STRUCTURE BEFORE FLEXIBILITY: MECHANISTICALLY TESTABLE FUNCTIONAL CONDUCTANCES FOR CONTINUOUS-TIME NETWORK DYNAMICS

Anonymous authors
Paper under double-blind review

ABSTRACT

Continuous-time neural models can fit trajectories accurately while assigning the wrong contributions to individual interactions. We introduce SP-KAN-LTC, which fixes the conductance and driving-force factorization on each known edge and learns only a positive bounded state-dependent conductance, represented by a univariate spline network. A matched two-by-two factorization-by-function-class comparison shows that factorization contributes far more to mechanism recovery than the edge function class. In an eight-edge controlled benchmark with known mechanisms, SP-KAN-LTC reduces edge-current and intervention-effect error by a median of 99.8% and 99.5% relative to graph-matched and capacity-matched Current-KAN, with intervention rank correlation rising from -0.70 to $+1.00$. Within the matched perceptron family, the corresponding reductions are 99.1% and 98.9%, showing that the attribution failure is not specific to the spline parameterization. Against MLP-Conductance, the flexible spline edge contributes a smaller and conditional effect: no consistent advantage is established at four interactions, whereas intervention-effect error improves by 65.9% at eight. On measured electrophysiology, every factorized run recovers the independently recorded synaptic current more accurately than every unfactorized run, though SP-KAN-LTC does not consistently outperform the matched perceptron. Aliasing, driving-force misspecification, hidden interaction memory and held-out optogenetic transfer delimit the interpretation: accurate trajectories do not guarantee identifiable or structurally valid edge mechanisms, and external transfer is not demonstrated on *C. elegans*. The results support a structure-before-flexibility principle: constraining what an interaction is allowed to mean matters more for mechanistic fidelity than the flexibility used to parameterize it.

1 INTRODUCTION

Continuous-time neural models fit a vector field to observed trajectories (Chen et al., 2018; Hasani et al., 2021; 2022), and the accuracy of that fit constrains only the aggregate derivative. When the derivative of a state is a sum over incoming edges, distinct edge-level decompositions reproduce the same sum along the trajectories actually visited, so a predictive fit does not by itself determine the contribution assigned to each edge. The distinction matters whenever the model is used to estimate interaction strengths or to predict the effect of an intervention rather than to forecast.

Constraining the form of the edge term narrows the admissible decompositions by fixing what an interaction is permitted to mean, at the cost of restricting the laws it can express. Conductance-based dynamics are the relevant instance. For states h_i and h_j, the interaction from j to i takes the form

$$I_{ij} = c_{ij}(h_j)\,(E_{ij} - h_i), \tag{1}$$

where c_{ij} is a nonnegative state-dependent interaction strength and $E_{ij} - h_i$ is an explicit driving force. Factorizations of this kind are standard in biologically structured dynamical models (Mi et al., 2022; Burghi et al., 2025), but in Liquid Time-Constant Network (LTC)-style neural conductance models the state dependence is carried by a restricted parametric gate, typically an affine sigmoid,

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.

so interaction laws outside that family cannot in general be represented directly by the edge gate (Appendix A.1). Whether c_{ij} can be made substantially more flexible without losing the structure that gives the decomposition its meaning is the question this work addresses.

Structure-Preserving KAN Liquid Time-Constant Network (SP-KAN-LTC) holds the factorization fixed and learns only c_{ij} (Figure 1), represented per edge by a univariate B-spline Kolmogorov–Arnold Network (KAN) (Liu et al., 2025b) composed with a bounded positive map. The learned function is confined to the interaction strength (Appendix A), while the signed factorized edge current, the additive per-edge decomposition of the state derivative and the sparse graph support are preserved rather than learned.

An edge quantity is called *mechanistically testable* here when its interpretation can be evaluated independently of trajectory fit, against a known interaction law, an independently measured current, or a held-out intervention. The term carries no claim that an individual edge is identifiable or causally recovered; whether it is depends on structural and excitation conditions that the experiments vary deliberately. The architecture assigns the learned object a fixed physical role; the data must determine whether that role is supported by the observations.

In the controlled benchmark, matched controls that vary the function class and the factorization separately at fixed graph and per-edge budget, together with four studies that remove one modeling assumption at a time, separate what the architecture contributes from what the data must supply (Figure 2).

1. **A functional-conductance architecture and its controlled isolation.** A continuous-time model that learns bounded nonlinear edge conductances while preserving the edge-current factorization, together with matched controls that separate the contribution of that factorization from the contribution of the edge function class.
2. **Evidence that structure dominates function class, with conditional gains from flexibility.** Mechanism recovery and intervention fidelity are governed primarily by the factorization; the flexible edge shows a conditional advantage that varies with the tested interaction count and is consistent with state-space excitation effects, but is not explained by law complexity alone, and no measured system here establishes a function-class advantage.
3. **Testable validity boundaries for edge-level claims.** Identifiability, misspecification, hidden-memory and external-transfer experiments state the conditions under which a learned edge law can be read mechanistically, and show that predictive accuracy certifies none of them.

2 Related Work

Continuous-time neural models parameterize a vector field and integrate it, decoupling the learned dynamics from the spacing of observations (Chen et al., 2018). Augmenting the state enlarges the class of representable flows (Dupont et al., 2019). LTC networks are already structured in the relevant sense, because the recurrent term is conductance-inspired and induces a state-dependent effective time scale (Hasani et al., 2021). Closed-form continuous-time networks approximate that solution analytically (Hasani et al., 2022). In the conductance-style native LTC baseline used here, the state dependence of each interaction is restricted to an affine-sigmoid gate. Related structural work constrains a different level of the system: second-order motion laws in equivariant graph dynamics (Liu et al., 2024) and momentum conservation in graph stochastic dynamics (Bishnoi et al., 2024). Both impose system-level structure, whereas the constraint studied here applies to an individual interaction. KANs supply the edge parameterization (Liu et al., 2025b), and have been used as the backbone of a Neural Ordinary Differential Equation (ODE) for symbolic recovery of dynamical laws (Koenig et al., 2024), with state-space structure for interpretable latent dynamics (Liu et al., 2025a), and in KAN-enhanced liquid networks for battery state-of-health estimation (Ma et al., 2026). These use KAN flexibility at the level of a vector field, a latent state or a predictor; here it is confined to one scalar conductance factor per edge.

The task addressed here is narrower than equation discovery, and the distinction determines which comparisons are informative. Universal physics-informed networks learn an unknown term of a partially specified differential equation from sparse noisy data and can subsequently recover a symbolic form for it (Podina et al., 2023). This work instead assumes the interaction template and estimates the functional law within it. Structure-preserving networks provide a complementary precedent,

making a scientifically privileged dynamical quantity explicit and evaluating it directly rather than only through trajectory error (Matsubara & Yaguchi, 2023). Graph-structured continuous-time models share the constraint of a known topology (Chamberlain et al., 2021; Huang et al., 2021) but parameterize propagation rather than the edge-specific law.

The conductance form of Equation 1 has an established record in biologically structured models. Connectome-constrained latent-variable models of whole-brain *C. elegans* activity found conductance-based synapses to predict held-out activity more accurately than current-based alternatives (Mi et al., 2022), and in a dynamic-clamp preparation reversal-potential and conductance priors improved prediction of synaptic currents withheld from training (Burghi et al., 2025). Concurrent work embeds Neural ODE components inside conductance-based neuron models to recover unknown ionic currents while preserving the remaining structure (Beck et al., 2026), and connectome-constrained linear dynamics explain part of the measured causal interaction structure, with shuffled-connectome controls performing worse (Creamer et al., 2024). What is added here is the isolation of the factorization from the scalar function class at fixed capacity, and evaluation of the resulting decomposition against interventions rather than against fit. A separate line preserves structure at system level through Hamiltonian (Greydanus et al., 2019), Lagrangian (Cranmer et al., 2020; Lutter et al., 2019) and port-Hamiltonian (Desai et al., 2021) parameterizations; the constraint here is local, preserving a conductance and a driving force per interaction without assuming a global energy function.

The identifiability literature supplies the standard adopted here for stating what a learned edge law can mean. Recovery of a latent causal model from interventional data is established only up to an equivalence class and under a stated faithfulness condition (Zhang et al., 2023); identification of a nonlinear transition function rests on sufficient variability and bijectivity assumptions declared in advance (Hızlı et al., 2025); and where identification is unavailable, treating nonidentifiability as the premise rather than as a failure is a productive stance (Pace et al., 2024). A representation aligned with a physically named quantity is not thereby causal, since the alignment may follow an invariance that is not itself causal (Yao et al., 2025). Flexible models admit multiple edge-level decompositions on restricted input support (Kuskova et al., 2026), a failure mode approached here through controlled aliasing of the driving-force geometry. Numerical claims are bounded in the same spirit, since resolution-decoupled parameterizations do not guarantee consistent behavior across discretizations (Gao et al., 2025).

3 SP-KAN-LTC

3.1 Functional-conductance dynamics

Each state evolves under a leak term, a sum of factorized interactions over its known incoming support, and any known external forcing,

$$\dot{h}_i = g_i^L \left(E_i^L - h_i\right) + \sum_{j \in \mathcal{N}(i)} c_{ij}(h_j) \left(E_{ij} - h_i\right) + I_i^{\text{known}}(u,t), \tag{2}$$

where $g_i^L > 0$ is the leak conductance, E_i^L the leak equilibrium, E_{ij} the fixed driving-force reference for interaction $j \to i$, $\mathcal{N}(i)$ the known active incoming support, and $I_i^{\text{known}}(u,t)$ a known external input when present. The recurrent summand is Equation 1: a learned state-dependent interaction strength $c_{ij}(h_j)$ multiplies an explicit driving force $E_{ij} - h_i$ that is not learned.

The learned factor is constrained to be positive and bounded,

$$c_{ij}(h_j) = c_{ij}^{\max}\, \sigma\big(\phi_{ij}^{\text{KAN}}(h_j)\big), \qquad 0 < c_{ij}(h_j) < c_{ij}^{\max}, \tag{3}$$

where ϕ_{ij}^{KAN} is a univariate KAN function of the source state, σ the logistic map, and $c_{ij}^{\max} > 0$ the per-edge conductance scale, which is fixed in the controlled benchmarks and learned within a bounded interval on the measured system. The learned function therefore governs interaction strength and nothing else: it does not directly emit an edge current or vector field, and positivity prevents it from reversing the sign imposed by the driving force. The canonical edge uses cubic splines on eight intervals, giving thirteen trainable parameters per active edge. A sensitivity analysis over grid resolution is reported in Appendix K.

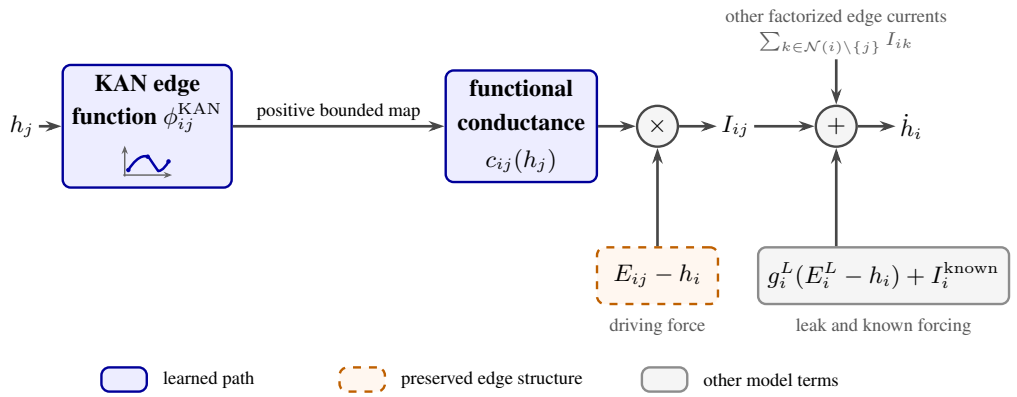


Figure 1: **SP-KAN-LTC architecture: the learned object is the conductance $c_{ij}(h_j)$, not the edge current I_{ij}.** A univariate KAN edge function is composed with a positive bounded map to produce a conductance that lies strictly between zero and a per-edge cap and multiplies the explicit driving force $E_{ij} - h_i$ to form the recurrent edge current. That current, the remaining factorized currents over the known support, and the leak and known forcing determine $\dot{h}_i$.

Positivity and boundedness carry over to the effective time constant, the reciprocal of the summed leak and interaction conductances; this property is inherited from native LTC networks rather than introduced here (Appendix A.2). The consequence for evaluation is that the same learned edge functions determine the interaction currents, the state derivative and the response time scale, so a single functional error propagates to every quantity the evaluation reads.

3.2 Mechanistic quantities

Several quantities are evaluated against ground truth where it exists. The conductance function is compared pointwise on a uniform grid over the observed source-state support, weighting every state equally, while the edge current of Equation 1 is evaluated along held-out trajectories, weighting by the visited states and by the realized driving force. The two disagree whenever approximation error concentrates outside the visited region, so both are reported rather than either being treated as the mechanism. Faithfulness is tested by intervention: an edge or identifiable edge group is removed from the learned and generating systems separately and the two changes in trajectory are compared in magnitude and in ordering, reported as an intervention-effect error with rank agreement. A conductance that looks plausible need not predict the consequence of removing it, and only this comparison tests that.

Normalized errors use benchmark-specific definitions, tabulated in Appendix M, and absolute values are compared only within compatible definitions.

3.3 Matched controls

Four controls separate the structural factorization from the scalar function class. In the controlled factorial comparison, every flexible model carries the same thirteen parameters per active edge. Multilayer Perceptron (MLP)-Conductance keeps Equation 1 intact and replaces ϕ_{ij}^{KAN} with a multilayer perceptron at the same thirteen parameters per active edge, varying the function class alone. Two graph-matched controls vary the factorization alone, one in each function class. Each keeps the known graph and the same per-edge budget, and receives the same two quantities, the source state and the driving force, through a two-input map $\Psi_{ij}(h_j, E_{ij} - h_i)$ that emits a signed edge current directly: Current-KAN parameterizes it with a single-interval, degree-one, width-one KAN composition, and Current-MLP with a one-hidden-layer width-three perceptron, each matching the same thirteen parameters per active edge. Both are therefore given the driving-force information without being required to multiply it by a positive source-dependent conductance, which is the structural difference the comparison is built to test (Appendix A.4). Crossing the two function classes with the presence or absence of the factorization completes a two-by-two design over the flexible models; Native LTC sits outside it as the restricted parent family. Native LTC supplies the parent conduc-

tance family and is not enlarged to match the flexible models, since its restricted gate is the object of comparison; outside the controlled neural setting the same gate is instantiated per system and reported as Sigmoid-Conductance. A near-oracle initialization of it separates optimization sensitivity from representational limitation where the generating law lies inside the native family.

3.4 SCOPE

SP-KAN-LTC preserves interaction structure and does not discover it. The canonical method assumes known graph support, so absent edges are never instantiated; the driving-force references E_{ij} are fixed and known in the mechanism-identification experiments rather than jointly inferred; the edge law is instantaneous in the source state and carries no latent edge-memory variable; and external forcing is treated as known wherever it enters Equation 2 explicitly. Each assumption bounds a different part of the mechanistic claim. We test these boundaries through driving-force misspecification, loss of individual-edge identifiability, hidden interaction memory, dense graph support, independently measured hidden currents, and transfer to held-out interventions. Where no edge law is uniquely observed, as on measured systems, the learned conductances are effective model interactions whose identification with literal biological conductances does not follow from the fit. The resulting edge decomposition is treated as a mechanistic hypothesis whose validity has to be established for the setting in which it is used.

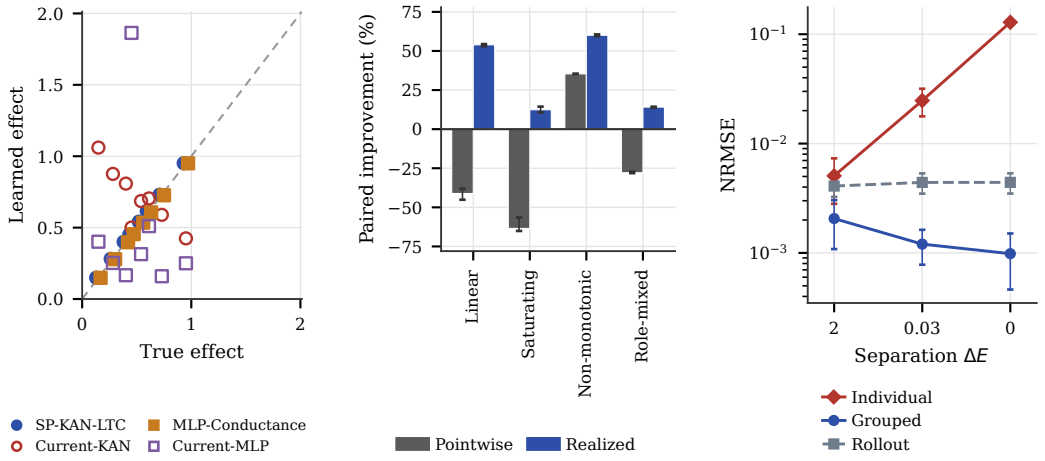
4 EVALUATION STRATEGY

The benchmarks trade access to ground truth against realism, and each gives up a different guarantee. The controlled benchmark supplies known interaction laws, a known graph and known driving forces, so every quantity in the readable chain has a target; four studies then remove one requirement at a time, taking away the separation between reversal references, the correctness of those references, clean observations, and the instantaneous edge assumption, the last in a Morris–Lecar system (Morris & Lecar, 1981). A sparsity ablation removes the known support itself. The empirical-topology benchmark keeps the mechanism known but simulates it on a measured hospital-ward contact network (Vanhems et al., 2013), at a scale where optimization becomes a factor in its own right. The measured half-center oscillator supplies an electrophysiological system whose interaction current is recorded independently and withheld from training and model selection; models are evaluated by transfer between two recorded trials in both directions. The *C. elegans* protocol supplies no observed edge law at all and asks whether a model fitted to observational calcium dynamics transfers to a held-out optogenetic intervention atlas (Randi et al., 2023), with neuron identities from a whole-animal atlas (Yemini et al., 2021) and synaptic polarity fixed before evaluation from molecular priors (Fenyves et al., 2020) and a train-only fallback.

The training seed is the unit of inference throughout. Models sharing a condition are compared seed by seed, and no comparison treats time points, trajectories or edges as independent replicates. Condition-level results are reported as mean $\pm$ standard deviation over seeds; matched contrasts are reported as seed-wise paired effect sizes with bootstrap confidence intervals and, where the design supports them, exact paired tests. Where a condition repeats across interaction-law regimes, the paired effect is aggregated within each seed before being summarized across seeds, so that the seed remains the unit rather than the regime. Directional win counts are descriptive only. Model selection uses validation data, and the reported evaluation uses held-out data. Normalized errors use denominators that differ across benchmarks, so magnitudes are compared only within a benchmark; the definitions are tabulated in Appendix M and the protocol for each benchmark in Appendix L. Numerical consistency of the controlled benchmark is assessed under a refined step size and an adaptive reference solver, confirming that the results do not depend on the integrator over the tested range and that the integrated trajectories respect the state bounds the continuous dynamics imply (Appendices A.3 and J).

Table 1: Controlled architectural isolation at eight active edges under mixed-shape laws, in a design that crosses factorization with edge function class. The four flexible models share the graph, the per-edge budget and the same two edge-level quantities; Native LTC is the restricted parent family and is not capacity-matched. Mean ± standard deviation over ten seeds. Conductance is undefined for the unfactorized models, and the intervention-effect error of Current-MLP is heavy-tailed, with a median of 0.760. Every factorial condition appears in Appendix B.

Model	Conductance	Rollout	Edge current	Interv. effect	Interv. ρ
SP-KAN-LTC	$\mathbf{0.0096 \pm 0.0014}$	$\mathbf{0.0028 \pm 0.0005}$	$\mathbf{0.0064 \pm 0.0010}$	$\mathbf{0.0036 \pm 0.0009}$	$\mathbf{1.00 \pm 0.00}$
MLP-Conductance	0.046 ± 0.043	0.011 ± 0.010	0.018 ± 0.016	0.012 ± 0.013	$\mathbf{1.00 \pm 0.00}$
Native LTC	0.300 ± 0.063	0.123 ± 0.033	0.207 ± 0.054	0.134 ± 0.045	0.988 ± 0.013
Current-MLP	N/A	0.019 ± 0.006	1.30 ± 0.23	1.74 ± 3.13	-0.307 ± 0.328
Current-KAN	N/A	0.032 ± 0.005	2.83 ± 0.29	0.873 ± 0.129	-0.700 ± 0.288



(a) Intervention fidelity. (b) Pointwise against realized. (c) Identifiability boundary.

Figure 2: Structural fidelity, conditional function-class effects, and the identifiability boundary. (a) Learned against true edge-removal effect in the controlled eight-edge benchmark, one marker per edge averaged over ten seeds. Shape denotes edge function class and fill denotes factorization; the dashed line is the identity relation. The two factorized models are offset horizontally by 0.02 in opposite directions, since their markers would otherwise coincide. (b) Seed-paired improvement of SP-KAN-LTC over MLP-Conductance on the 75-node contact topology, with bootstrap intervals; positive values favor SP-KAN-LTC. (c) At the three tested separations $\Delta E \in \{2, 0.03, 0\}$, individual conductance recovery fails as the references alias while the grouped current and the trajectory remain accurate.

5 Results: Structure, Flexibility, and Mechanistic Validity

5.1 Factorization governs mechanism fidelity across matched function classes

Removing the factorization degrades prediction and damages edge-level attribution far more severely. Table 1 reports the eight-edge mixed-shape condition, in which all four flexible models share the graph and the thirteen parameters per active edge; the separation between factorized and unfactorized models persists across all six conditions (Appendix B). Rollout error rises from 0.00280 ± 0.00054 to 0.0321 ± 0.0048 when the factorization is removed, a factor of eleven. Edge-current error rises from 0.00643 ± 0.00104 to 2.83 ± 0.29 and the correlation between learned and true intervention rankings moves from 1.00 ± 0.00 to -0.700 ± 0.288. The degradation in rollout is modest relative to the failure in edge-current and intervention recovery: seed-wise paired reductions are 99.8% and 99.5% on those two quantities against 90.9% on rollout.

An edge score read from the model itself would not expose this, since it is computed from the same decomposition that is wrong; only a comparison against an external intervention target does.

The same pattern holds in the other function class, separating a factorization effect from a function-class effect. Current-MLP predicts better than Current-KAN, at 0.019 ± 0.006 against 0.032 ± 0.005, and also improves several attribution quantities, but remains far from either factorized model: its edge-current error is 1.30 ± 0.23 against 0.018 ± 0.016 for the matched factorized perceptron, and its intervention rank correlation stays negative at -0.307 ± 0.328. Within the perceptron family, the paired improvements of the factorized model over its unfactorized counterpart are 60.7% on rollout, 99.1% on edge current and 98.9% on intervention effect, against 90.9%, 99.8% and 99.5% within the spline family (Appendix B). Removing the factorization therefore costs mechanism recovery almost completely in both function classes, and changing the edge function class within the unfactorized family does not recover it. At empirical-network scale, Current-KAN also fails to reach a useful predictive fit, staying between 0.94 and 1.98 in test rollout error across all six conditions; a ten-rate search does not remove that failure, so the difficulty is a property of the tested architecture over the searched range rather than of one learning-rate choice (Appendix D).

Two controls probe optimization and graph support within the factorized family. Where the generating law lies inside the native family, near-oracle initialization reduces the median conductance error of Native LTC to 0.000601 against 0.0596, winning ten of ten paired seeds on all five metrics, so the binding constraint there is optimization rather than the restricted gate (Appendix F). Replacing the known sparse support with a dense graph degrades mechanism recovery in every regime. Under mixed-shape laws, sparse support reduces conductance and intervention-effect error by 94.5% and 99.1% relative to dense support, while its effect on prediction is regime-dependent: rollout improves under sparse support where the generating law is native or monotonic and shows no reliable difference under mixed-shape laws (-3.3%), so predictive accuracy does not distinguish the supports in the regime where mechanism recovery matters most (Appendix E).

5.2 KAN GAINS ARE NOT EXPLAINED BY LAW COMPLEXITY ALONE

In the controlled benchmark, the contribution of the KAN edge varies with interaction count and is not explained by interaction-law complexity alone. Aggregating the paired effect within each seed across the three law regimes, the median paired improvement in conductance error for SP-KAN-LTC over capacity-matched MLP-Conductance is -63.9% at four active edges (95% bootstrap CI $[-92.6, +10.2]$) and $+59.8\%$ at eight (95% CI $[+22.1, +66.5]$), with positive values favoring SP-KAN-LTC. Intervention-effect error behaves the same way, -12.2% (95% CI $[-139.4, +80.1]$) against $+65.9\%$ (95% CI $[+29.2, +86.3]$). The four-edge interval spans zero on both quantities, so no consistent class advantage is established at that count; the eight-edge interval excludes it. Within-seed summaries are used because the error distributions are skewed (Appendix M).

The empirical-topology results do not support a simple law-complexity explanation, under which additional edge flexibility would matter for rich or non-monotonic laws and be unnecessary under a linear one. Under the linear law on the 75-node contact topology, the median paired improvement in intervention error for SP-KAN-LTC over MLP-Conductance is $+85.0\%$ (95% CI $[+82.4, +86.9]$), which is the largest of the six conditions rather than the smallest. The advantage is also specific to mechanism recovery: under the same law the restricted-gate Sigmoid-Conductance model attains the lowest test rollout error of the four while its intervention error remains more than four times that of SP-KAN-LTC.

The same conditions separate two quantities that are often treated as one. Pointwise conductance error, measured on a uniform grid, favors MLP-Conductance in most interaction-law conditions, while dynamically realized edge-flow error, measured along visited trajectories, favors SP-KAN-LTC; the non-monotonic law is the single condition in which both favor SP-KAN-LTC. Under the linear law, the two run in opposite directions by a wide margin, -41.1% (95% CI $[-45.1, -38.1]$) on pointwise conductance against $+53.9\%$ (95% CI $[+53.0, +54.4]$) on realized edge flow. A model can therefore fit the conductance function less accurately where the metric samples uniformly and still reproduce the interaction current more accurately where the dynamics place state-space mass. This is consistent with the local spline basis concentrating approximation capacity on the visited region, though the experiments do not isolate the basis as the cause, and it means that a pointwise functional metric alone does not determine which model recovers the realized interaction.

Local degree does not reproduce the interaction-count effect: a separate comparison at degree four and degree eight leaves intervention error favoring MLP-Conductance at both, so system-wide interaction count and local node degree act differently and no degree threshold is claimed (Appendix D).

The measured half-center oscillator carries two interactions and shows no consistent class advantage either, as Section 5.4 reports. The flexible edge is therefore not a general improvement over a capacity-matched alternative.

5.3 Mechanistic interpretation is conditional on identifiability and structural correctness

Individual-edge recovery fails before prediction does when two interactions become indistinguishable. A pair of edges shares a source trajectory and their reversal references are brought together, so the separation ΔE controls how far the two can be told apart. At $\Delta E = 2$, the design is well conditioned, with a minimum singular value of 0.087, and SP-KAN-LTC recovers the focus pair to a conductance error of 0.00507 ± 0.00226. At exact aliasing, the minimum singular value becomes numerically zero and the focus-pair conductance error rises to 0.1285 ± 0.0045, an increase of about 2400%. Over the same sweep, rollout error changes by 8%, while grouped-current error does not degrade. At exact aliasing, grouped-current error is 0.00098 ± 0.00052 and rollout error is 0.00442 ± 0.00093. The observations therefore continue to determine the trajectory and the grouped interaction while ceasing to determine their decomposition. Appendix A.5 derives this degeneracy, which makes the level of aggregation at which an explanation is offered part of the claim rather than a presentational choice.

An incorrect driving force biases the mechanism well before it degrades the fit. Holding the generating system fixed and shifting every reversal reference by a fraction α of the state range, SP-KAN-LTC conductance error rises from 0.00963 ± 0.00140 at $\alpha = 0$ to 0.1056 ± 0.0017 at $\alpha = +0.05$ and 0.2127 ± 0.0047 at $\alpha = +0.10$, while rollout error rises only from 0.00280 ± 0.00054 to 0.0128 ± 0.0035. Relative to the unperturbed anchor, the largest shift raises conductance error by 2100% and rollout error by 360%, so the relative increase in conductance error is about six times larger than the relative increase in rollout error. The effect is close to symmetric in sign, 0.2111 ± 0.0044 at $\alpha = -0.10$, and MLP-Conductance behaves the same way, so it follows from the assumed interaction form rather than from the edge parameterization. Corrupting the training observations produces a weaker version of the same pattern, in which the clean-setting advantage of SP-KAN-LTC over MLP-Conductance is no longer systematic at the largest noise level tested (Appendix G).

The instantaneous edge law has a measurable operating envelope, and the experiments bracket its boundary rather than locating it. In the Morris–Lecar Hidden-Memory Stress Test, the true interaction acquires a latent time constant, with η the ratio of that constant to the state time scale. At $\eta = 0$, the assumption holds exactly and SP-KAN-LTC is the best of the four models, reaching an interaction-current error of 0.00651 ± 0.00161 against 0.0236 ± 0.0097 for MLP-Conductance, 0.161 ± 0.039 for the history-capable Closed-Form Continuous-Time Network (CfC), and 0.281 ± 0.044 for the augmented neural ODE, the last over nine seeds rather than ten. At $\eta = 0.10$, the smallest nonzero mismatch tested, the ordering has already reversed: CfC reaches 0.129 ± 0.025 on interaction current against 0.227 ± 0.010 for SP-KAN-LTC, with the same ordering on rollout and on intervention effect. Measured against CfC, SP-KAN-LTC moves from 96% lower interaction-current error at $\eta = 0$ to 76% higher at $\eta = 0.10$. Because the interval between the two levels was not sampled, the crossing point is not determined; what the experiment establishes is that the instantaneous edge law is inside its favorable regime at $\eta = 0$ and outside it by $\eta = 0.10$.

The three studies expose different failure modes rather than one pattern. Aliasing destroys individual-edge recovery while the grouped interaction and the trajectory stay accurate; misspecified driving forces degrade mechanism far more than prediction; hidden memory places the interaction outside the model class and degrades both. What the three share is that in none of them does predictive fit establish that an individual edge law is identifiable or structurally valid.

5.4 Measured systems and external transfer

On measured electrophysiology, the separation between factorized and unfactorized models holds in both tested function classes, while the comparison between the two flexible factorized models remains direction dependent. The synaptic current between the two neurons is recorded independently and withheld from training and model selection, so the mechanism target is observed rather than constructed. Averaged over the two transfer directions, hidden-current error is 0.4621 ± 0.0276 for MLP-Conductance, 0.4977 ± 0.0276 for SP-KAN-LTC and 0.5712 ± 0.0063 for Sigmoid-Conductance, against 1.1896 ± 0.0756 for Current-MLP and 1.3279 ± 0.1054 for Current-KAN. Replacing the unfactorized spline edge with a perceptron raises waveform correlation from -0.62 to $+0.44$ but leaves it far from every factorized model, so the factorization gap is not specific to the spline parameterization. Every factorized run recovers the current more accurately than every unfactorized run: the smallest unfactorized error is 1.053, 75% higher than the largest factorized error of 0.601. Voltage rollout error is tightly clustered across the five, from 0.057 to 0.061, while hidden-current error ranges from 0.462 to 1.328; comparable voltage fit therefore does not discriminate between decompositions whose hidden-current errors differ nearly threefold.

Between the two flexible factorized models, the result depends on the direction of transfer. Training on Trial 3 and testing on Trial 5 gives hidden-current errors of 0.456 ± 0.048 for SP-KAN-LTC and 0.456 ± 0.018 for MLP-Conductance, a paired difference of -0.0006 ± 0.0479 that is not distinguishable from zero under the exact sign-flip test. Training on Trial 5 and testing on Trial 3 gives 0.540 ± 0.024 against 0.468 ± 0.048, a paired difference of $+0.072 \pm 0.057$, corresponding to 15% higher hidden-current error for SP-KAN-LTC (exact sign-flip $p = 0.008$); eight of ten seeds favor MLP-Conductance. Waveform agreement remains high for both models in both directions, between 0.885 and 0.976. The absence of a consistent advantage here is compatible with the controlled low-interaction result, though the measured system differs in several other respects and the agreement is not evidence that interaction count explains it. The asymmetry between the two directions is examined in Appendix H through exploratory support diagnostics, which are descriptive and do not establish its cause.

The external transfer test returns a null result. The *C. elegans* protocol supplies no observed edge law and asks whether models fitted to observational calcium dynamics reproduce the effects recorded in a held-out optogenetic intervention atlas. Of six conditions trained, Current-KAN is numerically ineligible for the 30-second endpoint: it fails a numerical-integrity gate that uses no biological score, at every tested solver refinement, so no transfer value for it is reported. Among the five eligible conditions, target-wise rank agreement with the atlas lies between $+0.012 \pm 0.026$ and $+0.045 \pm 0.016$, and top-five overlap spans 0.236 to 0.287 against a random-set expectation of 0.267 computed over the evaluable candidate sets; SP-KAN-LTC has the lowest mean, at 0.236. No condition demonstrates reliable transfer, and SP-KAN-LTC on the true connectome does not demonstrate an advantage over the same model on a shuffled graph (Appendix I).

We therefore use this experiment as a bound on external validity rather than as evidence about mechanism recovery. The protocol asks a learned decomposition to transfer across a change of intervention modality in a system where the generating edge law is unknown, the excitation is uncontrolled and the connectome is an estimate, so a failure does not separate a limitation of the model from one of the protocol.

6 Limitations and Conclusion

SP-KAN-LTC assumes known graph support, correct driving-force references and an instantaneous edge law; the controlled studies also assume a conductance cap containing the generating law, and external transfer was not demonstrated on *C. elegans*. When driving-force signatures alias, individual edges become unidentifiable while prediction and the grouped interaction remain accurate, making aggregation level part of the claim. Misspecified references increase conductance error by about six times the relative increase in rollout error, so fit can substantially understate mechanism error. Hidden interaction memory moves the generating system outside the model class; the unsampled interval between mismatch levels brackets rather than locates that boundary. The function-class comparison is conditional: no consistent advantage is established at four controlled interactions or on any measured system, and law complexity alone does not explain where it helps. An edge decomposition therefore remains a hypothesis resting on identifiability and interaction form, not fit.

AI USE STATEMENT

Generative AI tools were used for grammar editing and feedback on the writing and presentation of the manuscript. All AI-assisted text was reviewed by the authors, who take responsibility for the final content of the paper.

REFERENCES

Jonas Beck, Michael Deistler, Dóra Viktória Molnár, Jakob H. Macke, and Philipp Berens. Learning hybrid biophysical neuron models with neural ODEs. *arXiv preprint arXiv:2606.16693*, 2026.

Suresh Bishnoi, Jayadeva, Sayan Ranu, and N. M. Anoop Krishnan. BroGNet: Momentum-conserving graph neural stochastic differential equation for learning brownian dynamics. In *International Conference on Learning Representations*, 2024. arXiv:2306.11435.

Thiago B. Burghi, Kyra Schapiro, Maria Ivanova, Huaxinyu Wang, Eve Marder, and Timothy O'Leary. Quantitative prediction of intracellular dynamics and synaptic currents in a small neural circuit. *Frontiers in Computational Neuroscience*, 19:1515194, 2025. doi: 10.3389/fncom.2025.1515194.

Benjamin P. Chamberlain, James Rowbottom, Maria Gorinova, Stefan Webb, Emanuele Rossi, and Michael M. Bronstein. GRAND: Graph neural diffusion. In *Proceedings of the 38th International Conference on Machine Learning (ICML)*, volume 139 of *Proceedings of Machine Learning Research*, pp. 1407–1418, 2021.

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.

Miles Cranmer, Sam Greydanus, Stephan Hoyer, Peter Battaglia, David Spergel, and Shirley Ho. Lagrangian neural networks. In *ICLR 2020 Workshop on Integration of Deep Neural Models and Differential Equations*, 2020.

Matthew S. Creamer, Andrew M. Leifer, and Jonathan W. Pillow. Bridging the gap between the connectome and whole-brain activity in C. elegans. *bioRxiv 2024.09.22.614271*, 2024. doi: 10.1101/2024.09.22.614271.

Shaan A. Desai, Marios Mattheakis, David Sondak, Pavlos Protopapas, and Stephen J. Roberts. Port-Hamiltonian neural networks for learning explicit time-dependent dynamical systems. *Physical Review E*, 104(3):034312, 2021. doi: 10.1103/PhysRevE.104.034312.

Emilien Dupont, Arnaud Doucet, and Yee Whye Teh. Augmented neural ODEs. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Bánk G. Fenyves, Gábor S. Szilágyi, Zsolt Vassy, Csaba Sőti, and Peter Csermely. Synaptic polarity and sign-balance prediction using gene expression data in the Caenorhabditis elegans chemical synapse neuronal connectome network. *PLOS Computational Biology*, 16(12):e1007974, 2020. doi: 10.1371/journal.pcbi.1007974.

Wenhan Gao, Ruichen Xu, Yuefan Deng, and Yi Liu. Discretization-invariance? On the discretization mismatch errors in neural operators. In *International Conference on Learning Representations*, 2025.

Samuel Greydanus, Misko Dzamba, and Jason Yosinski. Hamiltonian neural networks. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Ramin Hasani, Mathias Lechner, Alexander Amini, Daniela Rus, and Radu Grosu. Liquid time-constant networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 7657–7666, 2021. doi: 10.1609/aaai.v35i9.16936.

Ramin Hasani, Mathias Lechner, Alexander Amini, Lucas Liebenwein, Aaron Ray, Max Tschaikowski, Gerald Teschl, and Daniela Rus. Closed-form continuous-time neural networks. *Nature Machine Intelligence*, 4(11):992–1003, 2022. doi: 10.1038/s42256-022-00556-7.

Çağlar Hızlı, Çağatay Yıldız, Matthias Bethge, S. T. John, and Pekka Marttinen. Identifying latent state transitions in non-linear dynamical systems. In *International Conference on Learning Representations*, 2025. arXiv:2406.03337.

Zijie Huang, Yizhou Sun, and Wei Wang. Coupled graph ODE for learning interacting system dynamics. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, 2021.

Benjamin C. Koenig, Suyong Kim, and Sili Deng. KAN-ODEs: Kolmogorov–arnold network ordinary differential equations for learning dynamical systems and hidden physics. *Computer Methods in Applied Mechanics and Engineering*, 432:117397, 2024. doi: 10.1016/j.cma.2024.117397.

Valentina V. Kuskova, Dmitry Zaytsev, and Michael Coppedge. When are neural interaction discoveries real? identifiability, recoverability, and a pre-fit diagnostic. *arXiv preprint arXiv:2606.08390*, 2026.

Wei Liu, Kiran Bacsa, Loon Ching Tang, and Eleni Chatzi. Structured kolmogorov–arnold neural ODEs for interpretable learning and symbolic discovery of nonlinear dynamics. *arXiv preprint arXiv:2506.18339*, 2025a.

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. arXiv:2308.13212.

Ziming Liu, Yixuan Wang, Sachin Vaidya, Fabian Ruehle, James Halverson, Marin Soljačić, Thomas Y. Hou, and Max Tegmark. KAN: Kolmogorov–arnold networks. In *International Conference on Learning Representations (ICLR)*, 2025b.

Michael Lutter, Christian Ritter, and Jan Peters. Deep lagrangian networks: Using physics as model prior for deep learning. In *International Conference on Learning Representations (ICLR)*, 2019.

Wentao Ma, Zhuo Li, Peng Guo, Yang Li, and Badong Chen. Lithium-ion battery SOH estimation via mixture correntropy loss-based PCA and kolmogorov–arnold-enhanced liquid neural networks. *IEEE Transactions on Industrial Electronics*, pp. 1–13, 2026.

Takashi Matsubara and Takaharu Yaguchi. FINDE: Neural differential equations for finding and preserving invariant quantities. In *International Conference on Learning Representations*, 2023. arXiv:2210.00272.

Lu Mi, Richard Xu, Sridhama Prakhya, Albert Lin, Nir Shavit, Aravinthan D. T. Samuel, and Srinivas C. Turaga. Connectome-constrained latent variable models of whole-brain neural activity. In *International Conference on Learning Representations (ICLR)*, 2022.

Catherine Morris and Harold Lecar. Voltage oscillations in the barnacle giant muscle fiber. *Biophysical Journal*, 35(1):193–213, 1981. doi: 10.1016/S0006-3495(81)84782-0.

Alizée Pace, Hugo Yèche, Bernhard Schölkopf, Gunnar Rätsch, and Guy Tennenholtz. Delphic offline reinforcement learning under nonidentifiable hidden confounding. In *International Conference on Learning Representations*, 2024.

Lena Podina, Brydon Eastman, and Mohammad Kohandel. Universal physics-informed neural networks: Symbolic differential operator discovery with sparse data. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 27948–27956. PMLR, 2023.

Francesco Randi, Anuj K. Sharma, Sophie Dvali, and Andrew M. Leifer. Neural signal propagation atlas of Caenorhabditis elegans. *Nature*, 623:406–414, 2023. doi: 10.1038/s41586-023-06683-4.

Philippe Vanhems, Alain Barrat, Ciro Cattuto, Jean-François Pinton, Nagham Khanafer, Corinne Régis, Byeul-a Kim, Brigitte Comte, and Nicolas Voirin. Estimating potential infection transmission routes in hospital wards using wearable proximity sensors. *PLOS ONE*, 8(9):e73970, 2013. doi: 10.1371/journal.pone.0073970.

Dingling Yao, Dario Rancati, Riccardo Cadei, Marco Fumero, and Francesco Locatello. Unifying causal representation learning with the invariance principle. In *International Conference on Learning Representations*, 2025. arXiv:2409.02772.

Eviatar Yemini, Albert Lin, Amin Nejatbakhsh, Erdem Varol, Ruoxi Sun, Gonzalo E. Mena, Aravinthan D. T. Samuel, Liam Paninski, Vivek Venkatachalam, and Oliver Hobert. NeuroPAL: A multicolor atlas for whole-brain neuronal identification in C. elegans. *Cell*, 184(1):272–288.e11, 2021. doi: 10.1016/j.cell.2020.12.012.

Jiaqi Zhang, Kristjan Greenewald, Chandler Squires, Akash Srivastava, Karthikeyan Shanmugam, and Caroline Uhler. Identifiability guarantees for causal disentanglement from soft interventions. In *Advances in Neural Information Processing Systems*, volume 36, 2023. arXiv:2307.06250.

A Architecture and structural properties

A.1 From the native conductance gate to a learned one

Under the normalization used here, in which the membrane capacitance is absorbed so that the coefficient on $\dot{h}_i$ is unity, the conductance form of the dynamics is Equation 2. The parent LTC family and SP-KAN-LTC differ in one place only, the functional form of the edge conductance. The native edge uses an affine-sigmoid gate,

$$g_{ij}^{\text{native}}(h_j) = c_{ij}^{\max}\,\sigma(a_{ij}(h_j - b_{ij}))\,, \qquad I_{ij}^{\text{native}} = c_{ij}^{\max}\,\sigma(a_{ij}(h_j - b_{ij}))\,(E_{ij} - h_i), \quad (4)$$

with two learned scalars per edge, a slope a_{ij} and a center b_{ij}, since the conductance cap $c_{ij}^{\max}$ is fixed rather than learned, as it is for every factorized model in the controlled comparison. Both gates receive the source state after a fixed per-node normalization, which is a property of the input pipeline rather than of either edge parameterization and is therefore suppressed in the equations. The native gate is monotone in h_j, with its shape fixed up to a location and a slope. Replacing that gate with a univariate KAN composed with the same bounded positive map gives Equation 3 and

$$I_{ij} = c_{ij}^{\max}\,\sigma\big(\phi_{ij}^{\text{KAN}}(h_j)\big)\,(E_{ij} - h_i). \quad (5)$$

Comparing Equations 4 and 5, the driving force $E_{ij} - h_i$, the graph support, the additive decomposition of $\dot{h}_i$ and the sign structure of each interaction are identical. What changes is that $\sigma(a_{ij}(h_j - b_{ij}))$, monotone by construction and carrying two learned scalars, becomes $\sigma(\phi_{ij}^{\text{KAN}}(h_j))$, which carries thirteen and can be non-monotone in h_j while remaining a conductance. Relative to the native LTC parent, this is the only architectural substitution; the factorization itself is preserved rather than introduced, and is isolated separately against the unfactorized controls.

A.2 Positivity and bounded conductance

Because $0 < \sigma(x) < 1$ for all finite x, Equation 3 gives

$$0 < c_{ij}(h_j) < c_{ij}^{\max} \quad (6)$$

for every state, with no constraint on the sign or shape of ϕ_{ij}^{KAN}. Two consequences follow. The learned function cannot reverse the sign that the driving force assigns to an interaction, so an edge that is excitatory under the generating system cannot be represented as inhibitory by making the conductance negative. And the effective time constant, the reciprocal of $g_i^L + \sum_j c_{ij}(h_j)$, is bounded above by $1/g_i^L$ and below by the value obtained at the conductance caps. Both properties are inherited from the conductance form and hold for the native gate as well; neither is introduced by the KAN parameterization.

A.3 Forward invariance of the state interval

Suppose the external forcing vanishes and every equilibrium and reversal reference lies in a common interval, $E_i^L, E_{ij} \in [h_{\min}, h_{\max}]$ for all i and j. At $h_i = h_{\max}$, every driving-force term in Equation 2 satisfies $E_{ij} - h_i \le 0$ and $E_i^L - h_i \le 0$, and since all conductances are positive by

Equation 6, $\dot{h}_i \leq 0$. At $h_i = h_{\min}$, the same argument with reversed inequalities gives $\dot{h}_i \geq 0$. The box $[h_{\min}, h_{\max}]^N$ is therefore forward invariant for the continuous dynamics.

Two qualifications bound this. The argument requires the vanishing-forcing hypothesis; with $I_i^{\text{known}} \neq 0$ the same conclusion holds only under additional bounds on the forcing, which are not assumed here. More importantly, forward invariance of the continuous system does not imply that an explicit numerical discretization preserves the same bounds, since a fixed step can overshoot a boundary that the continuous flow approaches asymptotically. The absence of bound violations reported in Appendix J is a property of the integrated trajectories at the tested step sizes, not a consequence of the proposition above.

A.4 Factorized and unfactorized interaction models

The graph-matched controls are built to test the consequence of removing the factorization while preserving the graph, the parameter budget and the edge-level information, which is easiest to state side by side. SP-KAN-LTC represents each interaction as

$$I_{ij} = \underbrace{c_{ij}^{\max}\,\sigma\big(\phi_{ij}^{\text{KAN}}(h_j)\big)}_{\text{learned conductance}} \cdot \underbrace{(E_{ij} - h_i)}_{\text{fixed driving force}}, \tag{7}$$

while the unfactorized controls represent it as

$$I_{ij} = \Psi_{ij}(h_j,\ E_{ij} - h_i)\,, \tag{8}$$

a two-input map emitting a signed edge current directly, parameterized by a width-one KAN composition in Current-KAN and by a one-hidden-layer width-three perceptron in Current-MLP. Equations 7 and 8 give the controls the same graph, the same per-edge parameter budget and the same two quantities, including the driving force. What they are not given is the requirement that the driving force enter multiplicatively, scaled by a positive function of the source state alone. The comparison in Table 1 therefore does not confound the factorization with access to the driving-force information, while matching the graph and the parameter count. Matching the parameter count does not make the two classes representationally identical, and because the controlled generators use the factorized form, this comparison measures the value of imposing that structure at fixed graph, information and parameter count rather than comparing representationally equivalent classes.

A.5 Identifiability under coincident driving-force references

Consider a state h_i receiving two interactions whose conductances depend on the same source state h. The incoming current is

$$I_i = c_{i1}(h)\,(E_{i1} - h_i) + c_{i2}(h)\,(E_{i2} - h_i). \tag{9}$$

When the two reversal references coincide at $E_{i1} = E_{i2} = E$, this collapses to

$$I_i = [c_{i1}(h) + c_{i2}(h)]\,(E - h_i), \tag{10}$$

so the trajectory depends on the two conductances only through their sum. Comparing Equations 9 and 10, any pair $(c_{i1} + \delta,\ c_{i2} - \delta)$ with δ chosen so that both remain in $(0, c^{\max})$ produces identical dynamics, and no amount of trajectory data separates them. The ambiguity preserves the summed conductance, which is the only combination entering the dynamics at exact aliasing; in this benchmark the grouped interaction current is recovered accurately while the individual conductances are not.

The reported conditioning statistic is built from the same edge features the canonical model uses. For each incoming edge, the spline basis, the normalized source term and a bias are evaluated over the training states and multiplied by that edge's driving force $E_{ij} - h_i$. These features carry benign redundancy within a single edge, so each block is first reduced by singular value decomposition to an orthonormal basis for its observed column space; only those edge subspaces are concatenated across edges sharing a receiving state. The quoted statistic is the smallest singular value across those receiver-wise matrices, with the condition number computed on the same matrices. A value near zero therefore indicates overlap between the subspaces of different edges rather than redundancy inside one edge, which is the algebraic content of Equation 10.

Table 2: Controlled benchmark, two-by-two design over the flexible models, all conditions. Native LTC is the restricted parent family and sits outside this design. Mean ± standard deviation over ten training seeds. Because the intervention-effect error of Current-MLP is heavy-tailed, intervention medians are reported for every model alongside the mean ± standard deviation.

Condition	Model	Rollout	Edge current	Interv. effect	Interv. median	Interv. ρ
4 edges, Mixed-monotonic	Current-KAN	0.0109 ± 0.0035	2.05 ± 0.44	0.550 ± 0.132	0.536	-0.680 ± 0.301
4 edges, Mixed-monotonic	Current-MLP	0.0187 ± 0.0109	0.819 ± 0.176	1.84 ± 3.52	0.7562	0.120 ± 0.492
4 edges, Mixed-monotonic	MLP-Conductance	0.00456 ± 0.00789	0.00576 ± 0.01042	0.00255 ± 0.00566	0.0005362	1.00 ± 0.00
4 edges, Mixed-monotonic	SP-KAN-LTC	$(5.56 \pm 4.40) \times 10^{-4}$	0.00104 ± 0.00073	$(2.44 \pm 3.45) \times 10^{-4}$	0.000117	1.00 ± 0.00
4 edges, Mixed-shape	Current-KAN	0.0301 ± 0.0100	2.06 ± 0.37	0.775 ± 0.166	0.7367	-0.680 ± 0.301
4 edges, Mixed-shape	Current-MLP	0.0246 ± 0.0256	1.29 ± 0.46	0.810 ± 0.156	0.7432	-0.180 ± 0.561
4 edges, Mixed-shape	MLP-Conductance	0.00683 ± 0.01287	0.0206 ± 0.0510	0.00985 ± 0.02823	0.0008454	0.980 ± 0.063
4 edges, Mixed-shape	SP-KAN-LTC	0.00308 ± 0.00042	0.00620 ± 0.00102	0.00159 ± 0.00082	0.001332	1.00 ± 0.00
4 edges, Native-exact	Current-KAN	0.0130 ± 0.0058	1.92 ± 0.47	0.567 ± 0.245	0.5728	-0.640 ± 0.263
4 edges, Native-exact	Current-MLP	0.0167 ± 0.0139	0.834 ± 0.169	26.3 ± 80.6	0.8126	0.100 ± 0.483
4 edges, Native-exact	MLP-Conductance	$(6.46 \pm 1.69) \times 10^{-4}$	$(7.54 \pm 2.85) \times 10^{-4}$	$(2.25 \pm 1.23) \times 10^{-4}$	0.0001678	1.00 ± 0.00
4 edges, Native-exact	SP-KAN-LTC	$(6.47 \pm 2.12) \times 10^{-4}$	0.00133 ± 0.00087	$(5.67 \pm 5.90) \times 10^{-4}$	0.0004324	1.00 ± 0.00
8 edges, Mixed-monotonic	Current-KAN	0.0209 ± 0.0049	2.35 ± 0.40	0.792 ± 0.222	0.781	-0.112 ± 0.248
8 edges, Mixed-monotonic	Current-MLP	0.0154 ± 0.0065	1.09 ± 0.11	9645 ± 30202	0.784	-0.188 ± 0.444
8 edges, Mixed-monotonic	MLP-Conductance	0.00772 ± 0.00821	0.00718 ± 0.00873	0.00841 ± 0.01065	0.005883	0.988 ± 0.023
8 edges, Mixed-monotonic	SP-KAN-LTC	0.00109 ± 0.00123	0.00155 ± 0.00122	0.00137 ± 0.00115	0.001056	1.00 ± 0.00
8 edges, Mixed-shape	Current-KAN	0.0321 ± 0.0048	2.83 ± 0.29	0.873 ± 0.129	0.8085	-0.700 ± 0.288
8 edges, Mixed-shape	Current-MLP	0.0190 ± 0.0057	1.30 ± 0.23	1.74 ± 3.13	0.7598	-0.307 ± 0.328
8 edges, Mixed-shape	MLP-Conductance	0.0108 ± 0.0102	0.0176 ± 0.0160	0.0119 ± 0.0126	0.008126	1.00 ± 0.00
8 edges, Mixed-shape	SP-KAN-LTC	0.00280 ± 0.00054	0.00643 ± 0.00104	0.00358 ± 0.00088	0.003823	1.00 ± 0.00
8 edges, Native-exact	Current-KAN	0.0236 ± 0.0060	2.18 ± 0.32	0.741 ± 0.387	0.6344	-0.100 ± 0.352
8 edges, Native-exact	Current-MLP	0.0224 ± 0.0051	1.03 ± 0.14	$(2.50 \pm 6.46) \times 10^5$	2514.0	-0.171 ± 0.359
8 edges, Native-exact	MLP-Conductance	0.00411 ± 0.00325	0.00231 ± 0.00125	0.00385 ± 0.00268	0.003163	1.00 ± 0.00
8 edges, Native-exact	SP-KAN-LTC	0.00146 ± 0.00222	0.00108 ± 0.00079	0.00192 ± 0.00320	0.001081	1.00 ± 0.00

The sweep in Section 5.3 traverses this degeneracy by varying $\Delta E = |E_{i1} - E_{i2}|$. At $\Delta E > 0$, the two driving forces differ and, under the excitation this benchmark provides, the decomposition is empirically determined up to the conditioning quantified by the reported minimum singular value. Distinct reversal references do not by themselves guarantee identifiability without sufficient state-space variation. At $\Delta E = 0$, Equation 10 applies exactly and the decomposition is not determined at all. The measured collapse in individual-edge recovery alongside unchanged grouped-current and rollout error is the empirical counterpart of this algebra, and the architecture cannot repair it, because the information is absent from the observations rather than from the model.

B FACTORIZATION-BY-FUNCTION-CLASS DESIGN

The controlled benchmark crosses factorization with edge function class, giving four flexible-model combinations. This appendix reports the design in full. The unfactorized perceptron edge uses a learning rate of 0.1, selected by validation one-step error over six rates on three tuning seeds, with mechanism targets excluded from tuning.

Within the perceptron family, the paired improvements of the factorized model over its unfactorized counterpart are 60.7% on rollout (95% CI $[37.7, 79.0]$), 99.1% on edge current (95% CI $[97.6, 99.4]$) and 98.9% on intervention effect (95% CI $[97.5, 99.6]$), each over ten paired seeds. Table 2 reports every condition. Crossing the two axes separates a factorization effect from a function-class one. In the eight-edge mixed-shape condition, Current-MLP shows the same mechanism-level failure as Current-KAN while predicting better, at a rollout error of 0.019 ± 0.006 against 0.032 ± 0.005, so neither the spline parameterization nor the optimization behavior of Current-KAN alone accounts for the pattern. It also improves several attribution quantities relative to Current-KAN while remaining far from either factorized model: in the eight-edge mixed-shape condition both unfactorized controls carry negative intervention rank correlations, at -0.307 and -0.700, whereas both factorized models reach 1.00.

The intervention-effect error of Current-MLP is heavy-tailed, because some edge removals send the learned model into an unstable counterfactual rollout. Five of the six conditions have a median between 0.74 and 0.82 despite outliers reaching 1.2×10^1, 2.6×10^2 and 9.6×10^4. The eight-edge native-exact condition is different in kind rather than degree: its median is already 2514 and nine of ten seeds exceed unity, so there the instability is the typical outcome rather than an outlier. This is why the paired median relative improvement, rather than a ratio of means, is the inferential summary used here and in the main text: a single unstable counterfactual dominates a mean but moves a median by little.

Table 3: **Controlled Functional-Interaction Benchmark**: all conditions. Mean $\pm$ standard deviation over training seeds (n per row); the training seed is the statistical unit. NRMSE uses relative-L_2 normalization; magnitudes are not comparable across benchmarks.

Condition	Model	n	Conductance	Vector field	Rollout	Edge current	Interv. effect	Interv. ρ
4 edges, Native-exact	SP-KAN-LTC	10	0.00393 ± 0.00077	0.00355 ± 0.00100	$(6.47 \pm 2.12) \times 10^{-4}$	0.00133 ± 0.00087	$(5.67 \pm 5.90) \times 10^{-4}$	1.00 ± 0.00
	MLP-Conductance	10	0.00199 ± 0.00048	0.00240 ± 0.00065	$(6.46 \pm 1.69) \times 10^{-4}$	$(7.54 \pm 2.85) \times 10^{-4}$	$(2.25 \pm 1.23) \times 10^{-4}$	1.00 ± 0.00
	Native LTC	10	0.00129 ± 0.00049	0.00199 ± 0.00075	$(5.85 \pm 2.10) \times 10^{-4}$	0.00156 ± 0.00060	$(9.17 \pm 3.53) \times 10^{-4}$	1.00 ± 0.00
	Current-KAN	10	N/A	0.0541 ± 0.0222	0.0130 ± 0.0058	1.92 ± 0.47	0.567 ± 0.245	-0.640 ± 0.263
4 edges, Mixed-monotonic	SP-KAN-LTC	10	0.00504 ± 0.00206	0.00304 ± 0.00201	$(5.56 \pm 4.40) \times 10^{-4}$	0.00104 ± 0.00073	$(2.44 \pm 3.45) \times 10^{-4}$	1.00 ± 0.00
	MLP-Conductance	10	0.0117 ± 0.0165	0.0141 ± 0.0234	0.00456 ± 0.00789	0.00576 ± 0.01042	0.00255 ± 0.00566	1.00 ± 0.00
	Native LTC	10	0.0863 ± 0.0461	0.0874 ± 0.0568	0.0293 ± 0.0230	0.0448 ± 0.0413	0.0228 ± 0.0401	1.00 ± 0.00
	Current-KAN	10	N/A	0.0412 ± 0.0131	0.0109 ± 0.0035	2.05 ± 0.44	0.550 ± 0.132	-0.680 ± 0.301
4 edges, Mixed-shape	SP-KAN-LTC	10	0.0108 ± 0.0026	0.0101 ± 0.0013	0.00308 ± 0.00042	0.00620 ± 0.00102	0.00159 ± 0.00082	1.00 ± 0.00
	MLP-Conductance	10	0.0246 ± 0.0454	0.0187 ± 0.0329	0.00683 ± 0.01287	0.0206 ± 0.0510	0.00985 ± 0.02823	0.980 ± 0.063
	Native LTC	10	0.213 ± 0.073	0.177 ± 0.097	0.0665 ± 0.0405	0.148 ± 0.084	0.0801 ± 0.0725	0.960 ± 0.084
	Current-KAN	10	N/A	0.0908 ± 0.0256	0.0301 ± 0.0100	2.06 ± 0.37	0.775 ± 0.166	-0.680 ± 0.301
8 edges, Native-exact	SP-KAN-LTC	10	0.00396 ± 0.00097	0.00564 ± 0.00637	0.00146 ± 0.00222	0.00108 ± 0.00079	0.00192 ± 0.00320	1.00 ± 0.00
	MLP-Conductance	10	0.00543 ± 0.00199	0.00773 ± 0.00468	0.00411 ± 0.00325	0.00231 ± 0.00125	0.00385 ± 0.00268	1.00 ± 0.00
	Native LTC	10	0.0704 ± 0.0795	0.138 ± 0.134	0.0826 ± 0.0812	0.0401 ± 0.0515	0.0624 ± 0.0729	0.952 ± 0.083
	Current-KAN	10	N/A	0.0431 ± 0.0058	0.0236 ± 0.0060	2.18 ± 0.32	0.741 ± 0.387	-0.100 ± 0.352
8 edges, Mixed-monotonic	SP-KAN-LTC	10	0.00494 ± 0.00137	0.00498 ± 0.00482	0.00109 ± 0.00123	0.00155 ± 0.00122	0.00137 ± 0.00115	1.00 ± 0.00
	MLP-Conductance	10	0.0156 ± 0.0112	0.0183 ± 0.0168	0.00772 ± 0.00821	0.00718 ± 0.00873	0.00841 ± 0.01065	0.988 ± 0.023
	Native LTC	10	0.154 ± 0.041	0.177 ± 0.073	0.0899 ± 0.0415	0.0788 ± 0.0241	0.111 ± 0.043	0.921 ± 0.056
	Current-KAN	10	N/A	0.0496 ± 0.0087	0.0209 ± 0.0049	2.35 ± 0.40	0.792 ± 0.222	-0.112 ± 0.248
8 edges, Mixed-shape	SP-KAN-LTC	10	0.00963 ± 0.00140	0.0135 ± 0.0044	0.00280 ± 0.00054	0.00643 ± 0.00104	0.00358 ± 0.00088	1.00 ± 0.00
	MLP-Conductance	10	0.0458 ± 0.0429	0.0421 ± 0.0335	0.0108 ± 0.0102	0.0176 ± 0.0160	0.0119 ± 0.0126	1.00 ± 0.00
	Native LTC	10	0.300 ± 0.063	0.319 ± 0.077	0.123 ± 0.033	0.207 ± 0.054	0.134 ± 0.045	0.988 ± 0.013
	Current-KAN	10	N/A	0.118 ± 0.017	0.0321 ± 0.0048	2.83 ± 0.29	0.873 ± 0.129	-0.700 ± 0.288

A numerical sensitivity check on the representative eight-edge mixed-shape outlier gave intervention effects of 16.95154, 16.95147 and 16.95155 under fixed-step Runge–Kutta, a halved step, and an adaptive reference, differing by less than 10^{-4}. The response is therefore attributable to the learned counterfactual dynamics rather than to numerical integration over the tested settings, and the reported summaries include these outcomes. What the factorization changes is not that a counterfactual current must stay bounded, since $c_{ij}(h_j)(E_{ij} - h_i)$ grows with $|h_i|$ even under a bounded conductance, but that the gain multiplying the driving force is positive, bounded and tied to the driving-force geometry, whereas the unfactorized control learns a signed current map under no such constraint.

C Controlled Functional-Interaction Benchmark, all conditions

Table 3 reports the controlled conditions at full precision, including the vector-field error omitted from the main text; the factorization-by-function-class design is reported in Appendix B. The separation between the two flexible factorized models and Current-KAN persists across all six combinations, and differences between the two flexible models are smaller than the factorization gap and remain condition dependent. Two features of the table are worth separating. Vector-field error, omitted from the main text, tracks rollout error closely across every condition, which is expected since both read the aggregate derivative, and neither separates the models the way the mechanism-facing quantities do. The intervention rank correlation saturates at unity for all three factorized models in most conditions, so it distinguishes factorized from unfactorized architectures but carries no resolution among the factorized ones; the intervention-effect error is the quantity that does. Reporting both is what prevents a ceiling on one metric from being read as equivalence on the other.

D Empirical-Graph SIS Benchmark, all conditions

Table 4 reports all six conditions of the empirical-topology benchmark. The dissociation described in the main text is visible across the interaction-law conditions: pointwise conductance error and dynamically realized edge-flow error do not order the two flexible factorized models the same way, and the density arm behaves differently again, with intervention error favoring the matched multilayer perceptron at both degrees. The unfactorized control is the clearest feature of the table: its test rollout error stays between 0.94 and 1.98 in every condition, at least nine times that of every factorized model and more than forty times in the four full-graph law conditions, and its intervention rank correlation varies widely across conditions rather than degrading consistently. That combination is consistent with an optimization failure over the searched range rather than with a systematic

Table 4: Empirical-Graph SIS Benchmark, all six conditions. Mean ± standard deviation over ten training seeds, with the seed as the statistical unit. The contact topology is measured and the SIS dynamics on it are simulated, so this benchmark is semi-synthetic. Normalization is quantity-specific and fixed per benchmark, so magnitudes are not comparable across benchmarks.

Condition	Model	Test rollout	Interv. effect	Interv. ρ
Linear	SP-KAN-LTC	0.0174 ± 0.0006	0.0323 ± 0.0012	0.9929 ± 0.0115
Linear	MLP-Conductance	0.0140 ± 0.0007	0.2157 ± 0.0354	0.6952 ± 0.1598
Linear	Sigmoid-Conductance	0.00987 ± 0.00037	0.1344 ± 0.0192	0.8976 ± 0.0683
Linear	Current-KAN	0.9392 ± 0.0226	1.645 ± 0.194	0.3833 ± 0.2860
Saturating	SP-KAN-LTC	0.0115 ± 0.0004	0.0145 ± 0.0007	0.9929 ± 0.0115
Saturating	MLP-Conductance	0.0127 ± 0.0010	0.0432 ± 0.0097	0.9452 ± 0.0252
Saturating	Sigmoid-Conductance	0.0267 ± 0.0030	0.0857 ± 0.0214	0.9238 ± 0.0784
Saturating	Current-KAN	1.323 ± 0.050	2.806 ± 0.471	0.3976 ± 0.2731
Non-monotonic	SP-KAN-LTC	0.00615 ± 0.00002	0.0456 ± 0.0014	0.9762 ± 0.0000
Non-monotonic	MLP-Conductance	0.0155 ± 0.0003	0.1833 ± 0.0340	0.4476 ± 0.1392
Non-monotonic	Sigmoid-Conductance	0.0432 ± 0.0020	0.3768 ± 0.0484	0.5238 ± 0.2360
Non-monotonic	Current-KAN	1.982 ± 0.054	2.480 ± 0.288	0.2095 ± 0.2264
Role-mixed	SP-KAN-LTC	0.0120 ± 0.0001	0.1399 ± 0.0016	0.9976 ± 0.0075
Role-mixed	MLP-Conductance	0.0171 ± 0.0002	0.1461 ± 0.0228	0.8810 ± 0.0561
Role-mixed	Sigmoid-Conductance	0.0232 ± 0.0009	0.1855 ± 0.0446	0.8381 ± 0.0984
Role-mixed	Current-KAN	1.156 ± 0.030	2.733 ± 0.366	0.3952 ± 0.2363
Degree 4, role-mixed	SP-KAN-LTC	0.0128 ± 0.0016	0.1431 ± 0.0029	1.000 ± 0.000
Degree 4, role-mixed	MLP-Conductance	0.0240 ± 0.0016	0.1362 ± 0.0235	0.9881 ± 0.0231
Degree 4, role-mixed	Sigmoid-Conductance	0.1198 ± 0.0241	0.2201 ± 0.1328	0.9500 ± 0.0285
Degree 4, role-mixed	Current-KAN	1.084 ± 0.009	1.301 ± 0.050	0.9357 ± 0.0450
Degree 8, role-mixed	SP-KAN-LTC	0.0112 ± 0.0003	0.1558 ± 0.0025	0.9762 ± 0.0000
Degree 8, role-mixed	MLP-Conductance	0.0175 ± 0.0008	0.1301 ± 0.0210	0.9476 ± 0.0151
Degree 8, role-mixed	Sigmoid-Conductance	0.1096 ± 0.0173	0.1845 ± 0.0319	0.9286 ± 0.0275
Degree 8, role-mixed	Current-KAN	1.206 ± 0.038	2.386 ± 0.224	0.4405 ± 0.2309

bias, which is why the result is stated as a property of the tested architecture over the searched range rather than of unfactorized models in general.

E KNOWN-GRAPH SPARSITY ABLATION

Replacing the known sparse support with a dense graph asks what the graph assumption contributes and separates two effects that a single accuracy figure would merge. The dense model has strictly more edges available to fit the same trajectories, so it may distribute fitted interaction across edges that carry none in the generating system while still reproducing the aggregate derivative. Table 5 shows that mechanism-facing quantities degrade sharply under dense support in every regime. The separation from prediction is clearest under mixed-shape laws, where conductance, edge-current and intervention recovery each improve by more than ninety percent under sparse support, in every paired seed, while rollout shows no reliable difference.

The asymmetry follows from where the two kinds of error are measured. Rollout reads the summed derivative, which the dense model can reproduce even when the extra edges carry fitted interaction, whereas the conductance and intervention metrics read the decomposition, which is what those edges make non-unique. This is the same mechanism that the identifiability sweep isolates by a different route: there the observations fail to determine the decomposition because two edges become indistinguishable, here because edges that should carry nothing are free to carry something. In the mixed-shape sparsity condition, as under exact aliasing, predictive accuracy does not distinguish the incorrect decomposition.

The regime dependence of the rollout column is the part that bounds the claim. Under native-exact and mixed-monotonic laws, sparse support improves rollout as well as mechanism recovery, by 56.0% and 43.2% in median paired terms; only under mixed-shape laws does the rollout advantage

Table 5: Known-graph sparsity ablation. Median values over ten paired seeds for known sparse support against a dense graph, within each interaction-law regime. Improvement is the median relative reduction under sparse support; positive values favor it. Improvement is the median of the seed-wise relative reductions and therefore need not equal the relative reduction between the two displayed medians. The p value is from a directional exact sign-flip test.

Regime	Metric	Sparse	Dense	Improvement (%)	p
Native-exact	Conductance	0.004086	0.2388	+98.3	0.000977
Native-exact	Vector field	0.003006	0.0065	+44.1	0.000977
Native-exact	Rollout	0.0006826	0.001845	+56.0	0.000977
Native-exact	Edge current	0.0008912	0.2556	+99.6	0.000977
Native-exact	Intervention effect	0.001081	0.3246	+99.7	0.000977
Mixed-monotonic	Conductance	0.005117	0.226	+97.8	0.000977
Mixed-monotonic	Vector field	0.00325	0.006986	+54.0	0.00488
Mixed-monotonic	Rollout	0.0006832	0.00122	+43.2	0.00684
Mixed-monotonic	Edge current	0.001198	0.2699	+99.6	0.000977
Mixed-monotonic	Intervention effect	0.001056	0.3249	+99.7	0.000977
Mixed-shape	Conductance	0.009342	0.1745	+94.5	0.000977
Mixed-shape	Vector field	0.01223	0.0142	+13.1	0.00391
Mixed-shape	Rollout	0.00281	0.002647	-3.3	0.519
Mixed-shape	Edge current	0.006636	0.2172	+96.9	0.000977
Mixed-shape	Intervention effect	0.003823	0.3922	+99.1	0.000977

Table 6: Native-LTC initialization control in the native-exact regime, where the generating law lies inside the native family. Median values over ten paired seeds; positive improvement favors near-oracle initialization. Improvement is the median of the seed-wise relative reductions and therefore need not equal the relative reduction between the two displayed medians. The p value is from a directional exact sign-flip test.

Metric	Near-oracle	Random	Improvement (%)	p
Conductance	0.0006006	0.05957	+98.8	0.000977
Vector field	0.001252	0.1562	+99.2	0.000977
Rollout	0.0007722	0.08813	+99.1	0.000977
Edge current	0.0007781	0.02561	+96.2	0.000977
Intervention effect	0.000817	0.03538	+97.0	0.000977

disappear. The dense model absorbs a simple generating law into spurious edges at some cost to the trajectory and a mixed-shape law at almost no cost; the experiment does not isolate why. Predictive accuracy is therefore uninformative about the support precisely in the regime where mechanism recovery matters most. Nothing here shows that the support could be inferred from data; the ablation assumes it is given and measures what that assumption is worth.

F NATIVE-LTC INITIALIZATION CONTROL

The restricted affine-sigmoid gate of the parent family is deliberately not enlarged to match the flexible models, which raises the question of whether its weaker performance reflects representational limitation or optimization difficulty. The two are separable in the native-exact regime, where the generating law lies inside the native family and a well-initialized model should therefore be able to reach it. Table 6 compares near-oracle against random initialization on paired seeds.

Near-oracle initialization wins every paired seed on all five metrics and reduces median conductance error by roughly two orders of magnitude. In this regime, the binding constraint is therefore optimization and not capacity. The inference does not extend to the mixed-law regimes, where the native family is also functionally restricted and the two causes are no longer separable; the control identifies which explanation holds where it can, and is silent elsewhere.

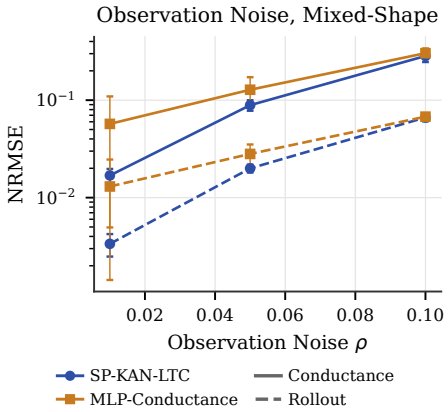


Figure 3: Observation-noise robustness in the mixed-shape regime. Mechanism error and prediction error against the noise level, with seed spread.

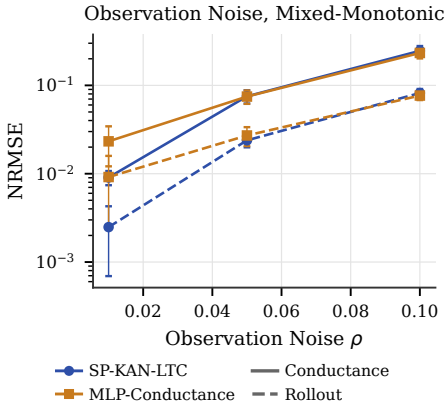


Figure 4: Observation-noise robustness in the mixed-monotonic regime.

G Observation-noise robustness

Corrupting the training observations while leaving the evaluation targets clean tests whether the advantage of the flexible edge survives when the signal it fits is degraded. The three interaction-law regimes are swept independently, since a flexible edge has more to lose where the true law is richer (Figures 3, 4 and 5).

The clean-setting ordering holds at the lowest noise level and does not survive to the highest, where the two models converge on both mechanism and prediction metrics. The convergence is not evidence that the edge parameterization stops mattering in principle; it is evidence that at this corruption level the data no longer distinguish the two function classes, which is a property of the observations rather than of the models. Because the evaluation targets remain clean, the effect is attributable to what the training signal supports and not to a degraded measurement of the outcome.

H Measured half-center oscillator, diagnostics

The two transfer directions give different answers about the flexible edge. The support diagnostics below may help explain that asymmetry; they are descriptive and do not enter model fitting or selection.

Table 7 reports all five models. The two directions are not symmetric. Training on Trial 3 and testing on Trial 5 shows no detectable difference between the two flexible factorized models under the paired test, while training on Trial 5 and testing on Trial 3 favors the multilayer perceptron in

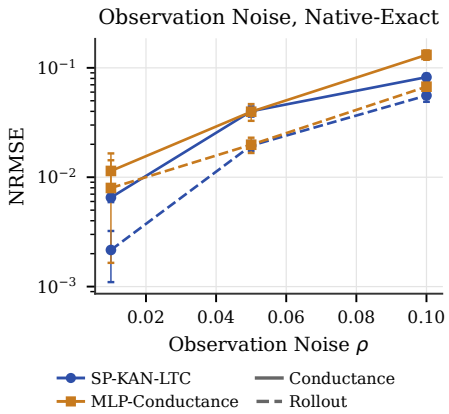


Figure 5: Observation-noise robustness in the native-exact regime, where the generating law lies inside the restricted family.

Table 7: Measured half-center oscillator, all five models. Mean ± standard deviation over ten seeds, with the two transfer directions averaged within each seed for the columns that span both. Hidden current is the error against the independently recorded synaptic current; voltage is the hundred-millisecond rollout error. Voltage rollout errors are tightly clustered across models, while hidden-current recovery separates the factorized from the unfactorized ones.

Model	Factorized	Hidden current	Waveform r	Voltage, 100 ms	Trial 3 to Trial 5	Trial 5 to Trial 3
SP-KAN-LTC	yes	0.4977 ± 0.0276	0.9260 ± 0.0399	0.0574 ± 0.0029	0.4559 ± 0.0485	0.5396 ± 0.0236
MLP-Conductance	yes	0.4621 ± 0.0276	0.9746 ± 0.0018	0.0589 ± 0.0052	0.4565 ± 0.0178	0.4677 ± 0.0483
Sigmoid-Conductance	yes	0.5712 ± 0.0063	0.9744 ± 0.0076	0.0609 ± 0.0048	0.5608 ± 0.0044	0.5817 ± 0.0104
Current-MLP	no	1.1896 ± 0.0756	0.4374 ± 0.2345	0.0575 ± 0.0019	1.2778 ± 0.1351	1.1013 ± 0.0419
Current-KAN	no	1.3279 ± 0.1054	-0.6227 ± 0.2647	0.0587 ± 0.0017	1.3853 ± 0.1333	1.2706 ± 0.0902

Table 8: Measured half-center oscillator, paired comparison on the independently recorded hidden current over ten seeds. The difference is SP-KAN-LTC minus MLP-Conductance, so a positive value indicates larger error for SP-KAN-LTC. The p value is from a two-sided exact sign-flip test.

Fold	Mean difference	SD	p
Trial 3 to Trial 5	-0.00058	0.04790	0.9707
Trial 5 to Trial 3	+0.07190	0.05712	0.007812

Table 9: Measured half-center oscillator, support diagnostics. Fraction of test inputs falling outside the range seen during training and outside the fixed spline grid. These fractions are properties of the fixed train and test split rather than of a fitted model, so they are identical across the ten seeds and carry no seed-to-seed deviation.

Fold	Edge	Outside training range	Outside spline grid
Trial 3 to Trial 5	0	0.0000	0.0505
Trial 3 to Trial 5	1	0.0000	0.0583
Trial 5 to Trial 3	0	0.0014	0.0714
Trial 5 to Trial 3	1	0.0003	0.0821

eight of ten paired seeds. Since the architectures, the budget and the protocol are identical in both directions, the asymmetry is associated with differences between the two trial directions rather than with architecture or parameter budget, and one candidate explanation is how the two trials populate the fixed spline domain.

Table 9 supports that reading without establishing it, and Table 8 gives the paired tests behind the comparison. Test inputs almost never leave the training range, but a non-negligible fraction falls outside the fixed spline grid in the direction where the flexible edge does worse. Outside the fixed

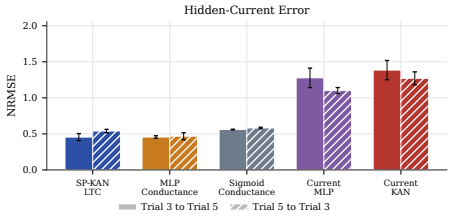


Figure 6: Hidden-current error by model and transfer direction on the measured half-center oscillator. Every factorized run lies below every run of either unfactorized control.

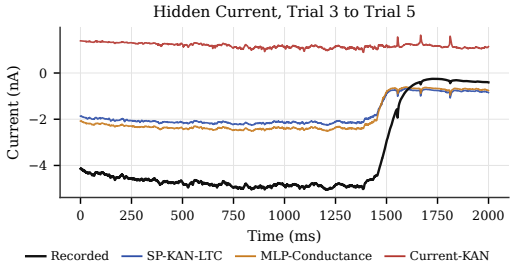


Figure 7: Recorded and learned hidden current, training on the first trial and testing on the second.

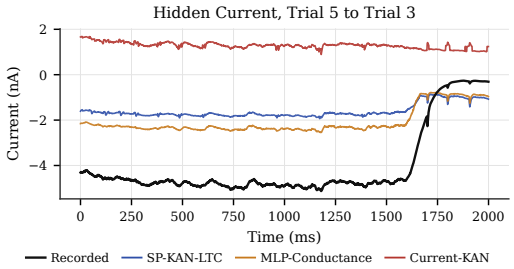


Figure 8: Recorded and learned hidden current, reverse direction.

spline grid, the B-spline contribution is inactive, leaving extrapolation to the global base activation and bias rather than to the fitted spline coefficients; the perceptron has no analogous grid boundary. This is consistent with the excitation account that the controlled and empirical-topology results also support, but there are only two directions and the diagnostic is correlational, so it is offered as a plausible reading rather than as a demonstrated cause.

Figure 6 summarizes the separation by model and transfer direction, and Figures 7 and 8 show the corresponding traces. The traces show SP-KAN-LTC and MLP-Conductance following the broad temporal structure of the recorded current while Current-KAN does not; Figure 6 gives the full five-model comparison. Neither flexible factorized model reproduces the current exactly, and substantial residual error remains in both directions. The evidence therefore supports a separation between the factorized and unfactorized models across both function classes, not accurate reconstruction of the recorded current.

I *C. elegans* INTERVENTION TRANSFER

Six conditions were trained and five scored. The same numerical-integrity gate applies to all six and uses no biological score; five pass, while Current-KAN fails at every tested solver refinement, leaving it numerically ineligible for the 30-second endpoint. Table 10 reports the eligible conditions, and Figures 9 and 10 show the two measures against chance.

Rank agreement remains near zero. Top-five overlap spans both sides of the random-set expectation, with SP-KAN-LTC at the lowest mean of 0.236 against an expectation of 0.267. None of the

Table 10: *C. elegans* intervention transfer at the 30-second endpoint, for the five numerically eligible conditions. Mean ± standard deviation over ten seeds. The random-set expectation for top-five overlap is 0.267.

Condition	Rank agreement	Top-five overlap
Linear ODE	$+0.0326 \pm 0.0207$	0.2873 ± 0.0123
MLP-Conductance	$+0.0452 \pm 0.0163$	0.2627 ± 0.0157
Sigmoid-Conductance	$+0.0124 \pm 0.0256$	0.2764 ± 0.0254
SP-KAN-LTC	$+0.0338 \pm 0.0148$	0.2364 ± 0.0105
SP-KAN-LTC (shuffled)	$+0.0155 \pm 0.0602$	0.2845 ± 0.0271

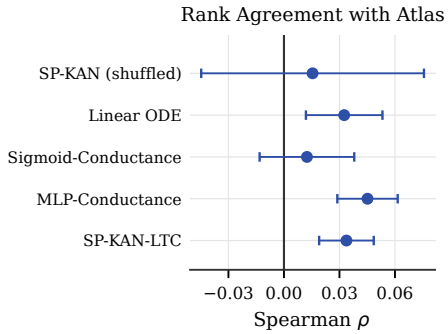


Figure 9: Target-wise rank agreement with the held-out intervention atlas for the five eligible conditions.

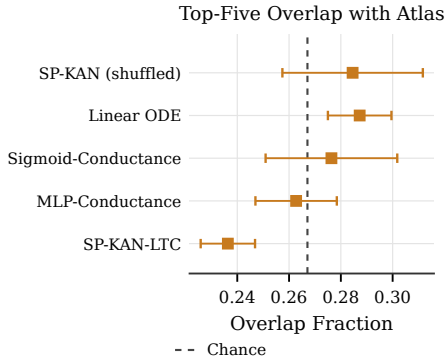


Figure 10: Top-five overlap with the atlas against the random-set expectation computed over the evaluable candidate sets.

eligible conditions demonstrates reliable transfer, and the ordering among them is not interpretable given this spread. The shuffled-graph control matters most here: it sits inside the same range as the true connectome, so the protocol provides no evidence that graph structure contributed to whatever agreement was obtained. A linear connectome-constrained model performs no worse than the conductance models, consistent with the absence of evidence that any model class recovers the intervention mechanism.

The null does not isolate a limitation of the architecture. The protocol asks for transfer across a change of intervention modality, from observational calcium dynamics to optogenetic perturbation, in a system where the generating edge law is unknown, the excitation in the training recordings is not controlled, and the connectome is itself an estimate.

Table 11: Cross-solver numerical consistency on the eight-edge mixed-shape condition, over ten seeds. Rollout error under fixed-step and refined-step Runge–Kutta integration and under an adaptive reference solver, with the disagreement between them and the step-refinement ratio.

Quantity	Mean ± SD	Minimum	Maximum
Rollout NRMSE, RK4 at dt	0.00280 ± 0.00054	0.00175	0.00368
Rollout NRMSE, RK4 at dt/2	0.00280 ± 0.00054	0.00175	0.00368
Rollout NRMSE, DOPRI5	0.00280 ± 0.00054	0.00175	0.00368
Disagreement, RK4 dt vs DOPRI5	$(4.25 \pm 2.54) \times 10^{-7}$	2.48×10^{-7}	1×10^{-6}
Disagreement, RK4 dt/2 vs DOPRI5	$(2.52 \pm 1.08) \times 10^{-8}$	1.51×10^{-8}	4.6×10^{-8}
Disagreement / model error	$(1.61 \pm 1.12) \times 10^{-4}$	6.74×10^{-5}	4.36×10^{-4}
Step-refinement ratio	16.3 ± 2.8	12.4	21.8
RK4 dt bound violations	0	0	0
RK4 dt over 2 bound violations	0	0	0
DOPRI5 bound violations	0	0	0

J Cross-solver numerical consistency

Continuous-time results can depend on the integrator, and a conclusion that changes with the step size is a property of the discretization rather than of the model. The eight-edge mixed-shape condition is therefore evaluated under a refined fixed step and under an adaptive reference solver.

Table 11 reports the comparison. The three settings agree to five significant figures, the disagreement between the fixed-step and reference solutions is four orders of magnitude below the model error itself, and no run violates the state bounds, whose continuous-time counterpart is given in Appendix A.3. The controlled-benchmark conclusions are therefore numerically consistent for the tested system, step sizes and solvers. That statement does not extend to discretization invariance in general, and it is not evidence that the system is non-stiff; both would require conditions this experiment does not establish.

K Edge-design sensitivity

The canonical edge uses cubic splines on eight intervals, giving thirteen trainable parameters, a count set by capacity matching against the multilayer perceptron. How sensitive the results are to that choice is examined here.

A higher-resolution grid does not yield a uniform benefit at four interactions. Table 12 gives the comparison.

Under the mixed-monotonic law, the variant is worse on conductance, edge current and intervention, winning four of ten paired seeds in each case, while under mixed-shape it improves conductance substantially. The higher-resolution variant can therefore help under mixed-shape laws without giving a uniform advantage across regimes; the thirteen-parameter canonical edge is used throughout the matched comparisons.

L Experimental protocols and training details

The settings below are those the experiments used.

L.1 Controlled benchmark

The generating systems have four states with known graph support at four or eight active edges, under three interaction-law regimes: native-exact, in which the true law lies inside the restricted parent family, mixed-monotonic, and mixed-shape. Each condition draws 64 training, 16 validation and 24 test trajectories of 120 steps at a fixed step of 0.05 with no jitter and is repeated over ten seeds. Each seed determines the generated training, validation and test trajectories as well as training randomness; models paired at the same seed share the same generated splits. Leak conductances, leak

Table 12: Edge-design sensitivity at four active interactions: a higher-resolution spline grid against the canonical thirteen-parameter edge. Median values over ten paired seeds; positive improvement favors the variant. Improvement is the median of the seed-wise relative reductions and therefore need not equal the relative reduction between the two displayed medians. The p value is from a two-sided exact sign-flip test.

Regime	Metric	Variant	Canonical	Improvement (%)	p
Mixed-monotonic	Conductance	0.004995	0.004491	-13.1	0.572
Mixed-monotonic	Edge current	0.0008492	0.000893	-6.1	0.594
Mixed-monotonic	Intervention effect	9.283×10^{-5}	0.000117	-8.4	0.812
Mixed-monotonic	Rollout	0.0003334	0.0004336	+4.3	0.514
Mixed-monotonic	Vector field	0.002472	0.002608	-5.8	0.605
Mixed-shape	Conductance	0.005766	0.01031	+42.0	0.00195
Mixed-shape	Edge current	0.003148	0.005977	+49.4	0.00195
Mixed-shape	Intervention effect	0.0006121	0.001332	+60.8	0.00195
Mixed-shape	Rollout	0.001479	0.00302	+52.8	0.00195
Mixed-shape	Vector field	0.005113	0.01029	+48.5	0.00195
Native-exact	Conductance	0.003852	0.003871	+11.9	0.236
Native-exact	Edge current	0.001007	0.001011	+26.4	0.541
Native-exact	Intervention effect	0.0004353	0.0004324	+12.6	0.711
Native-exact	Rollout	0.0005882	0.0006027	+14.3	0.0586
Native-exact	Vector field	0.003589	0.003301	+10.6	0.234

reversals and the external input are fixed per state and shared across models, and the conductance cap is 1.5 for every model that has one.

Models are trained on one-step state error with a batch size of 256 for at most 1000 epochs, stopping after 150 epochs without improvement, with gradient clipping at 5.0, no weight decay, and a plateau scheduler that halves the rate after 12 stagnant epochs down to a floor of 10^{-5}. Selection uses validation loss alone. Integration is fourth-order Runge–Kutta. Learning rates are fixed per model at 0.03 for SP-KAN-LTC, 0.1 for MLP-Conductance, 0.01 for Native LTC and 0.03 for Current-KAN, obtained from validation-only tuning with mechanism-recovery and test metrics excluded from selection; Current-MLP uses the validation-only procedure specified below. Edge functions use cubic splines on eight intervals over $[-1.5, 1.5]$ with a SiLU base activation; Current-KAN uses a single interval of degree one over the same grid. Intervention evaluation uses eight held-out rollout trajectories per edge.

The identifiability, misspecification, observation-noise and sparsity studies inherit this protocol, each varying one thing: the separation between reversal references, a signed shift applied to every reversal reference, the noise added to training observations while evaluation targets stay clean, and dense against known sparse support. The hidden-memory study does not inherit it and is described separately below.

L.2 Unfactorized perceptron control

Six rates from 10^{-3} to 0.3 are swept on three tuning seeds in a single condition, eight edges under mixed-shape laws, and the rate minimizing the median validation one-step error is selected, with ties broken by the mean and then by the smaller rate; mechanism targets are excluded from tuning. The selected rate of 0.1 is used for all $2 \times 3 \times 10 = 60$ evaluation runs, and the edge map matches the other unfactorized control in budget and inputs.

L.3 Empirical-topology benchmark

The topology is a measured hospital-ward contact network; the dynamics on it are simulated, so the benchmark is semi-synthetic rather than a record of transmission. Contact seconds are aggregated and edge weights transformed by $\log(1+x)$ on the contact count. Three graph variants are used: the full empirical graph and two role-stratified sparsifications targeting mean undirected degrees of four and eight, each required to stay connected and to preserve every node. Trajectories are simulated at a step of 0.05 over 120 steps under fourth-order Runge–Kutta. Each condition draws 48 training, 12

validation and 16 test trajectories. Training uses 96 windows of 20 steps per epoch with validation every five epochs, capped at 500 epochs for every model. Each model uses a learning rate selected by validation performance over six rates for the factorized models and ten for the unfactorized control. The simulated train, validation and test splits are fixed per condition; the ten training seeds vary model initialization and training-window sampling, with models paired at a seed using the same prepared splits. Intervention evaluation removes each of eight candidate contacts and compares 80-step rollouts over six held-out trajectories.

L.4 HIDDEN-MEMORY STRESS TEST

The hidden-memory study uses a Morris–Lecar system rather than the controlled generator and is not governed by the protocol above. Two neurons are observed through their voltages and recovery variables, and each reciprocal interaction carries a gating variable with its own dynamics,

$$I_e = \bar{g}_e \, s_e \, (E_e - V_{\mathrm{post}}), \qquad \tau_s \, \dot{s}_e = s_\infty(V_{\mathrm{pre}}) - s_e, \tag{11}$$

so the interaction is no longer a function of the presynaptic voltage alone. The mismatch level is $\eta = \tau_s/\tau_{\mathrm{ref}}$, where $\tau_{\mathrm{ref}} = 1/\phi$ and ϕ is the Morris–Lecar recovery-rate factor. We test $\eta \in \{0, 0.1, 0.5, 1\}$; $\eta = 0$ is implemented as the exact instantaneous limit $s_e = s_\infty(V_{\mathrm{pre}})$ rather than as a small value. All four models share the known intrinsic vector field and learn only the two reciprocal interaction currents, so they differ only in how the interaction is represented: SP-KAN-LTC and MLP-Conductance through an instantaneous conductance, a closed-form continuous-time network through recurrent state that reads out the two currents, and an augmented neural ODE through a persistent latent interaction state integrated alongside them. Each model uses a learning rate selected by validation performance over its own three-rate grid.

The reversal persists at the larger mismatch levels. At $\eta = 0.5$, the interaction-current error is 0.643 ± 0.047 for SP-KAN-LTC, 0.621 ± 0.039 for MLP-Conductance, 0.212 ± 0.058 for CfC and 0.190 ± 0.018 for the augmented neural ODE; at $\eta = 1$, the corresponding errors are 0.794 ± 0.028, 0.772 ± 0.073, 0.272 ± 0.051 and 0.184 ± 0.049. The history-capable baselines remain ahead of the factorized models at both levels. The main text discusses $\eta = 0$ and $\eta = 0.1$ because they bracket the crossing, whose location is not determined.

L.5 MEASURED HALF-CENTER OSCILLATOR

The recordings are the deposited dynamic-clamp dataset of Burghi et al. (2025), sampled at 0.1 ms, comprising two trials. Membrane voltages and injected currents are the observed variables, and the synaptic current between the two neurons is recorded separately and withheld from training and from model selection, which is what makes it a mechanism target rather than a fitted quantity. The graph consists of two reciprocal inhibitory edges, with reversal potentials fixed at -80 mV. One transfer direction trains on one trial and tests on the other; the second reverses the split. SP-KAN-LTC retains the same thirteen-parameter cubic-spline edge used elsewhere. The panel contains three conductance models and two unfactorized signed-current controls. For the conductance models, the maximum conductance is learned per edge within $[10, 500]$ nS rather than fixed, since the physical scale is not known in advance; Current-KAN and Current-MLP expose no conductance at all, and Current-MLP maps the same two inputs to a signed current through a one-hidden-layer width-three perceptron. Unlike the controlled factorial comparison, the measured comparison is not an exact per-edge parameter-count match: SP-KAN-LTC and MLP-Conductance each use a thirteen-parameter edge function together with one learned bounded maximum conductance per edge, whereas Current-KAN and Current-MLP each use thirteen parameters per edge, and Sigmoid-Conductance remains the lower-capacity restricted parent family at three parameters per edge. Intrinsic dynamics not attributable to the synapse are absorbed by a nuisance branch common to all five models: ten causal exponential voltage filters with time constants of 0.5, 1, 2, 5, 10, 20, 50, 100, 200 and 500 ms, whose outputs are passed with the instantaneous voltage into a per-neuron perceptron of widths $11 \to 32 \to 32 \to 1$ with tanh activations. Injected current enters separately. The same nuisance architecture is used for every model and is identically initialized across models at a given seed, so the architecture and initialization of that branch are controlled across comparisons, though each model optimizes its own copy.

L.6 External intervention transfer

Neuron identities come from the whole-animal atlas of Yemini et al. (2021) and the held-out intervention responses from the optogenetic atlas of Randi et al. (2023), which is never used in training. Twenty-four identified neurons are modeled, selected by coverage and then connectome degree, with worms split $70/15/15$ percent into training, validation and test sets and calcium traces centered and scaled using training-set statistics only, clipped to bound the effect of transients. Synaptic sign is fixed rather than learned, taken from the independent molecular prior where it resolves and from a train-only correlation fallback for unresolved edges; intervention outcomes are never used in this step. Each worm contributes eight baseline states, and the perturbation applied at evaluation is one training standard deviation of the state. The endpoint is the time-averaged response over 30 seconds under two-step reset semantics. Scoring requires at least eight overlapping neurons and at least three sources per ranked target.

Two numerical criteria are involved and should not be confused. Integration resolution is selected using validation trajectories alone, by comparison with doubled refinements and with independent stiff reference solutions at tolerances of 10^{-8} and 10^{-10}. Eligibility for the 30-second endpoint is determined separately, per model, by a numerical-consistency gate that compares each candidate resolution with its doubled refinement over the same window and uses no biological score. Passing requires trajectory relative error at or below 0.005 at the 95th percentile and 0.02 at the maximum, intervention-effect relative error at or below 0.01 at the 95th percentile, intervention-effect absolute error at or below 0.002 at the 95th percentile and 0.01 at the maximum, and zero guard violations. A model failing at every tested refinement receives no score rather than being evaluated at an unverified resolution. The gate was applied to all six trained conditions; five passed. Current-KAN failed at each successive refinement and is therefore reported as numerically ineligible, which is why five conditions rather than six appear in Appendix I.

L.7 Statistics

Metric definitions, the paired-effect construction, the bootstrap and the exact sign-flip tests are given in Appendix M and are not repeated here.

M Metric definitions and inferential procedure

Every reported error is normalized, and the normalization is set per benchmark and per quantity rather than shared across them. The mechanism-facing quantities of the controlled benchmark and the measured half-center oscillator use a relative L_2 error, $\|\hat{y} - y\|_2/\|y\|_2$, with the denominator computed on the same held-out evaluation set. Trajectory errors on the empirical-topology and Morris–Lecar benchmarks use a root-mean-square error divided by a standard deviation of the target rather than by its norm, which suits targets whose mean is far from zero: the empirical-topology rollout error divides by the standard deviation of the target trajectory, and the Morris–Lecar rollout and intervention quantities by the standard deviation of the training states, with interaction-current error standardized by the held-out current. Normalization puts conductance, current and state derivative on dimensionless scales within a benchmark, since they do not share units; differing denominators make magnitudes incomparable across benchmarks.

The core reported quantities are conductance where it is defined, edge current, rollout and intervention effect; the controlled benchmark additionally reports vector-field error. Conductance error compares c_{ij} against the generating law on a uniform grid spanning the observed source-state support, so every part of the support contributes equally regardless of how often the dynamics visit it. Edge-current error compares I_{ij} along held-out trajectories, so the same functional error is weighted by the visited states and by the realized driving force. The difference between these two is not a matter of precision: a model can hold an advantage on one and lose it on the other, and the empirical-topology results show that it does. Rollout error compares integrated trajectories from held-out initial conditions. Intervention-effect error is computed by removing an edge, or an identifiable group of edges, from the learned and generating systems separately and comparing the resulting change in trajectory; it is reported together with the rank correlation between learned and true effect orderings.

The training seed is the unit of inference. Unless otherwise noted, condition-level results are means with standard deviations over seeds; the paired ablation tables report medians instead, because their

per-seed ratios are skewed. Matched contrasts are computed per seed as the relative reduction $(e_b - e_a)/e_b$, where e_a and e_b are the errors of the two models on that seed, and summarized as the median over seeds with a 95% bootstrap percentile interval from $20{,}000$ resamples at a fixed random seed. The median is used rather than the mean because the per-seed error distributions are right-skewed, and a bootstrap interval rather than a parametric one because ten seeds do not support a distributional assumption. Where a condition repeats across interaction-law regimes, the median paired relative reduction across the three regimes is taken within each seed and then summarized across seeds, so that the number of independent units remains the number of seeds rather than the number of seed-by-regime cells. Directional win counts are descriptive only. Exact sign-flip tests are used where the design supports them: with ten paired differences all 2^{10} sign assignments are enumerated, so the reported p value is exact rather than asymptotic. The native-initialization and sparsity contrasts have a declared improvement direction and use directional tests, for which the smallest attainable value at ten pairs is $1/1024 = 0.000977$. The edge-design sensitivity and measured half-center oscillator comparisons have no declared direction and are tested two-sided against the absolute mean paired difference, where the smallest attainable value is $2/1024 = 0.00195$.

Model selection uses validation data, and the reported evaluation uses held-out data. None of this addresses the validity of the mechanism target itself. Where the target is generated, the comparison is exact by construction and the inferential procedure bounds only sampling variability across seeds; where it is measured, the target carries its own error, which these intervals do not represent.