PHYSICS-GUIDED PORT-GRAPH LEARNING FOR FASTENER PARAMETER IDENTIFICATION FROM VEHICLE RESPONSES

Anonymous authors
Paper under double-blind review

ABSTRACT

Identifying physical parameters from partial observations of coupled dynamical systems is a fundamental inverse problem, yet remains challenging when target effects are entangled with unknown excitation, unobserved states, and nuisance parameter variation. We study this setting in vehicle–track systems, where segment-shared fastener stiffness and damping are inferred from a single short record of vehicle-side responses without direct rail or slab measurements. We first assess local identifiability under track-irregularity and vehicle-parameter variation using profiled Fisher information. The analysis shows that the two target parameters retain distinct sensitivity directions and that their informative cues vary across frequency. Motivated by these observations, we propose a physics-guided port-graph learning framework that combines frequency-resolved magnitude and phase with the known mechanical interaction structure of the coupled system. A typed port graph propagates observed vehicle information through suspension, wheel–rail contact, and fastener interactions while representing unobserved track components as latent nodes. Experiments on simulated single-pass responses show improved parameter identification over strong spectral and graph-based baselines, with the clearest gains under short observation windows. Ablations further confirm the complementary roles of spectral resolution, phase information, and mechanically structured interaction. These results highlight physical interaction structure as an effective inductive bias for inverse identification from partial and temporally limited observations.

1 INTRODUCTION

Many dynamical systems reveal their physical properties only indirectly through the responses of interacting components (Kay, 1993; De Felice et al., 2024; Janny et al., 2024). When only a subset of system states is observable, the influence of a target parameter must propagate through the underlying coupling structure before it appears in the measured response. At the same time, unknown excitation, unobserved states, and uncertain system parameters can induce similar response variations, making inverse identification inherently ambiguous (Park et al., 2025; Yüksel et al., 2025; Casolo et al., 2026). This difficulty is particularly pronounced when observations are both indirect and temporally limited, since the available measurements may contain only a small portion of the dynamical evidence needed to distinguish different physical causes.

Vehicle–track interaction provides a representative example. As illustrated in Figure 1, onboard sensors provide accessible vehicle responses during a train passage, whereas the dynamic states of the rail and supporting track structure remain unobserved. Track properties nevertheless affect vehicle motion through wheel–rail contact and the coupled vehicle–track dynamics, which has motivated drive-by monitoring for infrastructure assessment (Yang et al., 2004; Cantero & Basu, 2015; Matsuoka et al., 2021; Shi et al., 2024).

We focus on jointly identifying segment-shared fastener stiffness k_f and damping c_f from vehicle-side responses collected during a single passage. Unlike previous fastener studies that primarily formulate condition assessment as discrete classification (Chen et al., 2022; Yuan et al., 2021), our goal is to recover continuous physical parameters without direct rail or slab measurements. The task

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.

presents two coupled difficulties: changes in k_f and c_f can be confounded with track irregularity and uncertain vehicle parameters, while a short single-pass record provides only limited dynamic evidence. This raises a basic recoverability question: *whether the two fastener parameters remain distinguishable after nuisance variation is taken into account*.

We address this question before constructing the estimator. A local sensitivity analysis based on profiled Fisher information shows that fastener stiffness and damping retain independent identification directions under the evaluated nuisance variations. Their sensitivities also vary substantially across frequency, and dynamic responses separate fastener stiffness from related suspension effects beyond their quasi-static relation. These observations motivate retaining fine-grained spectral evidence rather than compressing the response into coarse summaries. Accordingly, we represent each vehicle channel using frequency-resolved magnitude and circular phase, which provide complementary information for stiffness and damping.

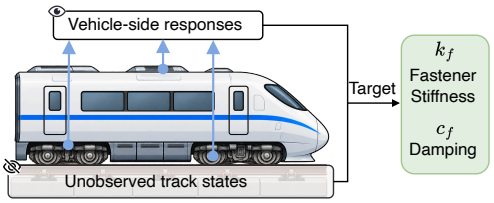

Figure 1: **Partial-observation fastener parameter identification from vehicle responses.** During a single passage, vehicle-side responses are observed while the dynamic states of the rail and slab remain inaccessible. The objective is to infer the segment-shared fastener stiffness k_f and damping c_f solely from vehicle measurements.

The remaining challenge is to relate these observed vehicle responses to the unobserved track components through which the target parameters act. Fastener effects propagate along known mechanical paths involving suspensions, wheel–rail contact, the rail, slabs, and fasteners (Zhai et al., 2009). We therefore encode the coupled system as a typed mechanical port graph, where observed vehicle nodes are initialized from response-dependent spectral features and unobserved rail and slab nodes are represented by sample-independent latent embeddings. Message passing along the known force-element topology enables interaction between measured and latent components without requiring direct track measurements. Together, the frequency-resolved representation and mechanical port graph form our *Physics-Guided Port-Graph Learning* framework for joint fastener parameter identification.

Experiments on simulated single-pass vehicle–track responses show that the proposed model improves parameter identification over strong spectral and graph-based baselines, with the clearest gains under short observation windows. This trend supports mechanical interaction structure as a useful inductive bias when the available observational evidence is limited.

Our contributions are threefold:

- We characterize the identifiability of fastener stiffness and damping from partial vehicle-side observations under nuisance variation.
- We propose a physics-guided port-graph model that couples frequency-resolved representations with known mechanical interaction structure.
- Experiments show improved fastener parameter identification, particularly under short observation windows where dynamic evidence is limited.

2 Related Work

2.1 Physical Parameter Identification from Partial Observations

Inferring hidden states or physical properties from partial observations has been studied through graph-based virtual sensing, learned physical simulators, joint state–parameter representations, generative models, and Bayesian inference (De Felice et al., 2024; Janny et al., 2024; Yang et al., 2022; Huang et al., 2024; Vadeboncoeur et al., 2026; MacKinlay et al., 2025). Related system-identification methods infer unknown dynamics or parameters using meta-learned context, contrastive representations, active interaction, and visual observations (Blanke & Lelarge, 2024; Nzoyem et al., 2025; Gonzalez Laiz et al., 2025; Memmel et al., 2024; Zhu et al., 2024; Liu et al., 2026; Lou et al., 2026). These approaches demonstrate that latent physical quantities can be recov-

ered from indirect observations, but differ substantially in the available sensing, trajectory length, or intervention assumptions.

For inverse problems governed by dynamical systems or differential equations, graph networks, inverse operators, sensitivity-aware learning, latent-space mappings, and generative approaches provide complementary mechanisms for parameter recovery (Zhao et al., 2022; Cho & Son, 2025; Behroozi et al., 2025; Gupta et al., 2025; Tauberschmidt et al., 2026). When sensing is fixed and only a short trajectory is available, however, recoverability itself becomes a central issue: finite observation length, nuisance variation, and dynamical structure determine whether target parameters can be distinguished (Kay, 1993; Park et al., 2025; Yüksel et al., 2025; Casolo et al., 2026). Our setting follows this regime and explicitly evaluates whether fastener stiffness and damping retain independent sensitivity directions before constructing the estimator.

2.2 Physics-Guided Graph and Spectral Representations

Physics-guided learning incorporates known system structure into objectives or architectures (Raissi et al., 2019; Karniadakis et al., 2021). Graph neural networks provide a natural representation for interacting physical components (Kipf & Welling, 2017; Gilmer et al., 2017; Battaglia et al., 2018), and have been extended to learned physical simulation, boundary-aware modeling, equivariant dynamics, PDE interactions, and Hamiltonian systems (Sanchez-Gonzalez et al., 2020; Wang et al., 2024; Liu et al., 2024; Zeng et al., 2025; Tan et al., 2026; Hoang et al., 2026). Related graph formulations have also encoded transport structure and network constraints (Hettige et al., 2024; Demirci et al., 2026). Most of these methods focus on forward dynamics, state reconstruction, or field prediction (Sanchez-Gonzalez et al., 2020; Wang et al., 2024; Zeng et al., 2025; Tan et al., 2026); in contrast, we use the known force-element topology as an inductive bias for inverse parameter identification from partial observations.

Frequency-aware representations provide a complementary way to preserve dynamical information that may be obscured in the time domain. Recent time-series models retain or decompose spectral structure for detection and forecasting (Wu et al., 2025; Huang et al., 2025). Our approach differs in that the spectral representation is tied directly to the inverse-identification problem: nuisance-profiled sensitivity analysis motivates retaining frequency-resolved information, while fastener impedance motivates preserving both magnitude and phase. These response features initialize the observed vehicle nodes and are subsequently exchanged with latent track nodes along typed mechanical connections.

3 Method

We estimate segment-shared fastener stiffness and damping from a single window of vehicle-side acceleration, while rail and slab responses remain unobserved. This inverse problem requires distinguishing parameter-dependent response variations from those induced by uncertain vehicle parameters and track irregularity. Our framework addresses this challenge by combining an identifiability-guided spectral representation with physics-guided port-graph learning. First, multi-channel accelerations are transformed into frequency-resolved features that retain both magnitude and phase, with the representation motivated by sensitivity analysis accounting for nuisance parameters. These features are then aggregated into body-wise embeddings and assigned to observed vehicle nodes. Next, typed message passing exchanges information between vehicle nodes and latent track nodes along suspension, wheel–rail contact, and fastener connections. The resulting graph representations are used to predict fastener stiffness and damping relative to their nominal values. Figure 2 summarizes the framework, and the following subsections detail the problem formulation, spectral representation, and graph-based estimator.

3.1 Problem Formulation

We consider the coupled vehicle–slab-track system of Zhai et al. Zhai et al. (2009), consisting of a 10-DOF vehicle, a modal rail, and free–free slabs connected through suspensions, wheel–rail contact, and fasteners. The segment-shared fastener parameters are represented as log multipliers of

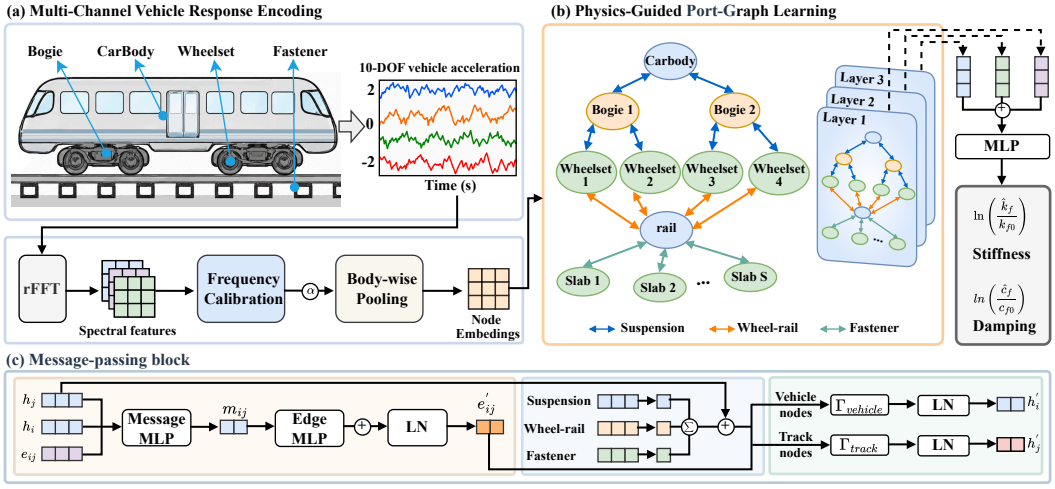

Figure 2: **Overview of the proposed framework.** (a) Multi-channel vehicle accelerations are transformed into frequency-resolved spectral representations and aggregated into body-wise node embeddings. (b) The vehicle–track system is organized as a typed mechanical port graph, where information propagates along suspension, wheel–rail contact, and fastener connections before joint prediction of fastener stiffness and damping. (c) Each message-passing block updates edge and node representations using typed interactions and type-wise averaging.

their nominal values k_{f0} and c_{f0}:

$$\boldsymbol{\theta} = \begin{bmatrix} \theta_k \\ \theta_c \end{bmatrix} = \begin{bmatrix} \ln(k_f/k_{f0}) \\ \ln(c_f/c_{f0}) \end{bmatrix} \in \mathbb{R}^2. \tag{1}$$

For an observation window of duration T, let $\mathbf{y}_T \in \mathbb{R}^{n_T}$ stack all vehicle acceleration samples. Rail and slab responses are unobserved. We model

$$\mathbf{y}_T = h_T(\boldsymbol{\theta}, \boldsymbol{\eta}) + \boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_{n_T}), \tag{2}$$

where $\boldsymbol{\eta} \in \mathbb{R}^p$ contains nine vehicle-parameter multipliers and the unknown track irregularity. The estimator predicts

$$\hat{\boldsymbol{\theta}} = f_{\boldsymbol{\phi}}(\mathbf{y}_T), \tag{3}$$

without using fastener parameters, time, or vehicle speed as inputs. The primary setting uses $T = 1.0\,\mathrm{s}$.

3.2 Identifiability-Guided Spectral Representation

Before constructing the filter-bank encoder, we first examine whether fastener stiffness and damping remain distinguishable from nuisance variation in the available vehicle-side responses. This analysis also reveals which response directions carry parameter-specific information and therefore guides the subsequent representation design.

At an evaluation point $(\boldsymbol{\theta}^\star, \boldsymbol{\eta}^\star)$, let

$$\mathbf{A} = \frac{\partial h_T}{\partial \boldsymbol{\theta}} \in \mathbb{R}^{n_T \times 2}, \qquad \mathbf{B} = \frac{\partial h_T}{\partial \boldsymbol{\eta}} \in \mathbb{R}^{n_T \times p} \tag{4}$$

denote the target and nuisance Jacobians. Projecting target sensitivities away from the nuisance subspace gives

$$\mathbf{P}_B = \mathbf{B}(\mathbf{B}^\top \mathbf{B})^\dagger \mathbf{B}^\top, \qquad \mathbf{S}_\theta = \mathbf{A}^\top (\mathbf{I}_{n_T} - \mathbf{P}_B)\mathbf{A}, \qquad \mathbf{I}_{\text{eff}} = \sigma^{-2}\mathbf{S}_\theta, \tag{5}$$

where $(\cdot)^\dagger$ denotes the Moore–Penrose pseudoinverse.

Theorem 1 (First-order identifiability under nuisance variation)**.** *At an evaluation point, the following conditions are equivalent:*

1. $\mathbf{S}_\theta \succ 0$;
2. $\operatorname{rank}[(\mathbf{I}_{n_T} - \mathbf{P}_B)\mathbf{A}] = 2$;
3. *every solution of* $\mathbf{A}\delta\boldsymbol{\theta} + \mathbf{B}\delta\boldsymbol{\eta} = \mathbf{0}$ *satisfies* $\delta\boldsymbol{\theta} = \mathbf{0}$.

Proof. Let

$$\mathbf{G} = (\mathbf{I}_{n_T} - \mathbf{P}_B)\,\mathbf{A}. \tag{6}$$

Since $\mathbf{I}_{n_T} - \mathbf{P}_B$ is a symmetric idempotent projector,

$$\mathbf{S}_\theta = \mathbf{G}^\top\mathbf{G}, \qquad \delta\boldsymbol{\theta}^\top\mathbf{S}_\theta\delta\boldsymbol{\theta} = \|\mathbf{G}\delta\boldsymbol{\theta}\|_2^2\,. \tag{7}$$

Hence, $\mathbf{S}_\theta \succ 0$ if and only if $\mathbf{G}$ has a trivial nullspace, which is equivalent to $\operatorname{rank}(\mathbf{G}) = 2$. Moreover, $\mathbf{G}\delta\boldsymbol{\theta} = \mathbf{0}$ holds if and only if $\mathbf{A}\delta\boldsymbol{\theta}$ lies in $\operatorname{col}(\mathbf{B})$, in which case there exists $\delta\boldsymbol{\eta}$ such that $\mathbf{A}\delta\boldsymbol{\theta} + \mathbf{B}\delta\boldsymbol{\eta} = \mathbf{0}$. □

For a differentiable representation $r : \mathbb{R}^{n_T} \to \mathbb{R}^m$ with Jacobian $\mathbf{R}$, identifiability after encoding similarly requires

$$\operatorname{rank}\left[(\mathbf{I}_m - \mathbf{P}_{RB})\mathbf{R}\mathbf{A}\right] = 2, \tag{8}$$

where $\mathbf{P}_{RB}$ projects onto $\operatorname{col}(\mathbf{RB})$.

For the evaluated vehicle–track configurations, $\mathbf{S}_\theta$ remains rank two under the tested nuisance settings. The target sensitivities also vary markedly with frequency; in particular, dynamic responses separate fastener stiffness from primary-suspension stiffness beyond their quasi-static relation. We therefore retain individual spectral components rather than coarse frequency summaries.

After zero-phase anti-alias filtering and resampling to $256\,\mathrm{Hz}$, the real FFT of vehicle channel j gives

$$\mathbf{X}_j = [X_{j,1}, \ldots, X_{j,M}]^\top \in \mathbb{C}^M, \qquad M = \left\lfloor \frac{N}{2} \right\rfloor, \tag{9}$$

after removing the DC component. With $\boldsymbol{\varphi}_j = \arg(\mathbf{X}_j)$, we form

$$\mathbf{q}_j = \begin{bmatrix} \log(1 + |\mathbf{X}_j|) \\ \cos\boldsymbol{\varphi}_j \\ \sin\boldsymbol{\varphi}_j \end{bmatrix} \in \mathbb{R}^{3M}. \tag{10}$$

Magnitude and phase are complementary because fastener stiffness and damping enter different quadratures of

$$Z_f(\omega) = k_f + i\omega c_f. \tag{11}$$

After standardization with training-set statistics, a projection shared across channels produces

$$\mathbf{h}_j = \mathrm{SiLU}\left(\mathbf{W}_{\mathrm{enc}}\widetilde{\mathbf{q}}_j + \mathbf{b}_{\mathrm{enc}}\right) \in \mathbb{R}^d. \tag{12}$$

A separate projection is trained for each observation duration because M varies with T. We refer to this shared per-channel projection as the filter-bank encoder.

3.3 Physics-Guided Port-Graph Learning

We represent the vehicle–track topology as a directed multigraph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. Nodes correspond to vehicle bodies, the rail, and slabs, while suspension, wheel–rail contact, and fastener elements define bidirectional typed edges.

Let $\mathcal{V}_{\mathrm{veh}}$ and $\mathcal{V}_{\mathrm{track}}$ denote the vehicle and track nodes, and $\mathcal{D}_i$ the acceleration channels assigned to vehicle node i. Node states are initialized as

$$\mathbf{v}_i^{(0)} = \begin{cases} \dfrac{1}{|\mathcal{D}_i|}\displaystyle\sum_{j\in\mathcal{D}_i} \mathbf{h}_j + \mathbf{p}_i, & i \in \mathcal{V}_{\mathrm{veh}}, \\ \mathbf{p}_i, & i \in \mathcal{V}_{\mathrm{track}}, \end{cases} \tag{13}$$

where $\mathbf{p}_i \in \mathbb{R}^d$ is a learnable node embedding. Thus, vehicle nodes contain sample-dependent observations, whereas rail and slab nodes remain latent.

For directed edge e, let $g(e) \in \{\text{susp}, \text{contact}, \text{fastener}\}$ denote its interaction type and $\rho(e)$ its direction. Its initial state is

$$\mathbf{z}_e^{(0)} = \mathbf{t}_{g(e)} + \mathbf{r}_{\rho(e)}, \tag{14}$$

with learnable type and direction embeddings.

At message-passing layer ℓ, edge $e = (s \to r)$ is updated by

$$\mathbf{m}_e^{(\ell)} = \Xi_\ell \left([\mathbf{v}_s^{(\ell)}; \mathbf{v}_r^{(\ell)}; \mathbf{z}_e^{(\ell)}] \right), \tag{15}$$

$$\mathbf{z}_e^{(\ell+1)} = \mathrm{LN}_{\mathrm{edge},\ell} \left(\mathbf{z}_e^{(\ell)} + \Phi_\ell [\mathbf{z}_e^{(\ell)}; \mathbf{m}_e^{(\ell)}] \right). \tag{16}$$

Because the numbers of suspension, contact, and fastener edges differ, incoming messages are averaged within each interaction type before aggregation:

$$\mathbf{a}_i^{(\ell)} = \sum_{g:\, |\mathcal{E}_{i,g}|>0} \frac{1}{|\mathcal{E}_{i,g}|} \sum_{e \in \mathcal{E}_{i,g}} \Psi_\ell \left(\mathbf{z}_e^{(\ell+1)} \right), \tag{17}$$

where $\mathcal{E}_{i,g}$ is the set of type-g edges entering node i. The node update is

$$\mathbf{v}_i^{(\ell+1)} = \mathrm{LN}_{\tau(i),\ell} \left(\mathbf{v}_i^{(\ell)} + \Gamma_{\tau(i),\ell} [\mathbf{v}_i^{(\ell)}; \mathbf{a}_i^{(\ell)}] \right), \tag{18}$$

where $\tau(i) \in \{\text{veh}, \text{track}\}$. The mappings Ξ_ℓ, Φ_ℓ, Ψ_ℓ, and $\Gamma_{\tau(i),\ell}$ are two-layer MLPs. We use $L = 3$ message-passing layers.

3.4 Joint Parameter Inference and Optimization

Let $\mathcal{E}_f^+$ denote the rail-to-slab fastener edges. After L graph layers, we compute

$$\bar{\mathbf{v}}_{\text{veh}}^{(L)} = \frac{1}{|\mathcal{V}_{\text{veh}}|} \sum_{i \in \mathcal{V}_{\text{veh}}} \mathbf{v}_i^{(L)}, \qquad \bar{\mathbf{z}}_f^{(L)} = \frac{1}{|\mathcal{E}_f^+|} \sum_{e \in \mathcal{E}_f^+} \mathbf{z}_e^{(L)}. \tag{19}$$

The target parameters are predicted from

$$\hat{\boldsymbol{\theta}} = H_{\boldsymbol{\phi}} \left([\bar{\mathbf{v}}_{\text{veh}}^{(L)}; \mathbf{v}_{\text{rail}}^{(L)}; \bar{\mathbf{z}}_f^{(L)}] \right) \in \mathbb{R}^2, \tag{20}$$

where $H_{\boldsymbol{\phi}} : \mathbb{R}^{3d} \to \mathbb{R}^2$ is a two-layer MLP.

For a minibatch $\mathcal{B}$, all learnable parameters are optimized using

$$\mathcal{L}(\boldsymbol{\phi}) = \frac{1}{2|\mathcal{B}|} \sum_{b \in \mathcal{B}} \sum_{q \in \{k,c\}} \ell_\delta \left(\hat{\theta}_q^{(b)} - \theta_q^{(b)} \right), \tag{21}$$

where ℓ_δ is the Huber loss with $\delta = 0.1$. Gaussian observation noise is resampled during training, and no gradients are propagated through the vehicle–track dynamics solver. Further implementation details are provided in Appendix C.

4 Experiments

4.1 Data, Baselines, and Evaluation

Data. We jointly estimate segment-shared fastener stiffness and damping from a single vehicle-acceleration passage. The data are generated by a coupled vehicle–slab-track simulator; its configuration and data-generation protocol are detailed in Appendix C.1. The simulator provides 2048 training, 128 validation, and 256 in-distribution test records, together with a held-out wheelset-mass test split. Each record contains up to 512 frames at 256 Hz over 2.0 s, and the primary evaluation uses the first 1.0 s.

Baselines. We compare against a parameter-matched spectral Transformer and ridge regression on the complete complex log-periodogram (complex-periodogram ridge). Additional neural baselines include a graph attention network, spectral MLP, and spectral CNN. The w/o-port-graph control and the conv. + pool, w/o-graph control are included to isolate the effects of the port graph and filter-bank encoder. The nominal-value predictor provides a task-scale reference. Model sizes are reported in Appendix B, and complete settings in Appendix C.

Evaluation & Implementation. We report log-RMSE separately for $\ln k_f$ and $\ln c_f$, with log-MAE provided in Appendix B. Learned models are reported as mean $\pm$ standard deviation over three training seeds. Primary comparisons use 10,000 paired bootstrap resamples over matched test records and training seeds; a difference is considered resolved when the corresponding 95% interval excludes zero. All neural models are implemented in PyTorch and optimized with AdamW using the same Huber objective. Experiments are conducted on four NVIDIA A100 GPUs.

4.2 Main Identification Results

Table 1 compares all estimators at the primary 1.0-s observation window. Relative to the parameter-matched spectral Transformer, our port-graph model reduces log-RMSE by approximately 16%, 10%, and 13% for in-distribution stiffness, in-distribution damping, and shifted stiffness, respectively; all three differences are resolved by paired bootstrap intervals. It also improves over complex-periodogram ridge in the same three settings. For shifted damping, the differences among the strongest methods remain unresolved. Overall, the results support the proposed representation and mechanical structure under the primary short-observation setting.

Table 1: Log-RMSE at 1.0 s. Learned entries are mean $\pm$ standard deviation over three training seeds. Bold and underline mark the lowest and second-lowest point estimates in each column.

Estimator	ID $\ln k_f$	ID $\ln c_f$	Shift $\ln k_f$	Shift $\ln c_f$
Port graph	**0.0957**$_{\pm.0030}$	**0.1294**$_{\pm.0031}$	**0.1138**$_{\pm.0031}$	<u>0.1715</u>$_{\pm.0031}$
Spectral Transformer	<u>0.1136</u>$_{\pm.0040}$	<u>0.1432</u>$_{\pm.0038}$	<u>0.1307</u>$_{\pm.0033}$	0.1829$_{\pm.0029}$
Graph attention network	0.1145$_{\pm.0022}$	0.1571$_{\pm.0037}$	0.1546$_{\pm.0091}$	0.1744$_{\pm.0011}$
Spectral MLP	0.1270$_{\pm.0030}$	0.1520$_{\pm.0033}$	0.1508$_{\pm.0021}$	0.1789$_{\pm.0040}$
Spectral CNN	0.1472$_{\pm.0021}$	0.1587$_{\pm.0008}$	0.1510$_{\pm.0060}$	0.1860$_{\pm.0073}$
w/o port graph	0.1171$_{\pm.0027}$	0.1588$_{\pm.0077}$	0.1381$_{\pm.0042}$	0.1751$_{\pm.0017}$
Conv. + pool, w/o graph	0.1479$_{\pm.0023}$	0.1692$_{\pm.0038}$	0.1564$_{\pm.0063}$	0.1853$_{\pm.0087}$
Complex-periodogram ridge	0.1235	0.1485	0.1347	**0.1709**
Nominal-value predictor	0.1820	0.1772	0.1794	0.1875

4.3 Component Analysis

Figure 3 examines successive changes to the encoder, graph interaction, and phase representation at the 1.0-s in-distribution setting. Replacing convolution and pooling with the frequency-resolved encoder primarily expands the recovered stiffness range, while adding the port graph further improves both targets and notably expands the damping range. Removing phase from the full model contracts the prediction region and increases damping error more strongly than stiffness error. These results indicate complementary contributions from frequency-resolved encoding, mechanical graph interaction, and phase information.

4.4 Spectral Representation

Table 2 isolates the effect of spectral representation using the same ridge readout. Replacing the handcrafted band summary with the full-resolution magnitude spectrum reduces stiffness log-RMSE by 29.0%. At fixed frequency resolution, adding circular phase reduces damping log-RMSE by 12.4% relative to magnitude alone, while stiffness changes little. These results are consistent with the sensitivity analysis in Section 3.2: fine-grained spectral resolution is important for stiffness identification, while phase provides complementary information for damping.

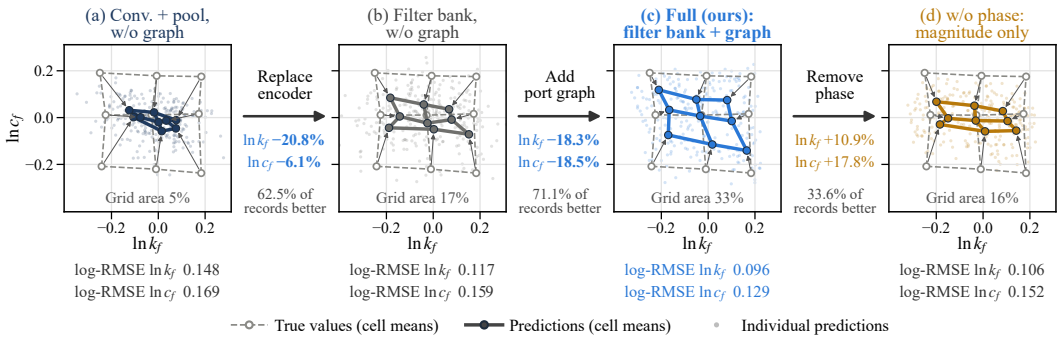

Figure 3: Component ablations on 1.0-s in-distribution records. (a) Conv. + pool, w/o graph; (b) filter-bank encoder, w/o port graph (the w/o-port-graph control in Table 1); (c) complete port graph; (d) w/o phase (magnitude only).

Table 2: Spectral-representation controls on full-window records.

Resolution	Retained content	Dim.	ln k_f	R^2	ln c_f	R^2
A. Summary granularity						
coarse	bands, coherence, phase	144	0.1558	0.2552	0.1486	0.2947
full	magnitude	2560	0.1106	0.6242	0.1513	0.2690
B. Content at full resolution						
full	magnitude	2560	0.1106	0.6242	0.1513	0.2690
full	circular phase	5120	0.1583	0.2306	0.1418	0.3585
full	magnitude + phase	7680	0.1107	0.6238	0.1325	0.4398

4.5 Effect of Observation Duration

Figure 4 evaluates models trained independently at eight observation durations. Below the full 2.0-s window, the port graph achieves 16 resolved gains over the spectral Transformer across 28 target–condition comparisons, with no resolved losses; all four comparisons favor the port graph at the shortest duration. Against complex-periodogram ridge, 21 of 32 comparisons favor the port graph and none favor ridge. As the observation window increases, the differences become less pronounced. This trend supports the role of mechanical structure as an inductive bias when the available dynamic evidence is limited.

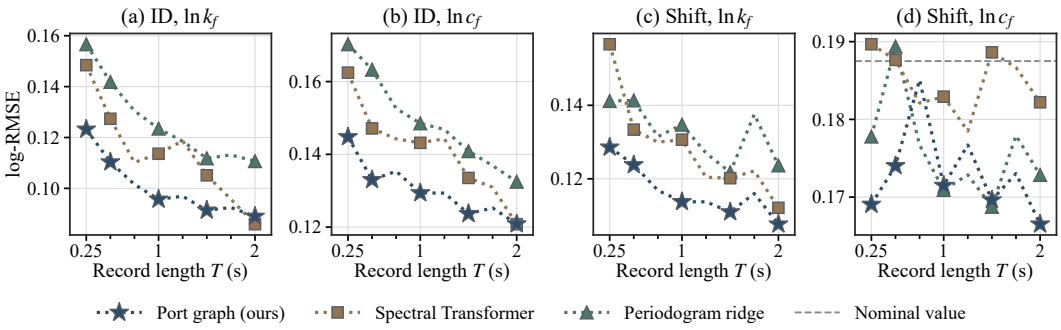

Figure 4: Log-RMSE over the eight-duration ladder. Markers highlight five durations while lines include all eight observations.

4.6 Effect of Test Noise

Figure 5 evaluates robustness to test noise while keeping the trained models fixed. Performance varies across the two target parameters: under strong noise, the port graph retains a clear advantage for damping, whereas its stiffness advantage does not persist. The w/o-port-graph control is also highly sensitive to a mismatch between training and test noise. These results indicate that the benefit

of structured interaction depends on both the target parameter and the observation-noise regime. Exact values and variability across noise realizations are reported in Appendix B.

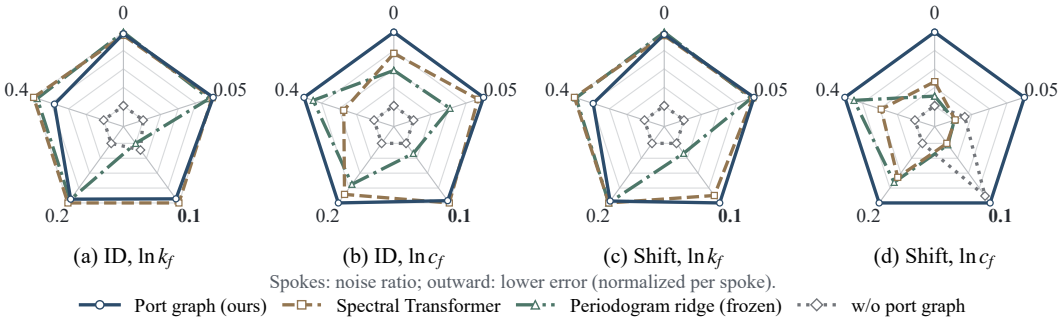

Figure 5: Test-noise sensitivity at the full window. Radar axes are test-noise ratios; each axis has its own error scale, with lower log-RMSE plotted farther outward. Panels cover in-distribution and wheelset-mass-shift stiffness and damping.

4.7 Parameter-Space Coverage

Figure 6 shows that the port-graph predictions cover a larger fraction of the ground-truth parameter support while remaining relatively compact. Their 5th–95th percentile spans recover 83.5% of the true stiffness range and 71.2% of the true damping range, whereas ridge predictions are more concentrated along damping and the Transformer covers a narrower stiffness range. The visualization indicates that the port-graph model's lower identification errors are accompanied by broad, well-aligned parameter-space coverage rather than indiscriminate dispersion, although some contraction toward the center remains.

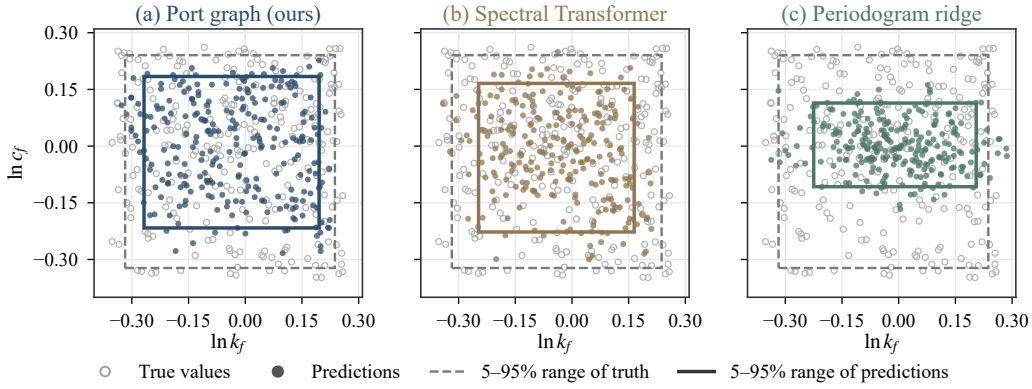

Figure 6: Parameter-space coverage on identical in-distribution records. Panels compare the port graph, spectral Transformer, and ridge. Open and filled points denote true and predicted parameters; dashed and solid rectangles delimit marginal 5th–95th percentile ranges.

5 Conclusion

We studied fastener stiffness and damping identification from partial vehicle-side observations in a coupled vehicle–track system. A profiled Fisher-information analysis showed that the two target parameters remained locally identifiable under the evaluated nuisance variations, while their frequency-dependent sensitivities motivated the use of fine-grained magnitude and phase representations. Based on these observations, we developed a physics-guided port-graph model that propagated vehicle-response information along known mechanical interaction paths while representing unobserved track components as latent nodes. Experiments showed improved identification over competitive baselines, especially under short observation windows.

AI USE STATEMENT

In this work, we did not use generative AI tools for research tasks requiring disclosure; we used generative AI tools only for language polishing, grammar checking, and organizing the presentation of the authors' research ideas. We reviewed all AI-assisted work to ensure that it accurately reflects the original methodology and results. We take responsibility for the final content of this work, including text produced with the aid of generative AI.

ETHICS STATEMENT

This work uses simulated vehicle–track responses and involves no human subjects or private personal data. The findings are limited to the evaluated synthetic settings; using predicted fastener properties to guide railway maintenance would require validation on field data and engineering review.

REPRODUCIBILITY STATEMENT

All experiments were run on four NVIDIA A100 GPUs. The complete code and execution environment are provided in the uploaded code supplement.

REFERENCES

Peter W. Battaglia, Jessica B. Hamrick, Victor Bapst, Alvaro Sanchez-Gonzalez, Vinicius Zambaldi, Mateusz Malinowski, Andrea Tacchetti, David Raposo, Adam Santoro, Ryan Faulkner, et al. Relational inductive biases, deep learning, and graph networks. *arXiv preprint arXiv:1806.01261*, 2018.

Abdolmehdi Behroozi, Chaopeng Shen, and Daniel Kifer. Sensitivity-constrained fourier neural operators for forward and inverse problems in parametric differential equations. In *International Conference on Learning Representations*, 2025.

Matthieu Blanke and Marc Lelarge. Interpretable meta-learning of physical systems. In *International Conference on Learning Representations*, 2024.

Daniel Cantero and Biswajit Basu. Railway infrastructure damage detection using wavelet transformed acceleration response of traversing vehicle. *Structural Control and Health Monitoring*, 22(1):62–70, 2015.

Cecilia Casolo, Sören Becker, and Niki Kilbertus. Identifiability challenges in sparse linear ordinary differential equations. In *International Conference on Learning Representations*, 2026.

Mei Chen, Wanming Zhai, Shengyang Zhu, Lei Xu, and Yu Sun. Vibration-based damage detection of rail fastener using fully convolutional networks. *Vehicle System Dynamics*, 60(7):2191–2210, 2022.

Sung Woong Cho and Hwijae Son. Physics-informed deep inverse operator networks for solving PDE inverse problems. In *International Conference on Learning Representations*, 2025.

Giovanni De Felice, Andrea Cini, Daniele Zambon, Vladimir Gusev, and Cesare Alippi. Graph-based virtual sensing from sparse and partial multivariate observations. In *International Conference on Learning Representations*, 2024.

Ege Demirci, Francesco Bullo, Ananthram Swami, and Ambuj K. Singh. FlowSymm: Physics-aware, symmetry-preserving graph attention for network flow completion. In *International Conference on Learning Representations*, 2026.

Justin Gilmer, Samuel S. Schoenholz, Patrick F. Riley, Oriol Vinyals, and George E. Dahl. Neural message passing for quantum chemistry. In *International Conference on Machine Learning*, pp. 1263–1272, 2017.

Rodrigo Gonzalez Laiz, Tobias Schmidt, and Steffen Schneider. Self-supervised contrastive learning performs non-linear system identification. In *International Conference on Learning Representations*, 2025.

Naveen Gupta, Medha Sawhney, Arka Daw, Youzuo Lin, and Anuj Karpatne. A unified framework for forward and inverse problems in subsurface imaging using latent space translations. In *International Conference on Learning Representations*, 2025.

Kethmi Hirushini Hettige, Jiahao Ji, Shili Xiang, Cheng Long, Gao Cong, and Jingyuan Wang. AirPhyNet: Harnessing physics-guided neural networks for air quality prediction. In *International Conference on Learning Representations*, 2024.

Tai Hoang, Alessandro Trenta, Alessio Gravina, Niklas Freymuth, Philipp Becker, Davide Bacciu, and Gerhard Neumann. Improving long-range interactions in graph neural simulators via hamiltonian dynamics. In *International Conference on Learning Representations*, 2026.

Jiahe Huang, Guandao Yang, Zichen Wang, and Jeong Joon Park. DiffusionPDE: Generative PDE-solving under partial observation. In *Advances in Neural Information Processing Systems*, volume 37, 2024.

Songtao Huang, Zhen Zhao, Can Li, and Lei Bai. TimeKAN: KAN-based frequency decomposition learning architecture for long-term time series forecasting. In *International Conference on Learning Representations*, 2025.

Steeven Janny, Madiha Nadri, Julie Digne, and Christian Wolf. Space and time continuous physics simulation from partial observations. In *International Conference on Learning Representations*, 2024.

George E. Karniadakis, Ioannis G. Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3:422–440, 2021.

Steven M. Kay. *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*. Prentice Hall, 1993.

Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations*, 2017.

Chunjiang Liu, Xiaoyuan Wang, Qingran Lin, Albert Xiao, Haoyu Chen, Shizheng Wen, Hao Zhang, Lu Qi, Ming-Hsuan Yang, Laszlo A. Jeni, Min Xu, and Yizhou Zhao. Multi-object system identification from videos. In *International Conference on Learning Representations*, 2026.

Yang Liu, Jiashun Cheng, Haihong Zhao, Tingyang Xu, Peilin Zhao, Fugee Tsung, Jia Li, and Yu Rong. SEGNO: Generalizing equivariant graph neural networks with physical inductive biases. In *International Conference on Learning Representations*, 2024.

Haozhe Lou, Mingtong Zhang, Haoran Geng, Hanyang Zhou, Sicheng He, Zhiyuan Gao, Siheng Zhao, Jiageng Mao, Pieter Abbeel, Jitendra Malik, Daniel Seita, and Yue Wang. D-REX: Differentiable real-to-sim-to-real engine for learning dexterous grasping. In *International Conference on Learning Representations*, 2026.

Dan MacKinlay, Russell Tsuchida, Daniel Pagendam, and Petra Kuhnert. Gaussian ensemble belief propagation for efficient inference in high-dimensional, black-box systems. In *International Conference on Learning Representations*, 2025.

Kodai Matsuoka, Hirofumi Tanaka, Kyohei Kawasaki, Claudio Somaschini, and Andrea Collina. Drive-by methodology to identify resonant bridges using track irregularity measured by high-speed trains. *Mechanical Systems and Signal Processing*, 158:107667, 2021.

Marius Memmel, Andrew Wagenmaker, Chuning Zhu, Dieter Fox, and Abhishek Gupta. ASID: Active exploration for system identification in robotic manipulation. In *International Conference on Learning Representations*, 2024.

Roussel Desmond Nzoyem, David Barton, and Tom Deakin. Neural context flows for meta-learning of dynamical systems. In *International Conference on Learning Representations*, 2025.

Hank Park, Grani A. Hanasusanto, and Yingying Li. Robust system identification: Finite-sample guarantees and connection to regularization. In *International Conference on Learning Representations*, 2025.

Maziar Raissi, Paris Perdikaris, and George E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019.

Alvaro Sanchez-Gonzalez, Jonathan Godwin, Tobias Pfaff, Rex Ying, Jure Leskovec, and Peter W. Battaglia. Learning to simulate complex physics with graph networks. In *International Conference on Machine Learning*, pp. 8459–8468, 2020.

Tao Shi, Ping Lou, and T. Y. Yang. Optimized machine learning methods for identifying the stiffness loss of CRTS-II slab track based on vehicle vibration signals. *Advanced Engineering Informatics*, 62:102886, 2024.

Yingjie Tan, Quanming Yao, and Yaqing Wang. DGNet: Discrete green networks for data-efficient learning of spatiotemporal PDEs. In *International Conference on Learning Representations*, 2026.

Jan Tauberschmidt, Sophie Fellenz, Sebastian Vollmer, and Andrew Duncan. Physics-constrained fine-tuning of flow-matching models for generation and inverse problems. In *International Conference on Learning Representations*, 2026.

Arnaud Vadeboncoeur, Gregory Duthé, Mark Girolami, and Eleni Chatzi. Geometric autoencoder priors for bayesian inversion: Learn first observe later. In *International Conference on Learning Representations*, 2026.

Haixin Wang, Jiaxin Li, Anubhav Dwivedi, Kentaro Hara, and Tailin Wu. BENO: Boundary-embedded neural operators for elliptic PDEs. In *International Conference on Learning Representations*, 2024.

Xingjian Wu, Xiangfei Qiu, Zhengyu Li, Yihang Wang, Jilin Hu, Chenjuan Guo, Hui Xiong, and Bin Yang. CATCH: Channel-aware multivariate time series anomaly detection via frequency patching. In *International Conference on Learning Representations*, 2025.

Tsung-Yen Yang, Justinian P. Rosca, Karthik Narasimhan, and Peter J. Ramadge. Learning physics constrained dynamics using autoencoders. In *Advances in Neural Information Processing Systems*, volume 35, 2022.

Y. B. Yang, C. W. Lin, and J. D. Yau. Extracting bridge frequencies from the dynamic response of a passing vehicle. *Journal of Sound and Vibration*, 272(3–5):471–493, 2004.

Zhandong Yuan, Shengyang Zhu, Xuancheng Yuan, and Wanming Zhai. Vibration-based damage detection of rail fastener clip using convolutional neural network: Experiment and simulation. *Engineering Failure Analysis*, 119:104906, 2021.

Oğuz Kaan Yüksel, Mathieu Even, and Nicolas Flammarion. Long-context linear system identification. In *International Conference on Learning Representations*, 2025.

Bocheng Zeng, Qi Wang, Mengtao Yan, Yang Liu, Ruizhi Chengze, Yi Zhang, Hongsheng Liu, Zidong Wang, and Hao Sun. PhyMPGN: Physics-encoded message passing graph network for spatiotemporal PDE systems. In *International Conference on Learning Representations*, 2025.

Wanming Zhai, Kaiyun Wang, and Chengbiao Cai. Fundamentals of vehicle–track coupled dynamics. *Vehicle System Dynamics*, 47(11):1349–1376, 2009.

Qingqing Zhao, David B. Lindell, and Gordon Wetzstein. Learning to solve PDE-constrained inverse problems with graph networks. In *Proceedings of the 39th International Conference on Machine Learning*, pp. 26895–26910, 2022.

Xiangming Zhu, Huayu Deng, Haochen Yuan, Yunbo Wang, and Xiaokang Yang. Latent intuitive physics: Learning to transfer hidden physics from a 3D video. In *International Conference on Learning Representations*, 2024.

A IDENTIFIABILITY DIAGNOSTICS

This appendix gives the nuisance construction, local information calculation, and numerical probes underlying Section 3.2.

A.1 NUISANCE MODEL

In Equation 2, $\boldsymbol{\eta}$ contains nine vehicle-parameter multipliers and a finite representation of track irregularity. For spatial frequency cells $\{\kappa_j\}_{j=1}^{K_c}$, an unknown-phase irregularity is represented as

$$r(x) = \sum_{j=1}^{K_c} \left[a_j \cos(\kappa_j x) + b_j \sin(\kappa_j x)\right]. \tag{22}$$

Thus each frequency cell contributes two real nuisance coefficients; the fixed-phase construction is obtained by retaining one coefficient per cell.

The time–bandwidth product $N_{\mathrm{TB}}(T) \simeq vT(\kappa_{\max} - \kappa_{\min})$ sets a nominal representation scale: approximately 41 cells at the 1.0-s operating point and 82 cells at the 2-s endpoint. We evaluate fixed-phase and sine–cosine banks at $K_c \in \{2/8/16\}$.

The noise scale is set by the estimator protocol: the global training-split frame-domain RMS of 2.750545 and the training noise ratio 0.1 give $\sigma = 0.275\,054\,5\,\mathrm{m\,s^{-2}}$, one scale for every Fisher bound below. It is also the operating point around which Figure 5 varies test-time noise.

A.2 FISHER INFORMATION

Linearizing Equation 2 at a physical state gives

$$\delta\mathbf{y} = \mathbf{A}\,\delta\boldsymbol{\theta} + \mathbf{B}\,\delta\boldsymbol{\eta} + \boldsymbol{\varepsilon}, \qquad \mathbf{A} = \frac{\partial \mathbf{h}_T}{\partial \boldsymbol{\theta}}, \quad \mathbf{B} = \frac{\partial \mathbf{h}_T}{\partial \boldsymbol{\eta}}. \tag{23}$$

For isotropic Gaussian noise, the joint Fisher matrix is

$$\mathcal{I} = \frac{1}{\sigma^2}\begin{bmatrix} \mathbf{A}^\top\mathbf{A} & \mathbf{A}^\top\mathbf{B} \\ \mathbf{B}^\top\mathbf{A} & \mathbf{B}^\top\mathbf{B} \end{bmatrix}. \tag{24}$$

The two targets are jointly locally identifiable whenever $\mathbf{S}_\theta$ has rank two. The corresponding marginal Cramér–Rao bounds are

$$\mathrm{Cov}(\widehat{\boldsymbol{\theta}}) \succeq \sigma^2 \mathbf{S}_\theta^{-1}. \tag{25}$$

For the registered two-cell, band-spanning nuisance bank at the 1.0-s operating point, we obtain

$$\mathbf{S}_\theta = \begin{bmatrix} 52.6248 & -34.8236 \\ -34.8236 & 30.4533 \end{bmatrix}, \qquad \lambda(\mathbf{S}_\theta) = \{4.9935, 78.0846\}. \tag{26}$$

Its condition number is 15.64 and the correlation implied by the inverse information is 0.8699. For the registered two-cell bank, the fixed-noise marginal Cramér–Rao standard deviations are $\sigma_{\ln k_f} = 0.076869$ and $\sigma_{\ln c_f} = 0.101048$. The nuisance matrix of this bank has rank 11/11.

The profiled target matrix has rank two at the nominal point and all four tested parameter-box corners in each of the 8 evaluated irregularity bases, and thus for all 32 basis–corner combinations. The smallest corner eigenvalue over those combinations is 3.4599, at the quadrature basis with $K_c = 8$ and multipliers $(1.15, 0.85)$ on (k_f, c_f). Within the registered two-cell bank, the fixed-noise corner bounds differ from the nominal values by at most 3.2% for $\ln k_f$ and 8.2% for $\ln c_f$. These results establish local joint identifiability across the evaluated basis, center, and corner set.

Wheelset mass is the leading nuisance group. Its sensitivity has $|\cos(\ln k_f, \ln m_w)| = 0.8614$ with the stiffness direction, and it remains the closest vehicle-parameter direction across the evaluated irregularity bases. Mechanically, wheelset mass enters the balance

$$m_w \ddot{z}_w = f_p - f_c,$$

at the interface through which contact forces are transmitted to the primary suspension. Its perturbation therefore changes the same vehicle-side transfer path that carries the fastener response.

The bounds in Equation 25 characterize local, record-level estimation at a specified physical state. Section 4 reports errors averaged over sampled target and nuisance distributions.

A.3 Static and Dynamic Sensitivities

A quasi-static reduction places the primary suspension stiffness k_p and fastener stiffness k_f in series,

$$k_{\mathrm{eq}} = \left(k_p^{-1} + k_f^{-1}\right)^{-1} = \frac{k_p k_f}{k_p + k_f}.$$

Their logarithmic sensitivities satisfy

$$\frac{\partial k_{\mathrm{eq}}/\partial \ln k_f}{\partial k_{\mathrm{eq}}/\partial \ln k_p} = \frac{k_p}{k_f} = 0.09408, \qquad |\cos(\mathbf{j}_{k_f}, \mathbf{j}_{k_p})| = 1, \tag{27}$$

because both perturbations act through the same scalar equivalent stiffness.

The moving dynamic response departs strongly from this limiting direction. A complementary steady-state frequency sweep gives a maximum observed cosine of 0.4615, while the sensitivity-norm ratio changes from 5.7241 at $0.5\,\mathrm{Hz}$ to 27.3820 at $41.5\,\mathrm{Hz}$ and 64.6113 at $64\,\mathrm{Hz}$. The direction and scale of the fastener sensitivity therefore vary across frequency.

Below $2\,\mathrm{Hz}$, the irregularity-driven fastener-to-suspension sensitivities have a magnitude ratio of 0.0211–0.0253 and an absolute cosine similarity of 0.5048–0.6655. The total-response sweep contains a separated sensitivity direction in the sinusoidal-sensitivity window ($24\,\mathrm{Hz}$ to $64\,\mathrm{Hz}$).

At the force-element level, stiffness and damping enter different quadratures of the fastener impedance. For harmonic relative displacement $\Delta z(t) = \Re\{\Delta Z(\omega) e^{i\omega t}\}$,

$$F_f(\omega) = (k_f + i\omega c_f)\Delta Z(\omega), \qquad \left.\frac{\partial F_f}{\partial \ln k_f}\right|_{\Delta Z} = k_f \Delta Z, \quad \left.\frac{\partial F_f}{\partial \ln c_f}\right|_{\Delta Z} = i\omega c_f \Delta Z. \tag{28}$$

These derivatives hold at fixed relative displacement. The vehicle-response derivatives also include changes in ΔZ and are represented by the columns of $\mathbf{A}$.

B Additional Results

All results in this appendix use the same simulation configuration as the main paper: 410 rail modes, 41 track slabs, $n_u = 830$, and $n_y = 1240$.

B.1 Extended One-Second Comparisons

Table 1 gives the primary log-RMSE comparison at a 1.0-s observation window. All methods share data splits and test-noise realizations; the linear baseline and constant predictor are deterministic. Each learned row is a separately trained estimator. Table 4 reports log-MAE and R^2 on the same records.

Table 3: Learned estimators at the primary 1.0-s window. Parameter counts come from the trained reports underlying Table 1 and the auxiliary untyped-edge control. Alternative widths were selected to match the port graph's capacity.

Estimator	Encoding and interaction	Parameters
Port graph	Filter bank, typed port graph	656,066
Spectral Transformer	Frequency tokens, attention	657,666
Graph attention network	Graph attention backbone	690,431
Spectral MLP	Spectral dense layers	683,714
Spectral CNN	Spectral convolution	659,970
Untyped edges	Filter bank, untyped graph	655,778
Bypass graph	Filter bank, direct readout	656,066
Conv. + pool, no graph	Convolution, pooled readout	664,642

At the 1.0-s window, a paired two-way bootstrap jointly resamples matched training seeds and test records (10,000 resamples, 95% percentile intervals; Table 5). The two factors are crossed, so one record resample is shared by every resampled seed. 3 of the four port-graph versus Transformer

Table 4: Additional metrics at the 1.0-s window on the same test records as Table 1. Learned rows report mean ± standard deviation over 3 seeds; deterministic rows have no seed spread. Lower log-MAE and higher R^2 are better.

Estimator	Metric	ID $\ln k_f$	ID $\ln c_f$	Shift $\ln k_f$	Shift $\ln c_f$
Port graph	log-MAE	$0.0758_{\pm 0.0014}$	$0.1057_{\pm 0.0033}$	$0.0929_{\pm 0.0045}$	$0.1395_{\pm 0.0033}$
Spectral Transformer	log-MAE	$0.0914_{\pm 0.0025}$	$0.1157_{\pm 0.0020}$	$0.1065_{\pm 0.0018}$	$0.1483_{\pm 0.0042}$
Graph attention network	log-MAE	$0.0913_{\pm 0.0020}$	$0.1307_{\pm 0.0019}$	$0.1258_{\pm 0.0058}$	$0.1477_{\pm 0.0028}$
Spectral MLP	log-MAE	$0.1032_{\pm 0.0026}$	$0.1202_{\pm 0.0024}$	$0.1204_{\pm 0.0028}$	$0.1468_{\pm 0.0036}$
Spectral CNN	log-MAE	$0.1217_{\pm 0.0021}$	$0.1294_{\pm 0.0017}$	$0.1238_{\pm 0.0045}$	$0.1556_{\pm 0.0064}$
Untyped edges	log-MAE	$0.0784_{\pm 0.0018}$	$0.1024_{\pm 0.0031}$	$0.0978_{\pm 0.0029}$	$0.1428_{\pm 0.0038}$
Bypass graph	log-MAE	$0.0952_{\pm 0.0041}$	$0.1283_{\pm 0.0038}$	$0.1138_{\pm 0.0045}$	$0.1419_{\pm 0.0033}$
Conv. + pool, no graph	log-MAE	$0.1203_{\pm 0.0017}$	$0.1416_{\pm 0.0008}$	$0.1280_{\pm 0.0065}$	$0.1569_{\pm 0.0047}$
Complex-periodogram ridge	log-MAE	0.1006	0.1239	0.1054	0.1450
Nominal value	log-MAE	0.1554	0.1488	0.1545	0.1595
Port graph	R^2	$0.7186_{\pm 0.0177}$	$0.4653_{\pm 0.0256}$	$0.5922_{\pm 0.0224}$	$0.1633_{\pm 0.0302}$
Spectral Transformer	R^2	$0.6035_{\pm 0.0277}$	$0.3454_{\pm 0.0351}$	$0.4621_{\pm 0.0269}$	$0.0481_{\pm 0.0302}$
Graph attention network	R^2	$0.5977_{\pm 0.0153}$	$0.2117_{\pm 0.0374}$	$0.2450_{\pm 0.0903}$	$0.1346_{\pm 0.0110}$
Spectral MLP	R^2	$0.5050_{\pm 0.0234}$	$0.2625_{\pm 0.0320}$	$0.2837_{\pm 0.0202}$	$0.0896_{\pm 0.0408}$
Spectral CNN	R^2	$0.3348_{\pm 0.0188}$	$0.1960_{\pm 0.0084}$	$0.2810_{\pm 0.0570}$	$0.0146_{\pm 0.0778}$
Untyped edges	R^2	$0.7030_{\pm 0.0055}$	$0.4888_{\pm 0.0139}$	$0.5521_{\pm 0.0335}$	$0.1288_{\pm 0.0446}$
Bypass graph	R^2	$0.5787_{\pm 0.0196}$	$0.1933_{\pm 0.0788}$	$0.3986_{\pm 0.0362}$	$0.1278_{\pm 0.0169}$
Conv. + pool, no graph	R^2	$0.3284_{\pm 0.0210}$	$0.0853_{\pm 0.0413}$	$0.2287_{\pm 0.0616}$	$0.0218_{\pm 0.0928}$
Complex-periodogram ridge	R^2	0.5318	0.2959	0.4282	0.1690
Nominal value	R^2	-0.0169	-0.0024	-0.0138	-0.0000

intervals exclude zero and favor the port graph (-0.0261 to -0.0102, -0.0245 to -0.0034, -0.0281 to -0.0059); the wheelset-mass damping interval -0.0241 to $+0.0016$ includes zero.

Table 5 also includes paired intervals for the auxiliary untyped-edge control. Against ridge regression on the complete complex periodogram, the port graph has a lower point estimate in three cells, with all 3 corresponding intervals below zero. Ridge is nominally lower on wheelset-mass damping ($\Delta = +0.0006$), but the paired interval -0.0115 to $+0.0128$ includes zero.

B.2 Observation-Duration Results

Table 6 uses the same paired record-and-seed uncertainty convention as the one-second primary comparison. Of the 28 cells below 2.00 s, 16 favor the port graph and 12 are unresolved. At the full window only shifted damping favors the graph. Against ridge, 21 of 32 cells favor the graph and the others are unresolved.

B.3 Test-Noise Results

A test-time noise sweep probes distribution mismatch around the training calibration. We re-evaluate trained models at 12 noise levels and reuse paired noise realizations across experimental arms. Figure 5 plots the resulting curves and Table 7 lists the numbers behind them. Of the 12 model–parameter combinations, 11 achieve their lowest error either at the training noise ratio 0.1 or at the preceding level. For the graph-free control, in-distribution stiffness error is 0.6352 at zero noise and 0.1053 at the training level, showing that noiseless input is itself outside the training distribution.

At the training level the spectral Transformer has marginally lower in-distribution stiffness and damping point estimates (0.0885 and 0.1184 against 0.0898 and 0.1192). At four times the training noise, the Transformer reaches 0.1710 against 0.2098 on stiffness, and frozen ridge reaches 0.1776. The port graph retains the lowest in-distribution damping point estimate there (0.1541 against 0.2490 and 0.1759). At zero noise the graph-free control reaches 0.6352 on in-distribution stiffness, against 0.1183 for the port graph.

Table 5: Paired uncertainty at the 1.0-s window for the learned estimators in Table 1, the auxiliary untyped-edge control, and the closed-form ridge. ΔRMSE is the port graph minus the compared estimator, so negative values favor the port graph. Percentile 95% intervals use 10,000 two-way bootstrap resamples of matched records and training seeds; ridge has a record axis only. An interval containing zero is unresolved.

Comparison	Split	Parameter	ΔRMSE	95% CI	Decision
graph − Transformer	ID	$\ln k_f$	-0.0179	-0.0261 to -0.0102	graph
graph − Transformer	ID	$\ln c_f$	-0.0138	-0.0245 to -0.0034	graph
graph − Transformer	wheelset	$\ln k_f$	-0.0169	-0.0281 to -0.0059	graph
graph − Transformer	wheelset	$\ln c_f$	-0.0114	-0.0241 to $+0.0016$	unresolved
graph − graph attention	ID	$\ln k_f$	-0.0187	-0.0277 to -0.0098	graph
graph − graph attention	ID	$\ln c_f$	-0.0277	-0.0386 to -0.0171	graph
graph − graph attention	wheelset	$\ln k_f$	-0.0410	-0.0573 to -0.0249	graph
graph − graph attention	wheelset	$\ln c_f$	-0.0029	-0.0196 to $+0.0138$	unresolved
graph − spectral MLP	ID	$\ln k_f$	-0.0312	-0.0426 to -0.0198	graph
graph − spectral MLP	ID	$\ln c_f$	-0.0226	-0.0373 to -0.0086	graph
graph − spectral MLP	wheelset	$\ln k_f$	-0.0370	-0.0529 to -0.0213	graph
graph − spectral MLP	wheelset	$\ln c_f$	-0.0074	-0.0260 to $+0.0103$	unresolved
graph − spectral CNN	ID	$\ln k_f$	-0.0515	-0.0627 to -0.0400	graph
graph − spectral CNN	ID	$\ln c_f$	-0.0293	-0.0417 to -0.0169	graph
graph − spectral CNN	wheelset	$\ln k_f$	-0.0373	-0.0548 to -0.0198	graph
graph − spectral CNN	wheelset	$\ln c_f$	-0.0146	-0.0331 to $+0.0027$	unresolved
graph − untyped edges	ID	$\ln k_f$	-0.0026	-0.0079 to $+0.0035$	unresolved
graph − untyped edges	ID	$\ln c_f$	$+0.0029$	-0.0045 to $+0.0101$	unresolved
graph − untyped edges	wheelset	$\ln k_f$	-0.0055	-0.0123 to $+0.0010$	unresolved
graph − untyped edges	wheelset	$\ln c_f$	-0.0035	-0.0120 to $+0.0054$	unresolved
graph − bypass graph	ID	$\ln k_f$	-0.0214	-0.0293 to -0.0132	graph
graph − bypass graph	ID	$\ln c_f$	-0.0296	-0.0433 to -0.0170	graph
graph − bypass graph	wheelset	$\ln k_f$	-0.0244	-0.0337 to -0.0151	graph
graph − bypass graph	wheelset	$\ln c_f$	-0.0036	-0.0159 to $+0.0086$	unresolved
graph − graph-free CNN	ID	$\ln k_f$	-0.0522	-0.0644 to -0.0392	graph
graph − graph-free CNN	ID	$\ln c_f$	-0.0399	-0.0530 to -0.0271	graph
graph − graph-free CNN	wheelset	$\ln k_f$	-0.0427	-0.0598 to -0.0248	graph
graph − graph-free CNN	wheelset	$\ln c_f$	-0.0139	-0.0331 to $+0.0036$	unresolved
graph − complex ridge	ID	$\ln k_f$	-0.0278	-0.0371 to -0.0183	graph
graph − complex ridge	ID	$\ln c_f$	-0.0191	-0.0299 to -0.0090	graph
graph − complex ridge	wheelset	$\ln k_f$	-0.0210	-0.0364 to -0.0065	graph
graph − complex ridge	wheelset	$\ln c_f$	$+0.0006$	-0.0115 to $+0.0128$	unresolved

Table 6: Record-and-seed paired bootstrap intervals for port graph minus spectral Transformer at each observation duration. Negative ΔRMSE favors the graph. Each interval uses 10,000 percentile resamples with matched records and training seeds. Intervals containing zero are unresolved.

Split	Target	T (s)	ΔRMSE	95% CI	Decision
ID	$\ln k_f$	0.25	-0.0252	$[-0.03703, -0.01346]$	graph
ID	$\ln k_f$	0.50	-0.0171	$[-0.02683, -0.00784]$	graph
ID	$\ln k_f$	0.75	-0.0087	$[-0.01690, -0.00040]$	graph
ID	$\ln k_f$	1.00	-0.0179	$[-0.02605, -0.01018]$	graph
ID	$\ln k_f$	1.25	-0.0218	$[-0.03146, -0.01250]$	graph
ID	$\ln k_f$	1.50	-0.0137	$[-0.02122, -0.00613]$	graph
ID	$\ln k_f$	1.75	-0.0032	$[-0.01068, +0.00435]$	unresolved
ID	$\ln k_f$	2.00	$+0.0030$	$[-0.00327, +0.00948]$	unresolved
ID	$\ln c_f$	0.25	-0.0176	$[-0.02784, -0.00728]$	graph
ID	$\ln c_f$	0.50	-0.0142	$[-0.02515, -0.00290]$	graph
ID	$\ln c_f$	0.75	-0.0090	$[-0.02136, +0.00289]$	unresolved
ID	$\ln c_f$	1.00	-0.0138	$[-0.02465, -0.00371]$	graph
ID	$\ln c_f$	1.25	-0.0149	$[-0.02512, -0.00484]$	graph
ID	$\ln c_f$	1.50	-0.0098	$[-0.02151, +0.00163]$	unresolved
ID	$\ln c_f$	1.75	-0.0056	$[-0.01668, +0.00625]$	unresolved
ID	$\ln c_f$	2.00	$+0.0000$	$[-0.01077, +0.01098]$	unresolved
Shift	$\ln k_f$	0.25	-0.0280	$[-0.04287, -0.01248]$	graph
Shift	$\ln k_f$	0.50	-0.0096	$[-0.02199, +0.00334]$	unresolved
Shift	$\ln k_f$	0.75	-0.0131	$[-0.02563, -0.00057]$	graph
Shift	$\ln k_f$	1.00	-0.0169	$[-0.02829, -0.00565]$	graph
Shift	$\ln k_f$	1.25	-0.0067	$[-0.01868, +0.00512]$	unresolved
Shift	$\ln k_f$	1.50	-0.0092	$[-0.01906, +0.00054]$	unresolved
Shift	$\ln k_f$	1.75	-0.0062	$[-0.01825, +0.00646]$	unresolved
Shift	$\ln k_f$	2.00	-0.0044	$[-0.01574, +0.00708]$	unresolved
Shift	$\ln c_f$	0.25	-0.0206	$[-0.03650, -0.00497]$	graph
Shift	$\ln c_f$	0.50	-0.0137	$[-0.03005, +0.00323]$	unresolved
Shift	$\ln c_f$	0.75	$+0.0030$	$[-0.01956, +0.02518]$	unresolved
Shift	$\ln c_f$	1.00	-0.0114	$[-0.02376, +0.00117]$	unresolved
Shift	$\ln c_f$	1.25	-0.0018	$[-0.01608, +0.01277]$	unresolved
Shift	$\ln c_f$	1.50	-0.0190	$[-0.03312, -0.00524]$	graph
Shift	$\ln c_f$	1.75	-0.0137	$[-0.02741, -0.00003]$	graph
Shift	$\ln c_f$	2.00	-0.0157	$[-0.03135, -0.00103]$	graph

Table 7: Test-time noise sweep in log-RMSE at the 2.00-s record. Models are trained with Gaussian augmentation at noise ratio 0.1, so the zero-noise column is also out of distribution. Superscript $\pm$ reports dispersion across 3 model seeds, and superscript $\circ$ reports dispersion across paired noise realizations. The ridge row uses the complete complex log-periodogram fit at the training noise ratio, with preprocessing and coefficients held fixed across test-noise levels, matching Figure 5.

estimator	0	0.05	0.1	0.2	0.4
proposed (port graph)					
$\ln kf$, in distribution	$\mathbf{0.1183}^{\pm 0.0105}_{\circ 0.0000}$	$\mathbf{0.0929}^{\pm 0.0021}_{\circ 0.0017}$	$\mathbf{0.0898}^{\pm 0.0008}_{\circ 0.0019}$	$\mathbf{0.1270}^{\pm 0.0107}_{\circ 0.0039}$	$\mathbf{0.2098}^{\pm 0.0093}_{\circ 0.0021}$
$\ln cf$, in distribution	$\mathbf{0.1261}^{\pm 0.0050}_{\circ 0.0000}$	$\mathbf{0.1183}^{\pm 0.0008}_{\circ 0.0015}$	$\mathbf{0.1192}^{\pm 0.0006}_{\circ 0.0026}$	$\mathbf{0.1313}^{\pm 0.0018}_{\circ 0.0029}$	$\mathbf{0.1541}^{\pm 0.0042}_{\circ 0.0033}$
$\ln kf$, wheelset-mass OOD	$\mathbf{0.1364}^{\pm 0.0091}_{\circ 0.0000}$	$\mathbf{0.1160}^{\pm 0.0007}_{\circ 0.0020}$	$\mathbf{0.1100}^{\pm 0.0040}_{\circ 0.0028}$	$\mathbf{0.1410}^{\pm 0.0134}_{\circ 0.0052}$	$\mathbf{0.2092}^{\pm 0.0081}_{\circ 0.0055}$
$\ln cf$, wheelset-mass OOD	$\mathbf{0.1601}^{\pm 0.0047}_{\circ 0.0000}$	$\mathbf{0.1614}^{\pm 0.0030}_{\circ 0.0017}$	$\mathbf{0.1680}^{\pm 0.0036}_{\circ 0.0021}$	$\mathbf{0.1694}^{\pm 0.0044}_{\circ 0.0036}$	$\mathbf{0.1726}^{\pm 0.0033}_{\circ 0.0041}$
spectral Transformer					
$\ln kf$, in distribution	$\mathbf{0.1295}^{\pm 0.0311}_{\circ 0.0000}$	$\mathbf{0.0962}^{\pm 0.0083}_{\circ 0.0017}$	$\mathbf{0.0885}^{\pm 0.0029}_{\circ 0.0028}$	$\mathbf{0.1192}^{\pm 0.0151}_{\circ 0.0025}$	$\mathbf{0.1710}^{\pm 0.0364}_{\circ 0.0026}$
$\ln cf$, in distribution	$\mathbf{0.1395}^{\pm 0.0183}_{\circ 0.0000}$	$\mathbf{0.1211}^{\pm 0.0084}_{\circ 0.0015}$	$\mathbf{0.1184}^{\pm 0.0049}_{\circ 0.0029}$	$\mathbf{0.1389}^{\pm 0.0040}_{\circ 0.0060}$	$\mathbf{0.2490}^{\pm 0.0157}_{\circ 0.0039}$
$\ln kf$, wheelset-mass OOD	$\mathbf{0.1443}^{\pm 0.0265}_{\circ 0.0000}$	$\mathbf{0.1196}^{\pm 0.0081}_{\circ 0.0028}$	$\mathbf{0.1127}^{\pm 0.0013}_{\circ 0.0028}$	$\mathbf{0.1375}^{\pm 0.0085}_{\circ 0.0034}$	$\mathbf{0.1766}^{\pm 0.0318}_{\circ 0.0092}$
$\ln cf$, wheelset-mass OOD	$\mathbf{0.1872}^{\pm 0.0130}_{\circ 0.0000}$	$\mathbf{0.1868}^{\pm 0.0048}_{\circ 0.0025}$	$\mathbf{0.1808}^{\pm 0.0018}_{\circ 0.0032}$	$\mathbf{0.1895}^{\pm 0.0102}_{\circ 0.0065}$	$\mathbf{0.2602}^{\pm 0.0133}_{\circ 0.0069}$
abl.: bypass the graph					
$\ln kf$, in distribution	$\mathbf{0.6352}^{\pm 0.1262}_{\circ 0.0000}$	$\mathbf{0.3086}^{\pm 0.0300}_{\circ 0.0023}$	$\mathbf{0.1053}^{\pm 0.0017}_{\circ 0.0024}$	$\mathbf{0.2425}^{\pm 0.0259}_{\circ 0.0034}$	$\mathbf{0.3016}^{\pm 0.0652}_{\circ 0.0020}$
$\ln cf$, in distribution	$\mathbf{0.1725}^{\pm 0.0272}_{\circ 0.0000}$	$\mathbf{0.1533}^{\pm 0.0101}_{\circ 0.0020}$	$\mathbf{0.1392}^{\pm 0.0028}_{\circ 0.0022}$	$\mathbf{0.1843}^{\pm 0.0312}_{\circ 0.0032}$	$\mathbf{0.3214}^{\pm 0.1150}_{\circ 0.0059}$
$\ln kf$, wheelset-mass OOD	$\mathbf{0.6475}^{\pm 0.1327}_{\circ 0.0000}$	$\mathbf{0.3181}^{\pm 0.0354}_{\circ 0.0028}$	$\mathbf{0.1316}^{\pm 0.0042}_{\circ 0.0049}$	$\mathbf{0.2466}^{\pm 0.0254}_{\circ 0.0040}$	$\mathbf{0.2985}^{\pm 0.0682}_{\circ 0.0020}$
$\ln cf$, wheelset-mass OOD	$\mathbf{0.2004}^{\pm 0.0298}_{\circ 0.0000}$	$\mathbf{0.1833}^{\pm 0.0097}_{\circ 0.0035}$	$\mathbf{0.1695}^{\pm 0.0011}_{\circ 0.0029}$	$\mathbf{0.2162}^{\pm 0.0366}_{\circ 0.0046}$	$\mathbf{0.3413}^{\pm 0.1132}_{\circ 0.0046}$
ridge on the complete log-periodogram (frozen at 0.1)					
$\ln kf$, in distribution	$\mathbf{0.1065}^{\pm - - -}_{\circ 0.0092}$	$\mathbf{0.1050}^{\pm - - -}_{\circ 0.0012}$	$\mathbf{0.1074}^{\pm - - -}_{\circ 0.0023}$	$\mathbf{0.1252}^{\pm - - -}_{\circ 0.0057}$	$\mathbf{0.1776}^{\pm - - -}_{\circ 0.0224}$
$\ln cf$, in distribution	$\mathbf{0.1502}^{\pm - - -}_{\circ 0.0251}$	$\mathbf{0.1351}^{\pm - - -}_{\circ 0.0029}$	$\mathbf{0.1357}^{\pm - - -}_{\circ 0.0039}$	$\mathbf{0.1479}^{\pm - - -}_{\circ 0.0087}$	$\mathbf{0.1759}^{\pm - - -}_{\circ 0.0165}$
$\ln kf$, wheelset-mass OOD	$\mathbf{0.1204}^{\pm - - -}_{\circ 0.0060}$	$\mathbf{0.1239}^{\pm - - -}_{\circ 0.0027}$	$\mathbf{0.1280}^{\pm - - -}_{\circ 0.0030}$	$\mathbf{0.1375}^{\pm - - -}_{\circ 0.0059}$	$\mathbf{0.1794}^{\pm - - -}_{\circ 0.0194}$
$\ln cf$, wheelset-mass OOD	$\mathbf{0.1951}^{\pm - - -}_{\circ 0.0210}$	$\mathbf{0.1871}^{\pm - - -}_{\circ 0.0081}$	$\mathbf{0.1805}^{\pm - - -}_{\circ 0.0098}$	$\mathbf{0.1856}^{\pm - - -}_{\circ 0.0220}$	$\mathbf{0.1947}^{\pm - - -}_{\circ 0.0192}$

C Implementation Details

C.1 Data and Evaluation Protocol

All records are generated by the coupled vehicle–slab-track simulator from Section 3.1; labels are simulator inputs, not quantities inferred from the responses. Each passage is integrated in double precision with time step $5.0 \times 10^{-5}\,\mathrm{s}$ using the explicit two-step Zhai scheme. After discarding the $0.2\,\mathrm{s}$ warm-up segment, the $10\,\mathrm{kHz}$ raw channels are downsampled into the model input. Excitation is a track-irregularity realization sampled independently for each example from a fixed standard spectrum. A reference solution uses 16 substeps within each recorded step. Relative to this reference, the chosen step size produces relative L_2 differences of 1.1×10^{-4} on vehicle-side records and 4.2×10^{-5} in the information band, with both decreasing monotonically under time-step refinement (Appendix Table 9).

The dataset contains 2048 training records, 128 validation records, 256 in-distribution test records, 128 wheelset-mass OOD records, and 128 unseen-speed records at $280\,\mathrm{km\,h^{-1}}$. Each sample independently draws stiffness and damping multipliers from $[0.7, 1.3]$, nine vehicle parameters from within $\pm 15\%$ of nominal values, and one irregularity realization; the in-distribution speed is $300\,\mathrm{km\,h^{-1}}$. The wheelset-mass split moves only that parameter to intervals 15% to 25% below or above nominal, while the other eight vehicle parameters remain in distribution. This attributes the shift to a single nuisance variable.

Independent Gaussian noise is added in the frame domain. Every sample uses one absolute standard deviation calibrated from a training-split reference RMS, with noise ratio $20\,\mathrm{dB}$, giving $\sigma = 0.2751$ ($0.275\,054\,\mathrm{m\,s^{-2}}$).

The Fisher calculation and learned estimators share the 1.0-s, $256\,\mathrm{Hz}$, 10-channel acquisition and the global training-split noise scale. Identification accuracy is measured by log-RMSE for both targets; Appendix Table 4 gives log-MAE and R^2 for the same residuals.

All learned models report the mean and standard deviation over the same 3 training seeds (11, 29, 47).

The spectral encoder standardizes each channel and coordinate using the training-split mean and standard deviation, floors the latter at 10^{-2}, and clips standardized values to ± 5. The training statistics are reused at validation and test time.

The linear baseline fits ridge regression to the complete complex log-periodogram from the same window. It shares preprocessing, paired noise realizations, and training-only standardization with the port graph, and selects a penalty separately for each target using only the validation split. The spectral Transformer receives the same calibrated complex features and is parameter matched to the port graph at the primary window (657,666 versus 656,066). The graph-free control retains the same encoder and width but replaces the graph summary with direct readout. A predictor that always returns nominal values anchors the error obtained without identification.

Table 1 additionally reports four parameter-matched alternatives and two structural controls, all retrained at the same half-window budget from the same code path. The alternatives use separately constructed Transformer, graph-attention, MLP, and CNN backbones; the controls test graph bypass with the filter bank retained and a joint encoder-and-graph substitution.

The primary identification comparison is reported at a 1.0-s window (Table 1); the complete observation-duration sweep is in Figure 4. The Fisher analysis of Section 3.2 uses this primary 1.0-s acquisition.

Table 8: Simulation configuration. Every reported value uses this configuration, so the listed graph dimensions are the dimensions used in the model runs.

Quantity	Value
Rail modes	410
Track slabs	41
Rail length	$264.45\,\mathrm{m}$
Degrees of freedom n_u	830
Force components n_y	1240
Graph nodes	49
Graph factors	420
Directed graph edges	840

The observation protocol archives a $2\,\mathrm{s}$ record after a $0.2\,\mathrm{s}$ warm-up, recording 512 frames at $256\,\mathrm{Hz}$ from 10 channels, and runs at $300\,\mathrm{km\,h^{-1}}$; the primary experiments take the first $1.0\,\mathrm{s}$ of each record, so 256 frames. Track slabs are $6.45\,\mathrm{m}$ long with 10 fasteners per slab. The simulator uses double precision. Study configurations, seeds, and optimizer settings are stored in version-controlled YAML files and in each report artifact, so the reported values can be reproduced without manually extracting settings from the paper.

Table 9: Trained configuration; all seeds agree on every entry. Shared training settings and the 2.00-s endpoint configuration. The parameter-count and validation-score rows refer to that endpoint; the primary 1.0-s model has 656,066 parameters. Remaining optimizer settings are stored in version-controlled study YAML and the same report files. The final four rows are simulator settings from the version-controlled physics configuration rather than training artifacts.

Setting	Value
Observation representation	`complex` (log magnitude + observed-spectrum circular phase)
Encoder	`filterbank`
Readout	`full`, normalization `off`
Aggregation	`group_mean`
Node width	96
Weight decay	0
Huber transition threshold	0.1
Noise augmentation	yes (per epoch, ratio 0.1)
Trainable parameters, 2.00 s	692930
Best validation log-RMSE, 2.00 s	$0.1025_{\pm.0010}$
Seeds	11, 29, 47
Integration scheme	explicit two-step Zhai (self-started first step, $\phi = \psi = 0$)
Integration step	5.0×10^{-5} s (0.5 / 0.5 coefficients)
Recording step	1.0×10^{-4} s, i.e. 10 kHz
Numerical precision	double (`float64`)