INFERRING NEURAL DYNAMICS WITH VARIATIONAL SCHRÖDINGER BRIDGES

Anonymous authors
Paper under double-blind review

ABSTRACT

Neural population activity evolves within low-dimensional dynamical spaces, yet recovering these dynamics from electrophysiological recordings remains difficult. This is particularly challenging for spontaneous neural activity, where there are no repeated trials or known inputs, and where the recorded network continuously transitions between internally generated and externally driven states. Here, we introduce NeuBridge, a variational extension of Schrödinger bridges to infer stochastic latent dynamics from neural population activity. We apply NeuBridge to large-scale Neuropixels recordings from the brainstem of sleeping mice, where neural activity evolves spontaneously across sleep-wake states and is occasionally perturbed by optogenetic stimulation of upstream populations. NeuBridge reconstructs firing rates more accurately than existing latent-dynamics methods and recovers latent dynamical structure across mice. Importantly, NeuBridge also estimates the posterior control cost required to move neural activity away from its intrinsic dynamics. Without being given the timing of laser stimulation, NeuBridge detects a clear increase in control cost during optogenetic perturbations. These results show that NeuBridge can recover low-dimensional stochastic dynamics from spontaneous neural activity and can quantify the external control effort required to explain perturbation-driven deviations from intrinsic dynamics.

1 INTRODUCTION

Neural population activity is increasingly understood to evolve within low-dimensional dynamical spaces. Inferring the underlying latent dynamics from high-density electrophysiological recordings or calcium imaging data could provide a compact description of how neural populations evolve over time and transition between local network states or global brain states. A growing number of latent dynamical models therefore aim to recover these low-dimensional trajectories and the dynamical rules that govern their evolution from high-dimensional neural activity.

However, a fundamental challenge remains: the observed activity of a recorded neural population reflects not only its intrinsic dynamics, but also inputs from unobserved populations, sensory signals, and other external perturbations. Consequently, the trajectory observed in a particular recording may deviate substantially from that predicted by the intrinsic population dynamics.

This problem is particularly pronounced for spontaneous neural activity, as encountered in the brain outside the context of a task or at its extreme during sleep. Unlike trial-based experiments, spontaneous recordings contain no common starting point, do not allow for averaging across trials, or lack information about inputs, either from the environment or from another brain area.

Instead, the network continuously traverses different regions of its state space, while intrinsically generated dynamics within the recorded population and externally driven changes are intermingled within the same continuous recording. Existing latent-dynamics approaches can infer nonlinear stochastic flow fields from neural population activity, and known task variables or external inputs can be explicitly incorporated when their timing is available. However, when such inputs are unknown, separating the intrinsic dynamics of the recorded population from externally driven deviations remains difficult. This is particularly challenging in spontaneous neural recordings, such as during sleep or unconstrained behaviors.

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.

Here, we introduce *Neubridge*, a variational latent-dynamics framework extended from Schrödinger bridges that is designed to infer autonomous stochastic dynamics from continuous, spontaneous recordings of neural population activity. NeuBridge separates a shared **reference process** R, defined by the intrinsic drift $f(z)$ and diffusion $\sigma(z)$, from an observation-informed **inference process** that adds a control term v_t to account for trajectory-specific deviations. This control term is designed to capture trajectory-specific deviations, including those caused by external influences that perturb the intrinsic dynamics. A path-space KL penalty constrains the control term, so that recurring structure across the recording is captured by the intrinsic dynamics, while only trajectory-specific deviations are explained by v_t.

We applied this method to Neuropixels recordings of neural populations in the brainstem (pons and midbrain) in mice spontaneously transitioning between different wake and sleep stages (non-rapid eye movement (NREM) sleep, and rapid eye movement (REM) sleep). These recordings provide a challenging test case: population activity evolves spontaneously across wakefulness, NREM, and REM sleep, while the network is occasionally perturbed through optogenetic stimulation of upstream neuronal populations. NeuBridge reconstructed neural activity more accurately than existing latent-dynamics approaches and recovered low-dimensional dynamical structure across animals. At the same time, the inferred control cost increased during optogenetic perturbations, providing evidence that NeuBridge could distinguish externally imposed perturbations from the intrinsic population dynamics.

Together, NeuBridge provides a framework for learning stochastic latent dynamics from spontaneous neural recordings while distinguishing intrinsically generated dynamics from externally driven perturbations.

2 Related work

Neural representation learning. Various methods have been proposed to learn low-dimensional latent variables from neural recordings. One of the earlier works on this is GPFA Yu et al. (2008), which models the latent variable with a Gaussian process prior. Pi-VAE Zhou & Wei (2020) is another approach that models a task-conditioned prior, and carries out variational inference to learn the latent representation of each time bin. CEBRA Schneider et al. (2023) alternatively does not use predefined priors, but utilizes contrastive learning to build its latent landscape from temporal relationships or behavioral variables. Those methods were proposed to infer a low-dimensional latent representation of the observed neural signals, but they do learn temporal dynamical rules of how those latent states transition.

Several recent studies use mask-modeling learning objectives, to extract representations that capture useful information from neural data. NDT2 Ye et al. (2023) masks neural activity to learn neural activity representations in a self-supervised manner. MtM Zhang et al. (2024), NEDS Zhang et al. (2025), and NeuroPaint Xia et al. (2026) extend masks, to include forecast tasks, within and across modality prediction, and unobserved brain region activity inference. However, those methods also do not explicitly learn latent dynamical rules or temporal dynamical structure of the observed data.

Neural dynamics inference. A separate class of methods has been developed to infer low-dimensional dynamics from neural population recordings. LFADS Pandarinath et al. (2018) uses a sequential variational autoencoder with recurrent neural networks to infer latent dynamics from individual trials, and is typically applied to trial-structured datasets, allowing smooth trajectory estimates without averaging across repetitions. However, LFADS encodes the dynamics implicitly in the recurrent generator rather than as an explicit flow field, making the inferred dynamical rules and characteristics like fixed points and attractors less directly accessible.

The rSLDS Linderman et al. (2017); Zoltowski et al. (2020) model approximates nonlinear population dynamics through transitions between discrete states, each governed by locally linear stochastic dynamics. When external inputs are known, they can be incorporated explicitly; otherwise, observed trajectories are modeled entirely through the learned switching dynamics and process noise. Thus, nonlinear dynamics in rSLDS are represented through a set of discrete, locally linear systems, without explicitly separating intrinsic dynamics from unknown perturbations. More recently, FINDR Kim et al. (2025) was developed to directly infer nonlinear stochastic flow fields from neural population activity. FINDR use variational inference to learn the conditional prior flow-field of a neural

system, as learned dynamical rules, and infer observation-based posterior. However, FINDR does not separately model shared intrinsic dynamics from observation-specific deviations.

NeuBridge instead separates a shared stochastic reference process from an additional control term that accounts for deviations in individual trajectories. The KL penalty constrains this control term, encouraging recurring structure across the recording to be captured by the reference process while allowing individual trajectories to deviate when required by the observations.

3 METHODS

3.1 TASK SETUP AND GENERATIVE MODEL

For neural time series data $X = x_{0:T}$, we want to infer its internal latent dynamics $Z = z_{0:T}$. We assume that the latent dynamics follow a stochastic differential equation (SDE), for which we use its Euler-Maruyama discretization,

$$z_{t+1} = z_t + f_\theta(z_t)\Delta t + \sigma_\theta(z_t)\sqrt{\Delta t}\varepsilon_t, \tag{1}$$

where $\varepsilon_t \sim \mathcal{N}(0, I)$. The Euler-Maruyama discretized SDE assumes that across a time bin Δt, a neural system's next state evolves from its current state, with the system state update represented by the drift term $f_\theta(z_t)$. We additionally assume that the internal uncertainty of where the system evolves is measured by the diffusion strength term $\sigma_\theta(z_t)$, which depends on the location z_t in the latent space, changing across the latent landscape.

Equation (1) represents the generative dynamics of our model. For a spontaneously evolving neural system, the drift term f_θ and diffusion term σ_θ in the Euler-Maruyama discretized SDE represents the internal dynamics rules that this system follows. If the system receives no external inputs and is therefore completely autonomous, then if we specify a start distribution z_0 based on the past N steps of history observation $H = x_{-N:0}$ (including x_0), we can predict the future latent trajectory rollout $z_{1:T}$. In this case, for a time span from steps 0 to T, the system's internal dynamical rollout can be formulated as the conditional prior,

$$R_\theta(Z|H) = r_\theta(z_0|H) \prod_{i=0}^{T-1} r_\theta(z_{t+1}|z_t). \tag{2}$$

We can then generate observations from the latent, where we assume a noisy observation with the observation likelihood $p_\theta(x_t|z_t)$, that projects each latent state z_t to an observed signal time bin x_t. Combining equation (2), the generative factorization can therefore be represented as,

$$p_\theta(x_{1:T}, Z|H) = R_\theta(Z|H) \prod_{t=1}^{T} p_\theta(x_t|z_t), \tag{3}$$

where in a prior rollout, latent trajectories can be autonomously generated from fixed dynamical rules and a start distribution, and each latent state will produce an observation.

3.2 BRIDGE-BASED VARIATIONAL INFERENCE

However, since even during spontaneous behaviors like sleep, the recorded population of neurons is never isolated from other brain areas or separated from the environment, their activity is not fully internally driven. For our observed neural population, the activity within a given time window is driven by both the internal drift f_θ and by additional influences not captured by these intrinsic dynamics, which we represent by control term v_ϕ. Such external influences may arise from other neural populations, sensory inputs, or optogenetic stimulation, etc. Therefore, an inferred latent posterior $q_{\theta,\phi}(Z|X)$ from each observation segment is under the influence of both the internal dynamics f_θ and the trajectory-specific control term v_ϕ.

We thus extend the Schrödinger bridge problem to formulate our question: Given pairs of observation-inferred start and end latent distributions, can we learn to infer the internal dynamics f_θ and rajectory-specific control v_ϕ required to explain deviations from these dynamics?

In the classical Schrödinger bridge problem, the goal is to identify an inference process Q that connects a prescribed pair of initial and terminal distributions, Q_0 and Q_T, while deviating as little as possible from a reference process R,

$$Q^* = \underset{Q:Q_0=\nu_0,Q_T=\nu_T}{\operatorname{argmin}} (\mathrm{KL}(Q||R)). \tag{4}$$

With observed signal $X = x_{0:T}$, we want to infer the latent posterior $q_{\theta,\phi}(Z|X)$. Here, we assume that the reference process R corresponds to the conditional prior $R_\theta(Z|H)$ as the dynamical rules of the latent segment and that the inference process Q is the observation-based posterior $q_{\theta,\phi}(Z|X)$. Then optimizing the Schrödinger bridge-based objective is particularly suitable for learning the dynamical landscape of a neural system: For a sufficiently long spontaneous neural recording, if we sample many recording segments with different start and end latent distributions $\{z_0, z_T\}$ throughout the recording, it will allow the inferred path distributions $\{R_\theta(\cdot|H^{(s)})\}_{s=1}^S$, where s indexes segments, to capture the dynamical rules governing the system over the recorded period.

Specifically for the inference process Q, we first encode the observed signal $X = x_{0:T}$ and observed preceding N-step (including the current time bin) history $H = x_{-N:0}$ into the segment context C_X and the history context C_H, respectively. We then first sample the inference z_0, z_T jointly from C_X. Then we can infer the interim states influenced by both f_θ and v_ϕ, with the inference factorization as,

$$Q_{\theta,\phi}(Z|X) = q_\theta(z_0, z_T|C_X) \prod_{t=0}^{T-2} q_{\theta,\phi}(z_{t+1}|z_t, z_0, z_T, C_X). \tag{5}$$

$q_{\theta,\phi}(z_{t+1}|z_t, z_0, z_T, C_X)$ in equation (5) describes our observation corrected dynamical transition process, computed by the following Euler–Maruyama discretization updated from equation (1),

$$z_{t+1} = z_t + (f_\theta(z_t) + v_\phi(z_t, z_0, z_T, C_X))\Delta t + \sigma_\theta(z_t)\sqrt{\Delta t}\varepsilon_t, \tag{6}$$

where $\varepsilon_t \sim \mathcal{N}(0, I)$.

For the reference process R, based on equation (2), we thus can obtain the autonomous reference R rollout as,

$$R_\theta(Z|H) = r_\theta(z_0|C_H) \prod_{i=0}^{T-1} r_\theta(z_{t+1}|z_t). \tag{7}$$

In this way, we extend the Schrödinger bridge problem with variational inference. We model the variational posterior as the inference process Q, and the prior as the reference process R, to simultaneously learn intrinsic dynamics and infer observation-corrected latent trajectories.

3.3 Learning Objectives

Having defined the variational prior and posterior, we can then train our model by minimizing the negative conditional variational evidence lower bound (ELBO),

$$\mathcal{L}_{\text{ELBO}} = -\mathbb{E}_Q \left[\sum_{t=1}^{T} \log p_\theta(x_t|z_t) \right] + D_{\text{KL}}(Q_{\theta,\phi}(Z|X)||R_\theta(Z|H)), \tag{8}$$

where the first term is optimized to learn the decoder p_θ that generates the observed X from the inferred posterior $Q_{\theta,\phi}(Z|X)$, while the second KL divergence term constrains the inferred posterior $Q_{\theta,\phi}(Z|X)$ to be close to the reference conditional prior $R_\theta(Z|H)$.

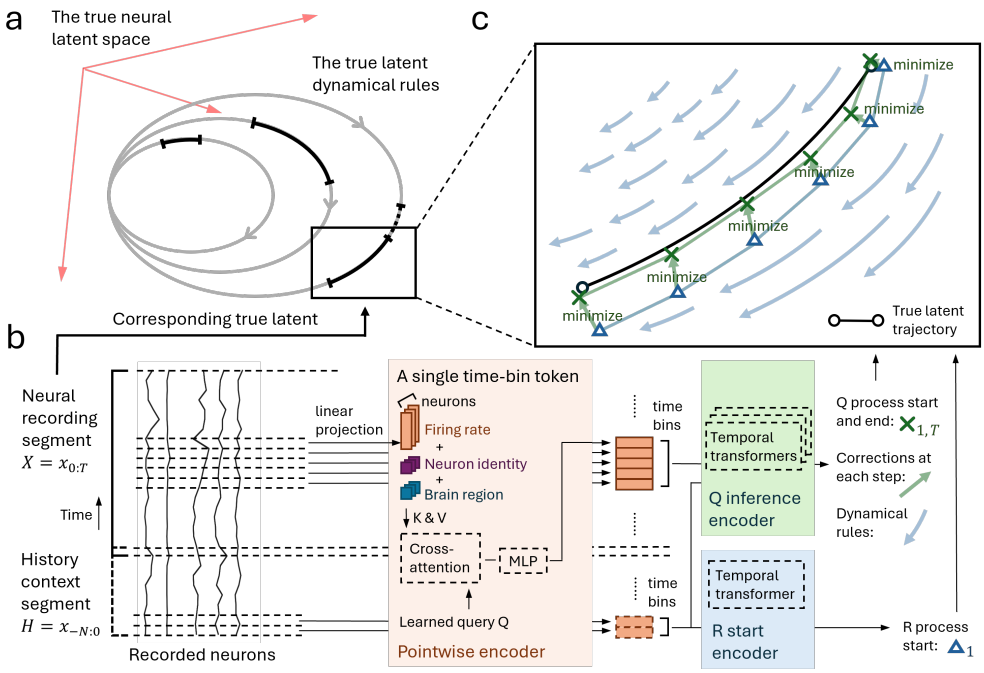


Figure 1: Schematic illustrating how NeuBridge simultaneously learns dynamical rules and infers corrected latent trajectories from recorded signals. **a**, We assume that a neural population follows low-dimensional latent dynamics, such that each recorded neural activity segment corresponds to a segment of a latent trajectory. NeuBridge takes an observed recorded segment of neural activity as input and jointly infers the latent trajectory and the dynamical rules of the latent space. **b**, NeuBridge first uses a cross-attention pointwise encoder to encode every time bin of the signal, whereby the obtained pointwise embeddings only have information of their own time bin. Each neuron's activity at each time bin is first linearly projected and combined with neuron-identity and brain-region embeddings. A pointwise cross-attention encoder then maps each time bin to a fixed-dimensional embedding e_t, allowing recordings with different numbers of neurons to be processed within a common framework. The single time-bin embeddings $e_{0:T}$ are then passed to the inference encoder Q to infer the start and end distributions z_0, z_T of the latent Q, the trajectory-specific correction v_t required at each step, and update the dynamical rules f_θ. In parallel, the R start encoder produces the R process start z_0 using only preceding history without future information. NeuBridge reconstructs the input neural signals from the inferred latent trajectory, while the model is penalized for deviations from the learned intrinsic dynamics. **c**, In this way, NeurBridge simultaneously infers dynamical rules and the external corrections needed to produce the observed activity patterns.

Meanwhile, the second KL term corresponds to the Schrödinger bridge objective in equation (4), which is learning to infer a Q process that deviates the least from the R process. Specifically, from the Euler-Maruyama discretized SDE in equation (6), the KL objective is the sum of the contribution from all interim transition corrections v_t plus the start and end boundary contributions, calculated as

$$D_{\mathrm{KL}}(Q_{\theta,\phi}||R_\phi) = \mathbb{E}_Q\left[\sum_{t=0}^{T-2}\frac{\Delta t}{2}v_t^\top D_\theta(z_t)^{-1}v_t\right] + \mathbb{E}_Q\left[\log\frac{q_\theta(z_0,z_T|C_X)}{r_\theta(z_0|C_H)r_\theta(z_T|z_{T-1})}\right], \quad D_\theta = \sigma_\theta\sigma_\theta^\top. \tag{9}$$

We additionally add a prior-start reconstruction term, which directly encourages the history-conditioned prior reference $R_\theta(z_0|H)$ to predict an initial latent state that is consistent with the observed neural activity at the beginning of a rollout segment. The overall objective of our model $\mathcal{L}_{\mathrm{NB}}$ is thus,

$$\mathcal{L}_{\text{start}} = -\mathbb{E}_{z_0 \sim R_\theta(Z|H)} \log p_\theta(x_0|z_0), \qquad \mathcal{L}_{\text{NB}} = \mathcal{L}_{\text{ELBO}} + \mathcal{L}_{\text{start}}. \tag{10}$$

3.4 Model architecture and parametrization

To implement the model across recordings with different numbers of recorded neurons, we first use a pointwise cross-attention encoder e_ϕ to map the input sequence $X = x_{0:T}$ and $H = x_{-N:0}$ into fixed-dimensional pointwise representations, $C_X = e_{0:T}$ and $C_H = e_{-N:0}$.

We then infer z_0 and z_T for the Q process, and z_0 for the R process. The posterior Q and prior R latent trajectories are subsequently inferred according to Eqs. (6) and (1), respectively, where f_θ and $v_{\theta,\phi}$ are parameterized by two temporal transformer encoders and σ_θ by an MLP.

For the decoder, we use a recording-specific linear projection for each recording to reconstruct the binned firing rates from the latent space. For a recording r,

$$\hat{x}_t^r = W^r z_t + b^r + \sigma_{\text{obs}} \varepsilon_t^r, \tag{11}$$

where W^r and b^r are the recording-specific decoder parameters, σ_{obs} is a learned global observation-noise parameter, and $\varepsilon_t^r \sim \mathcal{N}(0, I)$.

3.5 Model training

For model training schedules, we added a rollout horizon curriculum at the start of the training, and posterior guided rollout refinement at later stage of the training.

Since our model is trained with sampled segments of long autonomous neural recordings, immediately starting very very long segments (like 20 steps) may make model convergence difficult. NeuBridge also supports mixed length of segments in the training set. We therefore add the rollout horizon curriculum, where we sample each training segment's length from a rounded, clipped Gaussian, whose center moves smoothly from a short start length towards the target length. Typically, NeuBridge starts training with a training set of mostly 3-step long segments mixed with longer segments. The average segment length of the training set will gradually increase, and will eventually mostly sample, for example, 20-steps segments, with a few shorter-step samples.

The posterior guided rollout refinement is an additional refinement to the R rollout trajectories at the later part of the training process. After each epoch of the standard joint training of Q inference and R dynamics, we freeze a snapshot of Q as a teacher. We generate its inferred trajectory, and optimize R's autonomous rollout to match it with $\mathcal{L}_z = \frac{1}{T}\sum_{t=1}^{T} ||z_t^R - \text{stopgrad}(z_t^Q)||^2$. The rollout refinement objective can help the R process learn more accurate dynamics.

4 Experimental Results

4.1 Simulated Data: Lorenz dynamics

We first applied NeuBridge to simulated data with three-dimensional Lorenz latent dynamics, adapted from Lorenz (1963). The system follows non-linear latent dynamics and transitions non-periodically between two lobes of the latent space. We added small Gaussian progress noise ($\sigma_{\text{diff}} = 0.01$) to the latent dynamics, and used different linear projections from the latent with Gaussian ($\sigma_{\text{obs}} = 0.1$) noise to simulate 3 recordings with different numbers of simulated neurons. Each recording consisted of 8 training trajectories, 2 validation trajectories; each trajectory comprised 500 observed steps. The time bin of each step was 0.05 seconds.

We sample overlapping, 12-step long trials from the trajectories to obtain 11712 training samples and 2928 validation samples. For autonomous rollout of the reference process R, NeuBridge achieved a variance explained (R^2) of 0.805. For the inference process Q, which additionally includes the inferred correction term, NeuBridge achieved an R^2 of 0.895. Thus, as shown in Figure 2e, NeuBridge could effectively infer the latent dynamics from the simulated data and its autonomous rollout predictions remained accurate across long rollouts. Moreover, the observation-based correction increased the accuracy of the latent inference.

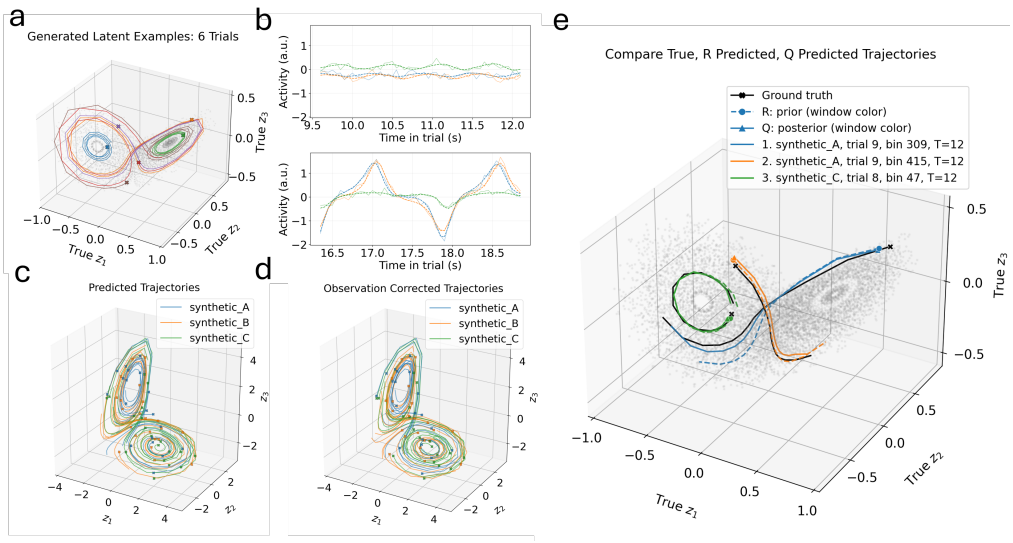


Figure 2: NeuBridge recovers Lorenz latent dynamics from noisy simulated neural data. **a**, Simulated latent Lorenz dynamics. We sampled 12-step trajectory segment to construct the dataset. The underlying dynamics are non-linear and irregular, with added small progression noise. **b**, Example simulation of neural population activity. **c**, Example rollouts of the reference process R. These trajectories are autonomous predictions generated from the learned dynamical rules. **d**, Example trajectories inferred by the Q process. These predictions include an observation-based correction to the dynamics of the reference process R. **e**, Example comparison between the true latent trajectory, the R dynamics predicted trajectory, and the Q observation inferred trajectory NeuBridge. Our model's latent for each recording is projected through an affine projection to its original simulated latent coordinate. The results show that NeuBridge can infer the correct autonomous dynamical rules across long rollouts, while the observation-based correction further improves trajectory inference.

4.2 Neuropixels Data: Disentangling the brainstem dynamics underlying sleep

We then applied NeuBridge to large-scale Neuropixels recordings from the brainstem (midbrain and pons) of sleeping mice from Lozano et al. (2026), where neural population activity evolves spontaneously across sleep-wake states. We analyzed spontaneous recordings from wild-type (WT) and eYFP control mice (where laser stimulation had no effect on neural activity), as well as recordings from the Opto group mice, in which REM sleep-promoting GABAergic neurons in the dorsomedial medulla (dmM) were optogenetically stimulated. Activation of dmM GABAergic neurons reliably promotes transitions from NREM to REM sleep, and these neurons project to pontine and midbrain regions sampled by our Neuropixels recordings. Their activation therefore provides a controlled external perturbation of the recorded population dynamics. In the optogenetic experiments (Opto group), laser stimulation was delivered at random intervals of approximately 15 min for 1 min and induced REM sleep in approximately 50% of stimulation trials. This dataset was therefore ideally suited to test whether NeuBridge can disentangle the intrinsic dynamics governing spontaneous neural population activity from externally driven perturbations, and quantify how such perturbations reshape latent dynamics and influence state transitions.

We first compared the reconstruction and predicted variance explained between NeuBridge and other methods with a dynamics-based latent. As shown in Figure 3c, NeuBridge outperformed other methods on both tasks. For the R process comparison, models only received 8 bins of history signal, and predicted the future $6, 12$, or 20 bins of firing rates using their learned latent dynamics. The R/prior reconstruction results evaluate how well the learned latent dynamical rules can explain the observed neural signals. NeuBridge performed slightly better than FINDR across all future rollout lengths, and substantially outperformed rSLDS and LFADS. For the Q process comparison, models can read the full segment of the observed signal as well as the 8 step history signal, and reconstructed the segment based on the low-dimensional latent Q/posterior trajectory. The Q/posterior reconstruc-

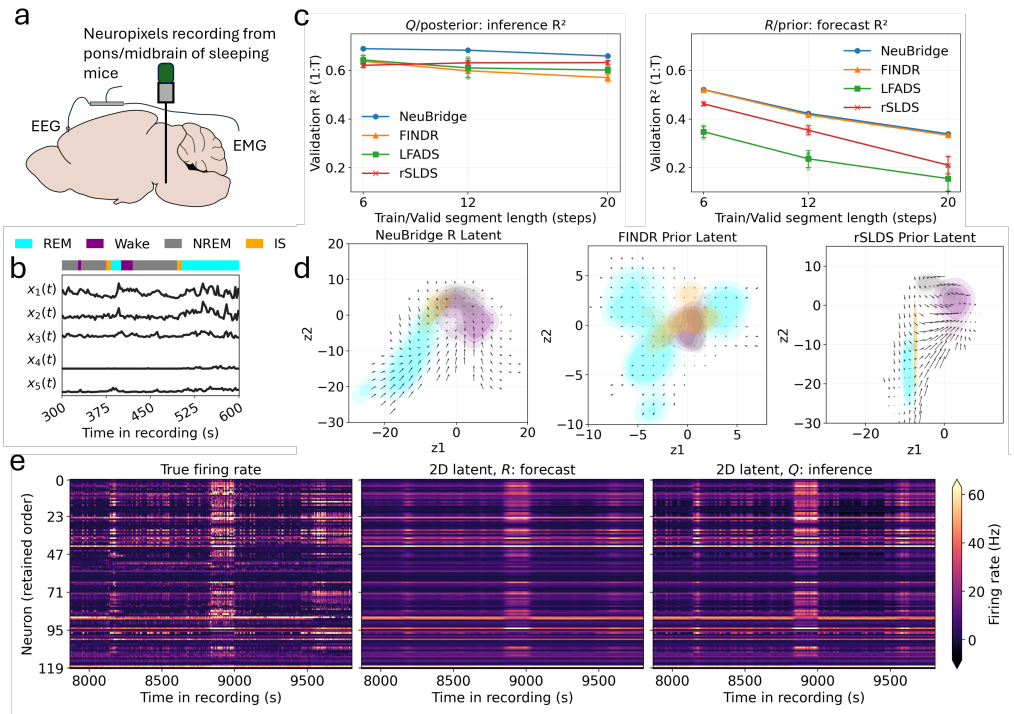


Figure 3: NeuBridge recovers latent dynamics from spontaneous Neuropixels recordings in sleeping mice. **a**, Lozano et al. (2026) used Neuropixels to record from midbrain/pontine regions in sleeping mice. Simultaneously recorded EEG/EMG signals were used to annotate sleep-wake stages: REM, Wake, NREM, and the intermediate stage (IS) between NREM and REM. **b**, Example Neuropixels recording. We convert the spiking activity into binned firing rates using 2.5-s bins. We then separated training and validation sets, followed by linear detrending, causal smoothing, and z-scoring. **c**, Comparison of NeuBridge with FINDR Kim et al. (2025), LFADS Pandarinath et al. (2018), and rSLDS Linderman et al. (2017); Zoltowski et al. (2020). These methods infer observation-based latent trajectories and dynamical rules of the latent space, allowing performece to be compared using variance explained by observation-inferred Q/posterior trajectories and by autonomous R/prior trajectories (that only depend on the accuracy of learned dynamical rules). Increasing rollout steps correspond to prediction further into the future. For Q/posterior trajectories, NeuBridge preserves more of the observed neural signal than the other models. For autonomous R/prior rollouts, NeuBridge performs slightly better than FINDR and substantially better than LFADS and rSLDS, indicating improved recovery of the latent dynamics. **d**, Latent flowfields learned by different methods. NeuBridge recovers the dynamical structure associated with transitions between different sleep-wake states and separates Wake, IS, and REM within the latent space. rSLDS learned similar dynamics, although more stretched, while FINDR shows less clearly structured latent dynamics. **e**, Reconstruction of firing rates. Example firing-rate heatmaps are shown for recorded neurons, and firing rates reconstructed from the autonomous R prediction and observation-based Q inference, using only a 2-dimentional latent trajectory. The R-based reconstructed firing rate capture the overal activity pattern, while the observation-corrected Q trajectory more closely firing rate fluctuations.

tion results thus evaluate how well the inferred latent trajectory explains the observed neural signal. NeuBridge performed better than all other models across all conditions.

NeuBridge also recovered the dynamical structure of sleep transitions, as shown in Figure 3d. Here we fitted the three models, NeuBridge, FINDR, rSLDS, that can infer flow-field dynamics, directly with a two-dimensional latent space, and visualized their flow-field dynamics. The 4 sleep stages were manually annotated using the EEG/EMG signals, which the model did not have access to. We used the R/prior start latent location and their corresponding sleep stage to fit the kernel density estimation (KDE) plot of each sleep stage. NeuBridge visualized the dynamical structure associated

with transitions between different sleep-wake states and separated Wake, IS, and REM within the latent space. rSLDS learned similar dynamics, although more stretched, while FINDR showed less clearly structured latent dynamics.

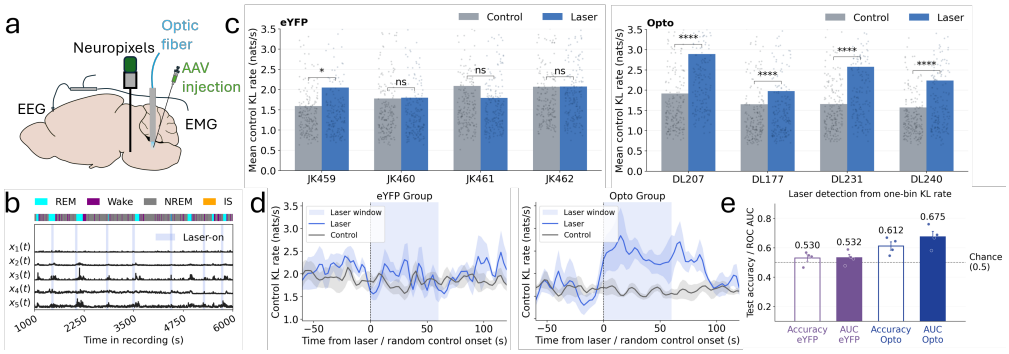


Figure 4: NeuBridge can separate intrinsic dynamics from external inputs **a**, For the Opto group mice, their REM sleep-promoting dmM GABAergic neurons were optogenetically stimulated.. **b**, Laser stimulation was delivered at random intervals of approximately 15 min for 1 min and induced REM sleep in approximately 50% of stimulation trials. **c**, Comparison of KL loss rate, which represents the control effort of the inferred correction term v, between laser and non-laser time bins. In eYFP control mice, where laser stimulation does not activate neurons, the KL rate during laser and non-laser bins have mostly no difference. In contrast, in the Opto group mice, where each laser stimulation excitates dmM GAD2 neurons and consequently promotes REM sleep, laser time bins' KL rate was significantly higher than for non-laser bins. **d**, Laser-triggered peristimulus time histogram (PSTH) of the KL rate, obtained by averaging the KL rate relative to laser onset . Blue lines represent the laser-triggered PSTH, gray lines depict randomly sampled non-laser segments (for each laser start we sampled 3 non-laser control trials). In the Opto group, NeuBridge immediately puts more effort to correct the inferred latent trajectory that deviates from prior dynamical rules. **e**, A linear classifier trained using only the scalar rate of each time bin from the Opto mice predicted whether a time-bin occurred during laser stimulation above chance level on held-out data. In contrast, KL rates in eYFP group did not allow for above-chance classification. Together, these results indicate that NeuBridge learns the intrinsic dynamical rules of a neural population, while separately inferring external perturbations that influence the latent dynamics.

NeuBridge could separate intrinsic dynamical rules from external laser stimuli in this experimental setting. NeuBridge explicitly modeled the control correction term v_t, representing the observation-based correction. Following equation (9) the control correction effort can be represented by the KL rate, which is the unit time value of the first KL objective term. The KL rate measures how much control correction is needed relative to how hard this latent location is to control. We can see from Figure 4, that during laser events, KL rate will significantly increased, while the KL rate itself had above chance predictability during the laser sttimulation intervals. Together, these results show that NeuBridge can identify periods in which the observed neural dynamics deviate from the learned intrinsic dynamics and require additional control, thereby detecting externally driven perturbations without prior knowledge of their timing.

5 Conclusions

NeuBridge provides a framework for learning low-dimensional stochastic dynamics from continuous spontaneous neural activity, where intrinsic dynamics and unobserved perturbations are intermingled. Across large-scale brainstem Neuropixels recordings, NeuBridge accurately reconstructed neural activity and recovered interpretable latent dynamics. Importantly, without being given the timing of optogenetic stimulation, the inferred control cost increased during perturbations, while no such increase was observed in control mice. This provides evidence that NeuBridge can identify when observed population trajectories cannot be fully explained by the learned intrinsic dynamics alone. These results show that NeuBridge can reveal both the intrinsic dynamical structure of neural populations and the points at which external influences reshape their trajectories.

AI USE STATEMENT

In this work, we used generative AI tools for Generate synthetic data sets, help develop theoretical models or conceptual frameworks, formulate mathematical claims, provide critical ingredients for proving mathematical claims, propose or refine hypotheses, design or provide feedback on research methodology or experiments, implement methods, clean and reformat dataset, support qualitative and thematic data analysis. We have not used generative AI tools to interpret results, and assist in the writing of proofs, assist with translation are not applicable to this work. Additionally, we used generative AI tools to summarize model specifications in the Appendix, to create or modify scientific figures or images, summarize or analyze existing literature, discover research topics or identify gaps, brainstorming, sourcing/searching for information, identify relevant literature, propose a title or keywords for a research paper. We have reviewed all AI-assisted work. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

The code for our model, compared baseline models, and code to reproduce all main data figures are included in the supplemental materials.

REFERENCES

Timothy Doyeon Kim, Thomas Zhihao Luo, Tankut Can, Kamesh Krishnamurthy, Jonathan W. Pillow, and Carlos D Brody. Flow-field inference from neural data using deep recurrent networks. In *Forty-second International Conference on Machine Learning*, 2025. URL https://openreview.net/forum?id=ZLC4B9oQWX.

Scott Linderman, Matthew Johnson, Andrew Miller, Ryan Adams, David Blei, and Liam Paninski. Bayesian learning and inference in recurrent switching linear dynamical systems. In *Artificial intelligence and statistics*, pp. 914–922. PMLR, 2017.

Edward N. Lorenz. Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, 20(2): 130–141, 1963. doi: 10.1175/1520-0469(1963)020⟨0130:DNF⟩2.0.CO;2.

David E Lozano, Jiso Hong, Xi Jin, Joseph A Stucynski, Christian K Machens, Shinjae Chung, and Franz Weber. Low-dimensional population dynamics in the brainstem gate rem sleep. *Nature neuroscience*, pp. 1–13, 2026.

Chethan Pandarinath, Daniel J O'Shea, Jasmine Collins, Rafal Jozefowicz, Sergey D Stavisky, Jonathan C Kao, Eric M Trautmann, Matthew T Kaufman, Stephen I Ryu, Leigh R Hochberg, et al. Inferring single-trial neural population dynamics using sequential auto-encoders. *Nature methods*, 15(10):805–815, 2018.

Steffen Schneider, Jin Hwa Lee, and Mackenzie Weygandt Mathis. Learnable latent embeddings for joint behavioural and neural analysis. *Nature*, 617(7960):360–368, 2023.

Ji Xia, Yizi Zhang, Shuqi Wang, Genevera Allen, Liam Paninski, Cole Hurwitz, and Kenneth Miller. Inpainting the neural picture: Inferring unrecorded brain area dynamics from multi-animal datasets. *Advances in Neural Information Processing Systems*, 38:79040–79072, 2026.

Joel Ye, Jennifer Collinger, Leila Wehbe, and Robert Gaunt. Neural data transformer 2: multi-context pretraining for neural spiking activity. *Advances in Neural Information Processing Systems*, 36:80352–80374, 2023.

Byron M Yu, John P Cunningham, Gopal Santhanam, Stephen Ryu, Krishna V Shenoy, and Maneesh Sahani. Gaussian-process factor analysis for low-dimensional single-trial analysis of neural population activity. *Advances in neural information processing systems*, 21, 2008.

Yizi Zhang, Yanchen Wang, Donato M Jiménez-Benetó, Zixuan Wang, Mehdi Azabou, Blake Richards, Renee Tung, Olivier Winter, Eva Dyer, Liam Paninski, et al. Towards a" universal translator" for neural dynamics at single-cell, single-spike resolution. *Advances in Neural Information Processing Systems*, 37:80495–80521, 2024.

Yizi Zhang, Yanchen Wang, Mehdi Azabou, Alexandre Andre, Zixuan Wang, Hanrui Lyu, International Brain Laboratory, Eva L Dyer, Liam Paninski, and Cole Lincoln Hurwitz. Neural encoding and decoding at scale. In *Forty-second International Conference on Machine Learning*, 2025. URL https://openreview.net/forum?id=vOdz3zhSCj.

Ding Zhou and Xue-Xin Wei. Learning identifiable and interpretable latent models of high-dimensional neural activity using pi-vae. *Advances in neural information processing systems*, 33:7234–7247, 2020.

David Zoltowski, Jonathan Pillow, and Scott Linderman. A general recurrent state space framework for modeling neural dynamics during decision-making. In *International Conference on Machine Learning*, pp. 11680–11691. PMLR, 2020.

A Appendix

A.1 Implementation specifications

NeuBridge and the neural baselines were implemented in PyTorch and trained using CUDA; rSLDS was fitted on CPU. Code was verified using Python 3.12.9, PyTorch 2.5.1, NumPy 2.2.0, and SciPy 1.16.3. For the main model comparisons, we performed three runs per configuration using random seeds $\{0, 1, 2\}$ and report the mean $\pm$ sample standard deviation across runs. We seeded the Python, NumPy, and PyTorch random-number generators, including CUDA generators when available. Individual run configurations and selected checkpoints are provided with the code. All CUDA experiments were run on a single Nvidia RTX 4090 GPU.

We compared models with eight-dimensional latent representations at prediction horizons $T \in \{6, 12, 20\}$, corresponding to 15, 30, and 50 seconds. NeuBridge, FINDR, and LFADS were trained using AdamW with learning rate 2×10^{-3}, batch size 256, no weight decay, and a maximum of 200 epochs. A 50-epoch segment-length curriculum gradually shifted sampling from short segments (starting at three steps) toward the target horizon, while retaining shorter segments thereafter. Gradient norms were clipped at 1 for NeuBridge and 5 for FINDR and LFADS.

NeuBridge used a transformer width of 32, four attention heads, and KL weight $\beta = 1$. From epoch 101 onward, each standard training epoch was followed by one reference-drift refinement pass with learning rate 5×10^{-4}, using latent rollout matching ($\mathcal{L}_z$) against a frozen posterior teacher. FINDR used encoder width 32, inference and drift widths of 50, and $\beta = 2$. LFADS used eight factors, a 64-unit generator, 32-unit encoders and controller, a 16-dimensional initial condition, two inferred inputs, dropout 0.05, and $\beta = 1$.

rSLDS used eight continuous dimensions and eight discrete regimes, with separate models fitted to each recording's contiguous training sequences using Laplace-EM. We allowed up to 200 iterations, with five initialization restarts of 25 iterations each. Early stopping monitored both validation inference and forecasting R^2, requiring both metrics to remain unimproved for 20 epochs once stopping became eligible. The reported comparisons use the checkpoint with the best validation forecasting R^2 for both inference and forecasting results.

A.2 Supplementary figures

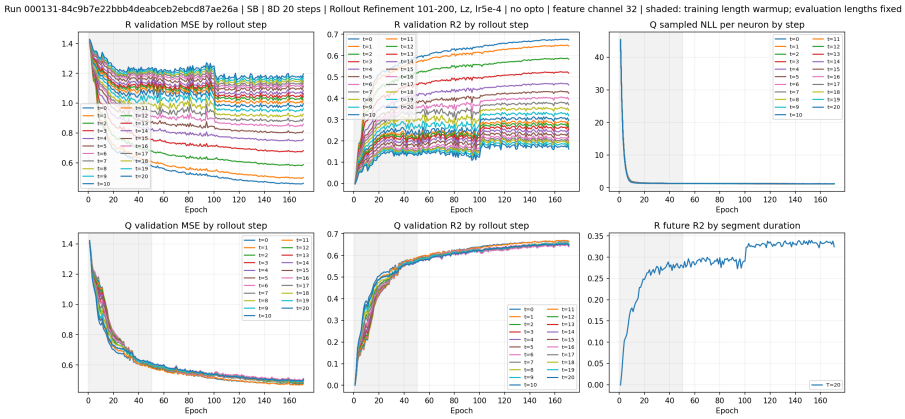


Figure 5: Example training-objective curves for a NeuBridge model. The first 50 epochs correspond to the rollout horizon curriculum, and the last 100 epochs to the rollout refinement. The R^2 scores remain similar for different steps of Q inference, since each Q step can access the full observed segment. In contrast, for the R process, we can see a clear loss of variance explained with each step of forecast into the future.