SparseEIS-Bench: Benchmarking Three- and Five-Point EIS Reconstruction with Measurement-Conditioned Operator-Homotopy Flow Matching

Anonymous authors
Paper under double-blind review

Abstract

Electrochemical impedance spectroscopy (EIS) provides rich frequency-domain information for battery management, but the time required for multifrequency scans limits its online use. We study ultra-sparse EIS reconstruction: recovering a complete impedance spectrum from only three to five complex-valued measurements. To support standardized research on this problem, we introduce SparseEIS-Bench, a large-scale benchmark comprising 25,090 measured spectra from 406 batteries across diverse chemistries, capacities, states of charge, states of health, and temperatures. SparseEIS-Bench provides a unified spectral representation, strictly disjoint Battery ID splits, and standardized three- and five-point reconstruction protocols for jointly evaluating reconstruction accuracy, physical consistency, probabilistic quality, and downstream utility. As a strong generative reference baseline, we further propose Measurement-Conditioned Operator-Homotopy Flow Matching (MOH-FM), which combines deterministic prediction with conditional residual generation and restores measured impedance values at the registered input positions. MOH-FM achieves aggregate complex root mean squared errors of $0.0465 \pm 0.0032\,\Omega$ and $0.0359 \pm 0.0025\,\Omega$ in the three- and five-point settings, respectively, yielding the lowest mean among the eight evaluated methods in both settings. Compared with linear residual flow matching, it reduces the mean aggregate error by 16.1% and 14.9%, respectively. The total scan-plus-inference time of both three- and five-point schemes is approximately 95% shorter than full-frequency scan time. Together, SparseEIS-Bench and MOH-FM establish a benchmark and modeling framework for rapid, measurement-efficient EIS reconstruction for practical battery management.

1 Introduction

1.1 Background and Motivation

Battery management requires accurate estimates of state of charge (SOC), state of health (SOH), and temperature, along with degradation and fault detection (Su & Chen, 2023; Jung et al., 2024; Dong et al., 2024; Birkl et al., 2017; Lu et al., 2023; Liu et al., 2026). Conventional voltage, current, and temperature signals and their equivalent models (Hu et al., 2012) offer limited insight into internal electrochemistry and are jointly affected by SOC, aging, and temperature. EIS provides complementary information on ohmic responses, interfacial charge transfer, and diffusion (Barsoukov & Macdonald, 2018; Lazanas & Prodromidis, 2023; Dong et al., 2024). However, scans are slow, especially at low frequencies, and require approximately stationary conditions. This motivates learning spectral priors across batteries and conditions to recover useful EIS representations from few observations.

We reconstruct the remaining response within each spectrum's measured frequency range from three or five complex impedance observations selected directly from raw scans. This underdetermined problem requires spectral priors and input consistency. To separate model capability from protocol differences, we fix observation budgets, disjoint Battery ID splits, observation locations, and

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

valid evaluation positions, assessing reconstruction quality together with physical and probabilistic reliability.

1.2 This Work and Contributions

We introduce SparseEIS-Bench, evaluate interpolation, structured deterministic methods, direct neural reconstruction, and residual generation, and propose MOH-FM as a generative reference baseline. Our contributions are:

- **A large-scale, independently measured data resource.** We construct SparseEIS-Bench, comprising 25,090 EIS spectra measured by the authors from 406 batteries and 1,218,599 raw frequency-point records. The data cover diverse cathode and anode chemistries, capacities, SOC, SOH, and temperature conditions, while retaining point-level measurement records, Battery IDs, and operating-condition labels.
- **Standardized tasks and evaluation protocols.** We use all 25,090 spectra to construct a fixed task collection with a unified dense target representation, validity masks, and three- and five-point input protocols that directly access raw measurement records. The benchmark strictly separates training, validation, and test batteries by Battery ID and fixes the data splits, per-spectrum observation masks, valid evaluation positions, and metrics. Cross-method comparisons thus evaluate batteries unseen during training within the same independently measured data domain, covering reconstruction accuracy, physical consistency, probabilistic quality, and downstream state estimation.
- **Systematic baselines and findings.** MOH-FM combines residual generation with observation projection. Its mean aggregate cRMSE is below the best deterministic interpolation baseline at both budgets, but lower error does not guarantee physical consistency.

2 Related Work

Battery learning and impedance reconstruction. BatteryML supports battery-degradation learning (Zhang et al., 2024); SOH models incorporate domain knowledge and physics (Lanubile et al., 2024; Wang et al., 2024). Public EIS data span SOC and temperature conditions (Buchicchio et al., 2022; Rashid et al., 2023), supporting degradation identification, health prediction, transfer, and hybrid estimation (Zhang et al., 2020; Li et al., 2023; Wu et al., 2024). Arrhenius-anchored thermometry combines ECM, DRT, and physics-constrained learning for cross-capacity estimation (Wang et al., 2027). DRT represents relaxation timescales (Schichlein et al., 2002; Ciucci & Chen, 2015), while equivalent circuits and KK tests support impedance analysis (Santoni et al., 2024; Boukamp, 1995; Schönleber et al., 2014). Duplex learning predicts condition-dependent spectra (Temiz et al., 2023), and partial spectra support SOH estimation (Xia et al., 2024). SparseEIS-Bench jointly evaluates error, physical consistency, and probabilistic quality of unobserved impedance within the valid band.

Generative paths and architectures. Diffusion and score models learn by denoising (Ho et al., 2020; Song et al., 2021b), with DDIM and EDM refining sampling and design (Song et al., 2021a; Karras et al., 2022). Neural ODEs and continuous normalizing flows establish continuous-time modeling (Chen et al., 2018; Grathwohl et al., 2019). FM, rectified flow, and stochastic interpolants learn transport by velocity regression (Lipman et al., 2023; Liu et al., 2023; Albergo & Vanden-Eijnden, 2023; Tong et al., 2024); geometric FM incorporates state-space structure (Chen & Lipman, 2024). Transformers, DiT, and SiT inform generative backbones and interpolation (Vaswani et al., 2017; Peebles & Xie, 2023; Ma et al., 2024).

Conditioning and structural priors. Diffusion inverse solvers use posterior guidance or measurement-subspace constraints (Chung et al., 2023; Song et al., 2023; Kawar et al., 2022). Masked autoencoders, PatchTST, and TimesNet study representations and sequence structure (He et al., 2022; Nie et al., 2023; Wu et al., 2023); FNO, DeepONet, and physics-informed learning address function mappings and structural priors (Li et al., 2021; Lu et al., 2021; Karniadakis et al., 2021). MOH-FM uses explicit fixed-grid operators without claiming cross-resolution generalization.

3 Benchmark Construction and Task Protocols

Measured data and annotations. SparseEIS-Bench comprises 406 batteries, 25,090 measured spectra, and 1,218,599 raw frequency-point records, covering LCO, LFP, NCA, and NCM chemistries across different capacities, SOC, SOH, and temperatures. Each spectrum corresponds to a complete frequency scan under a specific operating condition. Three- and five-point inputs are selected from its raw frequency-point records according to the protocol below. Table 1 summarizes label coverage in the benchmark.

Table 1: **Annotation coverage.** Label coverage across the 25,090 measured spectra.

	Spectra	Coverage	Range
Capacity	$\approx$ 25,033	99.77%	0.04–628 Ah
SOC	25,090	100.00%	0–100%
SOH	$\approx$ 24,843	99.02%	17.30–100%
Temperature	$\approx$ 24,590	98.01%	-40–83.8 °C

Figure 1 summarizes the benchmark's condition coverage and label structure in terms of electrode composition, capacity, temperature, SOC, and SOH. Figure 2 further illustrates measured impedance responses of representative batteries under different operating conditions. Together, these figures show substantial variation in spectral shape and scale across batteries and states.

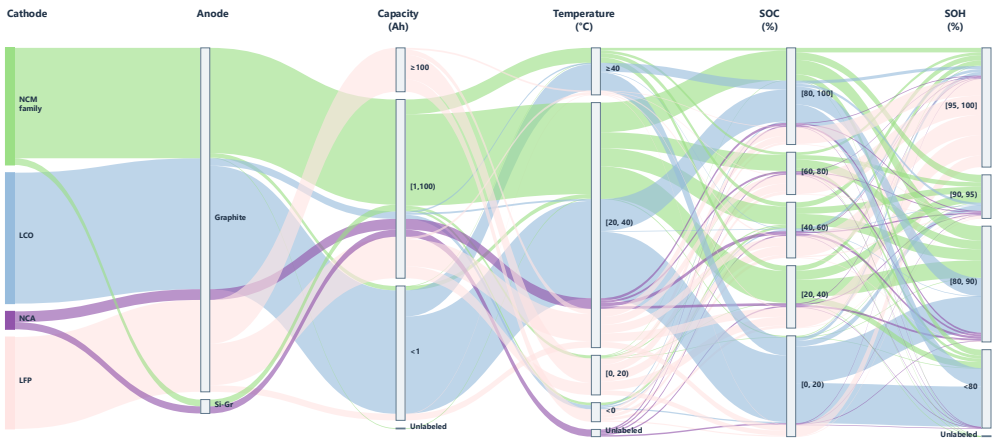

Figure 1: **Benchmark composition.** Electrode and state distributions in the fixed task collection. Band widths indicate the proportion of spectra, and connections between adjacent categories indicate label co-occurrence.

Measured responses. Figure 2 shows measured impedance responses under different battery operating conditions. In a Nyquist representation, the high-frequency real-axis intercept is typically associated with ohmic impedance from the electrolyte, current collectors, and contacts; the mid-frequency arc is primarily influenced by interfacial charge transfer and double-layer processes; and the low-frequency extension more strongly reflects mass transport and diffusion (Barsoukov & Macdonald, 2018; Lazanas & Prodromidis, 2023). Panels (a)–(c) show spectral sequences for the same cell as SOH, SOC, and temperature vary, respectively. Changes in curve position, arc width, and the low-frequency tail reflect condition-dependent variations in internal resistance, interfacial kinetics, and diffusion timescales. Panel (d) compares different cell chemistries at matched states, further illustrating differences in impedance scale and spectral shape associated with capacity and electrode composition. Similar condition dependence has been systematically observed in lithium-ion battery EIS studies (Waag et al., 2013; Dong et al., 2024).

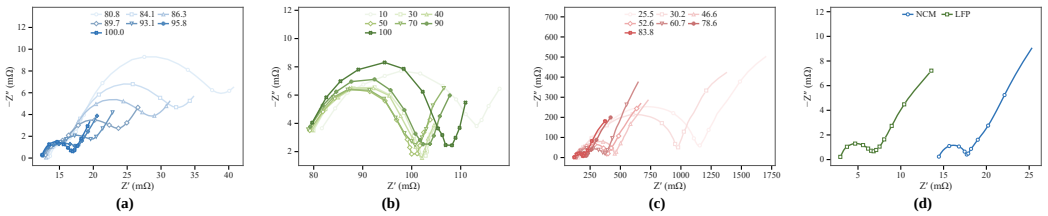

Figure 2: **Measured impedance responses across battery conditions.** (a) SOH; (b) SOC; (c) temperature; (d) cathode chemistry. Impedance is in mΩ. Panels (a)–(c) compare state sequences for the same cell, whereas (d) varies both capacity and electrode composition.

3.1 Unified Representation and Reconstruction Protocol

Measured spectra differ in their frequency ranges and sampling locations. To define a unified learning target, SparseEIS-Bench constructs a logarithmically spaced frequency grid with $L = 512$ positions from 0.01 Hz to 100 kHz and applies shape-preserving piecewise cubic Hermite interpolation (PCHIP) to the real and imaginary impedance components separately along $\log_{10} f$. PCHIP constructs reference values only within each spectrum's original measured frequency range, without extrapolation beyond it. Sparse inputs are selected from the raw frequency-point records retained for each spectrum. For spectrum n and observation budget $B \in \{3, 5\}$, we first place B logarithmically equidistant target locations between the lower and upper bounds of its original measured frequency range, then select the actual measurement nearest to each target in log-frequency distance. The complete selection formulas are given in Appendix Equations (15)–(16). The model receives the recorded raw frequencies and complex impedance values. Selected points are registered to the nearest valid positions on the unified grid to construct tensors and observation masks; if the nearest global grid position falls outside the spectrum's validity mask, the nearest valid position is used instead.

We split the fixed task collection strictly by Battery ID. All EIS spectra sharing a Battery ID belong to exactly one subset, yielding 18,285 training, 3,415 validation, and 3,390 test spectra, with pairwise disjoint Battery IDs across the three subsets. All methods share this split, the valid positions, and the three- and five-point observation masks predetermined for each spectrum. Complete fields and filtering rules are provided in Appendix A.

Let $z \in \mathbb{R}^{L\times 2}$ denote the two-channel reference target constructed by PCHIP, and let $v, m \in \{0, 1\}^L$ denote the validity mask and the mask obtained by registering the raw observations, respectively, with $m \leq v$. The condition y must include the selected raw measurement frequencies, complex impedance values, and masks, and may additionally include available operating-condition labels and their missingness indicators. For valid unobserved positions $U = \{i : v_i = 1, m_i = 0\}$, the primary reconstruction task is to learn the conditional distribution $p(z_U \mid y)$. The benchmark supports both sparse-impedance-only configurations and configurations with auxiliary operating conditions. All neural baselines in the main table that accept operating-condition information use the same fields and missingness indicators, while deterministic baselines use the same sparse impedance inputs according to their definitions.

Benchmark and baseline scope. In this paper, SparseEIS-Bench denotes the benchmark evaluation framework comprising the measured data resource, task definitions, disjoint Battery ID splits, predetermined per-spectrum observation masks, target representation, and evaluation metrics. Baselines are the methods implemented and compared within this unified framework. Interpolation, Eigenspectra ridge (Mänken et al., 2025), deterministic neural prediction, and residual flows form the baseline suite, with MOH-FM evaluated as a strong generative reference baseline.

4 A Measurement-Conditioned Reference Baseline

MOH-FM predicts a baseline spectrum with a deterministic branch $\mu_\theta(y)$, generates the residual $r = z - \mu_\theta(y)$, and restores input values. Figures 3 and 4 show the workflow and architecture. Its operator path uses the measurement-coverage quantities below.

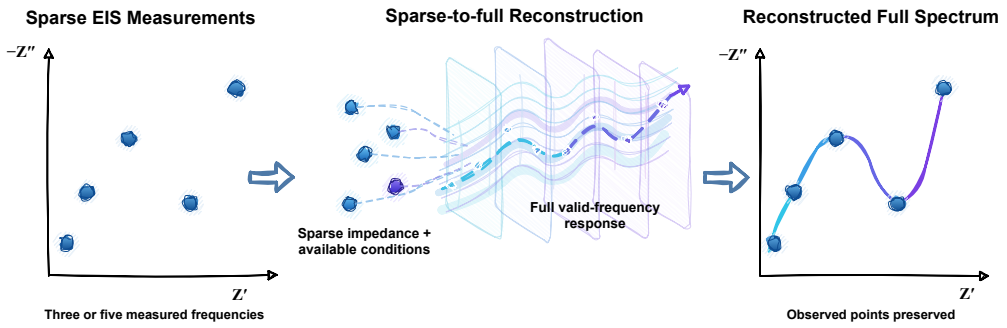

Figure 3: **Benchmark reconstruction workflow.** Three or five measured complex impedance values are selected from raw scan records according to a protocol fixed in advance and used to reconstruct the impedance response within the valid measured frequency range.

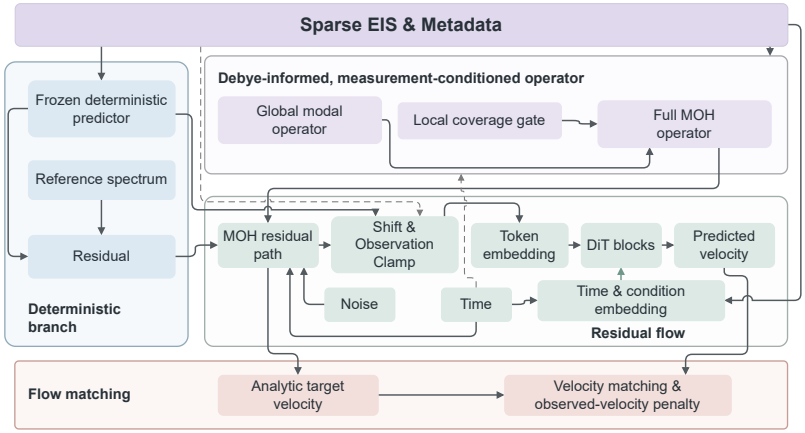

Figure 4: **MOH-FM reference baseline.** Deterministic prediction and conditional residual generation use a path determined by observation count, frequency span, and locations. Projection Π_y restores input impedance throughout integration, yielding $\hat{z} = \Pi_y(\mu_\theta(y) + \hat{r})$. Dashed lines indicate time and condition modulation.

Coverage quantities in the MOH-FM path. Let $\Omega, \mathcal{V}$ be observed and valid position sets, with sizes n_o, n_v and log-frequency spans Δ_o, Δ_v. The Debye kernel $h(u) = [2\cosh(u)]^{-1}$ gives:

$$I(y) = \mathrm{clip}_{[10^{-3},1]} \sqrt{\frac{n_o}{n_v}\left(\frac{1}{4} + \frac{3}{4}\frac{\Delta_o}{\Delta_v}\right)}, \tag{1}$$

$$C_i(y) = \mathrm{clip}_{[0,1]} \frac{\sum_{j\in\Omega} h^2(\ln f_i - \ln f_j)}{\sum_{j\in\mathcal{V}} h^2(\ln f_i - \ln f_j)}. \tag{2}$$

$I(y)$ summarizes count and span; $C_i(y)$ captures placement. Both use only frequencies and masks, excluding unobserved impedance and test errors.

Residual path. For generative time $t \in [0, 1]$, define

$$r_t = (1-t)\epsilon + tQ_y(t)r, \tag{3}$$

$$\epsilon \sim \mathcal{N}(0, I). \tag{4}$$

Linear residual FM sets $Q_y(t) = I$; MOH-FM uses a measurement-conditioned operator. Let $\mathcal{M}$ be the channel-wise orthonormal rFFT along log-frequency and D_k the normalized Debye-kernel

transfer magnitude:

$$s_k(y) = \text{clip}_{[0,1]}(I(y)D_k^2), \tag{5}$$

$$\beta_k(y) = 1 + \lambda[1 - s_k(y)]^p, \tag{6}$$

$$q_k(t, y) = t^{\beta_k(y)}, \tag{7}$$

$$Q_M(t, y) = \mathcal{M}^{-1} \,\text{diag}(q_k(t, y))\mathcal{M}. \tag{8}$$

The global residual coefficient is $tq_k = t^{1+\beta_k}$. Local corrections use $C_i(y)$:

$$A_y(t) = \text{diag}(\lambda_{\text{loc}} C_i(y)^\gamma t^{p_{\text{loc}}}), \tag{9}$$

$$Q_y(t) = Q_M(t, y) + A_y(t)[I - Q_M(t, y)]. \tag{10}$$

We use $(\lambda, p, \lambda_{\text{loc}}, p_{\text{loc}}, \gamma) = (1, 2, 0.1, 1, 1)$; D_k and implementation details appear in Appendix B.1. Coverage parameterizes the path; observation projection ensures input consistency.

Training and observation consistency. During residual-flow training, the deterministic branch is frozen, and the analytic target is $\dot{r}_t = -\epsilon + Q_y(t)r + t\dot{Q}_y(t)r$. Let Π_y restore impedance values on the observed set Ω. The velocity network receives $x_t = \Pi_y(\mu_\theta(y) + r_t)$ and is trained with the objective

$$\mathcal{L} = \mathbb{E}\left[\|v_\phi(x_t, t, y) - \dot{r}_t\|_U^2 + 0.01\|v_\phi(x_t, t, y)\|_\Omega^2\right], \tag{11}$$

where the norms are averaged over the corresponding positions and both channels. Inference uses 16 Heun integration steps with 32 velocity-field evaluations, applying projection at integration substeps and to the final output to obtain $\hat{z} = \Pi_y(\mu_\theta(y) + \hat{r})$. The two branches contain 23.59M parameters in total; the DiT and SwiGLU configurations are given in Appendix B (Peebles & Xie, 2023; Shazeer, 2020).

Analytic scope. The path connects noise and target-residual endpoints, with the following conditional Gaussian property.

Proposition 1 (Early-time conditional velocity variance)**.** *Suppose that the real modal residual $r_k \mid y \sim \mathcal{N}(\mu_k, \sigma_k^2)$ and $\epsilon_k \sim \mathcal{N}(0, 1)$ are conditionally independent. The unprojected path $\xi_{t,k} = b(t)\epsilon_k + a_k(t)r_k$ satisfies*

$$\text{Var}(\dot{\xi}_{t,k} \mid \xi_{t,k}, y) = \frac{\sigma_k^2(a_k\dot{b} - b\dot{a}_k)^2}{b^2 + a_k^2\sigma_k^2}. \tag{12}$$

When $b = 1 - t$, the early-time variance of the linear path $a_k = t$ approaches σ_k^2, whereas that of the power path $a_k = t^{1+\beta_k}$ with $\beta_k > 0$ approaches zero at rate $O(t^{2\beta_k})$.

This variance quantifies regression-target uncertainty given the path state; the proof and matrix extension appear in Appendix C.

5 Benchmark Results and Findings

Under the shared protocol, we compare reconstruction errors across eight baselines and assess reliability and utility through physical diagnostics, probabilistic scores, and downstream state estimation.

5.1 Reconstruction Quality under Fixed Observation Budgets

The primary metric, cRMSE, aggregates real and imaginary errors over all valid unobserved test positions, rather than averaging per-spectrum errors:

$$\text{cRMSE}_{\text{miss}} = \sqrt{\frac{\sum_n \sum_{i \in U_n} \left[(\hat{Z}'_{ni} - Z'_{ni})^2 + (\hat{Z}''_{ni} - Z''_{ni})^2\right]}{2\sum_n |U_n|}}. \tag{13}$$

Neural models share per-spectrum normalization: inputs and targets use a scale determined solely by raw sparse inputs. Table 2 compares eight configurations on the full held-out test set (approximately 13.5% of all spectra), using Equation (13) after restoring physical units.

Finding 1: Controlled comparisons show reconstruction differences. MOH-FM has the lowest mean aggregate cRMSE at both budgets, with three-seed means 16.1% and 14.9% below linear residual FM. Table 2 also includes predefined global-only MOH and scalar-power configurations. At three points, MOH-FM achieves $0.0465\,\Omega$, a 42.9% lower mean than the best deterministic interpolator, Log-linear ($0.0815\,\Omega$).

Table 2: **Held-out-battery reconstruction.** Aggregate cRMSE (Ω) on valid unobserved positions; lower is better. Neural methods report mean $\pm$ SD over three seeds; deterministic results are fixed. Bold marks the lowest mean. Global-only MOH differs from full MOH-FM only by omitting the local position correction.

Method	3 observations	5 observations
Deterministic baselines		
Log-linear	0.0815	0.0397
Log-PCHIP	0.0901	0.0372
Eigenspectra ridge	0.1225	0.0502
Neural reconstruction baselines		
Direct DiT	0.0971 ± 0.0069	0.0891 ± 0.0063
Linear residual FM	0.0554 ± 0.0038	0.0422 ± 0.0029
Matched scalar power	0.0897 ± 0.0062	0.0810 ± 0.0056
Global-only MOH	0.0496 ± 0.0034	0.0395 ± 0.0027
Full MOH-FM	$\mathbf{0.0465 \pm 0.0032}$	$\mathbf{0.0359 \pm 0.0025}$

Figure 5 shows two different representative spectra, using one reference checkpoint per budget. Three-point reconstruction recovers the broad arc and tail; the five-point example follows its reference arc more closely. Tail deviations remain in both. Registered positions restore measured inputs; aggregate cross-budget errors appear in Table 2. With one measurement cycle per frequency and 0.180 s of inference added to each three- or five-point scan, both examples reduce total scan-plus-inference time by approximately 95% relative to full-frequency scan time alone. These are case-specific estimates; frequencies and timing references appear in Appendix A.7.

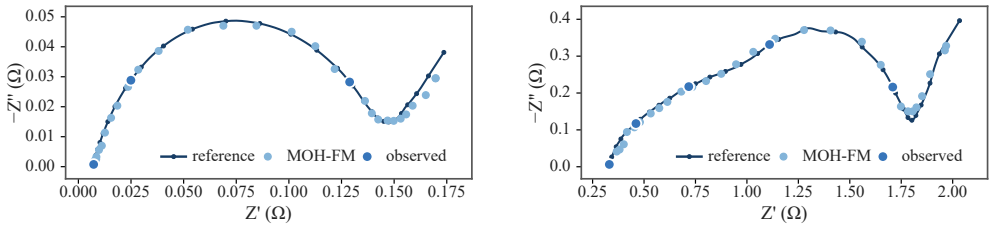

Figure 5: **Nyquist reconstruction.** (a) A three-observation example; (b) a five-observation example. Dark blue circles denote observed points.

5.2 Reconstruction Reliability

Finding 2: Accuracy does not ensure physical consistency or calibration. Table 3 evaluates raw outputs from one checkpoint per budget using Kramers–Kronig (KK) consistency, negative-real-part incidence, equivalent circuit model (ECM) parameters, and prediction intervals (Boukamp, 1995; Schönleber et al., 2014; Santoni et al., 2024; Gneiting & Raftery, 2007). At a relative Voigt projection residual threshold of 0.05, KK pass rates are 100%, 52.71%, and 90.38% for reference, three-point, and five-point spectra. Despite the lowest mean cRMSE, nearly half the spectra from the three-point checkpoint fail this criterion, exposing an accuracy–physical-consistency gap. Error alone therefore cannot establish a reliable replacement for full measurements. Five-point predictions also fall short of the reference pass rate, and finite-band KK testing does not fully verify passivity.

Table 3: **Physical consistency and probabilistic quality.** KK and 90% intervals use the full test set; ECM uses independent $\approx$ 17.7% subsets per budget, allowing descriptive comparisons only. MdAPE: median absolute percentage error; NLL: negative log-likelihood; CRPS: continuous ranked probability score. Negative-real-part denominators appear in the appendix.

Metric	3 pts	5 pts
Median KK residual	0.0497	0.0465
KK pass rate	52.71%	90.38%
Negative Re Z incidence (spectrum-level)	0.531%	0.501%
Valid points with negative Re Z (point-level)	0.00381%	0.00497%
ECM eligible-fit rate	99.17%	99.33%
R_0 / R_{ct} MdAPE (%)	3.51 / 5.18	1.83 / 2.48
τ / A_W MdAPE (%)	20.51 / 3.31	10.90 / 2.63
Empirical coverage (nominal 90%)	93.53%	97.81%
Normalized interval width	0.1532	0.1383
Normalized Gaussian NLL	-1.7371	-1.9882
Normalized CRPS	0.0236	0.0169

The three- and five-point R_0/R_{ct} MdAPEs are $3.51/5.18\%$ and $1.83/2.48\%$. The five-point checkpoint has narrower intervals and lower CRPS, but its 97.81% coverage for nominal 90% intervals deviates farther from nominal than the three-point result (93.53%). Improvements in some physical metrics and probabilistic scores thus coexist with conservative calibration.

5.3 Downstream State Estimation from Ultra-Sparse EIS Reconstruction

We compare state estimation from three-point reconstruction, five-point reconstruction, and reference EIS using identical Battery ID splits, Transformer architectures, and optimization. Each estimator starts from a checkpoint pretrained by masking training-battery data without state labels, then is fine-tuned separately for SOH, SOC, and temperature. Estimators also receive available electrode chemistries and rated capacity. SOH is current measured capacity divided by rated capacity; measured capacity supplies only low-weight auxiliary supervision, not an input, which qualifies interpretation of SOH results. The target state label is excluded from both reconstruction conditions and estimator inputs.

For input s and state q, normalize aggregate error $E^{(s)}_{q,d}$, $d \in \{\text{MAE}, \text{RMSE}\}$, by reference-EIS error:

$$R^{(s)}_{q,d} = 100 \times \frac{E^{(s)}_{q,d}}{E^{(\text{Reference EIS})}_{q,d}}. \tag{14}$$

The relative error of the reference EIS representation is therefore 100%, and lower values indicate smaller errors.

Downstream state-estimation results. Most configurations approach reference-EIS error; the largest gap is three-point SOC (Figure 6). Sparse reconstruction thus retains task-dependent utility.

Scope of the findings. Observation projection does not ensure physical consistency beyond measured points. Three-seed mean errors and single-checkpoint diagnostics have different aggregation levels; single-checkpoint and cross-budget comparisons remain descriptive. These metrics provide complementary evidence.

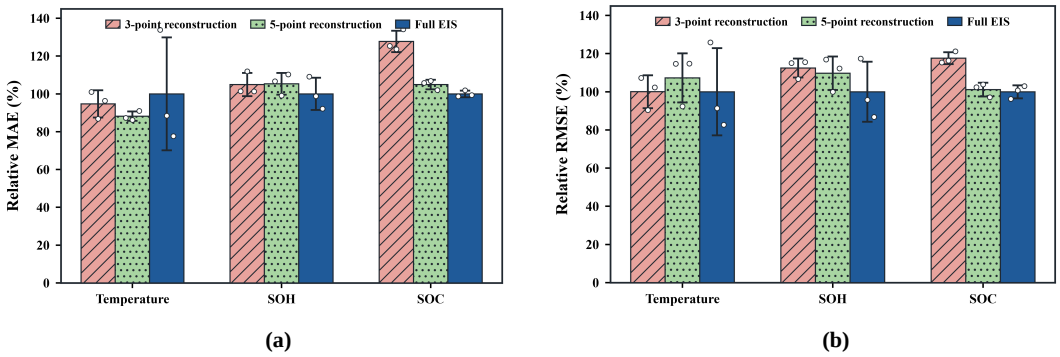

Figure 6: **Downstream state estimation.** (a) Relative MAE; (b) relative RMSE; lower is better. Temperature, SOH, and SOC errors from three- and five-point reconstructions are normalized by reference-EIS errors. Full EIS denotes the reference within the valid measured band.

6 Conclusion

This work establishes SparseEIS-Bench, a benchmark based on independently measured data, to provide a unified evaluation foundation for EIS reconstruction under limited observation budgets. The benchmark comprises 25,090 measured spectra from 406 batteries and defines three- and five-point input protocols that read directly from raw measurement records. Battery-disjoint data splits and a unified target representation enable comparable evaluation of different methods on unseen batteries. Evaluation covers reconstruction accuracy, physical consistency, probabilistic quality, and downstream state estimation, extending sparse-reconstruction research beyond error comparisons to a broader assessment of predictive reliability and practical utility.

The unified evaluation reveals interconnected benchmark findings: MOH-FM achieves lower mean aggregate cRMSE than the corresponding best deterministic interpolation baseline in both three- and five-point tasks. Downstream state-estimation performance using three- and five-point reconstructions approaches that of full-frequency EIS, while the total scan-plus-inference time of both schemes is approximately 95% shorter than full-frequency scan time. The baseline suite provides a common reference for current method capabilities, with MOH-FM serving as a strong generative reference baseline and a reproducible starting point for future comparisons. The findings are limited to held-out batteries within the same measured-data domain and the current task configurations. Future work can investigate ablations of impedance and operating-condition inputs, physics-constrained generation, cross-domain generalization, and validation with actual sparse acquisition.

AI USE STATEMENT

Generative AI assisted drafting, editing, literature review, code development, and data processing. It did not independently produce experimental results. The authors reviewed and verified all code, processing, and results and take responsibility for the study design, conclusions, citation accuracy, and manuscript.

REPRODUCIBILITY STATEMENT

Appendices A–D detail the benchmark, preprocessing, fixed splits, proofs, and metrics; Appendix A.6 specifies submission-stage data release and downstream protocols.

REFERENCES

Michael S. Albergo and Eric Vanden-Eijnden. Building normalizing flows with stochastic interpolants. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=li7qeBbCR1t.

Evgenij Barsoukov and J. Ross Macdonald (eds.). *Impedance Spectroscopy: Theory, Experiment, and Applications*. Wiley, 3 edition, 2018.

Christoph R. Birkl, Matthew R. Roberts, Euan McTurk, Peter G. Bruce, and David A. Howey. Degradation diagnostics for lithium ion cells. *Journal of Power Sources*, 341:373–386, 2017. doi: 10.1016/j.jpowsour.2016.12.011.

Bernard A. Boukamp. A linear Kronig–Kramers transform test for immittance data validation. *Journal of The Electrochemical Society*, 142(6):1885–1894, 1995. doi: 10.1149/1.2044210.

Emanuele Buchicchio, Alessio De Angelis, Francesco Santoni, Paolo Carbone, Francesco Bianconi, and Fabrizio Smeraldi. Dataset on broadband electrochemical impedance spectroscopy of lithium-ion batteries for different values of the state-of-charge. *Data in Brief*, 45:108589, 2022. doi: 10.1016/j.dib.2022.108589.

Ricky T. Q. Chen and Yaron Lipman. Flow matching on general geometries. In *International Conference on Learning Representations*, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/d1f9936d3be6997ffffab692977eebe6-Abstract-Conference.html.

Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David K. Duvenaud. Neural ordinary differential equations. In *Advances in Neural Information Processing Systems*, volume 31, 2018. URL https://papers.nips.cc/paper/2018/hash/69386f6bb1dfed68692a24c8686939b9-Abstract.html.

Hyungjin Chung, Jeongsol Kim, Michael T. McCann, Marc L. Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=OnD9zGAGT0k.

Francesco Ciucci and Chi Chen. Analysis of electrochemical impedance spectroscopy data using the distribution of relaxation times: A Bayesian and hierarchical Bayesian approach. *Electrochimica Acta*, 167:439–454, 2015. doi: 10.1016/j.electacta.2015.03.123.

Ming Dong, Xiaofeng Li, Zhang Yang, Yi Chang, Wangzeyu Liu, Yang Luo, Wanjun Lei, Ming Ren, and Chongxing Zhang. State of health (SOH) assessment for LIBs based on characteristic electrochemical impedance. *Journal of Power Sources*, 603:234386, 2024. doi: 10.1016/j.jpowsour.2024.234386.

Tilmann Gneiting and Adrian E. Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007. doi: 10.1198/016214506000001437.

Will Grathwohl, Ricky T. Q. Chen, Jesse Bettencourt, Ilya Sutskever, and David Duvenaud. FFJORD: Free-form continuous dynamics for scalable reversible generative models. In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=rJxgknCcK7.

Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16000–16009, 2022. URL https://openaccess.thecvf.com/content/CVPR2022/html/He_Masked_Autoencoders_Are_Scalable_Vision_Learners_CVPR_2022_paper.html.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, volume 33, pp. 6840–6851, 2020. URL https://papers.nips.cc/paper/2020/hash/4c5bcfec8584af0d967f1ab10179ca4b-Abstract.html.

Xiaosong Hu, Shengbo Li, and Huei Peng. A comparative study of equivalent circuit models for Li-ion batteries. *Journal of Power Sources*, 198:359–367, 2012. doi: 10.1016/j.jpowsour.2011.10.013.

Min Jae Jung, Sang-Gug Lee, and Kyung-Sik Choi. A new diagnostic indicator for lithium-ion batteries via electrochemical impedance spectroscopy: Harnessing the highest frequency peak in distribution of relaxation times. *Journal of Power Sources*, 611:234743, 2024. doi: 10.1016/j.jpowsour.2024.234743.

George Em Karniadakis, Ioannis G. Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440, 2021. doi: 10.1038/s42254-021-00314-5.

Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. In *Advances in Neural Information Processing Systems*, volume 35, pp. 26565–26577, 2022. URL https://papers.nips.cc/paper_files/paper/2022/hash/a98846e9d9cc01cfb87eb694d946ce6b-Abstract-Conference.html.

Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. In *Advances in Neural Information Processing Systems*, volume 35, pp. 23593–23606, 2022. URL https://papers.nips.cc/paper_files/paper/2022/hash/95504595b6169131b6ed6cd72eb05616-Abstract-Conference.html.

Andrea Lanubile, Pietro Bosoni, Gabriele Pozzato, Anirudh Allam, Matteo Acquarone, and Simona Onori. Domain knowledge-guided machine learning framework for state of health estimation in lithium-ion batteries. *Communications Engineering*, 3:168, 2024. doi: 10.1038/s44172-024-00304-2.

Alexandros Ch. Lazanas and Mamas I. Prodromidis. Electrochemical impedance spectroscopy—a tutorial. *ACS Measurement Science Au*, 3(3):162–193, 2023. doi: 10.1021/acsmeasuresciau.2c00070.

Yan Li, Mehdi Maleki, and Shadi Banitaan. State of health estimation of lithium-ion batteries using EIS measurement and transfer learning. *Journal of Energy Storage*, 73:109185, 2023. doi: 10.1016/j.est.2023.109185.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=c8P9NQVtmnO.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. In *International Conference on Learning Representations*, 2023. URL https://arxiv.org/abs/2210.02747.

Qingdi Liu, Yan Fu, Lishuo Liu, Yanke Lin, Jin Xin, Jianfeng Zhang, Cheng Liu, Lujia Pan, Dongxu Guo, Yuejiu Zheng, and Qiang Li. Battery fault: A comprehensive dataset and benchmark for battery fault diagnosis. In *International Conference on Learning Representations*, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/0d781fa5f639bf2caf728a68e9678362-Abstract-Conference.html.

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=XVjTT1nw5z.

Jiahuan Lu, Rui Xiong, Jinpeng Tian, Chenxu Wang, et al. Deep learning to estimate lithium-ion battery state of health without additional degradation experiments. *Nature Communications*, 14: 2760, 2023. doi: 10.1038/s41467-023-38458-w.

Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George Em Karniadakis. Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nature Machine Intelligence*, 3(3):218–229, 2021. doi: 10.1038/s42256-021-00302-5.

Nanye Ma, Mark Goldstein, Michael S. Albergo, Nicholas M. Boffi, Eric Vanden-Eijnden, and Saining Xie. SiT: Exploring flow and diffusion-based generative models with scalable interpolant transformers. In *European Conference on Computer Vision*, 2024. URL https://www.ecva.net/papers/eccv_2024/papers_ECCV/html/9828_ECCV_2024_paper.php.

Christian Mänken, Dominik Schäfer, and Rüdiger-A. Eichel. The exploitation of eigenspectra in electrochemical impedance spectroscopy: Reconstruction of spectra from sparse measurements. *Journal of Power Sources*, 628:235808, 2025. doi: 10.1016/j.jpowsour.2024.235808.

Yuqi Nie, Nam H. Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. A time series is worth 64 words: Long-term forecasting with transformers. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=Jbdc0vTOcol.

William Peebles and Saining Xie. Scalable diffusion models with transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 4195–4205, 2023. URL https://openaccess.thecvf.com/content/ICCV2023/html/Peebles_Scalable_Diffusion_Models_with_Transformers_ICCV_2023_paper.html.

Muhammad Rashid, Mona Faraji-Niri, Jonathan E. H. Sansom, Muhammad Sheikh, Dhammika Widanage, and James Marco. Dataset for rapid state of health estimation of lithium batteries using EIS and machine learning: Training and validation. *Data in Brief*, 48:109157, 2023. doi: 10.1016/j.dib.2023.109157.

Francesco Santoni, Alessio De Angelis, Antonio Moschitta, Paolo Carbone, Matteo Galeotti, Lucio Cinà, Corrado Giammanco, and Aldo Di Carlo. A guide to equivalent circuit fitting for impedance analysis and battery state estimation. *Journal of Energy Storage*, 82:110389, 2024. doi: 10.1016/j.est.2023.110389.

H. Schichlein, A. C. Müller, M. Voigts, A. Krügel, and E. Ivers-Tiffée. Deconvolution of electrochemical impedance spectra for the identification of electrode reaction mechanisms in solid oxide fuel cells. *Journal of Applied Electrochemistry*, 32(8):875–882, 2002. doi: 10.1023/A:1020599525160.

Michael Schönleber, Dino Klotz, and Ellen Ivers-Tiffée. A method for improving the robustness of linear Kramers–Kronig validity tests. *Electrochimica Acta*, 131:20–27, 2014. doi: 10.1016/j.electacta.2014.01.034.

Noam Shazeer. GLU variants improve transformer. *arXiv preprint arXiv:2002.05202*, 2020. URL https://arxiv.org/abs/2002.05202.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *International Conference on Learning Representations*, 2021a. URL https://openreview.net/forum?id=St1giarCHLP.

Jiaming Song, Arash Vahdat, Morteza Mardani, and Jan Kautz. Pseudoinverse-guided diffusion models for inverse problems. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=9_gsMA8MRKQ.

Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021b. URL https://openreview.net/forum?id=PxTIG12RRHS.

Tyng-Fwu Su and Kuo-Ching Chen. Rapid monitor of states of lithium-ion batteries through non-quasi-static electrochemical impedance spectroscopy and terminal voltage. *Journal of Power Sources*, 586:233641, 2023. doi: 10.1016/j.jpowsour.2023.233641.

Selcuk Temiz, Salim Erol, Hasan Kurban, and Mehmet M. Dalkilic. State of charge and temperature-dependent impedance spectra regeneration of lithium-ion battery by duplex learning modeling. *Journal of Energy Storage*, 64:107085, 2023. doi: 10.1016/j.est.2023.107085.

Alexander Tong, Kilian Fatras, Nikolay Malkin, Guillaume Huguet, Yanlei Zhang, Jarrid Rector-Brooks, Guy Wolf, and Yoshua Bengio. Improving and generalizing flow-based generative models with minibatch optimal transport. *Transactions on Machine Learning Research*, 2024. URL https://arxiv.org/abs/2302.00482.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30, 2017. URL https://papers.nips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html.

Wladislaw Waag, Stefan Käbitz, and Dirk Uwe Sauer. Experimental investigation of the lithium-ion battery impedance characteristic at various conditions and aging states and its influence on the application. *Applied Energy*, 102:885–897, 2013. doi: 10.1016/j.apenergy.2012.09.030.

Fujin Wang, Zhi Zhai, Zhibin Zhao, Yi Di, and Xuefeng Chen. Physics-informed neural network for lithium-ion battery degradation stable modeling and prognosis. *Nature Communications*, 15: 4332, 2024. doi: 10.1038/s41467-024-48779-z.

Ziding Wang, Xiongfei Du, Xuebing Han, Guangjun Qian, Xingyu Zhou, Shuoyuan Mao, Depeng Wang, Jianfeng Hua, and Minggao Ouyang. Arrhenius-anchored impedance thermometry for transferable LFP battery temperature estimation in grid-scale storage. *Applied Energy*, 427: 128785, 2027. doi: 10.1016/j.apenergy.2026.128785.

Haixu Wu, Tengge Hu, Yong Liu, Hang Zhou, Jianmin Wang, and Mingsheng Long. TimesNet: Temporal 2D-variation modeling for general time series analysis. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=ju_Uqw384Oq.

Jian Wu, Jinhao Meng, Mingqiang Lin, Wei Wang, Ji Wu, and Daniel-Ioan Stroe. Lithium-ion battery state of health estimation using a hybrid model with electrochemical impedance spectroscopy. *Reliability Engineering & System Safety*, 252:110450, 2024. doi: 10.1016/j.ress.2024.110450.

Bizhong Xia, Zhanpeng Qin, Hailing Fu, and Hongye Fu. Rapid estimation of battery state of health using partial electrochemical impedance spectra and interpretable machine learning. *Journal of Power Sources*, 603:234413, 2024. doi: 10.1016/j.jpowsour.2024.234413.

Han Zhang, Xiaofan Gui, Shun Zheng, Ziheng Lu, Yuqi Li, and Jiang Bian. BatteryML: An open-source platform for machine learning on battery degradation. In *International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=sxGugrYhP9.

Yunwei Zhang, Qiaochu Tang, Yao Zhang, Jiabin Wang, Ulrich Stimming, and Alpha A. Lee. Identifying degradation patterns of lithium ion batteries from impedance spectroscopy using machine learning. *Nature Communications*, 11:1706, 2020. doi: 10.1038/s41467-020-15235-7.

A Benchmark Data Card and Task Details

This appendix provides the data card for SparseEIS-Bench, including dataset scale, measurement and condition coverage, available labels, and protocols for reconstruction and downstream state-estimation tasks.

A.1 Dataset Summary

An EIS sample corresponds to a frequency scan of a battery under a specific operating condition. The main benchmark contains 25,090 measured spectra from 406 batteries, 1,218,599 raw frequency-point records, and 21 capacity–chemistry folders. Three- and five-point inputs are selected from these raw frequency-point records and do not constitute additional measurements. Table 4 summarizes the scale and measurement coverage of the main benchmark.

Table 4: **Dataset summary.** Scale and measurement coverage of the main benchmark.

Item	Value
Capacity–chemistry folders	21
Measured spectra	25,090
Raw measurement point records	1,218,599
Battery ID groups used for splitting	406
Folder-level nominal capacity range	0.04–628 Ah
Raw points per spectrum	14–89

The fixed benchmark contains 406 Battery IDs used for splitting. Battery ID formatting is standardized before splitting, after which Battery ID is used as the sole grouping key. All spectra with the same Battery ID belong to only one of the training, validation, or test subsets.

A.2 Chemistry and Condition Taxonomy

The measured spectra cover cathode chemistries including LCO, LFP, NCA, and NCM, as well as graphite and several graphite–silicon composite anodes. Table 5 retains the original category granularity from the data audit: NCM811 is listed separately as an explicit NCM subtype, and mixed folders are not reinterpreted as a single chemistry.

Table 5: **Chemistry composition of measured spectra.** Cathode and anode labels and their corresponding numbers of measured spectra.

Cathode	Spectra	Anode	Spectra
LCO	9,110	Graphite	24,123
NCM	7,804	Graphite–Si	500
LFP	6,393	Graphite–SiO_x	344
NCA	1,249	C–SiO	102
NCM811	344	Silicon–graphite	21
NCA/NCM mixed	190	–	–
Total	25,090	Total	25,090

SOC, SOH, capacity, and temperature are standardized to percent, percent, Ah, and degrees Celsius, respectively. Table 6 reports coverage across the measured spectra. Spectrum counts are derived from the weighted coverage rates in the field audit and are therefore reported as approximate values. A total of 24,343 measured spectra have all three labels: temperature, SOC, and SOH.

A.3 Metadata Schema and Measurement Provenance

Data curation retains raw measurement-record identifiers, sampling frequencies, and complex impedance values, while mapping operating-condition information that can be aligned to unified

Table 6: **Condition and target coverage.** Field availability across the 25,090 measured spectra.

Field	Coverage	Spectra	Benchmark role
SOC	100.00%	25,090	reconstruction condition / state-estimation target
SOH	99.02%	$\approx 24{,}843$	reconstruction condition / state-estimation target
Capacity	99.77%	$\approx 25{,}033$	coverage descriptor and optional target
Temperature	98.01%	$\approx 24{,}590$	reconstruction condition / state-estimation target

fields. Table 7 lists the main fields. Nyquist plots use the convention of plotting Z' against $-Z''$; `Im_Ohm` in the underlying data always denotes the imaginary impedance component itself.

Table 7: **Benchmark metadata schema.** Roles of raw records, grouping identifiers, and conditioning fields.

Field	Meaning	Benchmark handling
SpectrumID	Index of an individual EIS spectrum	Retrieval and auditing
Battery ID	Battery-level grouping identifier	Sole grouping key for training, validation, and test splits
f, Z', Z''	Raw frequency and complex impedance	Retain original samples
Cathode / anode	Electrode chemistry	Retain original categories; list mixed categories separately
Capacity	Capacity and value provenance	Retain measured, inferred, or nominal-estimate status
SOC / SOH / temperature	Measurement conditions and states	Standardize units and retain missingness masks

A.4 Curation and Unified Spectral Representation

After verifying measurement-record completeness, checking frequency validity, and constructing reference curves for the 25,090 measured spectra, all 25,090 spectra are included in the fixed reconstruction task. Each task sample retains both its raw frequency-point records and its unified reference representation. The unified representation uses $L = 512$ logarithmically spaced positions from 0.01 Hz to 100 kHz, with PCHIP applied separately to Z' and Z'' along $\log_{10} f$. PCHIP is evaluated only within each spectrum's original measurement range, without extrapolation. A grid position is included in the validity mask only if it satisfies both conditions: it lies within the original measurement range, and the gap between the adjacent measured frequencies bracketing it does not exceed 0.35 decades. The resulting validity mask v varies across spectra, and evaluation is restricted to positions with valid reference values.

This processing distinguishes raw measured frequency points, the unified learning grid, and per-spectrum validity masks. The unified 512-point array does not represent 512 direct measurements and contains no extrapolated responses beyond the original frequency range. Three- and five-point inputs are determined from the retained raw frequency-point records. For spectrum n and observation budget $B \in \{3, 5\}$, we first define

$$a_{n,b}^{(B)} = 10^{\log_{10} f_{n,\min}^{\text{raw}} + \frac{b}{B-1}\left(\log_{10} f_{n,\max}^{\text{raw}} - \log_{10} f_{n,\min}^{\text{raw}}\right)}, \quad b = 0, \ldots, B-1. \tag{15}$$

We then select the actual measurement nearest in log-frequency distance from the raw frequency-point records:

$$j_n(a_{n,b}^{(B)}) = \arg\min_j \left| \log_{10} f_{nj}^{\text{raw}} - \log_{10} a_{n,b}^{(B)} \right|. \tag{16}$$

The model receives the recorded $f^{\text{raw}}_{n,j_n(a_{n,b}^{(B)})}$ and $Z^{\text{raw}}_{n,j_n(a_{n,b}^{(B)})}$. Selected measurement points are registered to the nearest valid positions on the unified grid. If the nearest global grid position is outside the spectrum's validity mask, the nearest valid position is used instead, ensuring $m \leq v$.

A.5 Leakage-Resistant Data Splits

The main task splits training, validation, and test sets strictly by Battery ID, with counts reported in Table 8. All 25,090 spectra in the fixed task have stable Battery IDs. All spectra sharing a Battery ID, regardless of SOC, SOH, temperature, or cycling state, belong to a single subset, and the Battery IDs in the three subsets are pairwise disjoint.

Table 8: **Fixed battery-level split.** Spectrum counts and proportions in the fixed reconstruction task.

Subset	Spectra	Ratio	Grouping constraint
Train	18,285	72.9%	Battery ID only
Validation	3,415	13.6%	Battery ID only
Test	3,390	13.5%	Battery ID only
Total	25,090	100.0%	Mutually disjoint groups

A.6 Fixed Tasks and Release Boundary

For each spectrum, the reconstruction task provides a PCHIP reference target z, a validity mask v, a set of three or five measurements selected from the raw records, and the corresponding registration mask m, with $m \leq v$. Point selections are fixed in advance for each spectrum. All methods use the same data splits, valid positions, and raw input frequencies and impedance values; frequency points are not randomly resampled during evaluation. Model inputs must include actual measurement frequencies, complex impedance values, and masks. All neural baselines in the main table that accept operating conditions use the same auxiliary fields and missingness indicators, while deterministic baselines use the same sparse impedance inputs. The valid unobserved set is $U = \{i : v_i = 1, m_i = 0\}$. Reconstruction error is computed only on U, and final outputs restore the original input values at registered positions.

Downstream state-estimation protocol. Three-point reconstruction, five-point reconstruction, and reference EIS are each used as inputs for SOH, SOC, and temperature regression. All three tasks use fixed training, validation, and test splits with disjoint Battery IDs. The estimators use the same Transformer architecture described in the main text, are initialized from a checkpoint obtained by masked pretraining without state labels, and are fine-tuned separately for each target. The target label for each task is removed from both reconstruction conditions and state-estimator inputs. SOH is defined as the ratio of current capacity to rated capacity (capacity retention). The estimator uses rated capacity and available cathode and anode chemistries as conditioning features; current capacity is used only as an auxiliary supervision target with a small loss weight, not as an input feature. Because this auxiliary target is directly related to the definition of SOH, the SOH results should be interpreted in the context of this information setting. Evaluation reports MAE and RMSE relative to the reference EIS representation, as defined in Equation (14).

Downstream state estimation is defined separately from conditional reconstruction. When SOH, SOC, or temperature is the supervised target, that variable is excluded from the input conditions of the corresponding reconstruction task and from the state estimator's inputs; the splits remain strictly disjoint by Battery ID. Mixed-cathode folders are not used to draw conclusions about a single chemistry. The benchmark release retains the metadata and access instructions required to reproduce the tasks, and records are processed according to the data-management and de-identification requirements of this study.

Release boundary. At submission, we will make available a de-identified portion of the benchmark data, field definitions and processing documentation, Battery ID splits, predetermined per-spectrum three- and five-point observation masks, preprocessing and evaluation code, and the configurations required for reproduction. Upon acceptance, we will release all 25,090 processed measured spectra, together with the fixed splits, observation masks, reference baseline implementation, and complete experimental configurations. The scope of raw measurement-record release follows the applicable data-management and de-identification requirements.

Supplementary material. We provide anonymous supplementary material containing benchmark evaluation resources and examples of the core MOH model implementation. The benchmark subset uploaded with the supplementary material at submission contains 2,000 EIS spectra spanning 23 cohorts and includes training, validation, and test splits based on battery-level grouping. These numbers describe the subset provided at submission, not the full scale of the main benchmark. We also provide benchmark evaluation code and a simplified MOH implementation that illustrates key model components. This supplementary material primarily supports benchmark evaluation and facilitates understanding of the proposed method.

A.7 Measurement and Inference Time Estimates for Representative Cases

This section compares three- and five-point scan-plus-inference time against full-frequency scan time. For the two representative examples in Figure 5, the actual selected frequencies determine the three- and five-point scan times, to which per-spectrum inference time is added. The full-frequency baseline includes scan time only, with no inference time added. The reference full-frequency scan used for comparison runs from $10\,\mathrm{kHz}$ to $0.01\,\mathrm{Hz}$ with 10 logarithmic intervals per decade, covering six decades and 61 frequency points. This reference range is independent of the unified $0.01\,\mathrm{Hz}$ to $100\,\mathrm{kHz}$ grid used for modeling.

To account for the longer periods of low-frequency measurements, we approximate acquisition time by one complete sinusoidal cycle at each frequency, excluding instrument switching, settling, and communication overhead. The reference full-frequency scan time T_{full} includes only one measurement cycle at each of the 61 frequencies and is estimated as

$$T_{\text{full}} = \sum_{k=0}^{60} \frac{1}{f_k}, \qquad f_k = 10^{4-k/10}\,\mathrm{Hz}, \qquad T_{\text{full}} \approx 486.21\,\mathrm{s}. \tag{17}$$

The three actual frequencies selected for the representative example in Figure 5(a) are

$$\mathcal{F}_3 = \{0.12589,\ 1.9953,\ 158.49\}\,\mathrm{Hz}. \tag{18}$$

Summing one complete cycle at each of these frequencies gives an estimated per-spectrum measurement time of

$$T_{\text{meas},3} = \frac{1}{0.12589} + \frac{1}{1.9953} + \frac{1}{158.49} \approx 8.451\,\mathrm{s}. \tag{19}$$

Single-sample online timing records from the current implementation yield a median inference time of $T_{\text{inf},3} = 0.180\,\mathrm{s}$. This hardware-dependent measurement is used only for the present estimate and does not constitute a cross-device speed benchmark. The total three-point scan-plus-inference time and its savings relative to full-frequency scan time are therefore

$$T_{\text{total},3} = T_{\text{meas},3} + T_{\text{inf},3} \approx 8.631\,\mathrm{s}, \tag{20}$$

$$\eta_3 = \left(1 - \frac{T_{\text{total},3}}{T_{\text{full}}}\right) \times 100\%. \tag{21}$$

The five actual frequencies selected for the representative example in Figure 5(b) are

$$\mathcal{F}_5 = \{0.063096,\ 0.12589,\ 1.9953,\ 158.49,\ 794.33\}\,\mathrm{Hz}. \tag{22}$$

Summing one complete cycle at each of these frequencies gives an estimated per-spectrum measurement time of

$$T_{\text{meas},5} = \sum_{f_i \in \mathcal{F}_5} \frac{1}{f_i} \approx 24.301\,\mathrm{s}. \tag{23}$$

The available single-sample timing records likewise use $T_{\text{inf},5} = 0.180\,\mathrm{s}$ for the five-point scheme. The total five-point scan-plus-inference time and its savings relative to full-frequency scan time are therefore

$$T_{\text{total},5} = T_{\text{meas},5} + T_{\text{inf},5} \approx 24.481\,\mathrm{s}, \tag{24}$$

$$\eta_5 = \left(1 - \frac{T_{\text{total},5}}{T_{\text{full}}}\right) \times 100\%. \tag{25}$$

For these two representative examples, the three- and five-point scan times are each estimated from one measurement cycle per frequency and combined with 0.180 s of inference per spectrum. These totals are compared with the 61-point full-frequency scan time alone, which excludes inference. Using the smaller of the two budget-specific savings as a conservative summary, both schemes reduce total scan-plus-inference time by approximately 95% relative to full-frequency scan time.

B Reference-Baseline Implementation Details

B.1 Complete Reference-Baseline Definition

Residual representation and path coefficients. The deterministic branch $\mu_\theta(y)$ of Measurement-Conditioned Operator-Homotopy Flow Matching (MOH-FM) provides a conditional prediction and remains frozen during residual-flow training. The training residual $r = z - \mu_\theta(y)$ contains both prediction bias and conditional stochasticity. For generative time $t \in [0, 1]$, the path and prior noise are defined as

$$r_t = (1-t)\epsilon + tQ_y(t)r, \tag{26}$$

$$\epsilon \sim \mathcal{N}(0, I). \tag{27}$$

$\mathcal{M}$ applies an orthonormal rFFT to each of the two real-valued channels along the log-frequency axis, retaining $K = 257$ modes when $L = 512$. Both channels share real-valued clock coefficients and are mapped back to the frequency domain by irFFT. We refer to this log-frequency Fourier representation on a finite grid as the Mellin representation. Let $u_n = (n - L/2)\Delta_{\ln f}$, where $\Delta_{\ln f}$ is the spacing of the common grid along the natural-log-frequency axis, and define the Debye kernel $h(u) = [2\cosh(u)]^{-1}$. The normalized transfer magnitude of the discrete kernel is

$$D_k = \max\left(10^{-4}, \frac{|\,\mathrm{rFFT}(\mathrm{ifftshift}(h))_k|}{\max_\ell |\,\mathrm{rFFT}(\mathrm{ifftshift}(h))_\ell|}\right). \tag{28}$$

The global coverage score $I(y)$ is given by Equation (1) in the main text. Multiplying this score by D_k^2 yields mode-dependent strengths, exponents, and clocks:

$$s_k(y) = \mathrm{clip}_{[0,1]}(I(y)D_k^2), \tag{29}$$

$$\beta_k(y) = 1 + \lambda[1 - s_k(y)]^p, \tag{30}$$

$$q_k(t, y) = t^{\beta_k(y)}, \tag{31}$$

$$Q_M(t, y) = \mathcal{M}^{-1}\,\mathrm{diag}(q_k(t, y))\mathcal{M}. \tag{32}$$

The target-residual coefficient in the global branch is $a_k = tq_k = t^{1+\beta_k}$. The global term is constructed from the observation count and frequency span, while location-specific coverage enters the full path through the local operator term.

Location-resolved operator and training. Let Ω and $\mathcal{V}$ denote the index sets of observed and valid positions, respectively. Using the local coverage score in Equation (2) of the main text, we construct the local gate and full path operator:

$$A_y(t) = \mathrm{diag}(\lambda_{\mathrm{loc}} C_i(y)^\gamma t^{p_{\mathrm{loc}}}), \tag{33}$$

$$Q_y(t) = Q_M(t, y) + A_y(t)[I - Q_M(t, y)]. \tag{34}$$

Here, I is the identity operator, and $C_i(y)$ characterizes measurement coverage near frequency position i. The local term introduces a position-dependent correction to the global branch and forms part of the measurement-conditioned operator in Figure 4. The main configuration is $(\lambda, p, \lambda_{\mathrm{loc}}, p_{\mathrm{loc}}, \gamma) = (1, 2, 0.1, 1, 1)$. Because A_y and Q_M generally do not commute, the full operator generally cannot be diagonalized in the same Mellin basis.

Differentiating Equation (26) with respect to t gives the analytic target velocity and operator derivative:

$$\dot{r}_t = -\epsilon + Q_y(t)r + t\dot{Q}_y(t)r, \tag{35}$$

$$\dot{Q}_y = (I - A_y)\dot{Q}_M + \dot{A}_y(I - Q_M). \tag{36}$$

The Shift & Observation Clamp in Figure 4 transforms the path state into $x_t = \Pi_y(\mu_\theta(y) + r_t)$, which serves as the velocity-network input. For a nonempty set S, define $\|x\|_S^2 = (2|S|)^{-1} \sum_{i\in S} \sum_{c=1}^{2} x_{ic}^2$, where c indexes the real or imaginary channel. The training objective is

$$\mathcal{L} = \mathbb{E}_{t,\epsilon,(z,y)} \left[\|v_\phi(x_t, t, y) - \dot{r}_t\|_U^2 + 0.01 \|v_\phi(x_t, t, y)\|_\Omega^2 \right]. \tag{37}$$

The deterministic branch uses a reconstruction loss. Both branches have depth 8, width 256, 8 attention heads, and Fourier encodings with 32 frequency components, totaling 23.59M parameters. Each DiT block in Figure 4 contains self-attention, SwiGLU, two adaptive normalization operations, and gated residual connections (Peebles & Xie, 2023; Shazeer, 2020). Inference uses 16 Heun integration steps with 32 velocity-field evaluations. Observed impedance values are restored at integration substeps and in the final output:

$$\hat{z} = \Pi_y(\mu_\theta(y) + \hat{r}), \tag{38}$$

$$[\Pi_y(x)]_i = \begin{cases} y_i^{\text{obs}}, & i \in \Omega, \\ x_i, & i \notin \Omega. \end{cases} \tag{39}$$

Here, y_i^{obs} denotes the input observation restored at registered position $i \in \Omega$. This projection ensures numerical consistency after registration; the electrochemical properties of unobserved responses are assessed through independent diagnostics.

B.2 Implementation Conventions

Transforms and numerical conventions. The rFFT is applied separately to the real and imaginary channels. Real-valued gating coefficients preserve the conjugate symmetry required by the Fourier representation of real signals, and irFFT then returns real-valued outputs. The endpoint proof relies on the invertible transform within this real-signal representation and does not require rFFT to be invertible over arbitrary complex vectors. When computing the Debye kernel, u is first clipped to $[-30, 30]$ before evaluating the hyperbolic cosine, and D_k is lower-bounded by 10^{-4}. In Equation (1), Δ_v is numerically lower-bounded by 10^{-12}; when there is at most one observation, we set $\Delta_o = 0$. These coverage quantities depend only on the frequency range and masks.

Architecture. Both the deterministic and velocity branches have depth 8, width 256, 8 attention heads, and Fourier encodings with 32 components, totaling 23.59M parameters. The generative backbone uses DiT and SwiGLU (Peebles & Xie, 2023; Shazeer, 2020). Heun integration uses 16 steps, requiring 32 velocity-field evaluations per spectrum.

Ablation definitions and controlled variables. The deterministic prediction branch is trained first and then frozen during residual-flow training. For each random seed, Global-only MOH and full MOH-FM use the same global path parameters $\lambda = 1, p = 2$, paired model initializations, training steps, optimizer, and remaining training settings. Global-only MOH sets $\lambda_{\text{loc}} = 0$, retaining only the global MOH path, whereas full MOH-FM sets $(\lambda_{\text{loc}}, p_{\text{loc}}, \gamma) = (0.1, 1, 1)$ to introduce the local position-dependent correction. The only difference is whether the local coverage operator is enabled, making this a controlled ablation of the local operator. Linear residual FM sets $Q = I$. The scalar power path matched to the mean exponent is defined as

$$\bar{\beta}(y) = K^{-1} \sum_{k=1}^{K} \beta_k(y), \qquad Q_{\text{scalar}}(t, y) = t^{\bar{\beta}(y)} I. \tag{40}$$

The scalar power path uses a globally uniform time schedule.

Training and conditioning. Masked pretraining of the downstream estimator without state labels uses only data associated with training-set Battery IDs; validation and test batteries are excluded from pretraining. Reconstruction and velocity-regression losses are first averaged over the valid unobserved positions of each spectrum and then aggregated within a batch. The velocity constraint at input positions has a weight of 0.01. To ensure controlled comparisons, all neural models in Table 2 strictly share the Battery ID splits, predetermined per-spectrum three- and five-point input masks, frequency and impedance representations, normalization rules, valid unobserved evaluation set, and cRMSE definition. Each model uses a fixed training configuration, and the main-table metrics are

computed by the same evaluation procedure. Neural-model results report the mean and standard deviation of test-set aggregate cRMSE over repeated training with three random seeds. Deterministic baselines do not report this standard deviation because fixed inputs yield a unique output. When predicting a particular state, its label is removed from the reconstruction conditions. State estimators for three-point reconstruction, five-point reconstruction, and reference EIS are initialized from the same checkpoint obtained by masked pretraining without state labels and are then fine-tuned separately with supervision.

C Reference-Baseline Analysis and Diagnostics

C.1 Endpoint and Analytic Diagnostics

Proposition 2 (Endpoints and observation consistency)**.** *Let $\lambda \geq 0$, $p \geq 1$, $p_{\text{loc}} \geq 1$, $\gamma > 0$, and $0 \leq \lambda_{\text{loc}} \leq 1$. Equation* (26) *satisfies $r_0 = \epsilon, r_1 = r$. Equation* (39) *preserves all observed positions under exact arithmetic.*

Endpoints and observation consistency. Since $\beta_k \geq 1$, we have $q_k(0) = 0, q_k(1) = 1$, and hence $Q_M(1) = I$. Equation (34) further gives $Q_y(1) = I + A_y(1)(I - I) = I$, and $tQ_y(t) \to 0$ as $t \to 0^+$. Substitution into Equation (26) yields $r_0 = \epsilon$ and $r_1 = r$. Observation consistency follows directly from the piecewise definition in Equation (39). This projection result holds under exact arithmetic; in practical floating-point computation, the maximum observed-point error after denormalization is $1.19 \times 10^{-7}\,\Omega$.

Scalar conditional variance. Fix y and let $X = b\epsilon + ar$ and $V = \dot{b}\epsilon + \dot{a}r$. The jointly Gaussian covariance identities give

$$\text{Var}(X) = b^2 + a^2\sigma^2, \qquad \text{Var}(V) = \dot{b}^2 + \dot{a}^2\sigma^2, \tag{41}$$

$$\text{Cov}(V, X) = b\dot{b} + a\dot{a}\sigma^2. \tag{42}$$

The conditional variance formula for a joint Gaussian distribution gives $\text{Var}(V \mid X, y) = \text{Var}(V) - \text{Cov}(V, X)^2 / \text{Var}(X)$, which simplifies to Equation (12). The conditional mean μ does not affect this covariance calculation, so the result also holds for nonzero conditional means. For $a = t$ and $b = 1 - t$, the numerator in Equation (12) is σ^2, and the denominator approaches 1 as $t \to 0^+$. For $a = t^{1+\beta}$, we have

$$a\dot{b} - b\dot{a} = -t^{\beta}[(1 + \beta) - \beta t]. \tag{43}$$

Thus, as $t \to 0^+$, the conditional velocity variance approaches zero at rate $O(t^{2\beta})$. This result applies to the early-time local limit of a conditionally Gaussian path.

Multiple real modal coordinates. We expand the complex rFFT coefficients into energy-normalized real Fourier coordinates, using the corresponding single real coordinate for each endpoint mode. This allows Gaussian distributions to be analyzed in a real vector space without treating complex coefficients directly as independent real Gaussian variables. Let $r \mid y \sim \mathcal{N}(\mu, \Sigma)$ and $\epsilon \sim \mathcal{N}(0, I)$ be independent conditional on y. Write $X = B\epsilon + Tr$ and $V = \dot{B}\epsilon + \dot{T}r$. At times when C_{XX} is invertible,

$$C_{XX} = BB^{\top} + T\Sigma T^{\top}, \tag{44}$$

$$C_{VV} = \dot{B}\dot{B}^{\top} + \dot{T}\Sigma\dot{T}^{\top}, \tag{45}$$

$$C_{VX} = \dot{B}B^{\top} + \dot{T}\Sigma T^{\top}, \tag{46}$$

$$\text{Cov}(V \mid X, y) = C_{VV} - C_{VX}C_{XX}^{-1}C_{VX}^{\top}. \tag{47}$$

If Σ and the path matrices share a block-diagonal structure under the same coordinate grouping, Equation (47) decomposes by block. If they are all diagonal, each real coordinate satisfies Equation (12). A nonzero μ changes only the conditional mean, not the conditional covariance. For the full operator with a noncommuting local term, the matrix formula above can be evaluated using $B = (1 - t)I, T = tQ_y(t)$. However, the ordering of global gates alone is insufficient to determine a positive-semidefinite ordering of the conditional covariance matrices of different paths.

D Additional Evaluation and Diagnostics

KK and negative real impedance. Reconstruction performance, physical consistency, and probabilistic evaluation each have corresponding result files; the accompanying materials include per-spectrum predictions and aggregate metrics. The KK check uses finite-band Voigt projection, with a relative projection residual of at most 0.05 as the passing criterion. The median residual of the PCHIP reference representation is 0.001939. The point-level negative-real-part rates are 0.00381% and 0.00497% for three- and five-point predictions, respectively, defined as the number of valid predicted frequency points with a negative real component divided by the total number of valid predicted frequency points. The corresponding spectrum-level rates are 0.531% and 0.501%, defined as the number of predicted spectra containing at least one negative-real-part point divided by the total number of predicted spectra included in the check. This section uses Voigt projection to assess finite-band consistency of MOH-FM outputs.

ECM retention. We fit the following equivalent circuit model separately to reference and reconstructed spectra (Santoni et al., 2024):

$$Z_{\mathrm{ECM}}(\omega) = R_0 + \frac{R_{ct}}{1 + j\omega\tau} + \frac{A_W(1-j)}{\sqrt{2\omega}}, \qquad \omega = 2\pi f. \tag{48}$$

Here, j is the imaginary unit, and R_0, R_{ct}, τ, and A_W denote ohmic resistance, charge-transfer resistance, the relaxation time constant, and the diffusion-term coefficient, respectively. Parameters are kept positive through an exponential mapping, with log-parameters constrained to $[-20, 20]$. Optimization uses at most 250 function evaluations and requires at least 12 valid frequency positions. Approximately 17.7% of the test spectra are independently sampled for each input budget, so the three- and five-point subsets are unpaired. Inclusion requires successful fits for both the reference and predicted spectra and a relative reference-spectrum fitting residual of at most 0.15. MdAPE measures the relative deviation between parameters fitted to the two sets of spectra. The median fitting residuals for reference spectra in the three- and five-point settings are $0.05543/0.05293$, respectively, compared with $0.08398/0.07198$ for predicted spectra. In the order R_0, R_{ct}, τ, A_W, Spearman correlations are $(0.9632, 0.9901, 0.9400, 0.9890)$ in the three-point setting and $(0.9805, 0.9931, 0.9815, 0.9934)$ in the five-point setting.

Marginal uncertainty metrics. For each input budget, we perform 16 stochastic inference runs per test spectrum using the same corresponding reference checkpoint and construct intervals from the empirical 5% and 95% quantiles. At scoring time, interval width, Gaussian NLL, and CRPS are normalized using the two-channel RMS of the valid reference spectrum; inference continues to use the scale determined by the input impedance. Gaussian NLL is a Gaussian-approximation score based on the predictive sample mean and standard deviation, with the standard deviation lower-bounded by 10^{-3}. For either the real or imaginary component, let x_s denote the sth prediction and z the reference value. The sample estimate of CRPS is (Gneiting & Raftery, 2007)

$$\widehat{\mathrm{CRPS}} = \frac{1}{S}\sum_{s=1}^{S} |x_s - z| - \frac{1}{2S^2}\sum_{s=1}^{S}\sum_{s'=1}^{S} |x_s - x_{s'}|, \quad S = 16. \tag{49}$$