WHY DIRECTLY LEARNING PERIODIC TRAJECTORIES CAN FAIL

Anonymous authors
Paper under double-blind review

ABSTRACT

Operator learning of periodic solutions requires deciding how simulation data should be recorded and represented. A natural choice is to integrate long enough for transients to decay and record a window wide enough to contain at least one full period of all trajectories. We find that these conservative choices can make the resulting trajectories difficult to learn, even when the underlying periodic orbits vary regularly with system parameters. Unaligned trajectories generalize poorly even within the training distribution. Phase alignment substantially improves in-distribution generalization, but models trained on a fixed physical-time window still have large errors on trajectories with periods outside the training range. We explain both failures through a common mechanism: frequency differences accumulate over time, so the target phase varies rapidly with the parameters. Predictors that cannot track this variation incur a population MSE floor in both settings; for fixed window prediction, we also derive a per-sample lower bound. We then study one of the simplest representations that escape these floors: learning an aligned, normalized waveform and its period separately. We establish regularity of the decoupled targets under ODE assumptions and show experimentally that this approach avoids both failures in ODE systems and a PDE case study.

1 INTRODUCTION

Periodic dynamics are ubiquitous in science and engineering, from biological rhythms and chemical oscillations to mechanical systems and fluid flows (Vitaterna et al., 2001; Cassani et al., 2021; Chai et al., 2021; Juniper & Sujith, 2018). A central problem across these applications is to predict how the sustained periodic orbits change with system parameters.

Learning full trajectories and solution operators from system parameters is a standard problem in scientific machine learning (Karniadakis et al., 2021; Kovachki et al., 2024; Li et al., 2020). When the main interest is the long-time behavior rather than the transient trajectory, it is natural to learn the long-time solution directly. For systems converging to an equilibrium, one can learn the steady state (Marwah et al., 2023). Multiple steady states have also received some attention (Zhang et al., 2023). We study the analogous direct-learning problem when the long-time solution is periodic rather than stationary.

Unlike a steady state, a periodic orbit does not determine a unique trajectory target. The same orbit can be recorded from any starting phase, and its period usually varies with the system parameters. A trajectory target must therefore specify both a starting phase and a time parameterization. Although these choices do not change the orbit itself, they change the resulting parameter-to-trajectory map and may make accurate prediction difficult.

Related work. Similar issues appear in other fields that study oscillatory data. Uncertainty quantification uses constant phase interpolation, stochastic time warping, and rescaled time integration to approximate oscillatory trajectories (Witteveen & Bijl, 2008; Mai & Sudret, 2017; Le Maître et al., 2010). Functional data analysis uses curve registration and phase-and-amplitude separation to distinguish timing variation from waveform variation (Marron et al., 2015; Tucker et al., 2013; Ramsay & Li, 1998). Time series forecasting uses phase alignment across periodic windows (Wang et al., 2026), while other data driven methods learn asymptotic phase, coupling functions, or reduced phase dynamics (Takata et al., 2023; Hwang et al., 2026; Wilson, 2024).

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.

These works either modify the representation of oscillatory data or learn reduced phase dynamics. We study this representation problem for parameterized dynamical systems, where periodic trajectories are generated by a common family of equations rather than treated only as oscillatory curves. This setting connects the problem directly to scientific applications. It also provides useful mathematical structure to formalize how the orbit and period vary regularly with the system parameters. This allows us to ask not only which representation predicts well, but also what the failing representations have in common, how their shared property creates error floors, what property allows phase-waveform separation to avoid the same mechanism, and how we can find some principle to design future representations. Our contributions are summarized below:

- We identify two representation-induced failures in learning parameterized periodic trajectories via ablation studies. Phase misalignment obstructs iid generalization, whereas a fixed physical-time coordinate can remain accurate iid but fail when extrapolating to unseen periods.
- We identify a general phase-accumulation mechanism underlying both failures: coupling a parameter-dependent frequency to a large timescale amplifies small frequency differences into large phase variation. A unified theorem gives the resulting population error floor. Moreover, for fixed-window representation, a small distance from training period range can produce a large per-sample MSE.
- We identify a general way to escape this mechanism: the target should remove, control, or separately represent the accumulating phase. Phase alignment followed by separate learning of the normalized waveform and period is one instance of this principle. We establish continuity and neural-network approximability of this target for regular attracting limit cycles, and show how one full-state snapshot recovers physical-time predictions.

2 When Conservative Data Preparation Hurts

Motivation. Suppose we can generate training trajectories, either by numerically solving an ODE or PDE or by collecting data from a physical system whose parameters we can control. We must still decide when to begin recording and how long to record.

One natural choice is to wait until the same, conservatively late time for every parameter value, so that the trajectories are expected to have reached their periodic orbits. Another is to record over the same, conservatively wide physical-time window $[0, W]$, chosen to cover the longer periods we expect, including at parameter values not seen during training. Counterintuitively, we find that each choice can independently create a distinct generalization failure when learning periodic trajectories.

Formalization. We formalize this setting as a parameterized autonomous system:

$$\dot{x}(t) = f(x(t), p), \qquad x(0) = x_0, \qquad p \in \mathcal{P} \subset \mathbb{R}^m. \tag{1}$$

We assume that, given parameter p, each trajectory approaches a unique attracting periodic orbit $\mathcal{C}(p)$ with the period $T(p)$. Our goal is to learn how the resulting cycle varies with p, rather than the full trajectory from x_0, which also includes the initial approach to the cycle. We examine the two recording choices in the following ablations.

2.1 Two choices, two failure modes

To examine the two recording choices separately, we use the Goodwin oscillator (Gonze & Ruoff, 2021) and the higher-dimensional Goldbeter circadian model (GOLD95) (Goldbeter, 1995) as ordinary differential equation (ODE) testbeds. Both have stable periodic orbits in the selected parameter domains. We compare waveform targets with and without phase alignment and period normalization under IID and period-based out-of-distribution (OOD) splits. Dataset and model details are given in Appendices D and E. All reported waveform MSE values are computed after applying a Yeo–Johnson transform and standardizing each output variable using the training data.

Failure mode 1. We first establish a working baseline. Motivated by representations that separate timing from shape in other fields (Marron et al., 2015; Tucker et al., 2013; Ramsay & Li, 1998), we learn the period and waveform separately. For each trajectory, we extract one cycle, map it to a common phase interval, and align its waveform by matching peaks of a reference variable. The

resulting waveforms are learned accurately: mean IID validation MSE is 3.79×10^{-4} on Goodwin and 7.19×10^{-4} on GOLD95. We discuss how the learned period and waveform can be used for trajectory prediction in Section 5.3.

Does this success require both preprocessing steps? We first remove period normalization while retaining phase alignment. Somewhat surprisingly, the fixed-window targets remain easy to learn under IID splits: mean validation MSE is 5.48×10^{-5} on Goodwin and 8.28×10^{-4} on GOLD95. On Goodwin, the error is even lower than with period normalization.

Removing phase alignment produces a different result. On Goodwin, we integrate from a shared initial condition for a long fixed time and record without aligning the resulting waveforms. Mean IID validation MSE rises to 0.997 without period normalization and 0.970 with it, although training error remains below 1.1×10^{-3} in both cases. This test includes both the initial approach to the cycle and subsequent movement along it. To isolate the latter, we align the converged orbits of GOLD95 first, then evolve each for the same long time. Its unaligned targets again have mean IID validation MSE above 1. A common long integration after alignment is therefore sufficient to produce this failure, which we call *FM1*.

Failure mode 2. The IID ablations show that phase-aligned fixed-window targets can achieve low validation error without period normalization. But will this still hold when validation periods extend beyond the training range? Intuitively, it's possible because partial trajectory is still a trajectory.

Surprisingly, the ablation study tells a different story. We train on shorter-period trajectories and validate on longer-period ones, setting W to the longest training period. Fixed-window targets have mean OOD validation MSE of 0.228 on Goodwin and 0.200 on GOLD95; period normalization reduces these errors to 0.0104 and 0.0163, respectively. Thus, under the OOD split, lack of period normalization increases validation error by more than an order of magnitude. We call this failure mode 2 (*FM2*).

Table 1 reports the ablations over five training seeds. Both failure modes occur consistently across seeds. For FM1, unaligned IID targets have validation MSE near or above 1, whereas their phase-aligned counterparts remain below 10^{-3}. For FM2, fixed-window OOD validation MSE is more than 10-fold higher than the period-normalized MSE on both systems. The standard deviations and ranges across seeds are small relative to these gaps, and the MSE ranges of the compared settings do not overlap.

3 Error Floors for Parameterized Periodic Targets

"Coincidence" in two failure modes. The failed experiments in Section 2 share a feature: both use a large time scale in data preparation, namely the integration time c in FM1 and the observation-window width W in FM2. Does this shared feature explain the two failures? To investigate, we express their targets in a common form and analyze the resulting prediction errors.

Common form. For each parameter p, let $U(\cdot; p)$ denote a phase-aligned representation of the periodic orbit. The reference phase used for alignment may be chosen arbitrarily but is then fixed throughout the analysis. We write $x_p(t) = U(t\omega(p); p)$, where $\omega(p) = 1/T(p)$ and $U(\cdot; p)$ is one-periodic on $\mathbb{T} = \mathbb{R}/\mathbb{Z}$.

In FM1, shifting the aligned waveform by a common physical time c gives $F_c^{(1)}(p)(s) = U(s + c\omega(p); p)$. In FM2, recording it over $[0, W]$ and setting $t = Ws$ gives $F_W^{(2)}(p)(s) = U(Ws\omega(p); p)$. Both targets have the form

$$\Phi_\lambda(s, p) = \alpha(s) + \lambda\beta(s)\omega(p), \qquad F_\lambda(p)(s) = U\big(\Phi_\lambda(s, p); p\big), \tag{2}$$

with $(\alpha, \beta, \lambda) = (s, 1, c)$ for FM1 and $(\alpha, \beta, \lambda) = (0, s, W)$ for FM2. Hence $\nabla_p \Phi_\lambda(s, p) = \lambda\beta(s)\nabla\omega(p)$: in both cases, increasing the time scale amplifies phase variation across parameters.

To analyze this common form, let $X = \mathbb{R}^n$ for an ODE or $X = L^2(\mathcal{D}_x; \mathbb{R}^n)$ for a PDE, and let $\mathcal{H} = L^2([0, 1]; X)$ with $\|v\|_{\mathcal{H}}^2 = \int_0^1 \|v(s)\|_X^2 \, ds$. For each p, define the *phase mean* $\bar{U}(p) = \int_0^1 U(\tau; p) \, d\tau$, also viewed as a constant trajectory, and the *orbit variance* $V(p) = \int_0^1 \|U(\tau; p) - \bar{U}(p)\|_X^2 \, d\tau$, which measures variation along one orbit rather than across parameter samples. The *population phase variance* is $\sigma_{\text{phase}}^2 = \int_{\mathcal{P}} V(p) \, d\mu(p)$.

Table 1: ODE waveform ablations over five training seeds. Checkmarks indicate period normalization and phase alignment. Experiments under 5 training seeds are included. SD: standard deviation.

Goodwin					
Norm.	Align.	Train mean (SD)	Val mean (SD)	Val min	Val max
IID split					
✓	✓	3.36×10^{-4} (9.98×10^{-7})	3.79×10^{-4} (3.54×10^{-6})	3.74×10^{-4}	3.83×10^{-4}
×	✓	3.03×10^{-5} (1.55×10^{-6})	5.48×10^{-5} (4.59×10^{-6})	4.94×10^{-5}	6.17×10^{-5}
×	×	7.74×10^{-4} (8.74×10^{-5})	9.97×10^{-1} (1.96×10^{-2})	9.76×10^{-1}	1.03
✓	×	1.03×10^{-3} (4.40×10^{-5})	9.70×10^{-1} (1.56×10^{-2})	9.53×10^{-1}	9.91×10^{-1}
OOD split					
✓	✓	2.99×10^{-4} (6.07×10^{-7})	1.04×10^{-2} (1.89×10^{-3})	8.14×10^{-3}	1.25×10^{-2}
×	✓	2.84×10^{-4} (2.90×10^{-6})	2.28×10^{-1} (2.22×10^{-2})	1.94×10^{-1}	2.54×10^{-1}
GOLD95					
Norm.	Align.	Train mean (SD)	Val mean (SD)	Val min	Val max
IID split					
✓	✓	4.27×10^{-4} (7.72×10^{-6})	7.19×10^{-4} (9.07×10^{-6})	7.10×10^{-4}	7.32×10^{-4}
×	✓	3.08×10^{-4} (6.48×10^{-6})	8.28×10^{-4} (3.29×10^{-5})	7.99×10^{-4}	8.80×10^{-4}
×	×	9.75×10^{-4} (1.12×10^{-4})	1.328 (3.23×10^{-2})	1.303	1.378
✓	×	1.48×10^{-3} (3.68×10^{-5})	1.192 (1.18×10^{-2})	1.175	1.209
OOD split					
✓	✓	3.82×10^{-4} (7.62×10^{-6})	1.63×10^{-2} (3.63×10^{-3})	1.34×10^{-2}	2.21×10^{-2}
×	✓	3.19×10^{-4} (1.46×10^{-5})	2.00×10^{-1} (1.22×10^{-2})	1.85×10^{-1}	2.19×10^{-1}

Assumption 1 (Predictor and Target Regularity)**.** *(i)* $\mathcal{P} \subset \mathbb{R}^m$ is compact, has nonempty interior, and has Lebesgue-null boundary. *(ii)* ω has a positive C^1 extension to an open neighborhood of $\mathcal{P}$. *(iii)* $U : \mathbb{T} \times \mathcal{P} \to X$ is continuous and satisfies $\sup_{\tau\in\mathbb{T}} \|U(\tau; p) - U(\tau; q)\|_X \le K_U \|p - q\|_2$ for some $K_U < \infty$. *(iv)* μ has a continuous density $d\mu(p) = \varpi(p)\, dp$ on $\mathcal{P}$. *(v)* $\mu(\{p : \nabla\omega(p) = 0\}) = 0$.

3.1 Population error

We quantify the error caused by phase variation, under standard regularity assumptions and a bound on how fast the predictor varies with the parameter.

Theorem 1 (Population error floor and decomposition)**.** *Under Assumption 1, suppose* $\beta(s) \neq 0$ *almost everywhere. Let the trajectory predictor* $g_\lambda : \mathcal{P} \to \mathcal{H}$ *be Lipschitz with respect to the parameter* p*, with constant* $L_\lambda = o(\lambda)$*. Set* $\mathrm{MSE}_\lambda = \|g_\lambda - F_\lambda\|^2_{\mu,2}$*, where we define* $\|v\|^2_{\mu,2} = \int_{\mathcal{P}} \|v(p)\|^2_{\mathcal{H}}\, d\mu(p)$*. Then*

$$\liminf_{\lambda\to\infty} \mathrm{MSE}_\lambda \ge \sigma^2_{\mathrm{phase}}. \tag{3}$$

If MSE_λ *remains bounded, then*

$$\mathrm{MSE}_\lambda = \sigma^2_{\mathrm{phase}} + \|g_\lambda - \bar{U}\|^2_{\mu,2} + o(1). \tag{4}$$

Proof sketch. Use frequency as a local parameter coordinate and divide it into cells on which the predictor and fixed-phase waveform vary little, while the target completes many phase turns. Phase averaging removes the correlation between the predictor and the centered waveform $U - \bar{U}$. Expanding the squared error then gives (4). Full details are given in Appendix A.

Interpretation and insights. For a fixed architecture, if the layerwise Lipschitz constants remain uniformly bounded as λ grows, then L_λ is bounded and hence $L_\lambda = o(\lambda)$ (Gouk et al., 2021). Our training does not enforce this bound. The theorem is therefore conditional, and the experimental results are consistent with its predicted error floor.

A predictor that changes slowly with the parameters cannot match targets whose phases vary rapidly. In FM1, after a long common time shift, the initial phase changes rapidly with the parameter. In FM2, over a wide observation window, even a small frequency difference can produce a large phase

difference by the end of the window. Quantitatively, the theorem bounds the population MSE below by σ^2_{phase}, the orbit variance weighted by the parameter distribution. The decomposition further explains why amplitude-death-type outputs can be favored: predicting the constant phase mean $\bar{U}$ eliminates $\|g_\lambda - \bar{U}\|^2_{\mu,2}$.

Applications. Applying Theorem 1 to FM1 and FM2 gives

$$\liminf_{c\to\infty} \|g_c^{(1)} - F_c^{(1)}\|^2_{\mu,2} \geq \sigma^2_{\text{phase}}, \quad \liminf_{W\to\infty} \|g_W^{(2)} - F_W^{(2)}\|^2_{\mu_{\text{OOD}},2} \geq \sigma^2_{\text{phase}}. \tag{5}$$

Here, for second inequality, μ_{OOD} is a parameter distribution independent of W, with frequencies outside the training range, and $\sigma^2_{\text{phase}} = \int_{\mathcal{P}} V(p)\, d\mu_{\text{OOD}}(p)$. Numerical validation is given later in Section 4.

3.2 Per-sample OOD error

The FM2 experiments show that OOD prediction can fail. This leaves us curious about two questions: How far must a new sample's frequency lie outside the training range before its error becomes substantial? Which quantities determine the size of that error?

Fortunately, we can bound the per-sample error below using the orbit variance, the frequency gap, and the observation-window width.

To quantitatively derive the bound, let $\Omega_{\text{train}} = [\omega_-, \omega_+]$ be the training frequency range, and fix one endpoint ω_b. Define

$$\Gamma = \{p \in \mathcal{P} : \omega(p) = \omega_b\}, \qquad d_\omega(p) = |\omega(p) - \omega_b|, \qquad e_W(p) = \|g_W^{(2)}(p) - F_W^{(2)}(p)\|_{\mathcal{H}}.$$

Assumption 2 (FM2 boundary regularity)**.** There exists $r_0 \in \Gamma \cap \operatorname{int}\mathcal{P}$ such that $\nabla\omega(r_0) \neq 0$ and $V(r_0) > 0$. Moreover, there exist an open neighborhood $G \subset \operatorname{int}\mathcal{P}$ of r_0 and a constant $K_\tau < \infty$ such that $\|U(\tau + v; p) - U(\tau; p)\|_X \leq K_\tau |v|$ for all $p \in G$, $\tau \in \mathbb{T}$, and $v \in \mathbb{R}$.

Theorem 2 (FM2: per-sample OOD lower bound)**.** *Under Assumption 1(i)–(iii) and Assumption 2, suppose $g_W^{(2)}$ has parameter Lipschitz constant $L_W = o(W)$ and $e_W \to 0$ uniformly on a fixed neighborhood of r_0 in Γ. Then there exists a neighborhood $\mathcal{N} \subset G$ of r_0 such that, for every fixed $B > 0$,*

$$\frac{e_W(p)^2}{V(p)} \geq \min\{z(p)^2, 1\} - o(1), \qquad z(p) = W d_\omega(p), \tag{6}$$

uniformly over OOD points $p \in \mathcal{N}$ satisfying $0 < z(p) \leq B$, as $W \to \infty$.

Proof sketch. Compare each OOD prediction with its value at a nearby training boundary point, and freeze the waveform at the test parameter. Periodic averaging bounds the resulting trajectory discrepancy below by $V(p)\min\{z(p)^2, 1\} - o(1)$. Appendix B gives the full proof.

Interpretation. Surprisingly, when the observation window is wide, a sample just outside the training range can already be predicted poorly, even if a well-trained predictor has small error at the boundary frequency. Here is a straightforward explanation: Changing slowly with the parameters, a predictor gives nearly the same waveforms for a parameter on the training boundary and for an OOD parameter close to it. But the two true trajectories have slightly different frequencies. Far enough into a wide window, their phase difference keeps changing at every time step and can reach up to a full cycle, which makes the predictor fail. Numerical validation is given in Section 4.

Applicability. For Goodwin and GOLD95, the phase Lipschitz condition holds on a compact parameter region because $\partial_\tau U(\tau; p) = T(p) f(U(\tau; p); p)$, where both the period and governing equation, f, in (1) are continuous. For the point r_0, it suffices that varying a single parameter, with the others fixed, reaches ω_b inside the box with a nonzero partial derivative. We verified this numerically for both systems. Experimental details in Section F.6.

4 Numerical validation

Theorem 1 predicts that the population MSE has a floor once the burn-in shift c or the window length W is large enough. Theorem 2 predicts that, for test points close enough to the boundary of the training set, the error on individual samples remains large. We test both predictions numerically.

4.1 FM1: SHIFT SWEEP

We train each shift configuration with five seeds and report the lowest validation error reached during training (*Best val MSE*) and the final validation error (*Final val MSE*). Figure 1 shows that both are well below σ^2_{phase} at small shifts, but reach or exceed it at larger tested shifts on both systems. Training configurations and detailed results are given in Appendices F.1 and F.2.

At large shifts, Best val MSE is reached within the first four epochs and closely matches σ^2_{phase}: its mean across seeds exceeds σ^2_{phase} by at most 0.6%, with standard deviation at most 0.18% of σ^2_{phase}. By our error decomposition, a near-floor error means that the predictor is close to the orbit's phase average, with little variation in phase. After the early validation minimum, training MSE continues to decrease while validation MSE rises. Final val MSE reaches 1.69–1.82 times the floor, showing clear overfitting during continued training.

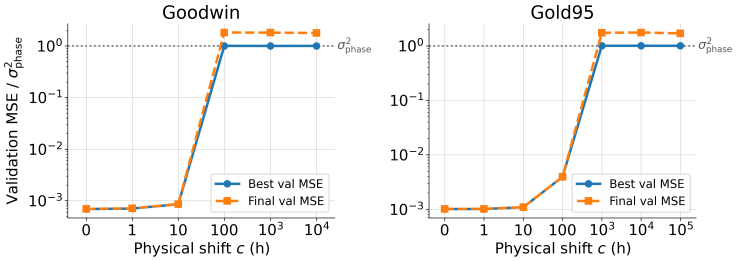


Figure 1: FM1 shift sweep. Validation MSE divided by σ^2_{phase} against the burn-in shift c. Markers show the mean over five seeds. The range across seeds is below 8% of the mean at every point except Gold95 at $c = 100$ h (14%), which lies in the transition. Dotted line: σ^2_{phase}.

4.2 FM2: WINDOW SWEEP

To test the predictions of Theorems 1 and 2 as the window length W grows, we sweep W on Goodwin and GOLD95 and train five seeds per W. Configuration details and per-W results are in Appendices F.3 and F.4.

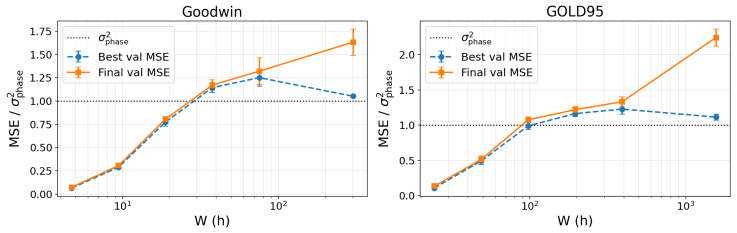


Figure 2: Population validation MSE normalized by σ^2_{phase} across the W sweep, for Goodwin (left) and GOLD95 (right). Error bars are one standard deviation over 5 seeds.

Population error. Theorem 1 predicts that, given a huge W, the population error stays above σ^2_{phase} for a predictor with limited parameter sensitivity. Figure 2 is consistent with this prediction. On both systems, the final validation MSE stays above σ^2_{phase}, given a large W, matching the FM1 results. The best validation MSE during training is closer to the floor but also stays above it, which suggests that failing to go below the floor is not due to insufficient training. Detailed W sweep results are given in Tables 9 and 10.

Per-sample OOD error. Theorem 2 gives the asymptotic lower bound $e_W(p)^2/V(p) \geq \min\{z^2, 1\} - o(1)$ near the training frequency boundary, where $z = W d_\omega(p)$ measures accumulated phase drift. We partition z into logarithmic bins and compute the mean and minimum of $e_W(p)^2/V(p)$ in each bin for each seed. Figure 3 shows a trend consistent with the asymptotic bound. At every tested W, the bin mean lies above $\min\{z^2, 1\}$. The bin minimum falls below this curve at

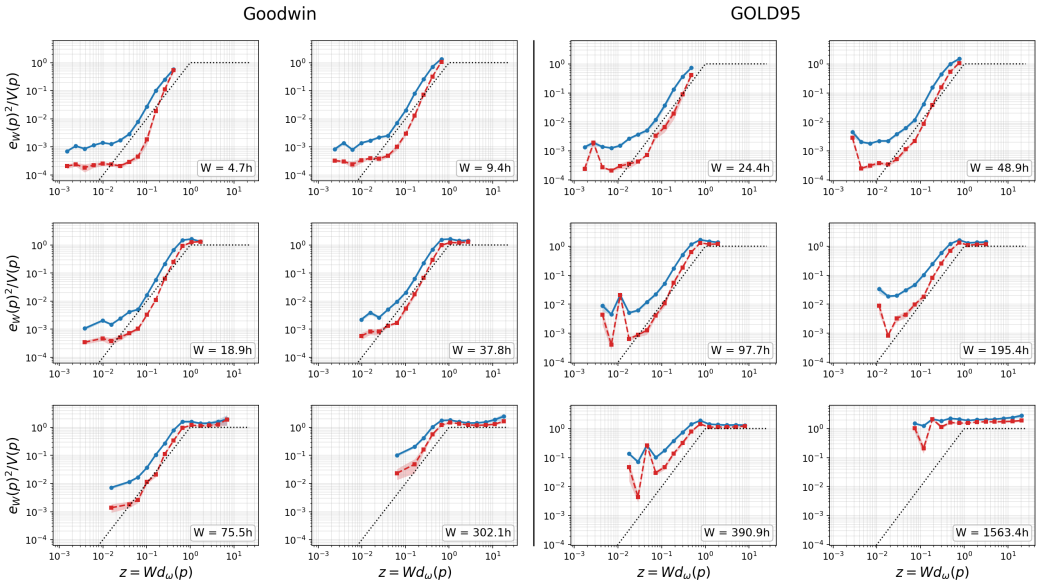


Figure 3: Per-sample OOD error for Goodwin (left) and GOLD95 (right). Blue: bin means; red: bin minima; shading: one standard deviation across seeds; dotted line: $\min\{z^2, 1\}$.

smaller windows, but lies above it for the large tested windows starting at $W = 302.1$ h for Goodwin and $W = 195.4$ h for GOLD95. The gap above the curve generally widens as W increases. This pattern recurs across seeds: the relative standard deviation across seeds is 2%–12% for the bin mean and 8%–22% for the bin minimum. The details of the bin partition and relative standard deviations are given in Appendix F.5.

5 Escaping the Error Floor: A Decoupled Target

Theorem 1 shows that a learning target (2) can develop an error floor. Our question is: how can we escape this floor? One immediate idea is to cap λ, or remove it from the phase term, so that $\lambda \to \infty$ no longer applies. The simplest case is to keep $\beta = 0$, so that no large-λ phase term enters the waveform representation, which is exactly the decoupled target we introduced in Section 2.

5.1 Decoupled Representation

Straightforwardly, the decoupled target learns the anchored waveform $U(s; p)$ and period $T(p)$ separately. In (2), its waveform component is exactly the choice $\alpha(s) = s$ and $\beta = 0$, so the large phase term is absent. For each ODE sample, we remove the transient, anchor phase at the unique maximum of a scalar observable a_{obs}, chosen as one state coordinate in our ODE experiments, and sample one cycle at fixed phases $\tau_1, \ldots, \tau_G$. With $U_G(p) = (U(\tau_1; p), \ldots, U(\tau_G; p))$, the two predictors are

$$\widehat{T}(p) = \mathcal{M}_\psi(p), \qquad \widehat{U}_G(p) = \mathcal{F}_\vartheta(p), \tag{7}$$

where $\mathcal{M}_\psi$ and $\mathcal{F}_\vartheta$ denote the period and waveform predictors. A periodic interpolant supplies values between phase samples. Neither target contains an uncontrolled observation phase.

5.2 ODE Regularity and Approximability

The following conditions ensure that the period and anchored waveform of (1) are regular; they are independent of the population assumptions used for the error floors.

Assumption 3 (Regular periodic setting)**.** The vector field satisfies $f \in C^2(\mathcal{D} \times \mathcal{O}; \mathbb{R}^n)$, and trajectories remain in $\mathcal{D}$. We assume $\mathcal{P}$ is compact. For every $p \in \mathcal{P}$, the ODE has exactly one limit cycle $\mathcal{C}(p) \subset \mathcal{D}$, which is attracting and hyperbolic with minimal period $T(p) > 0$, and no bifurcation,

loss of hyperbolicity, or branch switching occurs over $\mathcal{P}$. The observable $a_{\mathrm{obs}} \in C^2(\mathcal{D}; \mathbb{R})$ attains its maximum at exactly one phase per cycle, where its first time derivative vanishes and its second time derivative is strictly negative.

Theorem 3 (Approximability of decoupled ODE targets)**.** *Under Assumption 3, T and $U : \mathbb{T} \times \mathcal{P} \to \mathbb{R}^n$ are continuous. Hence U_G is continuous, and both T and U_G are uniformly approximable by feedforward networks with suitable continuous nonpolynomial activations.*

Proof sketch. Hyperbolicity continues the periodic orbit and its return time locally in p (Teschl, 2012, Lemma 6.9 and 12.7). The implicit function theorem continues the nondegenerate maximizing phase, while uniqueness makes the local anchors consistent. Thus T and the anchored waveform are continuous, and compactness gives the approximability (Hornik et al., 1989; Leshno et al., 1993). Appendix C gives the proof details.

5.3 How can we use decoupled targets?

Models for the two decoupled targets are straightforward to train and achieve small errors. For two ODE systems, the waveform errors are reported in Table 1, while period errors and training details are given in Appendix E.1. The remaining question is how to use these predictions to reconstruct a physical-time trajectory.

The learned period and waveform determine the periodic orbit but not the phase, so a single full-state snapshot y_{obs} suffices to recover a trajectory. An interpolant maps $\widehat{U}_G$ to $\widehat{U}$ on $[0, 1)$ between the sampled phases $\tau_1, \ldots, \tau_G$. We estimate $\widehat{\phi} \in \arg\min_{\phi \in [0,1)} \|\widehat{U}(\phi; p) - y_{\mathrm{obs}}\|_X$ and reconstruct $\widehat{x}(t) = \widehat{U}(\widehat{\phi} + t/\widehat{T}(p); p)$. In the exact setting, the phase is unique because a nonconstant minimal-period ODE orbit cannot visit the same state at two distinct phases. In practice, we use a coarse-to-fine search on the interpolated predicted waveform.

As shown in Table 2, a single snapshot is enough to reconstruct accurate trajectories on both systems across five training seeds, with a mean period error below 0.5%. The waveform MSE is computed after scaling each coordinate of the state to unit variance over the training set. For example, the mean MSE of $1.3{\times}10^{-4}$ on Goodwin means the prediction is typically off by about 1% of each coordinate's standard deviation, and the $7.6{\times}10^{-4}$ on Gold95 by about 3%. These errors measure tasks different from ablation studies and are not directly comparable with the validation MSE in Table 1. The search algorithm, metrics, and per-seed results are provided in Appendix G.

Table 2: Snapshot-based validation. Entries are mean $\pm$ standard deviation over five training seeds. Waveform MSE is measured in the encoded state space, where each state component has unit variance over the training set.

System	Statistic	Waveform MSE	Period error (%)
Goodwin	Mean	$(1.25 \pm 0.20){\times}10^{-4}$	0.099 ± 0.002
	Median	$(7.33 \pm 1.26){\times}10^{-5}$	0.071 ± 0.003
	95th	$(4.02 \pm 0.73){\times}10^{-4}$	0.255 ± 0.015
Gold95	Mean	$(7.61 \pm 0.37){\times}10^{-4}$	0.426 ± 0.017
	Median	$(2.04 \pm 0.12){\times}10^{-4}$	0.313 ± 0.016
	95th	$(4.17 \pm 0.35){\times}10^{-3}$	1.193 ± 0.069

6 Beyond ODEs: a FitzHugh–Nagumo PDE Case Study

We next examine the two failure modes in a more challenging setting: temporally periodic solutions of a partial differential equation (PDE). Taking $X = L^2(\mathcal{D}_x; \mathbb{R}^n)$ extends the FM1 and FM2 results directly to periodic PDE fields satisfying the assumptions of Section 3. We test this setting on a winding-one traveling-wave branch of the one-dimensional FitzHugh–Nagumo system adapted from Cebrián-Lacasa et al. (2024), with periodic boundary conditions. We align the waves by starting each one-period record at the time when the first spatial Fourier coefficient of u has zero phase. Further details are given in Appendix H.

Table 3: FitzHugh–Nagumo PDE representation ablations.

	Norm.	Align	Split	Train MSE	Val MSE
Main	✓	✓	IID	9.65×10^{-7}	9.25×10^{-7}
1	×	✓	IID	7.74×10^{-6}	7.61×10^{-6}
2	✓	×	IID	9.80×10^{-1}	1.00
3	×	×	IID	9.85×10^{-1}	1.00
4	✓	✓	OOD	3.07×10^{-6}	1.18×10^{-2}
5	×	✓	OOD	5.63×10^{-6}	2.66×10^{-1}

The results match the qualitative behavior of the ODE ablations in Table 1: both failure modes persist for spatial fields, while the decoupled method accurately learns the periodic solution. Extending the continuity theory for the decoupled representation to PDEs requires a separate treatment of traveling-wave branches and translation symmetry, which we leave for future work.

7 Discussion and Limitations

Regularity of the target maps. The error-floor analysis needs no explicit governing equation: it applies to any parameterized periodic function whose waveform and period maps satisfy the stated regularity, provided frequency varies with the parameters and the predictor's parameter Lipschitz constant grows more slowly than the time scale. The ODE result gives sufficient conditions for continuity of the decoupled targets, which can fail at bifurcations (Kuznetsov, 2023, Chapters 5). Analogous regularity for PDE systems remains future work.

OOD prediction remains difficult. Decoupling reduces OOD error in our experiments but does not eliminate it; waveform errors stay larger than their iid counterparts. Moreover, parameter extrapolation remains a general difficulty in operator learning (Zhu et al., 2023).

Structure-specific decoupled target. Our method assumes the period is determined only by the parameter. More complex periodic structure will need more than this simple escape. In atrial fibrillation, for example, period length varies by region of the heart (Skanes et al., 1998), so the period is no longer a scalar. Such settings will need decoupled targets built for their specific structure.

Operator learning beyond trajectory targets. The operator to be learned need not map parameters directly to trajectories on a domain: other targets include steady states (Marwah et al., 2023; Zhang et al., 2023), Green's functions for PDE parameters (Melchers et al., 2026), and Newton solvers for nonlinear PDEs (Hao et al., 2024). Future work could design operator-learning targets for a wider range of physical quantities.

8 Conclusion

The main lesson of this paper is that, for periodic systems, the error can be determined mainly by the learning target rather than by training. A periodic orbit can depend regularly on the parameters while the trajectory recorded from it does not. In this case, learning the target requires a predictor whose sensitivity to the parameters grows comparably with the time scale. If the sensitivity grows more slowly, the error floor remains. This large time scale comes from how the data are generated, through long transient removal or a wide observation window. Such choices are usually treated as preprocessing details. Our results suggest they deserve the same scrutiny as the model, and that errors reported on periodic benchmarks may partly reflect these choices.

More broadly, rather than learning the full trajectory, we can learn targets chosen from what the dynamics tell us about the system. For a periodic orbit, the parameters determine only the period and the waveform. The initial phase is a choice of when observation starts. Learning period and waveform separately, therefore, is a natural choice. Other systems have other structure or invariants and will need their own targets. A general principle for choosing such targets across dynamical systems remains an open question.

AI USAGE STATEMENT

In this work, we used generative AI tools to formulate mathematical claims, provide critical ingredients for proving mathematical claims, assist in the writing of proofs, implement methods, and clean and reformat data sets. We have not used generative AI tools to generate synthetic data sets, help develop theoretical models or conceptual frameworks, propose or refine hypotheses, design or provide feedback on research methodology or experiments, assist with translation, or interpret results. Qualitative and thematic data analysis is not applicable to this work. Additionally, we used generative AI tools to identify relevant literature, edit the paper to improve readability, and propose a title and keywords.

Required disclosure details. *Formulate mathematical claims:* LLMs suggested ways to weaken some technical assumptions, which the authors checked and adopted. *Provide critical ingredients for proving mathematical claims and assist in the writing of proofs:* The authors designed the local to global proof strategy; LLMs provided substantial help in deriving the local approximation results and were used to make notation consistent across the proofs. *Implement methods and clean and reformat data sets:* LLMs assisted in writing experiment scripts, including the data generation and data processing scripts. *Generate synthetic data sets:* The authors ran numerical ODE and PDE solvers using AI-assisted scripts; no data were generated directly by an LLM. *Theoretical framework, hypotheses, experimental design, and interpretation:* The formalization of the research problem, the common mechanism of the two failure modes, the research question (inspired by the authors' own previous work on periodic systems), the design of all experiments, and the interpretation of all results are the authors' own.

We have reviewed all AI assisted work. All proofs were verified line by line by the authors, and all LLM assisted code was tested. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

All assumptions are stated in the corresponding sections; complete proofs of Theorem 1, Theorem 2, and Theorem 3 appear in Appendices A, B, and C. The ODE systems, Sobol sampling box, and the integration, period-detection, and acceptance procedure are given in Appendix D; the FitzHugh–Nagumo PDE setup is in Appendix H. All ablation experiment details are in Appendix E. The snapshot-conditioned reconstruction details, including the phase search algorithm, its evaluation metric, and per-seed results, are in Appendix G. The FM1 shift sweep and FM2 window sweep architecture and protocol are specified in Appendix F. All data-generation, training, and analysis code is available at `https://anonymous.4open.science/r/period_failure-7DC5`.

REFERENCES

Andrea Cassani, Alessandro Monteverde, and Marco Piumetti. Belousov-zhabotinsky type reactions: The non-linear behavior of chemical systems. *Journal of mathematical chemistry*, 59(3):792–826, 2021.

Daniel Cebrián-Lacasa, Pedro Parra-Rivas, Daniel Ruiz-Reynés, and Lendert Gelens. Six decades of the fitzhugh–nagumo model: A guide through its spatio-temporal dynamics and influence across disciplines. *Physics Reports*, 1096:1–39, 2024.

Yuyang Chai, Wei Gao, Benjamin Ankay, Fengming Li, and Chuanzeng Zhang. Aeroelastic analysis and flutter control of wings and panels: A review. *International Journal of Mechanical System Dynamics*, 1(1):5–34, 2021.

Albert Goldbeter. A model for circadian oscillations in the drosophila period protein (per). *Proceedings of the Royal Society B: Biological Sciences*, 261(1362):319–324, 09 1995.

Didier Gonze and Peter Ruoff. The goodwin oscillator and its legacy. *Acta Biotheoretica*, 69(4): 857–874, 2021.

Henry Gouk, Eibe Frank, Bernhard Pfahringer, and Michael J Cree. Regularisation of neural networks by enforcing lipschitz continuity. *Machine Learning*, 110(2):393–416, 2021.

Wenrui Hao, Xinliang Liu, and Yahong Yang. Newton informed neural operator for solving nonlinear partial differential equations. *Advances in neural information processing systems*, 37:120832–120860, 2024.

Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5):359–366, 1989.

Gyuyoung Hwang, Hyeontae Jo, and Jae Kyoung Kim. Data-driven inference of coupling functions in oscillatory systems via physics-informed neural networks. *Physical Review Research*, 8(1): 013259, 2026.

Matthew P Juniper and Raman I Sujith. Sensitivity and nonlinearity of thermoacoustic oscillations. *Annual Review of Fluid Mechanics*, 50:661–689, 2018.

George Em Karniadakis, Ioannis G Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440, 2021.

Nikola B Kovachki, Samuel Lanthaler, and Andrew M Stuart. Operator learning: Algorithms and analysis. *Handbook of Numerical Analysis*, 25:419–467, 2024.

Yuri A. Kuznetsov. *Bifurcations of Equilibria and Periodic Orbits in n-Dimensional Dynamical Systems*, pp. 175–228. Springer International Publishing, Cham, 2023. ISBN 978-3-031-22007-4. doi: 10.1007/978-3-031-22007-4_5.

Olivier P Le Maître, Lionel Mathelin, Omar M Knio, and M Yousuff Hussaini. Asynchronous time integration for polynomial chaos expansion of uncertain periodic dynamics. *Discrete and Continuous Dynamical Systems*, 28(1):199–226, 2010.

Moshe Leshno, Vladimir Ya Lin, Allan Pinkus, and Shimon Schocken. Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. *Neural networks*, 6(6):861–867, 1993.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. *arXiv preprint arXiv:2010.08895*, 2020.

Chu V Mai and Bruno Sudret. Surrogate models for oscillatory systems using sparse polynomial chaos expansions and stochastic time warping. *SIAM/ASA Journal on Uncertainty Quantification*, 5(1):540–571, 2017.

J. S. Marron, James O. Ramsay, Laura M. Sangalli, and Anuj Srivastava. Functional data analysis of amplitude and phase variation. *Statistical Science*, 30(4):468–484, 2015.

Tanya Marwah, Ashwini Pokle, J Zico Kolter, Zachary Lipton, Jianfeng Lu, and Andrej Risteski. Deep equilibrium based neural operators for steady-state pdes. *Advances in Neural Information Processing Systems*, 36:15716–15737, 2023.

HA Melchers, JHM Prins, and Michael RA Abdelmalik. Neural green's operators for parametric partial differential equations. *Computer Methods in Applied Mechanics and Engineering*, 455: 118893, 2026.

James O Ramsay and Xiaochun Li. Curve registration. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 60(2):351–363, 1998.

Allan C. Skanes, Ravi Mandapati, Omer Berenfeld, Jorge M. Davidenko, and José Jalife. Spatiotemporal periodicity during atrial fibrillation in the isolated sheep heart. *Circulation*, 98(12): 1236–1248, 1998. doi: 10.1161/01.CIR.98.12.1236. URL https://www.ahajournals.org/doi/abs/10.1161/01.CIR.98.12.1236.

Shohei Takata, Yuzuru Kato, and Hiroya Nakao. Definition and data-driven reconstruction of asymptotic phase and amplitudes of stochastic oscillators via koopman operator theory. In *IUTAM symposium on Nonlinear dynamics for design of mechanical systems across different length/time scales*, pp. 141–153. Springer, 2023.

Gerald Teschl. *Ordinary differential equations and dynamical systems*, volume 140. American Mathematical Soc., 2012.

J Derek Tucker, Wei Wu, and Anuj Srivastava. Generative models for functional data using phase and amplitude separation. *Computational Statistics & Data Analysis*, 61:50–66, 2013.

Martha Hotz Vitaterna, Joseph S Takahashi, and Fred W Turek. Overview of circadian rhythms. *Alcohol research & health*, 25(2):85, 2001.

Min Wang, Hua Wang, and Fan Zhang. Aligntime: Interperiodic phase alignment sampling for time-series forecasting. *Information Processing & Management*, 63(1):104296, 2026.

Dan Wilson. Data-driven model identification using forcing-induced limit cycles. *Physica D: Nonlinear Phenomena*, 459:134013, 2024.

Jeroen AS Witteveen and Hester Bijl. An alternative unsteady adaptive stochastic finite elements formulation based on interpolation at constant phase. *Computer Methods in Applied Mechanics and Engineering*, 198(3-4):578–591, 2008.

Yimeng Zhang, Alexander Cloninger, Bo Li, and Xiaochuan Tian. A neural network kernel decomposition for learning multiple steady states in parameterized dynamical systems. *arXiv preprint arXiv:2312.10315*, 2023.

Min Zhu, Handi Zhang, Anran Jiao, George Em Karniadakis, and Lu Lu. Reliable extrapolation of deep neural operators informed by physics or sparse observations. *Computer Methods in Applied Mechanics and Engineering*, 412:116064, 2023.

A Unified Phase Averaging

A.1 Setup and lemmas

Recall that $\mathcal{H} = L^2([0,1]; X)$, and define

$$\bar{U}(p) = \int_0^1 U(\tau; p)\, d\tau, \qquad V(p) = \int_0^1 \|U(\tau; p) - \bar{U}(p)\|_X^2\, d\tau.$$

The population phase variance is

$$\sigma_{\text{phase}}^2 = \int_{\mathcal{P}} V(p)\, d\mu(p).$$

We first give an elementary periodic-averaging estimate.

Lemma A.1 (Average of a zero-mean periodic function)**.** *Let Z be a Banach space and let $g : \mathbb{T} \to Z$ be continuous and satisfy*

$$\int_0^1 g(\tau)\, d\tau = 0.$$

Then, for every $\theta \in \mathbb{T}$, $q > 0$, and $\xi \in \mathbb{R}$,

$$\left\| \int_0^1 g(\theta + qz\xi)\, dz \right\|_Z \le \sup_{\tau \in \mathbb{T}} \|g(\tau)\|_Z \min\{1, (q|\xi|)^{-1}\},$$

where the second term in the minimum is interpreted as $+\infty$ when $\xi = 0$.

Proof. Set $G = \sup_{\tau \in \mathbb{T}} \|g(\tau)\|_Z$ and $L = q|\xi|$. The integral is bounded by G, which proves the claim when $L < 1$. If $L \ge 1$, write $L = n + \rho$, where n is an integer and $0 \le \rho < 1$. Changing variables and canceling the n complete periods gives

$$\left\| \int_0^1 g(\theta + qz\xi)\, dz \right\|_Z = \frac{1}{L} \left\| \int_{\theta_0 + n}^{\theta_0 + n + \rho} g(t)\, dt \right\|_Z \le \frac{\rho G}{L} \le \frac{G}{L},$$

where $\theta_0 = \theta$ if $\xi > 0$, and $\theta_0 = \theta - L$ if $\xi < 0$. □

The following direct consequence gives the two estimates used below.

Corollary A.1 (Uniform phase-averaging bounds)**.** *Under the assumptions of Theorem 1, let*

$$v(\tau; p) = U(\tau; p) - \bar{U}(p), \qquad M = \sup_{\tau, p} \|U(\tau; p)\|_X.$$

For $q > 0$, define

$$\eta_q(s) = \begin{cases} \min\{1, (q|\beta(s)|)^{-1}\}, & \beta(s) \neq 0, \\ 1, & \beta(s) = 0, \end{cases} \qquad J_\beta(q) = \int_0^1 \eta_q(s)\, ds.$$

Then $J_\beta(q) \to 0$ as $q \to \infty$. Moreover, for every $p \in \mathcal{P}$ and every measurable $\theta : [0,1] \to \mathbb{T}$,

$$\left\| \int_0^1 v\big(\theta(\cdot) + qz\beta(\cdot); p\big)\, dz \right\|_{\mathcal{H}} \le 2M J_\beta(q)^{1/2}, \tag{8}$$

$$\left| \int_0^1 \int_0^1 \left[\|v(\theta(s) + qz\beta(s); p)\|_X^2 - V(p) \right] dz\, ds \right| \le 4M^2 J_\beta(q). \tag{9}$$

Proof. Since $\beta \neq 0$ almost everywhere, $\eta_q(s) \to 0$ for almost every s. Since $0 \le \eta_q \le 1$, dominated convergence gives $J_\beta(q) \to 0$.

For fixed p, both $v(\cdot; p)$ and $\|v(\cdot; p)\|_X^2 - V(p)$ have zero mean, and their norms are bounded by $2M$ and $4M^2$, respectively. Apply Lemma A.1 to each function for fixed s. The resulting pointwise bounds are $2M\eta_q(s)$ and $4M^2\eta_q(s)$. Squaring the first and using $\eta_q^2 \le \eta_q$ proves (8); integrating the second proves (9). □

We now use frequency as a local parameter coordinate to apply these bounds over the parameter domain. We first identify the limiting variance and then prove that the oscillatory part is asymptotically orthogonal to any sufficiently slow predictor.

Lemma A.2 (Convergence to the phase variance)**.** *Under the assumptions of Theorem 1, let* $b_\lambda = F_\lambda - \bar{U}$*. Then*

$$\|b_\lambda\|_{\mu,2}^2 \longrightarrow \sigma_{\text{phase}}^2 = \int_{\mathcal{P}} V(p)\, d\mu(p).$$

Proof. Let $v = U - \bar{U}$ and recall that $M = \sup_{\tau,p}\|U(\tau;p)\|_X$, so $\|v\|_X \le 2M$ and $b_\lambda(p)(s) = v\big(\alpha(s) + \lambda\omega(p)\beta(s); p\big)$. By Assumption 1(iii), v is Lipschitz in p with constant $L_v = 2K_U$. Define

$$f_\lambda(p) = \|b_\lambda(p)\|_{\mathcal{H}}^2 - V(p).$$

Then $|f_\lambda| \le 8M^2$ and $\int_{\mathcal{P}} f_\lambda\, d\mu = \|b_\lambda\|_{\mu,2}^2 - \sigma_{\text{phase}}^2$, so it suffices to show that this integral tends to zero. We first derive local result then extend it globally.

Local coordinates. The set $\mathcal{R} = \{p \in \operatorname{int}\mathcal{P} : \nabla\omega(p) \neq 0,\ \varpi(p) > 0\}$ has full μ-measure. Near each point of $\mathcal{R}$, the inverse function theorem gives (after relabeling coordinates if necessary) $p = \Psi(y, r) \in C^1$ with $y = (p_1, \dots, p_{m-1})$ and $r = \omega(p)$. Fix a compact box $Q = B \times [a, b]$ under new coordinate, let L_Ψ be a Lipschitz constant of Ψ in r on Q, and let $0 \le \chi \le 1$ be continuous and supported in $\Psi(\operatorname{int} Q)$. With $\kappa(y, r) = \chi(\Psi(y, r))\, \varpi(\Psi(y, r))\, |\det D\Psi(y, r)|$ and $\widetilde{f}_\lambda = f_\lambda \circ \Psi$,

$$\int_{\mathcal{P}} \chi f_\lambda\, d\mu = \int_Q \kappa\, \widetilde{f}_\lambda\, dy\, dr.$$

Local limit. Partition $[a, b]$ into equal cells $I_j = [r_j, r_j + h]$ with $h = \Theta(\lambda^{-1/2})$, so that $h \to 0$ and $\lambda h \to \infty$ as $\lambda \to \infty$. On I_j, freeze the parameter but not the phase:

$$B_{\lambda,j}(y, r)(s) = v\big(\alpha(s) + \lambda r\beta(s); p_j(y)\big), \qquad \phi_{\lambda,j} = \|B_{\lambda,j}\|_{\mathcal{H}}^2 - V(p_j), \qquad p_j(y) = \Psi(y, r_j).$$

Writing $\kappa\widetilde{f}_\lambda = \kappa(\widetilde{f}_\lambda - \phi_{\lambda,j}) + (\kappa - \kappa(\cdot, r_j))\phi_{\lambda,j} + \kappa(\cdot, r_j)\phi_{\lambda,j}$ on each cell and summing over cells gives $\int_{\mathcal{P}} \chi f_\lambda\, d\mu = T_{\lambda,1} + T_{\lambda,2} + T_{\lambda,3}$, where

$$T_{\lambda,1} = \sum_j \int_B \int_{I_j} \kappa(y, r)\big(\widetilde{f}_\lambda - \phi_{\lambda,j}\big)(y, r)\, dr\, dy,$$

$$T_{\lambda,2} = \sum_j \int_B \int_{I_j} \big(\kappa(y, r) - \kappa(y, r_j)\big)\phi_{\lambda,j}(y, r)\, dr\, dy,$$

$$T_{\lambda,3} = \sum_j \int_B \kappa(y, r_j) \int_{I_j} \phi_{\lambda,j}(y, r)\, dr\, dy.$$

We show that each term tends to zero. Below, $|Q| = \int_Q dy\, dr$.

For $T_{\lambda,1}$, let $r \in I_j$. Apply $\big|\|e\|^2 - \|e'\|^2\big| \le 4M\|e - e'\|$, valid for $\|e\|, \|e'\| \le 2M$, to both terms of f_λ, and use $\|\Psi(y, r) - p_j(y)\| \le L_\Psi h$. This gives $|\widetilde{f}_\lambda(y, r) - \phi_{\lambda,j}(y, r)| \le 8ML_vL_\Psi h$, so

$$|T_{\lambda,1}| \le 8ML_vL_\Psi h \sup_Q \kappa\, |Q| \longrightarrow 0, \qquad h \to 0.$$

For $T_{\lambda,2}$, let $r \in I_j$. Then $|\kappa(y, r) - \kappa(y, r_j)| \le \omega_\kappa(h)$, where $\omega_\kappa(h) = \sup\{|\kappa(y, r) - \kappa(y, r')| : (y, r), (y, r') \in Q,\ |r - r'| \le h\}$. Since $|\phi_{\lambda,j}| \le 8M^2$,

$$|T_{\lambda,2}| \le 8M^2\, \omega_\kappa(h)\, |Q| \longrightarrow 0,$$

since κ is uniformly continuous on Q and $h \to 0$.

For $T_{\lambda,3}$, substitute $r = r_j + hz$ and apply (9) with $\theta = \alpha + \lambda r_j\beta$ and $q = \lambda h$. This gives $\big|\int_{I_j} \phi_{\lambda,j}(y, r)\, dr\big| \le 4M^2 h J_\beta(\lambda h)$, uniformly in y and j. Summing over the $(b - a)/h$ cells,

$$|T_{\lambda,3}| \le 4M^2 J_\beta(\lambda h) \sup_Q \kappa\, |Q| \longrightarrow 0,$$

since $\lambda h \to \infty$ and $J_\beta(q) \to 0$ as $q \to \infty$ by Corollary A.1. Hence $\int_{\mathcal{P}} \chi f_\lambda \, d\mu \to 0$.

Global results. For $\varepsilon > 0$, choose a compact $K \subset \mathcal{R}$ with $\mu(\mathcal{P} \setminus K) < \varepsilon$, a finite cover of K by the coordinates above, and continuous cutoffs χ_i supported in them with $S = \sum_i \chi_i = 1$ on K and $0 \le S \le 1$. The local limit gives $\int_{\mathcal{P}} S f_\lambda \, d\mu \to 0$, and $\left|\int_{\mathcal{P}} (1 - S) f_\lambda \, d\mu\right| \le 8M^2\varepsilon$. Hence $\limsup_{\lambda\to\infty} \left|\int_{\mathcal{P}} f_\lambda \, d\mu\right| \le 8M^2\varepsilon$ for every $\varepsilon > 0$, which proves the claim. □

Lemma A.3 (Asymptotically vanishing product)**.** *Under the assumptions of Theorem 1, let* $b_\lambda = F_\lambda - \bar{U}$. *Suppose that* $\sup_\lambda \|a_\lambda\|_{\mu,2} < \infty$ *and*

$$\|a_\lambda(p) - a_\lambda(p')\|_{\mathcal{H}} \le M_\lambda |p - p'| \quad (p, p' \in \mathcal{P}), \qquad M_\lambda = o(\lambda).$$

Then

$$\langle a_\lambda, b_\lambda \rangle_{L^2(\mu;\mathcal{H})} \longrightarrow 0.$$

Proof. Write $v(\tau; p) = U(\tau; p) - \bar{U}(p)$, and recall that $M = \sup_{\tau,p} \|U(\tau; p)\|_X$. Then $\|v(\tau; p)\|_X \le 2M$ and $b_\lambda(p)(s) = v(\alpha(s) + \lambda\omega(p)\beta(s); p)$.

Local coordinate. The regular set $\mathcal{R} = \{p \in \operatorname{int} \mathcal{P} : \nabla\omega(p) \ne 0,\ \varpi(p) > 0\}$ is open and has full μ-measure. For each point in $p^* \in \mathcal{R}$, choose a neighborhood contained in $\mathcal{R}$. After relabeling coordinates if necessary, the inverse function theorem gives local coordinates $y = (p_1, \dots, p_{m-1})$, $r = \omega(p)$, and $p = \Psi(y, r) \in C^1$.

Choose a compact box $Q = B \times [a, b]$ within this coordinate neighbor such that $p_* \in \Psi(\operatorname{int} Q)$ and $\Psi(Q) \subset \mathcal{R}$, and let $0 \le \chi \le 1$ be continuous and supported in $\Psi(\operatorname{int} Q)$. Set

$$\begin{aligned} w(y, r) &= \varpi(\Psi(y, r)) |\det D\Psi(y, r)|, & \kappa(y, r) &= \chi(\Psi(y, r)) w(y, r), \\ \widetilde{a}_\lambda(y, r) &= a_\lambda(\Psi(y, r)), & \widetilde{b}_\lambda(y, r) &= b_\lambda(\Psi(y, r)). \end{aligned}$$

Then $d\mu = w \, dy \, dr$, with $0 < \inf_Q w \le \sup_Q w < \infty$, and κ is bounded and uniformly continuous on Q. Let L_Ψ be a Lipschitz constant for Ψ in r, and let L_v be a uniform parameter Lipschitz constant for v.

Partition $[a, b]$ into equal cells $I_j = [r_j, r_j + h]$, choosing $h = \Theta([\lambda(1 + M_\lambda)]^{-1/2})$. Since $M_\lambda = o(\lambda)$, we have $h \to 0$, $M_\lambda h \to 0$, and $\lambda h \to \infty$. Define

$$\begin{aligned} a_j(y) &= \frac{1}{h} \int_{I_j} \widetilde{a}_\lambda(y, r) \, dr, p_j(y) = \Psi(y, r_j), \\ B_{\lambda,j}(y, r)(s) &= v\big(\alpha(s) + \lambda r \beta(s); p_j(y)\big). \end{aligned}$$

Local cross term. The cross term weighted by χ is

$$\begin{aligned} &\int_{\mathcal{P}} \chi(p) \langle a_\lambda(p), b_\lambda(p) \rangle_{\mathcal{H}} \, d\mu(p) \\ &= \sum_j \int_B \int_{I_j} \kappa(y, r) \langle \widetilde{a}_\lambda(y, r), \widetilde{b}_\lambda(y, r) \rangle_{\mathcal{H}} \, dr \, dy. \end{aligned} \tag{10}$$

On each cell, subtract the frozen integrand:

$$\begin{aligned} &\kappa(y, r) \langle \widetilde{a}_\lambda, \widetilde{b}_\lambda \rangle_{\mathcal{H}} - \kappa(y, r_j) \langle a_j, B_{\lambda,j} \rangle_{\mathcal{H}} \\ &= \kappa(y, r) \langle \widetilde{a}_\lambda - a_j, \widetilde{b}_\lambda \rangle_{\mathcal{H}} + \kappa(y, r) \langle a_j, \widetilde{b}_\lambda - B_{\lambda,j} \rangle_{\mathcal{H}} \\ &\quad + \big(\kappa(y, r) - \kappa(y, r_j)\big) \langle a_j, B_{\lambda,j} \rangle_{\mathcal{H}}. \end{aligned} \tag{11}$$

Denote the integrals of these three terms over all cells by $R_{\lambda,1}$, $R_{\lambda,2}$, and $R_{\lambda,3}$, respectively.

First residual. The definition of a_j and the Lipschitz bounds give, for $r \in I_j$,

$$\begin{aligned}
&\|\widetilde{a}_\lambda(y,r) - a_j(y)\|_{\mathcal{H}} = \left\| \frac{1}{h} \int_{I_j} \left[\widetilde{a}_\lambda(y,r) - \widetilde{a}_\lambda(y,r')\right] dr' \right\|_{\mathcal{H}} \\
\leq &\frac{1}{h} \int_{I_j} \|\widetilde{a}_\lambda(y,r) - \widetilde{a}_\lambda(y,r')\|_{\mathcal{H}}\, dr' = \frac{1}{h} \int_{I_j} \|a_\lambda(\Psi(y,r)) - a_\lambda(\Psi(y,r'))\|_{\mathcal{H}}\, dr' \\
\leq &\frac{M_\lambda}{h} \int_{I_j} |\Psi(y,r) - \Psi(y,r')|\, dr' \leq \frac{M_\lambda L_\Psi}{h} \int_{I_j} |r - r'|\, dr' \\
\leq &\frac{M_\lambda L_\Psi}{h}\, h^2 = L_\Psi M_\lambda h.
\end{aligned} \tag{12}$$

Using (12) and $\|\widetilde{b}_\lambda\|_{\mathcal{H}} \leq 2M$, we obtain

$$|R_{\lambda,1}| \leq 2M \sup_Q |\kappa|\, L_\Psi M_\lambda h\, |Q| \longrightarrow 0. \tag{13}$$

Second residual. The parameter Lipschitz bound on v gives, for $r \in I_j$,

$$\|\widetilde{b}_\lambda(y,r) - B_{\lambda,j}(y,r)\|_{\mathcal{H}} \leq L_v |\Psi(y,r) - \Psi(y,r_j)| \leq L_v L_\Psi h. \tag{14}$$

To bound the second residual, we need to control the integral of $\|a_j(y)\|_{\mathcal{H}}$ over all parameter cells. Here the $\mathcal{H}$-norm is taken in s, while the outer integrals are taken in y, r. Since $a_j(y)$ does not depend on r,

$$\sum_j \int_B \int_{I_j} \|a_j(y)\|_{\mathcal{H}}\, dr\, dy = \sum_j h \int_B \|a_j(y)\|_{\mathcal{H}}\, dy.$$

By the definition of a_j and the triangle inequality,

$$h\|a_j(y)\|_{\mathcal{H}} = \left\| \int_{I_j} \widetilde{a}_\lambda(y,r)\, dr \right\|_{\mathcal{H}} \leq \int_{I_j} \|\widetilde{a}_\lambda(y,r)\|_{\mathcal{H}}\, dr.$$

Thus, summing over the cells and using Cauchy–Schwarz,

$$\begin{aligned}
\sum_j h \int_B \|a_j(y)\|_{\mathcal{H}}\, dy &\leq \sum_j \int_B \int_{I_j} \|\widetilde{a}_\lambda(y,r)\|_{\mathcal{H}}\, dr\, dy \\
&= \int_Q \|\widetilde{a}_\lambda(y,r)\|_{\mathcal{H}}\, dy\, dr \\
&\leq |Q|^{1/2} \left(\int_Q \frac{\|\widetilde{a}_\lambda(y,r)\|_{\mathcal{H}}^2}{w(y,r)} w(y,r)\, dy\, dr \right)^{1/2} \\
&\leq \left(\frac{|Q|}{\inf_Q w} \right)^{1/2} \sup_\lambda \|a_\lambda\|_{\mu,2} < \infty.
\end{aligned} \tag{15}$$

Combining (14) and (15),

$$|R_{\lambda,2}| \leq \sup_Q |\kappa|\, L_v L_\Psi h \sum_j h \int_B \|a_j(y)\|_{\mathcal{H}}\, dy \longrightarrow 0. \tag{16}$$

Third residual. Uniform continuity of κ implies that $|\kappa(y,r) - \kappa(y,r_j)| \to 0$ with a bound independent of y, j for $r \in I_j$. Since $\|B_{\lambda,j}\|_{\mathcal{H}} \leq 2M$,

$$|R_{\lambda,3}| \leq 2M \sup_{\substack{j,\ y \in B \\ r \in I_j}} |\kappa(y,r) - \kappa(y,r_j)| \sum_j h \int_B \|a_j(y)\|_{\mathcal{H}}\, dy \longrightarrow 0 \tag{17}$$

by (15).

Frozen term. Substituting $r = r_j + hz$ and applying Corollary A.1 with $q = \lambda h$ and $\theta(s) = \alpha(s) + \lambda r_j \beta(s)$ modulo one gives

$$\left\| \int_{I_j} B_{\lambda,j}(y,r)\, dr \right\|_{\mathcal{H}} \le 2Mh J_\beta(\lambda h)^{1/2} = o(h). \tag{18}$$

The bound is independent of y, j. Therefore,

$$\begin{aligned} &\left| \sum_j \int_B \kappa(y,r_j) \left\langle a_j(y), \int_{I_j} B_{\lambda,j}(y,r)\, dr \right\rangle_{\mathcal{H}} dy \right| \\ &\le 2M \sup_Q |\kappa|\, J_\beta(\lambda h)^{1/2} \sum_j h \int_B \|a_j(y)\|_{\mathcal{H}}\, dy \longrightarrow 0 \end{aligned} \tag{19}$$

by (18) and (15). All these give

$$\int_{\mathcal{P}} \chi(p) \langle a_\lambda(p), b_\lambda(p) \rangle_{\mathcal{H}}\, d\mu(p) \longrightarrow 0. \tag{20}$$

From local to global. For any $\varepsilon > 0$, inner regularity gives a compact $K \subset \mathcal{R}$ such that $\mu(\mathcal{P} \setminus K) < \varepsilon$. The sets $\Psi(\operatorname{int} Q)$ constructed above cover K, so compactness gives a finite subcover. Choose nonnegative continuous cutoffs χ_i, supported in these sets, such that $S = \sum_i \chi_i = 1$ on K and $0 \le S \le 1$ on $\mathcal{P}$.

The term weighted by S tends to zero by (20). Since $1 - S$ vanishes on K, Cauchy–Schwarz and $\|b_\lambda\|_{\mathcal{H}} \le 2M$ give

$$\begin{aligned} &\left| \int_{\mathcal{P}} (1-S) \langle a_\lambda, b_\lambda \rangle_{\mathcal{H}}\, d\mu \right| \\ &\le \left(\int_{\mathcal{P}\setminus K} \|a_\lambda\|_{\mathcal{H}}^2\, d\mu \right)^{1/2} \left(\int_{\mathcal{P}\setminus K} \|b_\lambda\|_{\mathcal{H}}^2\, d\mu \right)^{1/2} \\ &\le 2M \left(\sup_\lambda \|a_\lambda\|_{\mu,2} \right) \varepsilon^{1/2}. \end{aligned} \tag{21}$$

By the triangle inequality, (20), and (21),

$$0 \le \limsup_{\lambda\to\infty} \left| \langle a_\lambda, b_\lambda \rangle_{L^2(\mu;\mathcal{H})} \right| \le 2M \left(\sup_\lambda \|a_\lambda\|_{\mu,2} \right) \varepsilon^{1/2}.$$

Letting $\varepsilon \downarrow 0$ proves the claim. □

A.2 Proof of Theorem 1

Proof. Assume first that MSE_λ remains bounded. Set $a_\lambda = g_\lambda - \bar{U}$ and $b_\lambda = F_\lambda - \bar{U}$. Since $a_\lambda = (g_\lambda - F_\lambda) + b_\lambda$ and $\|b_\lambda\|_{\mu,2} \le 2M$, where $M = \sup_{\tau,p} \|U(\tau;p)\|_X$, boundedness of the MSE gives $\sup_\lambda \|a_\lambda\|_{\mu,2} < \infty$. Moreover, $\mathrm{Lip}(a_\lambda) \le L_\lambda + K_U = o(\lambda)$. Lemmas A.2 and A.3, together with the Hilbert-space identity, give

$$\mathrm{MSE}_\lambda = \|a_\lambda\|_{\mu,2}^2 + \|b_\lambda\|_{\mu,2}^2 - 2\langle a_\lambda, b_\lambda \rangle = \|g_\lambda - \bar{U}\|_{\mu,2}^2 + \sigma_{\mathrm{phase}}^2 + o(1).$$

This proves the decomposition. If the lower limit of the MSE is infinite, the lower bound is immediate. Otherwise, choose a sequence tending to infinity along which the MSE converges to its finite lower limit. The decomposition applies on that bounded-error sequence and proves the bound. □

B FM2: Per-Sample OOD Error

B.1 Setup and lemmas

Throughout this appendix, assume Assumption 1 and Assumption 2. For $v : [0, W] \to X$, write

$$\|v\|_{W,2}^2 = \frac{1}{W} \int_0^W \|v(t)\|_X^2\, dt.$$

Thus $\|v\|_{W,2} = \|v(W\,\cdot)\|_{\mathcal{H}}$. Define accumulated difference

$$D_W(p) = \|U(\,\cdot\,\omega(p); p) - U(\,\cdot\,\omega_b; p)\|_{W,2}.$$

Define largest predictor on boundary in G

$$\varepsilon_W := \sup_{r \in \Gamma \cap G} e_W(r).$$

Lemma B.1 (Comparison with the training boundary)**.** *Under condition in Theorem 2, there exist a neighborhood $\mathcal{N}_0 \subset G$ of r_0 and a constant $C_\Gamma > 0$ such that*

$$e_W(p) \geq D_W(p) - \varepsilon_W - C_\Gamma(L_W + K_U)d_\omega(p) \qquad (p \in \mathcal{N}_0).$$

Proof. Let $\nu = \nabla\omega(r_0)/\|\nabla\omega(r_0)\|_2$, so that ω increases in direction ν at r_0. For p near r_0 with $\omega(p) < \omega_b$, we move along $p + h\nu$, $h \geq 0$, until ω reaches ω_b; the case $\omega(p) > \omega_b$ is identical with ν replaced by $-\nu$.

To control the rate, note that ω is C^1, so there are $\rho, \kappa > 0$ such that $B(r_0, \rho) \subset G$ and $\partial_\nu \omega \geq \kappa$ on $B(r_0, \rho)$. Let $\mathcal{N}_0$ be the set of p with $\|p - r_0\|_2 < \rho/2$ and $d_\omega(p) < \kappa\rho/2$. For $p \in \mathcal{N}_0$, since ω grows at rate at least κ along the ray from p in direction ν, it reaches ω_b at a point $r(p) \in \Gamma \cap G$ with

$$\|p - r(p)\|_2 \leq \kappa^{-1} d_\omega(p).$$

Write $r = r(p)$. Since $\omega(r) = \omega_b$, we have $F_W^{(2)}(r) = U(W\,\cdot\,\omega_b; r)$, so the triangle inequality gives

$$\begin{aligned} D_W(p) &\leq \|F_W^{(2)}(p) - g_W^{(2)}(p)\|_{\mathcal{H}} + \|g_W^{(2)}(p) - g_W^{(2)}(r)\|_{\mathcal{H}} + \|g_W^{(2)}(r) - F_W^{(2)}(r)\|_{\mathcal{H}} \\ &\quad + \|U(W\,\cdot\,\omega_b; r) - U(W\,\cdot\,\omega_b; p)\|_{\mathcal{H}} \\ &\leq e_W(p) + L_W\|p - r\|_2 + \varepsilon_W + K_U\|p - r\|_2. \end{aligned}$$

The claim follows from $\|p - r\|_2 \leq \kappa^{-1} d_\omega(p)$, with $C_\Gamma = \kappa^{-1}$. □

Lemma B.2 (Fast phase averaging)**.** *Let $f : \mathbb{T} \times [0,1] \to [0, B]$ satisfy $|f(\theta, s) - f(\theta, s')| \leq L|s - s'|$, and set $\bar{f}(s) = \int_{\mathbb{T}} f(\theta, s)\, d\theta$. Then for every $a > 0$,*

$$\Big|\int_0^1 f(as, s)\, ds - \int_0^1 \bar{f}(s)\, ds\Big| \leq \frac{L + B}{a}.$$

Proof. Let $m = \lfloor a \rfloor$, $s_j = j/a$, $I_j = [s_j, s_{j+1}]$. As s runs over I_j, as covers $\mathbb{T}$ exactly once, so $\int_{I_j} f(as, s_j)\, ds = \bar{f}(s_j)/a$. Replacing s_j by s in the second argument, on either side, costs at most $L\int_{I_j}(s - s_j)\, ds = L/(2a^2)$ each, so the $m \leq a$ full cells contribute at most L/a. The remaining interval has length less than $1/a$, and there $|f(as, s) - \bar{f}(s)| \leq B$. □

Lemma B.3 (Phase separation over a long window)**.** *There exists a constant $M > 0$ such that, for every $p \in G$ and $W > 0$, with $a = W\omega_b$ and $\xi = W(\omega(p) - \omega_b)$,*

$$D_W(p)^2 \geq V(p)\min\{\xi^2, 1\} - \frac{4MK_\tau|\xi|}{a} - \frac{4M^2}{a}.$$

Moreover, $D_W(p) \leq 2M$.

Proof. Write $u_p = U(\cdot; p) - \bar{U}(p)$. By continuity and compactness, $\|u_p(\tau)\|_X \leq M$ on $\mathbb{T} \times \mathcal{P}$, so $D_W(p) \leq 2M$. Define

$$A_p(v) = \int_0^1 \|u_p(\tau + v) - u_p(\tau)\|_X^2\, d\tau, \qquad f(\theta, s) = \|u_p(\theta + \xi s) - u_p(\theta)\|_X^2.$$

Substituting $t = Ws$ and using $Ws\omega_b = as$, $Ws\omega(p) = as + \xi s$ gives

$$D_W(p)^2 = \int_0^1 \|u_p(as + \xi s) - u_p(as)\|_X^2\, ds = \int_0^1 f(as, s)\, ds.$$

The function f is periodic with period one in θ, satisfies $0 \le f \le 4M^2$, has θ average $A_p(\xi s)$, and is $4MK_\tau|\xi|$ Lipschitz in s by the difference of squares and the phase Lipschitz bound. Lemma B.2 gives

$$D_W(p)^2 \ge \int_0^1 A_p(\xi s)\, ds - \frac{4MK_\tau|\xi|}{a} - \frac{4M^2}{a}.$$

Denote by $\widehat{u}_{p,n}$ the n-th Fourier coefficient of u_p in τ. Since u_p is centered, $\widehat{u}_{p,0} = 0$, and Parseval's identity gives

$$V(p) = \int_0^1 \|U(\tau;p) - \bar{U}(p)\|_X^2\, d\tau = \sum_{n\neq 0} \|\widehat{u}_{p,n}\|^2.$$

By the shift theorem, $u_p(\cdot + v)$ has coefficients $e^{2\pi i n v}\widehat{u}_{p,n}$, so $u_p(\cdot + v) - u_p$ has coefficients $(e^{2\pi i n v} - 1)\widehat{u}_{p,n}$. Since $|e^{2\pi i n v} - 1|^2 = 2(1 - \cos 2\pi n v)$, Parseval's identity applied to $u_p(\cdot + v) - u_p$ gives

$$A_p(v) = 2\sum_{n\neq 0} \|\widehat{u}_{p,n}\|^2 (1 - \cos 2\pi n v).$$

Integrating term by term, which Tonelli permits since every term is nonnegative,

$$\int_0^1 A_p(\xi s)\, ds = \sum_{n\neq 0} \|\widehat{u}_{p,n}\|^2\, 2\Big(1 - \tfrac{\sin(2\pi n\xi)}{2\pi n\xi}\Big) \ge \sum_{n\neq 0} \|\widehat{u}_{p,n}\|^2 \min\{(n\xi)^2, 1\} \ge V(p)\min\{\xi^2, 1\}.$$

The middle step uses $2\big(1 - \frac{\sin 2\pi x}{2\pi x}\big) \ge \min\{x^2, 1\}$ for all real x. When $|x| \le \frac{1}{2}$, write $1 - \frac{\sin 2\pi x}{2\pi x} = \int_0^1 (1 - \cos 2\pi x r)\, dr$ and apply $1 - \cos t \ge 2t^2/\pi^2$ for $|t| \le \pi$, which gives the lower bound $\frac{16}{3}x^2 \ge x^2$. When $|x| \ge \frac{1}{2}$, $\frac{\sin 2\pi x}{2\pi x} \le \frac{1}{\pi}$, so the left side is at least $2(1 - \frac{1}{\pi}) \ge 1$. The last step uses $|n| \ge 1$. Plug into inequality of $D_W(p)^2$, lemma is proved. □

B.2 Derivation of per-sample bound

Proof of Theorem 2. Let $\mathcal{N}_0$ be the neighborhood from Lemma B.1. Since V is continuous and $V(r_0) > 0$, choose a smaller neighborhood $\mathcal{N} \subset \mathcal{N}_0$ of r_0 and a constant $v_* > 0$ such that $V(p) \ge v_*$ on $\mathcal{N}$.

Fix $B > 0$ and consider OOD points $p \in \mathcal{N}$ with $0 < W d_\omega(p) \le B$. Lemma B.1 gives

$$e_W(p) \ge D_W(p) - \delta_{W,B}, \qquad \delta_{W,B} := \varepsilon_W + C_\Gamma(L_W + K_U)\frac{B}{W}.$$

Since $\varepsilon_W \to 0$ and $L_W = o(W)$, we have $\delta_{W,B} \to 0$. Moreover,

$$e_W(p)^2 \ge D_W(p)^2 - 2D_W(p)\delta_{W,B} \ge D_W(p)^2 - 4M\delta_{W,B}.$$

The first inequality follows by squaring $e_W(p) \ge D_W(p) - \delta_{W,B}$ when $D_W(p) \ge \delta_{W,B}$, and holds trivially otherwise because its right side is negative. The second uses $D_W(p) \le 2M$.

Set $a = W\omega_b$ and $\xi = W(\omega(p) - \omega_b)$. Since $|\xi| = W d_\omega(p) \le B$, Lemma B.3 implies

$$D_W(p)^2 \ge V(p)\min\{(W d_\omega(p))^2, 1\} - \frac{C_B}{W}, \qquad C_B = \frac{4MK_\tau B + 4M^2}{\omega_b}.$$

Combining the two inequalities and dividing by $V(p) \ge v_*$ yields

$$\frac{e_W(p)^2}{V(p)} \ge \min\{(W d_\omega(p))^2, 1\} - \frac{C_B/W + 4M\delta_{W,B}}{v_*}.$$

The final term tends to zero independently of p. □

C Proof of Theorem 3

Write $\Phi(t, x, p)$ for the flow of $\dot{x} = f(x, p)$.

Proof. Fix $p_0 \in \mathcal{P}$ and a point $x_0 \in \mathcal{C}(p_0)$, and choose a transverse section Σ through x_0. By Teschl (2012, Lemma 12.7), the assumption $f \in C^2$ and hyperbolicity give a local C^2 continuation $x_\Sigma(p)$ of the corresponding fixed point of the Poincaré map. Moreover, Teschl (2012, Lemma 6.9) applied to the augmented system $\dot{x} = f(x, p)$ and $\dot{p} = 0$ shows that the local return time $\tau_\Sigma(x, p)$ is C^2 jointly in (x, p). Therefore,

$$T(p) = \tau_\Sigma(x_\Sigma(p), p)$$

is C^2 locally. After shrinking the parameter neighborhood if necessary, this first return is the minimal period. The unique-cycle assumption identifies the continued orbit with $\mathcal{C}(p)$, so the local definitions agree and T is continuous on $\mathcal{P}$.

Define the local unanchored waveform by

$$\widetilde{U}(\tau, p) = \Phi\big(\tau T(p), x_\Sigma(p), p\big).$$

Since the flow is C^2, $\widetilde{U}$ is jointly C^2 in (τ, p). Let $A(\tau, p) = a_{\mathrm{obs}}(\widetilde{U}(\tau, p))$. Since $a_{\mathrm{obs}} \in C^2$, the function A is C^2. At p_0, its unique maximizing phase is a nondegenerate critical point, so the implicit function theorem gives a local C^1 continuation $\tau_*(p)$. The strict gap between this maximum and the values outside a small phase neighborhood ensures that $\tau_*(p)$ remains the unique global maximizer for nearby p. Hence

$$U(\tau; p) = \widetilde{U}(\tau + \tau_*(p), p)$$

is jointly continuous locally. Uniqueness of the anchor makes these local definitions agree, so U is jointly continuous on $\mathbb{T} \times \mathcal{P}$.

It follows that U_G is continuous. Since $\mathcal{P}$ is compact, coordinatewise universal approximation gives uniform approximability of T and U_G (Hornik et al., 1989; Leshno et al., 1993). □

D ODE Systems and Data Generation

D.1 Goodwin oscillator

The Goodwin oscillator (Gonze & Ruoff, 2021) is

$$\frac{dX}{dt} = \frac{a_1}{1 + (Z/K)^{n_H}} - b_1 X, \qquad \frac{dY}{dt} = \alpha_1 X - \beta_1 Y, \qquad \frac{dZ}{dt} = \gamma_1 Y - \delta_1 Z, \tag{22}$$

with parameter vector

$$p = (a_1, K, b_1, \alpha_1, \beta_1, \gamma_1, \delta_1, n_H) \in \mathbb{R}^8. \tag{23}$$

The X coordinate is used for peak anchoring.

D.2 Goldbeter 1995 circadian model

The GOLD95 state is $x = (M, P_0, P_1, P_2, P_N)$, representing *per* mRNA, unphosphorylated, mono- and bisphosphorylated cytosolic PER, and nuclear PER (Goldbeter, 1995). Its equations are

$$\frac{dM}{dt} = v_s \frac{K_I^{n_H}}{K_I^{n_H} + P_N^{n_H}} - v_m \frac{M}{K_m + M}, \tag{24}$$

$$\frac{dP_0}{dt} = k_s M - V_1 \frac{P_0}{K + P_0} + V_2 \frac{P_1}{K + P_1}, \tag{25}$$

$$\frac{dP_1}{dt} = V_1 \frac{P_0}{K + P_0} - V_2 \frac{P_1}{K + P_1} - V_3 \frac{P_1}{K + P_1} + V_4 \frac{P_2}{K + P_2}, \tag{26}$$

$$\frac{dP_2}{dt} = V_3 \frac{P_1}{K + P_1} - V_4 \frac{P_2}{K + P_2} - k_1 P_2 + k_2 P_N - v_d \frac{P_2}{K_d + P_2}, \tag{27}$$

$$\frac{dP_N}{dt} = k_1 P_2 - k_2 P_N. \tag{28}$$

Following Goldbeter (1995), the four phosphorylation Michaelis constants are set equal to K, and the Hill coefficient is fixed at $n_H = 4$. The free parameter vector is

$$p = (v_s, v_m, K_m, k_s, v_d, K_d, k_1, k_2, K_I, K, V_1, V_2, V_3, V_4) \in \mathbb{R}^{14}. \tag{29}$$

The M coordinate is used for peak anchoring.

D.3 Sampling, integration, and acceptance

Candidate parameter vectors are drawn from a scrambled Sobol sequence over the box in Table 4. Each ODE is integrated until transient dynamics decay, the asymptotic period is estimated from successive peaks, and one peak-anchored period is resampled to 128 uniform phase points with refinement factor 5. Candidates are rejected if integration fails, no sustained oscillation is detected, or period detection fails.

For Goodwin, 5,000 trajectories are accepted from 5,062 candidates (1.2% rejection), with $T \in [3.62, 7.28]$ h and mean $T = 4.94$ h. For GOLD95, 10,000 are accepted from 13,958 candidates (28.4% rejection), with $T \in [16.01, 40.71]$ h and mean $T = 25.05$ h. The GOLD95 rejections comprise 2,318 period-detection failures and 1,640 oscillation-criterion failures.

Solver and tolerance settings, burn-in horizons, initial conditions, peak-detection and oscillation thresholds, and all random seeds are specified in the released data-generation code.

Table 4: Sobol sampling box. Every interval is $\pm 30\%$ of the baseline value, except the Goodwin Hill exponent, which is sampled on $n_H \in [10, 14]$.

Goodwin				GOLD95			
Param	Baseline	Lower	Upper	Param	Baseline	Lower	Upper
a_1	8.37	5.86	10.88	v_s	0.76	0.53	0.99
K	1.37	0.96	1.78	v_m	0.65	0.46	0.85
b_1	1.00	0.70	1.30	K_m	0.50	0.35	0.65
α_1	1.00	0.70	1.30	k_s	0.38	0.27	0.49
β_1	0.60	0.42	0.78	v_d	0.95	0.67	1.24
γ_1	1.00	0.70	1.30	K_d	0.20	0.14	0.26
δ_1	0.80	0.56	1.04	k_1	1.90	1.33	2.47
n_H	12.0	10.0	14.0	k_2	1.30	0.91	1.69
				K_I	1.00	0.70	1.30
				K	2.00	1.40	2.60
				V_1	3.20	2.24	4.16
				V_2	1.58	1.11	2.05
				V_3	5.00	3.50	6.50
				V_4	2.50	1.75	3.25

E Training Setup and Ablation Constructions

E.1 Main decoupled training setup

For each system, the main model trains a period MLP and a waveform FNO separately. We use a fixed $80/20$ train–validation split generated with split seed 0, giving $4{,}000/1{,}000$ samples for Goodwin and $8{,}000/2{,}000$ samples for GOLD95. Each model is trained with training seeds $0, 1, 2, 3, 4$; these seeds affect weight initialization and batch shuffling but not the data split. Unless stated otherwise, the ablation experiments use the same model capacity and output grid.

Table 5: Default configuration for the period and waveform models.

Model	Architecture	Grid	Epochs	LR	Min. LR
Period MLP	depth 2, width 128	–	100	5×10^{-4}	10^{-6}
Waveform FNO	16 modes, 128 channels	128	100	10^{-3}	10^{-5}

Across the five training seeds, the period MLP has mean IID validation RMSE 0.0077 h for Goodwin and 0.165 h for GOLD95.

All models are trained with Adam, cosine learning-rate annealing, batch size 32, no weight decay, and no dropout. For the period MLP, the parameters and periods are standardized. For the waveform FNO, the input coordinates and parameters are standardized, while non-negative state variables are power-transformed. All scalers are fitted on the training split only.

Table 1 reports losses at the final training epoch. The mean, standard deviation, minimum, and maximum validation losses are computed across the five training seeds. Remaining implementation details are provided in the released code.

E.2 ABLATION IMPLEMENTATION

Unless stated otherwise, all models use the default training settings and the 128-point grid in Table 5. Each experiment is trained for 100 epochs using training seeds $0, 1, 2, 3, 4$. Each seed controls model weight initialization (`torch.manual_seed`). The IID experiments use the same $80/20$ split generated with split seed 0. The period-based OOD splits are fixed across training seeds. Table 6 maps the original experiment numbers to their preprocessing choices and validation splits.

Table 6: Constructions used in the ODE ablations.

Row	Norm.	Align.	Split	Waveform target
Main	✓	✓	IID	Aligned cycle
1	×	✓	IID	Aligned fixed window
2	✓	×	IID	Unaligned cycle
3	×	×	IID	Unaligned fixed window
4	✓	✓	OOD	Aligned cycle
5	×	✓	OOD	Aligned fixed window

Aligned fixed window (row 1). For Goodwin, each trajectory is integrated from the same fixed initial condition for 500 h. The first subsequent peak is located, and the ODE is then integrated directly over a 10 h window. For GOLD95, the peak state from the existing aligned waveform is used as the initial condition, and the ODE is integrated directly over a 100 h window without an additional burn-in. Both trajectories are sampled at 128 uniform points without periodic tiling. The FNO input is $(t/W, p)$, so phase alignment is retained but time is not period-normalized.

Unaligned cycle (row 2). For Goodwin, every parameterized system starts from the same fixed initial condition and is integrated for $t_{\mathrm{burn}} = 500$ h. One period is then recorded directly from the integration endpoint. Its initial phase therefore has a portion from the parameter-dependent transient. For GOLD95, each peak-aligned state is instead advanced by the same 500 h, giving the controlled phase shift $\phi_0(p) = t_{\mathrm{burn}}/T(p) \pmod 1$, after which one period is recorded without realignment. In both systems, the recorded trajectory is sampled on $\tau = t/T(p) \in [0, 1]$. The target is period-normalized but not phase-aligned.

Unaligned fixed window (row 3). The unaligned one-period trajectories from row 2 are first interpolated using periodic cubic splines. The resulting continuous waveforms are then repeated over a fixed physical-time window and sampled at 128 uniform points. We use $W = 10$ h for Goodwin and $W = 100$ h for GOLD95. The FNO input is $(t/W, p)$, with neither phase alignment nor period normalization.

Aligned cycle under the OOD split (row 4). The targets are peak-aligned, period-normalized waveforms. For Goodwin, the $2{,}000$ shortest-period samples form the training set ($T \in [3.62, 4.72]$ h), and the 500 longest-period samples form the validation set ($T \in [5.77, 7.28]$ h). For GOLD95, training uses $T < 0.6T_{\max} = 24.43$ h ($n_{\mathrm{train}} = 4{,}822$), whereas validation uses $T \geq 24.43$ h ($n_{\mathrm{val}} = 5{,}178$, $T \in [24.44, 40.71]$ h).

Aligned fixed window under the OOD split (row 5). The period splits are identical to row 4. The window is the largest training period: $W = 4.72$ h for Goodwin and $W = 24.43$ h for GOLD95. Using linear interpolation, training trajectories are tiled to fill the window, whereas validation trajectories are clipped to their first W hours. The validation targets cover 0.65–0.82 of a cycle for Goodwin and 0.60–1.00 cycle for GOLD95. Phase alignment is retained.

F FM1 AND FM2 VALIDATION EXPERIMENTS

F.1 FM1 SHIFT SWEEP: CONFIGURATION

Data. Each sample pairs a parameter p with one period of its orbit on a normalized time grid. For a shift c, the target period starts c hours after the peak. Inputs are the same for every c; only the targets change. We use $c \in \{0, 1, 10, 10^2, 10^3, 10^4\}$ hours for Goodwin and $c \in \{0, 1, 10, 10^2, 10^3, 10^4, 10^5\}$ hours for Gold95.

Split and scaling. A single 80/20 train and validation split, fixed by split seed 0, is shared by all c and all seeds. This gives $4000/1000$ samples for Goodwin and $8000/2000$ for Gold95. Inputs are standardized, and targets are encoded by a power transform fitted on the training targets separately at each c. All losses are computed in this encoded space.

Phase variance. For each validation orbit, we encode the base orbit on 512 uniform phases and take the mean over phases and state variables of its squared deviation from the phase average. σ^2_{phase} is the average of this quantity over the validation set, with the same batch averaging as the validation loss. Because the encoding depends on c, σ^2_{phase} is computed separately at each c.

Model and training. The FNO takes (τ, p) as input channels and outputs all state variables, with 16 Fourier modes and 128 hidden channels. We train with the MSE loss, Adam at initial learning rate 10^{-3}, cosine annealing to 10^{-5} over 100 epochs, and batch size 32. Seeds 0 to 4 set the weight initialization and batch order (`torch.manual_seed`). The validation loss is evaluated after every epoch: Final val MSE is its value after the last epoch, and Best val MSE is its minimum over all epochs.

F.2 FM1 SHIFT SWEEP: DETAILED RESULTS

Detailed FM1 sweep results are summarized in Table 7. For each of the five seeds in rows marked †, the best validation checkpoint occurs in epochs 0–3. Afterward, training loss continues to decrease while validation MSE rises. Best val MSE reports the error at that checkpoint, and Final val MSE reports the error at the last epoch.

Table 7: FM1 shift sweep, mean ± std over five seeds. Val MSE is divided by σ^2_{phase} at the same c; the σ^2_{phase} column gives its unnormalized value. Best epoch is the range over seeds, zero based.

System	c (h)	σ^2_{phase}	Best val MSE/σ^2_{phase}	Best epoch	Final val MSE/σ^2_{phase}
Goodwin	0	0.5705	$6.870\times10^{-4} \pm 7.9\times10^{-6}$	94–99	$6.921\times10^{-4} \pm 1.30\times10^{-5}$
	1	0.5730	$7.034\times10^{-4} \pm 8.7\times10^{-6}$	96–99	$7.099\times10^{-4} \pm 1.09\times10^{-5}$
	10	0.5718	$8.539\times10^{-4} \pm 2.03\times10^{-5}$	96–99	$8.618\times10^{-4} \pm 2.57\times10^{-5}$
	100	0.5721	1.0055 ± 0.0018	0–1	1.8241 ± 0.0293 †
	10^3	0.5721	1.0013 ± 0.0007	1–3	1.8134 ± 0.0199 †
	10^4	0.5720	1.0034 ± 0.0005	0–1	1.7855 ± 0.0338 †
Gold95	0	0.7433	$1.0059\times10^{-3} \pm 1.90\times10^{-5}$	79–98	$1.0135\times10^{-3} \pm 1.99\times10^{-5}$
	1	0.7427	$1.0090\times10^{-3} \pm 1.23\times10^{-5}$	79–98	$1.0171\times10^{-3} \pm 1.43\times10^{-5}$
	10	0.7410	$1.0868\times10^{-3} \pm 1.60\times10^{-5}$	94–98	$1.0915\times10^{-3} \pm 1.56\times10^{-5}$
	100	0.7423	$3.918\times10^{-3} \pm 1.70\times10^{-4}$	95–98	$3.941\times10^{-3} \pm 1.87\times10^{-4}$
	10^3	0.7422	1.0030 ± 0.0004	0–2	1.7362 ± 0.0185 †
	10^4	0.7423	1.0030 ± 0.0009	0–2	1.7566 ± 0.0476 †
	10^5	0.7422	1.0008 ± 0.0009	0–3	1.6935 ± 0.0326 †

F.3 FM2 WINDOW SWEEP: CONFIGURATION

Network and optimization. We use the neuralop 1D FNO implementation with a single fixed hidden width and neuralop's default depth and activation. Input channels are $1 + n_{\text{params}}$ (a time encoding plus the parameter vector, broadcast along the grid) and output channels are n_{vars}. Goodwin: $n_{\text{vars}} = 3$, $n_{\text{params}} = 8$. GOLD95: $n_{\text{vars}} = 5$, $n_{\text{params}} = 14$. Table 8 lists the training hyperparameters, held fixed across the sweep.

Table 8: FM2 window sweep: training hyperparameters (identical for both systems).

Hidden channels	128
Optimizer	Adam
Learning rate	10^{-3}
Weight decay	0
LR schedule	cosine annealing, $\eta_{\min} = 10^{-5}$
Epochs	100
Effective batch size	32 (no gradient accumulation needed)
Validation frequency	every epoch

Window sweep. For each system we fix $W_0 =$ cutoff and sweep $W \in W_0 \cdot \{1, 2, 4, 8, 16, 64\}$, holding $dt = W_0/\text{base_grid}$ constant so the grid scales as base_grid $\cdot$ factor (base_grid $= 128$ for both systems). Goodwin uses a cutoff of 4.72h; GOLD95's cutoff is 24.428h. The resulting W values are:

- Goodwin: $\{4.72, 9.44, 18.88, 37.76, 75.52, 302.08\}$h
- GOLD95: $\{24.43, 48.86, 97.71, 195.43, 390.85, 1563.40\}$h

Each (W, system) combination is trained with 5 seeds (0–4). Each seed (0–4) sets weight initialization (`torch.manual_seed`) and, via an independent `torch.Generator`, minibatch shuffling.

Fixed spectral width. For each system, one n_{modes} is chosen (1728 for Goodwin, 1984 for GOLD95) at $2\times$ headroom over the harmonic demand at the largest W, and is held fixed across every W in that system's sweep (neuralop's SpectralConv slices this down to the FFT length at smaller W, so a single fixed value is valid throughout). This gives 865 kept rfft frequencies for Goodwin and 993 for GOLD95.

Split and OOD definition. Training/validation is a period-based split: samples with period $T <$ cutoff are training, $T \geq$ cutoff are validation (OOD in period, and hence in the window-mismatch variable z). Goodwin: $N = 5000$ total, 2020 train / 2980 val. GOLD95: $N = 10000$ total, 4822 train / 5178 val. The boundary and hardest-OOD reference sets used for the per-sample tracking use $n_{\text{boundary}} = 20$, $n_{\text{hard}} = 20$; the variance floor guard is $v_{\min_\text{frac}} = 0.1$; windows are not endpoint-inclusive.

F.4 FM2 WINDOW SWEEP: DETAILED RESULTS

Tables 9 and 10 report the validation MSE normalized by σ^2_{phase} for each W. Final and best-validation MSE are shown as mean $\pm$ one standard deviation over five seeds.

Table 9: Goodwin: W sweep, five seeds per W.

W (h)	Final MSE/σ^2_{phase}	Best MSE/σ^2_{phase}
4.72	0.073 ± 0.009	0.064 ± 0.005
9.44	0.305 ± 0.007	0.289 ± 0.014
18.88	0.805 ± 0.027	0.774 ± 0.043
37.76	1.171 ± 0.058	1.144 ± 0.052
75.52	1.321 ± 0.143	1.251 ± 0.090
302.08	1.633 ± 0.141	1.051 ± 0.012

Table 10: GOLD95: W sweep, five seeds per W.

W (h)	Final MSE/σ^2_{phase}	Best MSE/σ^2_{phase}
24.43	0.135 ± 0.015	0.109 ± 0.009
48.86	0.514 ± 0.050	0.492 ± 0.048
97.71	1.076 ± 0.042	0.990 ± 0.050
195.43	1.219 ± 0.043	1.164 ± 0.042
390.85	1.333 ± 0.069	1.228 ± 0.072
1563.40	2.244 ± 0.121	1.114 ± 0.042

F.5 Per-sample analysis for FM2

For each system, we collect $z = W d_\omega(p)$ for every validation parameter p and every window W in the sweep. Let $z_{\min}$ and $z_{\max}$ be the smallest and largest of these values, which occur at the shortest and longest window, respectively. We use the bin edges $b_k = z_{\min}\, r^k$ for $k = 0, \ldots, 21$ with $r = (z_{\max}/z_{\min})^{1/21}$, and bin k is $[b_k, b_{k+1})$. This gives $z_{\min} = 1.18 \times 10^{-3}$, $z_{\max} = 22.5$ and $r = 1.60$ for Goodwin, and $z_{\min} = 1.37 \times 10^{-3}$, $z_{\max} = 25.7$ and $r = 1.60$ for GOLD95. The same edges are used for every W.

The validation set is fixed, so every seed is evaluated on the same parameters. We keep only validation parameters whose waveform variance satisfies $V(p) \geq 0.1\, V(p_\partial)$, where p_∂ is the training parameter with the longest period, so that the ratio $e_W(p)^2/V(p)$ is not dominated by waveforms with very small variance.

Table 11 summarizes the per-sample results for each W. "Bins below" counts the nonempty bins in which the bin minimum, averaged over seeds, lies below $\min\{z^2, 1\}$ at the bin center; the bin mean never does. The two RSD columns measure how much the bin mean and the bin minimum vary across seeds. For each bin, we divide the standard deviation over the 5 seeds by the average over the 5 seeds, and report the median of this ratio over bins.

Table 11: Per-sample results over the W sweep.

System	W (h)	Bins below	RSD of bin mean (%)	RSD of bin min (%)
Goodwin	4.72	6 / 13	8.8	16.8
	9.44	5 / 13	8.8	12.9
	18.88	6 / 13	6.3	12.5
	37.76	5 / 13	11.5	11.3
	75.52	2 / 13	8.7	14.7
	302.08	0 / 12	9.2	9.5
GOLD95	24.43	6 / 13	11.1	22.3
	48.86	4 / 13	11.1	13.9
	97.71	3 / 14	7.8	16.1
	195.43	0 / 13	8.4	9.0
	390.85	0 / 14	9.6	13.1
	1563.40	0 / 13	2.4	7.5

F.6 Applicability of Theorem 2

Parameter notations are given in Section D. Except one varied parameter, all other parameters are fixed at their baseline values. The period decreases monotonically in δ_1 for Goodwin and increases monotonically in v_d for GOLD95.

Table 12: Single parameter period sweeps used to check Assumption 2.

System	Varied	Range	Period (h)
Goodwin	δ_1	$[0.56, 1.04]$	$[4.36, 5.53]$
GOLD95	v_d	$[0.66, 1.26]$	$[21.58, 26.03]$

G SNAPSHOT-CONDITIONED PREDICTION DETAILS

The supplied evaluation selects one full-state snapshot on each post-transient validation orbit, predicts a period and waveform, and matches the snapshot by periodic interpolation in encoded state space. It uses $N_\phi = 100$ candidates per search level and $D_{\text{search}} = 3$ levels.

Algorithm 1 Snapshot-based physical-time prediction

1: Predict $\widehat{T}(p)$ and the waveform grid; construct interpolants $\widehat{U}$ in physical state units and $\widehat{U}_{\text{enc}}$ in encoded state space, and encode y_{obs}.
2: Set the initial search interval to one full cycle.
3: **for** $d = 1, \ldots, D_{\text{search}}$ **do**
4: Evaluate $\|\widehat{U}_{\text{enc}}(\phi; p) - y_{\text{obs,enc}}\|_2^2$ at N_ϕ equally spaced candidates.
5: Select a minimizing candidate $\widehat{\phi}_d$ and use its two neighboring candidates to define the next interval, treating phase modulo one.
6: **end for**
7: Return $\widehat{x}(t) = \widehat{U}(\widehat{\phi}_{D_{\text{search}}} + t/\widehat{T}(p); p)$ over the requested physical-time horizon.

Evaluation metric. As in ablation studies Appendix E, errors are measured in the encoded state space of the training output scaler E. For G output points, let $\tau_j = j/(G-1)$, $j = 0, \ldots, G-1$, and let $x_{\text{ref}}(t; p, y_{\text{obs}})$ be the ODE solution initialized at the snapshot. The reported per-sample waveform error is $\text{MSE}_{\text{cycle}}(p) = \frac{1}{nG} \sum_{j=0}^{G-1} \|\widehat{U}_{\text{enc}}((\widehat{\phi} + \tau_j) \bmod 1; p) - E(x_{\text{ref}}(T(p)\tau_j; p, y_{\text{obs}}))\|_2^2$, where n is the number of state components.

Per-seed results. Table 13 reports the results for each training seed. All seeds are evaluated on the same validation samples. Snapshot times are drawn at random once for each validation sample and shared by all seeds.

Table 13: Per-seed snapshot-based validation. Entries are mean / median / 95th percentile over validation samples.

System	Seed	waveform MSE	Period error (%)
Goodwin	0	1.43×10^{-4} / 8.29×10^{-5} / 5.03×10^{-4}	0.100 / 0.072 / 0.256
	1	9.88×10^{-5} / 5.71×10^{-5} / 3.37×10^{-4}	0.098 / 0.067 / 0.258
	2	1.47×10^{-4} / 8.38×10^{-5} / 4.56×10^{-4}	0.098 / 0.070 / 0.246
	3	1.14×10^{-4} / 6.23×10^{-5} / 3.53×10^{-4}	0.097 / 0.072 / 0.238
	4	1.21×10^{-4} / 8.05×10^{-5} / 3.64×10^{-4}	0.103 / 0.076 / 0.278
Gold95	0	7.32×10^{-4} / 1.88×10^{-4} / 3.95×10^{-3}	0.433 / 0.320 / 1.190
	1	7.79×10^{-4} / 2.15×10^{-4} / 4.01×10^{-3}	0.412 / 0.291 / 1.213
	2	8.08×10^{-4} / 1.94×10^{-4} / 4.77×10^{-3}	0.429 / 0.323 / 1.175
	3	7.73×10^{-4} / 2.09×10^{-4} / 4.20×10^{-3}	0.407 / 0.301 / 1.099
	4	7.14×10^{-4} / 2.14×10^{-4} / 3.92×10^{-3}	0.449 / 0.329 / 1.291

H FitzHugh–Nagumo PDE implementation details

PDE and parameter domain. We consider the one-dimensional FitzHugh–Nagumo reaction–diffusion system adapted from (Cebrián-Lacasa et al., 2024),

$$u_t = D_u u_{xx} - u^3 + u - v, \tag{30}$$

$$v_t = D_v v_{xx} + \varepsilon(u - bv + a), \tag{31}$$

on a periodic spatial domain $x \in [0, \ell]$. The fixed constants and parameter ranges are summarized in Table 14. We generate parameter vectors using a scrambled Sobol sequence and retain 5000 samples whose post-transient solutions lie on the winding-one traveling-wave branch.

Table 14: FitzHugh–Nagumo PDE and data-generation settings.

Quantity	Symbol	Value
Activator diffusion	D_u	1
Inhibitor diffusion	D_v	1
Domain length	ℓ	50
Fixed reaction parameter	a	0.1
Sampled parameter	b	$[0.35, 0.65]$
Sampled parameter	ε	$[0.035, 0.065]$
Spatial grid size	N_x	256
PDE transient burn-in	$T_{\mathrm{burn,PDE}}$	600
Fourier-phase probe window	T_{probe}	200
Retained samples	–	5000

Numerical simulation and initial condition. The spatial Laplacian is approximated by a centered second-order finite difference on the periodic grid. The resulting ODE system is integrated using an explicit fourth-order Runge–Kutta method, with a time step below the explicit diffusion stability limit.

For each parameter sample, we first compute the limit cycle of the spatially homogeneous FitzHugh–Nagumo kinetics. Different phases on this cycle are assigned to successive spatial locations, with one full phase cycle distributed around the periodic domain. The PDE is then integrated for $T_{\mathrm{burn,PDE}} = 600$ before the post-transient traveling wave is measured.

Period estimation and phase alignment. For a winding-one rigid traveling wave,

$$u_p(x, t) = Q_p(x + v_{\mathrm{tw}}(p)t),$$

the first spatial Fourier coefficient satisfies

$$\widehat{u}_{p,1}(t) = \widehat{u}_{p,1}(0) \exp\left(\frac{2\pi i v_{\mathrm{tw}}(p) t}{\ell}\right).$$

We estimate its angular velocity $\nu(p)$ by fitting a linear function to the unwrapped Fourier phase over a post-burn-in interval of length T_{probe}, and define

$$T(p) = \frac{2\pi}{|\nu(p)|}.$$

The canonical phase is fixed by

$$\arg \widehat{u}_{p,1} = 0 \pmod{2\pi}.$$

The solution is advanced to the next occurrence of this phase before the periodic field is recorded. This Fourier-phase condition plays the same role as the peak-based anchor used for the ODE systems.

Starting from the anchor, one complete period is recorded and resampled to a 128×128 grid in normalized time and space. The stored canonical field has shape $128 \times 128 \times 2$, where the final dimension corresponds to u and v.

Table 15: Period-based OOD split for the FitzHugh–Nagumo experiments.

Set	Samples	Period range	Cycles in W_{OOD}
Training	2000	37.05–42.94	1.00–1.16
Validation	500	52.43–55.85	0.77–0.82

Construction of learning targets. All target representations for a given parameter sample are constructed from the same phase-aligned one-period field. The ablations therefore change only the representation while keeping the underlying periodic solution fixed.

For the normalized representation, the phase-aligned periodic field is represented on

$$(\tau, x) \in [0, 1) \times [0, \ell)$$

using 128 points in normalized time and 128 points in space.

For the FM1 experiments, we impose a common observation burn-in

$$T_{\mathrm{burn,in}} = 10000,$$

which is distinct from the PDE transient burn-in $T_{\mathrm{burn,PDE}} = 600$. It determines the observation phase

$$\phi_0(p) = \frac{T_{\mathrm{burn,in}} \bmod T(p)}{T(p)}.$$

Experiment 2 records one normalized cycle from this phase, while Experiment 3 uses the same phase to construct a fixed physical-time window.

For the iid fixed-window experiments, we use $W = 100$, with 256 temporal points and 128 spatial points. This window contains approximately 1.79–2.70 periods over the dataset. All shifted and fixed-window targets are evaluated from the same phase-aligned periodic records using Fourier interpolation.

Data splits. The iid experiments use 4000 training and 1000 validation samples. The period-based OOD split is summarized in Table 15. Experiments 4 and 5 use exactly the same training and validation samples.

The remaining 2500 samples are unused. Experiment 4 predicts the phase-aligned normalized field. Experiment 5 predicts on the common physical-time window

$$W_{\mathrm{OOD}} = T_{\mathrm{train,max}} = 42.94.$$

Model and training configuration. The waveform predictor is a two-dimensional FNO. Its inputs are the temporal coordinate, spatial coordinate, and parameters (b, ε), and its outputs are the two fields u and v. For the normalized representation, the temporal coordinate is $\tau \in [0, 1)$; for the fixed-window representation, it is physical time. The period predictor is a separate MLP mapping $p = (b, \varepsilon)$ to $T(p)$. Their configurations are summarized in Table 16.

The input parameters are standardized using statistics from the training split. For the FNO, u and v are standardized separately using training-set statistics. The MSE values in Table 3 are final-epoch losses in this standardized output space. For the period MLP, both the parameters and the period are standardized during training.

Period prediction is evaluated in the original period scale using

$$\mathrm{NMSE}_T = \frac{\sum_i \left(\widehat{T}_i - T_i\right)^2}{\sum_i T_i^2}.$$

The period predictor achieves validation NMSE 2.72×10^{-9} on the iid split and 2.16×10^{-3} on the OOD split.

Table 16: FitzHugh–Nagumo model and training configuration.

Setting	Waveform FNO	Period MLP
Architecture	16 modes/dim., 64 channels	depth 2, width 128
Activation	–	GELU
Batch size	16	32
Epochs	50	200
Optimizer	Adam	Adam
Initial learning rate	10^{-3}	10^{-3}
Schedule	cosine annealing	cosine annealing
iid split	$4000/1000$	$4000/1000$
OOD split	$2000/500$	$2000/500$