CHIRP: RECOVERING NETWORK TOPOLOGY FROM COUPLING-RESPONSE PROFILES

Anonymous authors
Paper under double-blind review

ABSTRACT

Synchronized trajectories can arise from a direct link, an indirect path, or a shared source. Oscillator experiments often sweep coupling strength only to describe the approach to synchronization. We instead write each phase-difference coefficient as a dose-independent term plus a term that changes with the known dose, and rank candidate edges by the second term. A single dose leaves the two terms collinear; varying the dose separates them. We bound the coefficient error and the condition for recovering the edge set. Choices were fixed on five physical networks and applied once to fifteen held-out graphs of 28 chaotic electronic oscillators. The locked estimator ranks edges above each declared baseline. Cutting the raw traces to a 3 s budget before filtering still gives higher average precision than a single-dose model under that same raw-time budget. Otsu's rule selects a graph from the ranking without using the known number of edges. In synthetic trials, average precision falls in a confounded scenario that adds a constant-amplitude common drive together with other disturbances, and falls again when the dose response is nonlinear. These results suggest that a calibrated intervention can make topology easier to identify than an observational fit when the intervention is structurally relevant and approximately affine over the fitted range.

1 INTRODUCTION

Network structure is often inferred from the way nodes move together. For oscillators, that inference is not straightforward. Two nodes can co-vary because they are directly connected, because they communicate through another node, or because both respond to a common source. A reduced phase model can introduce another source of dependence through model error. As a result, functional connectivity may be dense even when the physical graph is sparse (Vera-Ávila et al., 2020a). More flexible dynamical reconstruction helps, but it still has to distinguish a direct interaction from several other mechanisms that produce similar trajectories (Casadiego et al., 2017; Singhal et al., 2026).

Many experiments provide an additional source of information. The operator changes a calibrated quantity—for example, a coupling conductance, stimulus gain, ligand concentration, or interaction range—and records the system at several settings. Such sweeps are commonly used to describe the approach to synchronization or to fit a separate network at each setting (Vera-Ávila et al., 2020b; Singhal et al., 2023). We use the sweep differently. Our assumption is that a genuine coupling contribution should change with the applied dose, whereas natural frequencies, sensor cross-talk, persistent common drive, and some forms of model mismatch need not.

We call the resulting procedure CHIRP (Coupling-Homotopy Interventional Response Profiling). For receiver i, we write

$$\dot{\phi}_i^{(e)} = \alpha_{ie} + \sum_{j \neq i} \big(c_{ij} + d_e b_{ij}\big) \sin(\phi_j^{(e)} - \phi_i^{(e)}) + \epsilon_i^{(e)}. \tag{1}$$

Here, d_e is the known dose. The environment-specific intercept absorbs frequency shifts, c_{ij} represents a dose-independent component, and b_{ij} is the response to the intervention. We rank undirected edges using the symmetrized magnitude of the two directed slopes. The nuisance coefficients are not assumed to be sparse; only the intervention response is used as structural evidence.

This paper makes four contributions.

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.

1. We formulate topology recovery as estimation of an intervention response rather than selection from a single observational fit.
2. We show why a single dose cannot separate the nuisance and response terms, derive a rank condition for dose variation, and give finite-noise support bounds.
3. We freeze the estimator, baselines, thresholds, and decision criteria on R1–R5 and then evaluate once on R6–R20 of a public benchmark (Vera-Ávila et al., 2020b).
4. We examine both favorable and unfavorable cases using equal unique-time budgets, alternative baselines, dose permutations, a label-free threshold, reduced intervention budgets, and 45 synthetic trials.

The result is undirected topology among measured nodes, under a known scalar intervention whose effect is approximately affine along a short path. Direction, hidden nodes, and disturbances that themselves move with the dose are outside that setting; Section 8 returns to them.

2 Related work

Network inference from dynamics. Inferring a network from trajectories is an old problem, from Granger causality and transfer entropy to phase-coupling direction and reviews of dynamics-based reconstruction (Granger, 1969; Schreiber, 2000; Rosenblum & Pikovsky, 2001; Timme & Casadiego, 2014; Wang et al., 2016). For oscillators the usual description is a phase reduction of a limit-cycle network (Kuramoto, 1984; Pikovsky et al., 2001; Acebrón et al., 2005; Pietras & Daffertshofer, 2019; Stankovski et al., 2017). Phase-dynamics reconstruction represents the coupling function with Fourier terms (Kralemann et al., 2011); ARNI searches for direct nonlinear interactions without fixing the node model (Casadiego et al., 2017); and later phase-reduction methods cover both synchronized and desynchronized regimes (Matsuki et al., 2025; Hwang et al., 2026; Su et al., 2026). Those methods enrich the observational model. CHIRP asks a different question: whether a candidate coefficient moves with a calibrated intervention.

The most relevant experimental comparisons use the same 20-network electronic-oscillator corpus. Vera-Avila et al. studied structural identifiability through statistical similarity (Vera-Ávila et al., 2020a), while Singhal et al. combined correlation screening with phase-model regression (Singhal et al., 2023). NIPS uses forced-delay embeddings and partial state measurements, and reports a mean AUROC of approximately 0.96 in the weak-coupling region of this benchmark (Singhal et al., 2026). That result uses a different implementation and a different score, so we compare average precision on the same raw files.

Multiple environments and interventions. Several environments can be pooled without a known dose. Heterogeneous-dataset reconstruction keeps a common Boolean network and lets coefficients vary by environment (Yue et al., 2021); SpReME uses a related sparse regression (Park et al., 2023). Invariant prediction and CausalKinetiX use environmental changes to identify predictors or kinetic relationships (Heinze-Deml et al., 2018; Pfister et al., 2019), and anchor regression targets robustness under distributional shifts (Rothenhäusler et al., 2021). Our restriction is more specific: the dose is known, it is interacted with every candidate edge, and the observational nuisance is left unrestricted. Topology is then read from the dose slope. Section 4 shows the identifiability benefit of this restriction.

An affine dependence on a measured scheduling variable is standard in linear parameter-varying identification (Petreczky et al., 2023), and response to a perturbation has long been used for biochemical interactions (Kholodenko et al., 2002). Here the scheduling variable is a coupling dose, the nuisance term in Equation 1 is left dense, and the held-out test is a set of physical oscillator graphs.

Why average precision. There are 42 edges among 378 possible undirected pairs in each physical graph. With this imbalance, AUROC can remain high even when the top-ranked edges are not especially useful. We therefore use average precision (AP) as the primary ranking metric; its random-ranking reference is $42/378 = 0.111$ (Saito & Rehmsmeier, 2015). We report Otsu's label-free threshold as a separate operating point so that graph construction does not require the true edge count (Otsu, 1979).

3 PROBLEM AND METHOD

We observe one scalar voltage $v_i^{(e,r)}(t)$ for node $i \in \{1, \dots, N\}$ in environment e and repeat r. The coupling dose d_e is known, but the graph $A \in \{0,1\}^{N \times N}$ is not. In the physical benchmark, the graph is undirected and has no self-loops. After filtering, we use a Hilbert transform to obtain the unwrapped phase $\phi_i(t)$. Rather than differentiate the phase at each sample, we use the finite increment

$$y_{ier}(t) = \frac{\phi_i(t + W/2) - \phi_i(t - W/2)}{W \Delta t} \in \mathbb{R}^{n_e}, \qquad x_{ij,er}(t) = \sin(\phi_j(t) - \phi_i(t)). \tag{2}$$

The window reduces pointwise differentiation noise, while still allowing some model mismatch to remain in the regression.

For each target node, CHIRP fits Equation 1. We subtract the within-environment mean from the response and from every feature, which removes the environment-specific term α_{ie}. If $X_e \in \mathbb{R}^{n_e \times (N-1)}$ denotes the centered phase-difference features and y_e the centered response, the design for environment e is

$$\begin{aligned} &Z_e = [X_e \;\; d_e X_e] \in \mathbb{R}^{n_e \times 2(N-1)}, \quad \vartheta_i = [c_i^\top \;\; b_i^\top]^\top \in \mathbb{R}^{2(N-1)}, \\ &D_i = \mathrm{diag}(s_i), \quad \widehat{\vartheta}_i = \arg\min_{\vartheta} \sum_e \|y_e - Z_e \vartheta\|_2^2 + \lambda_i \|D_i \vartheta\|_2^2. \end{aligned} \tag{3}$$

The diagonal of D_i holds the column standard deviations of the stacked refit design. An entry below 10^{-10} is replaced by 1. Ridge penalties are chosen on a chronological 60/20 split pooled over environments, and the selected model is refit on the first 80% of each recording. Equation 3 is that refit. The penalty acts on $D_i \vartheta$, and the returned coefficient is in the units of the centered features. Proposition 3 bounds this solution (Appendix B.9). The fit does not use graph labels or the known graph density. For an undirected pair, we define

$$s_{ij} = \tfrac{1}{2}\big(|\widehat{b}_{ij}| + |\widehat{b}_{ji}|\big), \qquad i < j. \tag{4}$$

The scores are evaluated with AP and AUROC. When a binary graph is required, Otsu's rule partitions the 378 scores by maximizing between-class variance.

Why fit the doses jointly? At a fixed dose, the phase-difference coefficient is $c + d_e b$. Fitting every dose independently and then taking a slope therefore estimates several highly correlated quantities and can magnify noise. Under unregularized least squares with equal Grams, that two-stage slope is algebraically identical to the joint estimator (Proposition 4). The physical designs do not satisfy those conditions: the phase-difference Gram changes with dose, and each separate fit selects its own ridge. The first reported two-stage number also skips the column standardization used by the joint fit. The joint design estimates one dose-independent component and one dose-response direction from all rows at once, with a single penalty. The two-stage slope and the separate-fit homotopy are included to test whether the resulting difference in estimators matters empirically.

4 IDENTIFIABILITY FROM A COUPLING HOMOTOPY

Section 3 fits one slope per candidate edge. The statements below say when that slope is identifiable, when the dose placement is optimal for an affine response, and when a two-stage slope would have returned the same number.

We first consider one receiver and suppress the receiver index and ridge penalty. Let $w_e > 0$ denote the number, or more generally the relative weight, of observations in environment e.

Proposition 1 (Dose variation breaks fixed-dose collinearity)**.** *Suppose $X_e^\top X_e = w_e G$ for all environments, with $G \succ 0$. The stacked design $Z = [Z_1^\top, \dots, Z_E^\top]^\top$ has full column rank if and only if*

$$\sum_e w_e (d_e - \bar{d}_w)^2 > 0, \qquad \bar{d}_w = \frac{\sum_e w_e d_e}{\sum_e w_e}. \tag{5}$$

At a single dose, c and b cannot be identified separately.

Proof. Let $S_k = \sum_e w_e d_e^k$. Then $Z^\top Z = \begin{bmatrix} S_0 & S_1 \\ S_1 & S_2 \end{bmatrix} \otimes G$. The determinant of the first factor is $S_0 S_2 - S_1^2 = S_0 \sum_e w_e (d_e - \bar{d}_w)^2$. Since G is positive definite, the stacked design has full rank exactly when the weighted dose variance is positive. If all doses are equal, $d_e X_e$ is a scalar multiple of X_e, so the two coefficient blocks are collinear. □

The equal-Gram assumption is used only to obtain this transparent expression. In practice, the relevant condition is $\lambda_{\min}(Z^\top Z) > 0$. The column scales in Equation 3 are positive, so replacing Z by ZD^{-1} does not change that rank condition. Increasing the spread of the doses improves the conditioning of the idealized problem, although very large doses can violate the affine response or the phase-reduction approximation. Appendix B.1 records that rank check. What the estimator uses is the slope, not the coefficient at one dose. Figure 1 is a constructed pair: the non-edge coefficient is the constant 1.2, and the edge coefficient is $0.2 + 16d$. The non-edge is larger at every protocol dose, including $d = .05$. The slope reverses that order.

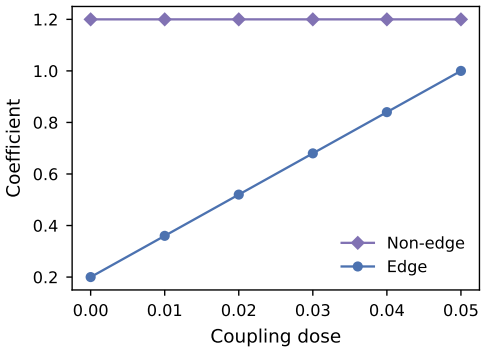

Figure 1: Constructed coefficients on the protocol doses. The non-edge stays at 1.2. The edge is $0.2 + 16d$, so a single dose ranks the non-edge first and the slope ranks the edge first. Not a recording.

Proposition 2 (Optimal affine dose spread)**.** *Under Proposition 1, homoskedastic noise with variance σ^2, and unregularized least squares,*

$$\mathrm{Cov}(\widehat{b}) = \frac{\sigma^2}{\sum_e w_e (d_e - \bar{d}_w)^2} G^{-1} \tag{6}$$

For a fixed total weight and doses restricted to $[d_{\min}, d_{\max}]$, this covariance is minimized in the Loewner order by placing equal total weight at the two endpoints.

Proof. Block inversion gives the stated covariance. The variance of a bounded variable is at most $(d_{\max} - d_{\min})^2/4$, with equality for equal mass at the two endpoints. □

Endpoint allocation is optimal only if the affine approximation remains credible. Interior doses are useful for checking curvature, which is why the physical protocol uses six doses rather than only the two endpoints (Appendix B.2).

Proposition 3 (Finite-noise coefficient error)**.** *Let D be the positive diagonal matrix in Equation 3, and write $\tilde{Z} = ZD^{-1}$. Under $y = Z\vartheta^\star + \epsilon$, the minimizer satisfies*

$$\|D(\widehat{\vartheta} - \vartheta^\star)\|_2 \le \frac{\|\tilde{Z}^\top \epsilon\|_2 + \lambda \|D\vartheta^\star\|_2}{\lambda_{\min}(\tilde{Z}^\top \tilde{Z}) + \lambda}. \tag{7}$$

Proof. Set $u = D\vartheta$. The objective is ordinary ridge regression of y on $\tilde{Z}$. Subtract $(\tilde{Z}^\top \tilde{Z} + \lambda I)u^\star$ from the normal equation and bound the inverse by its spectral norm. □

Dividing the left-hand side by a diagonal entry of D bounds that coordinate of $\widehat{\vartheta} - \vartheta^\star$. The support corollary uses such an allowance on the slope.

Corollary 1 (Support separation)**.** *If nonedges satisfy $b^\star_{ij} = b^\star_{ji} = 0$, every directed slope error is at most δ, and the smallest true symmetrized edge score is larger than 2δ, then any threshold strictly between δ and $s^\star_{\min} - \delta$ recovers the undirected support exactly.*

This result concerns coefficient separation. It does not say that Otsu's rule will land in the separating interval. A disturbance that moves with the dose enters the intervention coefficient and can be mistaken for an edge. Section 7 probes a different mismatch: a common drive whose amplitude is constant across doses, and a coupling gain that is not affine.

The same equal-Gram algebra that makes the rank condition transparent also makes a two-stage slope coincide with the joint estimator when the Grams share one shape.

Proposition 4 (Joint least squares versus two-stage slope). *Let $G_e = X_e^\top X_e$ and $\widehat{\beta}_e = G_e^{-1} X_e^\top y_e$, assuming each G_e is invertible. The unregularized joint estimator of (c, b) is the generalized least-squares fit of $\widehat{\beta}_e = c + d_e b$ with precision G_e. If $G_e = w_e G$ for a common $G \succ 0$, this fit coincides with weighted two-stage OLS of $\widehat{\beta}_e$ on d_e. Equal weights w_e recover unweighted two-stage OLS. The condition is sufficient, not necessary: with two environments both estimators interpolate the two coefficient vectors for any invertible Grams.*

Proof. The joint normal equations are

$$\sum_e G_e(c + d_e b) = \sum_e G_e \widehat{\beta}_e, \qquad \sum_e d_e G_e(c + d_e b) = \sum_e d_e G_e \widehat{\beta}_e, \tag{8}$$

which are the GLS equations for $\widehat{\beta}_e = c + d_e b$ with precision G_e. Substituting $G_e = w_e G$ and canceling the common positive-definite factor G leaves weighted least squares of $\widehat{\beta}_e$ on d_e. Equal w_e drop out of both sides. Necessity fails for $E = 2$. An intercept and a slope match the two vector observations exactly, and any positive-definite precision still selects that interpolant, so the slopes agree without G_1 and G_2 being scalar multiples of one matrix. With more than two environments, a Gram that is not of the form $w_e G$, or a ridge λ_e chosen inside environment e, can move the two-stage slope off the joint slope (Appendix B.4). □

5 FROZEN PHYSICAL EXPERIMENT

Dataset and split. The public benchmark contains 20 independently wired networks of 28 chaotic electronic Rössler oscillators. Each network has 42 undirected links, 101 coupling settings, three one-second repeats, and a sampling rate of 30 kHz (Vera-Ávila et al., 2020b). We used R1–R5 for development. Before reading any voltage data from R6–R20, we fixed the estimator, preprocessing, hyperparameter grid, baselines, metrics, random seed, threshold rule, and numerical advance criteria in a timestamped protocol. The protocol and analysis module were recorded with SHA-256 hashes (Appendix D). We then processed R6–R20 once. Because windows from one trajectory are dependent, the graph rather than the regression row is the independent unit.

Preprocessing and budget. The confirmatory fit uses six doses $d \in \{0, .01, .02, .03, .04, .05\}$ and all three repeats, which is 18 s of raw voltage. The single-dose baseline uses the three repeats at $d = .05$, which is 3 s. Matching the number of regression rows therefore still gave CHIRP six times as much raw trajectory. Filter cutoffs, window, stride, and the ridge grid are in Appendix D. After confirmation we repeated the fit on a 3 s raw budget: 0.5 s from repeat 1 at each dose, with the filter and Hilbert transform run only on that clip (Appendix C.1).

Baselines and controls. We compare the joint estimator with the following frozen alternatives: coefficients fitted separately at each positive dose and then averaged; a single-dose model at $d = .05$; the fitted nuisance coefficient c; a one-repeat version of CHIRP; path-averaged phase locking, absolute voltage correlation, and partial correlation; and a zero-dose weak model. Voltage correlation and the zero-dose model were included in the frozen protocol; their held-out numbers appear in Table 1 and Appendix C. A two-stage slope of separately fitted coefficients on dose was computed after the confirmatory run as a diagnostic, not as a frozen endpoint. To test whether the assigned dose ordering, rather than the mere pooling of traces, drives the ranking, we also permute dose labels 20 times for each network. Prefixes of the ordered dose path with two through six settings measure how much of the intervention path is needed. Paired one-sided Wilcoxon tests are performed across held-out graphs, and the primary 95% interval is obtained by bootstrapping graphs 20,000 times.

Frozen success criterion. Before evaluation, we required four conditions for advancing the main claim: held-out mean AP above .95, a bootstrap lower bound above .90, AP higher than both the single-dose and separate-fit baselines, and mean Otsu precision and recall above .85. These criteria were committed before the held-out data were analyzed.

Table 1: Held-out physical topology reconstruction on 15 graphs. Time is unique voltage duration. AP is the primary ranking metric; top-42 F1 is diagnostic because it uses the true graph density. Values are graph means.

Method	Time (s)	AP ↑	AUROC ↑	Top-42 F1 ↑
CHIRP	18	**0.9980**	**0.9997**	**0.9810**
CHIRP (3 s)†	3	0.9350	0.9835	0.8714
One-repeat CHIRP	6	0.9799	0.9939	0.9349
Separate-fit homotopy	15	0.9012	0.9583	0.8524
Two-stage slope†	18	0.8770	0.9405	0.8143
Two-stage, standardized†	18	0.9209	0.9728	0.8651
Single dose $d = .05$	3	0.8814	0.9444	0.8222
Dose-independent nuisance	18	0.2034	0.6318	0.2206
Path phase locking	15	0.3407	0.7809	0.3016
Path voltage correlation	15	0.2703	0.5910	0.2492
Path partial correlation	15	0.2998	0.6460	0.2810
Zero-dose weak model	3	0.1136	0.4659	0.0825

†Computed after the frozen run; not a frozen endpoint. The 3 s fit cuts the raw voltage to 0.5 s per dose before filtering. The unstandardized two-stage slope does not scale columns and uses a 20-value ridge grid. The standardized row uses the joint fit's column scaling and 23-value grid on the same 18 s recordings.

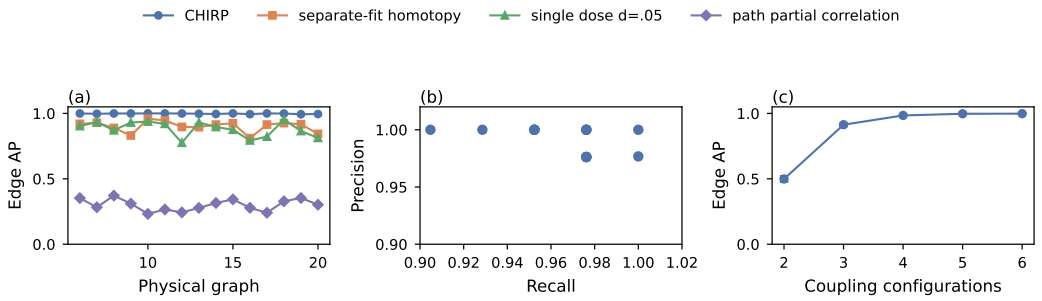

Figure 2: Held-out physical results. (a) Graph-level AP. (b) Precision and recall of the label-free Otsu graph. (c) Mean intervention-query curve with standard errors across graphs.

6 Physical Results

Primary endpoint. On the fifteen held-out graphs, CHIRP obtains mean AP 0.9980 (95% graph-bootstrap CI $[0.9968, 0.9990]$), mean AUROC 0.9997, and top-42 F1 0.9810 (Table 1). These values use the frozen 18 s protocol. Figure 2 separates that average into the graph-level ranking, the label-free Otsu graph, and the gain from adding coupling configurations.

The equal unique-time comparison is against the 3 s single-dose model (Figure 3). Cutting the raw voltage to 0.5 s at each dose before filtering gives AP 0.9350 (CI $[0.9184, 0.9513]$), versus 0.8814 for the single-dose baseline. The paired difference is positive on 13 of 15 graphs (one-sided exact Wilcoxon $p = 1.01 \times 10^{-3}$). The original 4,158 versus 4,155 row match is therefore not an equal-time comparison: it equalized regression rows by sampling 18 s six times more sparsely than 3 s. Using the nuisance coefficient alone gives mean AP 0.2034. Absolute voltage correlation and the zero-dose model give mean AP 0.2703 and 0.1136. Per-graph results and paired tests are reported in Appendix C.

All four pre-specified advance criteria are satisfied. The AP difference between the frozen 18 s CHIRP estimator and each declared baseline is positive on all 15 graphs; the one-sided exact Wilcoxon test gives $p = 3.05 \times 10^{-5}$ for each comparison.

A graph without knowing its size. Otsu selects 40.9 edges on average, at a mean score cutoff 146.0 (range $[127.3, 164.6]$), and gives precision 0.9921, recall 0.9651, and F1 0.9782. The Otsu graph therefore does not depend on supplying the true count of 42 edges. The frozen log-Gaussian

mixture rule was not replaced after the confirmation results were known. It uses a two-component fit on log scores and assigns a pair to the high component when its responsibility is at least $1/2$. On the held-out graphs, this rule corresponds to a mean score cutoff 1.257 (range $[0.348, 1.980]$). It selects 95.6 edges on average, with precision 0.4605, recall 1.0000, and F1 0.6243. Per-graph cutoffs are in Appendix C.

Dose ordering and intervention budget. Permuting the dose labels reduces mean AP to 0.4694. The observed ordering outperforms all 20 permutations on 15 of the 15 graphs. This control addresses the possibility that pooling several weak trajectories, rather than aligning them with the dose, explains the gain. The query curve in Figure 2 shows how performance changes as dose configurations are added. With one repeat, CHIRP obtains AP 0.9799. This ablation isolates the contribution of repeated traces. Mean AP increases from 0.4988 with the two smallest doses to 0.9136, 0.9844, 0.9974, and 0.9980 with three through six configurations, respectively.

Joint estimator versus two-stage slope. Fitting each environment separately, without column standardization, and regressing the fitted coefficients on dose gives AP 0.8770, compared with 0.9980 for the joint estimator; all 15 paired differences are positive ($p = 3.05 \times 10^{-5}$). That two-stage fit also uses a shorter ridge grid than the joint estimator, so the gap mixes the estimator with the penalty scaling. Repeating the slope on the same 18 s recordings, with each recording's columns divided by their standard deviation and with the joint fit's 23-value ridge grid, raises mean AP to 0.9209. The joint estimator is higher on all 15 held-out graphs. In Appendix Table 6 the gap runs from 0.0030 on R11 to 0.1805 on R12. Proposition 4 shows that the unregularized procedures coincide when the Grams share one shape. That condition is sufficient, not necessary. The physical designs do not meet it: the trace-normalized Grams differ by mean pairwise Frobenius gap 0.154, and each separate fit still chooses its own ridge. The remaining gap is that change of estimator, not a contradiction of the algebra that identifies b. Dropping the dose-independent block from the joint design, and keeping the column scaling and the ridge grid, gives mean AP 0.9980, the same reported value as the joint estimator. Across the 15 held-out graphs that score is higher on 7, lower on 4, and unchanged on 4.

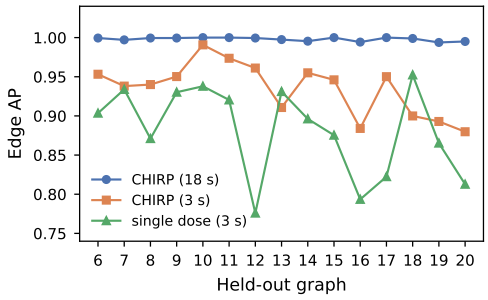

Figure 3: Equal unique-time comparison. Full-path CHIRP uses 18 s. The 3 s curve cuts raw voltage before filtering.

Post-confirmation sensitivities. The 3 s comparison, the two-stage slope, and the fit without the dose-independent block are exploratory and do not replace the primary 18 s endpoint. Using only the endpoint pair $(0, .05)$ gives AP 0.9624 when the same six traces and stride are used, compared with 0.4988 for the two-smallest-dose configuration. Omitting one of the six doses gives AP between 0.9890 and 0.9979.

A published analysis of the same oscillators reports about 0.96 mean AUROC in the weak-coupling region (Singhal et al., 2026). The implementation and the score differ, and the article materials we inspected did not include source code, so the comparison in this paper is average precision on shared raw files.

7 Synthetic assumption stress test

The physical benchmark is unusually favorable to a coupling-dose interpretation. We therefore examine the boundary of the method on 45 generated 28-node networks. The graphs are an Erdős–Rényi draw, a ring with added chords, and a star with added edges among the remaining nodes, five seeds of each, under three data-generating scenarios. In the *matched* scenario, the data follow the affine sine-response model. The *confounded* scenario adds a common drive whose amplitude is the same at every dose, together with a second harmonic, observation noise, and waveform distortion. In

the *nonlinear* scenario, the coupling gain is also nonlinear in dose. Hyperparameters are held fixed across all scenarios.

The decrease in AP from 1.000 to .642 and then .587 is a substantive loss, not a small robustness penalty. The edges are recovered when the data follow the assumed intervention model. Average precision falls when a shared drive, a second harmonic, and waveform distortion are added, and falls again when the coupling gain is nonlinear in dose. A known dose is not automatically a valid structural intervention. If it changes a common driver or a sensor artifact, the fitted slope b combines that effect with the effect of the physical edges. Extending the dose range can create another problem by making the local affine approximation inaccurate. Nonlinear dose bases, negative-control channels, and multiple independently targeted interventions are reasonable ways to address these issues, but none is evaluated here.

Table 2: Synthetic stress test. Each cell is the mean edge AP across 15 networks, with a 95% network-bootstrap interval.

Method	Matched	Confounded	Nonlinear
CHIRP	**1.000** [1.000, 1.000]	**.642** [.543, .738]	**.587** [.515, .656]
Single dose	.471 [.359, .588]	.196 [.164, .231]	.169 [.135, .206]
Separate slopes	.542 [.424, .659]	.235 [.194, .278]	.207 [.164, .255]
Phase locking	.447 [.380, .507]	.416 [.373, .456]	.471 [.421, .517]

8 Discussion and Limitations

CHIRP ranks edges by the part of the phase-difference coefficient that moves with a known dose. At one dose that part is collinear with the dose-independent term. With dose variation the two can be separated under a rank condition. The held-out circuits ask whether this survives chaotic electronics, filtering, and phase reduction.

The physical evidence is narrow. Every graph has 28 nodes and 42 undirected edges on one device family, the dose is one scalar, and symmetrization drops direction. A disturbance whose strength itself moves with the dose is not tested. A high AP does not mean the coupling magnitudes are accurate.

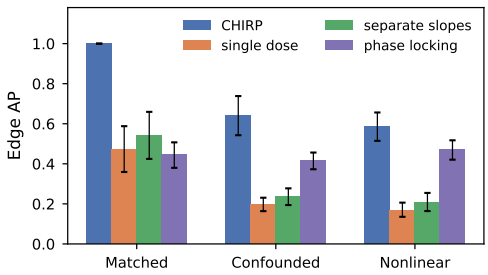

Figure 4: Synthetic stress test. Bars are mean edge AP; whiskers are 95% network-bootstrap intervals.

The frozen 18 s result is not a 3 s comparison. Under an equal 3 s raw-time budget, with the voltage cut before filtering, mean AP falls from 0.9980 versus 0.8814 to 0.9350 versus 0.8814. The paired difference is positive on 13 of 15 graphs. Permuting the dose labels lowers mean AP to 0.4694, and the recorded order wins on all 15 graphs, so the gain is not explained by pooling weak traces.

Otsu's rule cuts the ranking without being told that each network has 42 edges. It selects 40.9 edges on average, near that count, at precision 0.9921 and recall 0.9651. The frozen mixture rule is also label-free, but it selects 95.6 edges on average: recall stays 1.0000 while precision falls to 0.4605. That rule was not replaced after the held-out results were known. Synthetic average precision falls in the confounded scenario, which adds a constant-amplitude common drive together with a second harmonic, noise, and waveform distortion, and falls again when the dose response is nonlinear. What is identified is undirected topology among the measured nodes, for a scalar dose whose effect stays approximately affine on the fitted interval.

Reproducibility Statement

The full reproduction record is in Appendix D, including the public recordings, the locked analysis choices, the four runs, the stored artifacts, and SHA-256 hashes of the analysis modules.

ETHICS STATEMENT

This study uses publicly available measurements from electronic oscillator circuits and synthetic data. It does not involve human participants, personal data, or animals. The main ethical issue is the possibility of overstating computational results. We address this by distinguishing the development and confirmation data, reporting the settings in which the method fails, and acknowledging earlier work based on the same benchmark.

AI-USE STATEMENT

OpenAI Codex and Grammarly were used during manuscript preparation for grammar checking, prose editing, and identifying syntax errors in code-related text. We reviewed the suggested changes rather than accepting them automatically. We also checked the cited literature, tested the software, repeated the analyses, and retained the negative results. We take responsibility for the mathematical derivations, implementation, citations, data analysis, and scientific conclusions. The use of these tools was limited to writing and code-review assistance and did not replace our responsibility for the work.

REFERENCES

Juan A. Acebrón, L. L. Bonilla, Conrad J. Pérez Vicente, Félix Ritort, and Renato Spigler. The Kuramoto model: A simple paradigm for synchronization phenomena. *Reviews of Modern Physics*, 77:137–185, 2005. doi: 10.1103/RevModPhys.77.137.

Jose Casadiego, Mor Nitzan, Sarah Hallerberg, and Marc Timme. Model-free inference of direct network interactions from nonlinear collective dynamics. *Nature Communications*, 8:2192, 2017. doi: 10.1038/s41467-017-02288-4.

C. W. J. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3):424–438, 1969.

Seung-Yeal Ha and Jinwook Jung. Remarks on the slow relaxation for the fractional Kuramoto model for synchronization. *Journal of Mathematical Physics*, 59(3):032702, 2018. doi: 10.1063/1.5005865.

Seung-Yeal Ha, Jinwook Jung, and Peter Kuchling. Emergent behaviors in the fractional Kuramoto model with general network topologies and its hybridization. *Discrete and Continuous Dynamical Systems - B*, 29(3):1427–1452, 2024. doi: 10.3934/dcdsb.2023139.

Christina Heinze-Deml, Jonas Peters, and Nicolai Meinshausen. Invariant causal prediction for nonlinear models. *Journal of Causal Inference*, 6(2), 2018. doi: 10.1515/jci-2017-0016.

Gyuyoung Hwang, Hyeontae Jo, and Jae Kyoung Kim. Data-driven inference of coupling functions in oscillatory systems via physics-informed neural networks. *Physical Review Research*, 8:013259, 2026. doi: 10.1103/t8g2-n22v.

Boris N. Kholodenko, Anatoly Kiyatkin, Frank J. Bruggeman, Eduardo Sontag, Hans V. Westerhoff, and Jan B. Hoek. Untangling the wires: A strategy to trace functional interactions in signaling and gene networks. *Proceedings of the National Academy of Sciences*, 99(20):12841–12846, 2002. doi: 10.1073/pnas.192442699.

Bjoern Kralemann, Arkady Pikovsky, and Michael Rosenblum. Reconstructing phase dynamics of oscillator networks. *Chaos*, 21:025104, 2011. doi: 10.1063/1.3597647.

Yoshiki Kuramoto. *Chemical Oscillations, Waves, and Turbulence*. Springer, Berlin, 1984.

Akari Matsuki, Hiroshi Kori, and Ryota Kobayashi. Network inference applicable to both synchronous and desynchronous systems from oscillatory signals. *PLOS Complex Systems*, 2(9):e0000063, 2025. doi: 10.1371/journal.pcsy.0000063.

Nobuyuki Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1):62–66, 1979. doi: 10.1109/TSMC.1979.4310076.

MoonJeong Park, Youngbin Choi, and Dongwoo Kim. SpReME: Sparse regression for multi-environment dynamic systems. In *AAAI Workshop on When Machine Learning Meets Dynamical Systems: Theory and Applications*, 2023. URL https://openreview.net/forum?id=arD9T52XO2S.

Mihály Petreczky, Roland Tóth, and Guillaume Mercère. LPV-ARX representations of LPV state-space models with affine dependence. *Systems & Control Letters*, 173:105459, 2023. doi: 10.1016/j.sysconle.2023.105459.

Niklas Pfister, Stefan Bauer, and Jonas Peters. CausalKinetiX: Causal inference from dynamics with structured interventions. *Bioinformatics*, 35(14):i213–i221, 2019. doi: 10.1093/bioinformatics/btz325.

Bastian Pietras and Andreas Daffertshofer. Network dynamics of coupled oscillators and phase reduction techniques. *Physics Reports*, 819:1–105, 2019. doi: 10.1016/j.physrep.2019.06.001.

Arkady Pikovsky, Michael Rosenblum, and Jürgen Kurths. *Synchronization: A Universal Concept in Nonlinear Sciences*. Cambridge University Press, 2001.

Michael G. Rosenblum and Arkady S. Pikovsky. Detecting direction of coupling in interacting oscillators. *Physical Review E*, 64:045202, 2001. doi: 10.1103/PhysRevE.64.045202.

Dominik Rothenhäusler, Nicolai Meinshausen, Peter Bühlmann, and Jonas Peters. Anchor regression: Heterogeneous data meet causality. *Journal of the Royal Statistical Society: Series B*, 83(2): 215–246, 2021. doi: 10.1111/rssb.12398.

Takaya Saito and Marc Rehmsmeier. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLOS ONE*, 10(3):e0118432, 2015. doi: 10.1371/journal.pone.0118432.

Thomas Schreiber. Measuring information transfer. *Physical Review Letters*, 85(2):461–464, 2000. doi: 10.1103/PhysRevLett.85.461.

Bharat Singhal, Minh Vu, Shen Zeng, and Jr-Shin Li. Data-efficient inference of nonlinear oscillator networks. *IFAC-PapersOnLine*, 56(2):10089–10094, 2023. doi: 10.1016/j.ifacol.2023.10.879.

Bharat Singhal, István Z. Kiss, and Jr-Shin Li. NIPS: Network inference with partial state measurements using forced-delay embedding. *PNAS Nexus*, 5(1):pgaf397, 2026. doi: 10.1093/pnasnexus/pgaf397.

Tomislav Stankovski, Tiago Pereira, Peter V. E. McClintock, and Aneta Stefanovska. Coupling functions: Universal insights into dynamical interaction mechanisms. *Reviews of Modern Physics*, 89:045001, 2017. doi: 10.1103/RevModPhys.89.045001.

Weiwei Su, Shigefumi Hata, Hiroshi Kori, Hiroya Nakao, and Ryota Kobayashi. Distinguishing pairwise and higher-order interactions in coupled oscillators from time series. *Communications Physics*, 9:38, 2026. doi: 10.1038/s42005-025-02467-6.

Marc Timme and Jose Casadiego. Revealing networks from dynamics: an introduction. *Journal of Physics A: Mathematical and Theoretical*, 47(34):343001, 2014. doi: 10.1088/1751-8113/47/34/343001.

V. P. Vera-Ávila, Ricardo Sevilla-Escoboza, Joaquín Goñi, R. R. Rivera-Durón, and Javier M. Buldú. Identifiability of structural networks of nonlinear electronic oscillators. *Scientific Reports*, 10: 14668, 2020a. doi: 10.1038/s41598-020-71373-4.

V. P. Vera-Ávila, Ricardo Sevilla-Escoboza, A. A. Lozano-Sánchez, R. R. Rivera-Durón, and Javier M. Buldú. Experimental datasets of networks of nonlinear oscillators: Structure and dynamics during the path to synchronization. *Data in Brief*, 28:105012, 2020b. doi: 10.1016/j.dib.2019.105012.

Wen-Xu Wang, Ying-Cheng Lai, and Celso Grebogi. Data based identification and prediction of nonlinear and complex dynamical systems. *Physics Reports*, 644:1–76, 2016. doi: 10.1016/j.physrep.2016.06.004.

Zuogong Yue, Johan Thunberg, Wei Pan, Lennart Ljung, and Jorge Gonçalves. Dynamic network reconstruction from heterogeneous datasets. *Automatica*, 123:109339, 2021. doi: 10.1016/j.automatica.2020.109339.

A NOTATION

Tables 3 and 4 list every symbol used in the main text and in this appendix. The physical benchmark has $N = 28$ nodes and 42 undirected edges. Those two integers are protocol facts, not estimated parameters.

Table 3: Notation for the observations, the phase model, and the joint estimator.

Symbol	Meaning
N	Number of observed nodes.
A	Unknown undirected adjacency matrix in $\{0,1\}^{N\times N}$, with no self-loops.
i, j	Node indices.
e	Environment index; one coupling-dose setting.
r	Repeat index of a one-second recording at a fixed dose.
t	Time.
$v_i^{(e,r)}(t)$	Scalar voltage of node i in environment e and repeat r.
ϕ_i, $\phi_i^{(e)}$	Unwrapped Hilbert phase after the 700–1800 Hz bandpass.
Δt	Sampling interval ($1/30{,}000$ s in the physical data).
W	Differentiation window in samples ($W = 300$).
d_e	Known scalar coupling dose of environment e.
α_{ie}	Environment-specific intercept for receiver i; removed by within-environment centering.
c_{ij}	Dose-independent phase-difference coefficient from j to i.
b_{ij}	Dose-response coefficient (intervention slope) from j to i.
$\epsilon_i^{(e)}$	Residual of the reduced phase model.
$y_{ier}(t)$	Windowed phase increment (target velocity) of receiver i.
$x_{ij,er}(t)$	Feature $\sin(\phi_j(t) - \phi_i(t))$.
n_e	Number of regression rows in environment e.
X_e	Centered phase-difference design of size $n_e \times (N-1)$.
y_e	Centered response of the same length.
Z_e	Joint design $[X_e \;\; d_e X_e]$.
ϑ_i	Stacked coefficient $[c_i^\top \; b_i^\top]^\top \in \mathbb{R}^{2(N-1)}$.
λ_i	Ridge penalty for receiver i.
s_{ij}	Undirected edge score $\frac{1}{2}(\lvert\widehat{b}_{ij}\rvert + \lvert\widehat{b}_{ji}\rvert)$.
D_i	Diagonal matrix of column standard deviations of receiver i's stacked refit design. Entries below 10^{-10} are set to 1.
λ_e	Ridge penalty selected inside recording e for the two-stage first stage.
$\bar{d}$	Unweighted mean of the eighteen recording doses.

B COMPLETE PROTOCOL AND PROOF SCOPE

Section 5 repeats the frozen advance criteria. No frozen endpoint was added after R6–R20 had been inspected. The equal-Gram assumption in Proposition 1 is used only for the closed-form rank condition; with unequal Grams, full column rank of the stacked Z remains the necessary and sufficient condition for unregularized identification. Proposition 3 is a deterministic coefficient bound. It does not treat temporally dependent rows as independent observations, so inferential resampling is performed over graphs. Proposition 4 gives a sufficient condition under which a two-stage slope recovers the joint estimator. The physical oscillator designs do not meet that condition.

Table 4: Notation for identifiability statements and for label-free evaluation.

Symbol	Meaning
w_e	Relative weight, or row count, of environment e.
G, G_e	Phase-difference Gram; $G_e = X_e^\top X_e$. The closed-form rank identity uses $G_e = w_e G$.
Z	Stacked joint design $[Z_1^\top, \ldots, Z_E^\top]^\top$.
$\bar{d}_w$	Weighted mean dose $\sum_e w_e d_e / \sum_e w_e$.
S_k	Dose moment $\sum_e w_e d_e^k$.
σ^2	Homoskedastic residual variance in the unregularized bound.
$\vartheta^\star$	Population stacked coefficient.
ϵ	Residual vector in $y = Z\vartheta^\star + \epsilon$.
$\widehat{\beta}_e$	Environment-wise unregularized coefficient $(X_e^\top X_e)^{-1} X_e^\top y_e$.
$\widehat{\beta}_e(\lambda_e)$	Ridge coefficient of recording e, $(X_e^\top X_e + \lambda_e I)^{-1} X_e^\top y_e$. This is not the unregularized $\widehat{\beta}_e$.
δ	Uniform bound on directed slope error in the support corollary.
$s^\star_{\min}$	Smallest true symmetrized edge score.
τ	Label-free score cutoff (Otsu or mixture).
$\widehat{E}$	Edge set selected by a label-free rule.
AP	Edge average precision, macro-averaged over held-out graphs.
AUROC	Edge area under the ROC curve, likewise macro-averaged.

B.1 Dose variation

Figure 5 checks Proposition 1 on six equally weighted doses. A spread of zero pulls every dose to the common mean, so the design is a single dose. The dose-moment matrix is then singular, and its smallest eigenvalue is zero. Spreading the doses back to $d \in \{0, .01, .02, .03, .04, .05\}$ makes that eigenvalue positive. The weighted dose variance in Equation 5 vanishes at the same point and is positive once the doses differ. The two panels are the same condition drawn two ways. This construction assumes equal Grams. It does not use the physical recordings and is not a held-out endpoint.

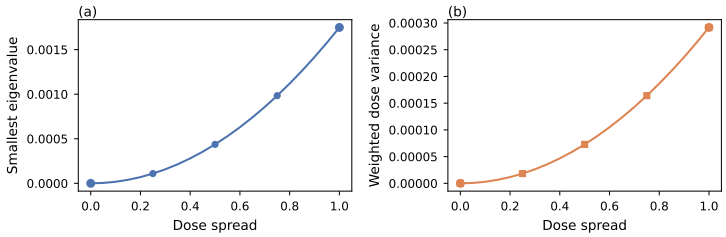

Figure 5: Dose variation and rank, for Proposition 1. Left: smallest eigenvalue of the 2×2 dose-moment matrix. Right: weighted dose variance. Both are zero when the doses collapse to one value, and both are positive at the six-dose protocol.

The equality $X_e^\top X_e = w_e G$ turns the stacked Gram into a Kronecker product. One environment contributes

$$Z_e^\top Z_e = \begin{bmatrix} w_e & w_e d_e \\ w_e d_e & w_e d_e^2 \end{bmatrix} \otimes G.$$

Summing over environments gives

$$Z^\top Z = \begin{bmatrix} S_0 & S_1 \\ S_1 & S_2 \end{bmatrix} \otimes G, \qquad S_k = \sum_e w_e d_e^k.$$

The eigenvalues of a Kronecker product are products of the two factors' eigenvalues. Because $G \succ 0$, the stacked design has full column rank exactly when the 2×2 moment matrix does. Its determinant factors as

$$S_0 S_2 - S_1^2 = S_0 \sum_e w_e (d_e - \bar{d}_w)^2.$$

The weight sum S_0 is positive, so the determinant is positive exactly when the weighted dose variance is positive. If every dose equals $\bar{d}_w$, then $d_e X_e$ is one common multiple of X_e, and the blocks that carry c and b are scalar multiples of each other.

B.2 Endpoint allocation

Figure 6 checks Proposition 2 on the interval $[0, .05]$. The quantity on the horizontal axis is $\sum_e w_e(d_e - \bar{d}_w)^2$. The covariance of $\widehat{b}$ is small when this factor is large. Putting equal weight on the two endpoints attains the maximum, $(d_{\max} - d_{\min})^2/4$. Four interior doses, and the six-dose protocol itself, sit strictly below that maximum. A single midpoint has factor zero: c and b are not separated. The physical protocol keeps the interior doses because the endpoint optimum assumes that the affine approximation still holds. This is a comparison of weights, not a held-out endpoint.

Fix the total weight at one and restrict every dose to $[d_{\min}, d_{\max}]$. Block inversion of the moment matrix makes $\mathrm{Cov}(\widehat{b})$ a positive multiple of the reciprocal of $\sum_e w_e(d_e - \bar{d}_w)^2$. A random variable on an interval of length $L = d_{\max} - d_{\min}$ has variance at most $L^2/4$. Equality holds when half the weight sits at each endpoint and nowhere else. Any dose placed strictly inside the interval lowers the variance factor, and therefore raises the covariance of $\widehat{b}$ in the Loewner order. The midpoint allocation in Figure 6 is the extreme: the factor is zero and b is not identified. The six-dose protocol spends weight inside the interval, so its factor is smaller than the endpoint maximum $(.05 - 0)^2/4$.

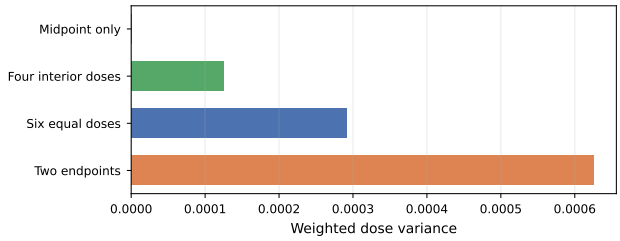

Figure 6: Dose allocations for Proposition 2. The axis is reversed: allocations run down the vertical axis and the weighted dose variance runs along the horizontal axis. The bottom bar is equal weight on the endpoints of $[0, .05]$. It has the largest weighted dose variance. The six-dose protocol is the third bar from the top. The midpoint allocation does not identify b.

Figure 7 checks the coefficient bound in Proposition 3 on a constructed one-receiver design at the six coupling doses $d \in \{0, .01, .02, .03, .04, .05\}$. Columns are scaled by their standard deviations, with the same floor of 10^{-10} as the joint estimator. The penalties are the 23 log-spaced values in $[10^{-4}, 10^{6}]$. The plotted error is $\|D(\widehat{\vartheta} - \vartheta^\star)\|_2$. For a random residual it lies on or below the right-hand side of Equation 7. The bound is attained when $D\vartheta^\star$ lies along the smallest eigenvector of $\tilde{Z}^\top \tilde{Z}$ and $\tilde{Z}^\top \epsilon$ is a nonpositive multiple of that vector. This check does not use the physical recordings and is not a held-out endpoint.

Set $u = D\vartheta$ and $\tilde{Z} = ZD^{-1}$. The objective in Equation 3 is ordinary ridge regression of y on $\tilde{Z}$. Its normal equation is $(\tilde{Z}^\top \tilde{Z} + \lambda I)\widehat{u} = \tilde{Z}^\top y$. Substitute $y = \tilde{Z}u^\star + \epsilon$, with $u^\star = D\vartheta^\star$, and move the population term to the left:

$$(\tilde{Z}^\top \tilde{Z} + \lambda I)(\widehat{u} - u^\star) = \tilde{Z}^\top \epsilon - \lambda u^\star.$$

The inverse of the left-hand matrix has spectral norm $1/(\lambda_{\min}(\tilde{Z}^\top \tilde{Z}) + \lambda)$. Taking Euclidean norms and using the triangle inequality on the right-hand side produces Equation 7, because $\|D(\widehat{\vartheta} - \vartheta^\star)\|_2 = \|\widehat{u} - u^\star\|_2$. The inequality is an equality when $u^\star$ is aligned with the smallest eigenvector of $\tilde{Z}^\top \tilde{Z}$ and $\tilde{Z}^\top \epsilon$ points in the opposite direction along that same vector, which is the sharp residual in Figure 7.

B.3 Support separation

Figure 8 checks the support corollary on constructed scores. Each directed slope is moved by the full error allowance, in the direction that makes separation hardest: nonedges are raised and edges are

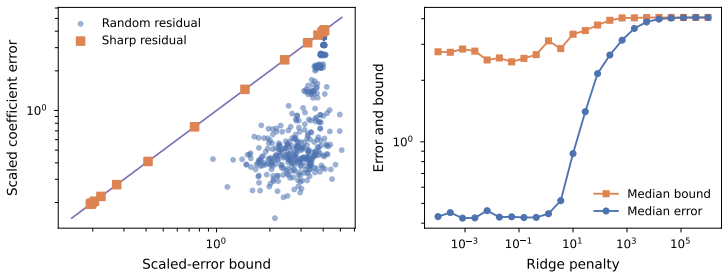

Figure 7: Finite-noise coefficient bound from Proposition 3. The error is the standardized coefficient error in Equation 7. Left: that error against the bound; the line is equality. Random residuals lie below the line, and sharp residuals lie on it. Right: median bound and median realized error over random residuals, against the ridge penalty. The median bound stays above the median error, and the two approach one another at large penalties.

lowered. The true edge score is larger than twice that allowance, so the worst-case nonedge still sits strictly below the worst-case edge. The shaded band is every threshold strictly between those two scores. Any threshold in the band keeps the nonedges out and the edges in. The figure does not run Otsu's rule, it does not use the physical recordings, and it is not a held-out endpoint.

Suppose $|\widehat{b}_{ij} - b^\star_{ij}| \le \delta$ and the same bound holds in the opposite direction. A nonedge has $b^\star_{ij} = b^\star_{ji} = 0$, so each absolute estimated coefficient is at most δ and the symmetrized score is at most δ. An edge whose true symmetrized score is $s^\star$ loses at most δ in each direction, so its estimated score is at least $s^\star - \delta$. If the smallest true edge score satisfies $s^\star_{\min} > 2\delta$, then $s^\star_{\min} - \delta > \delta$. The worst nonedge and the worst edge no longer meet. Every threshold strictly between them selects all of the edges and none of the nonedges.

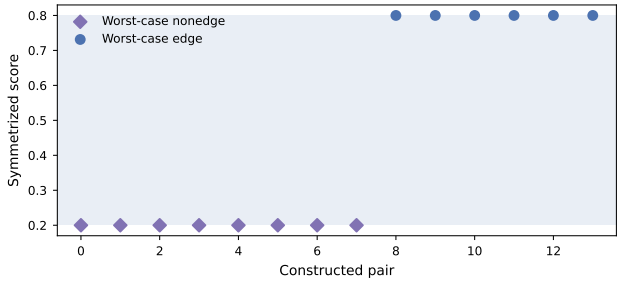

Figure 8: Support separation under a uniform directed-error allowance. Diamonds are worst-case nonedges. Circles are worst-case edges. The shaded band is the set of thresholds that recover the constructed support exactly.

B.4 When the two slopes agree

Figure 9 checks Proposition 4 on one constructed receiver. With the same phase-difference features at every dose, a residual does not separate the two slopes: the joint slope and the unweighted two-stage slope stay on the diagonal. Scaling the features with the dose moves the two-stage slope off that line. A separate ridge in each dose does the same, even when the features are shared and the residual is zero. The physical designs do both: the Grams change with dose, and each separate fit selects its own penalty. This check is not a held-out endpoint.

Equation 8 is generalized least squares of $\widehat{\beta}_e = c + d_e b$ with precision $G_e = X_e^\top X_e$. The three series in the figure separate the two conditions in that proposition.

If $G_e = wG$ for one matrix G and one common weight w, then G cancels from Equation 8. What remains is ordinary least squares of $\widehat{\beta}_e$ on d_e. A residual does not spoil the cancellation, because both the joint estimator and the two-stage slope see the same $\widehat{\beta}_e$. The circles stay on the diagonal.

If the features grow with the dose, G_e is not a scalar multiple of one matrix. The precision weights stay inside the normal equations. Unweighted least squares of the noisy environment-wise coefficients is then a different slope. The squares leave the diagonal.

A penalty λ_e chosen inside recording e replaces G_e by $G_e + \lambda_e I$ before $\widehat{\beta}_e$ is computed. The triangles keep one feature matrix and use no residual. Different penalties are enough to leave the diagonal. A shared Gram shape and one unregularized solve are sufficient for the identity. They are not necessary: with two environments both estimators interpolate the two coefficient vectors, so the slopes agree for any invertible Grams.

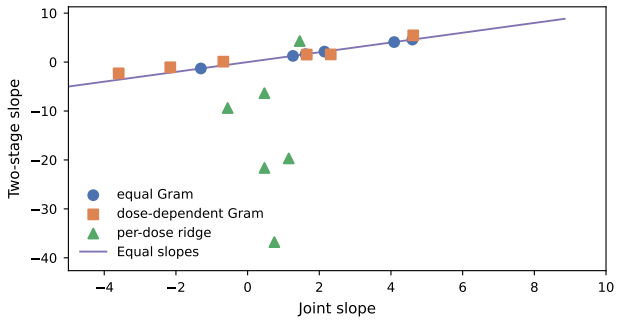

Figure 9: Joint slope against the two-stage slope, for Proposition 4. Circles use a common Gram and lie on the diagonal. Squares use a Gram that grows with dose. Triangles keep the common Gram but give each dose its own ridge penalty.

B.5 Trace-normalized Gram gap

Proposition 4 gives a sufficient condition for the two slopes to coincide. Grams that are not scalar multiples of one matrix can move them apart, but they need not. On each physical recording e and receiver i, let X_{ie} be the phase-difference features on every window of that recording, centered by the mean of those rows. This is the full windowed design, rather than the first 80% used in the joint refit. The recording Gram is the average of those receiver Grams, scaled by its trace:

$$G_e = \frac{1}{N}\sum_{i=1}^{N} X_{ie}^{\top} X_{ie}, \qquad \tilde{G}_e = G_e / \operatorname{tr}(G_e).$$

A trace below 10^{-15} is replaced by 10^{-15}. The reported gap is the mean Frobenius distance $\|\tilde{G}_e - \tilde{G}_{e'}\|_F$ over all unordered pairs among the eighteen recordings. That average is 0.154.

The estimator, preprocessing, ridge grids, baselines, metrics, seed, label-free rules, and advance criteria were locked on R1–R5, and each held-out edge list was opened only after every score for that graph had been built. Every recording drops 1000 samples at both ends. Separate-fit and single-dose regressions use a 20-value grid on $[10^{-4}, 10^{5}]$, and the undirected score is $s_{ij} = \frac{1}{2}(|\widehat{b}_{ij}| + |\widehat{b}_{ji}|)$.

B.6 From voltage to the phase increment

Each recording is a voltage trace sampled at 30,000 Hz. A fourth-order Butterworth band-pass, with cutoffs 700 Hz and 1800 Hz, is applied forward and backward, so the phase of the filter is zero. The Hilbert transform of that filtered trace supplies an analytic signal. Its angle is unwrapped along time and becomes $\phi_i(t)$.

Windows are then taken from the interior of the trace. The first and last 1000 samples are dropped, and each window is $W = 300$ samples long. The next window starts 120 samples later for the joint

estimator, and 20 samples later for a single-dose fit. The target on a window is the average phase derivative in Equation 2. The sine features are read at the window's midpoint, not at its endpoint. Nothing in this step removes α_{ie}. A constant drift in phase is a constant derivative, and the windowed quotient keeps that constant.

B.7 PATH PARTIAL CORRELATION

Path partial correlation is computed from the trimmed voltage. For one recording, drop 1000 samples at each end, form the Pearson correlation R across nodes, and let Θ be its Moore–Penrose inverse. The score matrix is

$$\Pi_{ij} = \left| \frac{\Theta_{ij}}{\sqrt{\Theta_{ii}\Theta_{jj}}} \right|,$$

with the diagonal set to 1. The path score is the mean of Π over the fifteen recordings at the positive doses $d \in \{.01, .02, .03, .04, .05\}$. The zero-dose recordings are left out of this average. Path phase locking and path voltage correlation use those same fifteen recordings. Phase locking is the absolute mean resultant length of the trimmed analytic phase, and voltage correlation is the absolute Pearson correlation of the trimmed voltage.

B.8 REMOVING THE ENVIRONMENT INTERCEPT

Equation 1 is an equation for the phase derivative. The regression does not use a sample-wise derivative. For a window of $W = 300$ samples starting at time t, the target is the forward increment

$$y_{ie}^{\mathrm{raw}}(t) = \frac{\phi_i(t+W) - \phi_i(t)}{W\Delta t},$$

and the phase-difference feature is read at the middle of that window. That increment is the average of $\dot{\phi}_i$ across the window. A constant α_{ie} averages to itself, so the window keeps the intercept. Equation 2 is the same quotient written from the window's midpoint. Removing α_{ie} is a later step.

On the windows that enter one solve the averaged equation is the constant term

$$y_e^{\mathrm{raw}} = \alpha_{ie}\mathbf{1} + X_e^{\mathrm{raw}}\beta_e + \epsilon_e, \qquad \beta_e = c_i + d_e b_i. \tag{9}$$

The intercept does not depend on time inside that recording. Let $\bar{y}_e$ be the mean of the target on the rows used in the solve, and let $\bar{x}_e^\top$ be the row of feature means on those same rows. Those means include the intercept,

$$\bar{y}_e = \alpha_{ie} + \bar{x}_e^\top \beta_e + \bar{\epsilon}_e.$$

Subtracting them from every row cancels α_{ie}:

$$y_e = X_e\beta_e + \epsilon_e^{\mathrm{c}}, \tag{10}$$

with $y_e = y_e^{\mathrm{raw}} - \bar{y}_e\mathbf{1}$ and $X_e = X_e^{\mathrm{raw}} - \mathbf{1}\bar{x}_e^\top$. The centered residual is ϵ_e with its mean removed. No column is retained for α_{ie}, and the coefficient vector is not asked to estimate it.

Because $\beta_e = c_i + d_e b_i$, the centered features split into the two blocks of the joint design,

$$X_e\beta_e = X_e c_i + d_e X_e b_i.$$

Centering is applied before the dose multiplies the second block, which is why $Z_e = [X_e \;\; d_e X_e]$ uses the centered X_e in both halves. The mean is computed inside the recording. An average taken across doses would mix unequal intercepts, and α_{ie} would not cancel.

Penalty selection and the later refit do not share one mean. Selection centers on the first 60% of the recording and applies that mean to the following validation rows. The refit throws that mean away, recomputes it on the first 80%, and solves only on those rows. In both solves the subtracted mean is the mean of the rows that define it, so the cancellation above applies to the rows in the normal equation.

A three-row arithmetic check makes the cancellation visible. Take $\alpha = 4$, $\beta = (2, -1)$, and

$$X^{\mathrm{raw}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}, \qquad y^{\mathrm{raw}} = \begin{bmatrix} 6 \\ 3 \\ 5 \end{bmatrix}.$$

The feature mean is $(2/3, 2/3)$ and the target mean is $14/3$. After subtracting those means, $X\beta$ equals the centered target on every row, and a regression with no intercept column returns β. The value 4 does not appear in the centered system. The same arithmetic is what each recording does with its own α_{ie}.

Leaving the rows uncentered does not give the same cancellation. The column of ones that would carry α_{ie} is absent, so a nonzero intercept is forced into $X^{\mathrm{raw}}\beta$. Part of α_{ie} is then absorbed by c_i, and the split between c_i and b_i moves with it. Pooling the mean across recordings fails for a different reason. Two recordings can share X and β and still have different intercepts. One subtracted mean cannot cancel both, and the leftover is a step that changes with the recording. That step is not proportional to d_e, but it is large enough to move the fitted slope. The implemented solve therefore recomputes the mean inside each recording, on the rows that enter that solve.

B.9 JOINT ESTIMATOR

For receiver i, each of the eighteen recordings is centered on its own first 80% of rows. The centered block enters as $Z_e = [X_e \;\; d_e X_e]$. Here Z and y are the refit blocks stacked, and D_i holds the column standard deviations of Z. Equation 3 penalizes $D_i\vartheta$. The penalty λ_i is chosen from 23 log-spaced values in $[10^{-4}, 10^{6}]$ on a pooled chronological 60/20 split. The refit uses the first 80% of each recording, and the minimizer is

$$\widehat{\vartheta}_i = D_i^{-1}\big(D_i^{-1}Z^\top Z D_i^{-1} + \lambda_i I\big)^{-1} D_i^{-1} Z^\top y. \tag{11}$$

The first half of $\widehat{\vartheta}_i$ is $\widehat{c}_i$ and the second half is $\widehat{b}_i$. The edge score is Equation 4. The same minimizer can be written as

$$\widehat{u} = \arg\min_u \|y - ZD_i^{-1}u\|_2^2 + \lambda_i\|u\|_2^2, \qquad \widehat{\vartheta}_i = D_i^{-1}\widehat{u}. \tag{12}$$

The returned c_i and b_i are in the units of the centered features. The penalty stays on the standardized coefficient. A column whose standard deviation is below 10^{-10} is given scale 1, so a nearly constant column is not divided by zero.

One exploratory fit keeps this split, these column scales, and this 23-value grid, but each recording contributes only $d_e X_e$. There is no dose-independent coefficient. On the held-out graphs its mean AP is 0.9980. It is not a frozen endpoint.

B.10 TWO-STAGE SLOPE

The two-stage score uses a separate ridge in every recording. Columns are not standardized. The penalty λ_e is chosen from the 20-value grid on $[10^{-4}, 10^{5}]$ by the validation error inside that recording. The coefficient is refit on the first 80% of its rows:

$$\widehat{\beta}_e(\lambda_e) = (X_e^\top X_e + \lambda_e I)^{-1} X_e^\top y_e. \tag{13}$$

Ordinary least squares then takes the slope of these eighteen coefficient matrices against dose. Each recording has equal weight:

$$\bar{d} = \frac{1}{18}\sum_{e=1}^{18} d_e, \qquad \widehat{b} = \frac{\sum_{e=1}^{18}(d_e - \bar{d})\,\widehat{\beta}_e(\lambda_e)}{\sum_{e=1}^{18}(d_e - \bar{d})^2}. \tag{14}$$

The regression intercept is not used. The undirected score is again Equation 4, applied to $\widehat{b}$. This slope is not the joint coefficient. Each recording has its own penalty, and the regression does not weight a recording by $X_e^\top X_e$.

A second two-stage score keeps these eighteen recordings and this unweighted slope, but standardizes columns inside each recording the way Appendix B.9 standardizes the joint design, and selects λ_e from the same 23-value grid on $[10^{-4}, 10^{6}]$. On the held-out graphs its mean AP is 0.9209. It is not a frozen endpoint.

Otsu, the primary label-free rule, cuts the 378 undirected scores. Sort them as $s_{(1)} \le \cdots \le s_{(m)}$ with $m = 378$. For a split after the first k scores, write $n_L = k$, $n_R = m - k$, and let μ_L and μ_R be the means of the two blocks. The split that is kept maximizes

$$n_L n_R (\mu_L - \mu_R)^2,$$

and the threshold is the midpoint between $s_{(k)}$ and $s_{(k+1)}$. A pair is kept when its score is strictly above that midpoint. The rule does not use the known edge count.

The frozen mixture rule was locked on R1–R5 and is not the primary graph. It fits two Gaussians to $\log s_{ij}$, after replacing a nonpositive score by a tiny positive floor so the logarithm exists. The two means start at the smallest and largest log-scores, the two variances start at the variance of the log-scores, and the two weights start at one half. Variances are not allowed to fall below 10^{-6}. Responsibilities, weights, means, and variances are then updated together, for at most 300 passes, and the loop stops when every parameter has moved by less than 10^{-10}. A pair is assigned to the high-mean component when its responsibility for that component is at least $1/2$. If the selected scores and the unselected scores do not overlap, the tabulated cutoff is the midpoint between the lowest selected score and the highest unselected score. If they overlap, the cutoff falls back to the lowest selected score. That cutoff is a summary of the assignment. The assignment itself is the responsibility rule, not the midpoint. The 3 s comparison cuts the raw voltage before filtering. Appendix C.1 records that sample order.

B.11 Synthetic trajectories

The stress test uses 28 nodes and 42 undirected edges, with no self-loops. The random family draws 42 distinct pairs uniformly from all unordered pairs. The small-world family starts from the cycle, including the edge between the first and last nodes, and then adds pairs whose circular distance is at least 2 until 42 edges are present. The hub family connects one node to the other 27 nodes and then adds edges among the remaining nodes until 42 edges are present. Section 7 uses these three constructions; they are not Watts–Strogatz or Barabási–Albert samples.

Edge weights are drawn independently from $[0.7, 1.3]$ and replaced by their symmetric part. Natural frequencies and common-drive gains are drawn independently from a normal distribution with mean 1 and standard deviation 0.25. The initial phase is uniform on $[-\pi, \pi]$.

Each trajectory is an Euler–Maruyama path with step 0.02, run for 4500 steps. At dose d,

$$\begin{aligned}\dot{\phi}_i = \omega_i + 5\Big(d + \eta\frac{d^2}{0.05}\Big)\sum_{j\neq i} A_{ij}w_{ij}\big(\sin(\phi_j - \phi_i - \ell) + h\sin 2(\phi_j - \phi_i)\big) \\ + \kappa g_i \sin\big(0.37t + 0.4\sin(0.11t)\big).\end{aligned} \tag{15}$$

The matched, confounded, and nonlinear scenarios set $(\kappa, h, \eta, \sigma, \ell, \rho)$ to $(0, 0, 0, 0.005, 0, 0)$, $(0.15, 0.10, 0, 0.010, 0.05, 0.03)$, and $(0.15, 0.12, 0.3, 0.012, 0.05, 0.03)$, respectively. The state update adds $\sigma\sqrt{0.02}$ times a standard normal vector. The first 500 steps are dropped. The retained phase is observed as $\phi + \rho\sin\phi$, plus an independent normal vector of scale σ, and then unwrapped along time. The common-drive amplitude κg_i is constant across doses. The dose enters through the coupling gain.

Regression rows use a forward window of 10 samples and a stride of 15, after dropping 100 samples at each end. The target is the phase increment over that window divided by 10×0.02, and the sine feature is read at the window midpoint. CHIRP then uses the same standardized joint ridge as the physical estimator, on six doses and three repeats. The single-dose score uses only $d = .05$. The separate-slope score fits each dose with its three repeats pooled, then takes the unweighted slope of those six coefficient vectors on dose. Phase locking averages the absolute mean resultant length over the positive doses. The 45 trajectories are drawn one family at a time, and within a family one scenario at a time. Let r be the number of trajectories already drawn and let $s \in \{0, 1, 2, 3, 4\}$ be the seed index inside that family and scenario. The generator seed is $10000 + 100s + 4r$, because each finished trajectory appends four method rows before the next seed is drawn.

C Additional held-out results

Top-42 F1 is included as a diagnostic because it uses the true graph density. Otsu is the primary label-free rule for producing a binary graph. We report the frozen two-Gaussian mixture threshold even though it performs worse; it was not replaced after the held-out results were known. The mixture cutoffs lie far below the Otsu cutoffs, which is why the mixture rule over-selects (mean 95.6 edges versus 42 true edges) while keeping recall 1.0000.

Table 5: Confirmatory aggregate results and paired tests against 18 s CHIRP. Values are taken from the frozen run artifact, except the 3 s row.

Comparison	Time (s)	Mean AP	one-sided exact Wilcoxon p
CHIRP	18	0.9980	–
CHIRP (3 s)[†]	3	0.9350	–
Separate-fit homotopy	15	0.9012	3.05×10^{-5}
Single dose $d = .05$	3	0.8814	3.05×10^{-5}
Dose-independent nuisance	18	0.2034	3.05×10^{-5}
Path phase locking	15	0.3407	3.05×10^{-5}
Path voltage correlation	15	0.2703	3.05×10^{-5}
Path partial correlation	15	0.2998	3.05×10^{-5}
Zero-dose weak model	3	0.1136	3.05×10^{-5}

[†] Versus the 3 s single-dose model, 3 s CHIRP has $p = 1.01 \times 10^{-3}$ (positive on 13 of 15 graphs).

Table 6: Per-graph held-out average precision. CHIRP uses 18 s. The 3 s CHIRP cuts the raw voltage before filtering. The single-dose and zero-dose models use 3 s. Two-stage does not scale columns. Std. two-stage uses the joint column scaling and the 23-value ridge grid, on the same 18 s recordings.

Graph	CHIRP	3 s CHIRP	Single	Separate	Two-stage	Std. two-stage	Volt. corr.	Zero-dose
R6	0.9994	0.9532	0.9035	0.9200	0.9104	0.9606	0.3178	0.1365
R7	0.9971	0.9381	0.9338	0.9293	0.9373	0.9502	0.2582	0.0985
R8	0.9994	0.9400	0.8710	0.8888	0.8570	0.9290	0.3084	0.0838
R9	0.9994	0.9502	0.9302	0.8312	0.8724	0.9255	0.2910	0.1246
R10	1.0000	0.9907	0.9378	0.9593	0.9293	0.9589	0.1968	0.0972
R11	1.0000	0.9736	0.9204	0.9462	0.9162	0.9970	0.2557	0.1185
R12	0.9994	0.9611	0.7759	0.8966	0.8492	0.8189	0.2252	0.1044
R13	0.9975	0.9107	0.9311	0.8952	0.8797	0.9286	0.2514	0.1153
R14	0.9954	0.9550	0.8962	0.9145	0.8656	0.9186	0.3034	0.1352
R15	1.0000	0.9461	0.8754	0.9248	0.8853	0.9323	0.2920	0.1148
R16	0.9943	0.8840	0.7934	0.8108	0.7835	0.8310	0.2522	0.1221
R17	1.0000	0.9501	0.8224	0.9145	0.8351	0.8728	0.2117	0.0881
R18	0.9989	0.9000	0.9523	0.9261	0.9320	0.9833	0.2735	0.1492
R19	0.9938	0.8929	0.8652	0.9172	0.8696	0.9175	0.3294	0.1162
R20	0.9951	0.8798	0.8127	0.8434	0.8320	0.8898	0.2872	0.0992
Mean	0.9980	0.9350	0.8814	0.9012	0.8770	0.9209	0.2703	0.1136

Table 7: Frozen label-free thresholds on held-out graphs. The mixture cutoff is the midpoint between the lowest selected score and the highest unselected score after assigning pairs with high-component responsibility at least $1/2$. Otsu is the primary rule.

	Log-Gaussian mixture					Otsu				
Graph	τ	$\lvert\widehat{E}\rvert$	P	R	F1	τ	$\lvert\widehat{E}\rvert$	P	R	F1
R6	1.970	92	0.4565	1.0000	0.6269	143.7	42	0.9762	0.9762	0.9762
R7	1.287	120	0.3500	1.0000	0.5185	152.2	42	0.9762	0.9762	0.9762
R8	0.663	110	0.3818	1.0000	0.5526	146.1	40	1.0000	0.9524	0.9756
R9	1.303	75	0.5600	1.0000	0.7179	147.4	41	1.0000	0.9762	0.9880
R10	1.854	69	0.6087	1.0000	0.7568	159.7	42	1.0000	1.0000	1.0000
R11	1.122	71	0.5915	1.0000	0.7434	164.6	39	1.0000	0.9286	0.9630
R12	0.406	110	0.3818	1.0000	0.5526	154.5	40	1.0000	0.9524	0.9756
R13	0.348	106	0.3962	1.0000	0.5676	140.1	42	0.9762	0.9762	0.9762
R14	1.078	108	0.3889	1.0000	0.5600	139.6	41	1.0000	0.9762	0.9880
R15	1.199	82	0.5122	1.0000	0.6774	149.9	41	1.0000	0.9762	0.9880
R16	1.980	113	0.3717	1.0000	0.5419	137.3	38	1.0000	0.9048	0.9500
R17	1.545	79	0.5316	1.0000	0.6942	160.8	40	1.0000	0.9524	0.9756
R18	1.029	92	0.4565	1.0000	0.6269	129.5	43	0.9767	1.0000	0.9882
R19	1.549	68	0.6176	1.0000	0.7636	127.3	40	1.0000	0.9524	0.9756
R20	1.517	139	0.3022	1.0000	0.4641	137.0	42	0.9762	0.9762	0.9762
Mean	1.257	95.6	0.4605	1.0000	0.6243	146.0	40.9	0.9921	0.9651	0.9782

C.1 WHERE THE 3 S BUDGET IS CUT

The 3 s fit uses repeat 1 at each of the six doses, then the same joint estimator as the frozen run. The window is $W = 300$ samples and the stride is 20, as in the single-dose baseline. Each source file is one second, hence 30,000 samples at 30,000 Hz. The fit keeps the first 15,000 samples and does not pass the rest to the filter. The fourth-order zero-phase Butterworth band, 700–1800 Hz, and the Hilbert unwrap then run on that clip alone. After the phase exists, 1000 samples are dropped at both ends of the short clip, and the interior is windowed with $W = 300$ and stride 20. There is no second cut in time. The opening and closing trim sits inside the 0.5 s budget, and the six clips yield 3,810 regression rows.

The single-dose baseline is already inside a 3 s raw budget. It takes the three one-second repeats at $d = .05$, filters each full second, and uses the same trim, window, and stride, for 4,155 rows. Its phase does not depend on voltage outside those three seconds. This comparison does not replace the frozen 18 s endpoint.

D REPRODUCTION RECORD

The physical recordings are the public 28-node electronic-oscillator trajectories of Vera-Avila et al. (Vera-Ávila et al., 2020b). Each network supplies six coupling doses $d \in \{0, .01, .02, .03, .04, .05\}$, three one-second repeats, and an undirected edge list of 42 links. Development graphs are R1–R5. Held-out graphs are R6–R20. Figure 10 draws each edge list. The frozen physical seed is 2027. The

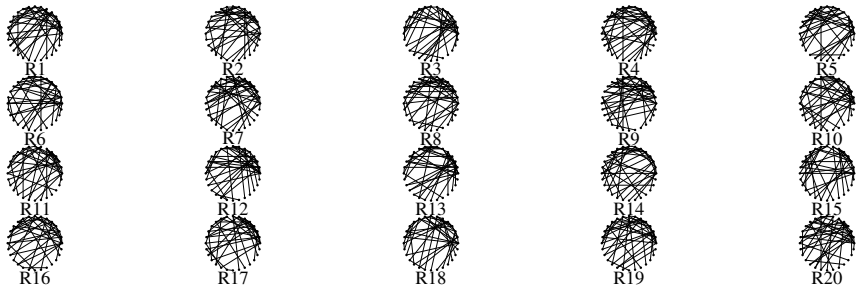

Figure 10: The twenty physical networks used in the benchmark. Each drawing has 28 nodes on a circle, in index order, and the 42 recorded undirected edges. R1–R5 are the development graphs. R6–R20 are held out. The circle is not a physical layout.

implementation uses NumPy, SciPy, Matplotlib, and PyYAML.

Four runs are required, in this order. Locked choices were fixed before held-out graphs were read: six doses, three repeats, a fourth-order zero-phase Butterworth band of 700–1800 Hz, Hilbert phase, window 300, joint-estimator stride 120, single-dose stride 20, a ridge grid of 23 values from 10^{-4} to 10^{6}, seed 2027, 28 nodes, 42 undirected edges, and the graph as the independent unit. Truth for a network is opened only after every score for that network exists.

Run 1 is the frozen physical confirmation and the only frozen endpoint. For each of the networks R1–R20, it scores: CHIRP; the dose-independent component; the separate-fit homotopy of the positive doses; the single-dose model at $d = .05$; the one-repeat CHIRP; path phase locking; path voltage correlation; path partial correlation; and the zero-dose weak model. Label-free graphs use the log-Gaussian mixture rule and Otsu. Dose labels are permuted 20 times at the ridge already selected for that network. Ordered prefixes of the dose path, from two through six configurations and always including the zero dose, form the intervention-query curve. The run records graph-level ranking metrics, label-free selections, permutation controls, the query curve, and regression-row diagnostics. Aggregates use a 20,000-resample graph bootstrap and paired one-sided Wilcoxon tests.

Run 2 is the synthetic assumption stress test. It is required and is not a frozen endpoint. It builds 45 trajectories from the three graph families in Appendix B.11, five seeds each, under the matched, confounded, and nonlinear scenarios defined there. It scores CHIRP, the single-dose model, separate slopes, and phase locking.

Run 3 is the post-confirmation leave-one-dose and two-stage slope analysis. It was added after R6–R20 had been processed. For each physical network, it scores the two-stage slope in Equations 13–14, CHIRP on the endpoint doses $\{0, .05\}$ only, and CHIRP with each of the six doses omitted in turn. A later pass repeats only the two-stage slope with the joint column scaling and the 23-value ridge grid, and fits the joint design with the dose-independent block removed. The joint comparison uses Equation 11. It does not replace the frozen endpoint.

Run 4 is the 3 s unique-time comparison and threshold diagnostics. It was also added after the confirmatory run. Appendix C.1 records the sample order. The 3 s fit is compared with the full-path CHIRP on 18 s and with the 3 s single-dose model. The same run records mixture and Otsu thresholds on the full-path scores and the mean pairwise gap between trace-normalized phase-difference Grams. It does not replace the frozen endpoint.

After those four runs, one redraw reads the stored numerical artifacts and writes the frozen-physical figure, the unique-time figure, and the synthetic-stress figure, without refitting. The same redraw writes the constructed dose-response example, and the checks for dose variation, endpoint allocation, the coefficient bound, support separation, and the two-stage identity. Those checks are not among the four runs.

Table 8 records SHA-256 hashes of the analysis modules used for the numbers. Confirmation results should be compared against these hashes together with the locked choices in Section 5.

Table 8: SHA-256 hashes of the analysis modules.

Module	SHA-256
Frozen physical estimator	`3b7a48aece0f0b8edd98c79391cfd152` `c76c4435adf2e6ac8f9a1f3a8647f5db`
Label-free mixture rule	`97cbb1442b8282df232abe155a9e825e` `37ec546aea56b2c772947f3a9f245c02`
Phase features and loaders	`4cc3393f4089d20d12696d37115d6bb6` `6a7f420263abaeb0d6479d45a31183f8`
Run-setting loader	`22350e79a2715f67c80948a11e44825b` `c743bd04125d74488b656fb3df9caeb9`
Two-stage and leave-one-dose	`cb6a57ff464c2bf99e631f17a2736dd6` `9a34ed2adc41ab737263b0012cab6989`
Equal unique-time comparison	`b6e5a3f2a07fddf4e63a7e2342e475be` `d4f1645cf5b43cc2f42264ca2d755fb3`
Synthetic stress test	`f33aca9a31c1c05924419fca43117777` `411b033ba63193be05b4db29968acc17`
Frozen physical settings	`8b6ec521499f18d0598976bd78c21df1` `871aaaff1daef62009cdc7c90e717802`
Synthetic settings	`b823b66856e00f99f5189a0f9ceadb61` `dc4c23d506a887d467a6ceff729698ea`
Leave-one-dose settings	`25a016720b10fef10fc4aceaff66c71d` `e9008f3f37f39cf93d020c4551f07c44`
Batch driver	`68b5012f5d79eeb3c890fa251dbde57a` `f9893dec1e149dbde658d7830ad61d81`

E FRACTIONAL-ORDER MODEL DECISION

The original project also considered a fractional Kuramoto model. Fractional dynamics are appropriate when the underlying system shows hereditary or power-law relaxation (Ha & Jung, 2018; Ha et al., 2024). The electronic benchmark does not provide evidence for that behavior. Adding a free fractional order would therefore introduce another source of flexibility without resolving the topology problem. We exclude that extension here. A meaningful test would require a preregistered comparison with the nested order-one model, using held-out trajectory likelihood together with relaxation diagnostics.