RAVEL: FACTORIZING MULTI-REGION NEURAL COMMUNICATION INTO REUSABLE PROPAGATION ROUTES AND DYNAMIC RECRUITMENT

Anonymous authors
Paper under double-blind review

ABSTRACT

Inter-regional neural interactions can change rapidly with behavior and internal state, but such changes are mechanistically ambiguous: they may reflect changing transmission properties, or changing recruitment of otherwise stable communication mechanisms. Existing multi-region models capture delayed and time-varying interactions, but do not explicitly separate these alternatives. Motivated by parallel biological communication routes and their state-dependent engagement, we introduce **RAVEL**, a conditional spectral model for directed communication in multi-region local field potential (LFP) recordings. RAVEL predicts target spectral activity from a source region and factorizes the source-to-target transfer into reusable delayed propagation components, with explicit transport delays and frequency-dependent filtering, and time-varying recruitment that determines when each component is expressed. In synthetic experiments, RAVEL recovers stable propagation physics and changing recruitment when that mechanism generates the data, while a matched dynamic-delay model is favored when transmission delays genuinely change. On Allen Brain Observatory LFP recordings, propagation dictionaries are reproducible across mice, transfer across behavioral contexts with little loss in held-out predictive performance, and exhibit behavior-dependent recruitment. Locally learned transfers can also be composed to predict a withheld long-range interaction without fitting the endpoint directly. RAVEL further performs competitively on complete-target cross-region prediction against established models. Together, these results support separating relatively stable propagation structure from its dynamic recruitment as an interpretable description of multi-region neural communication.

1 INTRODUCTION

Brain function depends on communication among distributed neuronal populations. Anatomical connectivity provides a relatively persistent substrate for this communication, whereas the effective influence between regions can change rapidly with sensory input, behavioral demands, and internal state (Friston, 2011; Battaglia et al., 2012; Fries, 2015; Kohn et al., 2020; Palmigiano et al., 2017; Niell & Stryker, 2010; McGinley et al., 2015). Behavior-dependent changes in inter-regional electrophysiological activity are also observed in human intracranial recordings spanning basal-ganglia and thalamic circuits (Kasiri et al., 2023; Seyyed Mousavi et al., 2025). This creates a basic ambiguity when an estimated inter-regional interaction changes over time. A change in phase relationship or apparent propagation delay may reflect a change in the transmission mechanism itself, or it may arise because the brain changes the relative engagement of multiple propagation components whose individual transmission properties remain stable. These explanations can produce similar observed coupling while implying different circuit organizations: one requires continual reconfiguration of the delay-bearing interaction, whereas the other predicts a reusable communication substrate whose recruitment changes with context.

This question is particularly natural for local field potential (LFP) recordings. LFPs arise from the superposition of population transmembrane currents and contain pronounced frequency-dependent structure in both amplitude and phase (Buzsáki et al., 2012; Einevoll et al., 2013; Pesaran et al., 2018). Frequency-specific correlations and phase relationships have therefore long been used to

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.

characterize large-scale neuronal interactions (Varela et al., 2001; Siegel et al., 2012; Fries, 2015; Bastos & Schoffelen, 2016). A frequency-domain source-to-target operator is especially useful in this setting: its magnitude captures frequency-dependent gain, its phase captures temporal relationships, and explicit propagation delays have a simple spectral signature. Classical approaches including directed transfer functions, partial directed coherence, spectral Granger methods, and spectral dynamic causal models have similarly used frequency-domain representations to study directed interactions (Geweke, 1982; Dhamala et al., 2008; Kaminski & Blinowska, 1991; Baccalá & Sameshima, 2001; Friston et al., 2012; Moran et al., 2009). Our goal is not to infer anatomical connectivity from spectral statistics alone, but to use this representation to distinguish which aspects of inter-regional transmission are stable from those that are dynamically expressed.

Recent latent-variable models have substantially advanced the analysis of multi-region neural communication. DLAG and mDLAG identify low-dimensional delayed signals shared across neuronal populations and characterize their direction of flow (Gokcen et al., 2022; 2023). MRM-GP represents frequency-dependent directional communication and phase delays, while the Adaptive Delay Model (ADM) explicitly allows communication delays themselves to vary over time (Li et al., 2024; 2025). MRDS-IR introduces interpretable parametric impulse-response channels between nonlinear regional dynamics, and MR-LFADS separates inter-regional communication from local dynamics and unobserved inputs (Dowling & Savin, 2025; Liu et al., 2025). Related work on reusable nonstationary dynamics expresses time-varying activity through combinations of simpler dynamical components or shared subcircuits (Mudrik et al., 2024; Chen et al., 2024; Mudrik et al., 2025). Together, these methods provide powerful descriptions of delayed and dynamic communication. However, they leave open a different mechanistic hypothesis: apparent time-varying communication may arise not because propagation delays themselves change, but because the relative recruitment of a small set of stable delayed propagation responses changes over time or context. RAVEL is designed to test this alternative factorization directly.

This factorization is motivated by known features of biological communication. Cortical areas can communicate through multiple parallel routes, including direct corticocortical and indirect cortico–thalamo–cortical pathways (Sherman & Guillery, 2011; Sherman, 2016); in the mouse visual system, direct intracortical and transthalamic pathways can convey distinct information to higher visual areas (Blot et al., 2021). At the same time, communication between fixed anatomical populations can be rapidly reconfigured through changes in synchronization and effective coupling (Fries, 2005; Battaglia et al., 2012; Palmigiano et al., 2017; Niell & Stryker, 2010; McGinley et al., 2015), and interactions between cortical populations can occupy selective low-dimensional communication subspaces (Semedo et al., 2019). These observations do not imply that an observationally inferred spectral component corresponds uniquely to an anatomical pathway. Rather, they motivate a route-level abstraction in which relatively stable propagation properties are separated from their dynamic engagement.

We introduce **RAVEL** (**R**eusable **A**toms with **V**arying **E**ngagement **L**evels), a conditional spectral analysis model for multi-region electrophysiology that makes this distinction explicit. RAVEL does not infer a global anatomical network; instead, for each ordered source–target region pair, it predicts target spectral activity from simultaneously observed source activity in low-dimensional communication coordinates. The source-to-target transfer operator is factorized into instantaneous and smooth zero-phase terms together with a small dictionary of delayed propagation atoms. Each atom has an explicit transport delay and a frequency-dependent one-pole response, while a nonnegative recruitment profile determines how strongly it is expressed across event time. The propagation dictionary can be shared across recordings, subjects, or contexts, while communication coordinates, spatial expression, nuisance coupling, and recruitment remain recording specific. RAVEL therefore asks whether a compact set of delayed propagation responses can remain reusable even when their expression changes.

A further advantage of the transfer-function representation is that it naturally supports composition. When neighboring regions are represented in compatible communication spaces, their source-to-target transfer operators can be multiplied along a candidate chain, providing a direct multi-hop prediction without fitting the endpoint relationship itself. We therefore test whether local transfers learned independently—including from disjoint sets of events—can predict a withheld long-range interaction. This asks whether locally identified propagation structure carries predictive information beyond the individual fitted edges, without requiring RAVEL to infer a global anatomical network or a unique causal pathway.

Our contributions are fourfold:

- We formulate a conditional spectral factorization that separates reusable delayed propagation physics from its dynamic recruitment while explicitly accounting for instantaneous and smooth zero-phase coupling prevalent in LFP data.
- We construct synthetic benchmarks that distinguish stable propagation with changing recruitment from genuinely time-varying transmission delay. This provides a calibrated test of the model assumption: RAVEL recovers the former mechanism, whereas a dynamic-delay model is favored when the delays truly change.
- Using Allen Institute LFP recordings, we test whether propagation dictionaries are reproducible across animals and transferable across behavioral contexts while retaining held-out predictive value, and we evaluate complete-target source-only prediction against established multi-region communication models.
- We test whether neighboring transfer functions learned from disjoint sets of observations can be composed to predict a withheld long-range interaction, providing a direct test of whether locally learned propagation structure carries predictive information beyond individually fitted edges and supports multi-hop prediction.

2 RAVEL: Reusable Atoms with Varying Engagement Levels

2.1 Conditional Spectral Communication

We introduce **RAVEL** (**R**eusable **A**toms with **V**arying **E**ngagement **L**evels), a conditional spectral model that explicitly separates stable propagation structure from its time-varying recruitment. Figure 1 provides an overview of the model and its factorization. RAVEL models directed communication separately for each ordered source–target region pair. For recording or subject s, trial n, event-time window w, taper k, and frequency f, let $\mathbf{X}_{s,n,w,k}(f) \in \mathbb{C}^{p_s}$ and $\mathbf{Y}_{s,n,w,k}(f) \in \mathbb{C}^{q_s}$ denote the complex Fourier coefficients of the source and target regions, respectively. RAVEL first projects these channel-space observations into d-dimensional communication coordinates,

$$\mathbf{x}_{s,n,w,k}(f) = \mathbf{U}_s^\top \mathbf{X}_{s,n,w,k}(f), \qquad \mathbf{y}_{s,n,w,k}(f) = \mathbf{V}_s^\top \mathbf{Y}_{s,n,w,k}(f), \tag{1}$$

where $\mathbf{U}_s \in \mathbb{R}^{p_s \times d}$ and $\mathbf{V}_s \in \mathbb{R}^{q_s \times d}$ are estimated using training data only.

The target communication coordinates are predicted conditionally on the simultaneously observed source coordinates,

$$\widehat{\mathbf{y}}_{s,n,w,k}(f) = \mathbf{H}_{s,w}(f)\mathbf{x}_{s,n,w,k}(f), \tag{2}$$

where $\mathbf{H}_{s,w}(f) \in \mathbb{C}^{d \times d}$ is a frequency- and event-time-dependent transfer matrix. Thus, RAVEL is a directed conditional spectral regression model. Each observed edge has its own fitted transfer; in the multi-region experiment, incident edges use one training-derived communication basis at their common region (Appendix A.2.1).

2.2 Stable Atoms, Dynamic Recruitment

RAVEL decomposes the conditional transfer matrix into zero-phase coupling and delayed propagation components,

$$\begin{aligned} \mathbf{H}_{s,w}(f) = \mathbf{B}_{s,0} &+ \sum_{q=1}^{Q} \phi_q(f)\mathbf{B}_{s,q} \\ &+ \sum_{m=1}^{M} r_{s,m}(w)\widetilde{h}_m(f)\mathbf{A}_{s,m}. \end{aligned} \tag{3}$$

The first term, $\mathbf{B}_{s,0} \in \mathbb{R}^{d \times d}$, captures unrestricted instantaneous coupling. The functions $\phi_q(f)$ are smooth real-valued, zero-phase functions of log frequency, with associated real rank-one matrices $\mathbf{B}_{s,q}$. These terms absorb instantaneous and smoothly frequency-dependent zero-phase relationships so that such structure need not be attributed to delayed propagation. We refer to the model containing

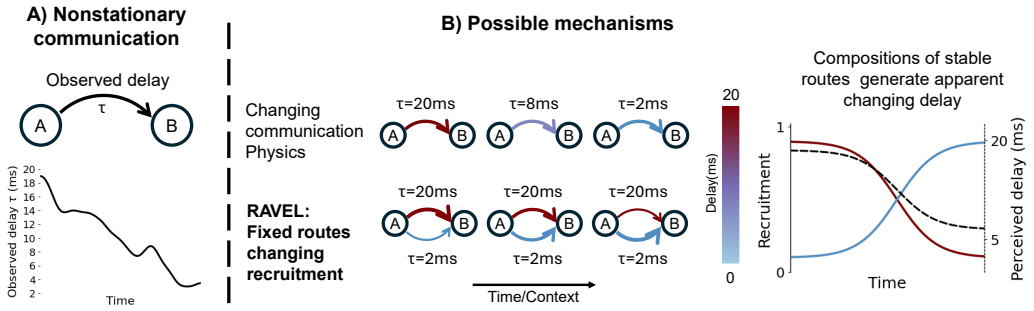


Figure 1: **Overview of RAVEL. A)** Inter-regional communication can exhibit an apparently time-varying delay. **B)** Such variation may reflect changing transmission delays or changing recruitment of fixed propagation motifs. RAVEL models the latter factorization: each motif has fixed propagation parameters, while its recruitment varies over time or context. Mixtures of motifs with different delays can therefore produce an apparently changing delay.

only these instantaneous and smooth zero-phase terms as the *nuisance model*; it provides a matched comparator for testing whether delayed propagation atoms add held-out predictive value.

Throughout the paper, we use propagation atom and propagation motif interchangeably to denote the same delayed model component. Propagation motif m is represented by a delayed first-order response

$$h_m(f) = \frac{\exp(-i2\pi f \tau_m/1000)}{1 + if/f_{c,m}}, \qquad f_{c,m} > 0, \tag{4}$$

where τ_m is an explicit transport delay in milliseconds and $f_{c,m}$ is the characteristic frequency of the first-order low-pass response. Since the pole itself contributes frequency-dependent phase, τ_m represents the explicit transport component rather than the total phase lag of the atom.

To remove arbitrary scaling between an atom and its spatial coefficient matrix, each response is normalized to unit root-mean-square magnitude over the analyzed frequency grid $\mathcal{F}$:

$$\widetilde{h}_m(f) = \frac{h_m(f)}{\sqrt{|\mathcal{F}|^{-1} \sum_{f' \in \mathcal{F}} |h_m(f')|^2}}. \tag{5}$$

The real rank-one matrix $\mathbf{A}_{s,m} \in \mathbb{R}^{d \times d}$ describes how motif m is expressed in the recording-specific source and target communication coordinates. Its recruitment,

$$r_{s,m}(w) \geq 0, \tag{6}$$

controls how strongly that propagation component is expressed at event-time window w.

A central assumption of RAVEL is therefore that the propagation response $(\tau_m, f_{c,m})$ is stable within the scope of a fitted dictionary, whereas its recruitment can vary over event time. Recruitment is estimated as a window-level profile shared across the trials entering a fit; it is not a separate free coefficient for every trial.

2.3 Recording-Specific Communication Coordinates

The communication subspaces are estimated separately for every recording and are not shared across subjects. Let $\mathbf{C}_{s,\ell}$ denote the lagged source–target cross-covariance computed from training residuals at temporal lag $\ell \in \mathcal{L}$. RAVEL forms

$$\mathbf{G}_s^x = \sum_{\ell \in \mathcal{L}} \mathbf{C}_{s,\ell} \mathbf{C}_{s,\ell}^\top, \qquad \mathbf{G}_s^y = \sum_{\ell \in \mathcal{L}} \mathbf{C}_{s,\ell}^\top \mathbf{C}_{s,\ell}. \tag{7}$$

The leading d eigenvectors of $\mathbf{G}_s^x$ and $\mathbf{G}_s^y$ define $\mathbf{U}_s$ and $\mathbf{V}_s$, respectively. This construction identifies low-dimensional source and target directions that participate strongly in lagged cross-region covariance without requiring a common channel geometry across recordings. For the multi-region experiment, we sum these covariance scores over each region's observed incident edges before selecting its axes (Appendix A.2.1).

2.4 Learning Reusable Propagation Physics

For a fixed atom dictionary

$$\mathcal{D} = \{(\tau_m, f_{c,m})\}_{m=1}^{M}, \tag{8}$$

RAVEL alternates between estimating the recording-specific transfer matrices and the recruitment profiles by minimizing complex source-to-target prediction error on the training data. The frequency-dependent bases carry the complex phase, while the spatial coefficient matrices are constrained to be real. The nuisance and propagation residue matrices are further restricted to rank one.

The atom dictionary is selected using held-out conditional prediction, rather than training reconstruction alone. Candidate dictionaries are compared against a matched nuisance model containing the instantaneous and smooth zero-phase terms but no delayed atoms. Multiple perturbed initializations are used to evaluate the stability of the recovered propagation responses. Full optimization, regularization, restart, and model-selection details are given in Appendix A.2.

When multiple recordings are fitted jointly, only the propagation dictionary $\mathcal{D}$ is shared. The preprocessing transformations, communication coordinates, $\mathbf{B}_{s,0}$, $\mathbf{B}_{s,q}$, $\mathbf{A}_{s,m}$, and $r_{s,m}(w)$ remain recording specific. Thus, pooling tests whether recordings share delayed propagation responses without requiring them to share electrode geometry, atom-specific spatial expression, nuisance coupling, or recruitment dynamics.

2.5 Complete-Target Prediction

Predicted communication coordinates are mapped back to the complete target-channel space as

$$\widehat{\mathbf{Y}}_{s,n,w,k}(f) = \mathbf{V}_s \widehat{\mathbf{y}}_{s,n,w,k}(f). \tag{9}$$

No target channel is supplied as a predictor in Eq. 2. All data-dependent preprocessing transformations and communication axes are estimated from training data and subsequently held fixed for validation and test data.

RAVEL therefore separates two levels of variation: the shared dictionary describes reusable propagation responses, whereas recording-specific recruitment describes when and how strongly those responses are expressed. The inferred atoms should be interpreted as reusable pairwise propagation components. Without additional anatomical or interventional constraints, they do not by themselves identify unique anatomical pathways.

3 Experiments

3.1 Dynamic Recruitment vs. Dynamic Delay

A central premise of RAVEL is that event-dependent changes in inter-regional communication can arise from changing *recruitment of stable propagation mechanisms*, rather than from changes in the propagation delays themselves. We therefore constructed a reciprocal synthetic benchmark that explicitly contrasts these two explanations. In both generative worlds, communication contained two delayed one-pole components and the observation dimensionality, trial count, noise statistics, preprocessing, and train/validation/test splits were held fixed. In the *dynamic-recruitment* world, the two propagation delays were fixed at 9.5 and 21.5 ms while their recruitment profiles varied smoothly over event time. In the *dynamic-delay* world, component amplitudes were static while the two physical delays changed smoothly from $7.5 \rightarrow 14.0$ ms and $19.0 \rightarrow 27.0$ ms, respectively. Both worlds used 45 Hz one-pole filters.

To isolate the dynamical assumption rather than model-order recovery, RAVEL was fit with the true atom count $M = 2$. We compared full RAVEL against (i) a static-recruitment RAVEL ablation, (ii) a nuisance-only model with no delayed atoms, and (iii) a matched two-component dynamic-delay model with time-invariant component gains but time-varying delays. We additionally fit the Adaptive Delay Model (ADM) (Li et al., 2025) using its native fitting and model-selection procedure. All methods were evaluated using the same sealed trials and complete-target source-only spectral evaluation. Full simulation and baseline details are given in Appendix A.3.

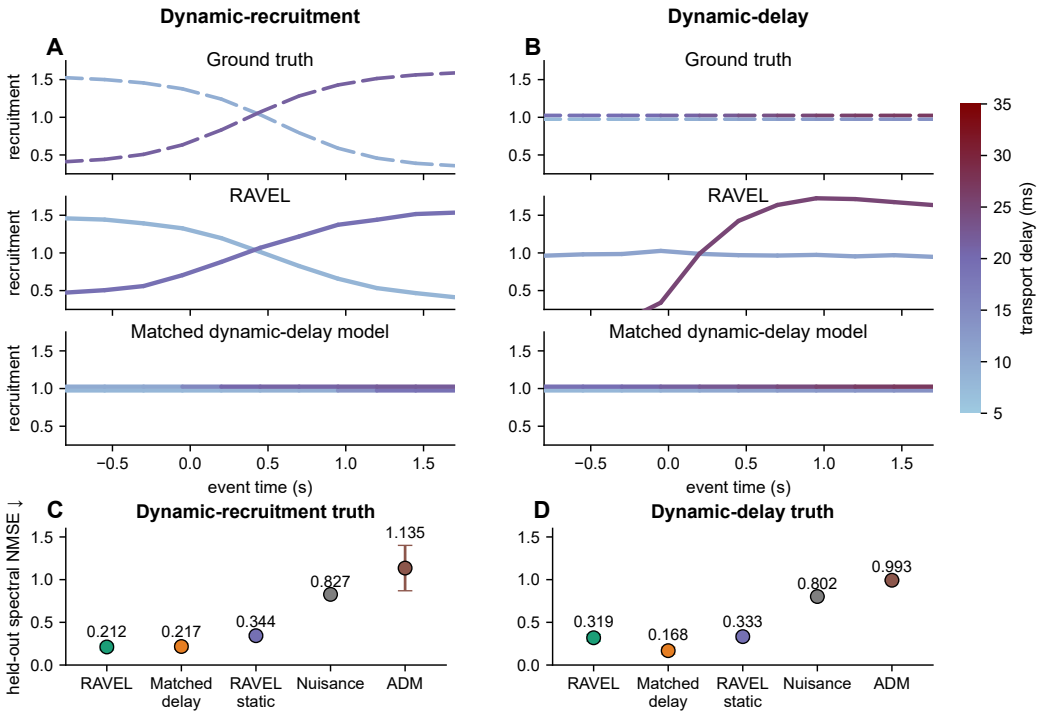


Figure 2: **Synthetic discrimination between dynamic recruitment and dynamic transmission delay. A–B)** Ground truth and fitted models in the dynamic-recruitment and dynamic-delay worlds, respectively. Line color denotes transport delay; dashed and solid curves denote ground truth and model estimates. **C–D)** Held-out complete-target spectral NMSE for the corresponding generative worlds (mean ± SD across three independently generated seeds; lower is better).

Under dynamic-recruitment truth, RAVEL recovered the generative mechanism even though held-out prediction alone did not uniquely identify it (Fig. 2A,C). Across the three seeds, RAVEL obtained an NMSE of 0.212 ± 0.034, similar to the more flexible matched dynamic-delay model (0.217 ± 0.033), and substantially below static-recruitment RAVEL (0.344 ± 0.023) and the nuisance-only model (0.827 ± 0.002). However, the two models explained these observations very differently. After matching recovered atoms to ground truth, RAVEL achieved a mean absolute delay error of 1.46 ms, mean complex atom-shape similarity of 0.955, and mean recruitment correlation of 0.999. In contrast, the matched dynamic-delay model attained similar predictive accuracy by introducing large temporal changes in delays that were in fact constant: its two inferred delay trajectories had RMSEs of 4.94 and 6.76 ms relative to the corresponding fixed ground-truth delays. Thus, comparable held-out prediction did not imply recovery of the correct communication mechanism.

The reciprocal dynamic-delay world provided a negative control for RAVEL's stable-physics assumption (Fig. 2B,D). When delays truly changed, the matched dynamic-delay model recovered them with a mean trajectory RMSE of 0.32 ms and mean trajectory correlation of 0.995, and its held-out NMSE decreased to 0.168 ± 0.040, compared with 0.319 ± 0.038 for RAVEL. Static-recruitment RAVEL performed similarly to full RAVEL (0.333 ± 0.032), indicating that allowing recruitment to vary does not eliminate the cost of imposing fixed propagation physics when the true delays move. ADM, fit and selected using its native objective, performed weakly under the complete-target source-only spectral evaluation in both worlds (1.135 ± 0.266 and 0.993 ± 0.009 NMSE). Because this evaluation differs from ADM's native objective, the result reflects performance on this source-only prediction task rather than its native delay-recovery setting. Together, these reciprocal simulations show that RAVEL preferentially recovers stable propagation components with dynamic recruitment, whereas the dynamic-delay model has a predictive and mechanistic advantage when transmission delays genuinely vary.

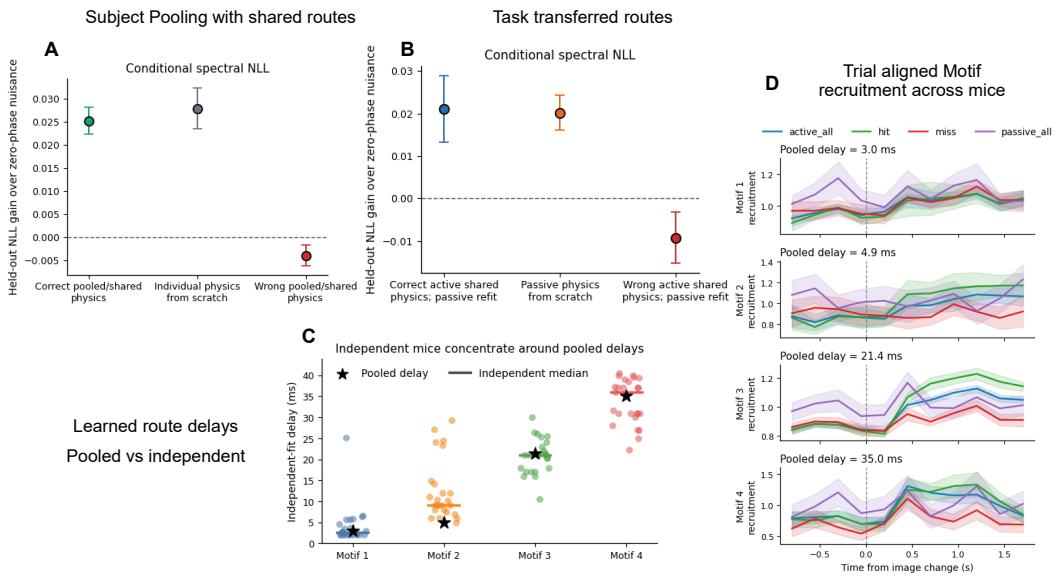


Figure 3: **Propagation physics across mice and behavioral contexts. A)** Held-out prediction with pooled/shared physics, independently fitted physics, and prespecified wrong-physics controls. Values show across-mouse mean ± SEM; the dashed line marks zero gain relative to the matched zero-phase nuisance model. **B)** Transfer of an active-trained shared dictionary to passive recordings, compared with passive-from-scratch and wrong-physics controls. **C)** Independently fitted mouse delays relative to the pooled dictionary; stars indicate pooled delays and horizontal bars indicate independent-fit medians. **D)** Trial-aligned recruitment of the four pooled motifs across active, hit, miss, and passive conditions (mean ± SEM across mice); the dashed line marks image-change onset.

3.2 Stable Physics Across Mice and Contexts

We analyzed publicly available Allen Brain Observatory Visual Behavior Neuropixels (VBN) recordings (Bennett et al., 2026; Bennett, 2024). For the VISam→VISpm connection, we evaluated shared propagation physics across 25 mice and two visual cortical regions (Details in Appendix A.4). Independent mouse fits recovered delays that clustered around the pooled dictionary. The pooled delays were 3.0, 4.9, 21.4, and 35.0 ms, compared with independent-fit median delays of 2.5, 9.0, 21.0, and 36.0 ms, respectively (Fig. 3). A single pooled dictionary also retained the predictive performance of independently fitted physics: the mean held-out NLL gain was 0.025 for pooled physics and 0.028 for independent fits, with no statistically significant difference. In contrast, the mean performance of eight prespecified wrong dictionaries was below the zero-phase nuisance control. This indicates that successful pooling depends on the learned propagation physics rather than only on refitting subject-specific parameters.

The learned dictionary also transferred from active to passive recordings. After freezing the active-trained physics and refitting the passive-specific parameters, the mean held-out NLL gain was 0.021, compared with 0.020 when the complete model was learned from passive data. This small numerical difference was not statistically significant, indicating that freezing the active-trained propagation physics preserved predictive performance relative to relearning the physics from scratch. Recruitment profiles were also repeatable across disjoint trial subsets (mean split-half correlation $r = 0.64$). Finally, recruitment differed with behavioral outcome. The clearest visual hit–miss separation occurred for Motif 3 (21.4 ms), whereas the fastest motif, Motif 1 (3.0 ms), showed little outcome-dependent separation. Together, these results support reuse of a common propagation dictionary across mice and contexts, with behavioral variation expressed through dynamic motif recruitment.

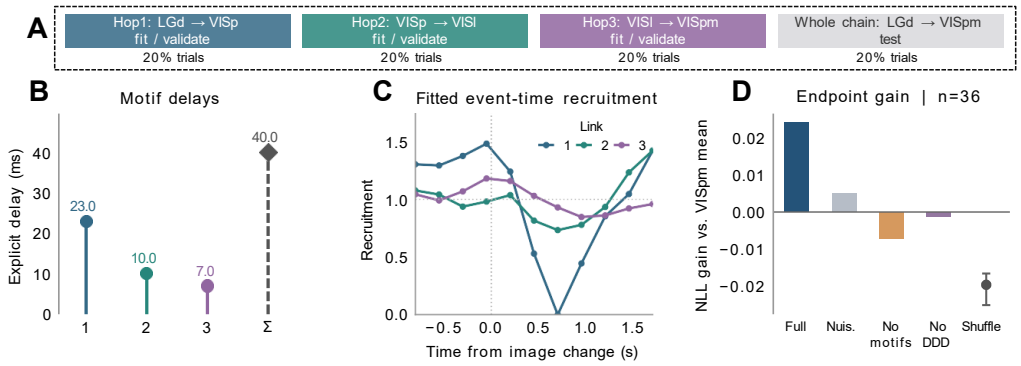


Figure 4: **Disjoint three-hop composition. A)** Each adjacent transfer is fitted on a separate chronological block, while LGd–VISpm is reserved for endpoint evaluation. **B)** Explicit transport delays of the three fitted motifs and their sum (Σ); the sum denotes the delay of the delayed$\times$delayed$\times$delayed (DDD) term rather than the complete transfer product. **C)** Event-time recruitment of the three fitted links. **D)** Held-out endpoint NLL gain relative to the training-derived VISpm mean for the full composition and controls. Nuis. is a refitted chain without delayed motifs; No motifs removes all delayed terms; No DDD removes only their three-link product; Shuffle permutes LGd test-trial identity. Positive values indicate improved prediction.

3.3 Composing Transfers for Long-Range Prediction

Using the same Allen Institute VBN dataset, we extracted the chain LGd $\rightarrow$ VISp $\rightarrow$ VISl $\rightarrow$ VISpm and tested whether its locally learned transfer functions predict an endpoint relation that was never fitted directly. In the strict disjoint setting, each adjacent pair was available for fitting on a different chronological block of active image-change sessions. No two links were fitted on the same events, and no LGd–VISpm observations entered fitting or model selection. On a separate held-out block, VISpm was predicted using LGd alone, with VISp and VISl masked, by composing the three transfers:

$$\widehat{\mathbf{y}}_{\mathrm{VISpm}}(w,f) = H_{\mathrm{VISl}\to\mathrm{VISpm}}(w,f)H_{\mathrm{VISp}\to\mathrm{VISl}}(w,f)H_{\mathrm{LGd}\to\mathrm{VISp}}(w,f)\mathbf{x}_{\mathrm{LGd}}(w,f). \quad (10)$$

As shown in Figure 4 (session 1108335514, mouse 571520), the composed model attained held-out 6–80 Hz spectral-residual $R^2 = 0.172$ on 36 events, compared with 0.058 for a refitted nuisance-only chain and 0.117 after removing the product of the three delayed terms. All three local links had positive held-out gains. The full chain improved negative log likelihood over the training-derived VISpm mean by 0.024, whereas shuffling LGd test-trial identities yielded a median gain of -0.020. These results demonstrate an advantage of RAVEL's spectral transfer-function parameterization: locally learned transfers can be composed to predict an unobserved endpoint relationship even under disjoint training. This demonstrates compositional endpoint prediction, not identification of a unique anatomical or causal route; details are provided in Appendix A.5.

3.4 Complete-Target Prediction Across Recording Sessions

We evaluated whether RAVEL can predict the activity of an entire cortical region from another region using Allen Brain Observatory Visual Coding LFP (Siegle et al., 2021; 2025). We analyzed VISp and VISam across eight recording sessions using drifting-grating trials at $135°$, with 2-s trials at 100 Hz and 10 TRAIN and 5 sealed TEST trials per session. We used a *complete-target* prediction task to isolate cross-region predictive information: to predict VISp from VISam, all VISam channels were observed while *all* VISp channels were withheld during inference, and vice versa.

We compared RAVEL with MRM-GP (Li et al., 2024) and DLAG (Gokcen et al., 2022). All methods used identical TRAIN/TEST splits and were evaluated on the same sealed TEST trials using spectral NMSE over 2–40 Hz. Although this is not the native evaluation task of all baselines, the same complete-target condition and spectral metric were applied to every method, while each retained its native training objective. Full configurations, session-level results, and statistical details are reported in Appendix A.6.

Table 1: **Cross-session complete-target prediction on Allen Visual Coding.** Spectral NMSE (mean ± SD across eight recording sessions at 135°; lower is better). No target-region channels are observed during inference.

Model	VISp ← VISam	VISam ← VISp
RAVEL	$\mathbf{0.702 \pm 0.190}$	$\mathbf{0.500 \pm 0.158}$
MRM-GP	0.776 ± 0.135	0.573 ± 0.175
DLAG	0.770 ± 0.131	0.759 ± 0.304

As shown in Table 1, RAVEL achieved the lowest mean NMSE in both prediction directions. For VISp←VISam, differences among the three models were not significant across sessions (Friedman $\chi^2 = 0.75$, $p = 0.687$), and RAVEL had lower NMSE than MRM-GP and DLAG in 5/8 and 4/8 sessions, respectively. The reverse direction showed a substantially stronger separation. For VISam←VISp, the omnibus difference across models was significant (Friedman $\chi^2 = 13.0$, $p = 0.0015$). RAVEL had lower NMSE than MRM-GP in 6/8 sessions and than DLAG in all 8/8 sessions. Exact paired sign-flip tests gave mean NMSE reductions of 0.073 relative to MRM-GP and 0.259 relative to DLAG. After Holm correction across the four directional pairwise comparisons, the RAVEL–DLAG difference remained significant ($p_{\mathrm{Holm}} = 0.0313$), whereas the RAVEL–MRM-GP comparison did not ($p_{\mathrm{Holm}} = 0.1406$).

Together, these results indicate that RAVEL is competitive with established latent communication models for VISp←VISam and provides a particularly consistent improvement for VISam←VISp across recording sessions. A complementary benchmark across all eight stimulus orientations of session 751348571, including ADM (Li et al., 2025), is reported in Appendix A.6.2. These predictive gains were also obtained with substantially lower empirical fitting cost: in a separate CPU-only synthetic scaling benchmark, ADM, DLAG, and MRM-GP required approximately 15.3×, 134.8×, and 170.0× longer fitting times than RAVEL, respectively, at the reference sequence length $T = 300$ (Appendix A.7).

4 Discussion and Limitations

RAVEL is a biologically inspired conditional spectral analysis model for multi-region electrophysiology that factorizes directed source-to-target interactions into reusable delayed propagation components and time-varying recruitment. This formulation tests whether apparent changes in interregional communication can be explained by changing engagement of stable propagation mechanisms rather than by continually changing transmission physics. Across simulations and Allen Institute recordings, the results support this factorization as a useful description of neural communication. Shared propagation dictionaries retained predictive value across animals and behavioral contexts while recruitment remained dynamic. Complete-target prediction further showed that this structured representation retains predictive value, while the multi-hop experiment provided a proof of principle that locally learned transfer functions can be composed to predict a withheld long-range relationship. More generally, these findings suggest that separating propagation physics from its dynamic engagement may provide an interpretable intermediate description between fixed anatomical connectivity and fully time-varying functional coupling.

Several limitations define the present scope of the work. The empirical results are restricted to mouse visual-system LFP recordings, and generalization to other species, systems, and modalities remains to be tested. Because the evidence is observational, inferred delays should be interpreted as predictive propagation components rather than uniquely identified anatomical pathways; perturbation experiments could directly compare RAVEL delays with stimulation-evoked response latencies measured along neural circuits. RAVEL also currently fits pairwise interactions, and the multi-hop experiment evaluates a prespecified chain rather than discovering network pathways. Finally, the delayed atoms use a restricted one-pole transfer function; richer filtering families may be needed when the complete transfer shape cannot be captured by this parameterization.

AI USE STATEMENT

We used OpenAI's ChatGPT as an assistive tool for code implementation, debugging, and manuscript editing. The scientific questions, model design, experiments, evaluation procedures, and conclusions were determined by the authors. All AI-assisted code and text used in the paper were reviewed and verified by the authors, who take full responsibility for the work.

ETHICS STATEMENT

This study uses synthetic data and previously collected, publicly available mouse electrophysiology recordings from the Allen Brain Observatory. No new human or animal experiments were conducted for this work. The original Allen Institute experiments were performed under the animal care and use procedures and approvals described by the data providers.

REPRODUCIBILITY STATEMENT

The appendix provides the model specification, preprocessing, optimization, data splits, synthetic-data generation, baseline configurations, transfer and pooling procedures, and evaluation metrics needed to reproduce the reported experiments. The Allen Institute datasets are publicly available. Code for RAVEL, synthetic experiments, baseline evaluation, and figure reproduction will be released in the anonymous repository accompanying the submission: `https://github.com/GitAnonym2/RAVEL2027`.

REFERENCES

Luiz A Baccalá and Koichi Sameshima. Partial directed coherence: a new concept in neural structure determination. *Biological cybernetics*, 84(6):463–474, 2001.

André M Bastos and Jan-Mathijs Schoffelen. A tutorial review of functional connectivity analysis methods and their interpretational pitfalls. *Frontiers in systems neuroscience*, 9:175, 2016.

Demian Battaglia, Annette Witt, Fred Wolf, and Theo Geisel. Dynamic effective connectivity of inter-areal brain circuits. *PLoS computational biology*, 8(3):e1002438, 2012.

Corbett Bennett. Allen institute - visual behavior - neuropixels, 2024. URL `https://doi.org/10.48324/dandi.000713/0.240702.1725`.

Corbett Bennett, Samuel D. Gale, Greggory Heller, Tamina K. Ramirez, Hannah Belski, Alex Piet, Omid Zobeiri, Adam Amster, Anton Arkhipov, Alex Cahoon, et al. Map of spiking activity underlying change detection in the mouse visual system. *Cell*, 189(15):4775–4791.e13, jul 2026. doi: 10.1016/j.cell.2026.06.025.

Antonin Blot, Morgane M Roth, Ioana Gasler, Mitra Javadzadeh, Fabia Imhof, and Sonja B Hofer. Visual intracortical and transthalamic pathways carry distinct information to cortical areas. *Neuron*, 109(12):1996–2008, 2021.

György Buzsáki, Costas A Anastassiou, and Christof Koch. The origin of extracellular fields and currents—eeg, ecog, lfp and spikes. *Nature reviews neuroscience*, 13(6):407–420, 2012.

Yenho Chen, Noga Mudrik, Kyle A Johnsen, Sankaraleengam Alagapan, Adam S Charles, and Christopher J Rozell. Probabilistic decomposed linear dynamical systems for robust discovery of latent neural dynamics. *Advances in Neural Information Processing Systems*, 37:104443–104470, 2024.

Mukeshwar Dhamala, Govindan Rangarajan, and Mingzhou Ding. Analyzing information flow in brain networks with nonparametric granger causality. *Neuroimage*, 41(2):354–362, 2008.

Matthew Dowling and Cristina Savin. Nonlinear multiregion neural dynamics with parametric impulse response communication channels. In *International Conference on Learning Representations*, 2025.

Gaute T Einevoll, Christoph Kayser, Nikos K Logothetis, and Stefano Panzeri. Modelling and analysis of local field potentials for studying the function of cortical circuits. *Nature Reviews Neuroscience*, 14(11):770–785, 2013.

Pascal Fries. A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends in cognitive sciences*, 9(10):474–480, 2005.

Pascal Fries. Rhythms for cognition: communication through coherence. *Neuron*, 88(1):220–235, 2015.

Karl J Friston. Functional and effective connectivity: a review. *Brain connectivity*, 1(1):13–36, 2011.

Karl J Friston, André Bastos, Vladimir Litvak, Klaas E Stephan, Pascal Fries, and Rosalyn J Moran. Dcm for complex-valued data: cross-spectra, coherence and phase-delays. *Neuroimage*, 59(1): 439–455, 2012.

John Geweke. Measurement of linear dependence and feedback between multiple time series. *Journal of the American statistical association*, 77(378):304–313, 1982.

Evren Gokcen, Anna I Jasper, João D Semedo, Amin Zandvakili, Adam Kohn, Christian K Machens, and Byron M Yu. Disentangling the flow of signals between populations of neurons. *Nature Computational Science*, 2(8):512–525, 2022.

Evren Gokcen, Anna Jasper, Alison Xu, Adam Kohn, Christian K Machens, and Byron M Yu. Uncovering motifs of concurrent signaling across multiple neuronal populations. *Advances in neural information processing systems*, 36:34711–34722, 2023.

Marcin Jan Kaminski and Katarzyna J Blinowska. A new method of the description of the information flow in the brain structures. *Biological cybernetics*, 65(3):203–210, 1991.

Maral Kasiri, Sina Javadzadeh, Jaya Nataraj, Seyyed Alireza Seyyed Mousavi, and Terence Sanger. Correlated activity in globus pallidus and thalamus during voluntary reaching movement in three children with primary dystonia. *Dystonia*, 2:11117, 2023. doi: 10.3389/dyst.2023.11117.

Adam Kohn, Anna I Jasper, João D Semedo, Evren Gokcen, Christian K Machens, and Byron M Yu. Principles of corticocortical communication: proposed schemes and design considerations. *Trends in Neurosciences*, 43(9):725–737, 2020.

Weihan Li, Chengrui Li, Yule Wang, and Anqi Wu. Multi-region markovian gaussian process: An efficient method to discover directional communications across multiple brain regions. *Proceedings of machine learning research*, 235:28112–39404, 2024.

Weihan Li, Yule Wang, Chengrui Li, and Anqi Wu. Learning time-varying multi-region brain communications via scalable markovian gaussian processes. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 36021–36041. PMLR, 2025.

Belle Liu, Jacob Sacks, and Matthew D Golub. Accurate identification of communication between multiple interacting neural populations. *Proceedings of machine learning research*, 267:39381, 2025.

Matthew J McGinley, Martin Vinck, Jacob Reimer, Renata Batista-Brito, Edward Zagha, Cathryn R Cadwell, Andreas S Tolias, Jessica A Cardin, and David A McCormick. Waking state: rapid variations modulate neural and behavioral responses. *Neuron*, 87(6):1143–1161, 2015.

Rosalyn J Moran, Klaas E Stephan, Thomas Seidenbecher, H-C Pape, Raymond J Dolan, and Karl J Friston. Dynamic causal models of steady-state responses. *Neuroimage*, 44(3):796–811, 2009.

Noga Mudrik, Yenho Chen, Eva Yezerets, Christopher J Rozell, and Adam S Charles. Decomposed linear dynamical systems (dlds) for learning the latent components of neural dynamics. *Journal of Machine Learning Research*, 25(59):1–44, 2024.

Noga Mudrik, Ryan Ly, Oliver Ruebel, and Adam S Charles. Creimbo: Cross-regional ensemble interactions in multi-view brain observations. The International Conference on Learning Representations, 2025.

Cristopher M Niell and Michael P Stryker. Modulation of visual responses by behavioral state in mouse visual cortex. *Neuron*, 65(4):472–479, 2010.

Agostina Palmigiano, Theo Geisel, Fred Wolf, and Demian Battaglia. Flexible information routing by transient synchrony. *Nature neuroscience*, 20(7):1014–1022, 2017.

Bijan Pesaran, Martin Vinck, Gaute T Einevoll, Anton Sirota, Pascal Fries, Markus Siegel, Wilson Truccolo, Charles E Schroeder, and Ramesh Srinivasan. Investigating large-scale brain dynamics using field potential recordings: analysis and interpretation. *Nature neuroscience*, 21(7):903–919, 2018.

João D Semedo, Amin Zandvakili, Christian K Machens, Byron M Yu, and Adam Kohn. Cortical areas interact through a communication subspace. *Neuron*, 102(1):249–259, 2019.

S Alireza Seyyed Mousavi, Sina Javadzadeh, Mehrnaz Asadi, Richard B Ivry, and Terence D Sanger. Behaviorally coupled oscillations reveal distinct timing roles of basal ganglia and cerebellothalamic circuits in dystonia. *Dystonia*, 4:15390, 2025. doi: 10.3389/dyst.2025.15390.

S Murray Sherman. Thalamus plays a central role in ongoing cortical functioning. *Nature neuroscience*, 19(4):533–541, 2016.

S Murray Sherman and RW Guillery. Distinct functions for direct and transthalamic corticocortical connections. *Journal of neurophysiology*, 106(3):1068–1077, 2011.

Markus Siegel, Tobias H Donner, and Andreas K Engel. Spectral fingerprints of large-scale neuronal interactions. *Nature Reviews Neuroscience*, 13(2):121–134, 2012.

Josh Siegle, Wayne Wakeman, Xiaoxuan Jia, Gregg Heller, Tamina Ramirez, Nile Graddis, Nicholas Mei, and Séverine Durand. Allen institute - visual coding - neuropixels (brain observatory 1.1 stimulus set), 2025. URL `https://doi.org/10.48324/dandi.000021/0.251116.2246`.

Joshua H Siegle, Xiaoxuan Jia, Séverine Durand, Sam Gale, Corbett Bennett, Nile Graddis, Greggory Heller, Tamina K Ramirez, Hannah Choi, Jennifer A Luviano, et al. Survey of spiking in the mouse visual system reveals functional hierarchy. *Nature*, 592(7852):86–92, 2021.

David J Thomson. Spectrum estimation and harmonic analysis. *Proceedings of the IEEE*, 70(9): 1055–1096, 1982.

Francisco Varela, Jean-Philippe Lachaux, Eugenio Rodriguez, and Jacques Martinerie. The brainweb: phase synchronization and large-scale integration. *Nature reviews neuroscience*, 2(4):229–239, 2001.

A Technical Details and Supplementary Experiments

A.1 Compute Hardware

Experiments were performed on a workstation equipped with an AMD Ryzen Threadripper PRO 7955WX 16-Core CPU, 255.1 GB RAM, and an NVIDIA RTX 5000 Ada Generation GPU with 32 GB memory. CPU or GPU execution was used according to the implementation described for each experiment.

A.2 RAVEL Implementation and Reproducibility Details

This appendix specifies the implementation used for the experiments in this paper. RAVEL separates a shared propagation dictionary from quantities that are estimated independently for each recording. We summarize the complete pipeline below, from preprocessing and communication-coordinate estimation through dictionary selection, pooling, and transfer.

A.2.1 Data Representation and Communication Coordinates

For recording or subject s, trial n, event-time window w, multitaper index k, and frequency f, let $\mathbf{X}_{s,n,w,k}(f) \in \mathbb{C}^{p_s}$ and $\mathbf{Y}_{s,n,w,k}(f) \in \mathbb{C}^{q_s}$ denote source- and target-region Fourier coefficients in channel space. Their d-dimensional communication coordinates are

$$\mathbf{x}_{s,n,w,k}(f) = \mathbf{U}_s^\top \mathbf{X}_{s,n,w,k}(f), \qquad \mathbf{y}_{s,n,w,k}(f) = \mathbf{V}_s^\top \mathbf{Y}_{s,n,w,k}(f). \tag{A.1}$$

Training-defined preprocessing. All data-dependent transformations are estimated using training trials and held fixed for validation, test, and transfer data. Raw epochs are constant-detrended, low-pass filtered at $100\,\mathrm{Hz}$ with a zero-phase fourth-order Butterworth filter, and zero-phase notch filtered at $60\,\mathrm{Hz}$ with $Q = 30$. Each channel is robustly standardized using its training-set median and $1.4826 \times \mathrm{MAD}$, with the training-set standard deviation used as a fallback when the robust scale is numerically zero. The mean event-locked waveform estimated from training trials is then subtracted from every trial, so the conditional model targets trial-resolved activity beyond the reproducible event-locked mean.

Spectral representation. The reference implementation uses data sampled at $500\,\mathrm{Hz}$. Complex Fourier coefficients are computed in $0.40\,\mathrm{s}$ windows advanced by $0.25\,\mathrm{s}$, using $K = 3$ DPSS tapers (Thomson, 1982) with time-bandwidth product $NW = 2.5$. The usual one-sided Fourier normalization is applied. Frequencies from 6 to $80\,\mathrm{Hz}$ are retained, excluding 58–62 Hz. Importantly, coefficients remain resolved by trial, window, taper, frequency, and channel rather than being averaged into a cross-spectral-density matrix.

Communication coordinates. Communication axes are estimated independently for each recording using training time-domain residuals. For lag ℓ, let

$$\mathbf{C}_{s,\ell} = \mathrm{Cov}\left[\mathbf{S}_s(t), \mathbf{T}_s(t+\ell)\right], \tag{A.2}$$

where $\mathbf{S}_s(t)$ and $\mathbf{T}_s(t)$ are the preprocessed source and target signals. Using

$$\mathcal{L} = \{-30, -20, -10, 0, 10, 20, 30\}\ \mathrm{ms},$$

RAVEL forms

$$\mathbf{G}_s^x = \sum_{\ell \in \mathcal{L}} \mathbf{C}_{s,\ell} \mathbf{C}_{s,\ell}^\top, \qquad \mathbf{G}_s^y = \sum_{\ell \in \mathcal{L}} \mathbf{C}_{s,\ell}^\top \mathbf{C}_{s,\ell}. \tag{A.3}$$

The leading d eigenvectors of $\mathbf{G}_s^x$ and $\mathbf{G}_s^y$ define $\mathbf{U}_s$ and $\mathbf{V}_s$, respectively. Eigenvector signs are fixed deterministically by requiring the largest-magnitude entry to be positive. Communication axes are recording specific and are never averaged across subjects.

Extension to a prespecified multi-region graph. For a recording s, let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be the specified directed graph of observed local interactions. For every edge $(i, j) \in \mathcal{E}$, compute the lagged training-residual cross-covariance $C_{s,ij,\ell} = \mathrm{Cov}[S_{s,i}(t), S_{s,j}(t+\ell)]$ using only the training events assigned to that edge. The communication axes for region i are the leading d eigenvectors of a single score that combines its observed neighbors:

$$\begin{aligned} G_{s,i} = &\sum_{(i,j)\in\mathcal{E}} \sum_{\ell \in \mathcal{L}} C_{s,ij,\ell} C_{s,ij,\ell}^\top \\ &+ \sum_{(j,i)\in\mathcal{E}} \sum_{\ell \in \mathcal{L}} C_{s,ji,\ell}^\top C_{s,ji,\ell}, \\ U_{s,i} = &\ \mathrm{eig}_d(G_{s,i}). \end{aligned} \tag{A.4}$$

Here eig_d returns the eigenvectors with the d largest eigenvalues. The same $U_{s,i}$ projects region i whether it is the target of one fitted edge or the source of the next. The training-derived preprocessing transform for region i is likewise shared by its incident edges. With a single edge, this construction reduces to Eq. A.3. All transforms and axes are estimated from local-edge training data and then held fixed; the unobserved endpoint pair is never used to construct them. Each local transfer is fitted with the same conditional model, nuisance terms, delayed atoms, and recruitment procedure as in the pairwise setting. Multi-hop prediction composes these locally fitted transfers in their shared regional coordinates; no graph-wide transfer or additional atom is fitted.

A.2.2 Conditional Transfer Model and Dynamic Recruitment

For recording s, event-time window w, and frequency f, the conditional source-to-target transfer operator is

$$\mathbf{H}_{s,w}(f) = \mathbf{B}_{s,0} + \sum_{q=1}^{Q} \phi_q(f)\mathbf{B}_{s,q} + \sum_{m=1}^{M} r_{s,m}(w)\widetilde{h}_m(f)\mathbf{A}_{s,m}. \tag{A.5}$$

The corresponding prediction is

$$\widehat{\mathbf{y}}_{s,n,w,k}(f) = \mathbf{H}_{s,w}(f)\mathbf{x}_{s,n,w,k}(f). \tag{A.6}$$

The constant matrix $\mathbf{B}_{s,0}$ captures unrestricted instantaneous coupling. The remaining zero-phase functions $\phi_q(f)$ are constructed from orthogonalized powers of standardized log frequency; the reference implementation uses $Q = 2$. Their coefficient matrices $\mathbf{B}_{s,q}$ are restricted to rank one.

Each delayed propagation atom is

$$h_m(f) = \frac{\exp(-i2\pi f\tau_m/1000)}{1 + if/f_{c,m}}, \qquad \tau_m > 0, \quad f_{c,m} > 0, \tag{A.7}$$

where τ_m is the explicit transport delay in milliseconds and $f_{c,m}$ is the characteristic frequency of the first-order response. Because the pole also contributes frequency-dependent phase, τ_m is not the complete phase delay of the atom. Each atom is normalized to unit root-mean-square magnitude over the analyzed frequency grid, and its recording-specific spatial residue $\mathbf{A}_{s,m} \in \mathbb{R}^{d\times d}$ is restricted to rank one.

Dynamic recruitment. The nonnegative coefficient $r_{s,m}(w)$ determines the event-time recruitment of atom m. Recruitment is shared across trials entering a fit; it is not a separate coefficient for every trial. For fixed atoms and coefficient matrices, recruitment is estimated by minimizing

$$\sum_{n,w,k,f} \left\| \mathbf{z}^{(m)}_{s,n,w,k}(f) - r_{s,m}(w)\mathbf{q}^{(m)}_{s,n,w,k}(f) \right\|_2^2 + \rho_r\kappa_{s,m}\sum_{w}\left[r_{s,m}(w) - 1\right]^2 + \rho_\Delta\kappa_{s,m}\sum_{w}\left[r_{s,m}(w+1) - r_{s,m}(w)\right]^2, \tag{A.8}$$

subject to

$$0 \le r_{s,m}(w) \le r_{\max}.$$

Here

$$\mathbf{q}^{(m)}_{s,n,w,k}(f) = \widetilde{h}_m(f)\mathbf{A}_{s,m}\mathbf{x}_{s,n,w,k}(f)$$

is the unit-recruitment contribution of atom m, and $\mathbf{z}^{(m)}$ is the target residual after subtracting all other model components. The reference values are $r_{\max} = 3$, $\rho_r = 0.04$, and $\rho_\Delta = 0.20$, with $\kappa_{s,m}$ set to the median atom-energy scale across event-time windows.

Because $\mathbf{A}_{s,m}$ and $r_{s,m}(w)$ have a multiplicative scale ambiguity, ordinary fitting normalizes each recruitment profile to unit temporal mean and absorbs the corresponding scale into $\mathbf{A}_{s,m}$. This normalization is omitted during strict recruitment-only transfer, where the source-context residue matrices are frozen.

A.2.3 Parameter Estimation, Data Splits, and Held-Out Scoring

For a fixed propagation dictionary

$$\mathcal{D} = \{(\tau_m, f_{c,m})\}_{m=1}^{M}, \tag{A.9}$$

recording-specific parameters are fitted by minimizing complex source-to-target prediction error on the training data,

$$\mathcal{L}_s^{\text{train}} = \sum_{n,w,k,f} \|\mathbf{y}_{s,n,w,k}(f) - \mathbf{H}_{s,w}(f)\mathbf{x}_{s,n,w,k}(f)\|_2^2. \tag{A.10}$$

Optimization alternates between coefficient updates and recruitment updates. Coefficient matrices are estimated by real-valued ridge regression with $\lambda_{\mathrm{ridge}} = 0.002$. After each update, nuisance and propagation residue matrices are projected to rank one using their leading singular component, whereas the instantaneous matrix $\mathbf{B}_{s,0}$ remains full rank. The implementation uses up to four coefficient sweeps per coefficient-fitting step and three rounds of alternating coefficient and recruitment updates, followed by a final coefficient fit.

Development and test splits. Trials are partitioned into

$$60\% \text{ training}, \qquad 10\% \text{ validation-select}, \qquad 10\% \text{ validation-gate}, \qquad 20\% \text{ test}.$$

Splits are stratified by outcome when labels are available; if a validation or test partition contains fewer than three trials, a deterministic unstratified split is used. Training data determine preprocessing, communication axes, recording-specific parameters, recruitment, the nuisance comparator, and the evaluation covariance. Validation-select data are used for dictionary optimization, validation-gate data for restart and atom-count selection, and the sealed test set is not used until all choices are fixed.

Matched zero-phase comparator and held-out scoring. For each candidate delayed model, RAVEL fits a matched nuisance model

$$\mathbf{H}_s^{\mathrm{nuis}}(f) = \mathbf{B}_{s,0} + \sum_{q=1}^{Q} \phi_q(f)\mathbf{B}_{s,q}, \tag{A.11}$$

using the same preprocessing, communication coordinates, regression procedure, and evaluation covariance as the complete model.

Predictions are mapped back to target-channel space as

$$\widehat{\mathbf{Y}}_{s,n,w,k}(f) = \mathbf{V}_s \widehat{\mathbf{y}}_{s,n,w,k}(f). \tag{A.12}$$

Held-out prediction is scored with a frequency-specific complex-Gaussian negative log likelihood using a training-derived target covariance. If $\widehat{\mathbf{\Psi}}_s(f)$ is the empirical training second moment, the evaluation covariance is

$$\mathbf{\Psi}_s(f) = (1-\alpha)\widehat{\mathbf{\Psi}}_s(f) + \alpha \operatorname{diag}\left[\widehat{\mathbf{\Psi}}_s(f)\right] + \epsilon_s(f)\mathbf{I}, \qquad \alpha = 0.25. \tag{A.13}$$

For recording s, improvement beyond the nuisance model is summarized by

$$\Delta_s(\mathcal{D}) = \mathrm{NLL}_s^{\mathrm{full}}(\mathcal{D}) - \mathrm{NLL}_s^{\mathrm{nuis}}, \tag{A.14}$$

so $\Delta_s(\mathcal{D}) < 0$ indicates that delayed propagation terms improve held-out prediction. For pooled dictionary selection, candidate dictionaries are ranked using

$$\mathcal{J}(\mathcal{D}) = \mathrm{median}_s\, \Delta_s(\mathcal{D}) + 0.25\ \mathrm{mean}_s\, \Delta_s(\mathcal{D}). \tag{A.15}$$

A.2.4 Dictionary Optimization, Stability, and Atom Count

The reference dictionary search uses

$$2 \le \tau_m \le 45 \text{ ms}, \qquad 5 \le f_{c,m} \le 75 \text{ Hz},$$

with a minimum $1.5\,\mathrm{ms}$ separation between atom delays. Atoms are ordered by increasing delay to remove label ambiguity.

The generic initialization families are

M	Initial delays (ms)	Initial f_c (Hz)
1	14	10
2	7, 23	12, 8
3	3, 13, 29	24, 11, 7
4	2.5, 9, 21, 37	35, 15, 9, 6

At each restart, delays are perturbed by Gaussian noise with standard deviation $3\,\mathrm{ms}$, and characteristic frequencies are perturbed multiplicatively in log space with standard deviation 0.15. Dictionary refinement uses deterministic coordinate-pattern search; all recording-specific parameters are refitted on training data before evaluating a proposed dictionary on validation-select data.

During development locking, four restarts are used with delay steps $4, 2, 1\,\mathrm{ms}$, and both delays and characteristic frequencies are eligible for refinement. During pooled fitting, three restarts are used with delay steps $2, 1\,\mathrm{ms}$. Both delay and $f_{c,m}$ are jittered at the beginning of a pooled restart, after which $f_{c,m}$ is held fixed while delay is locally refined.

Restart stability and consensus selection. Atoms across restarts are matched one-to-one with the Hungarian algorithm using complex atom-shape similarity. A matched atom is counted as supported when its normalized complex-response similarity is at least 0.90 and its delay differs by no more than $6\,\mathrm{ms}$; delay median absolute deviation is also recorded.

Restarts whose validation-select objective lies within one standard error of the best solution are treated as competitive. Among these solutions, a consensus medoid is identified using a distance that combines atom-shape dissimilarity with a 0.02-per-ms delay penalty. The validation-select best solution and the consensus medoid are then compared on validation-gate data; the consensus solution is retained when its mean validation-gate score is within one standard error of the better candidate.

Atom-count selection. When model order is selected from development data,

$$M \in \{1, 2, 3, 4\}.$$

Models within one standard error of the lowest validation-gate score are treated as competitive. RAVEL selects the smallest competitive M whose minimum atom restart-support fraction is at least 0.75. If none satisfies this stability criterion, the smallest atom count within one standard error is selected.

A.2.5 Pooling and Cross-Context Transfer

For S recordings, pooled RAVEL shares only the propagation dictionary

$$\mathcal{D} = \{(\tau_m, f_{c,m})\}_{m=1}^{M}.$$

Each recording retains its own

$$\mathbf{U}_s, \mathbf{V}_s, \mathbf{B}_{s,0}, \mathbf{B}_{s,q}, \mathbf{A}_{s,m}, r_{s,m}(w).$$

Pooling therefore tests whether delayed spectral responses are reusable without requiring recordings to share channel geometry, communication axes, spatial residues, nuisance coupling, or recruitment trajectories. In the independent comparison, each recording instead optimizes its own dictionary $\mathcal{D}_s$.

For cross-context transfer, a dictionary learned in source context A,

$$\mathcal{D}^A = \left\{(\tau_m^A, f_{c,m}^A)\right\}_{m=1}^{M},$$

is held fixed in target context B. In the primary frozen-physics adaptation, target-context nuisance coefficients, spatial residues, and recruitment are refitted while retaining the source-context preprocessing and communication axes. A stricter recruitment-only transfer additionally freezes the source-context coefficient matrices and estimates only $r_{s,m}^B(w)$. A static-recruitment control fixes $r_{s,m}^B(w) = 1$, while a target-context-from-scratch comparison relearns the propagation dictionary using the same fitting and validation procedure.

A.2.6 Interpretational Scope

RAVEL identifies components of a directed pairwise conditional predictor whose spectral structure is consistent with delayed first-order propagation and whose recruitment may vary over event time. The parameter τ_m describes the explicit pure-transport factor

$$\exp(-i2\pi f \tau_m / 1000),$$

and is therefore distinct from the complete frequency-dependent phase lag of $h_m(f)$, which also contains the phase contribution of the one-pole filter.

An inferred atom should be interpreted as a predictive delayed component only when it contributes held-out predictive value beyond the matched zero-phase comparator and is stable under the prescribed fitting procedure. Because RAVEL is an observational pairwise model, recovered atoms are reusable pairwise propagation components rather than uniquely identified anatomical pathways without additional anatomical or interventional evidence.

A.3 Synthetic Mechanism-Discrimination Benchmark

A.3.1 Experimental Objective and Model Order

This benchmark was designed to distinguish two mechanistically different explanations for event-dependent changes in cross-region communication: (i) stable propagation mechanisms whose recruitment changes over event time, and (ii) communication mechanisms whose physical transport delays themselves change over event time. The benchmark was not intended to test model-order selection. The generative process contained two delayed propagation components, so RAVEL was fit with the true atom count $M = 2$. ADM's across-region latent dimensionality K has a different interpretation and is not equivalent to RAVEL's atom count. We therefore did not enforce $K = 2$ as a capacity match. A preliminary fixed-$K = 2$ ADM fit was considered, but the final comparison used ADM's native model-selection procedure over $K \in \{1, 2, 3\}$, selecting K by validation marginal likelihood without access to the common held-out evaluator.

A.3.2 Synthetic Data Generation

Each Monte Carlo replicate consisted of two observed regions with six channels per region. We generated 40 trials sampled at 500 Hz from -1 to $+2$ s relative to the synthetic event, giving 1500 samples per trial. The same preprocessing and spectral representation used in the main method were applied to these raw synthetic epochs. Each replicate used a single low-dimensional source communication process $z(t)$ and independent random source and target channel loadings. The three reported seeds were

$$271101, \qquad 271102, \qquad 271103.$$

The generative mechanism was held fixed across seeds, while latent realizations, channel loadings, and observation noise were independently resampled.

For each replicate, trials were randomly divided into 60% training, 10% `val_select`, 10% `val_gate`, and 20% sealed test sets, corresponding to

$$24/4/4/8$$

trials. The test split was not used for fitting or model selection.

The broadband source process was generated by Fourier-domain filtering of white noise. Its spectral amplitude was proportional to $f^{-\alpha/2}$ above a 2 Hz low-frequency corner with $\alpha = 0.35$, together with a smooth high-frequency cutoff at 105 Hz. Target-local colored activity used $\alpha = 0.70$, a 2 Hz low-frequency corner, and a 90 Hz high-frequency cutoff. Source sensor noise had standard deviation 0.005, target-local latent noise had standard deviation 0.10, and target sensor noise had standard deviation 0.05. Channel loading magnitudes were perturbed independently using a log-normal factor with log-standard-deviation 0.10. A 0.75 s temporal pad was generated on each side before applying delays and filters to reduce edge artifacts.

Dynamic-recruitment world The latent target communication signal was generated as

$$y_{\mathrm{comm}}(t) = \sum_{m=1}^{2} g_m r_m(t) \left[h(\tau_m, p_m) * z\right](t), \tag{A.16}$$

where

$$h_m(f) = \frac{\exp(-i2\pi f \tau_m/1000)}{1 + if/p_m}. \tag{A.17}$$

The two transport delays and poles were fixed at

$$(\tau_1, \tau_2) = (9.5, 21.5)\ \text{ms}, \qquad (p_1, p_2) = (45, 45)\ \text{Hz},$$

with equal component gains

$$(g_1, g_2) = (0.70, 0.70).$$

Recruitment varied smoothly over the 11 event-time windows. Let $u \in [0, 1]$ denote normalized event-window position and define

$$s(u) = \sigma\left(\frac{u - 0.5}{0.13}\right), \qquad b(u) = \exp\left[-\frac{1}{2}\left(\frac{u - 0.5}{0.16}\right)^2\right]. \tag{A.18}$$

The unnormalized profiles were

$$\tilde{r}_1(u) = 0.38 + 1.30[1 - s(u)] + 0.10\, b(u), \tag{A.19}$$

$$\tilde{r}_2(u) = 0.38 + 1.30 s(u) + 0.075\, b(u), \tag{A.20}$$

and each profile was normalized to unit temporal mean. The window-level profiles were linearly interpolated to the 500 Hz sample grid before generating the target signal. Thus, the observed temporal change was produced by recruitment alone; the physical delays and poles were constant.

Dynamic-delay world The second world used the same dimensionality and preprocessing but reversed the source of temporal variation. Component gains were constant,

$$(g_1, g_2) = (0.75, 0.65),$$

and both poles were fixed at 45 Hz. Recruitment was therefore constant over event time. Instead, the transport delays changed smoothly according to normalized logistic trajectories:

$$\tau_m(u) = \tau_m^{\text{start}} + \left(\tau_m^{\text{end}} - \tau_m^{\text{start}}\right) q_m(u), \tag{A.21}$$

where $q_m(0) = 0$, $q_m(1) = 1$, and q_m is obtained by normalizing $\sigma[(u - 0.5)/\omega_m]$. We used

$$(\tau_1^{\text{start}}, \tau_2^{\text{start}}) = (7.5, 19.0)\ \text{ms},$$

$$(\tau_1^{\text{end}}, \tau_2^{\text{end}}) = (14.0, 27.0)\ \text{ms},$$

with logistic widths

$$(\omega_1, \omega_2) = (0.16, 0.14).$$

The window-level delay trajectories were interpolated to the sample grid and implemented using fractional temporal delays followed by the corresponding one-pole filter.

This synthetic experiment intentionally contains a relatively clean communication signal.

A.3.3 RAVEL Fitting and Ablations

Raw simulated epochs were passed through the same `build_route_session_from_epochs()` pipeline used by the main method. The synthetic benchmark used one communication coordinate ($d = 1$). RAVEL was fit using the frozen independent-session optimizer with $M = 2$. Each fit used the three native initialization strategies, full delay/pole refinement, and the standard `val_select`/`val_gate` restart-selection procedure. The sealed test trials were scored only after model selection.

Three RAVEL variants were evaluated:

1. **Full RAVEL**: two fixed delayed atoms with event-time-varying recruitment;
2. **Static-recruitment RAVEL**: the same two fitted atoms, but recruitment constrained to be time invariant;
3. **Nuisance only**: the zero-phase constant and smooth frequency-dependent nuisance terms with no delayed atoms.

These ablations distinguish the predictive contribution of delayed communication from the additional contribution of time-varying recruitment.

A.3.4 MATCHED DYNAMIC-DELAY COMPARATOR

To isolate the stable-delay versus dynamic-delay assumption without introducing a different observation model, we constructed a matched two-component alternative in the same communication coordinates used by RAVEL:

$$H(w, f) = H_{\text{zero}}(f) + a_1 h(f; \tau_1(w), p) + a_2 h(f; \tau_2(w), p). \tag{A.22}$$

Unlike RAVEL, the route gains a_1, a_2 are constant over event time while the two delay trajectories $\tau_1(w), \tau_2(w)$ are free to vary. The model retains the same zero-phase nuisance terms and is evaluated with the same held-out evaluator.

The common pole was selected from

$$p \in \{30, 45, 60\} \text{ Hz},$$

and delay-path smoothness from

$$\lambda \in \{0.05, 0.20, 0.80\}.$$

Delay trajectories were optimized on a 1 ms grid over the physical delay range, with a minimum 2 ms separation between the two components. Initialization used delays of 10 and 23 ms. Four alternating updates of static coefficients and delay paths were performed. Candidate pole/smoothness settings were selected by `val_select` NLL, after which `val_gate` and sealed-test scores were computed.

A.3.5 ADM FITTING AND MODEL SELECTION

ADM was fit using its native implementation rather than optimizing it with RAVEL's validation objective. For each seed and generative world, we fit candidate across-region latent dimensions

$$K \in \{1, 2, 3\}.$$

Each candidate used pCCA initialization followed by the package's rank-one deflation procedure (200 optimization iterations per round) and 300 Kalman–EM iterations. We used one within-region latent per region, `lag_across`$= 2$, `lag_within`$= 2$, initial kernel scales of 0.1, learning rate 5×10^{-3}, minimum learning rate 10^{-3}, and weight decay 0.01. Computations used double precision.

ADM model order was selected using the lowest mean native marginal NLL across `val_select` and `val_gate`; the test set was not opened during this selection. The selected K values were

$$(2, 1, 3)$$

for the three dynamic-recruitment replicates and

$$(3, 3, 2)$$

for the three dynamic-delay replicates.

For complete-target evaluation, all target-region observations were removed during inference: target values were set to zero and their observation variance was set to 10^{10}. ADM then reconstructed the complete target region from source-region observations. The predicted target time series were passed through the frozen RAVEL target transform and multitaper spectral representation before scoring with exactly the same held-out evaluator used for RAVEL and the matched dynamic-delay comparator. Thus, ADM retained its native training and model-selection objective, while only the final cross-model evaluation was shared.

As an implementation sanity check, we also ran the same ADM checkout in the same environment on ADM's native synthetic demonstration. On that native generative model, ADM obtained a reported delay RMSE of 0.160 and observation RMSE of 0.223, confirming that the installation could recover dynamic delays under its own generative assumptions. This sanity check was not used for hyperparameter selection.

A.3.6 EVALUATION AND MECHANISM-RECOVERY METRICS

All final predictive comparisons used the same eight sealed test trials and complete-target source-only spectral evaluation. We report NMSE as the primary predictive metric in the main text; NLL

and spectral R^2 are saved for every test trial and can be reported alongside NMSE in supplementary tables.

For the dynamic-recruitment world, fitted RAVEL atoms were matched to the two ground-truth atoms using Hungarian assignment on complex atom-shape similarity. We then measured absolute delay error, complex atom-shape similarity, and Pearson correlation between the matched true and inferred recruitment profiles. For the dynamic-delay comparator, fitted delay trajectories were Hungarian-matched to the ground-truth trajectories using trajectory RMSE.

For the dynamic-delay world, each moving ground-truth delay trajectory was matched either to the corresponding inferred dynamic-delay trajectory or to a fixed RAVEL atom. We report trajectory RMSE and, for the dynamic model, trajectory correlation.

Because the benchmark contains only three Monte Carlo seeds, results are reported descriptively as mean $\pm$ SD; we do not use the three seeds as the basis for asymptotic significance tests.

Table A.1: **Synthetic held-out complete-target prediction.** Spectral NMSE (mean $\pm$ SD across three independently generated seeds; lower is better).

Model	Stable recruitment	Dynamic delay
RAVEL	0.212 ± 0.034	0.319 ± 0.038
Matched dynamic delay	0.217 ± 0.033	$\mathbf{0.168 \pm 0.040}$
RAVEL, static recruitment	0.344 ± 0.023	0.333 ± 0.032
Nuisance only	0.827 ± 0.002	0.802 ± 0.003
ADM, native K selection	1.135 ± 0.266	0.993 ± 0.009

Table A.2: **Synthetic mechanism recovery.** Route-level values pool matched route estimates across the three seeds unless otherwise noted.

Truth/model	Metric	Result
Stable / RAVEL	absolute delay error	1.46 ± 0.87 ms
Stable / RAVEL	atom-shape similarity	0.955 ± 0.042
Stable / RAVEL	recruitment correlation	0.999 ± 0.001
Stable / matched delay	trajectory RMSE, route 1	4.94 ± 0.04 ms
Stable / matched delay	trajectory RMSE, route 2	6.76 ± 0.09 ms
Dynamic / matched delay	mean trajectory RMSE	0.321 ms
Dynamic / matched delay	mean trajectory correlation	0.995
Dynamic / RAVEL	truth-to-fixed-atom trajectory RMSE	5.78 ± 4.07 ms

Exact pole recovery was less identifiable than the combined complex transfer shape. Across the six matched RAVEL atoms in the dynamic-recruitment simulations, mean absolute pole error was 18.57 Hz, whereas mean complex atom-shape similarity remained 0.955. We therefore use transport-delay error, full atom-shape similarity, recruitment recovery, and held-out prediction as the primary recovery criteria rather than exact pole error alone.

A.4 Allen VBN: Pooled Propagation Physics, Cross-Context Transfer, and Recruitment

Cohort and preprocessing. The analysis in Fig. 3 used the directed VISam→VISpm connection in 25 Visual Behavior Neuropixels sessions from 25 mice (Table A.3). A session was eligible when both regions contained at least six valid LFP channels, and up to 12 channels per region were sampled approximately uniformly across probe depth. Active-task changes and passive image-change presentations were aligned separately; the first 30 s of each stimulus block were excluded, and epochs from -1 to 2 s around each event were retained. At most 180 active and 120 passive events were sampled approximately uniformly over time, with at least 35 aligned events required in each context. Preprocessing and spectral feature construction otherwise followed Appendix A.2.

Data splits. Active and passive events were split separately using the standard 60/10/10/20 RAVEL partitioning scheme (Appendix A.2); active-task splits were stratified by behavioral outcome when possible. The same splits were used for every model and control.

Table A.3: **Allen VBN cohort used for the VISam→VISpm analysis.** The analysis included one recording session from each of 25 mice. Superscript † denotes the six prespecified development sessions used for model-order selection. After model order was locked, the training partitions of all 25 sessions contributed to the final pooled fit.

Allen ecephys session IDs				
1052342277	1053709239†	1053718935	1055221968	1062755779
1064400234	1064415305	1065437523†	1065449881	1067588044†
1081079981	1090800639	1092283837	1093638203†	1095138995
1108334384	1108335514	1111013640	1115086689†	1116941914†
1118324999	1118327332	1124507277	1128520325	1130113579

Model-order selection and learned dictionary. The six development sessions identified in Table A.3 were used to select $M \in \{2, 3, 4\}$. Each candidate was fitted from four perturbed restarts. Within each candidate model order, the validation-gate procedure selected between the validation-select best restart and the consensus medoid. The resulting validation-gate mean $\Delta\mathrm{NLL} \pm \mathrm{SE}$ values were -0.004606 ± 0.003126, -0.009283 ± 0.002909, and -0.015191 ± 0.003284 for $M = 2$, $M = 3$, and $M = 4$, respectively, where more negative values indicate greater improvement over the zero-phase nuisance model.

The prespecified rule selected the smallest model within one standard error of the best validation-gate result whose minimum atom restart-support fraction was at least 0.75. Only $M = 4$ fell within the one-standard-error set, but its minimum restart-support fraction was 0.50. Because no model in the one-standard-error set satisfied the stability criterion, the prespecified fallback rule retained $M = 4$. The communication dimension was fixed at $d = 3$.

After locking M, the pooled dictionary was re-estimated using the active training partitions from all 25 sessions. The final learned dictionary was After locking M, the pooled dictionary was re-estimated using the active training partitions from all 25 sessions. The final learned dictionary was

$$\begin{aligned}\widehat{\mathcal{D}}_{\text{active}} = \big\{&(3.000\ \text{ms},\ 12.930\ \text{Hz}), (4.855\ \text{ms},\ 19.902\ \text{Hz}),\\ &(21.443\ \text{ms},\ 5.075\ \text{Hz}), (35.012\ \text{ms},\ 5.000\ \text{Hz})\big\},\end{aligned} \tag{A.23}$$

where each pair gives $(\tau_m, f_{c,m})$. where each pair gives $(\tau_m, f_{c,m})$.

After one-to-one atom matching, the fractions of independently fitted mice whose delays fell within 5 ms of the corresponding pooled delay were 0.96, 0.52, 0.84, and 0.72 for Motifs 1–4, respectively. Their median complex-response similarities to the pooled atoms were 0.980, 0.858, 0.902, and 0.839.

Cross-context transfer and wrong-physics controls. For active-to-passive transfer, the pooled active-trained dictionary, preprocessing transforms, and communication axes were frozen, while passive-specific coefficients and recruitment were refitted as described in Appendix A.2. The passive-from-scratch comparator instead relearned the pooled dictionary from passive training trials. For the passive wrong-physics control, each wrong dictionary was frozen in the same manner as the correct active-trained dictionary, while the same passive-specific parameters were refitted on the identical passive training trials.

Capacity-matched wrong-physics dictionaries. The wrong-physics controls were generated by a prespecified procedure using the learned dictionary but without consulting validation or test prediction scores. Candidate dictionaries contained $M = 4$ one-pole atoms, matching the capacity of the learned dictionary. For each candidate, delays were sampled uniformly from $[2, 45]$ ms and sorted, characteristic frequencies were sampled log-uniformly from $[5, 75]$ Hz, and candidates with less than 1.5 ms separation between adjacent delays were rejected. Generation used `numpy.random.default_rng` with deterministic seed 665054876, derived from the experiment base seed 20260921.

Candidate atoms were matched one-to-one to the learned atoms using normalized complex-response similarity over the analyzed frequency grid. A candidate dictionary was accepted only when (i) its mean matched similarity was at most 0.78, (ii) its maximum individual matched similarity was at

most 0.995, (iii) the candidate span explained at most 0.80 of the nuisance-residualized learned-kernel energy, and (iv) the condition number of its stacked real–imaginary kernel design was at most 500. The first eight candidates satisfying all criteria were retained. Their exact parameters are listed in Table A.4.

Table A.4: **Capacity-matched wrong-physics dictionaries.** Each atom is reported as (τ, f_c) in units of $(\mathrm{ms}, \mathrm{Hz})$. $\overline{s}$ is the mean matched complex-response similarity to the learned dictionary, and E_{span} is the fraction of nuisance-residualized learned kernel energy explained by the candidate span.

Control	Atom 1	Atom 2	Atom 3	Atom 4	$\overline{s}$	E_{span}
1	$(16.782, 43.845)$	$(18.987, 28.521)$	$(22.461, 32.429)$	$(41.349, 28.289)$	0.612	0.391
2	$(10.009, 74.664)$	$(14.664, 17.155)$	$(24.629, 21.271)$	$(29.960, 34.508)$	0.765	0.521
3	$(6.518, 6.799)$	$(22.825, 28.979)$	$(27.552, 44.150)$	$(35.106, 54.949)$	0.749	0.704
4	$(16.307, 5.965)$	$(26.365, 5.462)$	$(38.997, 24.479)$	$(40.883, 31.816)$	0.632	0.410
5	$(12.334, 11.483)$	$(23.872, 34.727)$	$(26.815, 31.876)$	$(38.087, 74.077)$	0.667	0.473
6	$(25.224, 68.445)$	$(31.460, 10.164)$	$(34.464, 74.224)$	$(36.293, 7.192)$	0.573	0.493
7	$(3.520, 47.992)$	$(13.303, 69.872)$	$(33.213, 62.497)$	$(37.532, 19.667)$	0.719	0.671
8	$(11.260, 7.417)$	$(28.935, 6.900)$	$(33.134, 7.420)$	$(43.773, 6.296)$	0.725	0.433

For each accepted dictionary, recording-specific nuisance coefficients, motif residues, and recruitment profiles were fitted using exactly the same training trials as the corresponding learned-physics model. Test performance was first averaged over the eight wrong dictionaries within each mouse, preserving the mouse as the statistical unit.

Recruitment analyses. For the hit–miss analysis, pooled physics and recording-specific residue matrices were held fixed and recruitment alone was refitted using the relevant non-test trials. Passive recruitment used the stricter recruitment-only transfer model, retaining the active residue matrices to preserve recruitment amplitude comparability across contexts. Split-half reliability was computed by independently refitting recruitment on two disjoint trial subsets and correlating the resulting time-by-motif profiles. The hit–miss recruitment comparison was descriptive; no inferential test over the recruitment time courses was performed. Motif 3 showed the clearest visual separation between the hit and miss profiles, whereas Motif 1 showed little outcome-dependent separation.

Statistical analysis. All model comparisons were paired within mouse ($n = 25$). For contrasts involving wrong physics, the eight wrong-dictionary scores were first averaged within mouse. For each of the six prespecified contrasts, we report the mean paired NLL difference, a 95% bootstrap percentile interval from 2,000 paired resamples, and a two-sided paired Wilcoxon signed-rank test. Benjamini–Hochberg correction was applied jointly across the six contrasts. In Table A.5, negative $\Delta\mathrm{NLL} = \mathrm{NLL}_A - \mathrm{NLL}_B$ favors model A.

Table A.5: **Mouse-level statistical comparisons for shared physics and cross-context transfer.** "Wrong" denotes the within-mouse average over the eight capacity-matched wrong dictionaries.

Contrast $A - B$	Mean ΔNLL	95% CI	p	BH-adjusted q
Correct pooled − wrong pooled	-0.02914	$[-0.03688, -0.02170]$	5.96×10^{-8}	1.19×10^{-7}
Independent − wrong pooled	-0.03181	$[-0.04161, -0.02345]$	5.96×10^{-8}	1.19×10^{-7}
Correct pooled − independent	0.00267	$[-0.00325, 0.01028]$	0.7712	0.7712
Frozen active − wrong frozen	-0.03033	$[-0.04355, -0.01594]$	3.76×10^{-4}	5.65×10^{-4}
Passive scratch − wrong frozen	-0.02942	$[-0.03647, -0.02258]$	5.96×10^{-8}	1.19×10^{-7}
Frozen active − passive scratch	-0.00091	$[-0.01466, 0.01497]$	0.2635	0.3162

The mean wrong-dictionary gain relative to the zero-phase nuisance model was -0.00388 in the active context and -0.00923 in the passive context. These values establish that the average wrong-physics performance lay below the nuisance reference; the formal inferential family in Table A.5 tests the learned models against the wrong-physics controls rather than testing each wrong-physics mean directly against zero.

A.5 Disjoint Three-Hop Composition: Preparation and Evaluation

Sessions and LFP preparation. We prespecified LGd, VISp, VISl, and VISpm and analyzed session 1108335514 using the strict disjoint arm, with a maximum of 12 selected LFP channels per region. For each region, the notebook selected a probe with valid region-assigned channels and sampled up to 12 channels across its depth. Active image-change trials with valid event times were retained after a 30 s block-start buffer; the full $[-1, 2]$ s window around each change had to lie within the recording block. At most 180 events were sampled evenly in time. LFP epochs spanned $[-1, 2]$ s relative to change, were interpolated to 500 Hz, demeaned within epoch, low-pass filtered at 100 Hz, and filtered with a zero-phase 60 Hz notch. Events were aligned by ID across all four regions.

Observation graph and local fits. Chronologically sorted aligned events were divided into five approximately equal blocks. Block 1 exposed LGd–VISp, block 2 VISp–VISl, block 3 VISl–VISpm, and block 4 only LGd and VISpm for the sealed endpoint test. Block 5 was unused in the disjoint arm. Each edge's block was divided deterministically into 60% training, 10% atom-selection validation, 10% validation gate, and 20% local test trials. Region-wise robust centering/scaling and mean event templates used only the training events of that region's incident edges. One $d = 3$ coordinate space was shared by each node's incident edges: its axes were the leading eigenvectors of the sum of the lagged cross-covariance scores computed separately on the observed training trials of each neighboring edge, at lags $-30, -20, -10, 0, 10, 20, 30$ ms. No LGd–VISpm cross-covariance or direct transfer was estimated. The nonobserved region/block combinations were masked before spectral features were constructed.

The existing Phase 0H v0.5 pairwise fitter was applied to each adjacent edge, with one delayed one-pole motif ($M = 1$) and the original constant and zero-phase frequency-basis terms. Its delay, pole, residue, and smooth event-time recruitment were estimated using local training data and selected by the local validation procedure. Recruitment is shared across the fitting trials at each event-relative time window; it is fixed across endpoint test events. Multitaper Fourier features used 0.40 s windows stepped by 0.25 s, three DPSS tapers with time-bandwidth product 2.5, and 6–80 Hz frequencies excluding 58–62 Hz. The three fitted transfers were multiplied in shared regional coordinates; only LGd was supplied at endpoint inference.

Scoring and controls. The endpoint reference is the channel-wise mean VISpm event-aligned waveform computed from third-link training trials. Its Fourier representation is subtracted from held-out VISpm before scoring, so a zero residual prediction corresponds to predicting that mean waveform rather than zero LFP. The same reference and frequency-specific, training-derived shrinkage covariance were used to score the full composition and every control on identical held-out events. The reported R^2 is the mean across events of $1 - \|Y - \widehat{Y}\|^2/\|Y\|^2$ in the retained Fourier representation, with Y the reference-subtracted VISpm channel data. A plotted likelihood gain is $\mathrm{NLL}_{\mathrm{reference}} - \mathrm{NLL}_{\mathrm{model}}$; thus, positive values indicate improvement over the training mean. A nuisance-only chain was independently refitted without delayed motifs. The No motifs and No DDD controls remove, without refitting, either all delayed components or specifically the product of the three delayed components. Twenty permutations of source test-trial identity formed the LGd-shuffle control with the full fit fixed. These comparisons test held-out prediction under the specified observation constraint; common inputs or other anatomical paths remain possible.

A.6 Allen Visual Coding Benchmarks

A.6.1 Primary Across-Session Benchmark

Sessions and evaluation condition. The primary cross-session benchmark used eight Allen Visual Coding recording sessions: 721123822, 751348571, 754829445, 755434585, 757216464, 760345702, 791319847, and 797828357. For this analysis, the stimulus condition was fixed to the 135° drifting-grating orientation, allowing recording session rather than orientation to serve as the paired unit of evaluation. Each session contributed 10 TRAIN and 5 sealed TEST trials for this orientation. Signals were represented as 2-s trials at 100 Hz.

The complete-target prediction task was evaluated independently in both directions. For VISp←VISam, all VISam channels were observed and all VISp channels were withheld during

inference. For VISam←VISp, all VISp channels were observed and all VISam channels were withheld. Therefore, prediction could not rely on simultaneously observed channels from the target region.

Preprocessing and evaluation. Signals were demeaned per trial and channel and standardized using TRAIN statistics only. Predictions and observations were trial-detrended, multiplied by a periodic Hann window, transformed with an rFFT, and evaluated over 2–40 Hz. The primary metric was spectral normalized mean-squared error,

$$\mathrm{NMSE} = \frac{\sum |\widehat{Y} - Y|^2}{\sum |Y|^2}.$$

We additionally report

$$R^2_{\mathrm{zero}} = 1 - \mathrm{NMSE}$$

and the magnitude of the complex spectral correlation

$$\rho_{\mathbb{C}} = \frac{|\langle Y, \widehat{Y} \rangle|}{\sqrt{\langle Y, Y \rangle \langle \widehat{Y}, \widehat{Y} \rangle}}.$$

The R^2_{zero} metric is an affine transformation of NMSE and is reported for interpretability, whereas complex spectral correlation provides a complementary scale-insensitive measure of spectral agreement (Table A.6).

Table A.6: **Secondary metrics for the cross-session complete-target benchmark.** Values are mean ± SD across eight recording sessions at 135°. Higher is better for both metrics.

	VISp ← VISam		VISam ← VISp	
Model	R^2_{zero}	Corr.	R^2_{zero}	Corr.
RAVEL	$\mathbf{0.298 \pm 0.190}$	$\mathbf{0.552 \pm 0.150}$	$\mathbf{0.500 \pm 0.158}$	$\mathbf{0.706 \pm 0.113}$
MRM-GP	0.224 ± 0.135	0.468 ± 0.136	0.427 ± 0.175	0.667 ± 0.117
DLAG	0.230 ± 0.131	0.488 ± 0.129	0.241 ± 0.304	0.645 ± 0.137

Per-session results. Table A.7 reports spectral NMSE for each individual recording session. RAVEL had lower NMSE than MRM-GP in 5/8 sessions and than DLAG in 4/8 sessions for VISp←VISam. For VISam←VISp, RAVEL had lower NMSE than MRM-GP in 6/8 sessions and than DLAG in all 8/8 sessions.

Table A.7: **Per-session spectral NMSE at 135°.** Lower is better. Each session is evaluated on its sealed TEST trials.

	VISp ← VISam			VISam ← VISp		
Session	RAVEL	MRM-GP	DLAG	RAVEL	MRM-GP	DLAG
721123822	0.532	0.525	0.516	0.346	0.332	0.358
751348571	0.713	0.709	0.697	0.353	0.420	0.519
754829445	**0.393**	0.686	0.707	**0.302**	0.501	0.582
755434585	0.892	0.897	**0.876**	**0.501**	0.533	0.834
757216464	0.991	0.908	**0.885**	**0.743**	0.787	1.206
760345702	**0.690**	0.748	0.762	0.521	**0.488**	0.537
791319847	**0.635**	0.831	0.810	**0.581**	0.790	0.964
797828357	**0.772**	0.903	0.910	**0.654**	0.734	1.076

Model selection and fitting. All model and architecture selection was performed using TRAIN data without access to sealed TEST trials. The configurations used for the benchmark are summarized in Table A.8. These configurations were frozen for the cross-session analysis.

RAVEL motif count was selected from $M \in \{1, 2, 3, 4\}$ using TRAIN data and was subsequently fixed to $M = 3$ for both prediction directions. ADM and DLAG capacities were selected from $n_{\mathrm{across}}, n_{\mathrm{within}} \in \{1, 2, 3, 4\}$, yielding $(1, 2)$ for ADM and $(1, 4)$ for DLAG. MRM-GP used the established Allen LFP configuration with one state, one across-region latent, and three within-region latents per region.

Table A.8: **Frozen configurations for the Allen Visual Coding benchmarks.** Capacity selection used TRAIN data only. ADM was included in the complementary single-session orientation benchmark described below.

Model	Frozen configuration	Training / selection
RAVEL	$d = 4,\ M = 3$	conditional spectral regression
MRM-GP	$K = 1$, 1 across, 3 within/region	native EM / ELBO objective
DLAG	1 across, 4 within/region	pCCA + exact EM
ADM	1 across, 2 within/region	marginal likelihood

Table A.9: **Exact paired tests across eight recording sessions.** Δ NMSE denotes $\mathrm{NMSE}_{\mathrm{baseline}} - \mathrm{NMSE}_{\mathrm{RAVEL}}$; positive values favor RAVEL. Holm correction is across the four directional pairwise comparisons.

Direction	Comparison	Δ NMSE	RAVEL wins	p	p_{Holm}
VISp ← VISam	RAVEL vs. MRM-GP	0.0738	5/8	0.1719	0.3438
VISp ← VISam	RAVEL vs. DLAG	0.0682	4/8	0.1953	0.3438
VISam ← VISp	RAVEL vs. MRM-GP	0.0731	6/8	0.0469	0.1406
VISam ← VISp	RAVEL vs. DLAG	0.2594	8/8	0.0078	**0.0313**

For the eight-session cross-session experiment, each method was fit once per session using a fixed training seed. Consequently, the dispersion reported in Tables 1 and A.6 is across recording sessions and should not be interpreted as optimization-seed variability.

Training-objective alignment. The benchmark is matched in data, TRAIN/TEST splits, complete-target prediction task, and evaluation metric, but not in training objective. RAVEL directly optimizes conditional source-to-target spectral prediction, whereas MRM-GP and DLAG retain their native generative objectives and are converted to source-only conditional predictors at evaluation. The same principle was used for ADM in the single-session benchmark below. We preserve the native baseline procedures rather than modifying their estimators to optimize the RAVEL objective.

Cross-session statistical analysis. Recording session was treated as the paired unit of analysis. For each baseline and direction, we formed

$$d_s = \mathrm{NMSE}_{\mathrm{baseline},s} - \mathrm{NMSE}_{\mathrm{RAVEL},s},$$

where positive values indicate lower NMSE for RAVEL. We used an exact two-sided paired sign-flip test across the eight sessions and applied Holm correction across the four RAVEL-versus-baseline directional comparisons. We additionally performed a Friedman omnibus test across RAVEL, MRM-GP, and DLAG separately for each prediction direction.

For VISp←VISam, the Friedman test did not indicate a model effect ($\chi^2 = 0.75$, $p = 0.6873$). For VISam←VISp, the omnibus model effect was significant ($\chi^2 = 13.0$, $p = 0.0015$). The subsequent paired tests showed that the RAVEL–DLAG difference remained significant after Holm correction, whereas the RAVEL–MRM-GP difference did not. Thus, the cross-session analysis supports comparable performance among the methods in the VISp←VISam direction and a particularly consistent reduction in prediction error for RAVEL relative to DLAG in the VISam←VISp direction.

A.6.2 Within-Session Orientation Robustness and ADM Comparison

As a complementary analysis, we present a benchmark on session 751348571 (the session used as a benchmark in MRM-GP), which contained 25 VISp and 22 VISam channels. Here, all eight drifting-grating orientations (0°, 45°, 90°, 135°, 180°, 225°, 270°, and 315°) were evaluated separately, with 10 TRAIN and 5 sealed TEST trials per orientation. This analysis tests whether the complete-target result is consistent across stimulus conditions within a fixed recording session and additionally includes ADM (Li et al., 2025) (Table A.10).

Secondary metrics for the same analysis are shown in Table A.11.

Table A.10: **Within-session complete-target prediction across stimulus orientations.** Spectral NMSE (mean ± SD across eight orientations in session 751348571; lower is better).

Model	VISp ← VISam	VISam ← VISp
RAVEL	$\mathbf{0.651 \pm 0.041}$	$\mathbf{0.381 \pm 0.028}$
MRM-GP	0.657 ± 0.037	0.415 ± 0.034
DLAG	0.654 ± 0.037	0.590 ± 0.073
ADM	0.975 ± 0.048	0.869 ± 0.053

Table A.11: **Secondary metrics for the within-session orientation benchmark.** Values are mean ± SD across eight stimulus orientations. Higher is better for both metrics.

	VISp ← VISam		VISam ← VISp	
Model	R^2_{zero}	Corr.	R^2_{zero}	Corr.
RAVEL	0.349 ± 0.041	0.592 ± 0.034	0.619 ± 0.028	0.793 ± 0.015
MRM-GP	0.343 ± 0.037	0.588 ± 0.030	0.585 ± 0.034	0.767 ± 0.022
DLAG	0.346 ± 0.037	0.596 ± 0.029	0.410 ± 0.073	0.736 ± 0.032
ADM	0.025 ± 0.048	0.237 ± 0.113	0.131 ± 0.053	0.448 ± 0.098

For this within-session analysis, orientation was treated as the paired condition. We tested

$$d_o = \text{NMSE}_{\text{baseline},o} - \text{NMSE}_{\text{RAVEL},o}$$

using an exact two-sided sign-flip test over all $2^8 = 256$ sign assignments. RAVEL versus ADM yielded mean NMSE reductions of 0.324 and 0.488 in the two prediction directions, with lower RAVEL NMSE in all 8/8 orientations ($p = 0.0078$ for each). Relative to MRM-GP, the corresponding reductions were 0.0063 (5/8 orientations; $p = 0.0703$) and 0.0341 (8/8; $p = 0.0078$). Relative to DLAG, the reductions were 0.0028 (6/8; $p = 0.5234$) and 0.2095 (8/8; $p = 0.0078$). Holm correction across all six directional comparisons gave adjusted $p = 0.0469$ for the four comparisons with raw $p = 0.0078$; the MRM-GP and DLAG comparisons in the VISp←VISam direction were not significant.

This complementary analysis therefore shows that the cross-region prediction pattern is not restricted to a single stimulus orientation. However, because the eight observations correspond to stimulus conditions within one recording session, these tests quantify within-session consistency rather than cross-session generalization. The eight-session analysis in Section 3.4 is therefore used as the primary benchmark in the main text.

A.7 Synthetic Computational-Efficiency Benchmark

To compare the computational cost of RAVEL with ADM, DLAG, and MRM-GP independently of predictive performance, we constructed a small synthetic two-region benchmark in which all methods received the same time-domain observations. The number of trials and observed channels was held fixed, while the number of time points per trial was varied over $T \in \{200, 300, 400\}$. Model capacities were fixed for all methods, and no hyperparameter search, model-selection sweep, or downstream prediction evaluation was included. Each configuration was fit five times, and wall-clock time is reported as the median with interquartile range.

All computational-efficiency experiments were run on CPU on the same workstation equipped with an AMD Ryzen Threadripper PRO 7955WX 16-Core processor and 255.1 GB RAM. The workstation also contained an NVIDIA RTX 5000 Ada Generation GPU with 32 GB memory, but GPU acceleration was not used for this benchmark.

RAVEL exhibited substantially lower empirical fitting time across all tested sequence lengths. At the reference setting $T = 300$, median fitting time was 0.277 s for RAVEL, compared with 4.239 s for ADM, 37.323 s for DLAG, and 47.083 s for MRM-GP, corresponding to approximately $15.3\times$, $134.8\times$, and $170.0\times$ longer fitting times, respectively. Across the tested range, RAVEL increased only from 0.270 s at $T = 200$ to 0.327 s at $T = 400$. Over the same range, ADM increased from 3.318 s to 5.757 s, DLAG from 15.594 s to 92.884 s, and MRM-GP from 36.519 s to 62.031 s.

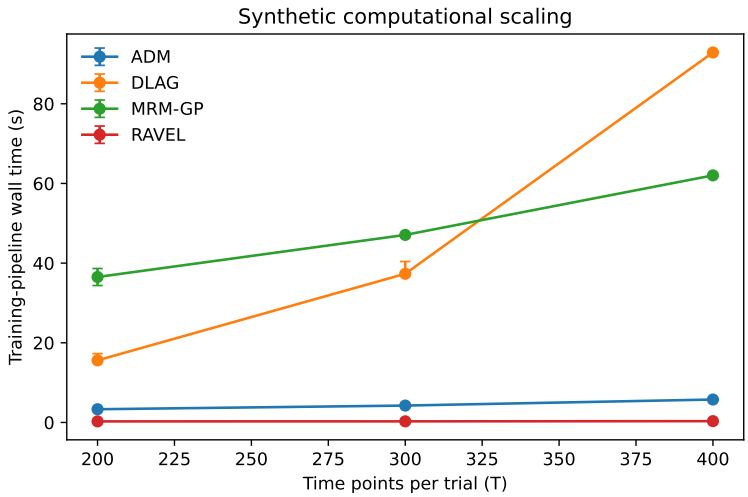


Figure A.1: **Synthetic computational scaling.** Training-pipeline wall-clock time for RAVEL, ADM, DLAG, and MRM-GP as a function of the number of time points per trial. All models were fit to identical synthetic two-region datasets using fixed model capacities and the same CPU hardware. Points show the median over five independent runs and error bars denote the interquartile range. Trial count and channel dimensionality were held fixed across sequence lengths. Timing includes each method's fitting pipeline but excludes downstream prediction and performance evaluation.

Table A.12: **Synthetic computational efficiency at $T = 300$.** Wall-clock fitting times are reported as the median and interquartile range over five independent CPU runs.

Model	Median fit time (s)	IQR (s)
RAVEL	**0.277**	0.271–0.280
ADM	4.239	4.205–4.241
DLAG	37.323	36.894–40.401
MRM-GP	47.083	46.507–47.402

These measurements reflect the empirical runtime of the specific implementations and fixed model configurations used here and should not be interpreted as estimates of formal asymptotic complexity.