HIERARCHICAL CYCLIC NEURAL ODES

Anonymous authors
Paper under double-blind review

ABSTRACT

Biological systems are inherently cyclical, operating through feedback loops where epigenetic modifications regulate transcription, driving protein synthesis and metabolic flux, which ultimately modulate the epigenome. Modeling these continuous, high-dimensional dynamics remains challenging because the underlying molecular layers are operating on disparate timescales. While standard Neural Ordinary Differential Equations (NODEs) are widely recognized for modeling continuous-time processes, they frequently experience timescale collapse and numerical stiffness when forced to map extreme multi-scale interactions. To stably learn these hierarchical oscillatory systems, the Hierarchical Cyclic Neural ODE is introduced. A Stuart-Landau limit-cycle prior is embedded into the model's latent vector field, constraining the neural network to focus primarily on the biological variations that deviate from an underlying rhythm. A cyclic small-gain analysis provides a theoretical guarantee that this architecture yields unconditionally bounded trajectories (preventing numerical instability) and exponential orbital attraction (ensuring perturbed states converge to a homeostatic orbit). Evaluated against leading continuous-time models including scNODE and Heavy Ball NODE, on a four-layer mouse liver atlas and a two-layer *Neurospora crassa* course in constant darkness, this architecture maintains stable trajectories through the forecast window and reduces held-out error on both. In contrast, acyclic ablations suffer from temporal mode collapse and amplify perturbations over a full cycle, confirming that closed-loop structural feedback is necessary to bound system energy. Leveraging these continuous dynamics, an analysis of the system's phase response demonstrates that a perturbation confined to a single layer produces a lasting phase shift in layers reachable only around the loop, a response that vanishes identically when the single return edge is removed.

1 INTRODUCTION

The regulatory architecture of a cell is a loop rather than a chain. The Central Dogma describes a unidirectional flow from DNA to RNA to protein (Crick, 1970), extended downstream by chromatin accessibility, transcripts, proteins and metabolic flux. To predict how cells evolve and respond to perturbations, standard integrative models typically encode this architecture as a unidirectional sequence (Argelaguet et al., 2020; Gayoso et al., 2021; Shaffiey & Pierobon, 2026). However, the final layer is not terminal. The enzymes that write and erase chromatin marks consume metabolites as obligate cofactors, including acetyl-CoA for histone acetyltransferases, S-adenosyl-methionine for methyltransferases, NAD^+ for sirtuins, and α-ketoglutarate for the TET and JmjC demethylases (Lu & Thompson, 2012). Metabolic state therefore sets the rate at which the epigenome can be modified, and the hierarchy closes on itself. This is a statement about causal structure rather than about timing. It holds in any cell, whether or not that cell's trajectory happens to be periodic.

Closing that edge changes what dynamics the computational model can express. A linear model without feedback can only track its input, whereas adding a feedback loop enables it to act as the generator of autonomous oscillation, from the Goodwin circuit to cellular clocks (Goldbeter, 1996; Alon, 2019). The natural attractor of a closed regulatory loop is therefore an orbit rather than a point, and that orbit is produced by the topology rather than imposed on it. Acyclic models remove both by severing the return edge which removes the feedback that generates the oscillation, and the linear damping used to stabilize the resulting unidirectional cascade (Shaffiey & Pierobon, 2026) drives the system at rest.

Omics measurements often arrive on irregular, modality-specific grids, and the quantity of interest is a continuous trajectory rather than a sequence of snapshots. Standard discrete-time sequence models

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.

(such as Transformers or linear forecasters) require artificially resampling or interpolating these grids, which introduces circularity and distorts the ground truth (Rubanova et al., 2019). By utilizing Neural Ordinary Differential Equations (NODEs) (Chen et al., 2018), we model the vector field directly. This not only allows us to evaluate the biological state at any arbitrary, unmeasured time point, but it also preserves the continuous causal mechanics required to perform *in-silico* structural interventions. The four molecular layers (epigenome, transcriptome, proteome, metabolome) do not share a timescale. Central metabolite pools turn over on the order of seconds (Milo & Phillips, 2015). Transcripts and proteins have median half-lives near 9 and 46 hours in mouse fibroblasts (Schwanhäusser et al., 2011). Chromatin modifications persist across hours to days (Zee et al., 2010). A single vector field spanning all four is stiff by construction (Curtiss & Hirschfelder, 1952; Hairer & Wanner, 1996), and the ratio of fastest to slowest rate in a learned multi-omics field has been measured above 10^6 (Shaffiey & Pierobon, 2026), with the resulting computational cost and established techniques addressing this detailed in Section 2.

The grey-box formulation of Rackauckas et al. (2020) fixes the part of the dynamics that domain knowledge determines, and learns the remainder. The Tri-Scale cascade of Shaffiey & Pierobon (2026) applies this kind of formulation to multi-omics as an acyclic coupling with per-modality timescales and first-order degradation kinetics. The damping is a point-attractor prior, so the only steady state available to the model is rest, and the guarantee it supports is per-layer, provable by induction because a cascade has a first layer driven by time alone. Closing the loop removes both the point-attractor and the per-layer induction, meaning the true biological architecture requires a stability guarantee that the prior theorem cannot recover.

The work presented in this paper retains the grey-box formulation and changes the physics inside it. In place of first-order degradation we embed a Stuart–Landau limit cycle in each layer and let the neural residual model only the deviation from that orbit, so that the system is stabilized toward sustained activity rather than toward rest. Any smooth system passing through a supercritical Hopf bifurcation reduces near onset to precisely this normal form (Kuramoto, 1984; Strogatz, 1994), so the functional form is determined by the bifurcation rather than selected. Rotational symmetry cancels the even-order terms, and a negative cubic is the lowest-order term available that can bound the amplitude. A closed loop generally sits near such a bifurcation, so once the system is modeled as a closed loop, Stuart-Landau is the standard equation that applies.

The stability analysis must change with the topology. Because a cycle lacks a first layer, the inductive approach of Shaffiey & Pierobon (2026) doesn't apply to cyclic topologies, requiring instead a small-gain Lyapunov analysis for interconnected systems (Dashkovskiy et al., 2007). The resulting condition constrains the *product* of the gains around the cycle, $\prod_k L_k < \prod_k A_k$, rather than each gain separately. This is a weaker condition than the diagonal dominance a cascade requires. An individual layer may exceed its own contraction rate provided the remaining layers compensate, and the acyclic condition is recovered as the special case in which every factor must hold alone. An individual layer may therefore stay volatile while the loop as a whole remains inside the bound, a behavior observed in every empirical run.

The model trains on the first cycle of a four-omics layer mouse liver time course and forecasts the second, a horizon twice the training window. It beats the strongest naive reference on 12 of 15 held-out observations. Paired ablations isolate the distinct contributions of the limit-cycle prior (preventing explosive divergence) and the closed-loop topology (preventing temporal mode collapse), while the cross-layer phase response isolates the return edge.

Contributions. (1) A closed-loop multi-omics Neural ODE that replaces dissipative kinetics with a Stuart–Landau limit-cycle prior, stabilizing the model toward a homeostatic orbit. (2) A cyclic small-gain theorem providing unconditional boundedness and regional exponential orbital attraction, enforced at initialization via spectral-norm estimation. (3) A cross-layer phase response analysis isolating the causal signal of the return edge, enabling structural interventions standard forecasters cannot perform.

2 Related Work

Continuous-time models and stiffness. Standard continuous models like NODEs (Chen et al., 2018) and Latent ODEs (Rubanova et al., 2019) degrade sharply under stiffness. Adaptive solvers are driven to intractable step sizes, and optimizers often escape this cost by learning over-smooth fields that fit slow variables while discarding fast ones. Existing mitigations operate on the vector field rather than

its structure. Jacobian regularization (Finlay et al., 2020) and Heavy Ball momentum with mechanical friction (Xia et al., 2021) shrink the magnitude of the dynamics uniformly, which is effective against stiffness but indiscriminate about what functional behaviors it removes.

Structural priors and grey-box dynamics. The universal differential equation or grey-box formulation fixes what domain knowledge determines and learns the remainder (Rackauckas et al., 2020), with physics-informed networks as a special case (Raissi et al., 2019). The Tri-Scale cascade of Shaffiey & Pierobon (2026) follows this pattern for multi-omics, fixing the molecular hierarchy as an acyclic coupling with per-modality timescales and first-order degradation kinetics. Its prior is dissipative, and the stability condition it admits, $\gamma_k > \tau_k L_k$, is per-layer and provable by induction because a cascade has a first layer; however, neither property survives closing the loop.

Oscillatory architectures and stability certificates. Oscillation appears inside neural sequence models, notably the coupled oscillatory RNN (Rusch & Mishra, 2021a) and its multiscale successor (Rusch & Mishra, 2021b), where damped forced oscillators keep gradients bounded over long horizons. There oscillation is a device for trainability in discrete time. Here it is a claim about the system being modeled, carried in continuous time by parameters with physical meaning. Alternatively, formal stability guarantees have been developed independently for learned continuous-time systems. Jointly learning a Lyapunov function with the field guarantees convergence to an equilibrium (Manek & Kolter, 2019), and recurrent equilibrium networks guarantee contraction between trajectories by construction (Revay et al., 2023). A Lyapunov function, meaning a quantity whose existence is enough to prove the stability claim, targets a point, and contraction to a unique trajectory rules out a limit cycle as surely as damping does. The notion we need is orbital stability, meaning stability of the invariant set rather than of any point or trajectory within it, and establishing it for an interconnected system requires a small-gain rather than a monolithic argument (Jiang et al., 1994; Dashkovskiy et al., 2007). To our knowledge no existing result covers a closed loop of learned, differently timescaled oscillators.

Multi-omics trajectory modeling and the missing return edge. Static multi-omics integration is well developed (Argelaguet et al., 2020; Gayoso et al., 2021), often relying on pseudotime or phase-inference (Anafi et al., 2017) to recover temporal structure. While scNODE (Zhang et al., 2024) predicts transcriptomic trajectories continuously, its monolithic latent state obscures the molecular hierarchy. Crucially, existing models share a strict directionality: information flows from chromatin to metabolites and stops. The return edge, where metabolites supply cofactors gating chromatin modification (Lu & Thompson, 2012), is entirely absent.

3 METHODOLOGY

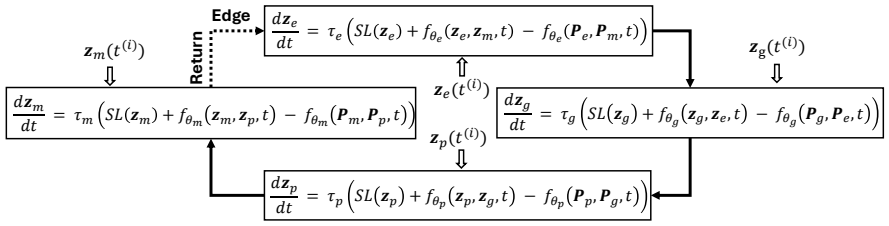


Figure 1: **Hierarchical Cyclic Neural ODE Architecture.** The continuous-time vector field forms a closed causal loop where upstream layers regulate downstream targets, and metabolic flux completes the cycle by governing epigenetic modification rates. Neural residuals are orbit-centered via projection ($\mathbf{P}$) onto the Stuart-Landau (SL) limit-cycle prior.

3.1 PROBLEM FORMULATION

Let dataset $\mathcal{D} = \{(\mathbf{x}_e(t^{(i)}), \mathbf{x}_g(t^{(i)}), \mathbf{x}_p(t^{(i)}), \mathbf{x}_m(t^{(i)}), t^{(i)})\}_{i=1}^{N}$ consist of aligned multi-omics observations, where N is the total number of samples, i indexes the individual samples, and $t^{(i)}$ represents the specific timepoint at which the i-th sample was measured. The timescale relationships within the data range from slowest to fastest: epigenomics $\mathbf{x}_e \in \mathbb{R}^{D_e}$, transcriptomics $\mathbf{x}_g \in \mathbb{R}^{D_g}$, proteomics $\mathbf{x}_p \in \mathbb{R}^{D_p}$, and metabolomics $\mathbf{x}_m \in \mathbb{R}^{D_m}$.

While formalized here as a four-layer multi-omics system (illustrated in Figure 1) to capture the complete cellular feedback loop, this hierarchical cyclic formulation generalizes to an arbitrary

K-layer closed-loop architecture (*e.g.*, the $K = 2$ transcript-protein loop of Section 4.1, or a $K = 3$ model mapping genes to proteins to metabolites).

To handle the high-dimensional observations $\mathbf{x}(t^{(i)})$ within dataset $\mathcal{D}$, the proposed model operates on a compressed latent representation, *i.e.*, the cell state $\mathbf{z}(t) = [\mathbf{z}_e(t), \mathbf{z}_g(t), \mathbf{z}_p(t), \mathbf{z}_m(t)]^\top$. An encoder $q_\phi(\mathbf{z}(0)|\mathbf{x}(0))$ initializes the latent state at $t = 0$, and a decoder $p_\psi(\mathbf{x}(t)|\mathbf{z}(t))$ maps the latent state at any arbitrary time t back to the observation space.

Each latent block $\mathbf{z}_k \in \mathbb{R}^{d_k}$ is the compressed representation of the corresponding observation $\mathbf{x}_k \in \mathbb{R}^{D_k}$, so $d_k \ll D_k$ throughout; we refer to d_k as the latent dimension of layer k. Each d_k is even, a requirement of the paired oscillator structure introduced in Section 3.2.

3.2 The Hierarchical Cyclic Vector Field

The mathematical foundation of this architecture addresses the fundamental biological limitations of standard Neural ODEs when modeling stiff cellular dynamics (Curtiss & Hirschfelder, 1952). Previous continuous-time models often utilize simple linear damping terms ($-\gamma\mathbf{z}$) to represent first-order molecular degradation and ensure stability (Alon, 2019; Khalil, 2002; Shaffiey & Pierobon, 2026). However, while mathematically stabilizing, linear damping inherently biases the vector field toward a static point-attractor, structurally failing to capture true biological homeostasis. Living cells are not static, their health is defined by sustained, autonomous rhythms, such as metabolic flux and circadian clocks (Goldbeter, 1996).

To faithfully model these dynamics without collapsing into stasis, we structure the continuous-time trajectory of our latent cell state using a Stuart–Landau limit-cycle prior coupled with a NODE continuous-depth transmission residual (Chen et al., 2018). We utilize this specific Stuart-Landau form because it is a fundamental mathematical template for stable oscillations. In dynamical systems theory, any system that develops a self-sustaining rhythm (a supercritical Hopf bifurcation) reduces to this exact equation (Kuramoto, 1984; Strogatz, 1994). Furthermore, to maintain a symmetric continuous cycle, even-order polynomial terms must mathematically cancel out. Therefore, a negative cubic term is the simplest mechanism available to bound the amplitude and prevent the system from diverging. Ultimately, by simply assuming that these biological layers oscillate, the core structure of our governing equations is mathematically determined rather than chosen.

Let $\mathbf{z}_k \in \mathbb{R}^{d_k}$ denote the latent state of the k-th omics layer, partitioned into $d_k/2$ coordinate pairs $\mathbf{z}_{k,j} = [x_{k,j}, y_{k,j}]^\top$. The latent width d_k must be even to accommodate this planar blocking. The Stuart-Landau prior acts blockwise, meaning all $d_k/2$ planar oscillators within a layer share the physical parameters $(\tau_k, \lambda_k, \omega_k, \gamma_k)$. We write the prior once for a representative pair, dropping j and setting $\mathbf{z}_k = [x_k, y_k]^\top$. In contrast, the neural residual $\hat{f}_{\theta_k}$ is not blockwise; it couples the full layer vectors $\mathbf{z}_k \in \mathbb{R}^{d_k}$ and $\mathbf{z}_{k-1} \in \mathbb{R}^{d_{k-1}}$ jointly. The equations are as follows:

$$\frac{dx_k}{dt} = \tau_k \left(\lambda_k x_k - \omega_k y_k - \gamma_k (x_k^2 + y_k^2) x_k + f_{\theta_k}^{(x)}(\mathbf{z}_k, \mathbf{z}_{k-1}, t) \right) \tag{1}$$

$$\frac{dy_k}{dt} = \tau_k \left(\lambda_k y_k + \omega_k x_k - \gamma_k (x_k^2 + y_k^2) y_k + f_{\theta_k}^{(y)}(\mathbf{z}_k, \mathbf{z}_{k-1}, t) \right) \tag{2}$$

where τ_k is referred to as a time scale, governing biological speed (such that a lower τ_k corresponds to slower dynamic, *e.g.*, epigenetics). Here, ω_k is the intrinsic oscillatory frequency, λ_k acts as a linear expansion rate simulating auto-catalytic molecular production, and γ_k acts as a non-linear damping parameter simulating resource-limited degradation. Together, λ_k and γ_k define the physical amplitude of the homeostatic limit cycle.

Because these parameters are shared across a layer's oscillators, all $d_k/2$ pairs within layer k target a common homeostatic radius $r_k^* = \sqrt{\lambda_k/\gamma_k}$, which is what permits the stability analysis of Section 3.4 to collapse each layer to a single scalar error norm.

f_{θ_k} is the NODE transmission residual processing upstream signals from layer $\mathbf{z}_{k-1}$. In standard acyclic models, this forms a feed-forward information cascade (*e.g.*, Proteins $\rightarrow$ Metabolites). However, true biological autonomy relies on closed-loop feedback (Lu & Thompson, 2012). Due to our cyclic topology (see Figure 1), the index wraps around modulo K, forming a structural causal loop where the fastest downstream layer explicitly acts as an input substrate to regulate the slowest upstream layer (*e.g.*, Metabolites $\rightarrow$ Epigenetics), encoding continuous metabolic feedback directly

into the vector field. Each layer takes a single upstream input, and while other edges exist in the biology, chromatin modification by proteins among them, the ring is not a claim that they are absent but the smallest topology that closes the hierarchy, which is what the analysis and the ablations isolate. The small-gain framework of Dashkovskiy et al. (2007) applies to any interconnection, with the condition becoming a bound on the gain around each cycle of the graph, so further edges can be added without changing the form of Theorem 1.

3.3 Orbit-Centered Parameterization

While the Stuart-Landau prior provides the physical template for oscillation, the unconstrained neural network f_{θ_k} risks inadvertently overpowering the limit cycle and destroying the system's underlying homeostasis. To guarantee that the neural network only models pathological or multi-scale deviations without competing with the biological physics, we enforce a strict geometric constraint.

For every layer k, we define $\mathbf{P}_k = \Pi_\Gamma(\mathbf{z}_k; \lambda_k, \gamma_k)$ as the radial projection of the current latent state onto its ideal homeostatic orbit. The neural transmission is then implemented as an orbit-centered residual:

$$f_{\theta_k}(\mathbf{z}_k, \mathbf{z}_{k-1}, t) = \hat{f}_{\theta_k}(\mathbf{z}_k, \mathbf{z}_{k-1}, t) - \hat{f}_{\theta_k}(\mathbf{P}_k, \mathbf{P}_{k-1}, t) \tag{3}$$

where $\hat{f}_{\theta_k}$ represents the base, unconstrained neural network layer.

Because this residual evaluates exactly to zero when the cell is perfectly on its limit cycle (i.e., when $\mathbf{z}_k = \mathbf{P}_k$), the parameterization mathematically guarantees that the network vanishes during perfect homeostasis. This structural constraint is the mechanism that allows the subsequent stability theorem to be proven, as it anchors the learned dynamics to the physical limit cycle.

3.4 Theoretical Stability Guarantee: Cyclic Small Gain Theorem

Closing the feedback loop in a multi-scale stiff system dramatically increases the risk of numerical explosion. We demonstrate that our architecture guarantees regional orbital stability by strictly bounding the neural residuals against the physical limit-cycle contraction.

Write $r_{k,j}$ for the radius of the j-th oscillator pair of layer k and $\mathbf{e}_k \in \mathbb{R}^{d_k/2}$ for that layer's vector of radial errors $r_{k,j} - r_k^*$. Layer k has its own limit-cycle set $\Gamma_k = \{\mathbf{z}_k \mid r_{k,j} = r_k^*\}$, a product of $d_k/2$ circles of radius r_k^*, and stacking the layers into $\mathbf{e} = (\mathbf{e}_1, \dots, \mathbf{e}_K)$ gives $\Gamma = \Gamma_1 \times \cdots \times \Gamma_K = \{\mathbf{e} = 0\}$, a torus of dimension $\sum_k d_k/2$ rather than a point. Each layer also admits an exogenous perturbation $\mathbf{u}_k$, zero in the unforced case.

Theorem 1 (Regional Input-to-State Orbital Stability)**.** *Consider the K-layer cyclic system of Section 3.2 with homeostatic radius $r_k^* = \sqrt{\lambda_k/\gamma_k}$, and let L_k, L_k^{self} be the Lipschitz constants of f_{θ_k} in $\mathbf{z}_{k-1}$ and $\mathbf{z}_k$. Fix $\mathcal{D}_\delta = \{\mathbf{z} : r_{k,j} \geq \delta_k > 0$ for all $k, j\}$, the states bounded away from the origin by an inner radius δ_k, and set*

$$A_k = \gamma_k \delta_k (\delta_k + r_k^*) - L_k^{\text{self}}, \qquad g_k = \frac{L_k}{A_k}, \qquad \rho = \left(\prod_{k=1}^{K} g_k \right)^{1/K}, \tag{4}$$

where A_k is layer k's net contraction rate, meaning its slowest restoring rate on $\mathcal{D}_\delta$ less the within-layer Lipschitz constant L_k^{self}, g_k is the incoming gain measured against that rate, and ρ, the loop gain, is the geometric mean of the g_k, which falls below one exactly when the product of the gains around the loop does. Assume $\delta_k < r_k^*$, $A_k > 0$ and $L_k > 0$ for all k. Then:

1. ***Unconditional boundedness.*** *Every trajectory enters a compact absorbing set in finite time, irrespective of the loop gain.*
2. ***Conditional attraction.*** *If $\rho < 1$, equivalently $\prod_k L_k < \prod_k A_k$, then Γ is exponentially attractive at rate $\mu = (1 - \rho) \min_k \tau_k A_k > 0$ on the region $\mathcal{B} = \{\mathbf{z} : V(\mathbf{e}) \leq c^\star\}$, where $c^\star = \min_k p_k (r_k^* - \delta_k)$ and the Lyapunov function V and weights p_k are defined in Equation 5 below. The set $\mathcal{B}$ is an inner estimate of the region of attraction rather than the full one, so trajectories starting outside it are not claimed to diverge. Appendix A shows $\mathcal{B}$ lies in $\mathcal{D}_\delta$ and is forward invariant.*
3. ***Input-to-state orbital stability.*** *Under the same hypothesis, let*

$$V(\mathbf{e}) = \sum_{k=1}^{K} p_k \|\mathbf{e}_k\|, \qquad p_1 = 1, \qquad p_{k+1} = \rho \, \frac{a_k}{b_{k+1}} \, p_k, \tag{5}$$

with $a_k = \tau_k A_k$ and $b_k = \tau_k L_k$. Then, for $\mathbf{z}(0) \in \mathcal{B}$,

$$V\big(\mathbf{e}(t)\big) \;\le\; e^{-\mu t}\, V\big(\mathbf{e}(0)\big) \;+\; \frac{1}{\mu} \sum_{k=1}^{K} p_k \tau_k \|\mathbf{u}_k\|_\infty. \tag{6}$$

The hypothesis $\rho < 1$ is independent of τ_k, which cancel in each g_k. The convergence rate μ does depend on them. Larger δ_k gives stronger contraction and a smaller certified region, and as $\delta_k \to r_k^*$ the condition approaches $\prod_k L_k < \prod_k (2\lambda_k - L_k^{\text{self}})$ while $\mathcal{B}$ shrinks toward Γ.

The full proof, which constructs a telescoping weighted Lyapunov function to bound the cross-layer gains around the cycle, is detailed in Appendix A.

Figure 4 reports ρ at every checkpoint of both training runs. It remains strictly below one from initialization onward in each case, ensuring the stability guarantee covers the entire optimization trajectory rather than the converged model alone. Furthermore, integrating the variational equation over one period at the trained optimum yields a transition matrix whose spectrum cleanly separates into neutral phase directions and directions that contract below numerical resolution, with no expanding direction on either dataset. In contrast, both ablations carry genuinely expanding directions. The full monodromy measurement is detailed in Appendix F.

3.5 Training Objective

The objective function is a composite loss balancing trajectory reconstruction and exact topological stability constraints, expressed as follows:

$$\mathcal{L} = \mathcal{L}_{\text{rec}} + \lambda_{\text{stab}} \mathcal{L}_{\text{stab}}\,, \tag{7}$$

where $\lambda_{\text{stab}} = 1.0$ acts as a scaling weight for the Small Gain stability penalty that applies an unscaled magnitude of the topological violation directly to the objective. The components $\mathcal{L}_{\text{rec}}$ and $\mathcal{L}_{\text{stab}}$ are detailed below.

Reconstruction loss $\mathcal{L}_{\text{rec}}$. The reconstruction loss is defined as the sum of Mean Squared Errors (MSE) between observed samples $\mathbf{x}_k(t^{(i)})$ and their reconstructed counterparts $\hat{\mathbf{x}}_k(t^{(i)})$ across all four hierarchical layers:

$$\begin{aligned}\mathcal{L}_{\text{rec}} = \sum_{i=1}^{N} \Big[&\text{MSE}(\mathbf{x}_e(t^{(i)}), \hat{\mathbf{x}}_e(t^{(i)})) + \text{MSE}(\mathbf{x}_g(t^{(i)}), \hat{\mathbf{x}}_g(t^{(i)})) \\ &+ \text{MSE}(\mathbf{x}_p(t^{(i)}), \hat{\mathbf{x}}_p(t^{(i)})) + \text{MSE}(\mathbf{x}_m(t^{(i)}), \hat{\mathbf{x}}_m(t^{(i)}))\Big]\,,\end{aligned} \tag{8}$$

where $\hat{\mathbf{x}}_k(t^{(i)}) = \boldsymbol{\mu}_{\psi,k}(\mathbf{z}_k(t^{(i)}))$ represents the mean of the decoded observation distribution at time $t^{(i)}$. Dimensional scaling factors may be applied to specifically upweight the metabolomic and epigenomic terms to counteract feature-space dimensionality imbalances, which is standard practice in multi-omics integration (Gayoso et al., 2021).

Small Gain Stability penalty $\mathcal{L}_{\text{stab}}$. To strictly enforce the Input-to-State Orbital Stability (ISOS) condition established in Theorem 1, the following soft constraint is added to penalize violations of the cyclic loop bound:

$$\mathcal{L}_{\text{stab}} = \max\left(0,\; \prod_{k=1}^{K} L_k \;-\; \prod_{k=1}^{K} \max\big(10^{-6},\, A_k(r_k^*)\big) \;+\; \epsilon\right)^2, \qquad A_k(r_k^*) = 2|\lambda_k| - L_k^{\text{self}}, \tag{9}$$

Here $\epsilon = 0.01$ is a safety margin and each contraction factor is floored at 10^{-6}, which keeps the product differentiable when a layer transiently violates its own bound. To practically enforce this stability condition during training, a Lipschitz monitoring module dynamically estimates L_k by computing the product of the spectral norms of the weight matrices in each neural sub-network f_{θ_k} (Delattre et al., 2023), using power iteration to find the largest singular value $\sigma(W)$ (Miyato et al., 2018). This topological regularization explicitly bounds the cumulative neural loop gain against the physical limit-cycle contraction rates, which is critical for maintaining robust stability in continuous-time models subject to extreme stiffness (De Marinis et al., 2025; Pauli et al., 2023).

4 EXPERIMENTS AND RESULTS

To empirically validate the theoretical stability guarantees and extrapolative capabilities of the Hierarchical Cyclic NODE, we benchmark the architecture against state-of-the-art continuous-time models on a multi-omics circadian dataset.

4.1 EXPERIMENTAL SETUP AND BASELINES

We assemble a four-layer mouse liver time course from three published studies. The transcriptome and epigenome come from Koike et al. (2012), the proteome from Robles et al. (2014), and the metabolome from the standard-chow arm of Eckel-Mahan et al. (2013). The epigenomic readout is CBP ChIP-seq. CBP is a histone acetyltransferase whose obligate substrate is acetyl-CoA, so it is the mechanistically appropriate readout for the metabolome to epigenome edge. Every layer is kept on the grid at which it was measured and nothing is resampled. A single latent trajectory is integrated over the union of the four grids, 24 timepoints in total, and each layer is scored only where it has an observation. All times are circadian time (CT) in hours, where CT0 is subjective dawn and CT0 to CT24 spans one subjective cycle, so CT24 onward is the second cycle. Table 4 gives the per-layer grids, feature counts and preprocessing.

Models train on the first cycle, up to CT21, and forecast the second, giving 26 training and 15 held-out observations. The epigenome has no observations past CT20, so no held-out error can be computed for that layer. Because the dynamics are recurrent we score against four naive references, the training mean, the last observation held flat, 24-hour periodic persistence and a linear extrapolation, and report ratios against the best of the four and against periodic persistence alone.

We benchmark against five continuous-time models, all sharing our encoder, decoder, optimizer, solver tolerance, epoch budget, and a fixed latent dimension of $d_k = 8$. We restrict the width to $d_k = 8$ because our orbit-centered formulation requires evaluating each sub-network twice per solver step; fixing this dimension ensures computational tractability while guaranteeing that our baseline comparison isolates the vector field architecture rather than model capacity. The baselines include: (1) **Vanilla NODE** (Chen et al., 2018), a standard continuous-depth network without structural priors; (2) **scNODE** (Zhang et al., 2024), a deep generative model for transcriptomic trajectories; (3) **Heavy Ball NODE (HBNODE)** (Xia et al., 2021), a second-order model utilizing momentum and physical friction to mitigate stiffness; (4) **Jacobian-Regularized NODE** (Finlay et al., 2020), which penalizes the Jacobian's Frobenius norm via Hutchinson's trace estimator; and (5) **Acyclic Cascade** (Shaffiey & Pierobon, 2026), a hierarchical cascade using linear damping ($-\gamma\mathbf{z}$) without cyclic feedback.

4.2 QUANTITATIVE BENCHMARK: FORECASTING BIOLOGICAL RHYTHMS

As shown in Table 1, standard continuous models struggle with long-term biological extrapolation when evaluated on their native sparse observational grids. While models like scNODE and Jacobian-Regularized NODE attempt to constrain the continuous-depth dynamics, they suffer from significant temporal drift and high variance across initializations when forecasting the held-out Day 2 trajectory, resulting in Proteomic Mean Squared Errors (MSE) of 21.314 ± 13.002 and 11.672 ± 4.737, respectively. Furthermore, Heavy Ball NODE (HBNODE), which utilizes classical mechanical momentum to tame stiffness, diverges severely across multiple layers (Gene MSE 8.417 ± 3.111, Proteomics MSE 21.924 ± 14.814). This demonstrates that standard mechanical friction priors are structurally incompatible with auto-catalytic biological oscillators.

Table 1: Per-layer 48-hour extrapolation MSE for cycle 2 on the native four-layer grid (averaged across 6 seeds $\pm$ standard deviation). Epigenomics has no Day 2 observations in this atlas, so no held-out error can be computed for that layer in any model.

Model	Metabolomics	Proteomics	Transcriptomics
Vanilla NODE	4.043 ± 0.755	11.261 ± 4.641	4.331 ± 1.458
scNODE	3.507 ± 0.521	21.314 ± 13.002	11.446 ± 12.766
Heavy Ball NODE	5.535 ± 2.222	21.924 ± 14.814	8.417 ± 3.111
Jacobian-Reg. NODE	4.141 ± 0.926	11.672 ± 4.737	4.473 ± 1.615
Acyclic Cascade	2.345 ± 0.403	2.961 ± 0.509	2.587 ± 0.564
Cyclic No-Prior	3.814 ± 0.833	24.311 ± 11.608	6.901 ± 1.406
Cyclic NODE (Ours)	$\mathbf{1.867 \pm 0.057}$	$\mathbf{2.048 \pm 0.025}$	$\mathbf{1.752 \pm 0.012}$

We repeat the benchmark on a second organism, a transcriptome and proteome time course from *Neurospora crassa* (Hurley et al., 2018a), 24 timepoints every 2 h across 48 h of constant darkness. The loop is $K = 2$, transcript to protein and back. Because the culture is free-running rather than light-entrained, the second cycle is not a repeat of the first. The ordering of Table 1 is reproduced and the margins widen, with transcriptome error down 58.7% and proteome error down 65.9% against the strongest continuous-time baseline. Appendix B gives the full table and the ablation, reported on a single seed.

The Acyclic Cascade (Shaffiey & Pierobon, 2026) reaches mean errors between 2.3 and 3.0, ahead of the generic baselines, but pays for its linear damping with the temporal mode collapse Section 4.3. In contrast, our Cyclic NODE strictly bounds the state without destroying the biological rhythm, lowering the error to under 2.1 in every layer while maintaining remarkably tight variance (± 0.057 or less). This represents an improvement over the strongest generic continuous-time baseline in every layer by more than 45%. The stability condition of Theorem 1 holds in every reported run.

4.3 Ablation Study: The Necessity of Prior and Topology

To isolate the specific mechanisms driving our model's performance, we perform a paired ablation study on the held-out Day 2 forecasting window (CT24–CT45). We evaluate the full Cyclic NODE against two structurally degraded variants initialized with identical random seeds to ensure a fair, unconfounded comparison: (1) **Cyclic No-Prior** retains the closed-loop topology but removes the Stuart-Landau prior, centering the neural residual on zero. Because there is no physical limit cycle, Theorem 1 does not apply, and the stability penalty is disabled; and (2) **Acyclic Cascade** replaces the Stuart-Landau prior with simple linear damping ($-\gamma \mathbf{z}$) and breaks the feedback loop (preventing metabolites from regulating epigenetics), recreating the topological constraints of prior biological cascades (Shaffiey & Pierobon, 2026).

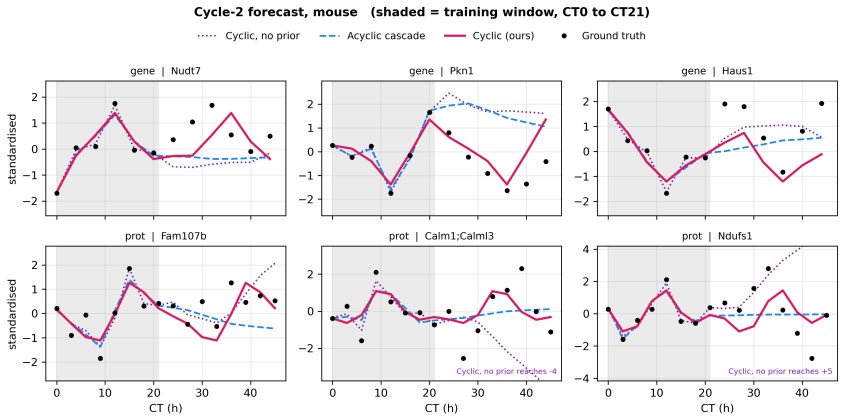


Figure 2: **Cycle-2 forecast, mouse liver atlas.** Cycle-2 forecast on the mouse liver atlas for the three highest-amplitude features of the transcriptome and proteome, ranked on the training window alone. Shaded is the training window CT0 to CT21 and every point to its right is held out. The acyclic cascade relaxes toward a fixed point on leaving that window. The variant without the prior leaves the plotted range and reaches the annotated values. The vertical axis represents the z-score.

Without the cubic term bounding radial amplitude, the unconstrained loop gain drives the trajectory to diverge during Day 2 forecasting, with proteomic MSE exceeding 50.0 by $t = 45$ h. The full model bounds the state and beats the No-Prior variant at all 15 of 15 held-out points, a geometric mean error ratio of 0.144. The Acyclic Cascade uses linear damping to prevent this divergence but falls into temporal mode collapse instead. With the loop severed, the vector field loses its autonomous rhythm, decays toward a point attractor, and drifts out of phase by Day 2. Re-closing the loop under the small-gain bound sustains the oscillation, beating the acyclic baseline at 13 of 15 forecast timepoints with a geometric mean ratio of 0.727. Figure 2 shows both failure modes on the held-out cycle.

4.4 Cross-Layer Phase Response and the Return Edge

Theorem 1 establishes the stability of the orbit Γ, not of any point on it. A transient displacement decays in the radial direction while phase, being neutrally stable along Γ, is left free. The system then returns to Γ at a phase that does not need to be the one where it originated from. A phase response

curve results, which is the standard measurement of circadian biology (Johnson, 1999), and a model converging to rest has no phase to shift and cannot produce such a shift.

The closed loop makes a second prediction. Because the orbit-centered residual of Section 3.3 vanishes identically on Γ, cross-layer coupling is active only while some layer is off its orbit. A displacement applied to one layer therefore activates the coupling for the duration of the transient, propagates around the loop, and leaves every layer it reaches at a shifted phase after the system has relaxed back onto Γ. Which layers it reaches is set by the topology alone.

We displace every oscillator pair of one layer at a given circadian phase, integrate for five further periods, and read the lasting shift in each layer against an unperturbed copy (see Appendix C for displacement size, averaging, and convergence checks). The full cross-layer phase response matrix and its corresponding heatmap visualization are detailed in Table 3 in Appendix C and Figure 3 below. The diagonal is each layer's own phase response curve, ranging from 0.1160 h for the metabolome to 0.8761 h for the epigenome under an identical relative displacement, following the learned timescales.

Six of the twelve off-diagonal entries have no causal path except through the return edge. All six are exactly zero when that edge is removed, against a solver noise floor of 6×10^{-6} h, while the six whose paths run forward through the hierarchy move by at most 0.02 h. Transmission attenuates with each hop around the cycle. In an acyclic cascade these entries are exactly zero, because no path exists. While Sections 4.2 and 4.3 establish that closing the loop improves forecasting accuracy, Table 3 reveals the structural feedback mechanism driving that improvement.

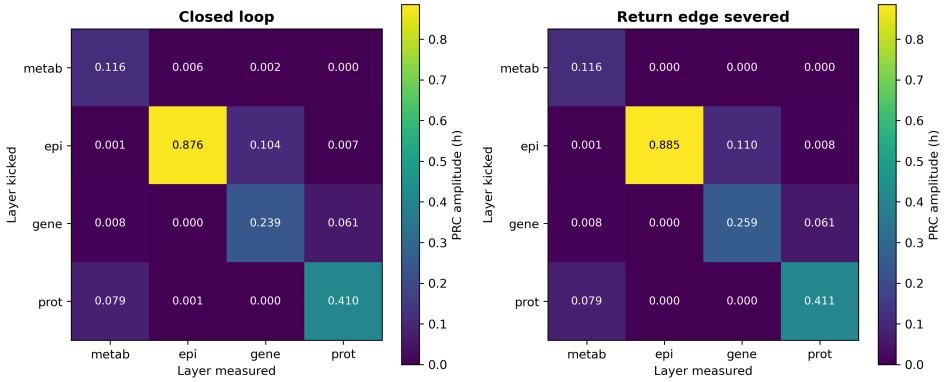


Figure 3: **Cross-layer phase response.** Left, the closed loop. Right, the identical trained weights with the metabolite to epigenome edge zeroed. Six entries collapse to exactly zero and the rest are unchanged, which localizes the effect to that single edge.

5 Conclusion

The regulatory architecture of a biological cell is generally understood as a loop of causally interacting molecular processes that occupy different layers, run on highly diverse timescales, and show inherent periodic behavior. Attempts to obtain continuous-time predictive models of these processes from multi-omics data often suffer from timescale collapse and limited accuracy. To address these problems, we introduced a Hierarchical Cyclic Neural ODE based on a Stuart–Landau limit-cycle prior. A cyclic small-gain theorem guarantees unconditional boundedness and regional exponential orbital attraction by constraining the product of gains around the loop. Because this condition is differentiable, it is enforced directly at initialization and holds throughout optimization.

Evaluated on a four-layer mouse liver atlas and a two-layer *Neurospora crassa* time course, our architecture accurately forecasts held-out cycles, outperforming Neural ODE baselines. Paired ablations demonstrate that removing the prior causes divergence, while breaking the loop induces temporal mode collapse. A cross-layer phase response analysis shows that every entry requiring the return edge falls to exactly zero once that edge is severed, which localizes cross-timescale transmission to the topology. Future extensions may explore learned periods, phase-sensitive couplings, and broader applications in networked dynamical systems (Dörfler & Bullo, 2014).

AI USE STATEMENT

In this work, we used generative AI tools for minor copy-editing (spelling and grammar) and LaTeX typesetting to adhere to conference formatting constraints. We take full responsibility for the scientific claims, mathematical proofs, and final content of this work.

REPRODUCIBILITY STATEMENT

To ensure full reproducibility, the complete step-by-step mathematical proof of the Cyclic Small Gain Theorem is provided in Appendix A. Comprehensive details regarding the multi-omics dataset construction, imputation, and grid alignments are documented in Appendix D. Furthermore, the forward integration and objective function mechanics are detailed, and the algorithm can be found in Appendix E. The fully reproducible Python source code, PyTorch model definitions, and preprocessed datasets are included in the supplementary materials.

REFERENCES

Uri Alon. *An Introduction to Systems Biology: Design Principles of Biological Circuits*. CRC Press, 2nd edition, 2019.

Ron C Anafi, Lauren J Francey, John B Hogenesch, and Junhyong Kim. Cyclops reveals human transcriptional rhythms in health and disease. *Proceedings of the National Academy of Sciences*, 114(20):5312–5317, 2017.

Ricard Argelaguet, Damien Arnol, Danila Bredikhin, Yonatan Deloro, Britta Velten, John C. Marioni, and Oliver Stegle. MOFA+: a statistical framework for comprehensive integration of multi-modal single-cell data. *Genome Biology*, 21(1):111, 2020. doi: 10.1186/s13059-020-02015-1.

Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David K. Duvenaud. Neural ordinary differential equations. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 31, pp. 6571–6583, 2018.

Francis Crick. Central dogma of molecular biology. *Nature*, 227(5258):561–563, 1970.

Charles F. Curtiss and Joseph O. Hirschfelder. Integration of stiff equations. *Proceedings of the National Academy of Sciences*, 38(3):235–243, 1952.

Sergey Dashkovskiy, Björn S. Rüffer, and Fabian R. Wirth. An ISS small gain theorem for general networks. *Mathematics of Control, Signals, and Systems*, 19(2):93–122, 2007.

Andrea De Marinis, Anton Savostianov, and Francesco Tudisco. Contractivity of neural ODEs: An eigenvalue optimization problem. *Mathematics of Computation*, 2025.

Blaise Delattre, Quentin Barthélemy, Alexandre Araujo, and Alexandre Allauzen. Efficient bound of Lipschitz constant for convolutional layers by Gram iteration. In *Proceedings of the 40th International Conference on Machine Learning (ICML)*, pp. 7513–7532. PMLR, 2023.

Florian Dörfler and Francesco Bullo. Synchronization in complex networks of phase oscillators: A survey. *Automatica*, 50(6):1539–1564, 2014.

Kristin L. Eckel-Mahan, Vishal R. Patel, Sara de Mateo, Ricardo Orozco-Solis, Nicholas J. Ceglia, Saurabh Sahar, Sherry A. Dilag-Penilla, Kenneth A. Dyar, Pierre Baldi, and Paolo Sassone-Corsi. Reprogramming of the circadian clock by nutritional challenge. *Cell*, 155(7):1464–1478, 2013.

Chris Finlay, Jörn-Henrik Jacobsen, Levon Nurbekyan, and Adam Oberman. How to train your neural ode: the world of jacobian and kinetic regularization. In *International Conference on Machine Learning*, pp. 3154–3164. PMLR, 2020.

Adam Gayoso, Zoë Steier, Romain Lopez, Jeffrey Regier, Kristopher L. Nazor, Aaron Streets, and Nir Yosef. Joint probabilistic modeling of single-cell multi-omic data with totalVI. *Nature Methods*, 18(3):272–282, 2021.

Albert Goldbeter. *Biochemical Oscillations and Cellular Rhythms: The Molecular Bases of Periodic and Chaotic Behaviour*. Cambridge University Press, 1996.

Ernst Hairer and Gerhard Wanner. *Solving Ordinary Differential Equations II: Stiff and Differential-Algebraic Problems*, volume 14 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, Heidelberg, 2nd edition, 1996.

Jennifer M. Hurley, Meaghan S. Jankowski, Hannah De los Santos, Alexander M. Crowell, Jennifer J. Loros, and Jay C. Dunlap. Circadian proteomic analysis uncovers mechanisms of post-transcriptional regulation in metabolic pathways. *Cell Systems*, 7(6):613–626, 2018a. doi: 10.1016/j.cels.2018.10.014.

Jennifer M. Hurley et al. Quantitative rna-seq and tmt-ms circadian timecourse data from Neurospora crassa wild type and delta csp-1 genetic backgrounds, 2018b.

Zhong-Ping Jiang, Andrew R. Teel, and Laurent Praly. Small-gain theorem for ISS systems and applications. *Mathematics of Control, Signals and Systems*, 7(2):95–120, 1994.

Carl Hirschie Johnson. Forty years of prcs: what have we learned? *Chronobiology International*, 16 (6):711–743, 1999. doi: 10.3109/07420529909016940.

Hassan K. Khalil. *Nonlinear Systems*. Prentice Hall, Upper Saddle River, NJ, 3rd edition, 2002.

Nobuya Koike, Seung-Hee Yoo, Hung-Chung Huang, Vivek Kumar, Choogon Lee, Tae-Kyung Kim, and Joseph S. Takahashi. Transcriptional architecture and chromatin landscape of the core circadian clock in mammals. *Science*, 338(6105):349–354, 2012.

Yoshiki Kuramoto. *Chemical Oscillations, Waves, and Turbulence*, volume 19 of *Springer Series in Synergetics*. Springer-Verlag, Berlin, Heidelberg, 1984.

Gioele La Manno, Ruslan Soldatov, Amit Zeisel, Emelie Braun, Hannah Hochgerner, Viktor Petukhov, Katja Lidschreiber, Maria E. Kastriti, Peter Lönnerberg, Alessandro Furlan, Jean Fan, Lars E. Borm, Zehua Liu, David van Bruggen, Jimin Guo, Xiaoling He, Roger Barker, Erik Sundström, Gonçalo Castelo-Branco, Patrick Cramer, Igor Adameyko, Sten Linnarsson, and Peter V. Kharchenko. RNA velocity of single cells. *Nature*, 560(7719):494–498, 2018.

Chao Lu and Craig B. Thompson. Metabolic regulation of gene expression by histone modification. *Nature Reviews Molecular Cell Biology*, 13(11):705–718, 2012.

Gaurav Manek and J. Zico Kolter. Learning stable deep dynamics models. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

Ron Milo and Rob Phillips. *Cell Biology by the Numbers*. Garland Science, Taylor and Francis Group, New York, 2015. ISBN 978-0-8153-4537-4.

Takeru Miyato, Toshiki Kataoka, Masanori Koyama, and Yuichi Yoshida. Spectral normalization for generative adversarial networks. In *International Conference on Learning Representations (ICLR)*, 2018.

Patricia Pauli, Dennis Gramlich, and Frank Allgöwer. Lipschitz constant estimation for 1D convolutional neural networks. In *Proceedings of the 5th Annual Learning for Dynamics and Control Conference (L4DC)*, pp. 1321–1332. PMLR, 2023.

Lawrence Perko. *Differential equations and dynamical systems*, volume 7. Springer Science & Business Media, 2001.

Christopher Rackauckas, Yingbo Ma, Julius Martensen, Collin Warner, Kirill Zubov, Rohit Supekar, Dominic Skinner, Ali Ramadhan, and Alan Edelman. Universal differential equations for scientific machine learning. *arXiv preprint arXiv:2001.04385*, 2020.

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

Max Revay, Ruigang Wang, and Ian R. Manchester. Recurrent equilibrium networks: Flexible dynamic models with guaranteed stability and robustness. *IEEE Transactions on Automatic Control*, 2023.

Maria S. Robles, Jürgen Cox, and Matthias Mann. In-vivo quantitative proteomics reveals a key contribution of post-transcriptional mechanisms to the circadian regulation of liver metabolism. *PLoS Genetics*, 10(1):e1004047, 2014.

Yulia Rubanova, Ricky T. Q. Chen, and David Duvenaud. Latent ordinary differential equations for irregularly-sampled time series. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

T. Konstantin Rusch and Siddhartha Mishra. Coupled oscillatory recurrent neural network (cornn): An accurate and architecture for learning long-term dependencies. In *International Conference on Learning Representations (ICLR)*, 2021a.

T. Konstantin Rusch and Siddhartha Mishra. Unicornn: A recurrent model for learning very long time dependencies. In *International Conference on Machine Learning (ICML)*, 2021b.

Björn Schwanhäusser, Dorothea Busse, Na Li, Gunnar Dittmar, Johannes Schuchhardt, Jana Wolf, Wei Chen, and Matthias Selbach. Global quantification of mammalian gene expression control. *Nature*, 473(7347):337–342, 2011. doi: 10.1038/nature10098.

Shohaib Shaffiey and Massimiliano Pierobon. Tri-scale neural ODEs for continuous multi-omics disease modeling. In *Proceedings of the 43rd International Conference on Machine Learning (ICML)*, volume 306, Seoul, South Korea, 2026. PMLR.

Eduardo D. Sontag. Smooth stabilization implies coprime factorization. *IEEE Transactions on Automatic Control*, 34(4):435–443, 1989.

Eduardo D. Sontag and Yuan Wang. On characterizations of the input-to-state stability property. *Systems & Control Letters*, 24(5):351–359, 1995.

Steven H. Strogatz. *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering*. Westview Press, Boulder, CO, 1994.

Hedi Xia, Vai Suliafu, Hangjie Ji, Tan Nguyen, Andrea Bertozzi, Stanley Osher, and Bao Wang. Heavy ball neural ordinary differential equations. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 34, pp. 18646–18659, 2021.

Barry M. Zee, Rebecca S. Levin, Bing Xu, Gary LeRoy, Ned S. Wingreen, and Benjamin A. Garcia. In vivo residue-specific histone methylation dynamics. *Journal of Biological Chemistry*, 285(5): 3341–3350, 2010.

Jiaqi Zhang, Kexin Zhang, and Ritambhara Singh. scNODE: Generative model for temporal single cell transcriptomic data prediction. *bioRxiv*, pp. 2023–11, 2024.

A Proof of Theorem 1

Proof. (i) Reduction to one scalar inequality per layer. For each oscillator pair, $r\,\mathrm{d}r/\mathrm{d}t = x\,\mathrm{d}x/\mathrm{d}t + y\,\mathrm{d}y/\mathrm{d}t$, in which the rotational contributions $x(-\omega_k y) + y(\omega_k x)$ cancel identically, so the frequencies never enter. This is the standard polar form of the supercritical Hopf normal form (Kuramoto, 1984; Strogatz, 1994). Since $\lambda_k = \gamma_k (r_k^*)^2$, the radial prior factors exactly as $\lambda_k r - \gamma_k r^3 = -\gamma_k r(r + r_k^*)\, e_{k,j}$, whose coefficient increases in r and is therefore minimized on $\mathcal{D}_\delta$ at $r = \delta_k$. Radial projection satisfies $\|\mathbf{z}_k - \mathbf{P}_k\| = \big|\|\mathbf{z}_k\| - r_k^*\big| = \|\mathbf{e}_k\|$ exactly, so the orbit-centered parameterization $\hat{f}_{\theta_k}(\mathbf{z}_k, \mathbf{z}_{k-1}, t) - \hat{f}_{\theta_k}(\mathbf{P}_k, \mathbf{P}_{k-1}, t)$ makes Γ invariant and converts Lipschitz continuity into $\|f_{\theta_k}\| \le L_k \|\mathbf{e}_{k-1}\| + L_k^{\text{self}} \|\mathbf{e}_k\|$, with L_k and L_k^{self} bounded in practice by products of spectral norms of the corresponding weight blocks (Miyato et al., 2018; Delattre et al., 2023). The norm $\|\mathbf{e}_k\|$ is not differentiable where $\mathbf{e}_k = 0$, which is Γ itself, so the estimate uses

the upper right Dini derivative $\mathrm{d}^+/\mathrm{d}t$, which exists along the trajectory and is what the comparison lemma requires. Together these give, for every layer,

$$\frac{\mathrm{d}^+}{\mathrm{d}t}\|\mathbf{e}_k\| \;\le\; -a_k\|\mathbf{e}_k\| + b_k\|\mathbf{e}_{k-1}\| + \tau_k\|\mathbf{u}_k\|, \qquad a_k = \tau_k A_k, \quad b_k = \tau_k L_k. \tag{10}$$

(ii) The weights close around the cycle. Applying the recursion equation 5 K times multiplies p_1 by $\rho^K \prod_k a_k / \prod_k b_k$, which equals 1 because $\rho^K = \prod_k g_k = \prod_k b_k / \prod_k a_k$ by equation 4. The weights are therefore consistent, and strictly positive since $a_k, b_k, \rho > 0$. This is the step that has no analogue in a cascade, since a chain terminates, whereas here the recursion must return to its starting value, and it does so precisely when $\rho^K = \prod_k g_k$.

(iii) Contraction. Multiplying equation 10 by p_k, summing over k, and reindexing the upstream sum,

$$\frac{\mathrm{d}^+ V}{\mathrm{d}t} \;\le\; \sum_{k=1}^{K} \big(-p_k a_k + p_{k+1} b_{k+1} \big)\|\mathbf{e}_k\| \;+\; \sum_{k=1}^{K} p_k \tau_k \|\mathbf{u}_k\|. \tag{11}$$

By construction $p_{k+1} b_{k+1} = \rho\, p_k a_k$, so every bracket equals $-(1-\rho)\, p_k a_k \le -\mu\, p_k$ with $\mu = (1-\rho)\min_k a_k = (1-\rho)\min_k \tau_k A_k$. Hence $\mathrm{d}^+V/\mathrm{d}t \le -\mu V + w$ with $w = \sum_k p_k \tau_k \|\mathbf{u}_k\|_\infty$, and $\mu > 0$ exactly when $\rho < 1$. The scalar comparison lemma (Khalil, 2002, Lemma 3.4) gives $V(t) \le e^{-\mu t} V(0) + w/\mu$ for as long as the trajectory remains in $\mathcal{D}_\delta$, which is equation 6, and setting $\mathbf{u} \equiv \mathbf{0}$ gives the unforced estimate. Estimate equation 6 is the class-$\mathcal{KL}$ plus class-$\mathcal{K}$ form of input-to-state stability (Sontag, 1989; Sontag & Wang, 1995) with the state norm replaced by the weighted distance to Γ.

(iv) The region $\mathcal{B}$ is forward invariant. Assume $\delta_k < r_k^*$ for every k, let $c^\star = \min_k p_k (r_k^* - \delta_k)$ and let $\mathcal{B} = \{\mathbf{z} : V(\mathbf{e}) \le c^\star\}$. If $V(\mathbf{e}) \le c^\star$ then $p_k\|\mathbf{e}_k\| \le c^\star$ for every k, so $\|\mathbf{e}_k\| \le c^\star/p_k \le r_k^* - \delta_k$, and since $|r_{k,j} - r_k^*| \le \|\mathbf{e}_k\|$ for each oscillator we obtain $r_{k,j} \ge \delta_k$, giving $\mathcal{B} \subseteq \mathcal{D}_\delta$. The estimate $\mathrm{d}^+V/\mathrm{d}t \le -\mu V + w$ of part (iii) is thus valid throughout $\mathcal{B}$, and on $\partial\mathcal{B}$, where $V = c^\star$, it gives $\mathrm{d}^+V/\mathrm{d}t \le -\mu c^\star + w < 0$ whenever $w < \mu c^\star$, which holds with no condition when $\mathbf{u} \equiv \mathbf{0}$. No trajectory leaves $\mathcal{B}$, so the estimate holds for all time and items 2 and 3 follow on $\mathcal{B}$.

(v) Boundedness. Item 1 follows from the radial equation in that every residual term is at most linear in the radii while $-\gamma_k r_{k,j}^3$ is cubic, so $\mathrm{d}r_{k,j}/\mathrm{d}t < 0$ outside a bounded ball, independently of ρ, which is ultimate boundedness in the sense of Khalil (2002, Thm. 4.18). No attraction claim is made for trajectories beginning outside $\mathcal{B}$. □

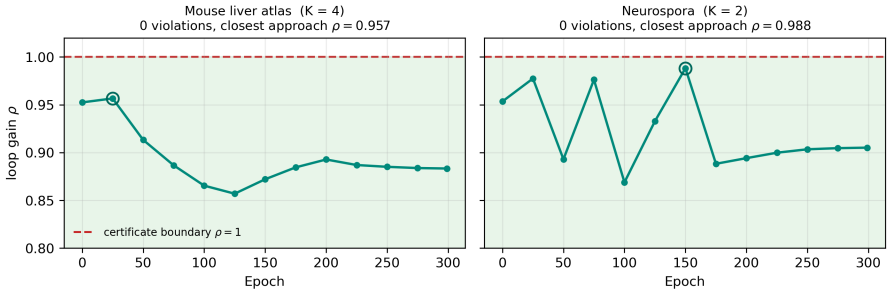


Figure 4: Loop gain $\rho = \big(\prod_k L_k / \prod_k A_k\big)^{1/K}$ at every checkpoint of training, on the mouse liver atlas at $K = 4$ (left) and *Neurospora crassa* at $K = 2$ (right). Shaded is $\rho < 1$, where Theorem 1 applies. Both runs satisfy it at initialization and at every subsequent check. On *Neurospora* the loop gain reaches 0.988 at epoch 150 and is driven back below one at the next checkpoint. Open circles mark the closest approach.

B Second Dataset: *Neurospora crassa* at $K = 2$

Why this dataset. There are two properties that makes it a suitable additional test. First, it has two molecular layers rather than four, so it instantiates the arbitrary-K claim of Section 3 on a real system rather than in principle. The loop is transcript to protein and back, the return edge being

transcription-factor feedback from the proteome onto transcription. Second, the culture free-runs in constant darkness. Its period drifts off 24 h and its amplitude damps across the second cycle, so forecasting cycle two is a genuine extrapolation. In an entrained time course the second day is close to a replicate of the first, and a model can score well by detecting periodicity alone.

Construction. We use the imputed wild-type matrices from the deposit of Hurley et al. (2018b) accompanying Hurley et al. (2018a), which report transcript and protein abundance at 24 timepoints every 2 h across 48 h in constant darkness, in biological triplicate. Replicates are averaged within each timepoint and the two layers are intersected on gene identity, giving 4741 genes carrying both a transcript and a protein trajectory on an identical grid. Both layers arrive centered per feature and receive no further transform beyond the train-only scaler used throughout. The grid is shifted by -2 h so that both layers observe $t = 0$, which the encoder requires. Nothing is resampled.

Protocol. Identical to Section 4.1 in every respect that can be held fixed. The same encoder and decoder, the same latent dimension of 8 per layer, the same optimizer, solver tolerance and epoch budget, and the same paired initialization stream across variants. The period is set to 22 h, the free-running period of *N. crassa*, so that 0 to 22 h in darkness is exactly one cycle of 12 training points and 24 to 46 h is a forecast cycle of 12 held-out points. Both layers have ground truth throughout the forecast window, giving 24 scored observations against 15 on the mouse liver atlas.

Table 2: Cycle-2 forecast MSE on the *Neurospora* course, all models on the identical task. Lower is better. The ordering of Table 1 is reproduced with wider margins.

Model	Transcriptome	Proteome	Mean
Vanilla NODE	7.915	4.586	6.251
scNODE	4.197	10.543	7.370
Heavy Ball NODE	10.709	47.262	28.985
Jacobian-Reg. NODE	8.126	4.499	6.312
Acyclic Cascade	2.208	**1.510**	1.859
Cyclic No-Prior	17.259	47.602	32.431
Cyclic NODE (Ours)	**1.732**	1.534	**1.633**

Results. Against the strongest continuous-time baseline on each layer, our model reduces transcriptome error by 58.7% and proteome error by 65.9%, and mean error across both layers by 73.9%. Both figures exceed the $> 26\%$ margin of Table 1, so the four-layer result is not an artefact of that particular atlas.

The two failure modes are the ones Section 1 anticipates. Heavy Ball NODE reaches 47.262 on the proteome, the same divergence it shows on the mouse liver atlas, which is what a mechanical friction prior does when the system it is applied to is autocatalytic. The no-prior cyclic ablation reaches 47.602 on the same layer, statistically indistinguishable from it, so the cubic amplitude bound and the friction term fail in the same way for the same reason. scNODE is the best of the generic baselines on the transcriptome at 4.197 and among the worst on the proteome at 10.543, consistent with a model designed for transcriptomic trajectories and then asked to carry a second modality in a monolithic latent state.

Ablation. Against the paired ablations, and with the identical initialization stream, the cyclic model beats the no-prior variant on $24/24$ held-out observations at a geometric mean error ratio of 0.077, a 92.3% reduction. That is a larger separation than the 0.144 obtained on the mouse liver atlas, so the unconditional boundedness claim of Theorem 1 item 1 is doing more work here, not less. Against the acyclic cascade the cyclic model wins $16/24$ observations at a geometric mean ratio of 0.892, with the transcriptome at 0.787 across $10/12$ points and the proteome a tie at 1.010. The proteome tie is visible in Table 2 as well, where the acyclic cascade is ahead by 1.6%, which is within run-to-run variation. Figure 5 shows the forecast trajectories for both layers.

Small-gain condition. The small-gain condition held at every checkpoint across all 300 epochs, with ρ ranging from 0.869 to 0.988 and no violations, and with zero skipped parameter updates.

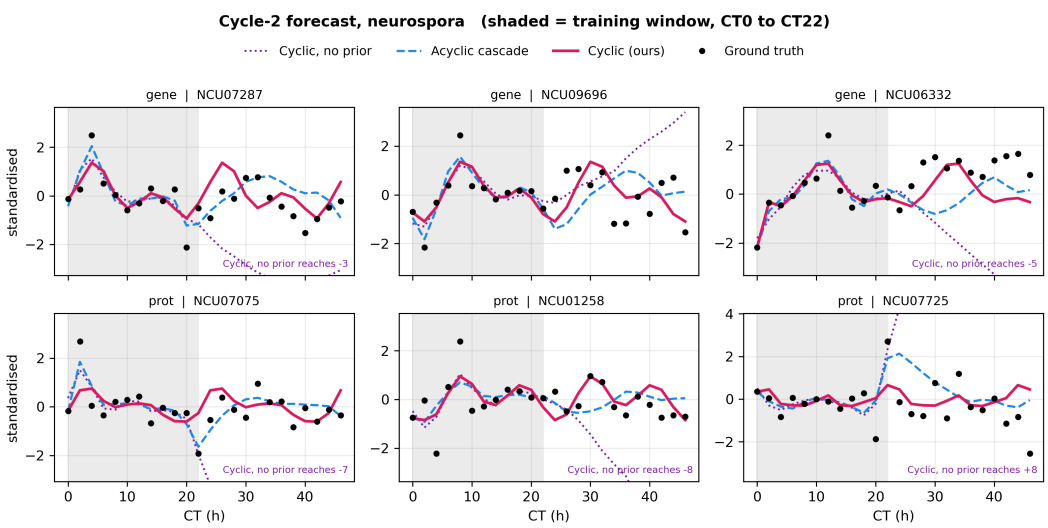


Figure 5: Cycle-2 forecast on the *Neurospora crassa* course in constant darkness, seed 0, for the three highest-amplitude transcripts and proteins, selected on the training window as in Figure 2. The shaded band is CT0 to CT22 and the remainder is held out. The same two failure modes appear at $K = 2$ that appear at $K = 4$. The acyclic cascade loses amplitude across the forecast window and the variant without the limit-cycle prior diverges past the plotted range. The full model remains bounded and rhythmic throughout, which is the behavior the theorem guarantees and the accuracy in Table 2 reflects. The vertical axis represents the z-score.

The condition is therefore not specific to $K = 4$, and the rescaling of Section 3.5 enforces it at initialization at $K = 2$ exactly as it does at $K = 4$.

C PHASE RESPONSE PROTOCOL

We integrate the trained model to its attractor, then at circadian phase ϕ displace every oscillator pair of one layer by $0.30\, r_k^*$ along that pair's x axis, and integrate for a further 120 h, roughly five periods. The lasting phase shift in each layer is the radius-weighted circular mean of the per-pair angular difference against an unperturbed copy read at the same absolute time, converted to hours. We repeat at 24 phases spanning one cycle and report the peak-to-trough amplitude. Each shift is read again one full period later and agrees to within 6×10^{-6} h, so it is permanent rather than still relaxing.

Rather than compare against a separately trained acyclic model, which would confound the topology with the retraining, we sever one edge in the trained cyclic model itself. Zeroing the cross-input weight block of the epigenomic sub-network removes the metabolite to epigenome connection and changes nothing else. Every remaining weight is identical, so any difference in the response matrix is attributable to that one edge.

Table 3: Cross-layer phase response amplitude in hours, closed loop against the same model with the return edge severed. Rows are the perturbed layer, columns the measured layer. **Bold** marks the six entries whose only causal path runs through the return edge, all exactly zero without it against a noise floor of 6×10^{-6} h. The remaining paths move by at most 0.02 h.

	Closed loop				**Return edge severed**			
Perturbed	Metab	Epi	Gene	Prot	Metab	Epi	Gene	Prot
Metab	0.1160	**0.0056**	**0.0016**	**0.0002**	0.1159	**0.0000**	**0.0000**	**0.0000**
Epi	0.0009	0.8761	0.1043	0.0066	0.0010	0.8850	0.1104	0.0085
Gene	0.0080	**0.0001**	0.2394	0.0608	0.0081	**0.0000**	0.2588	0.0607
Prot	0.0787	**0.0005**	**0.0002**	0.4101	0.0787	**0.0000**	**0.0000**	0.4111

D DATASET CONSTRUCTION

Table 4: Native layer grids after construction. Nothing is resampled at any stage of the pipeline. Feature counts are those the model sees, which for the transcriptome and proteome are one fewer than the published tables because a single duplicated gene symbol in each is averaged to one row.

Layer	Features	Points	Grid	Source
Metabolome	306	7	CT0-CT24, 4 h	Eckel-Mahan et al. (2013)
Epigenome (CBP)	652	6	CT0-CT20, 4 h	Koike et al. (2012)
Transcriptome	1370	12	CT0-CT44, 4 h	Koike et al. (2012)
Proteome	185	16	CT0-CT45, 3 h	Robles et al. (2014)
Union grid	–	24	CT0-CT45	–

Layer construction. The transcriptome is built from the intronic RPKM columns of the cycling-gene table of Koike et al. (2012), using intronic rather than exonic signal because it tracks nascent transcription and therefore the rate of the process being modelled rather than the steady-state pool **(La Manno et al., 2018)**. The epigenome is the CBP track of the same study's ChIP-seq peak table. Peak-level tag counts are restricted to the cycling gene set and summed per gene, which reduces 9113 raw peaks to 652 gene-level features, makes the epigenome to transcriptome edge gene-matched, and brings the layer to a width comparable to the others. CBP is selected from the eleven available ChIP targets because it is a histone acetyltransferase whose obligate substrate is acetyl-CoA, the metabolite that closes the loop in our topology(Lu & Thompson, 2012). The proteome is the cycling protein matrix of Robles et al. (2014), measured by quantitative mass spectrometry at 3-hourly resolution. The metabolome is the standard-chow arm of the panel of Eckel-Mahan et al. (2013), with the high-fat-diet arm excluded. In the transcriptome and the proteome one gene symbol appears twice, and those rows are averaged, giving 1370 and 185 features respectively.

Preprocessing. The transcriptome and the epigenome are heavy-tailed counts and receive a $\log(1+x)$ transform, without which a handful of high-abundance features dominate the mean squared error. The metabolome is a median-normalized ratio and receives a $\log$ transform. The proteome is already reported as a normalized ratio and is left untransformed. Each layer is then standardized per feature using means and variances fit on the training timepoints only, so no held-out observation contributes to the units in which error is reported.

Native grids. The proteome grid is 3-hourly and contains no CT4, CT8, CT16 or CT20. A pipeline that intersects all four layers onto a shared 6-point grid can only produce four of the six proteome columns by interpolation. Scoring a model against an interpolated target and then comparing it to a linear-interpolation baseline is circular, which is why every layer here is kept on its own grid and scored only where it has data.

Limitation of the assembled atlas. The four layers derive from three independent studies and therefore from three separate animal cohorts. They are aligned by circadian phase rather than measured in the same animals, so the cross-layer couplings the model learns are population-level phase relationships and not within-animal causal measurements. No published mouse liver time course measures all four layers in the same animals at this resolution. This affects the biological reading of the learned couplings. It does not affect the forecasting evaluation, which scores each layer against its own held-out observations from its own study.

E ALGORITHM: TRAINING THE HIERARCHICAL CYCLIC NEURAL ODE

Algorithm 1 details the forward integration and loss computation for a single training iteration, mirroring the explicit PyTorch implementation. The continuous-depth transmission residuals are centered on their respective limit-cycle orbits via radial projection, while spectral norms are monitored to enforce the Cyclic Small Gain constraint.

Algorithm 1 Hierarchical Cyclic Neural ODE Training

Input: Multi-omics snapshots $\{(\mathbf{x}_1(t^{(i)}), \ldots, \mathbf{x}_K(t^{(i)}), t^{(i)})\}$, safety margin $\epsilon = 0.01$.
Hyperparameters: $\lambda_{\text{stab}} = 1.0$
Initialize trainable parameters:
$\theta =$ Neural residual network weights $\{f_{\theta_k}\}_{k=1}^{K}$
$\Phi =$ Physical oscillator parameters $\{\tau_k, \lambda_k, \omega_k, \gamma_k\}_{k=1}^{K}$
$\psi =$ Encoder q_ϕ and decoder p_ψ weights
for epoch $= 1$ to E **do**
for batch $\in \mathcal{D}$ **do**
1. **Encode:** $\mathbf{z}(0) = [q_\phi(\mathbf{x}_1(0)), \ldots, q_\phi(\mathbf{x}_K(0))]^\top$.
2. **Integrate:** $\mathbf{z}(t)$ via Dopri5 solving the cyclic vector field (with k-1 wrapping modulo K):

$$\frac{d\mathbf{z}_k}{dt} = \tau_k\Big(\text{SL}(\mathbf{z}_k) + f_{\theta_k}(\mathbf{z}_k, \mathbf{z}_{k-1}, t) - f_{\theta_k}(\mathbf{P}_k, \mathbf{P}_{k-1}, t)\Big)$$

where SL is the Stuart-Landau prior and $\mathbf{P}_k$ *is the radial orbit projection* $\Pi_\Gamma(\mathbf{z}_k)$.
3. **Decode and compute** $\mathcal{L}_{\text{rec}}$: $\hat{\mathbf{x}}_k(t) = p_\psi(\mathbf{z}_k(t))$.
4. **Estimate neural gains:** Compute cross-layer gain L_k and intra-layer gain L_k^{self} via power iteration on the spectral norms of $\{f_{\theta_k}\}$.
5. **Compute stability penalty:** Evaluate physical contraction $A_k(r_k^*) = 2|\lambda_k| - L_k^{\text{self}}$,

$$\mathcal{L}_{\text{stab}} = \max\left(0, \prod_{k=1}^{K} L_k - \prod_{k=1}^{K} \max(10^{-6}, A_k) + \epsilon\right)^2$$

6. **Update** θ, Φ, ψ via AdamW on $\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{rec}} + \lambda_{\text{stab}}\mathcal{L}_{\text{stab}}$.
end for
end for

F MONODROMY ANALYSIS OF THE TRAINED MODELS

Theorem 1 claims orbital stability, which is a weaker and more delicate statement than asymptotic stability (Khalil, 2002). A trajectory displaced sideways off the invariant set must return to it, while a trajectory displaced along the set must not return, because that displacement is a re-timing rather than an error. The object that distinguishes the two is the state transition matrix of the variational equation over one period (Perko, 2001),

$$\Phi(t_0 + T, t_0) \;=\; \frac{\partial \mathbf{z}(t_0 + T)}{\partial \mathbf{z}(t_0)}, \tag{12}$$

whose eigenvalues μ_i we refer to as multipliers(Strogatz, 1994). A multiplier of modulus one is a neutral direction, a modulus below one is contraction over a cycle, and a modulus above one is a true instability.

Procedure. We take the trained checkpoint, integrate to $t_0 = T$ so that the initial transient has decayed, and form Φ by central differences on each latent coordinate, carrying all $2d_{\text{lat}}K$ perturbations and the unperturbed state through a single solver call. The model is cast to double precision and the solver tolerance tightened to 10^{-8}, since finite differences at the training tolerance would measure solver error rather than dynamics. Halving the step size changes the leading moduli by less than 10^{-8}, so the differences are in the linear regime. Nothing is retrained.

Two independent checks. Beyond the spectrum we report the relative return distance $\|\mathbf{z}(t_0+T) - \mathbf{z}(t_0)\|/\|\mathbf{z}(t_0)\|$ of the unperturbed trajectory, which tests whether the learned path closes after one period at all, and the gain and cosine obtained by pushing the velocity $f(t_0, \mathbf{z}(t_0))$ through Φ, which tests neutrality along the orbit without any eigendecomposition.

Reading the cyclic column. Every transverse direction contracts below what the measurement can resolve within a single cycle, and no direction expands. The largest transverse modulus we recover is 3.9×10^{-9} on the mouse atlas and 4.1×10^{-9} on *Neurospora*, both of which sit at the solver tolerance of 10^{-8} and at the level of the step-halving check, so they are floors on the measurement

Table 5: One-period monodromy of the trained models. The cyclic model closes to within 10^{-6} of a T-periodic orbit, contracts every transverse direction below numerical resolution, and carries no expanding direction on either dataset. Both ablations fail to close and carry expanding directions, and in both the expansion is not aligned with the orbit, since the tangent cosine is near zero.

Dataset	Model	return dist.	neutral	transverse $\lvert\mu\rvert$	expanding	$\max \lvert\mu\rvert$
3*Mouse ($K = 4$)	Cyclic (ours)	7.0×10^{-7}	16	3.9×10^{-9}	0	1.000
	Acyclic cascade	1.76	0	1.904	4	1.904
	Cyclic, no prior	2.55	0	3.715	14	3.715
3*_Neurospora_ ($K = 2$)	Cyclic (ours)	3.4×10^{-6}	8	4.1×10^{-9}	0	1.000
	Acyclic cascade	1.58	0	4.093	1	4.093
	Cyclic, no prior	3.08	0	3.276	6	3.276

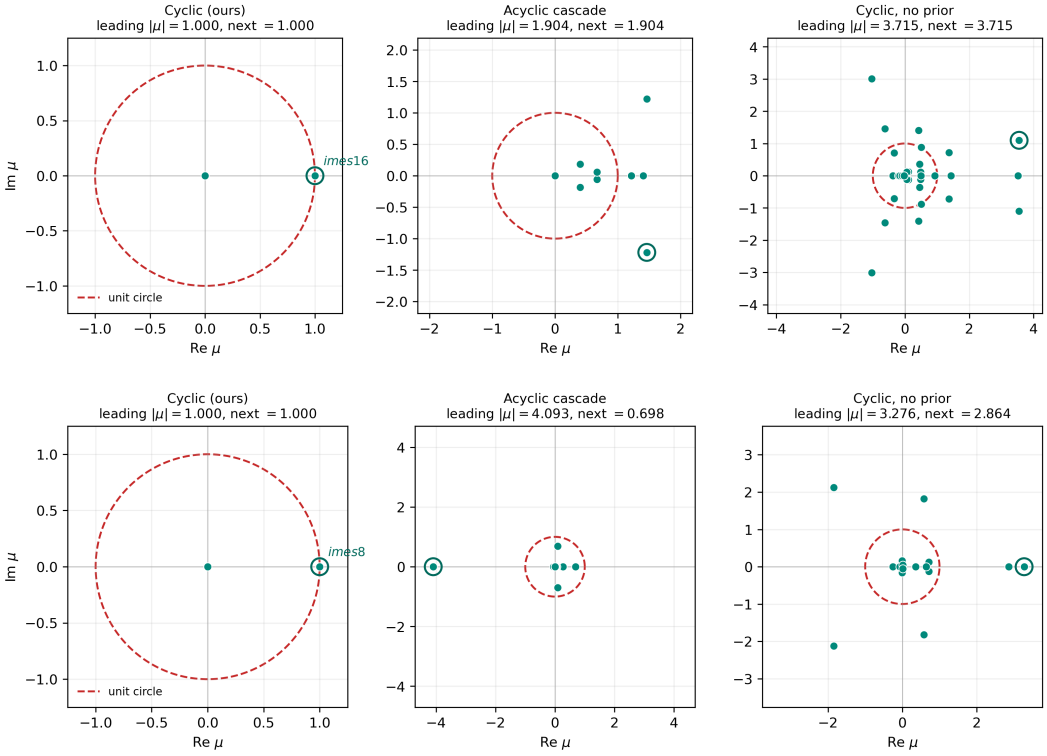


Figure 6: One-period multipliers against the unit circle, mouse liver atlas (top) and *Neurospora crassa* (bottom). For the cyclic model every multiplier lies either on the circle, one per planar oscillator and marked with its multiplicity, or at the origin, so perturbations either re-time the trajectory along Γ or vanish within a cycle. The acyclic cascade and the no-prior variant both place multipliers outside the circle, and in each case the tangent cosine of Table 5 is near zero, so that expansion is not a re-timing along any orbit.

rather than properties of the model. The conservative reading is contraction by at least eight orders of magnitude per cycle. This is the strongest form the orbital conclusion of Theorem 1 can take. The tangent test agrees independently, returning a gain and a cosine of 1.000 to five figures on both datasets. Figure 6 shows the spectra.

The invariant set is the torus Γ of Section 3.4. Each layer carries $d_{\mathrm{lat}}/2$ planar oscillators, so the architecture admits up to $Kd_{\mathrm{lat}}/2$ neutral phase directions, and we measure exactly that many, sixteen at $K = 4$ and eight at $K = 2$. The invariant set is therefore an attracting torus of that dimension, which is exactly the manifold $\Gamma = \{\mathbf{z} \mid r_{k,j} = r_k^*\}$ that Theorem 1 is stated over, and not an isolated closed curve. This follows from the construction. Differentiating the orbit-centered residual $\tilde{f} = \mathrm{net}(\mathbf{z}) - \mathrm{net}(\Pi\mathbf{z})$ at a point where $\mathbf{z} = \Pi\mathbf{z}$ gives

$$\frac{\partial \tilde{f}}{\partial \mathbf{z}} \;=\; J_{\mathrm{net}}\left(I - \frac{\partial \Pi}{\partial \mathbf{z}}\right), \tag{13}$$

and $I - \partial\Pi/\partial\mathbf{z}$ is the projector onto the radial direction, so the coupling responds to radial displacement and not to phase displacement. We measure the tangential response at 2.4×10^{-10} against a radial response of 0.65.

The coupling does not phase-lock the oscillators, and we make no such claim. The individual trajectory is nonetheless exactly T-periodic, as the return distance confirms, because the frozen-period parameterization fixes $\tau_k\omega_k = 2\pi/T$ for every oscillator, so all of them advance at a common rate with offsets set by the initial condition. Phase agreement across layers is therefore imposed by the parameterization and is not an emergent property we claim to have learned.

Why neutrality is the correct behavior here. A circadian oscillator that is phase-shifted stays shifted, which is what makes a phase response curve a meaningful measurement in the first place(Johnson, 1999). Had the model converged to an isolated orbit with a pinned phase, every shift would have decayed and the cross-layer responses of Section 4.4 would have been zero for reasons unrelated to topology. The neutral directions measured here are the mechanism behind those responses, and the exact zeros obtained when the return edge is removed show that the transfer between layers runs through the closed topology.

The caveat is that the learned vector field takes a normalized time input $t/t_{\max}$ and is therefore not exactly T-periodic, so these are eigenvalues of the one-period state transition matrix rather than Floquet multipliers in the strict sense. The interpretation is unchanged, since a modulus below one still means contraction across a cycle and a modulus at one still means neutrality.