3	Dataset Description and Characteristics	29

4.4	Preprocessing and Grid Alignment	29
4.5	Exploratory Data Analysis	30
4.5.1	Spectral Overview	31
4.5.2	Intensity Distribution	31
4.6	Physics-Informed Structural Priors	31
4.6.1	Fatigue Damage Evolution	35
4.7	Summary	36
5	Proposed Framework	39
5.1	Physics-Guided Teacher-Student Framework	39
5.2	Bayesian Spectral Deconvolution and Posterior Extraction	41
5.2.1	Methodological Design Space	41
5.3	Physics-Guided Synthetic Data Generation	44
5.3.1	The Posterior-Grounded Augmentation Problem	45
5.3.2	Mixed Generative Strategy	45
5.4	Amortized Neural Posterior Inference	48
5.4.1	Amortized Neural Posterior Estimation and Normalizing Flows	48
5.4.2	Domain and Parameter Transformations	49
5.4.3	Spectral Context Encoder	49
5.4.4	Conditional Masked Autoregressive Flow (cMAF)	50
5.4.5	Simulation-to-Reality Training Protocol	52
5.5	Summary	53
6	Implementation and Instantiation	54
6.1	Model Configuration Selection	54
6.1.1	Crystallographic Scope and Fitting Strategy	54
6.1.2	Structural Pooling Paradigm	55
6.1.3	Probabilistic Inference Engine	58

6.1.4	Peak Profile Morphology	62
6.2	Prior Specifications	62
6.2.1	Peak Position Priors	62
6.2.2	Peak Shape Priors	62
6.2.3	Background and Noise Model Priors	63
6.3	Neural Architecture and Training Hyperparameters	64
6.4	Software and Hardware Environment	68
6.5	Summary	69
7	Experimental Setup and Evaluation Methodology	70
7.1	Evaluation Dimensions	70
7.2	Task Definition	71
7.2.1	Spectral Inverse Problem	71
7.2.2	Downstream Classification	71
7.3	Feature Extraction	72
7.4	Cross-Validation Protocol	73
7.5	Evaluation Metrics	73
7.5.1	Spectral Fitting Quality	73
7.5.2	Posterior Quality and Physical Calibration	74
7.5.3	Computational Efficiency	75
7.5.4	Predictive Performance	75
7.5.5	Uncertainty Quantification	76
7.5.6	Physical Interpretability	77
7.6	Classification Models	77
7.7	Baselines	78
7.8	Summary	79

8	Experimental Results	80
8.1	Spectral Fitting Quality	80
8.1.1	Global Fitting	81
8.1.2	Localized Window Fitting	81
8.2	Posterior Quality	82
8.2.1	Physical Calibration	84
8.2.2	Posterior Geometry	85
8.3	Computational Efficiency	85
8.4	Downstream Classification Performance	86
8.4.1	Predictive Performance	86
8.4.2	Uncertainty Quantification	87
8.4.3	Physical Interpretability	90
8.5	Summary	90
9	Discussion and Analysis	92
9.1	Physical Validity of the Bayesian Teacher	92
9.2	Fidelity and Cost of Amortization	93
9.2.1	Bridging the Simulation-to-Reality Gap	93
9.2.2	Computational Cost of Amortization	94
9.3	Downstream Utility of Uncertainty-Aware Representations	94
9.3.1	Effect of Uncertainty Features	94
9.3.2	Calibration Quality	95
9.4	Physical Grounding of Classification Decisions	95
9.5	Generalization to Other Materials	95
9.6	Research Questions Revisited	96
9.7	Limitations	97

10 Conclusion	99
10.1 Summary of Research Outcomes	99
10.2 Broader Implications	100
10.3 Future Work	100
References	102

List of Figures

3.1	Forward model and inverse problem	21
3.2	Direct and two-stage formulations	23
4.1	Geometric illustration of Bragg’s Law	26
4.2	AA6061-T6 XRD spectral observation and Debye–Scherrer rings	27
4.3	Experimental data generation pipeline	27
4.4	Dog-bone fatigue specimen geometry	28
4.5	Domain coverage of all XRD scans by specimen	32
4.6	Raw XRD spectral profiles with FCC Bragg positions	33
4.7	Distribution of maximum photon intensity counts	33
4.8	Mean XRD spectrum and inter-scan variability	34
4.9	Log-scaled mean XRD intensity profile	34
4.10	Mean normalized XRD intensity profiles	35
4.11	Mean normalized XRD intensity profiles by damage state	37
5.1	Physics-Guided Teacher-Student Framework	40
5.2	cMAF overview	51
6.1	Representative global fitting	56
6.2	Representative localized window fitting	57
6.3	No-Pooling Individual Scan Model as a Directed Graphical Model	59
6.4	Hierarchical Partial-Pooling Directed Graphical Model	60

6.5	Detailed Hierarchical Directed Graphical Model	61
6.6	1D Spectral Context Encoder Architecture	65
6.7	Conditional Masked Autoregressive Flow Architecture	66
8.1	Global fitting PPC	81
8.2	Localized window fitting PPC	83
8.3	Teacher-Student PPC	84
8.4	SHAP summary plot	91

List of Tables

3.1	Summary of core mathematical notation	20
4.1	Dataset summary	30
6.1	Prior distributions	63
6.2	cMAF architecture specification	66
6.3	Training and posterior inference hyperparameters	67
6.4	Summary of the computational hardware and software environment	68
8.1	Localized window fitting quality	82
8.2	Spectral fitting quality	84
8.3	Posterior geometry discrepancy	85
8.4	Computational efficiency comparison	85
8.5	Downstream predictive performance	86
8.6	Per-specimen accuracy	88
8.7	Localized fitting downstream binary classification performance	88
8.8	Localized fitting per-specimen accuracy	89
8.9	Uncertainty quantification metrics	89

List of Abbreviations

ABC Approximate Bayesian Computation [14](#), [15](#)

ADVI Automatic Differentiation Variational Inference [1](#), [13](#), [14](#)

APT Automatic Posterior Transformation [15](#)

AUC Area Under the ROC Curve [76](#), [79](#), [86](#), [87](#)

cMAF Conditional Masked Autoregressive Flow [5](#), [17](#), [18](#), [39](#), [44](#), [50](#), [53](#), [55](#), [56](#), [64](#), [66](#), [71–75](#), [77](#), [78](#), [82](#), [84–87](#), [90](#), [93–97](#), [99](#)

CNF Conditional Normalizing Flow [17](#), [49](#)

DAG Directed Acyclic Graph [58–61](#)

ECE Expected Calibration Error [76](#), [79](#), [87](#), [89](#), [90](#), [95](#)

ELBO Evidence Lower Bound [13](#), [43](#)

FCC Face-Centered Cubic [3](#), [6](#), [9](#), [28](#), [31](#), [33–35](#), [42](#), [54–57](#), [61](#), [62](#), [69](#), [81–83](#), [93](#), [95](#), [98](#), [100](#)

FTIR Fourier-transform infrared spectroscopy [19](#)

FWHM Full-Width at Half-Maximum [6](#), [9](#), [14](#), [25](#), [44](#), [46](#), [47](#), [49](#), [61](#), [62](#), [75](#), [85](#), [90](#), [95](#), [97](#), [100](#)

IAF Inverse Autoregressive Flow [17](#)

LOGOCV Leave-One-Group-Out Cross-Validation [5](#), [6](#), [73](#), [75](#), [78–80](#), [84–86](#), [88–90](#), [92](#), [97](#)

MADE Masked Autoencoder for Distribution Estimation [17](#), [50](#), [64](#)

MAF Masked Autoregressive Flow [17](#)

MCMC Markov chain Monte Carlo [1](#), [3–5](#), [13](#), [14](#), [18](#), [24](#), [39](#), [43](#), [46](#), [48](#), [52](#), [58](#), [68–75](#), [77](#), [78](#), [80–87](#), [90–96](#)

MDN Mixture Density Network [15](#)

MLP Multi-Layer Perceptron [50](#)

MS Mass Spectrometry [19](#)

NLL Negative Log-Likelihood [48](#), [52](#)

NPE Neural Posterior Estimation [2](#), [14–16](#), [48](#)

NUTS No-U-Turn Sampler [4](#), [13](#), [14](#), [43](#), [58](#), [68](#)

PCHIP Piecewise Cubic Hermite Interpolating Polynomial [30](#), [31](#), [36](#)

PPC Posterior Predictive Checks [73](#), [80](#), [84](#)

pV pseudo-Voigt [9](#), [21](#), [44](#)

RMSE Root Mean Square Error [74](#), [79–82](#), [84](#), [92](#)

SBI Simulation-Based Inference [2](#), [8](#), [14](#), [15](#), [18](#), [99](#)

SHAP SHapley Additive exPlanations [6](#), [77](#), [79](#), [90](#), [95](#), [97](#), [100](#)

sPV Split Pseudo-Voigt [4](#), [5](#), [9](#), [62](#), [69](#), [81](#), [92](#), [95](#), [99](#)

XRD X-ray Diffraction [2–6](#), [8–10](#), [17–19](#), [25](#), [27](#), [29](#), [30](#), [32–36](#), [41](#), [44](#), [70](#), [80](#), [92](#), [96](#), [100](#)

Chapter 1

Introduction

Many scientific and engineering problems require inferring latent physical parameters from noisy, indirect measurements such as spectra, images, or time series. These are known as inverse problems, where rather than predicting observable outputs from known inputs, the task is to recover unobservable model parameters or system states from data. Inverse problems are, in general, ill-posed in the sense of Hadamard, meaning that solutions may fail to satisfy one or more of the classical conditions of existence, uniqueness, and stability under variations of the data. This ill-posedness reflects the fundamental difficulty that many different parameter configurations may be consistent with the same set of observations, and that small amounts of measurement noise can produce large errors in naive reconstructions.

In practice, data are often scarce, physical constraints are strong, and the measurements themselves are corrupted by noise from multiple sources. These compounding challenges mean that point estimates of unknown parameters, while computationally convenient, are insufficient for reliable parameter estimation. The consequences of ignoring uncertainty are especially severe in safety-critical applications, where a confidently incorrect prediction is significantly more dangerous than an honest admission of ambiguity.

Physics-informed probabilistic models address this by encoding domain constraints through structured priors and yielding posterior distributions that accurately quantify inferential uncertainty. While per-sample iterative inference methods, such as the sampling-based [Markov chain Monte Carlo \(MCMC\)](#) [13, 1] or the optimization-based [Automatic Differentiation Variational Inference \(ADVI\)](#) [17] provide accurate results, they must be rerun from scratch for each new observation. This makes computational throughput a fundamental bottleneck and precludes any real-time deployment scenario. To address this

limitation, [Simulation-Based Inference \(SBI\)](#) [6] methods have emerged as a principled alternative. These methods employ amortized [Neural Posterior Estimation \(NPE\)](#) [26, 11, 33] to provide near-instant posterior inference after a one-time offline training phase. Once trained, these methods, often implemented via conditional normalizing flows [34, 41], can transform the cost of inference from $\mathcal{O}(T)$ iterative steps to $\mathcal{O}(1)$ forward passes, enabling real-time deployment. However, standard [SBI](#) approaches rely on abundant forward simulations from a known generative model and commonly struggle to capture the subtle cross-parameter correlations present in real experimental data. When the synthetic training distribution fails to cover the true empirical data manifold, the resulting posteriors can be poorly calibrated and physically implausible. This failure mode is especially acute in domains where data acquisition is expensive and the gap between simulation and reality is large.

[X-ray Diffraction \(XRD\)](#) based structural health monitoring crystallizes all of these challenges simultaneously.

1.1 Motivating Application: Structural Health Monitoring

[XRD](#) analysis exemplifies this fundamental tension between physical fidelity and computational throughput. [XRD](#) is a non-destructive characterization technique that encodes rich microstructural information, including lattice parameters, dislocation density, and residual stress, within the angular positions, intensities, and morphologies of diffraction peaks. In the context of structural health monitoring, the progressive accumulation of fatigue damage in metallic alloys manifests as measurable changes in peak broadening, asymmetry, and position, providing a physically grounded signature of material degradation.

Machine learning has recently transformed the [XRD](#) analysis landscape, yet a comprehensive survey of the field [35] identifies four persistent and critical limitations.

1. **Data Scarcity:** high-quality labeled [XRD](#) datasets are prohibitively expensive to acquire, as each specimen requires controlled fatigue loading, precise diffractometer alignment, and expert annotation of the damage state.
2. **Lack of Interpretability:** black-box neural network predictions offer no physical insight into the crystallographic mechanisms driving the prediction, precluding scientific validation.

3. **Poor Generalization:** models trained on one material system or loading condition routinely fail on unseen configurations due to the absence of physics constraints in the learned representation.
4. **Absence of Physics Constraints:** models that ignore crystallographic principles, such as Bragg’s Law [2] and **Face-Centered Cubic (FCC)** selection rules, yield physically implausible predictions that cannot be trusted in safety-critical contexts.

Consequently, the materials science and engineering communities have actively called for hybrid approaches that combine the inference speed of machine learning with the physical accuracy of model-based analysis.

1.2 Problem Statement

This thesis addresses the following overarching problem: given a small, labeled dataset of **XRD** spectra acquired from fatigue-loaded aluminum alloy specimens, how can one perform fast, physically accurate, and uncertainty-aware inference of the underlying microstructural parameters, and subsequently use those inferred parameters to classify the structural damage state of each specimen?

This overarching problem consists of two interdependent subproblems.

- The first is the *spectral inverse problem*: for each observed spectrum, the model must recover the posterior distribution over the latent physical parameters (peak positions, widths, asymmetries, and intensities) in a manner that is consistent with crystallographic theory, accounts for heteroscedastic measurement noise, and propagates uncertainty rather than collapsing to a point estimate.
- The second is the *downstream classification problem*: the model must utilize the inferred posteriors, which encode both the estimated physical state and its associated uncertainty, as feature representations for a supervised classifier that generalizes robustly to unseen physical specimens.

The central challenge is that these two subproblems are in computational tension. Exact Bayesian inference via **MCMC** resolves the first subproblem with high fidelity, but at a computational cost that prohibits real-time deployment. Standard amortized inference methods resolve the throughput constraint but sacrifice physical fidelity when the

simulation-to-reality gap is large. An ideal solution must simultaneously satisfy the physical fidelity requirement of the inverse problem and the throughput requirement of the downstream task.

To resolve this tension and systematically evaluate the proposed solution, this research seeks to answer four fundamental questions:

1. **High-Fidelity Ground Truth:** Is it possible to establish a physically accurate, uncertainty-aware ground truth for microstructural parameters from a limited dataset of noisy [XRD](#) spectra?
2. **Amortized Inference:** Is it possible to distill the posterior geometry of an exact Bayesian model into a neural network to achieve real-time inference without sacrificing physical fidelity or suffering from simulation-to-reality distribution shifts?
3. **Downstream Utility:** Do the uncertainty-aware posteriors provide a robust feature representation for the binary classification of structural damage, and does this approach generalize effectively to unseen physical specimens?
4. **Physical Interpretability:** Can the resulting classification decisions be attributed to physically meaningful parameters and validated against established crystallographic principles?

1.3 Proposed Approach

This thesis proposes a Physics-Guided Teacher-Student Distillation Framework that operationalizes this hybrid vision by separating the concerns of physical fidelity and computational efficiency across two distinct models.

The **teacher** is a hierarchical Bayesian model that reformulates the spectral deconvolution approach of [21] using asymmetric [Split Pseudo-Voigt \(sPV\)](#) peak profiles, a heteroscedastic noise model, and partial pooling across all specimens under shared population-level hyperpriors. This model is fit to the real [XRD](#) dataset via [MCMC](#) utilizing the [No-U-Turn Sampler \(NUTS\)](#) [14]. This yields reference posteriors that are physically calibrated, asymptotically exact, and highly robust to the heterogeneous signal-to-noise conditions present in the dataset. The teacher operates at $\mathcal{O}(T)$ complexity and is not intended for operational deployment; its sole purpose is to produce a high-fidelity ground truth that supervises the neural student.

The **student** is a [Conditional Masked Autoregressive Flow \(cMAF\)](#) paired with a domain-specific 1D residual convolutional encoder. The student is trained via a three-phase protocol designed specifically to bridge the simulation-to-reality gap. In the first phase, the student is pre-trained on a synthetic corpus generated from the teacher’s posterior geometry using three complementary strategies: posterior jittering, Gaussian copula sampling, and forward model noise injection. In the second phase, the student is fine-tuned directly on the teacher’s reference posteriors from the empirical data, first with the encoder frozen and then end-to-end, utilizing a synthetic replay mechanism to prevent catastrophic forgetting of the physical constraints. After training, the student processes each new spectrum by sampling directly from the trained flow at $\mathcal{O}(1)$ complexity, drastically reducing inference latency while preserving the geometric fidelity of the teacher.

The inferred posteriors, represented as highly structured vectors containing posterior means and standard deviations, are subsequently used as feature representations for the downstream binary classification of structural damage (Nominal versus Damaged) under strict [Leave-One-Group-Out Cross-Validation \(LOGOCV\)](#).

1.4 Contributions

The primary contribution of this thesis is the successful design, implementation, and evaluation of an amortized neural inference framework for the safety-critical task of [XRD](#) fatigue damage assessment. The specific research contributions are as follows:

1. **Hierarchical Bayesian Spectral Deconvolution:** A fully specified hierarchical [MCMC](#) model is developed for the AA6061-T6 fatigue dataset, incorporating asymmetric [sPV](#) peak profiles, heteroscedastic noise modeling, and partial pooling across specimens. This model provides physically calibrated reference posteriors across a wide range of signal-to-noise conditions, achieving a near-perfect spectral reconstruction R^2 of 0.992.
2. **Physics-Guided Teacher-Student Distillation:** A generalizable framework is proposed for distilling the posterior geometry of an exact Bayesian teacher into an amortized neural student via correlation-preserving synthetic data generation and a staged simulation-to-reality fine-tuning protocol.
3. **Amortized Neural Posterior Inference:** A [cMAF](#) equipped with a multi-scale 1D residual convolutional encoder is designed and trained. The student achieves a

spectral reconstruction R^2 of 0.918 and a 1-Wasserstein distance of 3.08 from the teacher’s reference posterior, while successfully reducing inference latency by orders of magnitude.

4. **Empirical Validation on Structural Health Monitoring:** The framework is validated end-to-end on a downstream binary damage classification task under strict LOGOCV across eight physical specimens. The amortized student reaches a classification accuracy of 0.952, close to the exact Bayesian teacher (0.971) and above the BayesFlow and end-to-end CNN baselines. Furthermore, a localized window fitting ablation yields a peak diagnostic accuracy of 0.991.
5. **Physical Interpretability Analysis:** SHapley Additive exPlanations (SHAP) analysis identifies the left and right Full-Width at Half-Maximum (FWHM) parameters of the (311), (200), (422), and (111) FCC reflections as the dominant discriminative features. This provides necessary physical validation that the framework’s predictions are grounded in established fatigue-induced lattice distortion mechanisms rather than statistical artifacts.

1.5 Thesis Organization

The remainder of this thesis is organized as follows:

- **Chapter 2 (Background and Related Work):** Reviews the necessary theoretical background in three core areas: XRD physics, Bayesian spectral analysis, and simulation-based inference with normalizing flows.
- **Chapter 3 (Problem Formulation):** Formalizes the spectral inverse problem and the downstream classification task, defines the core mathematical notation, and outlines the structural challenges inherent to the dataset.
- **Chapter 4 (Data Acquisition and Preprocessing):** Describes the AA6061-T6 fatigue dataset, the experimental conditions under which the XRD spectra were acquired, and the preprocessing pipeline applied to produce normalized spectral representations.
- **Chapter 5 (Proposed Framework):** Presents the Physics-Guided Teacher-Student Distillation Framework in full generality, defining the hierarchical Bayesian teacher, the synthetic corpus generation strategies, and the neural student architecture.

- **Chapter 6 (Implementation and Instantiation)**: Details the concrete instantiation of the framework for the AA6061-T6 dataset, documenting the selected model configurations, prior specifications, architecture hyperparameters, and hardware environment.
- **Chapter 7 (Experimental Setup and Evaluation Methodology)**: Defines the complete evaluation methodology, including the cross-validation protocol, feature extraction procedures, formal metric definitions, and baseline models.
- **Chapter 8 (Experimental Results)**: Presents all quantitative findings organized by evaluation dimension, covering spectral fitting quality, posterior fidelity, computational efficiency, and downstream classification performance.
- **Chapter 9 (Discussion and Analysis)**: Interprets the experimental findings, analyzes the trade-offs of amortization, evaluates the physical interpretability of the classification decisions, and discusses the broader applicability of the framework.
- **Chapter 10 (Conclusion)**: Summarizes the core outcomes of the research and proposes specific avenues for future work.

Chapter 2

Background and Related Work

This chapter details the theoretical background and related literature necessary to develop the proposed inference framework. The review is organized into three specific areas. First, it outlines the physical principles of [XRD](#) and the established methods for characterizing fatigue damage, highlighting the practical limitations of current machine learning approaches in this domain. Second, it details Bayesian spectral deconvolution and hierarchical modeling as an accurate approach for enforcing crystallographic constraints and extracting calibrated uncertainty estimates from noisy spectra. Finally, it reviews [SBI](#) and normalizing flows, which provide the neural network techniques required to overcome the high computational cost of exact Bayesian sampling. Together, the intersection of these three domains defines the specific methodological gap addressed by this research.

2.1 X-ray Diffraction Physics and Damage Characterization

The integration of machine learning into structural health monitoring requires a thorough understanding of both the physical data generation process and the historical methods used to analyze it. This section reviews the foundational literature on [XRD](#) physics, traces the evolution of analytical fatigue damage assessment, and identifies the persistent limitations in current data-driven approaches that motivate this thesis.

2.1.1 Physical Foundations and Spectral Modeling

XRD serves as a non-destructive probe of internal crystalline microstructure. The foundational mechanism, formalized by Bragg [2], dictates that constructive interference from parallel atomic planes occurs strictly when the path difference between scattered waves is an integer multiple of the X-ray wavelength. This is mathematically defined as:

$$n\lambda = 2d_{hkl} \sin \theta, \quad (2.1)$$

where n is a positive integer, d_{hkl} is the interplanar spacing for the crystallographic plane indexed by Miller indices (h, k, l) , and θ is the glancing angle of incidence. Equation (2.1) ensures that peak positions act as a deterministic fingerprint of the crystal geometry. Furthermore, prior crystallographic studies establish that for an **FCC** lattice like AA6061-T6, destructive interference systematically eliminates reflections from mixed-parity planes. This selection rule strictly bounds the latent space to exactly K permitted reflections within the standard acquisition range, providing a rigid structural prior for subsequent inference models.

In quantitative diffraction line-profile analysis, literature standardizes the modeling of peak broadening using localized continuous functions. The **pseudo-Voigt (pV)** function is established as the standard mathematical approximation for this morphology [38]. The profile is defined as a linear combination of a Lorentzian $\mathcal{L}$ and a Gaussian $\mathcal{G}$ distribution, governed by a mixing parameter $\eta \in [0, 1]$:

$$\text{pV}(t; \mu, \Gamma, \eta) = \eta \mathcal{L}(t; \mu, \Gamma) + (1 - \eta) \mathcal{G}(t; \mu, \Gamma), \quad (2.2)$$

where μ is the peak center and Γ is the **FWHM**. While symmetric profiles are adequate for isotropic microstrain, capturing the complex, directional lattice distortions induced by fatigue requires decoupling the width parameters. Consequently, modern microstructural analyses frequently rely on an asymmetric **sPV** formulation [40] to independently parameterize the left and right flanks of the diffraction peak [25].

2.1.2 Diffraction Signatures of Fatigue Damage

The application of **XRD** to structural health monitoring is grounded in materials science literature linking fatigue to microscopic crystallographic defects. Early systematic studies [36] established that the superposition of dislocation sub-cells manifests directly in the measured spectrum as peak broadening and directional asymmetry.

Subsequent studies have thoroughly characterized this evolution. A three-stage progression of diffraction breadth over a component’s fatigue life was empirically identified as a rapid initial increase, a prolonged stable plateau, and a final sharp rise preceding fracture [37]. This was expanded by explicitly modeling the spatial distribution of dislocations, proving that the degree of breadth evolution is directly correlated with applied strain amplitude and cyclic history [23, 24]. This body of work conclusively validates that the latent morphological parameters of a diffraction peak encode the physical state of the material.

2.1.3 Analytical Models for Fatigue Life Assessment

Relying on these physical mechanisms, the engineering community has developed a robust lineage of analytical models for fatigue life assessment. It has been demonstrated that breadth measurements can be correlated with remaining life through analytical models calibrated on aluminum and steel specimens [42]. This paradigm was successfully extended to highly complex, safety-critical domains, including steel pipelines under pressure cycling [16] and oil-industry infrastructure [32, 31]. Furthermore, these hand-engineered empirical correlations continue to be refined in recent analytical frameworks [7].

More recently, XRD analysis has been demonstrated to be capable of capturing early signs of structural damage specifically within the aluminum alloy AA6061-T6 [10].

2.1.4 Machine Learning in XRD: Opportunities and Limitations

To overcome the limitations of manual profile fitting and simple scalar correlations, recent literature has increasingly explored machine learning for spectral analysis. Deep neural networks and other machine learning paradigms offer the capacity to map raw XRD spectra directly to damage classifications, phase identifications, and microstructural parameters with sub-second latency [35].

However, a comprehensive survey [35] evaluating machine learning methods in XRD highlights some critical limitations that currently inhibit their widespread deployment in safety-critical contexts:

1. **Data Scarcity:** High-quality labeled XRD datasets require exhaustive experimental procedures. Consequently, modern models are constrained to the small-data regime, frequently resulting in severe overfitting.

2. **Lack of Interpretability:** Standard neural networks function as black boxes. They provide no guarantee that their predictions are driven by valid crystallographic phenomena rather than statistical artifacts in the measurement noise.
3. **Poor Generalization:** Purely data-driven models demonstrate extreme fragility to distribution shifts, frequently failing when evaluated on unseen physical specimens or varying diffractometer alignments.
4. **Absence of Physics Constraints:** Current machine learning approaches are predominantly physics-agnostic, treating the diffraction spectrum as a generic signal without enforcing the structural constraints imposed by crystallographic theory. The disconnect between data-driven analysis and underlying physics has been identified as a source of incorrect conclusions and a barrier to the broader adoption of machine learning techniques in the field.

These identified gaps in the literature define the exact research problem addressed in this thesis. Moving beyond deterministic analytical models and unconstrained neural networks necessitates a hybrid methodology. The field requires a framework that enforces the physical constraints established by classical crystallography, quantifies the uncertainty inherent in the fatigue literature, and preserves the computational throughput promised by modern deep learning.

2.2 Bayesian Spectral Deconvolution and Hierarchical Modeling

The traditional approach to spectral deconvolution relies on deterministic optimization techniques, most commonly non-linear least squares, to recover the latent parameters of a forward physical model. While computationally inexpensive, deterministic methods converge to a single point estimate. This approach discards all information regarding parameter uncertainty and correlation, and it is consequently susceptible to overfitting when applied to noisy, overlapping diffraction peaks where multiple parameter configurations produce comparably small residuals. Bayesian spectral deconvolution resolves this by reframing peak fitting as a probabilistic inference task. Rather than searching for a single optimal parameter vector, the Bayesian framework infers the full posterior probability distribution over all latent physical parameters, conditioned on the observed spectrum and constrained by prior physical knowledge. This reframing provides calibrated uncertainty

estimates for each inferred quantity, supports principled model comparison via marginal likelihoods, and naturally incorporates domain knowledge through the prior distribution.

2.2.1 Bayesian Spectral Deconvolution

The application of Bayesian inference to spectral deconvolution was formalized in a series of contributions [21, 39]. The foundational work [21] treats spectral decomposition as a Bayesian model selection problem: given an observed spectrum, the goal is to simultaneously infer the number of constituent peaks K and the parameters of each peak function (position, width, and amplitude) from the data, without manual specification of K . The framework represents the spectrum as a superposition of symmetric basis functions, assigns prior distributions over peak parameters, and employs the exchange Monte Carlo method to sample from the posterior while escaping the local minima that plague the nonlinear, multimodal posteriors arising from multi-peak models [21]. Model selection over K is performed via the Bayes free energy, an approximation to the log marginal likelihood, which penalises model complexity and thereby prevents the introduction of spurious peaks. A subsequent extension [39] further demonstrated that the noise variance can be estimated as a hyperparameter jointly with K , removing the requirement for a pre-specified signal-to-noise assumption and eliminating a common source of misspecification in prior work.

More recent work has validated the broader direction of incorporating domain-specific physical priors into Bayesian spectral deconvolution. [15] extended the foundational framework to X-ray absorption near edge structure spectroscopy by introducing priors that explicitly discriminate between the physical properties of peaks in distinct energy domains, demonstrating that domain-specific prior distributions improve estimation accuracy and model evidence relative to physics-agnostic treatments. This result reinforces the design choice made in the teacher model to encode crystallographic constraints directly in the prior structure.

2.2.2 Hierarchical Models and Partial Pooling

A central challenge in experimental materials science is that data is intrinsically grouped. The dataset utilized in this work comprises multiple spectral scans acquired from a small set of physical specimens. If the spectral model is fitted independently to each scan, the small within-scan sample size produces high-variance parameter estimates, particularly for lower-intensity reflections where the signal-to-noise ratio is poor. Conversely, pooling all

scans into a single global model destroys the genuine between-specimen variability required to distinguish damage states.

Hierarchical Bayesian models provide a principled resolution to this tension through partial pooling [8]. In a hierarchical architecture, group-level parameters are treated as draws from a shared population distribution governed by hyperpriors. This structure induces an automatic regularization effect. Estimates for highly noisy, data-poor scans are shrunk toward the population mean in proportion to their uncertainty, preventing them from drifting into physically implausible configurations. Conversely, data-rich, high-intensity scans retain estimates close to their observed evidence while simultaneously informing the population-level hyperparameters that serve the noisier scans. This architecture is essential for obtaining well-calibrated posteriors across the heterogeneous signal-to-noise conditions present in real experimental data.

2.2.3 Probabilistic Inference: MCMC versus Variational Methods

Extracting the complex, high-dimensional posterior defined by a hierarchical spectral model requires a robust probabilistic inference engine. The choice between exact and approximate inference involves a fundamental trade-off between geometric fidelity and computational cost.

MCMC methods generate asymptotically exact samples from the target distribution [13, 1]. However, standard random-walk proposals lead to slow mixing and high autocorrelation in the highly correlated posteriors arising from overlapping diffraction peaks. Hamiltonian Monte Carlo exploits gradient information to simulate physical trajectories through the parameter space, suppressing random-walk behavior. The **NUTS** [14] extends this by using a recursive algorithm to adaptively determine the number of integration steps, halting automatically when the simulated trajectory begins to reverse direction. **NUTS** serves as the primary inference engine for the teacher model, providing the asymptotically exact reference posteriors required for neural distillation.

Stochastic Variational Inference offers a fast alternative by reformulating posterior computation as an optimization problem. **ADVI** [17] makes this broadly applicable by optimizing the **Evidence Lower Bound (ELBO)** via stochastic gradient ascent. While **ADVI** reduces fitting time from hours to minutes, standard **ADVI** relies on a mean-field Gaussian approximating family that factorizes the posterior across all parameters. This systematically underestimates posterior variance and destroys the cross-parameter correlations that are physically critical in spectral deconvolution (such as the coupling between **FWHM**

asymmetry and peak mixing ratios). Therefore, while [ADVI](#) is highly valuable as a rapid diagnostic tool during model development, exact inference via [NUTS](#) is strictly mandated for the final teacher model to preserve the physical geometry of the posterior.

2.3 Simulation-Based Inference and Normalizing Flows

The hierarchical Bayesian model described in [section 2.2](#) produces high-fidelity reference posteriors, but at a computational cost that is fundamentally incompatible with operational deployment. Fitting the teacher model to a single new spectrum requires substantial time of [MCMC](#) sampling, whereas structural health monitoring demands near-instantaneous inference at deployment time. [SBI](#) provides the principled framework for resolving this tension by separating the concerns of inferential fidelity and computational efficiency across an offline training phase and an online inference phase. This section reviews the [SBI](#) literature, the [NPE](#) methods that operationalize it, and the normalizing flow architectures that implement the conditional density estimators required by the student model.

2.3.1 Simulation-Based Inference

Many scientific and engineering domains possess high-fidelity forward simulators that can generate synthetic observations from a given set of physical parameters, yet for which the associated likelihood function cannot be evaluated pointwise. This setting, commonly referred to as the implicit-likelihood or likelihood-free regime, renders standard [MCMC](#) and variational inference inapplicable in their canonical forms, since both require explicit likelihood evaluations. Historically, the standard resolution to this problem was [Approximate Bayesian Computation \(ABC\)](#). This approach operates by sampling parameters from a prior, simulating synthetic observations, and retaining only those parameters that produce simulations falling within a predefined distance tolerance of the empirical data. While conceptually straightforward, [ABC](#) requires enormous volumes of simulations, discards the vast majority of them, and suffers from severe scaling limitations in high-dimensional parameter spaces. Furthermore, [ABC](#) is a non-amortized method, meaning the entire costly sampling procedure must be repeated from scratch for every new observation. A comprehensive review of the [SBI](#) field identifies these primary shortcomings of classical approaches [6]: the large number of forward simulations required, the information loss introduced by reducing

data to low-dimensional summary statistics, and the per-observation computational waste incurred by non-amortized methods.

Amortized neural density estimation directly addresses the third limitation by training a neural network offline to approximate the posterior across all datasets consistent with the forward model [6]. After a one-time training phase, inference for any new observation reduces to direct sampling from the trained network, transforming the cost from $\mathcal{O}(T)$ iterative steps to $\mathcal{O}(1)$ forward passes. This amortization principle is the foundation of the teacher-student distillation framework proposed in this work.

2.3.2 Neural Posterior Estimation

NPE is the branch of **SBI** that directly targets the posterior $p(\mathbf{z} \mid \mathbf{x})$ by training a parametric conditional density estimator on simulated $(\mathbf{z}, \mathbf{x})$ pairs. The first **NPE** method was introduced by replacing the accept-reject kernel of **ABC** with a neural network trained to approximate the full posterior density [26]. This approach eliminates the tolerance hyperparameter ε that governs **ABC** approximation quality, concentrates simulations on the posterior-relevant region of parameter space via sequential proposals, and requires significantly fewer model evaluations than **ABC**. Early implementations employed **Mixture Density Networks (MDNs)** as the density estimator, but these networks possess limited expressive capacity and struggle to recover the complex, highly correlated geometries of physical posteriors.

The subsequent generation of **NPE** methods addressed this by introducing an importance-weighted training loss that allows the proposal distribution to differ from the prior while still targeting the correct posterior, without the need for post-hoc normalization [19]. However, importance-weighting introduces high-variance gradients that destabilize training as the proposal diverges from the prior. This instability was resolved with **Automatic Posterior Transformation (APT)**, also identified in the literature as the culmination of Sequential **NPE** [11]. **APT** derives a reparameterized training objective that remains valid under arbitrary, dynamically updated proposal distributions without relying on importance weights, enabling the stable incorporation of expressive normalizing flow architectures as the posterior estimator family.

The practical realization of globally amortized **NPE** for complex scientific models was demonstrated in the BayesFlow framework [33]. BayesFlow pairs an invertible inference network with a jointly trained summary network that learns to compress raw observational data into informative embeddings, eliminating the historical reliance on hand-crafted summary statistics. This architecture trains the networks to approximate the posterior across

all plausible datasets from a given generative model, enabling rapid posterior inference for any new observation after a single offline training phase. The teacher-student distillation framework instantiates this amortization principle and architectural strategy in a novel setting. The proposed student utilizes a 1D residual convolutional encoder to embed raw spectra. However, instead of relying on a conventional forward simulator, the teacher is a fully specified hierarchical Bayesian model fit to real experimental data, and the student is trained to reproduce the teacher’s reference posteriors to bridge the simulation-to-reality gap.

2.3.3 Normalizing Flows and Conditional Density Estimation

To overcome the limited expressiveness of Mixture Density Networks, the conditional density estimators used in [NPE](#) are now most commonly implemented as normalizing flows. These generative models provide a principled mathematical framework for exact log-density evaluation by constructing complex probability distributions through the composition of a sequence of invertible transformations applied to a simple base density [\[34\]](#). Formally, let $\mathbf{u} \sim p_{\mathbf{u}}(\mathbf{u})$ be a base random variable, typically a standard Gaussian, and let $f : \mathbb{R}^D \rightarrow \mathbb{R}^D$ be an invertible, differentiable transformation. The density of the transformed variable $\mathbf{z} = f(\mathbf{u})$ is given by the change-of-variables formula:

$$\log p_{\mathbf{z}}(\mathbf{z}) = \log p_{\mathbf{u}}(f^{-1}(\mathbf{z})) + \log \left| \det \frac{\partial f^{-1}}{\partial \mathbf{z}} \right|, \quad (2.3)$$

where the log-absolute Jacobian determinant accounts for the volume change induced by the transformation [\[34\]](#). By composing F such transformations, a flow of arbitrary expressiveness can be constructed. The expressiveness of the resulting distribution is determined by the choice of transformation family, while tractability requires that both f^{-1} and the Jacobian determinant are efficiently computable [\[27\]](#).

The autoregressive flow family has been identified as the most expressive tractable class of transformations for continuous density estimation [\[27\]](#). In an autoregressive flow, each output dimension z_d is generated as a deterministic function of u_d and all preceding dimensions $z_{1:d-1}$. This factorization produces a lower-triangular Jacobian whose determinant is the product of its diagonal entries, making the log-determinant computation linear in dimensionality D rather than cubic [\[27\]](#).

2.3.4 Masked Autoregressive Flow and MADE

The architectural prerequisite for an autoregressive normalizing flow is an efficient neural network that enforces the autoregressive ordering constraint. [Masked Autoencoder for Distribution Estimation \(MADE\)](#) was introduced to impose this constraint through a binary masking scheme applied to the weight matrices of a standard feedforward network [9]. Each output unit $\hat{z}_d$ is guaranteed to receive information only from input units $z_{1:d-1}$ by zeroing all weight connections that would otherwise violate the ordering. [MADE](#) therefore produces the full set of conditional distributions $\{p(z_d \mid z_{1:d-1})\}_{d=1}^D$ in a single forward pass, making it significantly faster than alternative autoregressive architectures that require D sequential passes [9].

Building directly on [MADE](#), the [Masked Autoregressive Flow \(MAF\)](#) constructs a normalizing flow by stacking multiple [MADE](#) blocks, each modeling the random noise variables used internally by the preceding block [28]. This results in a composed transformation whose Jacobian is lower-triangular by construction, enabling exact and efficient log-density evaluation in a single forward pass through all layers. By permuting the autoregressive ordering between successive [MADE](#) blocks, the flow ensures that all pairs of latent dimensions interact across the depth of the network, preventing the composite transformation from factorizing into independent subgroups [28]. This provides the precise architectural basis for the [cMAF](#) student model in this thesis. A critical design trade-off distinguishes [MAF](#) from its architectural counterpart, the [Inverse Autoregressive Flow \(IAF\)](#), [MAF](#) supports fast density evaluation but requires D sequential passes for sampling, while [IAF](#) supports fast sampling but requires D sequential passes for density evaluation [27]. Since the student model’s primary task during training is to evaluate and maximize the log-density of the teacher’s reference posteriors, [MAF](#) is the architecturally appropriate choice.

The student model further requires that the flow be conditioned on the observed [XRD](#) spectrum so that it models the conditional posterior $p(\mathbf{z} \mid \mathbf{x})$ rather than a marginal density. [Conditional Normalizing Flows \(CNFs\)](#), in which the base-density-to-output mapping is conditioned on an auxiliary input, have been shown to support exact likelihood-based training, efficient conditional sampling, and stable optimization without mode collapse [41]. In the [cMAF](#) student proposed in this work, conditioning is implemented by incorporating the dense spectrum embedding $\mathbf{h}$, produced by the 1D residual convolutional encoder, into the affine parameters of each [MADE](#) block. This allows the flow’s learned bijection to vary as a function of the input spectrum, producing a distinct posterior geometry for each observed diffraction pattern.

2.4 Summary

This chapter has reviewed the theoretical foundations and related literature across three distinct domains required for this research. First, the physical principles of [XRD](#) and fatigue damage evolution establish that peak morphology contains critical, physically grounded signatures of structural degradation. However, current data-driven approaches struggle with data scarcity, lack of interpretability, and the absence of physical constraints. Second, hierarchical Bayesian spectral deconvolution resolves these limitations by enforcing crystallographic priors and quantifying inferential uncertainty through exact [MCMC](#) sampling, but at a computational cost that prohibits real-time deployment. Finally, [SBI](#), powered by autoregressive normalizing flows, offers a mathematically accurate mechanism to amortize this computational burden, transforming iterative sampling into direct sampling from a trained neural network. The intersection of these three domains defines the methodological gap addressed in this work. The following chapters detail how this work operationalizes this intersection through a physics-guided teacher-student framework, distilling the exact, empirically grounded posteriors of a hierarchical Bayesian teacher into a [cMAF](#) student.

Chapter 3

Problem Formulation

We formulate the problem as a supervised classification task where each specimen is represented by multiple spectral measurements generated by a latent physics-based mechanism. This is motivated by the structural commonality found in scientific functional data; modalities such as Raman spectroscopy, [Fourier-transform infrared spectroscopy \(FTIR\)](#), [Mass Spectrometry \(MS\)](#), and [XRD](#) all share a fundamental additive parametric form, defined by the superposition of background, signal peaks, and noise. The problem is decomposed into a spectral inverse problem and a downstream classification task, and [section 3.5](#) motivates this decomposition over a direct mapping from spectra to labels.

3.1 Notation

[Table 3.1](#) summarizes the core mathematical symbols used throughout this thesis. Quantities defined here are referenced consistently across the chapters.

3.2 Grouped Sample Structure

Let there be S distinct physical sources (e.g., specimens or subjects) indexed by $s \in \{1, \dots, S\}$. We observe a total of N spectral measurements (scans) indexed by $i \in \{1, \dots, N\}$. Since multiple measurements may originate from the same physical source, we define a known grouping map $g : \{1, \dots, N\} \rightarrow \{1, \dots, S\}$, where $g(i) = s$ indicates that measurement i was collected from source s . Each scan i carries its own class label

Table 3.1: Summary of core mathematical notation.

Symbol	Description
S	Number of distinct physical sources (specimens)
N	Total number of spectral measurements (scans)
M_i	Number of sampled domain points in scan i
$g(i)$	Grouping map assigning scan i to its source specimen
$\mathbf{x}_i$	Observed spectral measurement for scan i
$\mathbf{z}_i$	Latent physical parameters for scan i
y_i	Class label for scan i
K	Number of latent spectral components (peaks)
$\mathcal{T}$	Continuous spectral domain (e.g., diffraction angle)

y_i , which reflects the state of its source at the time of measurement, so a single source can contribute scans to more than one class. Scans from the same source are not independent observations, since they are measurements of the same physical object and share source-specific characteristics.

3.3 Spectral Observations

Each observation $\mathbf{x}_i$ consists of discrete measurements sampled over a continuous spectral domain $\mathcal{T}$ (e.g., diffraction angle, frequency, or time). We represent these as ordered pairs of domain coordinates and response values,

$$\mathbf{x}_i := \{(t_j^{(i)}, v_j^{(i)})\}_{j=1}^{M_i}, \quad (3.1)$$

where $t \in \mathcal{T}$ is the domain coordinate, $v \in \mathbb{R}$ is the measured response intensity, and M_i is the number of sampled points, which may vary across observations due to experimental constraints.

3.4 Inverse Problem

We assume the data generation process follows a physics-based additive structure common to spectroscopic techniques. As illustrated in [Figure 3.1](#), the observed signal is composed

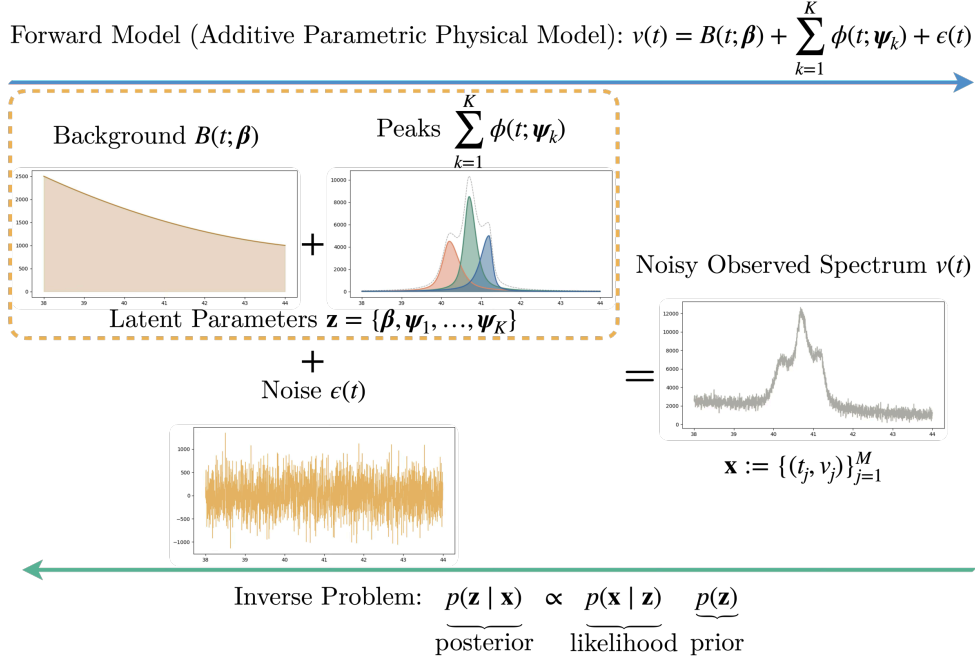


Figure 3.1: The forward model (top) constructs a spectrum $v(t)$ by superimposing a parameterized background $B(t; \boldsymbol{\beta})$, localized peak profiles $\phi(t; \boldsymbol{\psi}_k)$, and noise $\epsilon(t)$. The corresponding inverse problem (bottom) seeks to infer the full posterior distribution $p(\mathbf{z} | \mathbf{x})$ over the latent physical parameters $\mathbf{z}$ from the noisy spectral observation $\mathbf{x}$.

of a slowly varying background and a superposition of localized peaks,

$$v_i(t) = \underbrace{B(t; \boldsymbol{\beta}_i)}_{\text{Background}} + \sum_{k=1}^K \underbrace{\phi(t; \boldsymbol{\psi}_{i,k})}_{\text{Peak } k} + \underbrace{\epsilon_i(t)}_{\text{Noise}}, \quad (3.2)$$

where B is a baseline function parameterized by $\boldsymbol{\beta}_i$, ϕ is a peak profile function (e.g., [pV](#)) parameterized by $\boldsymbol{\psi}_{i,k}$, and $\epsilon_i(t)$ represents measurement noise. The complete set of latent physical parameters for scan i is denoted $\mathbf{z}_i = \{\boldsymbol{\beta}_i, \boldsymbol{\psi}_{i,1}, \dots, \boldsymbol{\psi}_{i,K}\}$.

The inverse problem requires inferring plausible parameter values $\mathbf{z}_i$ from the observed measurement $\mathbf{x}_i$, formalized via Bayes' theorem,

$$\underbrace{p(\mathbf{z}_i | \mathbf{x}_i)}_{\text{posterior}} \propto \underbrace{p(\mathbf{x}_i | \mathbf{z}_i)}_{\text{likelihood}} \underbrace{p(\mathbf{z}_i)}_{\text{prior}}. \quad (3.3)$$

Here, the *likelihood* $p(\mathbf{x}_i | \mathbf{z}_i)$ quantifies how well the parameters explain the observed

spectrum via Equation 3.2, while the *prior* $p(\mathbf{z}_i)$ encodes physical constraints (e.g., non-negativity, crystallographic bounds). The resulting *posterior* $p(\mathbf{z}_i \mid \mathbf{x}_i)$ captures the uncertainty over the physical state given the noisy observations.

3.5 Problem Statement

Each scan i is associated with a class label $y_i \in \{1, \dots, C\}$ (e.g., nominal versus damaged). We assume the label is a function of the latent physical parameters $\mathbf{z}$ rather than of raw signal artifacts. Given the dataset $\mathcal{D} = \{(\mathbf{x}_i, y_i, g(i))\}_{i=1}^N$, the objective is twofold:

1. Solve the inverse problem to estimate the latent parameters $\mathbf{z}_i$ (or their posteriors) for each scan i .
2. Learn a mapping $f : \mathbf{z} \rightarrow \{1, \dots, C\}$ that generalizes to unseen sources.

The classification task can be approached in two ways, as illustrated in Figure 3.2. In the direct formulation, a single model $c : \mathbf{x} \rightarrow \{1, \dots, C\}$ maps the observed spectrum to the class label and treats the physical process that generated the data as a black box. In the two-stage formulation adopted in this thesis, the inverse problem is solved first to recover the posterior $p(\mathbf{z}_i \mid \mathbf{x}_i)$ over the latent physical parameters, and a separate mapping $f : \mathbf{z} \rightarrow \{1, \dots, C\}$ then predicts the class label from these parameters. The two-stage formulation is preferred for the following reasons:

- **Scientific Meaning:** The latent parameters $\mathbf{z}_i$ describe the physical state of the source and are of scientific interest independently of the classification task.
- **Interpretability:** Classification decisions can be traced back to physically meaningful parameters rather than to raw spectral features.
- **Incorporation of Physical Knowledge:** Physical constraints enter the inverse problem through the forward model and the priors.
- **Uncertainty Quantification:** The posterior quantifies the uncertainty of each inferred parameter, which can be propagated to the downstream prediction.
- **Validation:** The inferred parameters can be checked against the observed data through the forward model and against established physical knowledge.

- **Reusability:** The same inferred parameters can serve other downstream tasks without solving the inverse problem again.
- **Low-Dimensional Representation:** The classifier operates on a small set of physically structured parameters instead of the full spectrum.

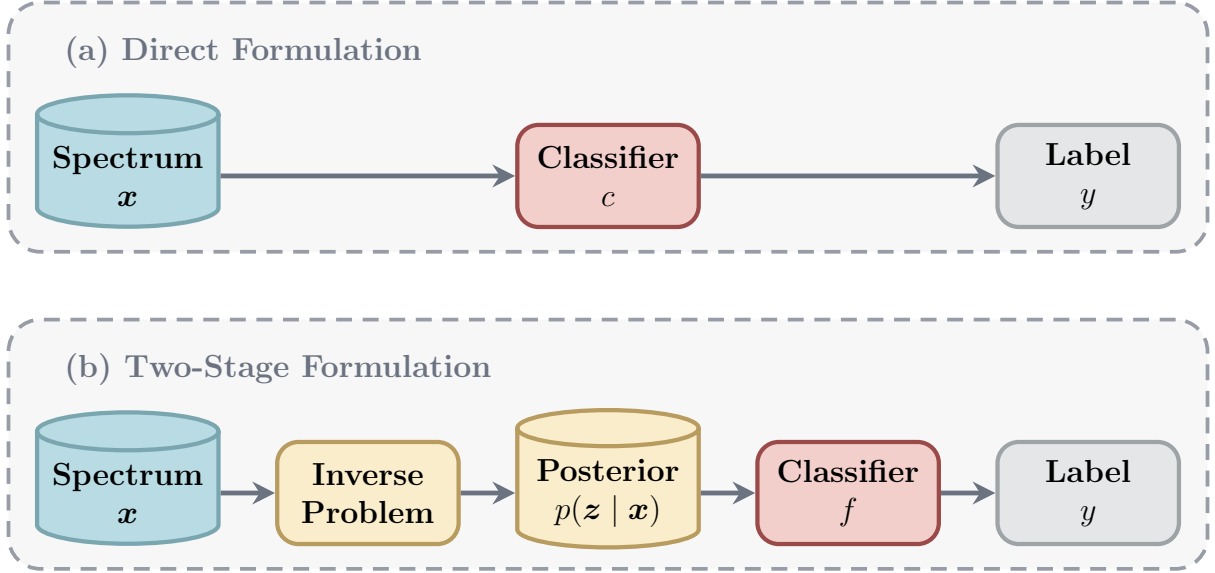


Figure 3.2: (a) The direct formulation maps the observed spectrum to the class label with a single model. (b) The two-stage formulation first solves the inverse problem for the posterior over the latent physical parameters and then predicts the class label from these parameters.

3.6 Key Challenges

The inverse problem of Equation 3.3 is high-dimensional with overlapping peaks and noise, necessitating strong regularization via informative priors to ensure identifiable solutions. Solving this ill-posed problem requires addressing several simultaneous statistical and computational constraints common to scientific settings:

Data Scarcity: High-fidelity scientific datasets are constrained to the small-data regime ($N \sim 10^2$) due to the prohibitive cost of acquisition (e.g., hours of synchrotron beam time or mechanical cycling per sample).

Grouped Distribution Shift: Multiple scans from the same source share source-specific signatures; proper evaluation therefore requires grouped cross-validation to assess generalization to unseen sources rather than to measurement replicates.

Uncertainty Quantification: Reliable decision-making in safety-critical domains requires not only accurate predictions but also calibrated confidence estimates.

Inference Latency: Deployment in high-throughput production environments necessitates near-instantaneous inference, precluding expensive iterative optimization techniques such as [MCMC](#) at test time.

Interpretability: The decision boundary must be explainable in terms of physically meaningful parameters (e.g., peak shifts or asymmetry) rather than black-box features.

3.7 Summary

This chapter has established the formal mathematical setting for the thesis. Spectral measurements are modeled as grouped observations of a shared physics-based generative process, with class labels assigned to individual scans and scans grouped by their source specimen. The core inference task is to recover the posterior over latent physical parameters $\mathbf{z}_i$ from each observation $\mathbf{x}_i$, subject to the ill-posedness, data scarcity, and grouped dependence structure identified above. The classification task is formulated in two stages, in which the inverse problem is solved first and the class label is then predicted from the inferred physical parameters. The following chapter introduces the dataset used to instantiate this formulation.

Chapter 4

Data Acquisition and Preprocessing

This chapter instantiates the mathematical framework established in [chapter 3](#) using experimental observations of AA6061-T6 aluminum alloy. It covers the material science principles that motivate the measurement approach, the experimental protocol used to generate labeled specimens, the construction and preprocessing of the dataset, and an exploratory analysis of the resulting spectral observations.

4.1 Material Science Background

The fundamental mechanism for non-destructive microstructural assessment relies on the interaction of X-rays with the crystal lattice, governed by Bragg’s Law ([Figure 4.1](#)). By monitoring diffraction patterns, fatigue-induced damage can be quantified through the evolution of peak profiles. While peak shifts are used to measure macrostresses, broadening of the diffraction profile (quantified via the [FWHM](#), half-value breadth, or integral breadth) characterizes micro-plasticity and the accumulation of dislocation density [[36](#), [20](#)].

This broadening serves as a direct indicator of damage evolution and follows a characteristic three-stage process [[37](#)]. Diffraction breadths undergo a rapid increase during the initial 15% to 20% of total fatigue life, corresponding to rapid work hardening and a rise in dislocation density. A stable plateau phase follows for the majority of the cyclic life, before a final sharp increase in broadening occurs immediately prior to failure [[23](#), [24](#)]. By correlating measured changes in peak breadth with applied strain and cyclic history, analytical models have been developed to predict the remaining fatigue life of materials including aluminum 2024-T3 and steel pipelines [[42](#), [16](#), [32](#), [31](#), [7](#)]. More recently, [XRD](#) analysis has

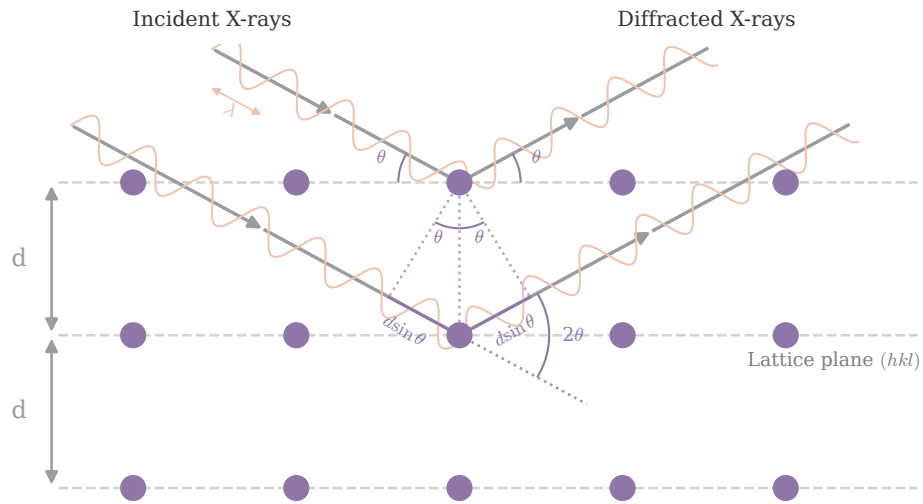


Figure 4.1: Geometric illustration of Bragg's Law. Incident X-rays scatter off parallel atomic planes separated by interplanar spacing d . Constructive interference occurs when the path difference $2d \sin \theta$ is an integer multiple of the wavelength λ , expressed as $n\lambda = 2d_{hkl} \sin \theta$. The total diffraction angle recorded by the detector is 2θ .

been demonstrated to capture early signs of structural damage in AA6061-T6 aluminum alloy components and to predict fracture initiation under uniaxial fatigue loading [10]. A representative XRD measurement from an AA6061-T6 specimen is shown in Figure 4.2, illustrating both the raw 2D Debye–Scherrer diffraction rings and the 1D spectral profile $I(2\theta)$.

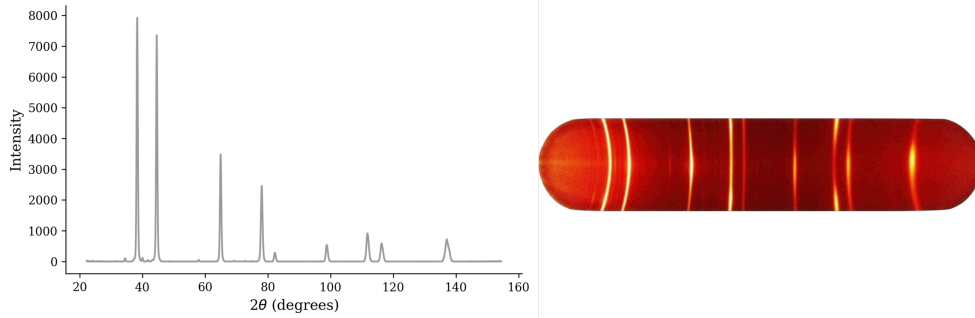


Figure 4.2: (left) 1D spectral profile $I(2\theta)$ obtained from an AA6061-T6 specimen. (right) Representative Debye–Scherrer diffraction rings collected from the same specimen.

4.2 Data Generation

The physical experimental workflow was carried out in three sequential phases: specimen acquisition, damage induction via mechanical cycling, and XRD characterization. Figure 4.3 provides a conceptual overview of this integrated pipeline.

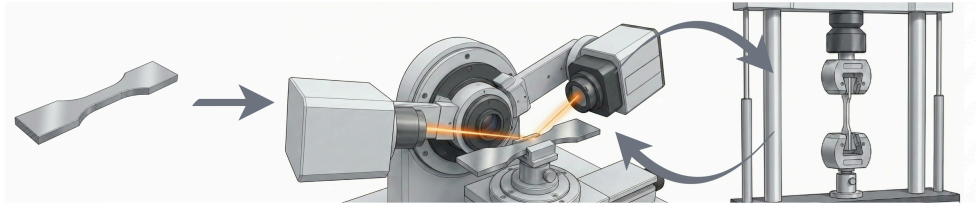


Figure 4.3: Overview of the experimental data generation workflow. Standard dog-bone specimens (left) are subjected to interrupted uniaxial fatigue cycling in a servo-hydraulic load frame (right). At prescribed life fractions (0%, 50%, and 75% N_f), specimens are characterized using X-ray diffraction (center) to capture the evolution of lattice microstructure and diffraction peak broadening.

4.2.1 Specimen Acquisition

Fatigue specimens were precision-machined from 25.4 mm extruded AA6061-T6 bars into standard dog-bone geometry (Figure 4.4). This alloy was selected for its stable and well-characterized FCC microstructure, which yields a predictable set of diffraction peaks amenable to physics-informed modeling. Specimen dimensions were designed to ensure consistent failure modes localized within the gauge length.

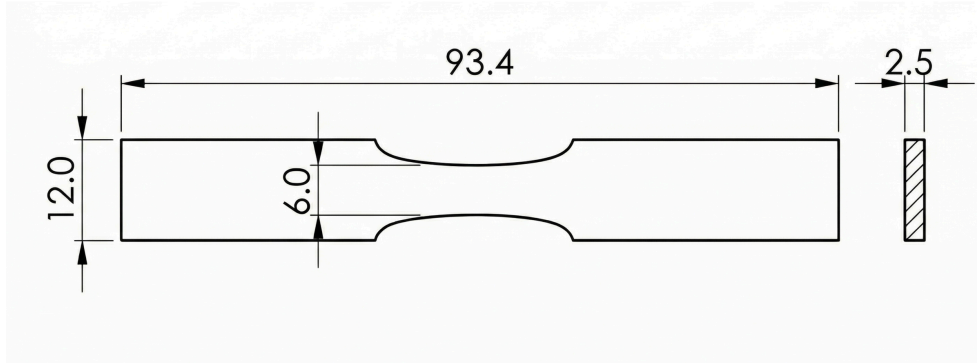


Figure 4.4: Geometry of the dog-bone fatigue specimens (units in mm).

4.2.2 Damage Induction

Specimens were subjected to stress-controlled uniaxial fatigue cycling using an MTS 810 servo-hydraulic system, operated in fully reversed mode ($R = -1$, ± 50 kN) at a stress amplitude of 200 MPa. Fatigue cycling was interrupted at prescribed fractions of the total fatigue life N_f to produce three labeled physical states:

1. *Pristine* (0% N_f): the undamaged baseline, prior to any mechanical cycling.
2. *Mid-Life* (50% N_f): intermediate damage accumulation, corresponding to the stable plateau phase of the fatigue breadth evolution.
3. *Late-Life* (75% N_f): advanced damage approaching fracture initiation, corresponding to the onset of the final breadth increase.

4.2.3 XRD Characterization

Lattice microstructures were scanned using a Bruker D8-Discover diffractometer equipped with a VÅNTEC-500 area detector (135 mm radius). The X-ray tube operated at 40 kV and 40 mA utilizing Cu $K\alpha$ radiation ($\lambda = 1.5406 \text{ \AA}$). Each scan was produced by integrating 24 frames at 40 s exposure per frame, yielding a high signal-to-noise ratio suitable for precise peak-profile characterization. The diffraction angle range covered was $2\theta \in [22.22^\circ, 154.45^\circ]$. In the spectral model, the domain coordinate t corresponds to the diffraction angle 2θ , and the response v corresponds to the measured photon intensity counts I .

4.3 Dataset Description and Characteristics

Table 4.1 summarizes the key properties of the assembled dataset. The full dataset comprises $N = 106$ XRD scans collected from $S = 8$ distinct physical specimens across the three damage states described in subsection 4.2.2. Scans were acquired in the pristine state and at several stages of fatigue cycling, and this work uses the pristine scans and the scans acquired at 50 % and 75 % of the fatigue life. Each scan is labeled by the damage state of the specimen at the time of measurement.

Binary Label Reformulation. The three physical states are collected to track the progression of fatigue damage across the specimen lifecycle. For the downstream classification task, however, the objective is to distinguish structurally sound specimens from those exhibiting any degree of fatigue-induced damage. Accordingly, the Mid-Life and Late-Life states are merged into a single Damaged class, and the Pristine state forms the Nominal class. This binary formulation reflects the practical decision boundary relevant to structural health monitoring, determining whether a component has sustained microstructural damage, rather than staging its precise degree of progression.

4.4 Preprocessing and Grid Alignment

Raw spectral measurements arrive as variable-length sequences $\mathbf{x}_i = \{(t_j^{(i)}, v_j^{(i)})\}_{j=1}^{M_i}$ with heterogeneous domain supports. To enable vectorized Bayesian and neural inference, all observations are standardized onto a uniform coordinate grid spanning the global overlapping support $\mathcal{T}_{\text{common}} = [t_{\min}, t_{\max}]$. The target grid resolution is adaptively determined

Table 4.1: Dataset summary. Three physical damage states are collected experimentally and collapsed into two classes for the downstream binary classification task.

Property	Value
Specimens (S)	8
Total scans (N)	106
<i>Physical damage states (collected)</i>	
Pristine (0 % N_f)	51 scans
Mid-Life (50 % N_f)	28 scans
Late-Life (75 % N_f)	27 scans
<i>Binary classification labels (downstream task)</i>	
Nominal (Pristine)	51 scans
Damaged (Mid + Late-Life)	55 scans
<i>XRD acquisition parameters</i>	
Diffraction angle range	$2\theta \in [22.22^\circ, 154.45^\circ]$
Radiation source	Cu $K\alpha$, $\lambda = 1.5406 \text{ \AA}$
Detector	Bruker VÅNTEC-500, 135 mm radius

from the empirical step sizes of the raw scans, bounded by a maximum dimension of $L_{\max} = 4,000$ points.

Response intensities $v^{(i)}$ are projected onto this grid via [Piecewise Cubic Hermite Interpolating Polynomial \(PCHIP\)](#) interpolation. Unlike standard cubic splines, [PCHIP](#) explicitly preserves signal monotonicity, guaranteeing topological fidelity and strictly preventing the introduction of non-physical oscillatory artifacts around sharp spectral peaks.

[Figure 4.8](#) shows the resulting mean spectrum and $\pm 1\sigma$ inter-scan variability envelope across all $N = 106$ aligned observations.

The resulting aligned dataset serves as the common input representation for both the Bayesian and neural inference pipelines described in [chapter 5](#).

4.5 Exploratory Data Analysis

Prior to model development, an exploratory analysis of the spectral dataset was conducted to characterize its structure, variability, and class-level properties. Each subsection below addresses a specific empirical question about the data.

4.5.1 Spectral Overview

Figure 4.5 illustrates the 2θ domain coverage for all $N = 106$ scans, grouped and color-coded by specimen. Every observation spans the full acquisition range $\mathcal{T}_{\text{common}} = [22.2^\circ, 154.4^\circ]$ without truncation or missing segments, confirming that grid alignment via PCHIP interpolation is lossless across the entire dataset. Dashed vertical lines mark the global domain boundaries $t_{\min} = 22.2^\circ$ and $t_{\max} = 154.4^\circ$ that define the common grid $\mathcal{T}_{\text{common}}$.

Figure 4.6 shows the raw intensity profiles $I(2\theta)$ for all $S = 8$ specimens overlaid on a common axis, with dashed vertical lines indicating the theoretical Bragg positions for the $K = 9$ FCC-permitted reflections. Nine distinct diffraction peaks are present across the full 2θ range, with inter-specimen variation in absolute intensity visible across all reflections.

4.5.2 Intensity Distribution

Figure 4.7 shows the distribution of maximum photon intensity counts across all $N = 106$ scans. The distribution is right-skewed, with most scans falling below 8,000 counts and a tail extending to approximately 26,000 counts. The median maximum intensity is 6,288.7 counts.

4.6 Physics-Informed Structural Priors

A central computational challenge in spectral decomposition is determining the appropriate number of latent components K and separating them from background scattering. Without additional constraints, this constitutes an ill-posed problem. This ill-posedness is mitigated by introducing a structural prior grounded in crystallographic principles: Bragg’s Law ($n\lambda = 2d_{hkl}\sin\theta$) combined with FCC reflection selection rules strictly permits exactly $K = 9$ fundamental reflections for this material system, corresponding to Miller indices (111), (200), (220), (311), (222), (400), (331), (420), and (422).

Peak position priors $\mu_{\text{theory},k}$ are initialized to the corresponding theoretical Bragg angles. As confirmed in Figure 4.8, these theoretical positions align with the observed mean peak locations across all specimens and damage states, validating the crystallographic prior. To further inspect the alignment of lower-intensity reflections and the variance of the background signal, the mean intensity profile is presented on a logarithmic scale in Figure 4.9. Enforcing these constraints ensures that the resulting models remain physically valid and interpretable.

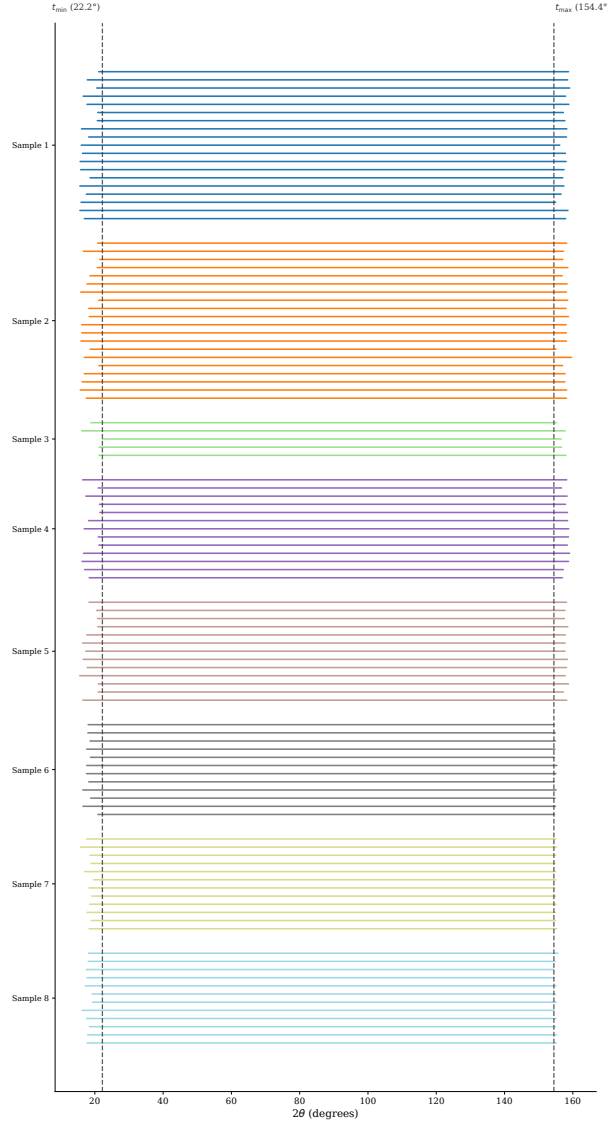


Figure 4.5: Domain coverage of all $N = 106$ XRD scans, grouped and color-coded by specimen (Samples 1–8). Each horizontal bar represents one scan. Dashed vertical lines indicate the global domain boundaries $t_{\min} = 22.2^\circ$ and $t_{\max} = 154.4^\circ$.

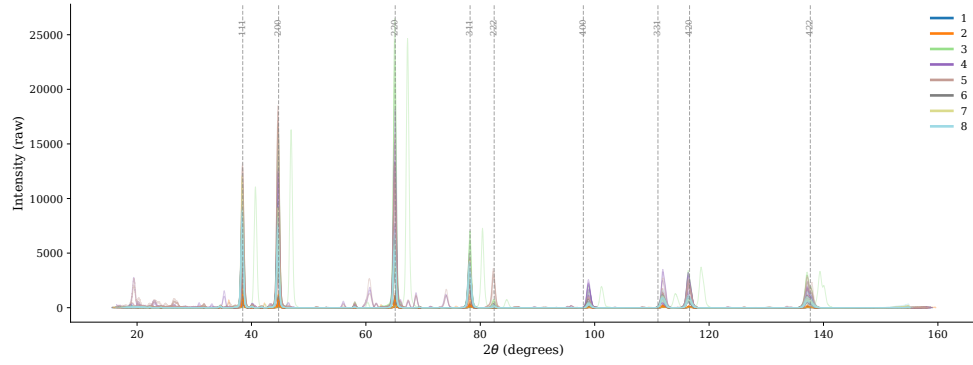


Figure 4.6: Raw XRD spectral profiles $I(2\theta)$ for all $S = 8$ specimens, color-coded by specimen (Samples 1–8). Dashed vertical lines indicate the theoretical Bragg positions for the $K = 9$ FCC-permitted reflections, with Miller indices labeled.

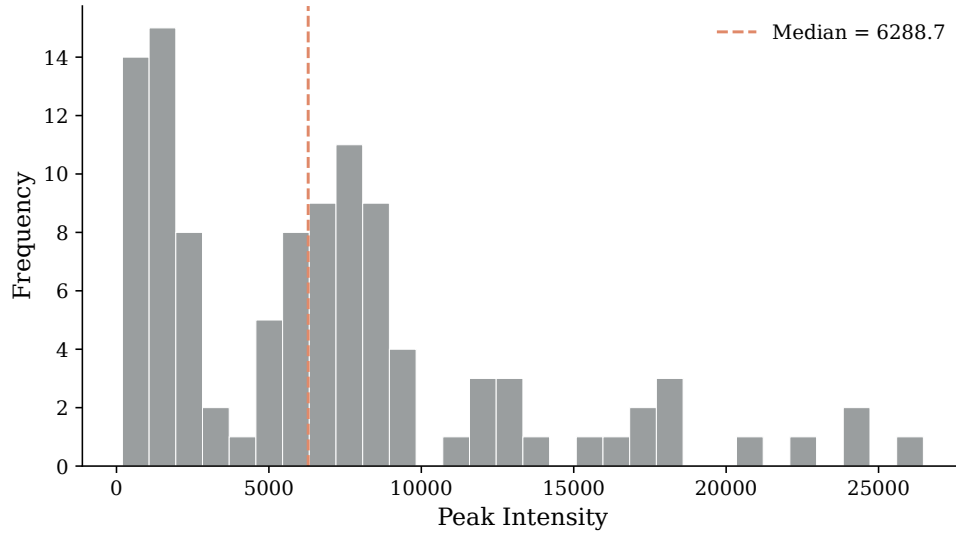


Figure 4.7: Distribution of maximum photon intensity counts across all $N = 106$ XRD scans. The dashed vertical line indicates the median value of 6,288.7 counts.

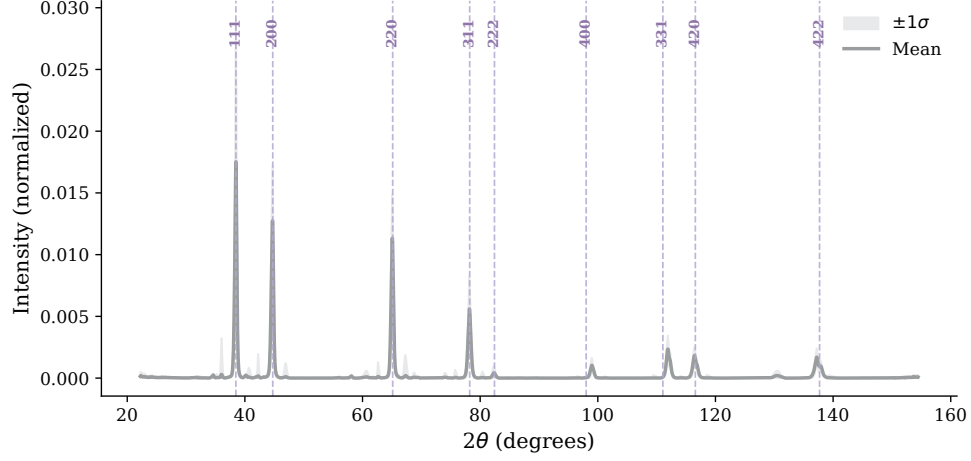


Figure 4.8: Mean XRD spectrum and $\pm 1\sigma$ inter-scan variability across all $N = 106$ aligned observations. Theoretical Bragg positions for the $K = 9$ permitted FCC reflections are overlaid as dashed vertical lines, with Miller indices labeled to indicate specific crystallographic planes.

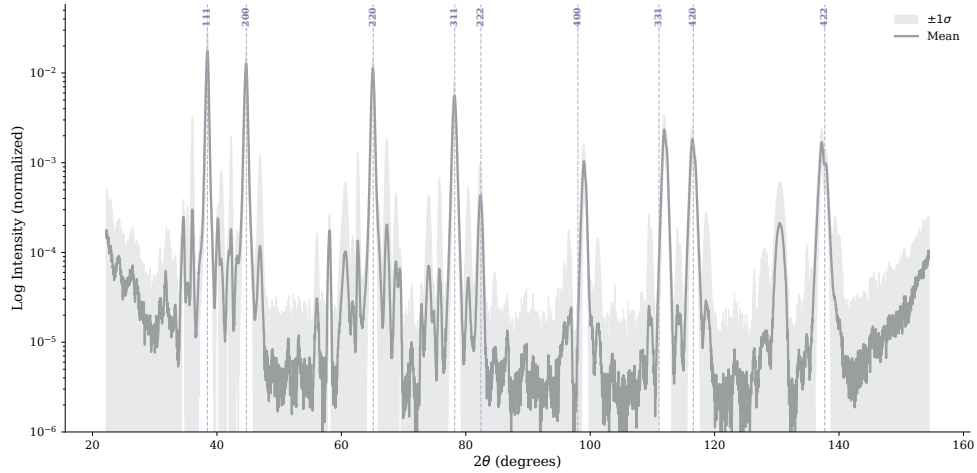


Figure 4.9: Log-scaled mean XRD intensity profile and $\pm 1\sigma$ variability envelope. The logarithmic representation highlights the high signal-to-noise ratio and confirms the presence of lower-intensity FCC reflections at higher 2θ angles, consistent with theoretical Miller indices.

4.6.1 Fatigue Damage Evolution

To examine the direct observable effects of mechanical cycling on the crystalline lattice, a localized analysis of the $K = 9$ fundamental FCC reflections was performed. Figure 4.10 presents the mean normalized intensity profiles extracted over a $\pm 3^\circ$ window centered on each theoretical Bragg angle, stratified by damage state: Pristine (0% N_f), Mid-Life (50% N_f), and Late-Life (75% N_f).

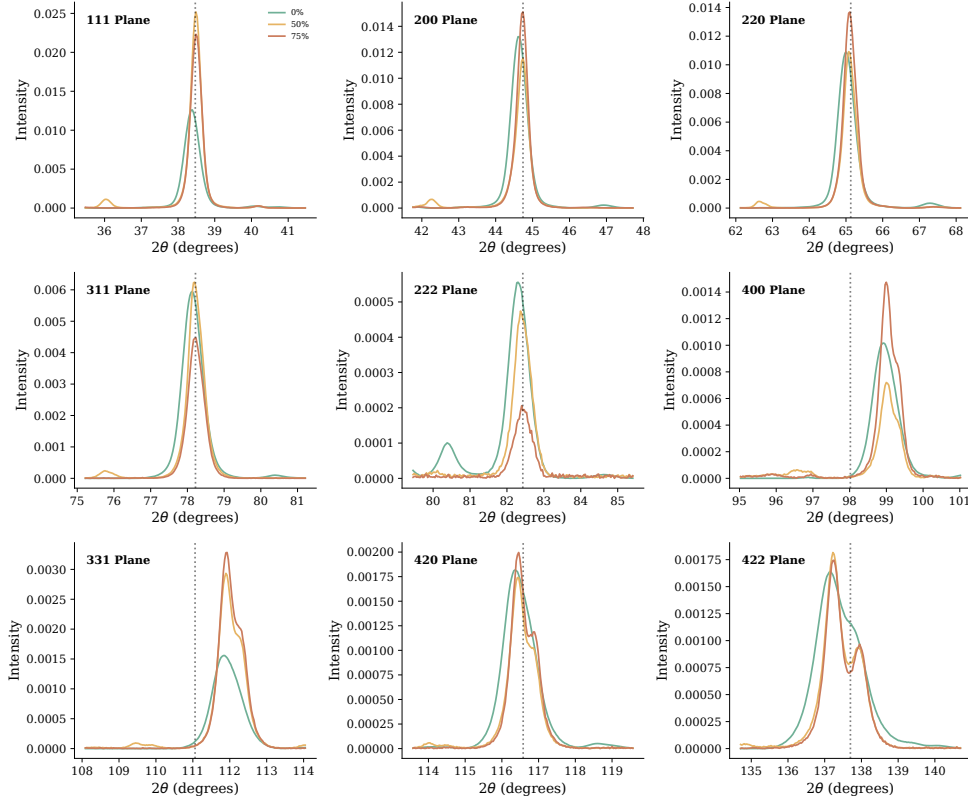


Figure 4.10: Mean normalized XRD intensity profiles in a $\pm 3^\circ$ window around each of the $K = 9$ FCC-permitted Bragg reflections for AA6061-T6, stratified by damage state: Pristine (0% N_f), Mid-Life (50% N_f), and Late-Life (75% N_f). The progressive evolution of peak breadth and amplitude reflects the accumulation of fatigue-induced microstructural damage. Dotted vertical lines indicate the theoretical Bragg positions.

Consistent with the material science principles outlined in section 4.1, the accumulation of fatigue-induced micro-plasticity and dislocation density manifests directly in the

experimental data. These manifestations include variations in maximum peak intensity, localized profile broadening, and subtle peak shifts relative to the theoretical Bragg positions (indicated by the vertical dotted lines). This visually confirms that the structural degradation of the AA6061-T6 lattice is systematically encoded within the peak profile geometries, providing a sound empirical basis for the downstream classification task.

While Figure 4.10 examines localized distortions at individual Bragg reflections, a macroscopic view of the full 2θ domain on a logarithmic scale reveals complementary global trends. Figure 4.11 presents the mean normalized intensity profiles for the three damage states across the complete acquisition range $\mathcal{T}_{\text{common}}$.

A systematic elevation of the inter-peak baseline is apparent in the Mid-Life and Late-Life spectra relative to the Pristine state. To maintain visual focus on resolvable physical signals, the y-axis lower bound is constrained to 10^{-6} .

This global background shift is consistent with an increase in diffuse, non-Bragg scattering. As fatigue-induced microstructural disorder accumulates, localized strains and defect networks disrupt the periodicity of the lattice, scattering incident X-rays incoherently. Taken together with the localized peak broadening, this trend confirms that fatigue damage produces measurable structural degradation across the full spatial domain of the spectral data, providing comprehensive empirical support for downstream modeling.

4.7 Summary

This chapter has introduced the dataset that instantiates the abstract problem formulation of chapter 3. The AA6061-T6 fatigue specimens provide $N = 106$ XRD scans from $S = 8$ distinct specimens across three damage states, reformulated into a binary Nominal versus Damaged classification task. The PCHIP preprocessing pipeline maps all variable-length raw scans onto a common 4,000-point grid without coverage loss or non-physical interpolation artifacts. The exploratory analysis establishes four empirical properties that directly shape subsequent modeling choices.

- Complete domain coverage enables a shared grid representation
- Observed peak-theory agreement validates physics-informed component constraints.
- Wide intensity dynamic range motivates per-scan normalization.
- Characteristics such as localized peak broadening confirm that damage is systematically encoded in the spectral data.

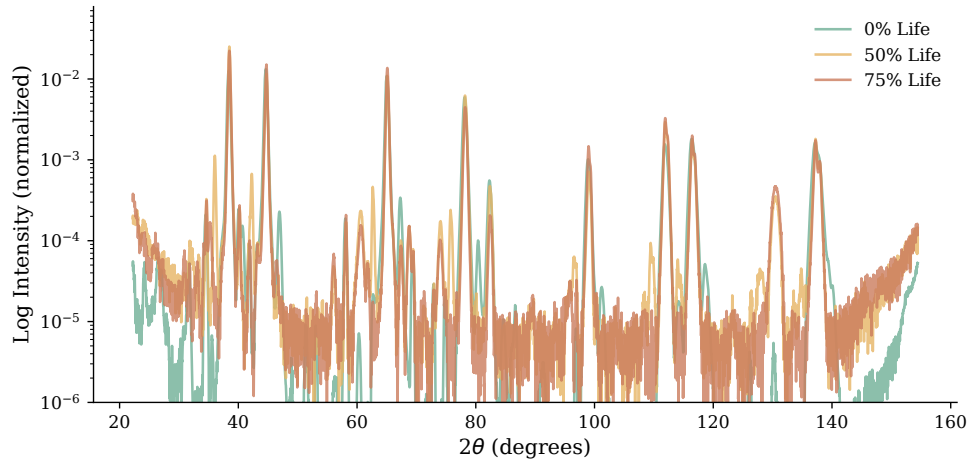


Figure 4.11: Mean normalized X-ray diffraction intensity profiles on a logarithmic scale, stratified by damage state: Pristine (0 % N_f), Mid-Life (50 % N_f), and Late-Life (75 % N_f). A systematic elevation of the inter-peak background baseline is apparent in the damaged states, indicative of increased diffuse scattering. The y-axis lower bound is set to 10^{-6} , consistent with the noise floor of the normalized spectra, to exclude near-zero interpolation artifacts.

The following chapter builds on these properties to develop the physics-guided teacher-student framework that performs inference over the latent parameters $\mathbf{z}_i$.

Chapter 5

Proposed Framework

This chapter details the theoretical architecture of the proposed solution, defining the abstract boundaries and mathematical formulations of the inference models before detailing their specific application.

5.1 Physics-Guided Teacher-Student Framework

We propose a unified framework that decomposes the spectral inverse problem into two complementary pipelines (Figure 5.1). All measurements first pass through a preprocessing module that aligns variable-length scans onto a standardized grid.

- **Bayesian Pipeline (Teacher):** To address data scarcity and ill-posedness, we fit a hierarchical Bayesian model via [MCMC](#) with physics-informed priors and partial pooling. This produces high-fidelity posteriors that serve as the ground truth for the neural pipeline.
- **Neural Pipeline (Student):** Bayesian knowledge is distilled in two stages. First, [MCMC](#) posteriors are used as an empirical joint distribution and expanded into a dense synthetic corpus via physics-aware augmentation. Second, a [cMAF](#) is trained on this corpus and fine-tuned on real data to bridge the simulation-to-reality gap. Together, these pipelines ensure the downstream task inherits accurate uncertainty quantification at near-instantaneous inference latency.

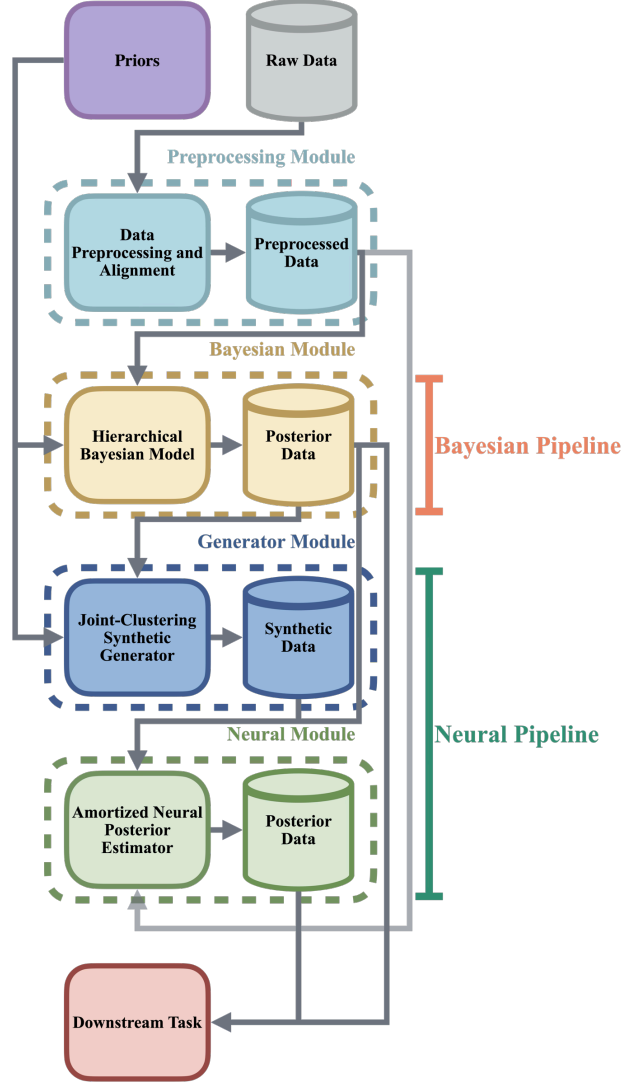


Figure 5.1: Four-module pipeline: **(1) Preprocessing:** Raw Data $\rightarrow$ Preprocessed Data. **(2) Bayesian Teacher:** Preprocessed Data + Priors $\rightarrow$ High-Fidelity Posteriors. **(3) Synthetic Generator:** Posteriors $\rightarrow$ Correlation-Preserving Synthetic Data. **(4) Neural Student:** Synthetic Data (pre-training) + Real Data (fine-tuning) $\rightarrow$ Amortized Posterior Estimator (cMAF).

5.2 Bayesian Spectral Deconvolution and Posterior Extraction

Spectral deconvolution is the process of decomposing an observed composite signal into its constituent components by recovering the latent parameters of a generative forward model. In the deterministic setting, this yields point estimates, which are prone to overfitting and provide no measure of parameter uncertainty. Bayesian spectral deconvolution reframes this as a probabilistic inference problem. Rather than seeking a single best-fit parameter vector, we infer the full posterior distribution $p(\mathbf{z}_i \mid \mathbf{x}_i)$ over the latent physical parameters given the observed spectrum, as formalized in [Equation 3.3](#). This yields calibrated uncertainty estimates that propagate through all downstream analyses.

When the dataset contains multiple scans drawn from distinct physical specimens, per-scan evidence is often insufficient to fully constrain the inference, particularly under low signal-to-noise conditions. A hierarchical extension addresses this by coupling the per-scan posteriors through shared population-level hyperpriors. This allows the model to borrow statistical strength across all N scans and regularize parameter estimates in regimes where individual scan evidence alone is insufficient.

Constructing a robust probabilistic model for [XRD](#) requires navigating a complex methodological design space. To establish the most physically accurate and computationally reliable inference pipeline, we formalize the available design options and systematically evaluate them to select the optimal model configuration.

5.2.1 Methodological Design Space

The architecture of the Bayesian teacher model is defined across four primary methodological dimensions.

Crystallographic Scope and Spectral Domain

The first set of design choices determines which reflections are included in the spectral model and over what angular range the forward model is fit. The choice of phase set and the choice of spectral fitting strategy are treated as independent axes where each phase configuration can be combined with either a full or reduced peak subset, a global or localized fitting window, or both.

- **Pure FCC Aluminum Matrix.** The simplest configuration restricts the latent components exclusively to the fundamental FCC reflections of the aluminum matrix. The structural prior is enforced by fixing $K = 9$ and centering the peak-position priors $\mu_{\text{theory},k}$ at the theoretical Bragg angles for the (111), (200), (220), (311), (222), (400), (331), (420), and (422) planes. This configuration isolates the contribution of the aluminum matrix and serves as the reference for subsequent ablations.
- **Extended Set Including Secondary Phases.** AA6061-T6 is not a single-phase system; it contains Mg_2Si precipitates and other intermetallic constituents inherent to the 6000-series alloy. An extended configuration augments the FCC peak set with additional reflections attributed to these secondary phases, instantiating the forward model with a superset of peaks while retaining the same additive structure and noise model. This allows the inferred posteriors to capture microstructural changes arising from both the aluminum matrix and the precipitate network.
- **Reduced Subset of Dominant Peaks.** Rather than fitting all reflections in the chosen phase set, this configuration restricts inference to the strongest peaks in the intensity profile. Focusing on the highest-intensity reflections reduces the dimensionality of the latent parameter space and evaluates whether a compact, high-confidence subset of peak-shape parameters retains sufficient information for the downstream task.
- **Localized Window Fitting.** In global fitting, the forward model must simultaneously account for all peaks and a smooth background that varies continuously across the full acquisition domain $\mathcal{T}_{\text{common}}$. In localized fitting, this complexity is reduced by restricting the fitting procedure to a narrow angular window centered on the theoretical Bragg position $\mu_{\text{theory},k}$ for each peak k . The window width is chosen to fully enclose each reflection while remaining narrow enough to exclude adjacent secondary-phase features. Within each window, a local instance of the forward model is fit, making the inferred parameters sensitive primarily to peak morphology rather than to variations in the global background.

Structural Pooling Paradigm

The second dimension of variation concerns how the latent parameters are coupled across scans, corresponding to different assumptions about inter-specimen parameter structure.

- **Individual Scan Models Without Pooling.** In the simplest formulation, the forward model is applied independently to each scan, with scan-level parameters

drawn directly from priors. Each scan is treated as a self-contained inference problem, and no information is propagated between scans during fitting. Under favorable signal-to-noise conditions this is sufficient, but when a scan is corrupted by noise or low intensity, the inferred parameters have no external anchor and may converge to values inconsistent with the physical system.

- **Hierarchical Model With Partial Pooling.** In the hierarchical formulation, all N scans are modeled jointly under a shared set of population-level hyperparameters. Scan-specific parameters (e.g., peak positions, widths, and areas) are treated as draws from hyperpriors that encode crystallographic structure and expected inter-specimen variability. When a scan provides weak or ambiguous evidence, partial pooling regularizes its parameter estimates toward the population mean $\mu_{\text{global},k}$, reducing the risk of physically implausible solutions.

Probabilistic Inference Engines

The third dimension concerns the algorithm used to approximate the posterior.

- **Stochastic Variational Inference (AutoNormal).** The simplest variational approach constructs an approximate posterior $q_{\lambda}(\mathbf{z})$ by maximizing the [ELBO](#) using an `AutoNormal` guide, which parameterizes a Gaussian with a per-parameter mean and diagonal covariance. This fully factorized mean-field approximation assumes independence among all latent dimensions, making it the least expressive but most computationally tractable of the three methods.
- **Stochastic Variational Inference (AutoLowRankMultivariateNormal).** To relax the independence assumption, the `AutoLowRankMultivariateNormal` guide augments the diagonal covariance with a low-rank correction matrix. This captures the leading off-diagonal posterior correlations among latent parameters while keeping the number of free variational parameters tractable.
- **Markov Chain Monte Carlo With NUTS.** The most accurate inference method employs [MCMC](#) via the [NUTS](#) for asymptotically exact Hamiltonian Monte Carlo sampling of the joint parameter space. Gradient-informed proposals with adaptive path lengths yield samples from the true posterior distribution without introducing variational bias.

Peak Profile Morphology

The final dimension dictates the mathematical representation of the diffraction peaks. In [XRD](#) analysis, peak shapes are predominantly modeled using a [pV](#) function, defined as a linear combination of Gaussian and Lorentzian profiles governed by a mixing parameter η . The choice of how the width of this profile is parameterized directly determines the capacity of the model to capture complex microstructural deformations.

- **Symmetric Pseudo-Voigt Profile.** A standard [pV](#) parameterization enforces symmetric left and right [FWHM](#) values, represented by a single parameter Γ . The symmetric profile ϕ_{sym} evaluated at domain coordinate t is defined as:

$$\phi_{\text{sym}}(t; \mu, \Gamma, \eta) = \eta \mathcal{L}(t; \mu, \Gamma) + (1 - \eta) \mathcal{G}(t; \mu, \Gamma) \quad (5.1)$$

where μ is the peak center position, $\eta \in (0, 1)$ is the Lorentzian mixing fraction, $\mathcal{L}$ denotes the Lorentzian profile function, and $\mathcal{G}$ denotes the Gaussian profile function. Because a single width parameter Γ governs the entire spatial domain of the peak, the profile decays at an identical rate on both sides of the center.

- **Asymmetric Split Pseudo-Voigt Profile.** An asymmetric parameterization explicitly decouples the peak width into independent left (Γ_L) and right (Γ_R) [FWHM](#) parameters. This additional degree of freedom allows the model to capture directional peak asymmetry and differing tail geometries on either side of the centroid. The split profile ϕ_{split} is defined piecewise as:

$$\phi_{\text{split}}(t; \mu, \Gamma_L, \Gamma_R, \eta) = \eta \mathcal{L}(t; \mu, \gamma(t)) + (1 - \eta) \mathcal{G}(t; \mu, \gamma(t)) \quad (5.2)$$

where the position-dependent width parameter $\gamma(t)$ is assigned as:

$$\gamma(t) = \begin{cases} \Gamma_L & \text{if } t < \mu \\ \Gamma_R & \text{if } t \geq \mu \end{cases} \quad (5.3)$$

By allowing Γ_L and Γ_R to vary independently during probabilistic inference, the Bayesian model can directly quantify the asymmetric dislocation accumulation that characterizes advanced structural damage.

5.3 Physics-Guided Synthetic Data Generation

Training highly expressive neural estimators, such as [cMAF](#), necessitates large volumes of paired observations and latent parameters. In the domain of scientific inverse problems,

empirical data is almost exclusively restricted to the small-data regime. Direct training on the limited empirical scans alone is insufficient because the number of observed specimens is small relative to the dimensionality of the latent parameter space. Direct augmentation by independent resampling from each parameter’s marginal posterior discards all off-diagonal structure in the posterior covariance, generating parameter vectors that are statistically consistent in their marginals but physically inconsistent in their joint geometry. To ensure the neural student learns a physically valid mapping, we integrate a physics-guided generative protocol that leverages the high-fidelity posteriors extracted by the Bayesian teacher. This section formalizes the synthetic generation problem and introduces the design space of augmentation strategies evaluated in this work.

5.3.1 The Posterior-Grounded Augmentation Problem

Let $\boldsymbol{\theta}_i \in \mathbb{R}^D$ denote the physical latent parameter vector for scan i . Bayesian spectral deconvolution and inference over the full dataset yields a collection of posterior chains:

$$\mathcal{P} = \{\boldsymbol{\theta}_i^{(m)}\}_{m=1}^{M_{\text{post}}}, \quad i = 1, \dots, N \quad (5.4)$$

where m indexes posterior samples and i indexes scans. Collectively, the pool $\mathcal{P}$ constitutes an empirical approximation to the joint parameter distribution $p(\boldsymbol{\theta} \mid \mathbf{x}_{1:N})$ as observed in the real specimen population.

A naive augmentation strategy samples each component of $\boldsymbol{\theta}$ independently from its marginal posterior, discarding all off-diagonal structure in the posterior covariance. This approach is problematic because the physical parameters of a diffraction pattern are mathematically and physically coupled. The synthetic generation problem is therefore to construct a procedure $\mathcal{G} : \mathcal{P} \rightarrow \tilde{\mathcal{P}}$ that produces an expanded set of parameter vectors $\tilde{\mathcal{P}}$ satisfying two criteria:

1. The marginal distributions of all physical parameters remain consistent with the observed posterior pool, and
2. The joint correlation structure encoded in $\mathcal{P}$ is preserved in $\tilde{\mathcal{P}}$.

5.3.2 Mixed Generative Strategy

To satisfy the requirements of generating a dense, correlation-preserving synthetic corpus without violating crystallographic constraints, we expand the empirical posterior pool $\mathcal{P}$

into a corpus of paired observations and latent parameters $\tilde{\mathcal{D}} = \{(\mathbf{x}_n^{\text{synth}}, \boldsymbol{\theta}_n^{\text{synth}})\}_{n=1}^{N_{\text{synth}}}$ via three complementary mechanisms. The first two mechanisms are specifically designed to satisfy the criteria of maintaining the empirical marginal distributions and preserving the joint correlation structure encoded in the [MCMC](#) posterior pool. The third mechanism focuses on observational robustness by generating novel spectral realizations from exact parameter configurations. These mechanisms are blended as a weighted combination to balance local interpolation, structured extrapolation, and noise resilience.

To counteract the imbalanced specimen representation inherent in the small-data regime, each mechanism draws base samples from the posterior pool using inverse-frequency weighting. Let $g(i) \in \{1, \dots, S\}$ denote the specimen group for scan i , and let $N_{g(i)}$ represent the total number of scans originating from that same specimen. The normalized sampling weight w_i assigned to each scan i is mathematically defined as:

$$w_i = \frac{1/N_{g(i)}}{\sum_{j=1}^N (1/N_{g(j)})} \quad (5.5)$$

This formulation ensures that the probability of sampling from a specific scan is inversely proportional to the frequency of its parent specimen in the dataset. This prevents the neural student from artificially overweighting overrepresented physical samples, ensuring a balanced structural prior.

Mechanism 1: Empirical Parameter Jittering

The first mechanism expands existing posterior samples by injecting parameter-type-aware noise while strictly respecting the physical boundaries of the parameter space. A base parameter vector $\boldsymbol{\theta}$ is sampled from the pooled posterior using the weights defined in [Equation 5.5](#), and specific noise profiles are applied according to the transformation class of each parameter group.

For peak center positions $\mu_{i,k}$, which are real-valued but bounded by the spectral domain $\mathcal{T}_{\text{common}}$, additive noise is applied:

$$\tilde{\mu}_{i,k} = \mu_{i,k} + \xi_{\mu}, \quad \xi_{\mu} \sim \mathcal{N}(0, \sigma_{\mu}^2), \quad (5.6)$$

where the variance hyperparameter σ_{μ}^2 controls the spread, and the result is clipped by a margin δ to prevent positions from drifting outside the valid spectral domain.

For strictly positive parameters, including the left and right [FWHM](#) values $\Gamma_{L,i,k}$ and $\Gamma_{R,i,k}$, peak areas $A_{i,k}$, and noise scale parameters, multiplicative log-normal noise is applied

to ensure non-negativity is preserved:

$$\tilde{z}_d = z_d \cdot \exp(\xi_d), \quad \xi_d \sim \mathcal{N}(0, \sigma_d^2). \quad (5.7)$$

For **FWHM** parameters specifically, the left and right widths are modified with a shared base noise augmented by independent asymmetry noise applied in opposing directions. This allows the synthetic corpus to capture a broader range of asymmetric peak geometries while preserving the fundamental correlation between the two width parameters.

For fractional parameters such as the Lorentzian mixing fraction $\eta_{i,k} \in (0, 1)$, additive Gaussian noise is applied with clipping enforced at the valid boundaries to ensure physical consistency.

Mechanism 2: Gaussian Copula Generation

The second mechanism extrapolates beyond the empirical support of the posterior pool while preserving the full rank correlation structure among all latent parameters. An empirical Gaussian copula [22] is fitted to a subset of the posterior pool.

The fitting procedure transforms each marginal dimension to the normal scale via the empirical probability integral transform [18]. Let $\mathbf{R} \in \mathbb{R}^{n \times D}$ denote the rank matrix of the pool, where R_{ij} is the rank of the j -th parameter of the i -th sample. The normal-quantile-transformed matrix is given by:

$$Z_{ij} = \Phi^{-1}\left(\frac{R_{ij} + 1}{n + 1}\right), \quad (5.8)$$

where Φ^{-1} is the standard normal quantile function and an offset is applied to avoid boundary singularities. A regularized covariance matrix $\hat{\Sigma}_{\text{LW}}$ is estimated from $\mathbf{Z}$ using a shrinkage estimator, which stabilizes the correlation structure in the high-dimensional, limited-sample regime. New samples are drawn from $\mathcal{N}(\bar{\mathbf{z}}, \hat{\Sigma}_{\text{LW}})$ and inverted through the empirical quantile function of each marginal dimension to recover samples in the original parameter space.

Mechanism 3: Forward Noise Simulation

The third mechanism directly leverages the physical forward model to generate synthetic observations with fresh realizations of instrument noise. A parameter vector $\boldsymbol{\theta}^{\text{synth}}$ is drawn exactly from the posterior pool without applying any parameter noise. This configuration

is passed through the additive forward model to produce a noise-free spectrum $\hat{v}(t)$. A fresh independent realization of the heteroscedastic noise process is then added:

$$x^{\text{synth}}(t) = \hat{v}(t; \boldsymbol{\theta}^{\text{synth}}) + \epsilon(t), \quad \epsilon(t) \sim \mathcal{N}(0, \sigma_{\text{base}}^2 + \alpha^2 \max(0, \hat{v}(t))) . \quad (5.9)$$

Rather than generating diverse parameter configurations, this mechanism generates diverse spectral realizations of the exact same physical parameters. This regularizes the downstream encoder, enforcing robustness to stochastic variation in photon counts at a fixed physical state.

5.4 Amortized Neural Posterior Inference

The hierarchical Bayesian model provides accurate uncertainty quantification but requires a joint optimization over all N scans simultaneously due to the partial pooling of population-level hyperpriors. This coupling precludes per-scan decomposition at test time, making iterative inference structurally incompatible with real-time deployment. Amortized inference resolves this incompatibility by shifting the computational burden entirely to an offline training phase, reducing inference at test time to direct sampling from a trained neural network.

5.4.1 Amortized Neural Posterior Estimation and Normalizing Flows

Amortized [NPE](#) replaces per-sample iterative optimization with a learned inference network $q_{\phi}(\mathbf{z} \mid \mathbf{x})$ that approximates the posterior for any observation $\mathbf{x}$. The objective is to learn a global set of network parameters ϕ such that the generated density closely matches the reference [MCMC](#) posterior across the entire data distribution. This is achieved by minimizing the expected [Negative Log-Likelihood \(NLL\)](#) of the true posterior parameters under the approximate neural distribution:

$$\phi^* = \arg \min_{\phi} \mathbb{E}_{\mathbf{x} \sim p(\mathbf{x})} [D_{\text{KL}}(p(\mathbf{z} \mid \mathbf{x}) \parallel q_{\phi}(\mathbf{z} \mid \mathbf{x}))] . \quad (5.10)$$

In this framework, the objective is computed directly using the [MCMC](#) teacher’s posterior samples, which serve as the exact draws from the reference distribution $p(\mathbf{z} \mid \mathbf{x})$.

To construct a sufficiently expressive approximate posterior capable of capturing the correlated, multi-modal geometry of the parameter space, $q_{\phi}(\mathbf{z} \mid \mathbf{x})$ is parameterized as

a [CNF](#). A normalizing flow defines a bijective transformation $f_\phi : \mathbb{R}^D \rightarrow \mathbb{R}^D$ between a simple base distribution $p_0(\mathbf{u}) = \mathcal{N}(\mathbf{0}, \mathbf{I})$ and the target distribution. The log-density of the transformed variable $\mathbf{z} = f_\phi(\mathbf{u})$ is obtained via the change of variables formula:

$$\log q_\phi(\mathbf{z} \mid \mathbf{x}) = \log p_0(f_\phi^{-1}(\mathbf{z})) + \log \left| \det \frac{\partial f_\phi^{-1}}{\partial \mathbf{z}} \right|. \quad (5.11)$$

5.4.2 Domain and Parameter Transformations

Before being presented to the flow, both the input spectra and the target parameters are transformed to ensure numerical stability and satisfy the unconstrained domain $\mathbb{R}^D$ required by normalizing flows. Input spectra $\mathbf{x}_i$ are normalized to $[0, 1]$ via per-sample maximum intensity scaling. The scaling constant is retained so that physical units can be recovered at inference time.

Bounded physical parameters are mapped to the unconstrained real line via parameter-type-aware bijections. For fractional parameters such as $\eta_{i,k}$, a standard logit transformation is applied. For parameters with defined upper bounds $z_{\max}$ (such as [FWHM](#) values and peak amplitudes), a scaled logit transformation is utilized:

$$\tilde{z} = \log \left(\frac{z/z_{\max}}{1 - z/z_{\max}} \right), \quad (5.12)$$

where $z_{\max}$ is defined by an empirical upper bound scaled with a margin of safety to accommodate out-of-sample values. All transformed parameters are then standardized to zero mean and unit variance using statistics computed on the synthetic training corpus, yielding the final unconstrained target vector $\tilde{\mathbf{z}}$.

5.4.3 Spectral Context Encoder

Conditioning the normalizing flow on an observed spectrum requires a fixed-dimensional context embedding. Passing a high-dimensional raw spectrum directly as a conditioning input to each flow block would result in an impractical number of parameters. Instead, a 1D residual convolutional encoder $h_\theta : \mathbb{R}^L \rightarrow \mathbb{R}^{d_h}$ compresses the input spectrum into a compact context vector $\mathbf{h} = h_\theta(\mathbf{x})$.

The encoder architecture utilizes a convolutional stem followed by successive stages of residual blocks with progressively increasing channel widths. Max-pooling operations are

inserted between stages to systematically downsample the spatial resolution. Following the final residual stage, adaptive average pooling collapses the spatial dimension, and the resulting representation is projected through a [Multi-Layer Perceptron \(MLP\)](#) to yield the final context embedding. The multi-scale receptive field hierarchy enables the encoder to capture both the broad spectral background envelope and the fine-grained morphology of individual peaks within a single forward pass.

5.4.4 Conditional Masked Autoregressive Flow (cMAF)

The approximate posterior $q_\phi(\tilde{\mathbf{z}} \mid \mathbf{h})$ is parameterized as a [cMAF](#), as illustrated in [Figure 5.2](#). The flow maps the standard Gaussian base distribution to the target distribution through a series of F invertible autoregressive transformations:

$$\tilde{\mathbf{z}} = T_F \circ \dots \circ T_1(\mathbf{u}; \mathbf{h}), \quad \mathbf{u} \sim p_0(\mathbf{u}). \quad (5.13)$$

Each transformation T_k relies on a [MADE](#) block. In the normalizing direction (from target to base), the inverse transformation T_k^{-1} parameterizes an affine autoregressive coupling:

$$u_d = (\tilde{z}_d - s_d(\tilde{\mathbf{z}}_{<d}, \mathbf{h})) \cdot \exp(-\ell_d(\tilde{\mathbf{z}}_{<d}, \mathbf{h})), \quad (5.14)$$

where s_d and ℓ_d are the shift and log-scale functions conditioned on all preceding data dimensions $\tilde{\mathbf{z}}_{<d}$ and the context $\mathbf{h}$. Because the shift and scale depend on the data rather than the base distribution, the normalizing pass is highly parallelizable. Conversely, the generative mapping T_k requires sequential inversion:

$$\tilde{z}_d = u_d \cdot \exp(\ell_d(\tilde{\mathbf{z}}_{<d}, \mathbf{h})) + s_d(\tilde{\mathbf{z}}_{<d}, \mathbf{h}). \quad (5.15)$$

The log-determinant of the Jacobian for the normalizing pass of a single block is efficiently computed as the sum of the log-scales:

$$\log \left| \det \frac{\partial T_k^{-1}}{\partial \tilde{\mathbf{z}}} \right| = - \sum_{d=1}^D \ell_d(\tilde{\mathbf{z}}_{<d}, \mathbf{h}). \quad (5.16)$$

To ensure dense interactions between all parameter dimensions across the composed flow, the ordering of dimensions is explicitly reversed between consecutive blocks. This reversal ensures that each dimension alternates between acting as a conditioned and a conditioning variable across successive transformations, maximizing the flow’s expressivity.

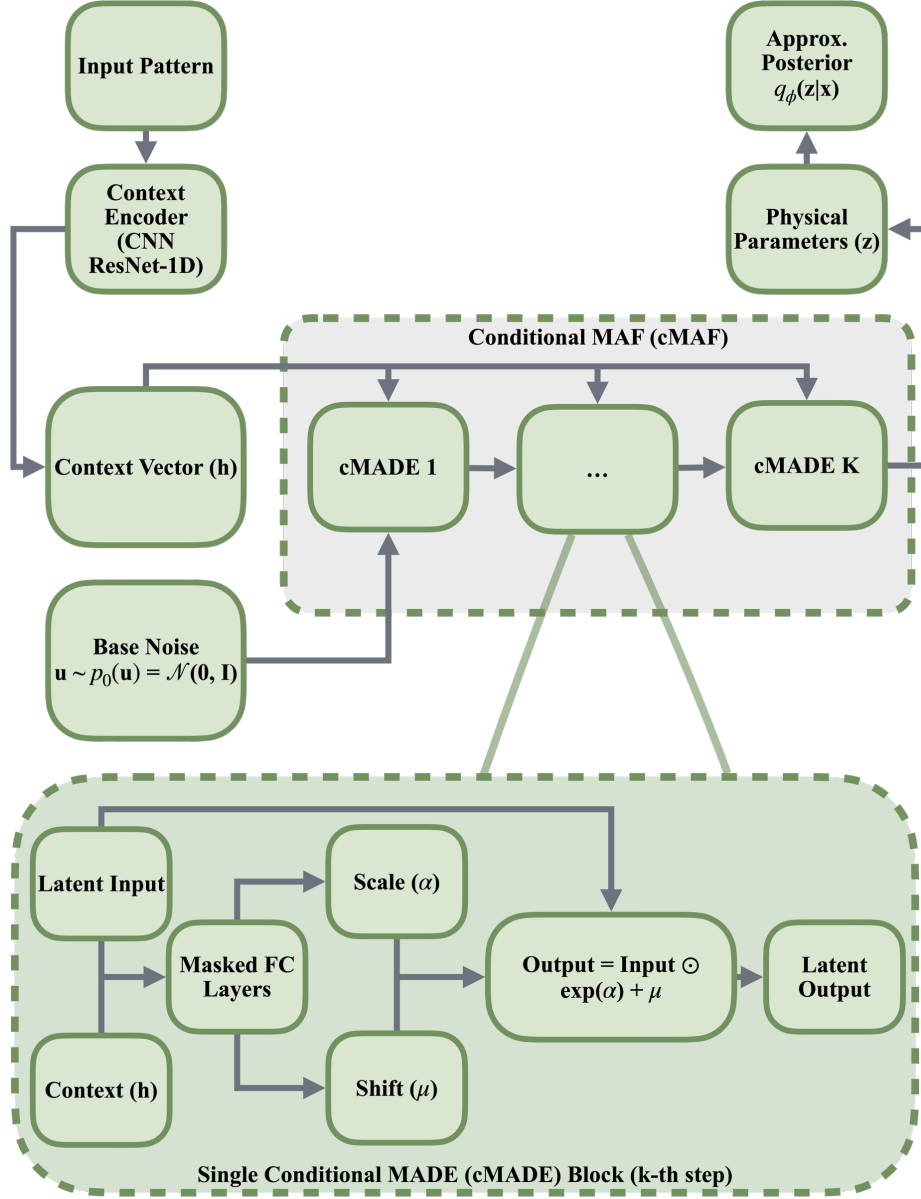


Figure 5.2: Architecture of the conditional masked autoregressive flow. The spectral encoder maps the input pattern to a context vector $\mathbf{h}$, which conditions each cMADE block. Starting from base noise $\mathbf{u}$, each block applies an affine autoregressive transformation whose shift μ and log-scale α , denoted s_d and ℓ_d in Equation 5.15, are predicted by masked fully connected layers. The number of blocks K corresponds to F in the text.

5.4.5 Simulation-to-Reality Training Protocol

Training a normalizing flow exclusively on the synthetic corpus introduces a distribution shift, as the forward simulator cannot perfectly replicate all sources of instrument specific variation. Conversely, training directly on the limited empirical data risks severe overfitting. A staged transfer learning protocol is therefore employed to bridge the simulation-to-reality gap.

Phase 1: Synthetic Pre-training. The full model (both encoder and flow layers) is trained on the synthetic corpus to learn the global topology of the parameter space. The objective is to minimize the expected [NLL](#) on synthetic $(\mathbf{x}^{\text{synth}}, \tilde{\boldsymbol{\theta}}^{\text{synth}})$ pairs:

$$\mathcal{L}_{\text{synth}}(\phi, \theta) = \mathbb{E}_{(\mathbf{x}, \tilde{\boldsymbol{\theta}}) \sim p_{\text{synth}}} \left[-\log q_{\phi}(\tilde{\boldsymbol{\theta}} \mid h_{\theta}(\mathbf{x})) \right]. \quad (5.17)$$

Optimization proceeds using gradient descent techniques with dynamic learning rate scheduling and early stopping based on a held-out synthetic validation set.

Phase 2a: Frozen-Encoder Fine-Tuning. In the first phase of real-data adaptation, the encoder weights θ are frozen and only the flow layers are updated. This prevents the robust feature representations learned from the synthetic corpus from being immediately overwritten. Training minimizes the [NLL](#) with respect to the [MCMC](#) teacher draws from the empirical data:

$$\mathcal{L}_{\text{real}}(\phi) = \mathbb{E}_{i \sim \mathcal{D}_{\text{real}}, \tilde{\boldsymbol{\theta}} \sim q_{\text{MCMC}}(\cdot \mid \mathbf{x}_i)} \left[-\log q_{\phi}(\tilde{\boldsymbol{\theta}} \mid h_{\theta}(\mathbf{x}_i)) \right]. \quad (5.18)$$

As in Phase 2b, this objective is combined with the synthetic replay term of [Equation 5.19](#).

Phase 2b: End-to-End Fine-Tuning with Synthetic Replay. In the final phase, the encoder is unfrozen, and the full network is optimized end-to-end. To prevent catastrophic forgetting of the physical constraints learned during pre-training, a synthetic replay penalty is incorporated into the training objective:

$$\mathcal{L}_{\text{total}}(\phi, \theta) = \mathcal{L}_{\text{real}}(\phi, \theta) + \lambda_{\text{replay}} \mathcal{L}_{\text{synth}}(\phi, \theta), \quad (5.19)$$

where λ_{replay} is a scaling hyperparameter. This three-phase protocol ensures the model inherits the physical prior structure encoded in the synthetic corpus, adapts its flow geometry to the empirical posterior distribution, and retains crystallographic consistency throughout deployment.

5.5 Summary

This chapter has presented the theoretical architecture of the physics-guided teacher-student framework. The Bayesian pipeline produces high-fidelity hierarchical posteriors, serving as the reference ground truth for all downstream evaluation. The synthetic generation pipeline expands these posteriors into a dense training corpus through three complementary mechanisms that preserve both the marginal distributions and the joint correlation structure of the latent parameters. Finally, the neural pipeline distills this knowledge into a [cMAF](#) student, trained through a staged transfer learning protocol that bridges the simulation-to-reality gap while preventing catastrophic forgetting. At inference time, the optimized encoder and flow map a raw spectrum to thousands of physically constrained posterior samples by direct sampling, without iterative inference.

Chapter 6

Implementation and Instantiation

This chapter details the concrete instantiation of the theoretical framework introduced in [chapter 5](#) for the AA6061-T6 dataset described in [chapter 4](#). The abstract design space is resolved into a single, fully specified model configuration through a principled selection process driven by domain knowledge and targeted diagnostic evaluations. The selected configuration is then fully specified with prior distributions, architecture hyperparameters, and the software and hardware environment used for all experiments.

6.1 Model Configuration Selection

The methodological design space defines four independent dimensions along which the Bayesian teacher can be configured: crystallographic scope and spectral domain, structural pooling paradigm, probabilistic inference engine, and peak profile morphology. This section documents the selection made along each dimension and the domain-driven reasoning that eliminates suboptimal alternatives.

6.1.1 Crystallographic Scope and Fitting Strategy

Phase Set. The fundamental [FCC](#) aluminum matrix configuration with $K = 9$ reflections is selected as the crystallographic scope. While AA6061-T6 contains secondary precipitate phases (primarily Mg_2Si), their diffraction signatures exhibit significant inter-specimen variability driven by microstructural heterogeneity. The specific secondary peaks present in a given scan depend on precipitate morphology, orientation, and local chemistry,

which vary across the eight specimens in the dataset and with the location on each specimen. Augmenting the forward model with a static, globally fixed set of secondary-phase reflections is therefore not viable. When a secondary-phase peak is included in the forward model but absent from a given observation, the sampler is forced to explain the absence by driving the corresponding amplitude or area parameter toward zero. In a hierarchical model with partial pooling, this creates systematic pressure on the population-level hyper-priors that distorts the estimates for all scans. This convergence pathology is avoided by restricting the model to the nine FCC reflections, which are physically guaranteed to be present in every scan by the FCC selection rules and the Bragg condition.

Fitting Strategy. Both global and localized window fitting strategies are employed in the final pipeline, with each serving a distinct and critical role. As explained in [subsection 5.2.1](#), certain scans exhibit residual intensity features between the nine FCC peaks that are not explained by the pure aluminum model. Localized window fitting addresses this by restricting the likelihood evaluation to a narrow angular window of $\pm 3^\circ$ centered on each theoretical Bragg angle $\mu_{\text{theory},k}$, as illustrated in [Figure 6.2](#). Within each window, a local instance of the forward model is fit independently, making the inferred peak parameters insensitive to inter-peak features outside the window. This produces tighter, more physically accurate posteriors for the peak morphology parameters, which are utilized specifically for diagnostic validation.

However, as illustrated in [Figure 6.1](#), global fitting spans the full acquisition domain $\mathcal{T}_{\text{common}}$ and produces a unified background parameterization in the form of a smooth fourth-degree Chebyshev polynomial with five coefficients. This parameterization is consistent across all nine reflections and captures the slowly varying baseline contribution from diffuse scattering and instrument response. Because the cMAF student is trained to ingest the full spectral profile and simultaneously predict all structural components, it strictly requires this coherent, full-spectrum parameterization. Therefore, the global fit posteriors serve as the reference ground-truth targets for the cMAF teacher-student distillation and provide the foundation for synthetic data generation, whereas the localized fits provide essential diagnostic validation of the peak parameters.

6.1.2 Structural Pooling Paradigm

Hierarchical partial pooling is selected over the individual scan paradigm. The primary motivation is the heterogeneous signal-to-noise ratio observed across the dataset. Peak intensities span nearly an order of magnitude across the observations, with the distribution

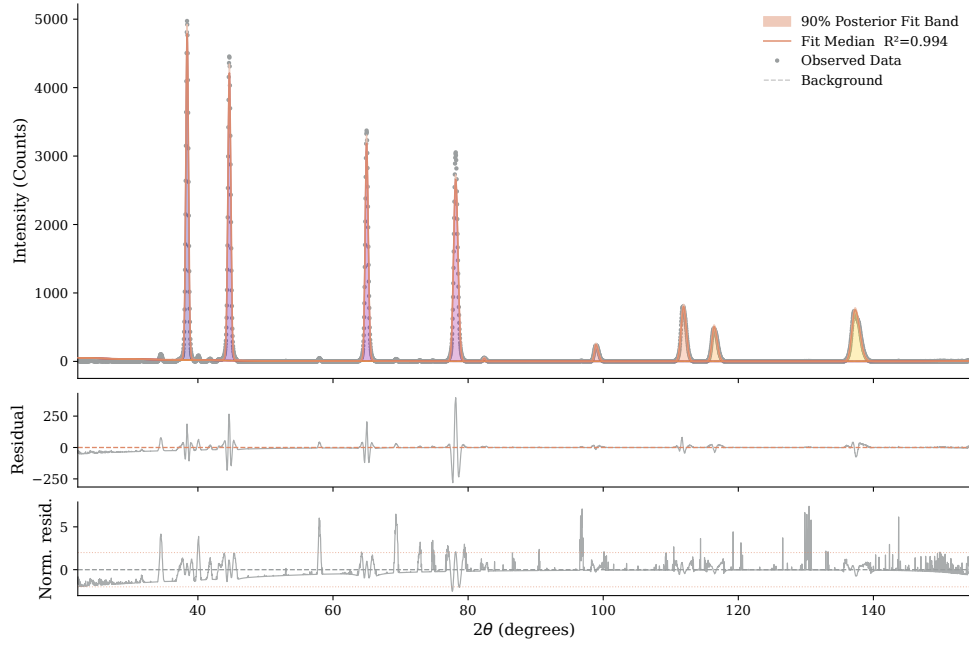


Figure 6.1: Representative global fit utilizing the $K = 9$ fundamental FCC reflections across the full acquisition domain $\mathcal{T}_{\text{common}}$. This full-spectrum parameterization provides the coherent target required for cMAF distillation.

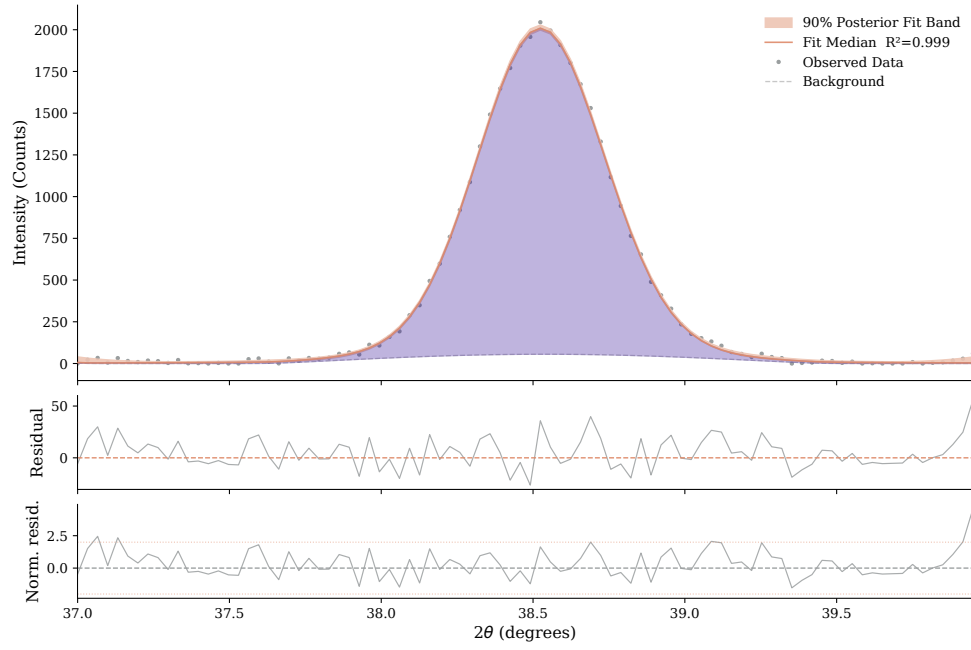


Figure 6.2: Representative localized window fit for a single FCC reflection. The likelihood evaluation is restricted to a $\pm 3^\circ$ window centered on the theoretical Bragg position $\mu_{\text{theory},k}$, providing diagnostic validation of the peak morphology.

heavily right-skewed. Scans in the low-intensity tail may have insufficient photon counts at higher-angle reflections to constrain the corresponding peak morphology parameters from per-scan evidence alone.

Hierarchical partial pooling resolves this by modeling all scans jointly under shared population-level hyperpriors. For each peak k , the population mean position $\mu_{\text{global},k}$ is placed under a hyperprior centered on the theoretical Bragg angle $\mu_{\text{theory},k}$, and scan-level positions are modeled as perturbations around this shared mean:

$$\mu_{\text{global},k} \sim \mathcal{N}(\mu_{\text{theory},k}, \tau^2), \quad (6.1)$$

$$\mu_{i,k} \sim \mathcal{N}(\mu_{\text{global},k}, \sigma_{\mu,k}^2). \quad (6.2)$$

When a scan is heavily corrupted by noise, the posterior for $\mu_{i,k}$ is regularized toward $\mu_{\text{global},k}$ rather than drifting into physically implausible regions. The degree of shrinkage is automatically determined by the relative informativeness of the scan-level likelihood versus the population-level hyperprior.

A simplified plate notation diagram is used to provide the statistical graphical model for each pooling paradigm. [Figure 6.3](#) depicts the individual scan model without pooling, where each scan is treated as an independent inference problem with no shared parameters. [Figure 6.4](#) illustrates the hierarchical model with partial pooling, where global hyperparameters $\psi_{\text{global},k}$ enforce statistical sharing across scans. A detailed unpacked [Directed Acyclic Graph \(DAG\)](#) ([Figure 6.5](#)) further breaks down the peak parameters into their physical components to explicitly show how crystallographic principles are encoded in the model structure.

6.1.3 Probabilistic Inference Engine

During prototyping and initial experimental runs, stochastic variational inference was used due to its computational efficiency and faster turnaround, but for the final pipeline exact inference via [MCMC](#) utilizing the [NUTS](#) was selected over stochastic variational inference configurations. Variational methods optimize a proxy objective and are mathematically known to systematically underestimate posterior variance in highly correlated, nonlinear parameter spaces. A teacher that underestimates uncertainty propagates this compression downstream, producing a student with artificially narrow predictive intervals. Because the primary objective of this Bayesian pipeline is to serve as an exact, unbiased ground-truth teacher for the neural distillation process, exact inference via [MCMC](#) with [NUTS](#) is therefore selected as the required probabilistic engine.

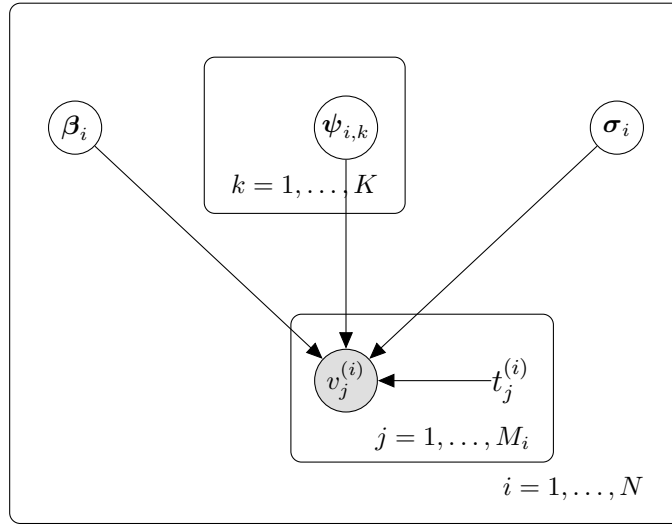


Figure 6.3: **DAG** of the individual Bayesian teacher model without pooling. The latent parameters for the background (β_i), peak morphology ($\psi_{i,k}$), and heteroscedastic noise (σ_i) are drawn directly from fixed, unmodeled priors for each scan i , treating each measurement as a self-contained inference problem. These parameters map the domain coordinates $t_j^{(i)}$ to the observed intensities $v_j^{(i)}$.

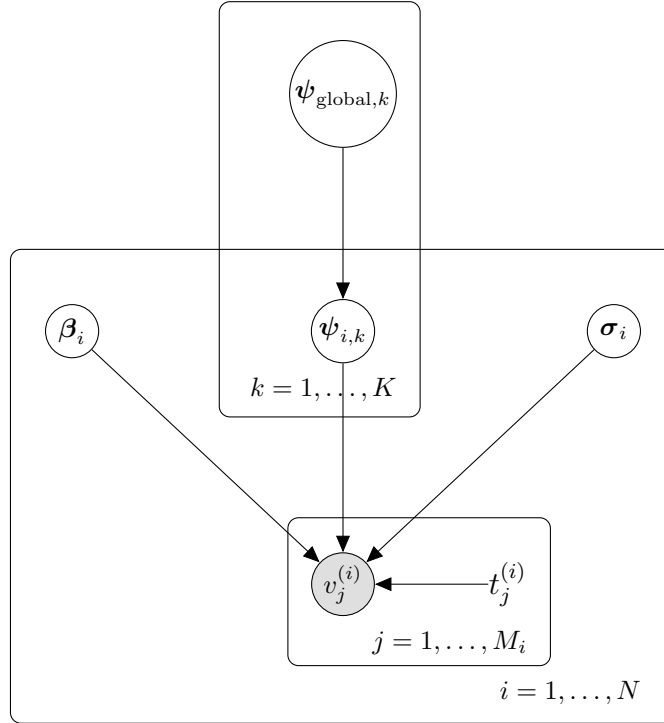


Figure 6.4: **DAG** of the hierarchical Bayesian teacher model. In contrast to the unpooled individual model, global hyperparameters $\psi_{\text{global},k}$ enforce partial pooling across the dataset. This allows the scan-specific peak parameters $\psi_{i,k}$ to borrow statistical strength from the population, regularizing the estimates in regimes with low signal-to-noise. Together with the background β_i and heteroscedastic noise parameters σ_i , these parameters map the domain coordinates $t_j^{(i)}$ to the observed intensities $v_j^{(i)}$.

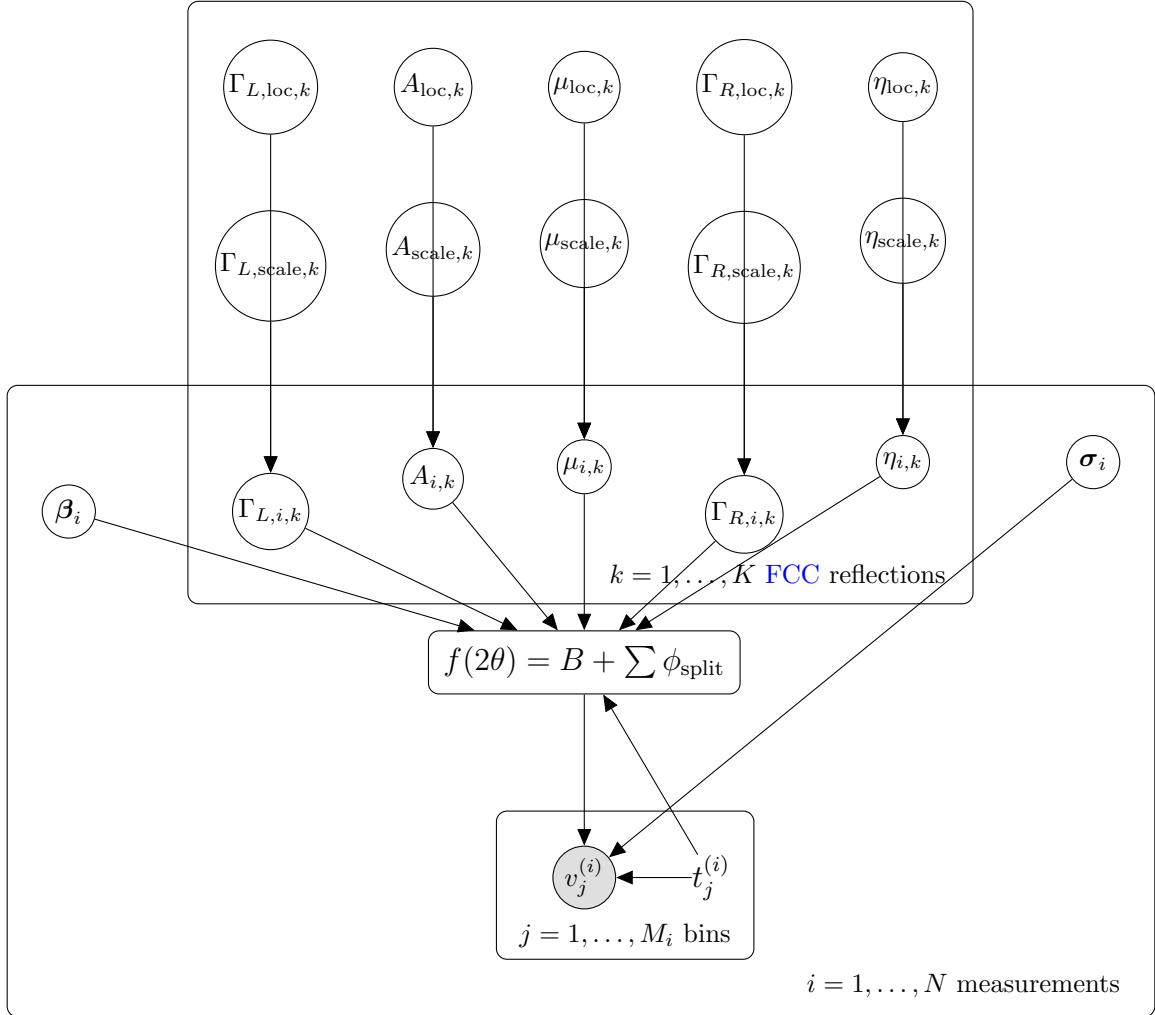


Figure 6.5: Detailed conceptual **DAG** of the hierarchical Bayesian teacher model. By explicitly unpacking the physical peak parameters (Area A , Position μ , **FWHM** Γ , and Shape η), the diagram highlights exactly how crystallographic principles are enforced across all N measurements.

6.1.4 Peak Profile Morphology

The asymmetric **sPV** profile is selected over the symmetric baseline. Fatigue-induced dislocation accumulation intrinsically manifests as asymmetric peak broadening and lattice strain. The left and right wings of a diffraction peak respond differently to the tensile and compressive micro-strains, producing an asymmetric intensity distribution around the peak center that cannot be captured by a single width parameter. Capturing this directional skew by decoupling the left and right **FWHM** parameters is the better option for preserving the microstructural signatures required for downstream classification.

6.2 Prior Specifications

The Bayesian teacher requires prior distributions over all latent parameters. All priors are designed to be weakly informative. They are tight enough to enforce physical plausibility but broad enough to accommodate the full range of observed peak morphologies across the scans and damage states.

6.2.1 Peak Position Priors

Peak center positions are assigned a two-level hierarchical prior. The population-level hyperprior is centered on the theoretical Bragg angle $\mu_{\text{theory},k}$ derived from Bragg’s Law and the known **FCC** lattice parameter of AA6061-T6 ($a = 4.050 \text{ \AA}$). A standard deviation of $\tau = 0.15^\circ$ is used to permit population-level shifts due to systematic macrostress while preventing unphysical drift. The scan-level deviation is assigned a half-normal prior on $\sigma_{\mu,k}$ to enforce non-negativity of the spread parameter. [Table 6.1](#) summarizes all prior distributions.

6.2.2 Peak Shape Priors

Rather than placing priors directly on peak amplitude, the model is parameterized using the integrated peak area $A_{i,k}$, which is assigned a half-normal prior. This orthogonal parameterization decouples integrated intensity from peak broadening, reducing posterior correlation and improving sampler efficiency. Left and right **FWHM** parameters $\Gamma_{L,i,k}$ and $\Gamma_{R,i,k}$ are assigned independent half-normal priors with a common scale of 0.5° . The Lorentzian mixing fraction $\eta_{i,k}$ is assigned a Beta(2, 2) prior centered at 0.5, reflecting

prior neutrality between Gaussian-dominant and Lorentzian-dominant peak shapes while penalizing values near 0 or 1.

6.2.3 Background and Noise Model Priors

The background is modeled as a fourth-degree Chebyshev polynomial with five coefficients, chosen for its numerical orthogonality over the domain, which minimizes the condition number during sampling. The constant term is assigned a half-normal prior with scale 0.5, and the four higher-order coefficients are assigned zero-mean normal priors with standard deviations σ_j of 0.05, 0.02, 0.01, and 0.005, favoring a smooth background. Positivity of the reconstructed background is enforced via a softplus transformation applied to the polynomial evaluation.

The heteroscedastic noise model is parameterized by a baseline scale $\sigma_{\text{base},i}$ and an intensity-dependent scaling factor α_i , which governs Poisson-like variance growth. Both parameters are assigned strictly positive half-normal priors with small scales ($\sigma_{\text{base}} \sim \mathcal{HN}(0.01)$, $\alpha \sim \mathcal{HN}(0.05)$) that encode the expectation of low absolute noise levels while permitting the sampler to adapt to the empirically observed noise floor.

Table 6.1: Prior distributions for all latent parameters of the hierarchical Bayesian teacher. $\mathcal{HN}(\sigma)$ denotes a half-normal distribution with scale σ . Parameters are defined on the normalized spectrum after per-sample maximum intensity scaling.

Parameter	Symbol	Prior	Constraint
Population peak position	$\mu_{\text{global},k}$	$\mathcal{N}(\mu_{\text{theory},k}, 0.15^2)$	$\in \mathcal{T}_{\text{common}}$
Scan peak position	$\mu_{i,k}$	$\mathcal{N}(\mu_{\text{global},k}, \sigma_{\mu,k}^2)$	$\in \mathcal{T}_{\text{common}}$
Position deviation scale	$\sigma_{\mu,k}$	$\mathcal{HN}(0.10)$	> 0
Integrated peak area	$A_{i,k}$	$\mathcal{HN}(0.30)$	> 0
Left FWHM	$\Gamma_{L,i,k}$	$\mathcal{HN}(0.50)$	$(0.005^\circ, 3.0^\circ)$
Right FWHM	$\Gamma_{R,i,k}$	$\mathcal{HN}(0.50)$	$(0.005^\circ, 3.0^\circ)$
Lorentzian mixing fraction	$\eta_{i,k}$	Beta(2, 2)	$(0, 1)$
Chebyshev constant term	$\beta_{i,0}$	$\mathcal{HN}(0.50)$	> 0
Higher-order coefficients	$\beta_{i,1:4}$	$\mathcal{N}(0, \sigma_j^2)$	$\mathbb{R}$
Baseline noise scale	$\sigma_{\text{base},i}$	$\mathcal{HN}(0.01)$	> 0
Intensity noise scale	α_i	$\mathcal{HN}(0.05)$	> 0

6.3 Neural Architecture and Training Hyperparameters

This section specifies the concrete architecture of the [cMAF](#) student and the hyperparameters of the three-phase training protocol introduced in [chapter 5](#).

The architecture and training hyperparameters of the [cMAF](#) student were selected through an empirical search. A range of candidate configurations was defined, covering the encoder depth and channel widths, the number and hidden dimension of the [MADE](#) blocks, the learning rates of each training phase, and the composition of the synthetic corpus. Each candidate was trained under the protocol described in [subsection 5.4.5](#), and the configuration achieving the best performance on held-out validation data was retained.

Spectral Context Encoder. The 1D residual convolutional encoder $h_\theta : \mathbb{R}^{4000} \rightarrow \mathbb{R}^{256}$ consists of a convolutional stem followed by four residual stages and a projection head. The four residual stages operate at channel widths of 32, 64, 128, and 256 respectively, each comprising two sequential residual blocks. Max-pooling layers with stride 2 are inserted after stages 1, 2, and 3 to progressively halve the spatial resolution. Kernel sizes decrease across stages (15, 7, 5, 3) to maintain a multi-scale receptive field. Following the final residual stage, adaptive average pooling collapses the spatial dimension to 1. This is projected through a two-layer MLP with a hidden dimension of 256, ReLU activation, and a dropout rate of 0.10, yielding the final context embedding $\mathbf{h} \in \mathbb{R}^{256}$. The architecture is visualized in [Figure 6.6](#).

Conditional Masked Autoregressive Flow. The [cMAF](#) consists of $F = 12$ [MADE](#) blocks, each with a hidden dimension of 512. The context is concatenated with the input to the first masked linear layer, with full connectivity from context to all hidden units. Log-scale outputs are clamped to $[-5, 5]$ to prevent numerical instability. Dimension ordering is reversed between consecutive blocks to ensure that each parameter dimension alternates between acting as a conditioned and a conditioning variable. The target parameter dimensionality is $D = 52$, corresponding to the $K = 9$ peaks, each parameterized by 5 values $(\mu, A, \Gamma_L, \Gamma_R, \eta)$, plus 5 Chebyshev background coefficients, 1 baseline noise scale, and 1 intensity noise scale parameter. The architecture is visualized in [Figure 6.7](#).

[Table 6.2](#) summarizes the architectural specifications of both the spectral context encoder and the [cMAF](#). [Table 6.3](#) details the hyperparameters for each phase of training, including synthetic pre-training, frozen-encoder fine-tuning, end-to-end fine-tuning, and posterior inference.

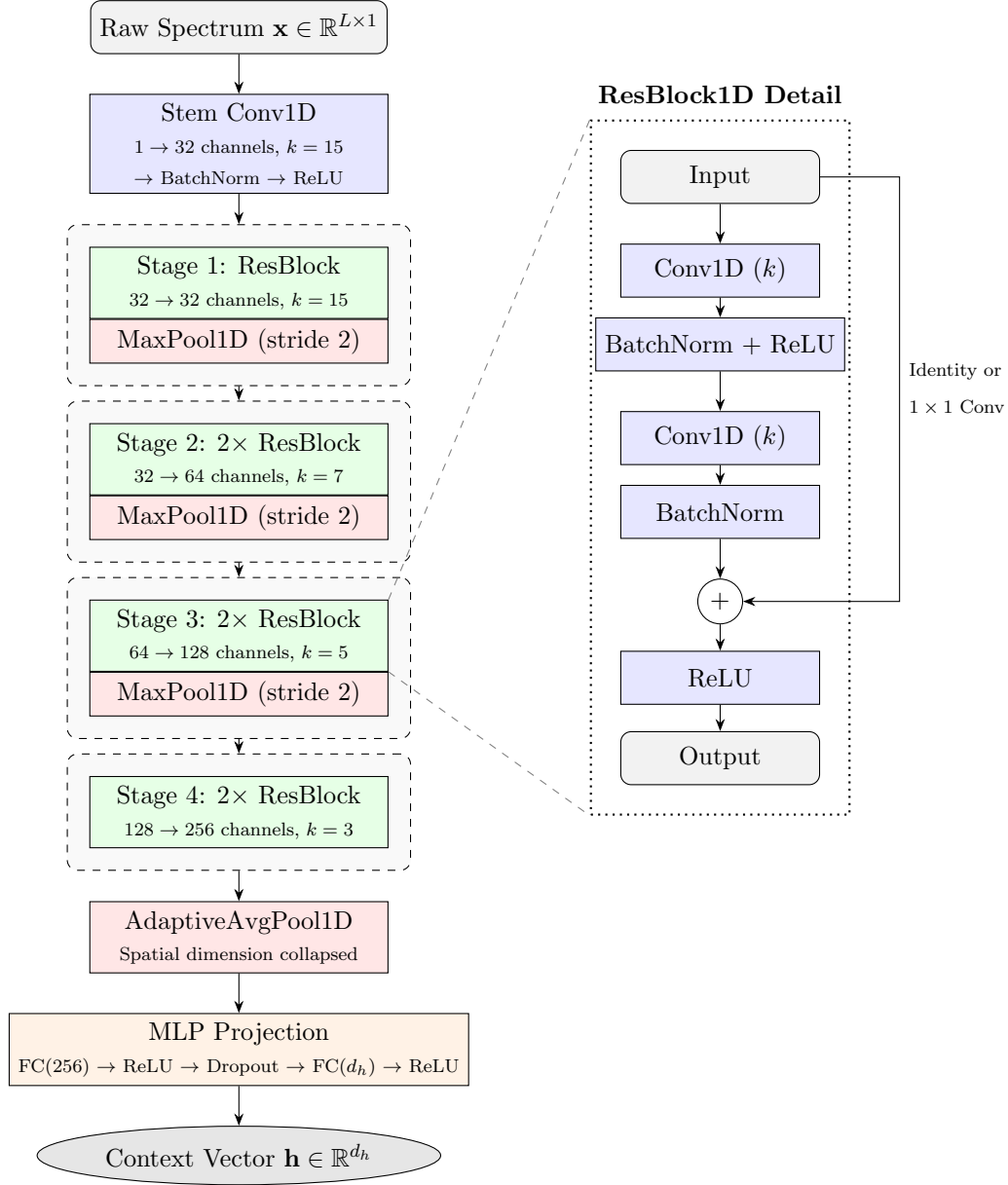


Figure 6.6: Architecture of the 1D Spectral Context Encoder. (Left) The macroscopic multi-scale hierarchy, progressively downsampling spatial resolution via max-pooling while increasing channel depth. (Right) The internal structure of the custom **ResBlock1D** modules, featuring standard post-activation batch normalization and a conditional 1×1 convolutional skip connection for dimension matching.

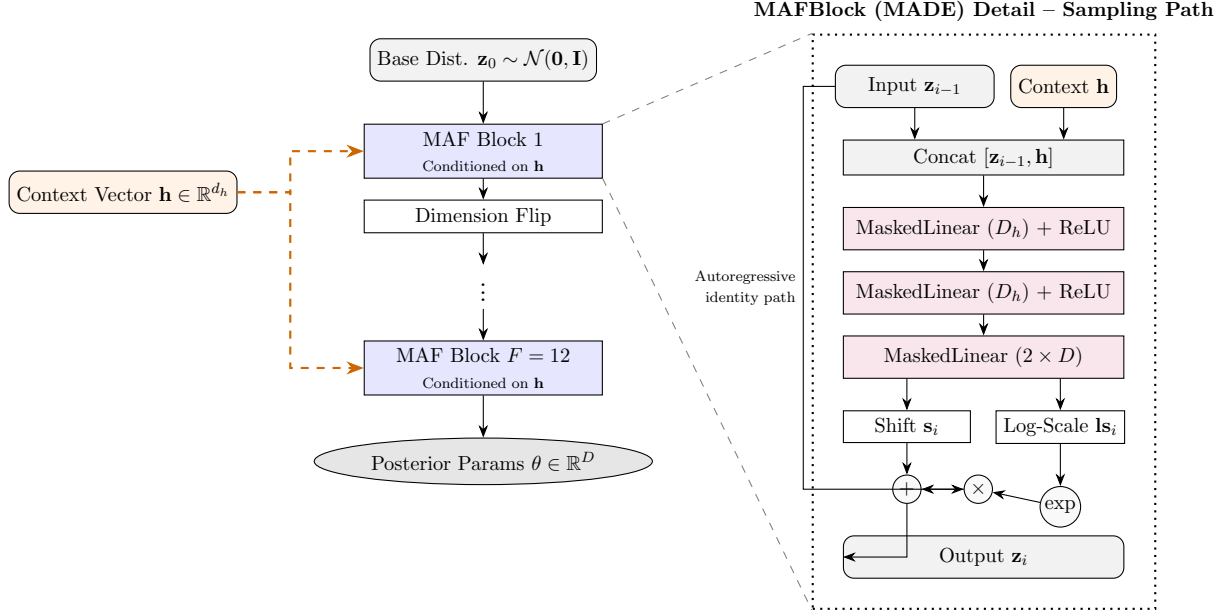


Figure 6.7: Architecture of the **cMAF**. (Left) The macroscopic generation process, transforming samples from a standard Gaussian into target posterior parameters via a sequence of affine transformations and dimension flips, conditioned globally on the Context Vector. (Right) The internal structure of a single MAF Block, utilizing Masked Linear layers to strictly enforce the autoregressive property during sequential sampling.

Table 6.2: **cMAF** architecture specification.

Component	Specification
<i>Spectral Context Encoder</i>	
Input dimension	$L = 4,000$
Residual stage channels	$[32, 64, 128, 256]$
Kernel sizes per stage	$[15, 7, 5, 3]$
Context embedding dimension	$d_h = 256$
Dropout rate	0.10
<i>Normalizing Flow</i>	
Number of MADE blocks	$F = 12$
Hidden dimension per block	512
Target parameter dimension	$D = 52$
Log-scale clamp range	$[-5, 5]$

Table 6.3: Training and posterior inference hyperparameters.

Component	Specification
<i>Phase 1: Synthetic Pre-training</i>	
Synthetic corpus size per fold	$N_{\text{synth}} = 10,000$
Corpus composition	40% jittering, 35% copula, 25% forward noise
Batch size	64
Optimizer	AdamW
Learning rate	10^{-3}
Max epochs	100
Early stopping patience	15 epochs
<i>Phase 2a: Frozen-Encoder Fine-Tuning</i>	
Encoder	frozen
Optimizer	Adam
Learning rate	5×10^{-5}
Epochs	4
Synthetic replay weight	$\lambda_{\text{replay}} = 0.10$
Teacher draws per scan per step	$K_T = 32$
<i>Phase 2b: End-to-End Fine-Tuning</i>	
Encoder	unfrozen
Optimizer	Adam
Learning rate	3×10^{-5}
Max epochs	50
Early stopping patience	15 epochs
Synthetic replay weight	$\lambda_{\text{replay}} = 0.10$
<i>Posterior Inference</i>	
Posterior draws per scan	1,000

6.4 Software and Hardware Environment

All experiments were conducted on a server equipped with dual Intel Xeon Gold 5420+ processors and a single NVIDIA H100 NVL graphics processing unit with 94 GB of VRAM, running CUDA 12.8, with the full configuration summarized in Table 6.4. Both the Bayesian and neural pipelines utilized full hardware acceleration throughout training and inference.

Table 6.4: Summary of the computational hardware and software environment utilized for model training and inference.

Component	Specification
<i>Central Processing Unit (CPU)</i>	
Processor Model	Dual Intel Xeon Gold 5420+
Core Topology	2 Sockets, 28 Cores per socket (112 Threads total)
System Memory (RAM)	1.0 TiB
<i>Graphics Processing Unit (GPU)</i>	
GPU Model	1 × NVIDIA H100 NVL
Video Memory (VRAM)	94 GB per GPU
<i>Software Environment</i>	
Operating System	Ubuntu 24.04.4 LTS (Kernel 6.8.0-110-generic)
NVIDIA Driver Version	570.195.03
CUDA Toolkit Version	12.8

The Bayesian pipeline was implemented using NumPyro [30] and JAX, with the NUTS sampler provided by NumPyro’s MCMC interface. JAX’s just-in-time compilation and vectorized map (`vmap`) operations were used to accelerate the forward model evaluations required at each Hamiltonian step. The neural pipeline was implemented in PyTorch, with all network components and training loops constructed using native `torch.nn` modules. Posterior traces were serialized to NetCDF format via ArviZ for compatibility with standard probabilistic programming diagnostics. A fixed global random seed of 42 was applied consistently across Python, NumPy [12], and PyTorch [29] random number generators, with CUDA deterministic mode enabled to ensure full reproducibility.

6.5 Summary

This chapter has resolved the abstract design space into a fully specified model configuration for the AA6061-T6 dataset. The selected configuration adopts the pure FCC matrix scope with dual global and localized fitting, hierarchical partial pooling, exact MCMC inference, and the asymmetric sPV profile. The selection of each dimension is justified by the physical properties of the dataset and the requirement that the Bayesian teacher serve as an unbiased reference posterior for neural distillation. Full prior specifications, architecture hyperparameters, and the three-phase training protocol are documented for reproducibility. The following chapter defines the evaluation methodology used to formally assess the instantiated framework.

Chapter 7

Experimental Setup and Evaluation Methodology

This chapter defines the complete evaluation methodology used to assess the instantiated framework. It specifies how the framework is evaluated across multiple dimensions: the quality of the spectral fits produced by the Bayesian teacher, the fidelity of the posteriors produced by both the teacher and the neural student, the computational cost of each inference paradigm, and the utility of those posteriors for the downstream structural damage classification task. All experimental decisions documented here are established prior to observing the empirical results reported in [chapter 8](#).

7.1 Evaluation Dimensions

The framework is evaluated across multiple dimensions that reflect the sequential structure of the pipeline:

1. **Spectral Fitting Quality:** This dimension assesses how accurately the Bayesian forward model reconstructs the observed [XRD](#) spectra from the inferred posteriors.
2. **Posterior Quality:** This dimension evaluates the geometric fidelity of the neural student’s approximate posterior relative to the [MCMC](#) teacher’s reference posterior. It is measured across two axes:
 - **Physical Calibration:** Assesses how well the posterior samples, when passed through the deterministic forward model, reconstruct the observed spectra. This

measures the degree to which the inferred posteriors are consistent with the underlying physical data.

- **Posterior Geometry:** Quantifies how closely the neural student’s posterior distribution matches the exact [MCMC](#) teacher’s posterior.
3. **Computational Efficiency:** Evaluates the inference time required by each methodology.
 4. **Downstream Classification Performance:** This dimension determines whether the inferred posteriors, when utilized as feature representations, support accurate and well-calibrated damage state prediction on unseen physical specimens. It is divided into three analytical focuses:
 - **Predictive Performance:** Evaluates the accuracy, discriminative power, and class balance of the classifier’s predictions.
 - **Uncertainty Quantification:** Assesses the calibration of the classifier’s predicted probabilities, measuring whether the confidence of predictions aligns with their empirical correctness.
 - **Physical Interpretability:** Verifies whether the classifier’s decision boundary is grounded in physically meaningful parameters consistent with established fatigue damage mechanisms, rather than relying on statistical artifacts.

7.2 Task Definition

7.2.1 Spectral Inverse Problem

The primary task is to solve the spectral inverse problem for each scan i : given the observed preprocessed spectrum $\mathbf{x}_i \in \mathbb{R}^L$, infer the full posterior distribution $p(\mathbf{z}_i \mid \mathbf{x}_i)$ over the $D = 52$ latent physical parameters defined in [section 6.2](#). The Bayesian teacher solves this problem iteratively via hierarchical [MCMC](#), while the neural student solves it amortizedly by sampling directly from the trained [cMAF](#). The posterior quality evaluation measures how closely the student recovers the teacher’s reference solution.

7.2.2 Downstream Classification

The downstream task is formulated as binary classification. Each scan i is assigned a label $y_i \in \{0, 1\}$, where $y_i = 0$ denotes the *Nominal* class (Pristine scans, $N_{\text{nom}} = 51$) and

$y_i = 1$ denotes the *Damaged* class (Mid-Life and Late-Life scans combined, $N_{\text{dmg}} = 55$). Mid-Life and Late-Life damage states are merged into a single Damaged class because the primary operational requirement for structural health monitoring is the detection of any microstructural degradation from the pristine baseline, rather than the fine-grained staging of damage progression. The label describes the damage state of the specimen at the time of the scan, so a specimen can contribute scans to more than one class.

7.3 Feature Extraction

The inference engine produces $M_{\text{post}} = 1,000$ posterior samples $\{\mathbf{z}_i^{(m)}\}_{m=1}^{M_{\text{post}}}$ for each scan i . Rather than passing raw posterior chains into a classifier, two scalar feature sets are derived from these samples to provide structured representations of the posterior geometry.

Point-Estimate Features (X_{mean}). The first feature set is constructed by computing the posterior mean for the relevant latent parameters:

$$X_{\text{mean},i} = \frac{1}{M_{\text{post}}} \sum_{m=1}^{M_{\text{post}}} \mathbf{z}_i^{(m)} \in \mathbb{R}^{45}, \quad (7.1)$$

where the 45-dimensional vector is formed by retaining strictly the $K \times 5$ peak morphology parameters (position, area, left FWHM, right FWHM, and Lorentzian mixing fraction for each of the nine reflections). The background and noise parameters are discarded as they carry no direct microstructural interpretation. This feature set establishes a deterministic baseline that does not propagate posterior uncertainty into the classifier.

Distributional Features (X_{dist}). The second feature set augments the posterior means with the posterior standard deviations:

$$X_{\text{dist},i} = [X_{\text{mean},i}, \text{std}(\{\mathbf{z}_i^{(m)}\}_{m=1}^{M_{\text{post}}})] \in \mathbb{R}^{90}, \quad (7.2)$$

where std denotes the per-parameter empirical standard deviation across the posterior samples. This feature set encodes both the estimated physical state and the inferential uncertainty associated with each parameter, enabling the evaluation of whether propagated structural uncertainty improves downstream predictive performance and calibration.

Both feature sets are extracted identically from the [MCMC](#) teacher and the [cMAF](#) student to ensure a direct comparison of their downstream utility.

7.4 Cross-Validation Protocol

The grouped data structure of the dataset (multiple scans per specimen) invalidates standard randomized train-test splitting. If scans from the same specimen appear in both the training and test sets, the classifier is evaluated on data that share specimen-specific characteristics with the training observations. This data leakage inflates performance metrics and produces optimistic estimates that fail to generalize to unseen specimens. To enforce a strict generalization requirement, **LOGOCV** is adopted as the evaluation protocol.

In **LOGOCV**, the dataset is partitioned into $S = 8$ folds, corresponding to the distinct physical specimens. In each fold s , all scans from specimen s are held out as the test set, and the classifier is trained exclusively on the scans from the remaining $S - 1 = 7$ specimens. This ensures that every test prediction is made on a physical specimen entirely unseen during classifier training. The procedure is repeated for all S specimens, yielding a complete out-of-fold prediction for every scan in the dataset.

Crucially, both the **MCMC** teacher and the **cMAF** student are trained within this **LOGOCV** framework. For each fold, the population-level hyperpriors of the hierarchical teacher are estimated from the training specimens only, so that they are not informed by the held-out specimen. The synthetic corpus of the student is then generated from the posterior pool of these training specimens, and fine-tuning is performed solely on the training scans. Consequently, posterior samples for the held-out specimen are generated by a student that has never observed that specimen at any stage of training, ensuring that the resulting posterior fidelity and classification metrics reflect generalization to an unseen specimen.

7.5 Evaluation Metrics

The evaluation metrics are organized to correspond to the dimensions defined in [section 7.1](#).

7.5.1 Spectral Fitting Quality

Fitting quality is assessed via **Posterior Predictive Checks (PPC)**. For each scan i , M_{ppc} posterior samples are drawn and passed through the deterministic forward model to generate a distribution of reconstructed spectra. The median reconstruction is compared against the observed spectrum using two scalar metrics:

- **Coefficient of Determination (R^2):** Measures the proportion of variance in the observed spectrum explained by the median posterior reconstruction:

$$R^2 = 1 - \frac{\sum_{j=1}^L (v_j - \hat{v}_j)^2}{\sum_{j=1}^L (v_j - \bar{v})^2}, \quad (7.3)$$

where v_j are the observed intensities, $\hat{v}_j$ are the median reconstructed intensities, and $\bar{v}$ is the mean of the observed intensities. Values closer to 1 indicate a more accurate spectral reconstruction.

- **Root Mean Square Error:** Measures the absolute magnitude of reconstruction residuals on the normalized intensity scale:

$$\text{RMSE} = \sqrt{\frac{1}{L} \sum_{j=1}^L (v_j - \hat{v}_j)^2}. \quad (7.4)$$

Lower values indicate tighter agreement between the generative model and the empirical data.

7.5.2 Posterior Quality and Physical Calibration

Posterior quality is assessed along two complementary axes: geometric fidelity to the [MCMC](#) reference, and physical calibration of the reconstructed spectra.

Physical Calibration. Posterior predictive checks utilizing R^2 and [Root Mean Square Error \(RMSE\)](#) are applied to both the [MCMC](#) teacher and the [cMAF](#) student. Comparing these values quantifies the degree of physical calibration lost during the distillation process.

Posterior Geometry. To measure how closely the student’s posterior approximates the geometry of the teacher’s posterior, two distribution-level discrepancy metrics are utilized:

- **1-Wasserstein Distance (W_1):** Also known as the Earth Mover’s Distance, W_1 measures the minimum cost of transporting the probability mass of one distribution to match another:

$$W_1(p, q) = \inf_{\gamma \in \Gamma(p, q)} \mathbb{E}_{(\mathbf{z}, \mathbf{z}') \sim \gamma} [\|\mathbf{z} - \mathbf{z}'\|], \quad (7.5)$$

where $\Gamma(p, q)$ denotes the set of all joint distributions with marginals p and q . Lower values indicate closer agreement between the teacher and student posterior distributions.

- **Maximum Mean Discrepancy (MMD²):** This is a kernel-based two-sample test statistic that measures the difference between two distributions by comparing their mean embeddings in a reproducing kernel Hilbert space:

$$\text{MMD}^2(p, q) = \mathbb{E}_{\mathbf{z}, \mathbf{z}' \sim p}[k(\mathbf{z}, \mathbf{z}')] - 2 \mathbb{E}_{\mathbf{z} \sim p, \mathbf{z}' \sim q}[k(\mathbf{z}, \mathbf{z}')] + \mathbb{E}_{\mathbf{z}, \mathbf{z}' \sim q}[k(\mathbf{z}, \mathbf{z}')], \quad (7.6)$$

where $k(\cdot, \cdot)$ is a positive definite kernel function. Lower values indicate superior distributional alignment.

Both W_1 and MMD^2 are computed between the posterior samples of the **MCMC** teacher and the **cMAF** student for each test scan, on the peak positions and **FWHM** parameters.

7.5.3 Computational Efficiency

Inference time per evaluation set is measured for both the teacher and the student under identical hardware conditions. The **MCMC** teacher operates at iterative complexity $\mathcal{O}(T)$, where T is the total number of Hamiltonian steps across all chains, and cannot be decomposed per scan due to the hierarchical coupling. The **cMAF** student operates at amortized complexity $\mathcal{O}(1)$, processing each scan independently by sampling directly from the trained flow. Both inference times are reported in their natural units (minutes for the teacher, milliseconds for the student) to make the practical deployment contrast concrete.

7.5.4 Predictive Performance

Downstream classification performance is evaluated using three metrics computed on the pooled **LOGOCV** out-of-fold predictions, in which each scan is predicted once by the classifier trained without its specimen:

- **Accuracy:** The proportion of correctly classified scans across all test folds.
- **Macro-F1 Score:** The unweighted mean of per-class F1 scores, defined as the harmonic mean of precision and recall for each class. Macro-F1 is reported in addition to accuracy because the cost of misclassification is asymmetric in the structural health monitoring context (where false negatives are significantly more consequential than false positives).

- **Area Under the ROC Curve (AUC):** The area under the Receiver Operating Characteristic curve, measuring the classifier’s ability to rank damaged scans above nominal scans across all decision thresholds. [Area Under the ROC Curve \(AUC\)](#) provides a global, threshold-independent measure of discriminative power. Pooling also allows [AUC](#) to be computed when a test fold contains scans from a single class, for which a per-fold [AUC](#) is undefined.

In addition, accuracy is computed for each held-out specimen and summarized by its mean and standard deviation across the S folds. Macro-F1 is not computed per fold, since it is not meaningful for a test fold that contains scans from a single class.

All three metrics are reported for each combination of inference engine and feature set to enable a systematic ablation of the contribution of uncertainty features to downstream performance.

7.5.5 Uncertainty Quantification

Two metrics, computed on the same pooled predictions, are used to assess the calibration of the classifier’s predicted probabilities:

- **Expected Calibration Error:** [Expected Calibration Error \(ECE\)](#) measures the alignment between predicted class probabilities and empirical frequencies by partitioning predictions into B equally spaced confidence bins and computing the weighted average of the absolute calibration gap:

$$\text{ECE} = \sum_{b=1}^B \frac{|B_b|}{N} |\text{acc}(B_b) - \text{conf}(B_b)|, \quad (7.7)$$

where $|B_b|$ is the number of predictions in bin b , $\text{acc}(B_b)$ is the empirical accuracy in that bin, and $\text{conf}(B_b)$ is the mean predicted confidence. A perfectly calibrated model yields an [ECE](#) of zero.

- **Brier Score:** The Brier Score measures the mean squared difference between the predicted probability $\hat{p}_i$ and the true binary label y_i :

$$\text{BS} = \frac{1}{N} \sum_{i=1}^N (\hat{p}_i - y_i)^2. \quad (7.8)$$

Lower values indicate predictions that are simultaneously more accurate and properly confident when correct.

7.5.6 Physical Interpretability

To verify that the classifier’s decision boundary is grounded in physically meaningful parameters rather than statistical artifacts, [SHAP](#) values are computed for classifiers trained on the [MCMC](#) teacher’s posteriors. The teacher is selected over the [cMAF](#) student because its posteriors depend only on the forward model and posterior inference, whereas the student’s posteriors are additionally shaped by the synthetic training corpus and the three-phase distillation protocol. Restricting the analysis to the teacher allows the identified features to be attributed to the physical parameters rather than to artifacts introduced during distillation. [SHAP](#) values assign each input feature a contribution score representing its marginal impact on the model output, computed via the Shapley value framework from cooperative game theory. This analysis identifies which peak morphology parameters drive the classification decision and validates that the identified features are consistent with the established physical mechanisms of fatigue-induced lattice distortion.

7.6 Classification Models

The posterior-derived feature vectors are evaluated using three standard supervised classifiers. These classifiers span a range of model capacities and inductive biases, enabling the assessment of whether the feature representations are linearly separable or require complex non-linear decision boundaries.

- **Logistic Regression (LR):** A linear classifier that models the log-odds of class membership as a linear function of the input features. Logistic Regression [\[5\]](#) serves as the linear baseline and provides a direct measure of the linear separability of the posterior feature space. L2 regularization is applied, with the regularization strength selected via inner cross-validation on the training folds.
- **Random Forest (RF):** An ensemble of decision trees trained via bootstrap aggregation. Random Forest [\[3\]](#) captures non-linear feature interactions and provides built-in feature importance estimates that complement the [SHAP](#) analysis. The number of trees and maximum depth are treated as hyperparameters.
- **XGBoost:** A gradient-boosted ensemble of decision trees that iteratively fits residuals. XGBoost [\[4\]](#) is included as the highest-capacity classifier, capable of capturing complex non-linear relationships that simpler models may miss. Learning rate, maximum depth, and the number of estimators are treated as hyperparameters.

All classifiers are trained and evaluated strictly within the [LOGOCV](#) framework. Hyperparameter selection for each classifier is performed via an inner cross-validation loop on the training folds only, ensuring no test-set information leaks into model selection. The identical hyperparameter search procedure is applied across all combinations of inference engine and feature set to ensure fair comparison.

7.7 Baselines

Two baselines and one reference are included to contextualize the performance of the proposed framework.

BayesFlow Baseline. A BayesFlow normalizing flow is trained as a standard amortized inference baseline, using the identical [LOGOCV](#) structure as the proposed [cMAF](#) student. BayesFlow employs a summary network to compress the observed spectrum and a coupling-flow posterior estimator, but it does not incorporate the domain-specific residual convolutional encoder or the three-phase simulation-to-reality fine-tuning protocol. This baseline isolates the contribution of the physics-guided architectural choices and the fine-tuning protocol. Downstream classification performance is reported for BayesFlow under the same metrics as the [cMAF](#) student.

End-to-End CNN Baseline. An end-to-end convolutional classifier is trained as a black-box baseline that bypasses posterior inference entirely, corresponding to the direct formulation described in [section 3.5](#). It uses the same 1D residual convolutional encoder as the [cMAF](#) student ([section 6.3](#)), followed by a linear classification head, and is trained directly on the preprocessed spectra to predict the damage class. It is trained under the same [LOGOCV](#) structure, with early stopping performed on held-out training specimens that together contain both classes. This baseline isolates the contribution of the physics-based posterior representation, since it has access to the same spectra but no physical model, no posterior uncertainty, and no synthetic training data.

MCMC Teacher Reference. The [MCMC](#) teacher’s posteriors are treated as the reference upper bound for posterior quality, since the student is trained to reproduce them. For downstream classification, the teacher’s performance serves as the reference against which the student is compared. Any performance gap between the [cMAF](#) student and the [MCMC](#) teacher directly quantifies the cost of amortization.

7.8 Summary

This chapter has defined the complete evaluation methodology across multiple dimensions of the framework assessment: spectral fitting quality, posterior fidelity, computational efficiency, and downstream classification performance. The strict [LOGOCV](#) protocol ensures that all performance estimates reflect true generalization to physically unseen specimens. Metrics are formally defined at three levels of granularity: physical calibration (R^2 and [RMSE](#)), posterior geometry (W_1 , MMD^2), and classification utility (Accuracy, Macro-F1, [AUC](#), [ECE](#), Brier Score, and [SHAP](#)). The following chapter applies this methodology to present the quantitative findings and experimental results.

Chapter 8

Experimental Results

This chapter presents the quantified findings of the framework proposed in [chapter 5](#), instantiated in [chapter 6](#), and evaluated under the methodology defined in [chapter 7](#). Results are organized to follow the evaluation dimensions in sequence: spectral fitting quality, posterior quality, computational efficiency, and downstream classification performance. All metrics are computed under the [LOGOCV](#) protocol across the $S = 8$ physical specimens of the AA6061-T6 dataset. As described in [chapter 6](#), an additional experiment utilizing localized window fitting is conducted. The results for this localized approach are reported specifically within the spectral fitting and predictive performance sections. The interpretation and broader implications of these results are reserved for the subsequent discussion in [chapter 9](#).

8.1 Spectral Fitting Quality

The foundational requirement of the proposed framework is that the hierarchical Bayesian model must provide an accurate generative description of the observed [XRD](#) data. Spectral fitting quality is assessed exclusively for the [MCMC](#) teacher via [PPC](#). This compares the median reconstructed spectrum against the observed photon counts using the Coefficient of Determination (R^2) and [RMSE](#) on the normalized intensity scale.

8.1.1 Global Fitting

Under the global fitting paradigm, the **MCMC** teacher spans the full acquisition domain $\mathcal{T}_{\text{common}}$. The teacher achieves near-perfect spectral reconstruction, recording an R^2 of 0.992 and an **RMSE** of 0.007. This confirms that the hierarchical forward model, utilizing asymmetric **sPV** profiles alongside a heteroscedastic noise parameterization, successfully captures the physical variations across all damage states. **Figure 8.1** visualizes this fit, demonstrating that the 90% posterior predictive intervals consistently enclose the observed photon counts. This high fidelity establishes the global **MCMC** posteriors as a robust, physically calibrated ground truth for subsequent neural distillation.

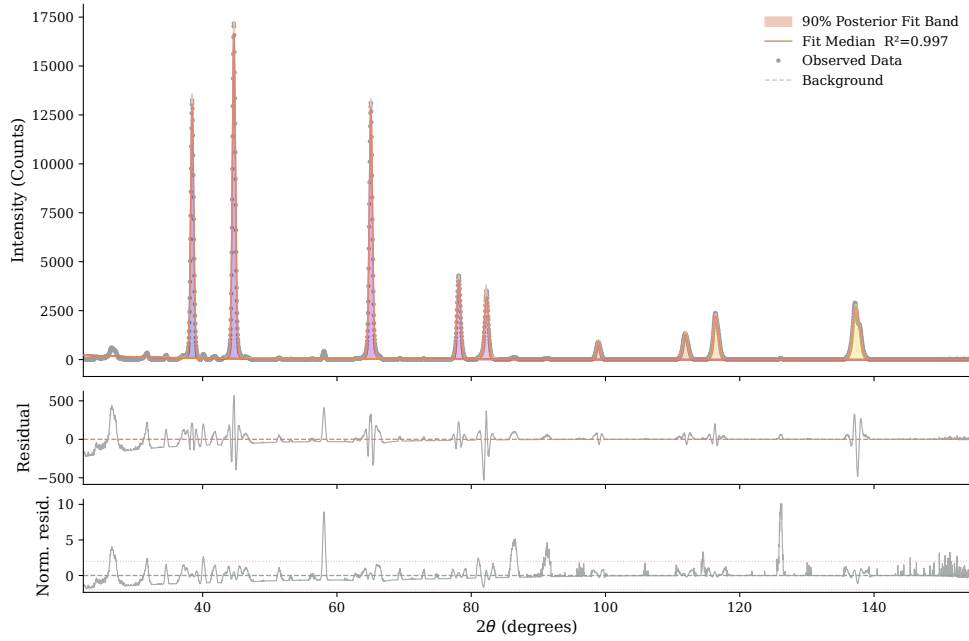


Figure 8.1: Posterior predictive checks for the **MCMC** teacher under the global fitting paradigm. The solid line denotes the median reconstruction and the shaded region denotes the 90% posterior predictive interval. Observed photon counts are shown as dots.

8.1.2 Localized Window Fitting

To address potential residual intensity features between the primary **FCC** peaks that are not captured by the pure aluminum model, a localized window fitting experiment was

conducted. This approach restricts the likelihood evaluation to a $\pm 3^\circ$ angular window centered exclusively on each target reflection. Figure 8.2 presents the posterior predictive reconstructions for these isolated regions. By rendering the model mathematically blind to inconsistent secondary phases outside the targeted window, the localized fits produce highly stable marginal distributions for individual peaks.

To quantitatively assess the fidelity of this localized approach, the reconstruction metrics were computed individually for each of the nine fundamental FCC reflections. Table 8.1 summarizes the median R^2 and Normalized RMSE aggregated across all scans for each reflection.

Table 8.1: Spectral fitting quality for the localized window fitting paradigm. Metrics are computed independently for each of the nine fundamental FCC reflections and reported as medians across all $N = 106$ scans.

Reflection (hkl)	Median $R^2 \uparrow$	Median Norm. RMSE $\downarrow$
(111)	1.000	0.0028
(200)	1.000	0.0024
(220)	1.000	0.0034
(311)	1.000	0.0037
(222)	0.988	0.0334
(400)	0.997	0.0140
(331)	0.999	0.0118
(420)	0.997	0.0185
(422)	0.966	0.0697
Overall Aggregate	0.999	0.0111

8.2 Posterior Quality

Once the MCMC teacher is validated as a ground-truth reference, the cMAF student must be evaluated on its ability to accurately recover this posterior geometry while significantly accelerating inference time.

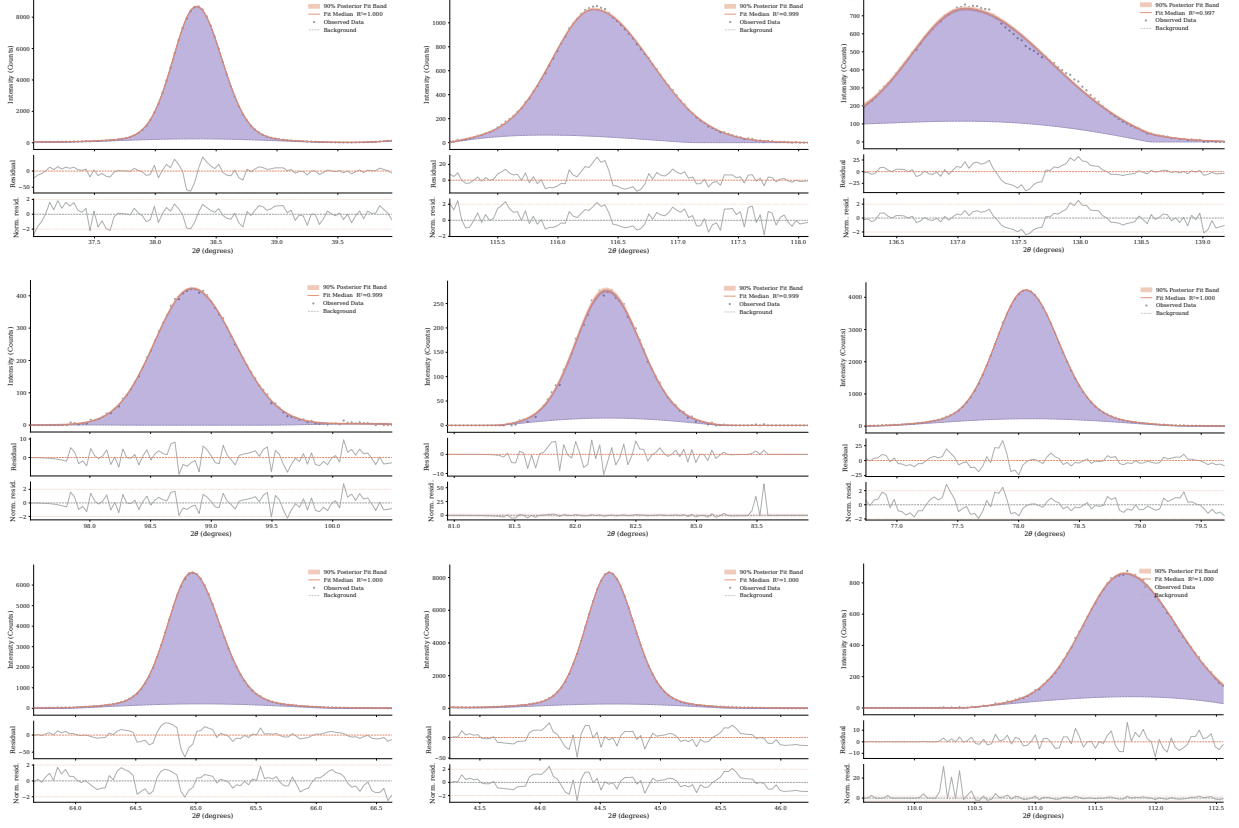


Figure 8.2: Posterior predictive checks for the **MCMC** teacher under the localized window fitting paradigm. The 3×3 grid represents the isolated fitting regions for the fundamental **FCC** reflections. The solid line denotes the median reconstruction and the shaded region denotes the 90% posterior predictive interval.

8.2.1 Physical Calibration

To verify that the neural student retains the physical meaning of the extracted posteriors, **PPC** metrics are applied directly to the **cMAF** student’s output. The student achieves an R^2 of 0.918 and an **RMSE** of 0.020, as reported in Table 8.2. While this represents a modest reduction in reconstruction fidelity relative to the exact teacher ($R^2 = 0.992$), the student successfully preserves the capacity to reconstruct the original photon counts. Figure 8.3 visually compares the reconstructions. The **cMAF** student’s posterior predictive intervals remain well-formed and physically plausible.

Table 8.2: Spectral fitting quality via Posterior Predictive Checks. Metrics are aggregated across all $N = 106$ scans under **LOGOCV**. R^2 closer to 1 and **RMSE** closer to 0 indicate higher fidelity reconstruction.

Method	$R^2 \uparrow$	RMSE $\downarrow$
MCMC Teacher	0.992	0.007
cMAF Student	0.918	0.020

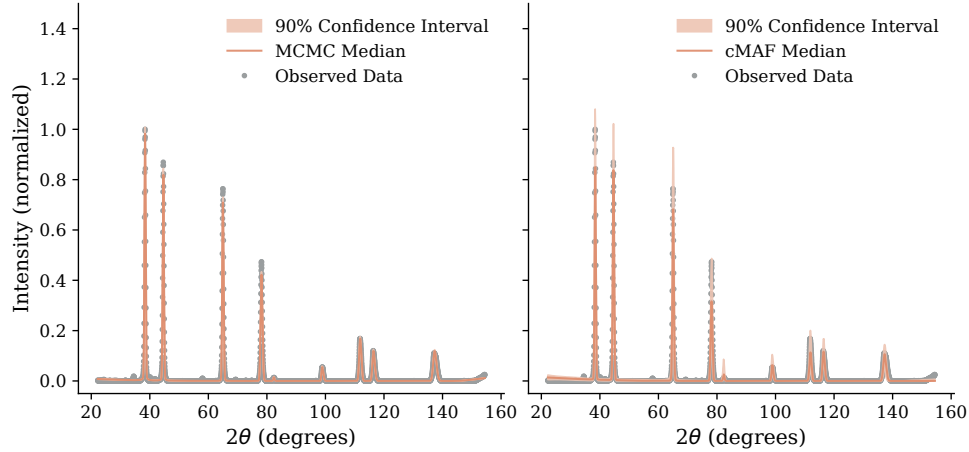


Figure 8.3: Posterior predictive checks comparing the **MCMC** teacher (left column) and the **cMAF** student (right column). The solid line denotes the median reconstruction and the shaded region denotes the 90% posterior predictive interval.

8.2.2 Posterior Geometry

The geometric fidelity of the amortized student posteriors relative to the [MCMC](#) reference is quantified using mathematical discrepancy metrics, reported in [Table 8.3](#).

Table 8.3: Posterior geometry discrepancy. Lower values of W_1 and MMD^2 indicate closer geometric agreement with the exact [MCMC](#) teacher’s posterior distribution. Both metrics are computed on the peak positions and [FWHM](#) parameters.

Method	1-Wasserstein (W_1) ↓	MMD^2 ↓
MCMC Teacher	Reference	
cMAF Student	3.08	0.395

The [cMAF](#) student achieves a W_1 of 3.08 and an MMD^2 of 0.395 relative to the teacher.

8.3 Computational Efficiency

The core justification for neural distillation is the reduction of computational latency at inference time while providing comparable performance. [Table 8.4](#) reports the inference time required to process an evaluation set, measured under identical hardware conditions on an NVIDIA H100 NVL GPU.

Table 8.4: Computational efficiency comparison. Inference time is reported per evaluation set (all scans in a single [LOGOCV](#) test fold). The [MCMC](#) teacher requires joint processing due to hierarchical coupling, whereas the [cMAF](#) student processes scans independently.

Method	Inference Time ↓	Complexity	Per-Scan Decomposable
MCMC Teacher	103.4 min	$\mathcal{O}(T)$	No
cMAF Student	20.07 ms	$\mathcal{O}(1)$	Yes

The hierarchical [MCMC](#) teacher requires 103.4 minutes per evaluation set. This latency is structurally unavoidable because the partial pooling mechanism forces joint optimization over all scans at $\mathcal{O}(T)$ complexity. By contrast, the [cMAF](#) student successfully amortizes this population-level structure during offline training. At test time, the student processes each scan independently by sampling directly from the trained flow, reducing the inference latency to just 20.07 milliseconds at $\mathcal{O}(1)$ complexity, enabling real-time diagnostic capabilities.

8.4 Downstream Classification Performance

The downstream utility of the inferred posteriors is evaluated on the binary structural damage classification task under the LOGOCV protocol. To isolate the effects of uncertainty propagation, results are reported independently for the point-estimate features (X_{mean}) and the distributional features (X_{dist}). Furthermore, the standard BayesFlow architecture is reintroduced here as a simulation-based discriminative baseline.

8.4.1 Predictive Performance

Global Fitting Framework

The predictive classification results for the primary global framework, including Accuracy, Macro-F1, and AUC, are detailed in Table 8.5.

Table 8.5: Downstream predictive performance under LOGOCV using the global fitting paradigm. Metrics are computed on the pooled out-of-fold predictions of all test folds. The CNN baseline operates directly on the preprocessed spectra and therefore has a single feature set. Best results per inference source and feature set are shown in **bold**.

Inference Source	Classifier	Mean Features (X_{mean})			Mean + Std Features (X_{dist})		
		Acc $\uparrow$	Macro-F1 $\uparrow$	AUC $\uparrow$	Acc $\uparrow$	Macro-F1 $\uparrow$	AUC $\uparrow$
MCMC Teacher	Logistic Regression	0.962	0.962	0.990	0.971	0.971	0.980
	Random Forest	0.971	0.971	0.964	0.962	0.962	0.973
	XGBoost	0.942	0.942	0.948	0.942	0.942	0.944
cMAF Student	Logistic Regression	0.817	0.817	0.914	0.837	0.836	0.906
	Random Forest	0.952	0.952	0.977	0.952	0.952	0.981
	XGBoost	0.952	0.952	0.960	0.952	0.952	0.960
BayesFlow Baseline	Logistic Regression	0.856	0.856	0.895	0.808	0.807	0.826
	Random Forest	0.875	0.875	0.894	0.865	0.865	0.871
	XGBoost	0.865	0.865	0.923	0.837	0.837	0.887
CNN Baseline	End-to-end	0.311	0.311	0.486	–	–	–

The exact MCMC teacher establishes the reference downstream performance, achieving a peak accuracy and Macro-F1 of 0.971 with Logistic Regression on X_{dist} and Random Forest on X_{mean} . The proposed cMAF student achieves 0.952 accuracy and Macro-F1 with both Random Forest and XGBoost on either feature set, 0.019 below the teacher. Notably,

the **cMAF** student experiences a degradation in performance when evaluated with the linear Logistic Regression classifier (0.817 accuracy).

Adding posterior standard deviations (X_{dist}) does not consistently improve accuracy. It raises the Logistic Regression accuracy of both the **cMAF** student (from 0.817 to 0.837) and the **MCMC** teacher (from 0.962 to 0.971), but lowers the teacher’s Random Forest accuracy from 0.971 to 0.962 and the accuracy of all three BayesFlow classifiers. It leaves the student’s Random Forest and XGBoost accuracy unchanged at 0.952.

Both baselines fall below the best teacher and student configurations. The BayesFlow baseline reaches accuracies between 0.808 and 0.875 and **AUC** values between 0.826 and 0.923. The end-to-end CNN, which shares the encoder architecture of the student but is trained directly on the preprocessed spectra, reaches an accuracy of 0.311 and an **AUC** of 0.486.

Table 8.6 reports the accuracy of each configuration for every held-out specimen, together with the mean and standard deviation across specimens. The mean per-specimen accuracy ranges from 0.942 to 0.971 for the teacher, from 0.792 to 0.950 for the student, and from 0.795 to 0.859 for the BayesFlow baseline, and is 0.404 for the CNN baseline.

Localized Fitting Framework

Table 8.7 reports the predictive performance metrics generated utilizing the posteriors extracted from the localized window fitting experiment.

By isolating the high-confidence peak morphology features from broader background noise and secondary phase anomalies, the localized fitting approach yields exceptional discriminative capacity. The XGBoost classifier reaches an accuracy of 0.991 utilizing both feature sets, the highest accuracy observed across all configurations.

Table 8.8 reports the corresponding per-specimen accuracy. The mean per-specimen accuracy ranges from 0.937 to 0.990.

8.4.2 Uncertainty Quantification

The quality of the calibrated confidence estimates for the downstream predictions is presented in **Table 8.9**.

The **MCMC** teacher and the **cMAF** student reach comparable calibration, with **ECE** values between 0.071 and 0.123 for the teacher and between 0.072 and 0.107 for the student.

Table 8.6: Per-specimen accuracy under LOGOCV using the global fitting paradigm. Each column reports the accuracy on the scans of one held-out specimen, and the last column reports the mean and standard deviation across the held-out specimens. The corresponding pooled accuracies are reported in Table 8.5.

Inference Source	Classifier	Features	Held-out Specimen								Mean $\pm$ SD
			1	2	3	4	5	6	7	8	
MCMC Teacher	Logistic Regression	X_{mean}	1.000	1.000	1.000	0.846	0.846	1.000	1.000	1.000	0.962 ± 0.071
		X_{dist}	1.000	1.000	1.000	0.923	0.846	1.000	1.000	1.000	0.971 ± 0.057
	Random Forest	X_{mean}	1.000	1.000	1.000	0.923	0.846	1.000	1.000	1.000	0.971 ± 0.057
		X_{dist}	1.000	1.000	1.000	0.846	0.846	1.000	1.000	1.000	0.962 ± 0.071
	XGBoost	X_{mean}	1.000	1.000	1.000	1.000	0.538	1.000	1.000	1.000	0.942 ± 0.163
		X_{dist}	1.000	1.000	1.000	1.000	0.538	1.000	1.000	1.000	0.942 ± 0.163
cMAF Student	Logistic Regression	X_{mean}	1.000	1.000	0.750	0.538	0.385	1.000	0.667	1.000	0.792 ± 0.245
		X_{dist}	1.000	1.000	1.000	0.538	0.385	1.000	0.750	1.000	0.834 ± 0.249
	Random Forest	X_{mean}	1.000	1.000	1.000	0.923	0.923	1.000	0.750	1.000	0.950 ± 0.088
		X_{dist}	1.000	1.000	1.000	0.923	0.923	1.000	0.750	1.000	0.950 ± 0.088
	XGBoost	X_{mean}	1.000	1.000	1.000	0.923	0.923	1.000	0.750	1.000	0.950 ± 0.088
		X_{dist}	1.000	1.000	1.000	0.923	0.923	1.000	0.750	1.000	0.950 ± 0.088
BayesFlow Baseline	Logistic Regression	X_{mean}	1.000	0.895	1.000	0.538	0.769	0.917	0.833	0.917	0.859 ± 0.151
		X_{dist}	0.895	0.895	0.750	0.538	0.615	1.000	0.833	0.833	0.795 ± 0.154
	Random Forest	X_{mean}	1.000	1.000	0.750	0.615	0.615	1.000	0.833	1.000	0.852 ± 0.173
		X_{dist}	1.000	1.000	0.750	0.538	0.615	1.000	0.833	1.000	0.842 ± 0.190
	XGBoost	X_{mean}	1.000	1.000	0.750	0.538	0.692	0.833	0.917	1.000	0.841 ± 0.171
		X_{dist}	1.000	1.000	1.000	0.538	0.615	0.750	0.917	0.833	0.832 ± 0.182
CNN Baseline	End-to-end	–	0.000	0.000	1.000	0.615	0.615	0.000	1.000	0.000	0.404 ± 0.456

Table 8.7: Localized fitting downstream binary classification performance (Nominal versus Damaged) evaluated under LOGOCV. Metrics are computed on the pooled out-of-fold predictions of all test folds. Best result per column is shown in **bold**.

Classifier	Mean Features (X_{mean})			Mean + Std Features (X_{dist})		
	Acc $\uparrow$	Macro-F1 $\uparrow$	AUC $\uparrow$	Acc $\uparrow$	Macro-F1 $\uparrow$	AUC $\uparrow$
Logistic Regression	0.953	0.953	0.993	0.953	0.953	0.993
Random Forest	0.981	0.981	0.998	0.981	0.981	0.998
XGBoost	0.991	0.991	1.000	0.991	0.991	1.000

Table 8.8: Per-specimen accuracy under LOGOCV using the localized fitting paradigm. Each column reports the accuracy on the scans of one held-out specimen, and the last column reports the mean and standard deviation across the held-out specimens. Specimens are numbered as in Table 8.6, and the corresponding pooled accuracies are reported in Table 8.7.

Classifier	Features	Held-out Specimen								Mean $\pm$ SD
		1	2	3	4	5	6	7	8	
Logistic Regression	X_{mean}	1.000	1.000	0.800	0.923	0.769	1.000	1.000	1.000	0.937 ± 0.098
	X_{dist}	1.000	1.000	0.800	0.923	0.769	1.000	1.000	1.000	0.937 ± 0.098
Random Forest	X_{mean}	1.000	1.000	0.800	1.000	0.923	1.000	1.000	1.000	0.965 ± 0.072
	X_{dist}	1.000	1.000	0.800	1.000	0.923	1.000	1.000	1.000	0.965 ± 0.072
XGBoost	X_{mean}	1.000	1.000	1.000	1.000	1.000	0.917	1.000	1.000	0.990 ± 0.029
	X_{dist}	1.000	1.000	1.000	1.000	1.000	0.917	1.000	1.000	0.990 ± 0.029

Table 8.9: Uncertainty quantification metrics under LOGOCV, computed on the pooled out-of-fold predictions. Lower ECE indicates better alignment between predicted probabilities and empirical frequencies. Lower Brier Score indicates more accurate and confident predictions. Best result per column is shown in **bold**.

Inference Source	Classifier	Mean Features (X_{mean})		Mean + Std Features (X_{dist})	
		ECE $\downarrow$	Brier $\downarrow$	ECE $\downarrow$	Brier $\downarrow$
MCMC Teacher	Logistic Regression	0.074	0.036	0.071	0.036
	Random Forest	0.101	0.051	0.087	0.048
	XGBoost	0.123	0.073	0.122	0.073
cMAF Student	Logistic Regression	0.093	0.114	0.072	0.115
	Random Forest	0.106	0.065	0.099	0.060
	XGBoost	0.099	0.053	0.107	0.053
BayesFlow Baseline	Logistic Regression	0.116	0.111	0.157	0.158
	Random Forest	0.103	0.111	0.157	0.132
	XGBoost	0.068	0.105	0.110	0.119
CNN Baseline	End-to-end	0.370	0.286	–	–

The teacher attains the lowest Brier score (0.036, Logistic Regression), while the lowest Brier score of the student is 0.053 (XGBoost). The BayesFlow baseline shows ECE values between 0.068 and 0.157 and Brier scores between 0.105 and 0.158, higher than the best teacher and student values, and the CNN baseline shows the poorest calibration (ECE 0.370, Brier 0.286).

8.4.3 Physical Interpretability

To verify the physical validity of the decision boundaries driving the predictive performance, SHAP values were computed for the MCMC teacher’s XGBoost classifier. As described in subsection 7.5.6, the teacher is selected for interpretability analysis because its posteriors are free from the effects of the distillation process. Figure 8.4 visualizes the SHAP summary plot, ranking features by their mean absolute SHAP value across all LOGOCV training and test fold predictions.

The SHAP analysis consistently identifies the left and right FWHM parameters of the (311), (200), (422), and (111) reflections (specifically the `fwhml_311`, `fwhmr_311`, `fwhmr_200`, `fwhmr_422`, `fwhml_422`, and `fwhml_111` physical parameters) as the dominant features contributing to the damage classification. High FWHM values strongly drive the model’s predictions toward the damaged class. The stability of these feature rankings across both the point-estimate and distributional feature sets indicates that the models rely on consistent structural features.

8.5 Summary

This chapter has reported the quantitative findings of the framework evaluation across all evaluation dimensions. The MCMC teacher achieves near-perfect spectral reconstruction ($R^2 = 0.992$) and serves as the reference posterior for all downstream tasks. The cMAF student achieves $R^2 = 0.918$, $W_1 = 3.08$, and $\text{MMD}^2 = 0.395$ relative to the teacher, while reducing inference time from 103.4 minutes to 20.07 milliseconds. In the downstream classification task, the amortized posteriors reach an accuracy of 0.952, close to the exact Bayesian teacher (0.971) and above the BayesFlow (0.875) and CNN (0.311) baselines, with calibration comparable to the teacher. Furthermore, the localized fitting approach isolates key microstructural parameters to achieve a peak diagnostic accuracy of 0.991. The SHAP analysis confirms that these predictions rely on fundamental physical parameters. The implications of these results are explored in detail in chapter 9.

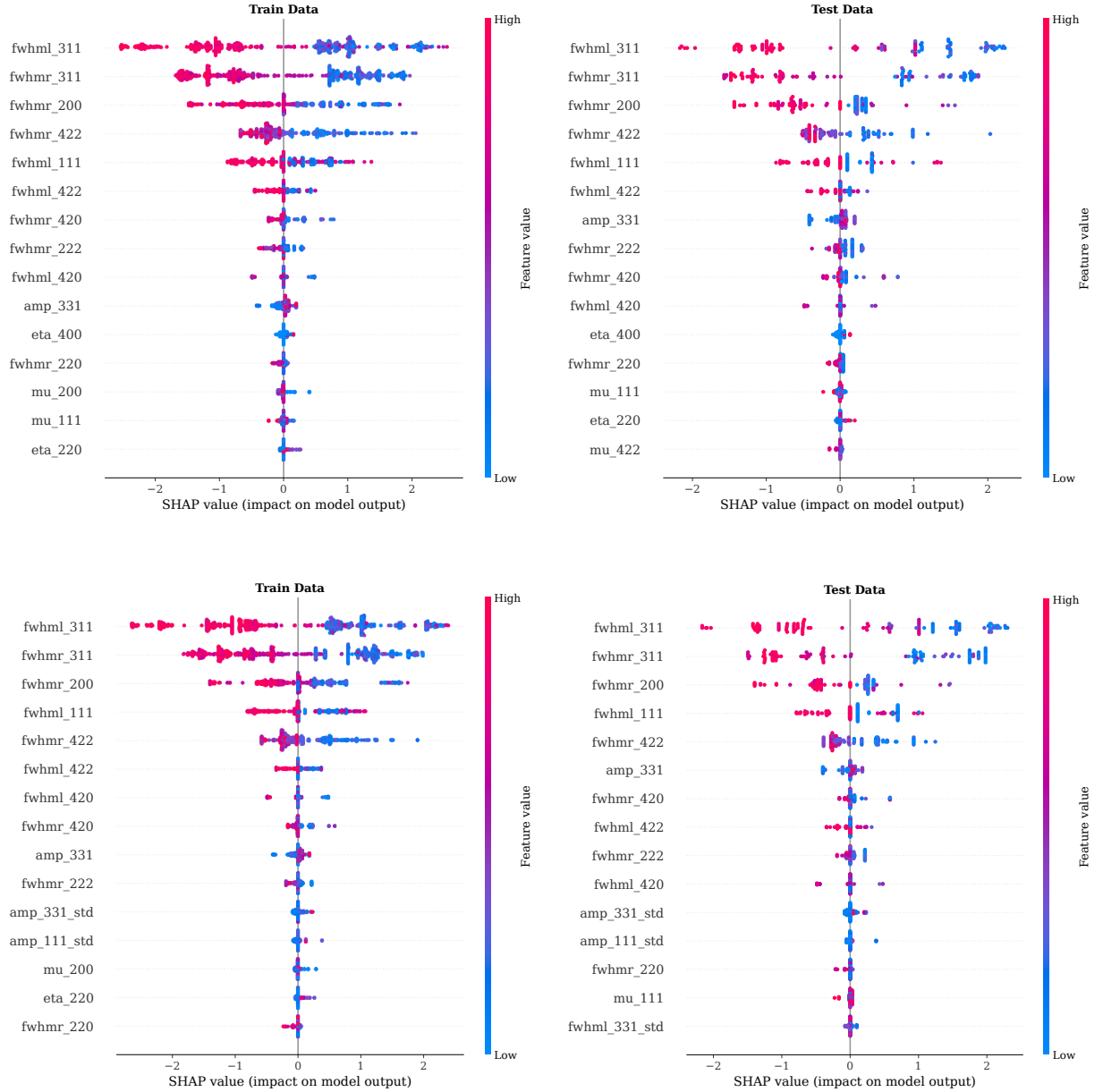


Figure 8.4: SHAP feature importance for the [MCMC](#) teacher XGBoost classifier under X_{mean} (top row) and X_{dist} (bottom row), evaluated on training (left) and testing (right) folds. Each dot represents one scan; color encodes normalized feature value (red: high, blue: low).

Chapter 9

Discussion and Analysis

This chapter interprets the quantitative findings reported in [chapter 8](#) within the context of the core problem formulated in [chapter 3](#). The analysis proceeds across four central themes: the physical validity of the Bayesian teacher’s posteriors, the fidelity and computational cost of neural amortization, the downstream utility of uncertainty-aware feature representations, and the physical grounding of the classification decisions. The chapter concludes by discussing the generalization of the framework to other materials, explicitly revisiting the research questions of this thesis, and outlining the limitations of the proposed framework.

9.1 Physical Validity of the Bayesian Teacher

The hierarchical [MCMC](#) teacher achieves a spectral reconstruction R^2 of 0.992 and an [RMSE](#) of 0.007 across all $N = 106$ scans under [LOGOCV](#), with 90% posterior predictive intervals consistently enclosing the observed photon counts. These results confirm that the combination of asymmetric [sPV](#) profiles, a heteroscedastic noise model, and hierarchical partial pooling is sufficient to provide an accurate generative description of the AA6061-T6 [XRD](#) dataset across all damage states. The near-unity R^2 establishes the teacher as a credible, physically calibrated ground truth, validating the foundational premise of the distillation framework: a high-fidelity Bayesian reference posterior can indeed be constructed from the available scarce data.

The role of partial pooling in achieving this fidelity deserves specific attention. The AA6061-T6 dataset exhibits a heavily right-skewed intensity distribution, featuring a small

number of high-intensity scans and a long tail of low-count observations. Under an individual scan paradigm, scans in this low-intensity tail would lack sufficient photon counts to constrain the higher-angle reflections, producing diffuse posteriors that artificially drift toward the prior boundaries. The hierarchical partial pooling mechanism automatically regularizes these weak-evidence scans toward the population-level hyperpriors, effectively borrowing statistical strength from all scans. The uniformly high R^2 across all damage states demonstrates that this regularization is highly effective. The teacher does not sacrifice accuracy on low signal to noise ratio scans to achieve high accuracy on high signal to noise ratio scans.

The localized window fitting experiment provides complementary evidence for the physical validity of the posteriors. By restricting the likelihood evaluation to a $\pm 3^\circ$ window centered on each FCC reflection, the localized fits isolate the peak morphology posteriors from the influence of inter-peak residual features arising from unmodeled secondary phases. The resulting downstream classification accuracy of 0.991 with XGBoost, the highest observed across all configurations, indicates that the localized posteriors carry a more concentrated microstructural signal than the global posteriors. This suggests that the modest contribution of secondary-phase residuals in the global fit introduces a small amount of noise into the fine-grained peak parameter estimates. While this noise does not materially degrade the teacher’s global spectral reconstruction quality, eliminating it maximizes the discriminative signal required for optimal damage classification.

9.2 Fidelity and Cost of Amortization

The central operational question for the neural student is whether the massive gains in computational efficiency come at an acceptable cost to physical fidelity. The empirical results provide a clear, quantified answer.

9.2.1 Bridging the Simulation-to-Reality Gap

A pervasive challenge in applying deep learning to scientific inverse problems is the simulation-to-reality gap. When models are trained exclusively on idealized forward simulations, they frequently fail upon encountering the complex, correlated noise structures of empirical data. The proposed cMAF student addresses this gap, achieving $R^2 = 0.918$ and $W_1 = 3.08$. This is attributed to the physics-guided three-phase training protocol. Rather than relying on naive forward simulation, the framework generated a synthetic training corpus that actively preserved the cross-parameter correlations of the exact MCMC posteriors

via Gaussian copulas and targeted parameter jittering. Furthermore, the protocol allowed the network to learn the broad topological structure of the crystallographic space during pre-training before adapting to the specific empirical noise floor during the frozen-encoder fine-tuning phase. The synthetic replay penalty successfully prevented the network from catastrophically forgetting the strict physical boundaries of the parameter space.

9.2.2 Computational Cost of Amortization

The inference latency reduction from 103.4 minutes (MCMC teacher) to 20.07 milliseconds (cMAF student) represents a speedup factor of approximately $309,000\times$. This reduction has a profound structural interpretation. The MCMC teacher’s latency is fundamentally tied to the hierarchical partial pooling mechanism. Because all N scans are modeled jointly under shared hyperpriors, the teacher cannot decompose inference across scans. Every new observation conceptually requires re-running the full hierarchical sampler to update the population-level hyperpriors.

The cMAF student, by contrast, absorbs this population-level structure into its encoder weights during training. At test time, each scan is processed independently by sampling directly from the trained flow, making the student’s inference latency invariant to both dataset size and the number of MCMC chains conventionally required for convergence. This transformation from $\mathcal{O}(T)$ to $\mathcal{O}(1)$ complexity makes the student uniquely suitable for deployment contexts that are structurally inaccessible to the teacher, such as streaming sensor data processing or high-throughput laboratory screening workflows.

9.3 Downstream Utility of Uncertainty-Aware Representations

9.3.1 Effect of Uncertainty Features

Adding posterior standard deviations (X_{dist}) to the posterior means (X_{mean}) does not consistently improve classification. With Logistic Regression, it raises the accuracy of both the cMAF student (0.817 to 0.837) and the MCMC teacher (0.962 to 0.971), whereas it lowers the accuracy of the teacher’s Random Forest (0.971 to 0.962) and of all three BayesFlow classifiers. For the student’s Random Forest and XGBoost classifiers, both feature sets yield the same accuracy (0.952). The posterior means therefore carry most of the discriminative signal.

9.3.2 Calibration Quality

The [cMAF](#) student and the [MCMC](#) teacher reach comparable [ECE](#) values (0.072 to 0.107 and 0.071 to 0.123, respectively), while the teacher retains a lower Brier score (0.036 versus 0.053) that reflects its higher accuracy. The [ECE](#) values of the BayesFlow baseline overlap with those of the teacher and the student, but its Brier scores (0.105 to 0.158) are higher than the best teacher and student values. The distillation therefore preserves the calibration of the downstream probabilities.

9.4 Physical Grounding of Classification Decisions

To trust a predictive model, domain experts must verify that the model is making decisions based on valid physical phenomena.

The [SHAP](#) analysis of the [MCMC](#) teacher’s XGBoost classifier provides direct evidence that the decisions are grounded in physically meaningful microstructural parameters. The dominant features identified are the left and right [FWHM](#) parameters of the (311), (200), (422), and (111) [FCC](#) reflections, with high [FWHM](#) values consistently driving predictions toward the Damaged class. These mathematical findings align perfectly with established materials science theory.

The asymmetric nature of the broadening, captured by the independently parameterized left and right [FWHM](#) parameters, further validates the choice of the [sPV](#) profile. The [SHAP](#) analysis shows that both $\Gamma_{L,i,k}$ and $\Gamma_{R,i,k}$ contribute independently to the classification.

9.5 Generalization to Other Materials

The framework separates components that are specific to AA6061-T6 from components that are independent of the material system. Apart from assuming a fixed number of peaks, the problem formulation in [chapter 3](#), the hierarchical partial pooling, the correlation-preserving synthetic generation, and the three-phase training protocol make no assumption about the material. The material enters the framework through the forward model, specifically through the set of permitted reflections, their theoretical Bragg positions, and the priors placed on the peak parameters.

In its current form, the forward model assumes that the number of peaks K and their approximate positions are known in advance. This assumption holds for materials with a well-defined crystal structure, where Bragg’s Law and the selection rules fix the characteristic reflections at specific 2θ values. Extending the framework to another crystalline material therefore requires recomputing the permitted reflections from its crystal structure and lattice parameters. A different number of reflections can affect the computational cost of [MCMC](#) sampling but does not change the structure of the framework.

The assumption does not hold for materials without a fixed set of reflections, such as semi-crystalline polymers. Extending the framework to such cases would require changes to three components:

- **Teacher:** The number of peaks, their positions, and their shapes must be inferred jointly from the data, since they cannot be fixed in advance. This could be achieved through Bayesian model selection over K , with hierarchical priors that accommodate peaks present in only a subset of scans.
- **Synthetic Generation:** The Gaussian copula is fitted to parameter vectors of fixed dimension D . When K varies across scans, the posterior pool contains vectors of different dimensions, and the copula must be adapted accordingly.
- **Student:** The output dimension of the [cMAF](#) student is tied to K . A variable number of peaks would require a student that can represent parameter vectors of varying dimension.

9.6 Research Questions Revisited

The experimental findings can now be mapped directly onto the core problems established in this thesis.

Is it possible to establish a physically accurate, uncertainty-aware ground truth for microstructural parameters from a limited dataset of noisy [XRD](#) spectra?

Yes. The [MCMC](#) teacher achieves $R^2 = 0.992$ with well-formed posterior predictive intervals across all damage states and signal to noise ratio regimes. The partial pooling mechanism successfully regularizes the low signal to noise ratio scans toward the population-level hyperpriors without degrading the fits of high signal to noise ratio observations.

Is it possible to distill the posterior geometry of an exact Bayesian model into a neural network to achieve real-time inference without sacrificing physical fidelity or suffering from simulation-to-reality distribution shifts? Yes. The [cMAF](#) student achieves $R^2 = 0.918$ and $W_1 = 3.08$. The distillation reduces inference latency by orders of magnitude to 20.07 milliseconds.

Do the uncertainty-aware posteriors provide a robust feature representation for the binary classification of structural damage, and does this approach generalize effectively to unseen physical specimens? Yes. Under strict [LOGOCV](#) across distinct physical specimens, the amortized student reaches an accuracy of 0.952, close to the exact Bayesian teacher (0.971) and well above the BayesFlow (0.875) and end-to-end CNN (0.311) baselines, with calibration comparable to the teacher. Furthermore, the localized fitting ablation shows that refining the peak parameter posteriors raises diagnostic accuracy to 0.991.

Can the resulting classification decisions be attributed to physically meaningful parameters and validated against established crystallographic principles? Yes. [SHAP](#) analysis consistently identifies the left and right [FWHM](#) parameters of the (311), (200), (422), and (111) reflections as the dominant discriminative features, in direct alignment with the physical theory of strain broadening.

9.7 Limitations

Several limitations of the present study provide context for these results.

Dataset Scale. The AA6061-T6 dataset comprises $N = 106$ scans from $S = 8$ physical specimens, which is relatively small for deep learning standards. While the [LOGOCV](#) protocol maximizes the use of available data and enforces a strict generalization requirement, the evaluation is conducted on a single material system under a specific fatigue loading condition. The generalizability of the framework to other alloy systems or distinct diffractometer configurations remains to be empirically demonstrated.

Binary Classification Formulation. The downstream classification task merges Mid-Life and Late-Life scans into a single Damaged class to focus on the operationally relevant

detection objective. This formulation does not assess the framework’s ability to discriminate between progressive damage stages, which would be required for a full prognostic life-prediction capability.

Fixed Crystallographic Scope. The model is strictly restricted to the nine fundamental FCC reflections of the aluminum matrix, excluding secondary precipitate phases. While this choice is mathematically well-motivated to ensure sampler convergence, it means that any discriminative information encoded exclusively in the precipitate phase signatures is not utilized by the framework.

Chapter 10

Conclusion

This chapter summarizes the key research outcomes, discusses the broader implications of the findings, and outlines promising directions for future work. The limitations identified in this study provide a roadmap for advancing this research program in subsequent investigations.

10.1 Summary of Research Outcomes

This research successfully bridged the simulation-to-reality gap that typically hinders standard [SBI](#) methods in data-scarce empirical settings. By formalizing a Physics-Guided Teacher-Student Distillation Framework, the methodology decoupled the requirement for physical exactness from the requirement for operational speed.

The hierarchical Bayesian teacher, utilizing asymmetric [sPV](#) peak profiles, a heteroscedastic noise model, and partial pooling across all specimens, proved highly effective at extracting ground-truth posterior geometries from the AA6061-T6 fatigue dataset. By enforcing crystallographic constraints and hierarchical partial pooling, the teacher achieved a near-perfect spectral reconstruction ($R^2 = 0.992$) across a wide range of signal-to-noise ratios. This fidelity established the teacher as a robust, physically calibrated ground truth.

To overcome the iterative computational bottleneck of the teacher, this structural knowledge was distilled into a [cMAF](#) student. The success of this distillation was driven by a three-phase training protocol and a correlation-preserving synthetic generation strategy. The resulting amortized student achieved a 1-Wasserstein distance of 3.08 to the exact posterior and reduced inference latency from over 100 minutes to 20.07 milliseconds. This

acceleration transforms the inference paradigm from an iterative process requiring joint processing of all scans to amortized sampling that processes scans independently, enabling real-time deployment in contexts previously inaccessible to the teacher.

The downstream validation confirmed that the inferred posteriors provide an effective feature representation for damage detection. The amortized student reached a classification accuracy of 0.952, close to the exact Bayesian teacher (0.971), with comparable probability calibration. Furthermore, the [SHAP](#) analysis verified that the model’s predictions rely on fundamental physical parameters, such as the left and right [FWHM](#) parameters of several [FCC](#) reflections, consistent with the physics of fatigue degradation.

10.2 Broader Implications

Beyond the specific application to [XRD](#) fatigue analysis, this work demonstrates a general design principle for applying probabilistic machine learning to scientific inverse problems. The teacher-student distillation paradigm separates the requirements of physical accuracy from computational efficiency, allowing each to be addressed by the most appropriate tool. The exact Bayesian model provides the accuracy that domain constraints demand, while the neural student provides the throughput that operational deployment requires.

Additionally, propagating inferential uncertainty into the feature representation did not consistently improve classification. Posterior standard deviations improved Logistic Regression accuracy for both the student and the teacher, but reduced the accuracy of the teacher’s Random Forest and of all BayesFlow classifiers.

10.3 Future Work

The limitations identified in this study provide a roadmap for future research.

Prognostic Life Prediction. The present framework addresses the binary detection of structural damage. A natural extension is the three-class formulation (Pristine, Mid-Life, Late-Life), which would support prognostic life prediction in addition to diagnostic damage detection. The ultimate objective is to achieve a continuous regression of fatigue life remaining, which would require a more granular dataset with finer damage state resolution. Future research should leverage larger, time-resolved datasets to map the trajectory of latent peak parameters as material degradation evolves.

Material Generalization. The framework has so far been evaluated only on AA6061-T6. Extending it to materials without a fixed set of reflections, such as semi-crystalline polymers, would require inferring the number of peaks jointly with their shapes, as discussed in [section 9.5](#).

Online and Continual Learning. The current three-phase training protocol is offline. Extending this framework to a continual learning setting, where the student updates its posterior estimates incrementally as new spectra are acquired during component service life, would be a major leap forward.

References

- [1] Christophe Andrieu, Nando De Freitas, Arnaud Doucet, and Michael I. Jordan. An Introduction to MCMC for Machine Learning. *Machine Learning*, 50(1-2):5–43, January 2003.
- [2] William Henry Bragg and William Lawrence Bragg. The reflection of X-rays by crystals. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 88(605):428–438, 1913.
- [3] Leo Breiman. Random Forests. *Machine Learning*, 45(1):5–32, October 2001.
- [4] Tianqi Chen and Carlos Guestrin. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–794, San Francisco California USA, August 2016. ACM.
- [5] David R. Cox. The regression analysis of binary sequences. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 20(2):215–232, 1958.
- [6] Kyle Cranmer, Johann Brehmer, and Gilles Louppe. The frontier of simulation-based inference. *Proceedings of the National Academy of Sciences*, 117(48):30055–30062, December 2020.
- [7] Geovana Drumond, Bianca Pinheiro, Ilson Pasqualino, Francine Roudet, and Didier Chicot. Fatigue Life Prediction of Steel Pipelines Based on X-ray Diffraction Analyses. *Journal of Materials Engineering and Performance*, 31(1):801–813, January 2022.
- [8] Andrew Gelman, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin. *Bayesian Data Analysis*. Chapman and Hall/CRC, third edition, 2013.

- [9] Mathieu Germain, Karol Gregor, Iain Murray, and Hugo Larochelle. MADE: Masked autoencoder for distribution estimation. In *International Conference on Machine Learning*, pages 881–889. PMLR, 2015.
- [10] Sepehr Ghazimorady, Sina Negarandeh, Yi Chen, Melika Morafegh, Verena Kantere, and Hamid Jahed. X-ray diffraction-based estimation of remaining fatigue life in AA6061-T6 for “additive remanufacturing®”. In *Proceedings of the Holistic Innovation in Additive Manufacturing (HI-AM) Conference*, volume 1, 2025.
- [11] David Greenberg, Marcel Nonnenmacher, and Jakob Macke. Automatic posterior transformation for likelihood-free inference. In *International Conference on Machine Learning*, pages 2404–2414. PMLR, 2019.
- [12] Charles R. Harris, K. Jarrod Millman, Stéfan J. Van Der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. Van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández Del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. Array programming with NumPy. *Nature*, 585(7825):357–362, September 2020.
- [13] W. Keith Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.
- [14] Matthew D. Hoffman and Andrew Gelman. The No-U-Turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.*, 15(1):1593–1623, 2014.
- [15] Shuhei Kashiwamura, Shun Katakami, Ryo Yamagami, Kazunori Iwamitsu, Hiroyuki Kumazoe, Kenji Nagata, Toshihiro Okajima, Ichiro Akai, and Masato Okada. Bayesian Spectral Deconvolution of X-Ray Absorption Near Edge Structure Discriminating between High- and Low-Energy Domains. *Journal of the Physical Society of Japan*, 91(7):074009, July 2022.
- [16] Soon Ho Kim, Seok Swoo Cho, and Jung-Hyen Park. Prediction of Fatigue Life in 2024-T3 Aluminum Using X-ray half-value breadth. *International Journal of Precision Engineering and Manufacturing*, 3(2):78–86, 2002.
- [17] Alp Kucukelbir, Dustin Tran, Rajesh Ranganath, Andrew Gelman, and David M. Blei. Automatic differentiation variational inference. *Journal of machine learning research*, 18(14):1–45, 2017.

- [18] Han Liu, John Lafferty, and Larry Wasserman. The nonparanormal: Semiparametric estimation of high dimensional undirected graphs. *Journal of Machine Learning Research*, 10(10), 2009.
- [19] Jan-Matthis Lueckmann, Pedro J Goncalves, Giacomo Bassetto, Kaan Öcal, Marcel Nonnenmacher, and Jakob H Macke. Flexible statistical inference for mechanistic models of neural dynamics. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.
- [20] G. Maeder, J. L. Lebrun, and J. M. Sprauel. Present possibilities for the X-ray diffraction method of stress measurement. *NdT International*, 14(5):235–247, 1981.
- [21] Kenji Nagata, Seiji Sugita, and Masato Okada. Bayesian spectral deconvolution with the exchange Monte Carlo method. *Neural Networks*, 28:82–89, 2012.
- [22] Roger B. Nelsen. Dependence. In *An Introduction to Copulas*, pages 157–225. Springer New York, New York, NY, 2006.
- [23] R. N. Pangborn, S. Weissmann, and I. R. Kramer. PREDICTION OF FATIGUE LIFE BY X-RAY DIFFRACTION METHODS. *Fatigue & Fracture of Engineering Materials & Structures*, 1(3):363–369, April 1979.
- [24] R. N. Pangborn, S. Weissmann, and I. R. Kramer. Dislocation distribution and prediction of fatigue damage. *Metallurgical Transactions A*, 12(1):109–120, January 1981.
- [25] Juan Pantoja-Cortes, Florentino Sanchez-Bajo, and Angel L. Ortiz. A line-broadening analysis model for the microstructural characterization of nanocrystalline materials from asymmetric x-ray diffraction peaks. *Journal of Physics: Condensed Matter*, 24(21):215301, 2012.
- [26] George Papamakarios and Iain Murray. Fast ε -free inference of simulation models with bayesian conditional density estimation. *Advances in neural information processing systems*, 29, 2016.
- [27] George Papamakarios, Eric Nalisnick, Danilo Jimenez Rezende, Shakir Mohamed, and Balaji Lakshminarayanan. Normalizing Flows for Probabilistic Modeling and Inference. *Journal of Machine Learning Research*, 22:1–64, 2021.
- [28] George Papamakarios, Theo Pavlakou, and Iain Murray. Masked autoregressive flow for density estimation. *Advances in neural information processing systems*, 30, 2017.

- [29] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, and Luca Antiga. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.
- [30] Du Phan, Neeraj Pradhan, and Martin Jankowiak. Composable Effects for Flexible and Accelerated Probabilistic Programming in NumPyro, December 2019.
- [31] Bianca Pinheiro, Jacky Lesage, Ilson Pasqualino, Nouredine Benseddiq, and Edoardo Bemporad. Toward a Fatigue Life Assessment of Steel Pipes Based on X-Ray Diffraction Measurements. In *International Conference on Offshore Mechanics and Arctic Engineering*, volume 56529, page V05BT04A027. American Society of Mechanical Engineers, 2015.
- [32] BLJPIBEBE Pinheiro, Jacky Lesage, I. Pasqualino, E. Bemporad, and Nouredine Benseddiq. X-ray diffraction study of microstructural changes during fatigue damage initiation in pipe steels: Role of the initial dislocation structure. *Materials Science and Engineering: A*, 580:1–12, 2013.
- [33] Stefan T. Radev, Ulf K. Mertens, Andreas Voss, Lynton Ardizzone, and Ulrich Köthe. BayesFlow: Learning complex stochastic models with invertible neural networks. *IEEE transactions on neural networks and learning systems*, 33(4):1452–1466, 2020.
- [34] Danilo Rezende and Shakir Mohamed. Variational inference with normalizing flows. In *International Conference on Machine Learning*, pages 1530–1538. PMLR, 2015.
- [35] Vasile-Adrian Surdu and Romuald Gy\Horgy. X-ray diffraction data analysis by machine learning methods—a review. *Applied Sciences*, 13(17):9992, 2023.
- [36] Shuji Taira. X-ray-diffraction approach for studies on fatigue and creep: The philosophy of the application of the X-ray-diffraction approach for studies on the mechanical behavior of materials is discussed by taking cases of studies of fatigue and creep as examples. *Experimental Mechanics*, 13(11):449–463, November 1973.
- [37] K Tanaka. Recent x-ray diffraction studies of metal fatigue in Japan. *Journal of Strain Analysis*, 10(1):32–41, January 1975.
- [38] P. Thompson, D. E. Cox, and J. B. Hastings. Rietveld refinement of Debye–Scherrer synchrotron X-ray data from Al₂O₃. *Applied Crystallography*, 20(2):79–83, 1987.

- [39] Satoru Tokuda, Kenji Nagata, and Masato Okada. Simultaneous Estimation of Noise Variance and Number of Peaks in Bayesian Spectral Deconvolution. *Journal of the Physical Society of Japan*, 86(2):024001, February 2017.
- [40] H. Toraya. Array-type universal profile function for powder pattern fitting. *Applied Crystallography*, 23(6):485–491, 1990.
- [41] Christina Winkler, Daniel Worrall, Emiel Hoogeboom, and Max Welling. Learning Likelihoods with Conditional Normalizing Flows, November 2019.
- [42] Sam Y. Zamrik and Robert N. Pangborn. Fatigue damage assessment using X-ray diffraction and life prediction methodology. *Nuclear engineering and design*, 116(3):407–413, 1989.