Universal Approximation Theorems for Dynamical Systems with Infinite-Time Horizon Guarantees

Ábel Ságodi and Il Memming Park
Champalimaud Foundation, Lisbon, Portugal
{abel.sagodi, memming.park}@research.fchampalimaud.org

Abstract

Universal approximation theorems establish the expressive capacity of neural network architectures. A trained dynamical model can pass every finite-time test yet fail in the long run in three ways: it can settle to the wrong steady state, drift out of phase, or collapse a continuum of memory states into isolated ones. We consider models whose vector field is parameterized expressively enough to approximate any target vector field together with its Jacobian, e.g., Neural ODEs. We prove that such models can avoid all three failures, matching a target system's trajectories over the infinite-time horizon $[0, \infty)$ on all but a small set of initial conditions (ε-δ closeness: trajectory error within ε except for initial conditions of measure $< \delta$). We cover three target classes, the dynamical primitives of decision-making, rhythm generation, and analog memory: structurally stable systems with a finite number of hyperbolic periodic orbits (fixed points or limit cycles via exact period matching), and continuous attractors that are normally hyperbolic manifolds of fixed points or of isochronous limit cycles (via discretization). A temporal generalization bound, ε-δ closeness implies time-averaged L^p error $\leq \varepsilon^p + \delta \cdot D^p$ for all $t \geq 0$, connects these guarantees to training metrics. This is the first universal approximation framework for multistable dynamics over an unbounded horizon.

1 Introduction

Universal approximation results provide rigorous guarantees about the expressive capacity of neural networks. For dynamical systems, Recurrent Neural Networks (RNNs) are widely cited as universal approximators[60], justifying their deployment in models of neural computation[49,199]. However, a fundamental gap exists between this theoretical promise and biological reality. Existing guarantees are strictly limited to *finite* time horizons or systems with a globally stable equilibrium (the fading memory property). This restriction explicitly excludes **multistability**, i.e., the coexistence of multiple attractors, which is the dynamical basis of essential cognitive functions. Decision-making relies on selecting among distinct basins of attraction; working memory requires self-sustaining persistent activity; and neural oscillations (limit cycles) drive rhythmic motor control[21,101,160,191]. Consequently, current theories fail to address the very dynamical regimes required for computation. Extending guarantees to infinite time is crucial for **temporal generalization**: ensuring models coherently replicate dynamics over indefinite durations rather than finite windows. Three fundamental failure modes prevent a naive extension of finite-time results:

- **B-type error** (Basin mismatch): Small approximation errors near separatrices push trajectories into incorrect basins of attraction.
- **P-type error** (Phase drift): For limit cycles, minute period mismatches cause unbounded phase divergence as $t \to \infty$.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

- **D-type error** (Discretization): Continuous attractors are not structurally stable; generic perturbations destroy their continua of marginally stable fixed points.

These topological obstructions call for more than standard Grönwall-based error analysis, which yields exponentially growing bounds unusable for infinite time.

This work establishes universal approximation results for multistable dynamical systems over infinite time horizons. Our approach exploits structural stability theory: Morse-Smale systems (a significant subset of all structurally stable systems [13]) are robust to small C^1 perturbations, enabling infinite-time bounds. For limit cycles, we additionally require *exact period matching* via a localized correction procedure. Continuous attractors are not structurally stable, so we *tile* them with hyperbolic attractors spaced closer than ε and approximate the resulting Morse-Smale system.

Our analysis adopts a learning-theoretic framework: we define **target classes** $\mathcal{F}$ of dynamical systems to be approximated, a **hypothesis class** $\hat{\mathcal{F}}$ of Neural ODEs (Definition 1), and an **approximation criterion** (ε-δ closeness, Definition 9). Universal approximation means: for any $f \in \mathcal{F}$ and any $\varepsilon, \delta > 0$, there exists $\hat{f} \in \hat{\mathcal{F}}$ achieving ε-δ closeness over infinite time.

1.1 Main Results (Informal)

We establish universal approximation for three target classes. Throughout, ε-δ closeness means: the volume of initial conditions with trajectory error exceeding ε is less than δ (Definition 9).

> **Theorem FP** (Fixed Points, Informal; see Theorem 3). *For any Morse-Smale system with hyperbolic fixed points and any $\varepsilon, \delta > 0$, there exists a finite-size Neural ODE that is ε-δ close to the target for all $t \in [0, \infty)$.*

> **Theorem LC** (Limit Cycles, Informal; see Theorem 4). *For Morse-Smale systems with hyperbolic limit cycles, Neural ODEs achieve ε-δ closeness via exact period matching through localized vector field scaling.*

> **Theorem CA** (Continuous Attractors, Informal; see Theorem 5). *For systems with normally hyperbolic continuous attractors (line attractors, ring attractors, isochronous cylinders), Neural ODEs achieve ε-δ closeness via tiling: approximation by a dense grid of discrete attractors with spacing $< \varepsilon$.*

Finally, we establish a temporal generalization bound to bridge these guarantees to training metrics:

> **Theorem TG** (Temporal Generalization, Informal; see Theorem 6). *If $\hat{\varphi}$ is ε-δ close to φ, then the time-averaged L^p error satisfies $\mathcal{E}_{p,\infty} \leq \varepsilon^p + \delta \cdot D^p$, where D is the domain diameter. This bridges topological guarantees to practical training metrics like Mean Squared Error (MSE).*

1.2 Contributions

1. We establish **universal approximation results for multistable dynamical systems over infinite time horizons** (Theorems 3–5).
2. We prove that for each (ε, δ) a **Neural ODE of finite width** suffices (Section 4), while no single width suffices for all pairs.
3. We introduce **exact period matching** via localized vector field scaling to eliminate P-type error for limit cycles (Theorem 4).
4. We derive a **temporal generalization bound** linking ε-δ closeness to L^p error (Theorem 6).

Figure 1 illustrates the landscape of universal approximation results. Prior infinite-time theories were limited to fading memory systems (a single global attractor). Our results extend to Morse-Smale systems and to the normally hyperbolic continuous attractors of $\mathcal{F}_{\mathrm{CA}}$ [166].

Scope. Our theorems cover Morse-Smale systems (finitely many hyperbolic fixed points and periodic orbits) and continuous attractors that are normally hyperbolic manifolds of fixed points or

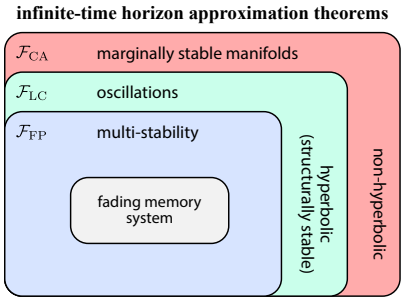


Figure 1: The landscape of universal approximation for dynamical systems. Existing infinite-time results require the fading memory property (FMP), excluding multistability; FMP systems with a hyperbolic equilibrium lie in $\mathcal{F}_{\mathrm{FP}}$. We prove Neural ODEs are dense in Morse-Smale systems ($\mathcal{F}_{\mathrm{FP}}$, $\mathcal{F}_{\mathrm{LC}}$) and normally hyperbolic continuous attractors ($\mathcal{F}_{\mathrm{CA}}$), subject to isochrony for oscillatory manifolds.

of isochronous limit cycles. They exclude chaotic targets, for which a positive Lyapunov exponent makes the infinite-horizon bound vacuous. We do not claim Morse-Smale systems are generic; they are open and dense in dimension two [154], but not dense in dimension three and higher [175]. Our proofs rely on openness rather than density: a learned field that is Morse-Smale keeps its qualitative dynamics under weight perturbation, which is what makes the guarantee meaningful. Our third target class, continuous attractors, is structurally unstable and a genericity argument would exclude it; we include it because it is a computational primitive, and Theorem 5 shows it remains approximable.

1.3 Related Work

The theoretical capability of neural networks to model dynamical systems is well-established, but existing results largely bifurcate into two regimes: finite-time approximation for general systems, and infinite-time approximation for restricted classes of stable systems.

Universal approximation for RNNs on finite time horizons was established by Funahashi and Nakamura [60] and Li [119], however, these guarantees rely on Grönwall inequalities where error bounds grow exponentially with time, rendering them vacuous for long-term behavior [181]. This lineage was expanded by Kimura and Nakano [105] and proven for various architectures by Chow and Li [32], extending to time-variant systems with fixed initial states [122]. This limitation is universal across architectures: it applies to discrete-time recurrence [2,100], time-variant systems [122], and Neural ODEs [24]. While Neural ODEs are universal approximators for homeomorphisms [211], flow maps [120], and functions on manifolds via geometric controllability [51], these results only approximate the diffeomorphism ϕ_t for a fixed t, providing no guarantee of topological equivalence or attractor preservation as $t \to \infty$. Recent work enforces exact fixed-point constraints for provable universality [148], but does not address the infinite-horizon trajectory guarantees or the limit-cycle and continuous-attractor cases we treat.

For infinite time, existing results require the *fading memory property* [14], which implies convergence to a unique equilibrium and excludes multistability. Echo State Networks [96] and reservoir computing [70] achieve infinite-time universality only for fading memory systems. The fading memory property ensures that the influence of past inputs decays asymptotically, allowing for infinite-time approximation of filters [66] and State Space Models [200]. Similarly, Nakamura and Nakagawa [144] extended approximation to infinite horizons only for systems with a globally asymptotically stable equilibrium. Hanson and Raginsky [76] provided infinite-time results based on uniform asymptotic incremental stability, a condition that forces all trajectories to converge to each other. Finally, while Hart et al. [77] showed that ESNs can form attracting submanifolds topologically conjugate to structurally stable diffeomorphisms, their result is restricted to discrete-time maps and requires embedding the dynamics into a high-dimensional reservoir state space. Recent Neural ODE architectures instead build in convergence by construction [137], but the guaranteed convergence is again to a single stable equilibrium, precluding the multistability our target classes require. Explicit universal approximation guarantees for multistable continuous-time flows in their native state space are lacking.

Alternatively, operator-theoretic approaches attempt to globalize dynamics by lifting the state to an infinite-dimensional space where evolution becomes linear. Koopman operator theory [109,139] lifts dynamics to a space of observables, while Carleman linearization embeds polynomial nonlinearities into an infinite system of moments [19,57]. While finite-dimensional truncations like Extended Dynamic Mode Decomposition (EDMD) [203] can capture global features, they theoretically require infinite width to eliminate closure error and generally lack rigorous uniform trajectory guarantees [16].

Koopman eigenfunctions have also been used to identify the separatrices themselves [38], the measure-zero loci where our B-type error concentrates.

Our work extends these results to infinite time for structurally stable systems. For a comprehensive review of the literature on universal approximation for dynamical systems, including finite-time results, fading memory systems, and reservoir computing, see App. G.

Remark 1 (Direct vs. embedding approximation). Finite-dimensional RNNs face a structural limitation: they cannot express an arbitrary number of attractors without increasing dimension [60,92,93,206]. Consequently, RNN universality requires *embedding* the target into a higher-dimensional latent space. In contrast, Neural ODEs enable *direct* approximation: the learned vector field operates in the same state space as the target, preserving the original coordinates and attractor geometry. This distinction is crucial for mechanistic interpretability of learned dynamics.

2 Preliminaries

Let $\mathcal{X} \subset \mathbb{R}^n$ be a bounded open set. We denote by $\mathfrak{X}^1(\mathcal{X})$ the space of C^1 vector fields $f : \mathcal{X} \to \mathbb{R}^n$ whose trajectories remain in $\mathcal{X}$ for all $t \geq 0$. The **flow** $\varphi : \mathbb{R}_{\geq 0} \times \mathcal{X} \to \mathcal{X}$ satisfies $\frac{d}{dt}\varphi(t, x_0) = f(\varphi(t, x_0))$ with $\varphi(0, x_0) = x_0$. We identify each system $\dot{x} = f(x)$ with its vector field f. Standard definitions from dynamical systems theory (ω-limit sets, non-wandering sets, transversal intersections, separatrices, tubular neighborhoods) are given in App. A.

Definition 1 (Hypothesis class with C^1 UAP)**.** A class $\hat{\mathcal{F}}$ of C^1 functions $\hat{f} : \mathcal{X} \to \mathbb{R}^n$ has the C^1 **universal approximation property** if for every $f \in C^1(\mathcal{X}, \mathbb{R}^n)$ and every $\eta > 0$, there exists $\hat{f} \in \hat{\mathcal{F}}$ with $\|f - \hat{f}\|_{C^1} < \eta$, where $\|g\|_{C^1} \coloneqq \sup_x \|g(x)\| + \sup_x \|Dg(x)\|_{\mathrm{op}}$ controls both function values and derivatives.

Remark 2. Standard feedforward networks with smooth activations (e.g., tanh, sigmoid) achieve C^r approximation for any $r \geq 0$ on compact domains [87], with explicit width and depth bounds for the joint function-and-derivative (C^1) error in the tanh case [40]. The property is not tied to one activation: periodic-activation networks [174], Kolmogorov-Arnold spline networks [128], and piecewise-polynomial (ReLU^k) networks that approximate a function together with its derivatives [8] all satisfy Definition 1; even plain ReLU networks qualify after mollification.

Definition 2 (Neural ODE)**.** Let $\{f_\theta\}_{\theta \in \Theta}$ be a family of C^1 vector fields $f_\theta : \mathcal{X} \to \mathbb{R}^n$ computed by a neural network architecture, with θ ranging over the weights and biases of all widths. A **Neural ODE** is the dynamical system $\dot{x} = f_\theta(x)$ for some $\theta \in \Theta$, and $\hat{\mathcal{F}} = \{f_\theta\}_{\theta \in \Theta}$ is its hypothesis class.

Definition 3 (Hyperbolic fixed point)**.** A fixed point x^* of $\dot{x} = f(x)$ is **hyperbolic** if all eigenvalues of the Jacobian $Df(x^*)$ have nonzero real part.

Definition 4 (Hyperbolic periodic orbit)**.** A periodic orbit γ with period $T > 0$ is **hyperbolic** if no eigenvalue of the Poincaré return map (Definition 24) on a transverse section has modulus 1.

Definition 5 (Morse-Smale system)**.** A C^1 vector field f is **Morse-Smale** if: (1) the non-wandering set $\Omega(f)$ is a finite collection of hyperbolic fixed points and hyperbolic periodic orbits, where x is *non-wandering* if for every neighborhood U of x and every $T > 0$ there exists $t > T$ with $\varphi(t, U) \cap U \neq \emptyset$; and (2) all stable and unstable manifolds intersect *transversally*: at every intersection point p, the tangent spaces satisfy $T_pW^s + T_pW^u = \mathbb{R}^n$.

Theorem 1 (Palis-Smale [149])**.** *Every Morse-Smale system is structurally stable: small C^1 perturbations yield topologically equivalent flows.*

Remark 3 (Morse-Smale systems). Morse-Smale systems are "generic well-behaved" dynamics: all trajectories converge to finitely many stable fixed points or limit cycles, with no degenerate connections (homoclinic tangencies, heteroclinic cycles) that would cause bifurcations. Condition (1) rules out chaotic attractors; condition (2) ensures structural stability. Morse-Smale systems are not the only structurally stable class (Axiom A systems [176] with strange attractors are also structurally stable), but our trajectory-tracking analysis does not extend to chaotic dynamics, where sensitive dependence makes trajectory-level identifiability fundamentally limited [61] and data-driven Neural ODEs instead recover the attractor's statistics rather than individual trajectories [29,126].

Theorem 2 (Stable Manifold Theorem [30])**.** *Let x^* be a hyperbolic fixed point of a C^1 vector field f with k stable eigenvalues (negative real part) and $n - k$ unstable eigenvalues. Then:*

1. *The **stable manifold** $W^s(x^*) = \{x : \varphi(t, x) \to x^* \text{ as } t \to +\infty\}$ is a C^1 embedded submanifold of dimension k.*

2. *The **unstable manifold** $W^u(x^*)$ is a C^1 embedded submanifold of dimension $n - k$.*
3. *Trajectories in $W^s(x^*)$ converge to x^* exponentially.*

Remark 4. For saddle points ($0 < k < n$), the stable manifold has dimension $< n$, hence Lebesgue measure zero. Since basin boundaries in Morse-Smale systems are finite unions of stable manifolds of saddles, they have measure zero, a key fact for controlling B-type error.

Definition 6 (Normally hyperbolic invariant manifold)**.** A compact invariant manifold $\mathcal{M}$ is **normally hyperbolic** if the linearized flow contracts/expands transverse to $\mathcal{M}$ at exponential rates dominating any expansion/contraction tangent to $\mathcal{M}$[55].

Definition 7 (Strictly inward-pointing)**.** Let $\mathcal{X} \subset \mathbb{R}^n$ have smooth boundary $\partial\mathcal{X}$ with outward unit normal $\nu(x)$. A vector field f is **strictly inward-pointing** at $\partial\mathcal{X}$ if there exists $\gamma > 0$ such that $f(x) \cdot \nu(x) < -\gamma$ for all $x \in \partial\mathcal{X}$.

Remark 5. The strictly inward-pointing condition ensures that any C^1-close approximation $\hat{f}$ with $\|f - \hat{f}\|_{C^0} < \gamma$ automatically satisfies $\hat{f}(x) \cdot \nu(x) < 0$ on $\partial\mathcal{X}$, so trajectories of $\hat{f}$ remain in $\mathcal{X}$ (Lemma 13). This can always be achieved by choosing $\mathcal{X}$ appropriately (e.g., a sublevel set of a Lyapunov-like function).

3 Approximation Metric

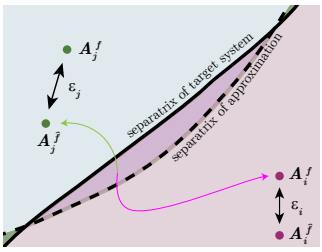


Figure 2: Basin mismatch (B-type error) in a bistable system. The target has stable fixed points A_i^f and A_j^f separated by a separatrix (solid). The approximation has nearby fixed points with a shifted separatrix (dashed). Initial conditions in the shaded region converge to different attractors in target vs. approximation, controlled by the δ parameter. Near separatrices, perfect trajectory tracking is impossible; the δ term captures this unavoidable failure region.

Standard approximation metrics (uniform norm, L^p norms) are unsuitable for infinite-time dynamical systems: near separatrices, trajectories can diverge to different attractors regardless of vector field closeness. We adopt an ε-δ framework that decouples precision from reliability.

A distinct class of approximation errors, which we term *basin errors*, arises when the target and the approximation assign the same initial condition x_0 to different attractors (i.e., $x_0 \in \mathrm{BoA}(A_i^f)$ but $x_0 \in \mathrm{BoA}(A_j^{\hat{f}})$ for $i \neq j$). For such points, the trajectory error does not remain bounded by a small ε but converges to the distance between the two distinct attractors $\|A_i - A_j\|$, see Fig. 2. Our metric accepts this inherent limitation via the δ parameter.

Definition 8 (ε-volume error)**.** For flows $\varphi, \hat{\varphi} : \mathbb{R}_{\geq 0} \times \mathcal{X} \to \mathcal{X}$, the **$\varepsilon$-volume error** is:

$$\|\varphi - \hat{\varphi}\|_\varepsilon := \frac{1}{\mathrm{vol}(\mathcal{X})} \int_{\mathcal{X}} \mathbb{1}\left[\sup_{t \geq 0} \|\varphi(t, x_0) - \hat{\varphi}(t, x_0)\| \geq \varepsilon\right] dx_0. \tag{1}$$

Definition 9 (ε-δ closeness)**.** Flows φ and $\hat{\varphi}$ are **ε-δ close** if $\|\varphi - \hat{\varphi}\|_\varepsilon < \delta$.

Viewed as a function of ε, $\delta(\varepsilon) := \|\varphi - \hat{\varphi}\|_\varepsilon$ is the tail of the worst-case error $S(x_0) = \sup_{t \geq 0} \|\varphi(t, x_0) - \hat{\varphi}(t, x_0)\|$ under the uniform measure on initial conditions, and the flows are ε-δ close when this curve passes below (ε, δ). Its crossing with the diagonal is the Ky Fan distance (Remark 15), and the pair (ε, δ) plays the role of accuracy and confidence in PAC learning[193], here over initial conditions; for large ε, δ measures basin mismatch, the model-versus-target analogue of basin stability[138].

This (ε, δ)-closeness is more than a practical heuristic; it induces a rigorous topology on the space of dynamical systems. Specifically, it corresponds to the topology of *convergence in measure* with respect to the trajectory supremum norm, which is metrizable via the Ky Fan metric. We provide the topological proofs and formal basis properties in App. B.

4 Main Results

4.1 Target Classes

Definition 10 (Target class $\mathcal{F}_{\mathrm{FP}}$)**.** $\mathcal{F}_{\mathrm{FP}} \subset \mathfrak{X}^1(\mathcal{X})$ is the set of Morse-Smale vector fields whose ω-limit sets (Definition 17) consist solely of hyperbolic fixed points.

Definition 11 (Target class $\mathcal{F}_{\mathrm{LC}}$)**.** $\mathcal{F}_{\mathrm{LC}} \subset \mathfrak{X}^1(\mathcal{X})$ is the set of Morse-Smale vector fields with at least one hyperbolic limit cycle attractor (Definition 18).

Definition 12 (Isochronous manifold)**.** A normally hyperbolic invariant manifold $\mathcal{M}$ foliated by periodic orbits $\{\gamma_\theta\}_{\theta\in\Theta}$ is **isochronous** if all orbits have the same period: $T(\gamma_\theta) = T$ for all $\theta \in \Theta$.

Definition 13 (Target class $\mathcal{F}_{\mathrm{CA}}$)**.** $\mathcal{F}_{\mathrm{CA}} \subset \mathfrak{X}^1(\mathcal{X})$ is the set of vector fields whose ω-limit sets are: (i) hyperbolic fixed points, (ii) hyperbolic periodic orbits, (iii) normally hyperbolic manifolds of fixed points, or (iv) isochronous normally hyperbolic manifolds of limit cycles (Definition 12).

Remark 6 (Isochrony requirement). Condition (iv) prevents P-type error from compounding D-type error. Tiling an oscillatory manifold by discrete limit cycles makes each tile inherit its local period, so a frequency gradient (periods varying along the attractor) leaves neighboring tiles mismatched and drives unbounded phase drift as $t \to \infty$; isochrony lets every tile be retuned to a common period through the correction of Theorem 4. For example, a cylinder $S^1 \times [0, 1]$ whose orbit at height z has period $T(z)$ is isochronous only when $T(z)$ is constant.

4.2 Universal Approximation for Fixed Points

Theorem 3 (UAP for $\mathcal{F}_{\mathrm{FP}}$)**.** *Let $f \in \mathcal{F}_{\mathrm{FP}}$ with flow $\varphi : \mathbb{R}_{\geq 0} \times \mathcal{X} \to \mathcal{X}$, and let $\hat{\mathcal{F}}$ be a hypothesis class with the C^1 universal approximation property (Definition 1). Assume f is strictly inward-pointing at $\partial\mathcal{X}$ (Definition 7).*

For all $\varepsilon > 0$ and all $\delta > 0$, there exists $\hat{f} \in \hat{\mathcal{F}}$ with flow $\hat{\varphi}$ such that: $\|\varphi - \hat{\varphi}\|_\varepsilon < \delta$.

Proof Sketch. The proof proceeds in three steps (full proof in App. C):

Step 1: Structural stability radius. By Theorem 1, there exists $\eta_0 > 0$ such that $\|f - \hat{f}\|_{C^1} < \eta_0$ implies topological equivalence via homeomorphism h with $\|h - \mathrm{id}\|_{C^0} \to 0$ as $\eta_0 \to 0$.

Step 2: Basin error control. Separatrices S have measure zero (codim-≥ 1 saddle stable manifolds), so $\mu(N_\eta(S)) \to 0$ as $\eta \to 0$. Pick $\eta > 0$ with $\mu(N_\eta(S)) < \delta$; by Lemma 12, $\mathcal{E}_{\mathrm{basin}} \subset N_\eta(S)$ for $\|f - \hat{f}\|_{C^1}$ small, so $\mu(\mathcal{E}_{\mathrm{basin}}) < \delta$.

Step 3: Trajectory bounds in valid set. Define $\mathcal{V} \coloneqq \mathcal{X} \setminus N_\eta(S)$, depending on f, η alone. By Lemma 11, $T^*(\eta) < \infty$ and the constants K_1, K are determined a priori. For $x_0 \in \mathcal{V}$, with $D\varphi_{s,t}$ the flow Jacobian:

$$\|\varphi(t, x_0) - \hat{\varphi}(t, x_0)\| \leq \int_0^t \|D\varphi_{s,t}\| \cdot \|f - \hat{f}\|_{C^0}\, ds \leq \frac{K\nu}{\lambda},$$

where $\lambda > 0$ is the slowest contraction rate (smallest $|\mathrm{Re}(\mu)|$ for stable eigenvalues μ at attractors) and $K \geq 1$ bounds transient growth before trajectories enter the linearization regime. Exponential contraction near hyperbolic attractors makes the integral converge, giving a bound uniform in $t \geq 0$.

By UAP, choose $\hat{f} \in \hat{\mathcal{F}}$ with $\|f - \hat{f}\|_{C^1} < \min(\eta, \lambda\varepsilon/K)$ to achieve $\|\varphi - \hat{\varphi}\|_\varepsilon < \delta$. □

Remark 7 (Interpretation). Theorem 3 states that multistable systems with multiple stable equilibria can be approximated with arbitrary precision (ε) and reliability ($1 - \delta$) over infinite time. The key insight is that structural stability confines basin errors to a thin layer near separatrices, and exponential contraction near attractors bounds trajectory error uniformly in time.

Remark 8 (Basin error scaling). For Morse-Smale systems with codimension-1 separatrices, the basin error volume scales linearly with the C^1 perturbation: $\mu(\mathcal{E}_{\mathrm{basin}}) \leq C_S \cdot \|f - \hat{f}\|_{C^1}$, where C_S depends on the surface area of the separatrices and the transversality of the flow. This allows choosing $\eta_{\mathrm{vol}} = \delta/C_S$ to satisfy the reliability constraint.

Remark 9 (Quantitative scaling). The required C^1 approximation accuracy η_0 scales as: $\eta_0 = O(\min\{\delta, \lambda/\varepsilon K\})$, where λ is the slowest contraction rate at attractors, and $K \geq 1$ bounds transient growth.

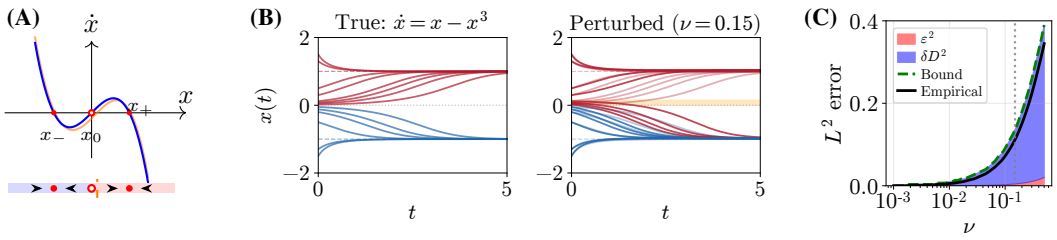


Figure 3: The 1D bistable system (Example 1). **(A)** Vector field $f(x) = x - x^3$ with stable $x_\pm = \pm 1$ (filled) and unstable $x_0 = 0$ (open); phase line shows basins with perturbed separatrix (dashed orange). **(B)** Trajectories from various initial conditions; orange region marks ICs that switch basins under perturbation $\hat{f} = f - \nu \cos(2x)$. **(C)** Time-averaged L^2 error vs. perturbation size ν: empirical error (black) stays below bound $\varepsilon^2 + \delta D^2$ (dashed, Theorem 6), decomposed into trajectory error (red) and B-type basin error (blue).

Example 1 (1D bistable system)**.** Consider the double-well system $\dot{x} = f(x) = x - x^3$ on $\mathcal{X} = [-2, 2]$ (Figure 3). This has stable fixed points at $x_\pm = \pm 1$ and an unstable fixed point (separatrix) at $x_0 = 0$. The basins are $\mathrm{BoA}(x_-) = [-2, 0)$ and $\mathrm{BoA}(x_+) = (0, 2]$. **Parameters:** The Jacobian $f'(x) = 1 - 3x^2$ gives $f'(\pm 1) = -2$, so $\lambda = 2$. The Lipschitz constant on $\mathcal{X}$ is $L = \max |f'| = \max |1 - 3x^2| = 11$ (at $x = \pm 2$).

Approximation: Let $\hat{f}(x) = x - x^3 + \nu g(x)$ for some bounded perturbation g with $\|g\|_{C^1} \leq 1$. The perturbed separatrix shifts by $O(\nu)$: if $\hat{x}_0$ satisfies $\hat{f}(\hat{x}_0) = 0$, then $|\hat{x}_0| \leq C\nu$ for small ν.

Error bounds: For the valid set $\mathcal{V} = \mathcal{X} \setminus (-\eta, \eta)$ with $\eta > C\nu$: 1. basin error: $\mu(\mathcal{E}_{\text{basin}}) \leq 2\eta$ (the strip around $x = 0$) and 2. trajectory error: $\sup_t \|\varphi - \hat{\varphi}\| \leq \nu \left(\frac{e^{LT^*} - 1}{L} + \frac{C}{\lambda} \right)$ (cf. Lemma 11, with $C \approx 1$ in 1D). Choosing $\eta = \delta/2$ and $\nu = \min(\eta/C, \lambda\varepsilon/K)$ achieves ε-δ closeness.

4.3 Universal Approximation for Limit Cycles

Theorem 4 (UAP for $\mathcal{F}_{\text{LC}}$)**.** *Let $f \in \mathcal{F}_{\text{LC}}$ with hyperbolic limit cycles $\gamma_1, \ldots, \gamma_k$ having periods $T_1, \ldots, T_k > 0$. Assume f is strictly inward-pointing at $\partial\mathcal{X}$. Let $\hat{\mathcal{F}}$ be a hypothesis class with the C^1 universal approximation property satisfying:*

- *For $k = 1$: $\hat{\mathcal{F}}$ is closed under scalar multiplication (always satisfied by neural networks).*
- *For $k \in \mathbb{N}$: the cycles are well-separated ($\min_{i \neq j} \operatorname{dist}(\gamma_i, \gamma_j) > 0$). The localized period correction is then realized under the C^1 UAP alone (Remark 11).*

For all $\varepsilon > 0$ and all $\delta > 0$, there exists $\hat{f} \in \hat{\mathcal{F}}$ with flow $\hat{\varphi}$ such that: $\|\varphi - \hat{\varphi}\|_\varepsilon < \delta$.

Proof Sketch. The key challenge is P-type error: even with $\|f - \hat{f}\|_{C^1}$ small, period mismatch causes unbounded phase drift. We eliminate this via exact period matching (full proof in App. D).

Step 1: Base approximation. Choose $\tilde{f} \in \hat{\mathcal{F}}$ with $\|f - \tilde{f}\|_{C^1} < \eta$. By Fenichel's persistence theorem, $\tilde{f}$ has limit cycles $\tilde{\gamma}_i$ with periods $\tilde{T}_i$ satisfying $|\tilde{T}_i - T_i| \leq L_T \eta$.

Step 2: Period correction. For a single cycle, global scaling $\hat{f} = c^* \tilde{f}$ with $c^* = \tilde{T}/T$ achieves exact period matching: the cycle $\hat{\gamma} = \tilde{\gamma}$ has period exactly T.

For multiple cycles with different correction factors, use localized bump functions ψ_i supported in disjoint tubular neighborhoods N_i of radius $r < \frac{1}{2} \min_{i \neq j} \operatorname{dist}(\gamma_i, \gamma_j)$, with $\psi_i = 1$ on $\tilde{\gamma}_i$. Define $\hat{f} = (1 + \sum_i \alpha_i^* \psi_i) \tilde{f}$ where $\alpha_i^* = (\tilde{T}_i - T_i)/T_i$. The separation condition ensures the supports are disjoint, so each cycle is corrected independently.

Step 3: Infinite-time bound. With exact period matching, the transient phase offset accumulated during $[0, T^*]$ is *frozen* for all $t > T^*$. The asymptotic error is bounded by geometric proximity $d_H(\gamma, \hat{\gamma}) + O(\|f - \hat{f}\|_{C^1})$, which can be made $< \varepsilon$. □

Remark 10 (Interpretation). Theorem 4 shows that limit cycle systems require a two-stage approximation: first C^1-approximate the vector field, then correct the period exactly. Without period correction, even perfect C^1 approximation yields unbounded phase drift as $t \to \infty$ (P-type error). The bump function construction localizes corrections to avoid interference between cycles.

Remark 11 (Localized period correction: realizability). For $k > 1$ limit cycles (or isochronous manifolds in Theorem 5), the period correction is localized to disjoint tubular neighborhoods N_i of the cycles. It suffices that $\hat{\mathcal{F}}$ contain localized corrections $\hat{\mathbf{\Phi}}_i$ that are C^1-close to the ideal fields (equal to $\tilde{f}$ on N_i and vanishing outside a slightly larger neighborhood): by Lemma 18 (Bump Function Realizability), the C^1 UAP provides such $\hat{\mathbf{\Phi}}_i \in \hat{\mathcal{F}}$ to any tolerance $\min(\nu, \zeta)$ in C^1 norm, giving on-cycle alignment below ν on $\tilde{\gamma}_i$ and off-cycle leakage below ζ on $\tilde{\gamma}_j$ for $j \neq i$. Whenever these tolerances satisfy $\nu + (N-1)\,\zeta < \frac{1}{C_Z}$ with $C_Z = \max_i \sup_{t \in [0, \tilde{T}_i]} \|Z_i(t)\|$ (with Z_i the adjoint, i.e. phase-response, solutions), the period Jacobian is strictly diagonally dominant, hence invertible by Lemma 21 (Robust Invertibility), and the inverse function theorem yields the exact period correction. Smooth-activation networks meet the condition with sufficient width; compactly supported activations (windowed ReLU, RBF, B-spline) meet it exactly [7,34,46,47,52,79,82,97,159].

4.4 Universal Approximation for Continuous Attractors

Continuous attractors are not structurally stable: generic perturbations collapse the manifold into discrete points, causing **D-type error** (Discretization)[166]. We address this via a tiling strategy that accepts bounded D-type error.

Theorem 5 (UAP for $\mathcal{F}_{\mathrm{CA}}$)**.** *Let $f \in \mathcal{F}_{\mathrm{CA}}$ with normally hyperbolic continuous attractor $\mathcal{M}$ (Definition 6). Assume f is strictly inward-pointing at $\partial\mathcal{X}$. Let $\hat{\mathcal{F}}$ be a hypothesis class with the C^1 universal approximation property satisfying: (1) if $\mathcal{M}$ consists of fixed points: UAP suffices; (2) if $\mathcal{M}$ is foliated by isochronous limit cycles: the localized period correction is realized under the C^1 UAP alone (Remark 11); no separate condition on $\hat{\mathcal{F}}$ is needed. Then, for all $\varepsilon > 0$ and all $\delta > 0$, there exists $\hat{f} \in \hat{\mathcal{F}}$ with flow $\hat{\varphi}$ such that: $\|\varphi - \hat{\varphi}\|_\varepsilon < \delta$.*

Proof Sketch. Continuous attractors are not structurally stable: generic perturbations collapse the continuum into discrete attractors (D-type error). We use a **tiling strategy** (full proof in App. E).

Manifolds of fixed points: Construct an intermediate system f_{tile} with discrete stable fixed points $\{x_1, \ldots, x_k\} \subset \mathcal{M}$ separated by saddles, with spacing $< \varepsilon/2$. Since $f_{\mathrm{tile}} \in \mathcal{F}_{\mathrm{FP}}$, Theorem 3 applies. The maximum error is bounded by the tiling resolution.

Isochronous manifolds of limit cycles: Tile the continuous attractor with a grid of discrete stable limit cycles $\{\gamma_1, \ldots, \gamma_N\}$ covering $\mathcal{M}$ with spacing $< \varepsilon/2$. Apply Theorem 4 with period correction. The isochrony condition ensures P-type error does not compound D-type error. □

5 Temporal Generalization

A key practical implication is that ε-δ closeness implies bounded L^p error over infinite time, connecting topological guarantees to training metrics.

Definition 14 (Time-averaged L^p error)**.** For flows $\varphi, \hat{\varphi}$ on bounded domain $\mathcal{X}$:

$$\mathcal{E}_{p,\infty}(\varphi, \hat{\varphi}) := \frac{1}{\mathrm{vol}(\mathcal{X})} \int_{\mathcal{X}} \left(\limsup_{T \to \infty} \frac{1}{T} \int_0^T \|\varphi(t, x_0) - \hat{\varphi}(t, x_0)\|^p \, dt \right) dx_0.$$

Theorem 6 (L^p error bound)**.** *Let $\varphi, \hat{\varphi} : \mathbb{R}_{\geq 0} \times \mathcal{X} \to \mathcal{X}$ be continuous flows on bounded domain $\mathcal{X} \subset \mathbb{R}^n$ with diameter $D := \sup_{x,y \in \mathcal{X}} \|x - y\| < \infty$. If $\hat{\varphi}$ is ε-δ close to φ (Definition 9), then for all $p \geq 1$: $\mathcal{E}_{p,\infty}(\varphi, \hat{\varphi}) \leq \varepsilon^p + \delta \cdot D^p$.*

Proof. Partition $\mathcal{X}$ into the "good" set $G = \{x_0 : \sup_{t \geq 0} \|\varphi(t, x_0) - \hat{\varphi}(t, x_0)\| \leq \varepsilon\}$ and "bad" set $B = \mathcal{X} \setminus G$. By ε-δ closeness, $\mathrm{vol}(B) < \delta \cdot \mathrm{vol}(\mathcal{X})$. On G: error bounded by ε for all time, so time-average $\leq \varepsilon^p$. On B: error bounded by diameter D, so time-average $\leq D^p$. Summing: $\mathcal{E}_{p,\infty} \leq \frac{\mathrm{vol}(G)}{\mathrm{vol}(\mathcal{X})} \varepsilon^p + \frac{\mathrm{vol}(B)}{\mathrm{vol}(\mathcal{X})} D^p \leq \varepsilon^p + \delta \cdot D^p$. □

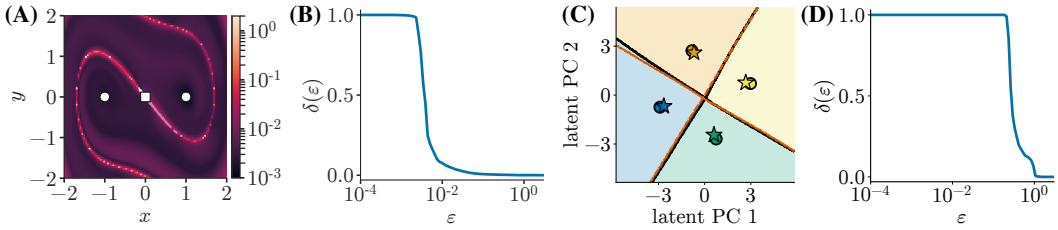


Figure 4: ε-δ closeness on a bistable system and on a real trained network. **(A)** Sup-error map for a Neural ODE fit to the damped Duffing system: the error concentrates on a thin collar around the separatrix (B-type); basin agreement 99.96%. **(B)** Its $\delta(\varepsilon)$ curve stays small ($\delta(0.1) = 0.007$), dominated by the separatrix collar. **(C)** Real versus fitted separatrices and fixed points on a 2D slice through four memory states: the trained network's basin boundaries (solid) and the Neural ODE's (dashed) coincide over 98% of the slice, and the network's fixed points (circles) sit close to the Neural ODE's (stars), so the approximation preserves the basin partition ($\mathcal{F}_{\mathrm{FP}}$), as the true and learned separatrices coincide in (A). **(D)** That Neural ODE's $\delta(\varepsilon)$ curve against the trained network, on the network's own readout (99.2% per-bit accuracy).

Corollary 7 (MSE bound)**.** *Under the assumptions of Theorem 6:* $\mathrm{MSE}_\infty(\varphi, \hat{\varphi}) \le \varepsilon^2 + \delta \cdot D^2$.

Remark 12. Theorem 6 establishes that topological guarantees (via ε-δ closeness) imply bounds on practical metrics. The converse is false: low MSE does not guarantee topological correctness, as a model can achieve low average error while failing to capture asymptotic stability or periodicity.

6 Numerical experiments

We validate the three guarantees and the basin / phase / discretization taxonomy on trained Neural ODEs and on real task-trained networks. App. F gives full details, additional architectures, and the budget and width scaling.[1]

ε-δ closeness and the failure modes. We illustrate the metric on the bistable Duffing system (Figure 4A,B). A width-64 Neural ODE reproduces both basins with 99.96% agreement; the trajectory sup-error is of order 10^{-3} across both basins and rises to the basin diameter only on a thin collar around the separatrix (B-type), so $\delta(\varepsilon)$ stays small ($\delta(0.1) = 0.007$). App. F illustrates the other two modes: on the Van der Pol oscillator a residual period error of 1.25×10^{-3} produces phase drift that reaches the cycle diameter near $t \approx 6 \times 10^3$, invisible to a 45-cycle validation window (P-type) and removed by the exact period matching of Theorem 4; and on a ring attractor the continuum discretizes into 10 alternating sinks and saddles while the ring itself persists (D-type).

Architecture generality and real networks. On a three-dimensional target with eight fixed points, a smooth-activation MLP, a C^1 RNN field, and a radial-basis-function field each recover all eight attractors with $\hat{\delta}(0.3) \le 6 \times 10^{-3}$, as Theorem 3 predicts for any C^1-dense class. Two real task-trained networks fall in the target classes: a gated recurrent network trained on the 3-bit flip-flop task [183,197] has stable fixed points at all eight memory states ($\mathcal{F}_{\mathrm{FP}}$), and a Neural ODE inherits all eight (autonomous basin agreement 99.8%) and reproduces the readout at 99.2% per-bit accuracy (Figure 4C,D); a rectified-tanh network trained to integrate angular velocity [166] holds heading on a ring that the finite network tiles into 12 hyperbolic sinks alternating with 12 saddles, a structurally stable approximation of the continuum whose fixed points a Neural ODE inherits ($\mathcal{F}_{\mathrm{CA}}$, D-type).

7 Discussion

Summary. We established universal approximation theorems for multistable dynamical systems over infinite time horizons. Our results apply to any Neural ODE architecture with the universal approximation property and provide explicit ε-δ guarantees linking vector field approximation to

[1] Code reproducing all experiments is available at `https://github.com/Asagodi/uap-infinite-time-code`.

trajectory bounds. For systems with stable fixed points ($\mathcal{F}_{\mathrm{FP}}$), trajectory errors decay to zero on the good set; for limit cycles ($\mathcal{F}_{\mathrm{LC}}$), errors stabilize at a small residual set by the cycle's geometric distance and phase mismatch. Finally, for continuous attractors ($\mathcal{F}_{\mathrm{CA}}$), we show that arbitrary precision is achievable via dense tiling of discrete attractors, overcoming their structural instability. Furthermore, we established a rigorous link between our topological ε-δ condition and the squared error loss used in training.

For finite-time approximation, almost any architecture suffices, including linear state-space models. However, their temporal generalization behavior can vary wildly, rendering finite-time guarantees inadequate for accurate, reliable long-time modeling of a system [110,183]. If we require the causal relations in a mechanism [17,18,103] to be captured by a vector field, then infinite-time horizon results provide for mechanistic interpretability guarantees at the state space level. This distinguishes Neural ODEs from finite-dimensional RNNs [74,146], which require embedding-type approximations that transform the state space and introduce extra state variables [197]; Neural ODEs instead enable direct approximation of the vector field in the original coordinates. This also makes XFADS [43] more expressive than LFADS [150,184], enabling topological reconstruction of multistable attractors.

Beyond existence proofs, our framework provides a practical taxonomy for analyzing model failure in computational neuroscience. By classifying errors into basin-type, period-type, and trajectory-type, researchers can diagnose specifically *why* a model fails to match a true biological implementation. For example, a Basin-type error corresponds directly to a failure in reliable memory storage, whereas Period-type errors reflect the fragility of oscillatory binding. Furthermore, this framework introduces a novel metric for the complexity of learning a computation, quantified by geometric properties such as the total size of separatrices required to define decision boundaries. This enables a more comprehensive and unified approach to comparing models of neural computation, replacing simple goodness-of-fit metrics with geometric evaluations of dynamic fidelity.

Limitations. *Learnability gap.* Our theorems guarantee existence of approximations, not their learnability. The loss landscape for infinite-horizon systems is prone to exploding gradients [161], and practical training requires specialized algorithms such as multiple shooting, teacher forcing [84] or homotopy-based training [107], along with strong inductive biases.

Fragility of period matching. Exact isochrony is not structurally stable: in the space of vector fields, those with a prescribed period T form a measure-zero set. Thus, while Neural ODEs are universal approximators for limit cycles, they are not robust implementations for tasks requiring infinite-time phase locking. To mitigate this, one can parametrize the period explicitly, separating attractor geometry from the flow speed on it [165].

Finite-time sensitivity. Our guarantee is asymptotic, constraining the basin and phase structure as $t \to \infty$; the finite-time error is instead set by how strongly the flow amplifies perturbations before settling. Even in Morse-Smale systems this amplification is largest on the B-type collar around separatrices, where finite-time Lyapunov exponents are large [170], and non-normal linearizations can add substantial transient growth, quantified by pseudospectra [192], without changing the asymptotic guarantee. Near saddles, the same sensitivity lets small inputs steer trajectories between basins [167].

Structurally unstable and chaotic systems. Our uniform convergence metric is ill-suited for chaotic systems due to sensitive dependence on initial conditions [85]. We term this **C-type error**: trajectories diverge exponentially even with accurate local approximation. Validating chaotic models therefore requires invariant set reconstruction: matching Lyapunov spectra [78], minimizing attractor Hausdorff distance [84], or shadowing [86]. *Quasiperiodic flows on tori* similarly suffer frequency locking under generic perturbations [72,73,152], while *center manifolds* and *homoclinic cycles* are determined by higher-order terms or precise manifold intersections.

Stochastic dynamics. Real-world systems are invariably subject to noise. Recent work has begun to establish universal approximation properties for stochastic reservoirs and filters [11,50,66], approximation guarantees for the distance between path measures [4,28] and flow matching techniques [127,141].

Non-autonomous systems. Extending infinite-time guarantees to input-driven dynamics remains open and requires pullback attractors rather than static invariant sets [106]. A continuous-time recurrent network can approximately realize any trajectory on a finite interval, and on an infinite one for periodic input [144]. Finite-time approximation of input-driven systems has been established for Neural ODEs [107,120,209] and neural flows [9].

Acknowledgments and Disclosure of Funding

This work was supported by NIH RF1-DA056404 and the Portuguese Recovery and Resilience Plan (PPR), through project number 62, Center for Responsible AI, and the Portuguese national funds, through FCT (Fundação para a Ciência e a Tecnologia) in the context of the project UIDB/04443/2020. The authors declare no competing interests.

References

[1] Andrei Agrachev and Andrey Sarychev. Control on the manifolds of mappings with a view to the deep learning. *Journal of Dynamical and Control Systems*, 28(4):989–1008, 2022.

[2] Miguel Aguiar, Amritam Das, and Karl H Johansson. Universal approximation of flows of control systems by recurrent neural networks. *arXiv preprint arXiv:2304.00352*, 2023.

[3] Mohammed Ashraf and Ayman A El-Badawy. Frequency-domain Neural ODEs for modeling non-linear dynamical systems. *arXiv preprint arXiv:2606.22075*, 2026.

[4] Julio Backhoff-Veraguas, Sigrid Källblad, and Benjamin A Robinson. Adapted Wasserstein distance between the laws of SDEs. *arXiv preprint arXiv:2209.03243*, 2022.

[5] Shaojie Bai. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271*, 2018.

[6] Shaojie Bai, J Zico Kolter, and Vladlen Koltun. Deep equilibrium models. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

[7] Jonathan T Barron. Continuously differentiable exponential linear units. *arXiv preprint arXiv:1704.07483*, 2017.

[8] Denis Belomestny, Alexey Naumov, Nikita Puchkin, and Sergey Samsonov. Simultaneous approximation of a smooth function and its derivatives by deep neural networks with piecewise-polynomial activations. *Neural Networks*, 161:242–253, 2023.

[9] Marin Biloš, Johanna Sommer, Syama Sundar Rangapuram, Tim Januschowski, and Stephan Günnemann. Neural flows: Efficient alternative to neural ODEs. *Advances in neural information processing systems*, 34:21325–21337, 2021.

[10] Adrian N Bishop. Universal time-uniform trajectory approximation for random dynamical systems with recurrent neural networks. *arXiv preprint arXiv:2211.08018*, 2022.

[11] Adrian N Bishop and Edwin V Bonilla. Recurrent neural networks and universal approximation of Bayesian filters. In *International Conference on Artificial Intelligence and Statistics*, pages 6956–6967. PMLR, 2023.

[12] Adrian N Bishop, Edwin V Bonilla, and Pierre Del Moral. Recurrent neural networks and (time-uniform) universal approximation of Bayesian filters. *IEEE Transactions on Automatic Control*, 2026.

[13] Christian Bonatti. Survey towards a global view of dynamical systems, for the C^1-topology. *Ergodic Theory and Dynamical Systems*, 31(4):959–993, 2011.

[14] Stephen Boyd and Leon Chua. Fading memory and the problem of approximating nonlinear operators with Volterra series. *IEEE Transactions on Circuits and Systems*, 32(11):1150–1161, 1985.

[15] Michael S Branicky. Universal computation and other capabilities of hybrid and continuous dynamical systems. *Theoretical computer science*, 138(1):67–100, 1995.

[16] Steven L Brunton, Marko Budišić, Eurika Kaiser, and J Nathan Kutz. Modern Koopman theory for dynamical systems. *arXiv preprint arXiv:2102.12086*, 2021.

[17] Rosa Cao and Daniel Yamins. Explanatory models in neuroscience: Part 1–taking mechanistic abstraction seriously. *arXiv preprint arXiv:2104.01490*, 2021.

[18] Rosa Cao and Daniel Yamins. Explanatory models in neuroscience: Part 2–constraint-based intelligibility. *arXiv preprint arXiv:2104.01489*, 2021.

[19] Torsten Carleman. Application de la théorie des équations intégrales linéaires aux systèmes d'équations différentielles non linéaires. *Acta Mathematica*, 59(1):63–87, 1932.

[20] Elena Celledoni, Brynjulf Owren, Chong Shen, Baige Xu, and Takaharu Yaguchi. UEPI: Universal energy-behavior-preserving integrators for energy conservative/dissipative differential equations. *Advances in Neural Information Processing Systems*, 38:43717–43751, 2026.

[21] Bo Chang, Minmin Chen, Eldad Haber, and Ed H Chi. AntisymmetricRNN: A dynamical system view on recurrent neural networks. *arXiv preprint arXiv:1902.09689*, 2019.

[22] Rishidev Chaudhuri, Berk Gerçek, Biraj Pandey, Adrien Peyrache, and Ila Fiete. The intrinsic attractor manifold and population dynamics of a canonical cognitive circuit across waking and sleep. *Nature neuroscience*, 22(9):1512–1520, 2019.

[23] F. C. Chen and H. K. Khalil. Adaptive control of nonlinear systems using neural networks. *International Journal of Control*, 55(6):1299–1317, 1992.

[24] Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. *Advances in neural information processing systems*, 31, 2018.

[25] Sheng Chen and Stephen A Billings. Neural networks for nonlinear dynamic system modelling and identification. *International Journal of Control*, 56(2):319–346, 1992.

[26] T. Chen and H. Chen. Approximations of continuous functionals by neural networks with application to dynamic systems. *IEEE Transactions on Neural Networks*, 4(6):910–918, 1993. doi: 10.1109/72.258503.

[27] T. Chen and H. Chen. Universal approximation to nonlinear operators by neural networks with arbitrary activation functions and its application to dynamical systems. *IEEE Transactions on Neural Networks*, 6(4):911–917, 1995. doi: 10.1109/72.388949.

[28] Xiuqiong Chen, Yangtianze Tao, Wenjie Xu, and Stephen Shing-Toung Yau. Recurrent neural networks are universal approximators with stochastic inputs. *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

[29] Anthony G. Chesebro, David Hofmann, Vaibhav Dixit, Earl K. Miller, Richard H. Granger, Alan Edelman, Christopher V. Rackauckas, Lilianne R. Mujica-Parodi, and Helmut H. Strey. Scientific machine learning of chaotic systems learns reduced-order equations for neural populations, 2026. URL `https://arxiv.org/abs/2507.03631`.

[30] Carmen Chicone. *Ordinary Differential Equations with Applications*. Springer, New York, NY, 2006.

[31] Ju-Yeop Choi, Hugh F Van Landingham, and Stanoje Bingulac. A constructive approach for nonlinear system identification using multilayer perceptrons. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 26(2):307–312, 1996.

[32] Tommy WS Chow and Xiao-Dong Li. Modeling of continuous time dynamical systems with input by recurrent neural networks. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, 47(4):575–578, 2000.

[33] L Chua and D Green. A qualitative analysis of the behavior of dynamic nonlinear networks: Steady-state solutions of nonautonomous networks. *IEEE Transactions on Circuits and Systems*, 23(9):530–550, 1976.

[34] Djork-Arné Clevert, Thomas Unterthiner, and Sepp Hochreiter. Fast and accurate deep network learning by exponential linear units (ELUs). In *International Conference on Learning Representations*, 2016.

[35] Michael A Cohen. The construction of arbitrary stable dynamics in nonlinear neural networks. *Neural Networks*, 5(1):83–103, 1992.

[36] Bernard D Coleman and Victor J Mizel. On the general theory of fading memory. *Archive for Rational Mechanics and Analysis*, 29:18–31, 1968.

[37] Christa Cuchiero, Martin Larsson, and Josef Teichmann. Deep neural networks, generic universal interpolation, and controlled ODEs. *SIAM Journal on Mathematics of Data Science*, 2(3):901–919, 2020.

[38] Kabir V Dabholkar and Omri Barak. Finding separatrices of dynamical flows with Deep Koopman Eigenfunctions. *arXiv preprint arXiv:2505.15231*, 2025.

[39] M. F. Danca and G. Chen. Approximation and decomposition of attractors of a hopfield neural network system. *Chaos, Solitons & Fractals*, 186:115213, 2024.

[40] Tim De Ryck, Samuel Lanthaler, and Siddhartha Mishra. On the approximation of functions by tanh neural networks. *Neural Networks*, 143:732–750, 2021.

[41] Mostafa Dehghani, Stephan Gouws, Oriol Vinyals, Jakob Uszkoreit, and Łukasz Kaiser. Universal transformers. *arXiv preprint arXiv:1807.03819*, 2018.

[42] Michael Dellnitz and Oliver Junge. On the approximation of complicated dynamical behavior. *SIAM Journal on Numerical Analysis*, 36(2):491–515, 1999.

[43] Matthew Dowling, Yuan Zhao, and Il Memming Park. eXponential FAmily dynamical systems (XFADS): Large-scale nonlinear gaussian state-space modeling. In *Advances in Neural Information Processing Systems (NeurIPS)*, December 2024. doi: 10.52202/079017-0430. URL https://openreview.net/forum?id=Ln8ogihZ2S.

[44] K. Doya. Universality of fully connected recurrent neural networks. Technical Report 1, Dept. of Biology, UCSD, 1993.

[45] Laura N Driscoll, Krishna Shenoy, and David Sussillo. Flexible multitask computation in recurrent networks utilizes shared dynamical motifs. *Nature Neuroscience*, 27(7):1349–1363, 2024.

[46] Shiv Ram Dubey, Satish Kumar Singh, and Bidyut Baran Chaudhuri. Activation functions in deep learning: A comprehensive survey and benchmark. *Neurocomputing*, 2022.

[47] Włodzisław Duch and Norbert Jankowski. Survey of neural transfer functions. *Neural computing surveys*, 2(1):163–212, 1999.

[48] Richard M. Dudley. *Real Analysis and Probability*, volume 74 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2 edition, 2002.

[49] Daniel Durstewitz, Georgia Koppe, and Max Ingo Thurm. Reconstructing computational system dynamics from neural data with recurrent neural networks. *Nature Reviews Neuroscience*, 24(11):693–710, 2023.

[50] Peter J Ehlers, Hendra I Nurdin, and Daniel Soh. Stochastic reservoir computers. *Nature Communications*, 16(1):1–11, 2025.

[51] Karthik Elamvazhuthi, Xuechen Zhang, Samet Oymak, and Fabio Pasqualetti. Learning on manifolds: Universal approximations properties using geometric controllability conditions for Neural ODEs. In *Learning for Dynamics and Control Conference*, pages 1–11. PMLR, 2023.

[52] Stefan Elfwing, Eiji Uchibe, and Kenji Doya. Sigmoid-weighted linear units for neural network function approximation in reinforcement learning. *Neural Networks*, 107:3–11, 2018.

[53] Ahmed ElGazzar and Marcel van Gerven. Universal differential equations as a common modeling language for neuroscience. *arXiv preprint arXiv:2403.14510*, 2024.

[54] Bard Ermentrout and David Hillel Terman. *Mathematical foundations of neuroscience*, volume 35. Springer, 2010.

[55] Neil Fenichel. Persistence and smoothness of invariant manifolds for flows. *Indiana University Mathematics Journal*, 21(3):193–226, 1971. ISSN 00222518, 19435258. URL http://www.jstor.org/stable/24890380.

[56] Michel Fliess. Vers une approche algébrique des systèmes non linéaires en temps discret. In A. Bensoussan and J. L. Lions, editors, *Analysis and Optimization of Systems*, pages 594–603, Berlin, Heidelberg, 1980. Springer Berlin Heidelberg. ISBN 978-3-540-38489-2.

[57] Marcelo Forets and Amaury Pouly. Explicit error bounds for Carleman linearization. *arXiv preprint arXiv:1711.02552*, 2017.

[58] Marcel O. Franz and Bernhard Schölkopf. A unifying view of Wiener and Volterra theory and polynomial kernel regression. *Neural Computation*, 18(12):3097–3118, 2006.

[59] Ken-Ichi Funahashi. On the approximate realization of continuous mappings by neural networks. *Neural Networks*, 2(3):183–192, 1989.

[60] Ken-ichi Funahashi and Yuichi Nakamura. Approximation of dynamical systems by continuous time recurrent neural networks. *Neural Networks*, 6(6):801–806, 1993.

[61] Matteo Gallo, Fabio Anselmi, and Paolo Lazzari. Attractor geometry determines the identifiability limits of system discovery. *arXiv preprint arXiv:2607.18490*, 2026.

[62] Freddy Rafael Garces, Victor Manuel Becerra, Chandrasekhar Kambhampati, and Kevin Warwick. *Strategies for feedback linearisation: A dynamic neural network approach*. Springer Science & Business Media, 2012.

[63] Antoine Girard and George J Pappas. Hierarchical control system design using approximate simulation. *Automatica*, 45(2):566–571, 2009.

[64] Antoine Girard, A Agung Julius, and George J Pappas. Approximate simulation relations for hybrid systems. *Discrete event dynamic systems*, 18(2):163–179, 2008.

[65] Matthew D Golub and David Sussillo. FixedPointFinder: A Tensorflow toolbox for identifying and characterizing fixed points in recurrent neural networks. *Journal of Open Source Software*, 3(31):1003, 2018.

[66] Lukas Gonon and Juan-Pablo Ortega. Reservoir computing universality with stochastic inputs. *IEEE Transactions on Neural Networks and Learning Systems*, 31(1):100–112, 2019.

[67] Lukas Gonon and Juan-Pablo Ortega. Fading memory echo state networks are universal. *Neural Networks*, 138:10–13, 2021.

[68] Lukas Gonon, Lyudmila Grigoryeva, and Juan-Pablo Ortega. Risk bounds for reservoir computing. *Journal of Machine Learning Research*, 21(240):1–61, 2020.

[69] Lukas Gonon, Lyudmila Grigoryeva, and Juan-Pablo Ortega. Approximation bounds for random neural networks and reservoir systems. *The Annals of Applied Probability*, 33(1): 28–69, 2023.

[70] Lyudmila Grigoryeva and Juan-Pablo Ortega. Echo state networks are universal. *Neural Networks*, 108:495–508, 2018.

[71] Lyudmila Grigoryeva and Juan-Pablo Ortega. Universal discrete-time reservoir computers with stochastic inputs and linear readouts using non-homogeneous state-affine systems. *Journal of Machine Learning Research*, 19(24):1–40, 2018.

[72] John Guckenheimer and Philip Holmes. *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*, volume 42 of *Applied Mathematical Sciences*. Springer, New York, 1983.

[73] John Guckenheimer and Philip Holmes. Structurally stable heteroclinic cycles. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 103, pages 189–192. Cambridge University Press, 1988.

[74] Mansura Habiba and Barak A Pearlmutter. Neural ordinary differential equation based recurrent neural network model. In *2020 31st Irish signals and systems conference (ISSC)*, pages 1–6. IEEE, 2020.

[75] Barbara Hammer. On the approximation capability of recurrent neural networks. *Neurocomputing*, 31(1-4):107–123, 2000.

[76] Joshua Hanson and Maxim Raginsky. Universal simulation of stable dynamical systems by recurrent neural nets. In *Learning for Dynamics and Control*, pages 384–392. PMLR, 2020.

[77] Allen Hart, James Hook, and Jonathan Dawes. Embedding and approximation theorems for echo state networks. *Neural Networks*, 128:234–247, 2020.

[78] Joseph D Hart. Attractor reconstruction with reservoir computers: The effect of the reservoir's conditional Lyapunov exponents on faithful attractor reconstruction. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 34(4), 2024.

[79] Md Mehedi Hasan, Md Ali Hossain, Azmain Yakin Srizon, and Abu Sayeed. Talu: A hybrid activation function combining tanh and rectified linear unit to enhance neural networks. *arXiv preprint arXiv:2305.04402*, 2023.

[80] Ramin M Hasani, Mathias Lechner, Alexander Amini, Daniela Rus, and Radu Grosu. Liquid time-constant recurrent neural networks as universal approximators. *arXiv preprint arXiv:1811.00321*, 2018.

[81] Shuhei Hayashi. Connecting invariant manifolds and the solution of the C^1 stability and Ω-stability conjectures for flows. *Annals of mathematics*, 145(1):81–137, 1997.

[82] Soufiane Hayou, Arnaud Doucet, and Judith Rousseau. On the impact of the activation function on deep neural networks training. In *International conference on machine learning*, pages 2672–2680. PMLR, 2019.

[83] Robert Hermann and Arthur Krener. Nonlinear controllability and observability. *IEEE Transactions on automatic control*, 22(5):728–740, 1977.

[84] F. Hess, Z. Monfared, M. Brenner, and D. Durstewitz. Generalized teacher forcing for learning chaotic dynamics. *arXiv preprint arXiv:2306.04406*, 2023.

[85] Morris W Hirsch and Bill Baird. Computing with dynamic attractors in neural networks. *Biosystems*, 34(1-3):173–195, 1995.

[86] Frank C Hoppensteadt. *Analysis and simulation of chaotic systems*, volume 94. Springer Science & Business Media, 2013.

[87] K. Hornik. Approximation capabilities of multilayer feedforward networks. *Neural Networks*, 4(2):251–257, 1991.

[88] Sen Hu. A proof of C^1 stability conjecture for three-dimensional flows. *Transactions of the American Mathematical Society*, 342(2):753–772, 1994.

[89] Chin-Wei Huang, David Krueger, Alexandre Lacoste, and Aaron Courville. Neural autoregressive flows. In *International conference on machine learning*, pages 2078–2087. PMLR, 2018.

[90] Zifeng Huang, Konstantin M Zuev, Yong Xia, and Michael Beer. Upper approximation bounds for neural oscillators. *arXiv preprint arXiv:2512.01015*, 2025.

[91] Yongkang Huo, Thomas Chaffey, and Rodolphe Sepulchre. Kernel modelling of fading memory systems. *arXiv preprint arXiv:2403.11945*, 2024.

[92] Sungmin Hwang, Viola Folli, Enrico Lanza, Giorgio Parisi, Giancarlo Ruocco, and Francesco Zamponi. On the number of limit cycles in asymmetric neural networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2019(5):053402, 2019.

[93] Sungmin Hwang, Enrico Lanza, Giorgio Parisi, Jacopo Rocchi, Giancarlo Ruocco, and Francesco Zamponi. On the number of limit cycles in diluted neural networks. *Journal of Statistical Physics*, 181:2304–2321, 2020.

[94] Piotr Indyk. Optimal simulation of automata by neural nets. In *Annual Symposium on Theoretical Aspects of Computer Science*, pages 337–348, Berlin, Heidelberg, March 1995. Springer Berlin Heidelberg.

[95] Isao Ishikawa, Takeshi Teshima, Koichi Tojo, Kenta Oono, Masahiro Ikeda, and Masashi Sugiyama. Universal approximation property of invertible neural networks. *Journal of Machine Learning Research*, 24(287):1–68, 2023.

[96] Herbert Jaeger. The "echo state" approach to analysing and training recurrent neural networks-with an erratum note. *Bonn, Germany: German National Research Center for Information Technology GMD Technical Report*, 148(34):13, 2001.

[97] Ameya D Jagtap and George Em Karniadakis. How important are activation functions in regression and classification? A survey, performance comparison, and future directions. *Journal of Machine Learning for Modeling and Computing*, 4(1), 2023.

[98] Priyank Jaini, Kira A Selby, and Yaoliang Yu. Sum-of-squares polynomial flow. In *International Conference on Machine Learning*, pages 3009–3018. PMLR, 2019.

[99] Haotian Jiang, Zhong Li, and Qianxiao Li. Approximation theory of convolutional architectures for time series modelling. In Marina Meila and Tong Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 4961–4970. PMLR, 18–24 Jul 2021. URL `https://proceedings.mlr.press/v139/jiang21d.html`.

[100] Liang Jin, Madan M Gupta, and Peter N Nikiforuk. Universal approximation using dynamic recurrent neural networks: Discrete-time version. In *Proceedings of ICNN'95-International Conference on Neural Networks*, volume 1, pages 403–408. IEEE, 1995.

[101] Anil Kag, Ziming Zhang, and Venkatesh Saligrama. RNNs incrementally evolving on an equilibrium manifold: A panacea for vanishing and exploding gradients? In *International Conference on Learning Representations*, 2020.

[102] Chandra Kambhampati, F Garces, and Kevin Warwick. Approximation of non-autonomous dynamic systems by continuous time recurrent neural networks. In *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium*, volume 1, pages 64–69. IEEE, 2000.

[103] David Michael Kaplan and Carl F Craver. The explanatory force of dynamical and mathematical models in neuroscience: A mechanistic perspective. *Philosophy of science*, 78(4): 601–627, 2011.

[104] Joe Kilian and Hava T Siegelmann. The dynamic universality of sigmoidal neural networks. *Information and computation*, 128(1):48–56, 1996.

[105] Masahiro Kimura and Ryohei Nakano. Learning dynamical systems by recurrent neural networks from orbits. *Neural Networks*, 11(9):1589–1599, 1998.

[106] Peter E. Kloeden and Martin Rasmussen. *Nonautonomous dynamical systems*. Number 176 in Mathematical Surveys and Monographs. American Mathematical Society, 2011.

[107] Joon-Hyuk Ko, Hankyul Koh, Nojun Park, and Wonho Jhe. Homotopy-based training of NeuralODEs for accurate dynamics discovery. *Advances in Neural Information Processing Systems*, 36:64725–64752, 2023.

[108] T. Konishi and Y. Kawahara. Stable invariant models via Koopman spectra. *Neural Networks*, 165:393–405, 2023.

[109] Bernard O Koopman. Hamiltonian systems and transformation in Hilbert space. *Proceedings of the National Academy of Sciences*, 17(5):315–318, 1931.

[110] Georgia Koppe, Hazem Toutounji, Peter Kirsch, Stefanie Lis, and Daniel Durstewitz. Identifying nonlinear dynamical systems via generative recurrent neural networks with applications to fMRI. *PLoS computational biology*, 15(8):e1007263, 2019.

[111] Samuel A Korsky and Robert C Berwick. On the computational power of RNNs. *arXiv preprint arXiv:1906.06349*, 2019.

[112] EB Kosmatopoulos and Manolis A Christodoulou. Structural properties of gradient recurrent high-order neural networks. *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, 42(9):592–603, 1995.

[113] Stefan C Kremer. On the computational power of Elman-style recurrent networks. *IEEE Transactions on Neural Networks*, 6(4):1000–1004, 1995.

[114] Christian Kuehn and Sara-Viola Kuntz. Embedding capabilities of Neural ODEs. *arXiv preprint arXiv:2308.01213*, 2023.

[115] Abhijeet Kumar, M Srinivas, and RBV Subramanyam. Time series analysis using memory enhanced liquid neural network. In *IFIP International Conference on Artificial Intelligence Applications and Innovations*, pages 330–341. Springer, 2025.

[116] Ivan Kupka. Contribution à la théorie des champs génériques. *Contributions to differential equations*, 2:457–484, 1963.

[117] Samuel Lanthaler, Konstantin Rusch, and Sascha Mishra. Neural oscillators are universal. *arXiv preprint arXiv:2305.08753*, 2023.

[118] Boyu Li, Robert Simon Fong, and Peter Tino. Simple cycle reservoirs are universal. *Journal of Machine Learning Research*, 25:1–28, 2024.

[119] Leong Kwan Li. Approximation theory and recurrent networks. In *[Proceedings 1992] IJCNN International Joint Conference on Neural Networks*, volume 2, pages 266–271. IEEE, 1992.

[120] Qianxiao Li, Ting Lin, and Zuowei Shen. Deep learning via dynamical systems: An approximation perspective. *Journal of the European Mathematical Society*, 25(5):1671–1709, 2022.

[121] W. Li and J. J. E. Slotine. Neural network control of unknown nonlinear systems. In *1989 American Control Conference*, pages 1136–1141. IEEE, June 1989.

[122] Xiao-Dong Li, John KL Ho, and Tommy WS Chow. Approximation of dynamical time-variant systems by continuous-time recurrent neural networks. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 52(10):656–660, 2005.

[123] Zhong Li, Jiequn Han, Weinan E, and Qianxiao Li. On the curse of memory in recurrent neural networks: Approximation and optimization analysis. In *International Conference on Learning Representations*, 2021. URL `https://openreview.net/forum?id=8Sqhl-nF50`.

[124] Zhong Li, Haotian Jiang, and Qianxiao Li. On the approximation properties of recurrent encoder-decoder architectures. In *International Conference on Learning Representations*, 2021.

[125] Zhong Li, Jiequn Han, E Weinan, and Qianxiao Li. Approximation and optimization theory for linear continuous-time recurrent neural networks. *Journal of Machine Learning Research*, 23(42):1–85, 2022.

[126] Alec J Linot and Michael D Graham. Data-driven reduced-order modeling of spatiotemporal chaos with neural ordinary differential equations. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 32(7), 2022.

[127] Yaron Lipman, Marton Havasi, Peter Holderrieth, Neta Shaul, Matt Le, Brian Karrer, Ricky TQ Chen, David Lopez-Paz, Heli Ben-Hamu, and Itai Gat. Flow matching guide and code. *arXiv preprint arXiv:2412.06264*, 2024.

[128] Ziming Liu, Yixuan Wang, Sachin Vaidya, Fabian Ruehle, James Halverson, Marin Soljačić, Thomas Y Hou, and Max Tegmark. KAN: Kolmogorov-Arnold Networks. *arXiv preprint arXiv:2404.19756*, 2024.

[129] Wolfgang Maass and Henry Markram. On the computational power of circuits of spiking neurons. *Journal of computer and system sciences*, 69(4):593–616, 2004.

[130] Wolfgang Maass, Prashant Joshi, and Eduardo D Sontag. Computational aspects of feedback in neural circuits. *PLoS computational biology*, 3(1):e165, 2007.

[131] I. G. Malkin. *Some Problems in the Theory of Nonlinear Oscillations*. U.S. Atomic Energy Commission, Technical Information Service, 1956. Translated from the Russian.

[132] Ricardo Mañé. A proof of the C^1 stability conjecture. *Publications Mathématiques de l'IHÉS*, 66:161–210, 1987.

[133] Gandhi Manjunath. Stability and memory-loss go hand-in-hand: Three results in dynamics and computation. *Proceedings of the Royal Society A*, 476(2242):20200563, 2020.

[134] Gandhi Manjunath and Herbert Jaeger. Echo state property linked to an input: Exploring a fundamental characteristic of recurrent neural networks. *Neural Computation*, 25(3):671–696, 2013.

[135] Stefano Massaroli, Michael Poli, Jinkyoo Park, Atsushi Yamashita, and Hajime Asama. Dissecting Neural ODEs. *Advances in Neural Information Processing Systems*, 33:3952–3963, 2020.

[136] Michael B Matthews. Approximating nonlinear fading-memory operators using neural network models. *Circuits, Systems and Signal Processing*, 12:279–307, 1993.

[137] Wenjie Mei, Dongzhe Zheng, and Shihua Li. ControlSynth neural ODEs: Modeling dynamical systems with guaranteed convergence. *Advances in Neural Information Processing Systems*, 37:99232–99261, 2024.

[138] Peter J Menck, Jobst Heitzig, Norbert Marwan, and Jürgen Kurths. How basin stability complements the linear-stability paradigm. *Nature physics*, 9(2):89–92, 2013.

[139] Igor Mezić. Spectral properties of dynamical systems, model reduction and decompositions. *Nonlinear Dynamics*, 41:309–325, 2005.

[140] John Miller and Moritz Hardt. Stable recurrent models. *arXiv preprint arXiv:1805.10369*, 2018.

[141] Gal Mishne, Zhengchao Wan, Qingsong Wang, and Yusu Wang. Elucidating flow matching ODE dynamics with respect to data geometries. *arXiv preprint arXiv:2412.18730*, 2024.

[142] Cristopher Moore. Unpredictability and undecidability in dynamical systems. *Physical Review Letters*, 64(20):2354, 1990.

[143] Cristopher Moore. Finite-dimensional analog computers: Flows, maps, and recurrent neural networks. In *1st International Conference on Unconventional Models of Computation-UMC*, volume 98, pages 59–71, 1998.

[144] Yuichi Nakamura and Masahiro Nakagawa. Approximation capability of continuous time recurrent neural networks for non-autonomous dynamical systems. In *Artificial Neural Networks–ICANN 2009: 19th International Conference, Limassol, Cyprus, September 14-17, 2009, Proceedings, Part II 19*, pages 593–602. Springer, 2009.

[145] S Kumpati Narendra and Kannan Parthasarathy. Identification and control of dynamical systems using neural networks. *IEEE Transactions on Neural Networks*, 1(1):4–27, 1990.

[146] Murphy Yuezhen Niu, Lior Horesh, and Isaac Chuang. Recurrent neural networks in the eye of differential equations. *arXiv preprint arXiv:1904.12933*, 2019.

[147] Antonio Orvieto, Soham De, Caglar Gulcehre, Razvan Pascanu, and Samuel L Smith. On the universality of linear recurrences followed by nonlinear projections. *arXiv preprint arXiv:2307.11888*, 2023.

[148] Feliciano Giuseppe Pacifico, Duccio Fanelli, Lorenzo Buffoni, Lorenzo Chicchi, Diego Febbe, and Raffaele Marino. Exact fixed-point constraints in Neural-ODEs with provable universality. *arXiv preprint arXiv:2605.10613*, 2026.

[149] Jacob Palis and Stephen Smale. Structural stability theorems. In *Global Analysis*, pages 223–231. Proc. Symp. Pure Math, 1970.

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

[151] Thomas Parisini and Riccardo Zoppoli. Neural approximations for infinite-horizon optimal control of nonlinear stochastic systems. *IEEE transactions on neural networks*, 9(6):1388–1408, 1998.

[152] Il Memming Park, Ábel Ságodi, and Piotr Aleksander Sokół. Persistent learning signals and working memory without continuous attractors. August 2023.

[153] Krzysztof Patan. Approximation of state-space trajectories by locally recurrent globally feed-forward neural networks. *Neural Networks*, 21(1):59–64, 2008.

[154] Marilia C Peixoto and Maurício M Peixoto. Structural stability in the plane with enlarged boundary conditions. *An. Acad. Brasil. Ci*, 31(2):135–160, 1959.

[155] Giordano Pola, Arjan J van der Schaft, and Maria Domenica Di Benedetto. Bisimulation theory for switching linear systems. In *2004 43rd IEEE Conference on Decision and Control (CDC)(IEEE Cat. No. 04CH37601)*, volume 2, pages 1406–1411. IEEE, 2004.

[156] Jordan B Pollack. The induction of dynamical recognizers. *Connectionist approaches to language learning*, pages 123–148, 1991.

[157] Demetri Psaltis, Athanasios Sideris, and Alan A Yamamura. A multilayered neural network controller. *IEEE control systems magazine*, 8(2):17–21, 1988.

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

[159] Prajit Ramachandran, Barret Zoph, and Quoc V Le. Searching for activation functions. *arXiv preprint arXiv:1710.05941*, 2017.

[160] PE Rapp. Why are so many biological systems periodic? *Progress in neurobiology*, 29(3): 261–273, 1987.

[161] Antônio H Ribeiro, Koen Tiels, Luis A Aguirre, and Thomas Schön. Beyond exploding and vanishing gradients: Analysing RNN training using attractors and smoothness. In *International conference on artificial intelligence and statistics*, pages 2370–2380. PMLR, 2020.

[162] Joel W Robbin. A structural stability theorem. *Annals of Mathematics*, 94(3):447–493, 1971.

[163] Clark R Robinson. Structural stability of C^1 flows. In *Dynamical Systems—Warwick 1974: Proceedings of a Symposium Held at the University of Warwick 1973/74*, pages 262–275. Springer, 1974.

[164] Domenec Ruiz-Balet and Enrique Zuazua. Neural ODE control for classification, approximation, and transport. *SIAM Review*, 65(3):735–773, 2023.

[165] Ábel Ságodi and Il Memming Park. Dynamical archetype analysis: Autonomous computation. July 2025. URL `http://arxiv.org/abs/2507.05505`.

[166] Ábel Ságodi, Guillermo Martín-Sánchez, Piotr Sokół, and Il Memming Park. Back to the continuous attractor. In *Advances in Neural Information Processing Systems (NeurIPS)*, December 2024. doi: 10.52202/079017-2136. URL `https://openreview.net/forum?id=fvG6ZHrH0B`.

[167] Ábel Ságodi, J. L. Pughe-Sanford, X. Ding, and D. B. Chklovskii. Data-enabled local reach-avoid control near saddle equilibria with a neuronal interpretation. In *65th IEEE Conference on Decision and Control (CDC)*, Honolulu, HI, USA, December 2026. To appear.

[168] Andrew R Sedler, Christopher Versteeg, and Chethan Pandarinath. Expressive architectures enhance interpretability of dynamics-based neural population models. *Neurons, behavior, data analysis, and theory*, 2023, 2023.

[169] Rodolphe Sepulchre. Fading memory [From the Editor]. *IEEE Control Systems Magazine*, 41(1):4–5, 2021.

[170] Shawn C. Shadden, Francois Lekien, and Jerrold E. Marsden. Definition and properties of Lagrangian coherent structures from finite-time Lyapunov exponents in two-dimensional aperiodic flows. *Physica D: Nonlinear Phenomena*, 212(3-4):271–304, 2005.

[171] Hava T Siegelmann and Eduardo D Sontag. On the computational power of neural nets. In *Proceedings of the fifth annual workshop on Computational learning theory*, pages 440–449, 1992.

[172] Hava T Siegelmann and Eduardo D Sontag. Analog computation via neural networks. *Theoretical Computer Science*, 131(2):331–360, 1994.

[173] Jiří Šíma and Pekka Orponen. Continuous-time symmetric Hopfield nets are computationally universal. *Neural Computation*, 15(3):693–733, 2003.

[174] Vincent Sitzmann, Julien Martel, Alexander Bergman, David Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. *Advances in neural information processing systems*, 33:7462–7473, 2020.

[175] Stephen Smale. Structurally stable systems are not dense. *American Journal of Mathematics*, 88(2):491–496, 1966.

[176] Stephen Smale. Differentiable dynamical systems. *Bulletin of the American mathematical Society*, 73(6):747–817, 1967.

[177] Chang hoon Song, Geonho Hwang, Jun ho Lee, and Myungjoo Kang. Minimal width for universal property of deep RNN. *Journal of Machine Learning Research*, 24(121):1–41, 2023.

[178] Eduardo D. Sontag. Polynomial response maps. In *Lecture Notes in Control and Information Sciences*, volume 13. Springer Verlag, Berlin, 1979.

[179] Eduardo D. Sontag. Realization theory of discrete-time nonlinear systems: Part I – The bounded case. *IEEE Transactions on Circuits and Systems*, 26(5):342–356, May 1979.

[180] Eduardo D Sontag. Neural nets as systems models and controllers. In *Proc. Seventh Yale Workshop on Adaptive and Learning Systems*, volume 73, 1992.

[181] Eduardo D Sontag. A learning result for continuous-time recurrent neural networks. *Systems & control letters*, 34(3):151–158, 1998.

[182] Alessandro Sperduti. On the computational power of recurrent neural networks for structures. *Neural Networks*, 10(3):395–400, 1997.

[183] David Sussillo and Omri Barak. Opening the black box: Low-dimensional dynamics in high-dimensional recurrent neural networks. *Neural Computation*, 25(3):626–649, 2013.

[184] David Sussillo, Rafal Jozefowicz, LF Abbott, and Chethan Pandarinath. LFADS-Latent Factor Analysis via Dynamical Systems. *arXiv preprint arXiv:1608.06315*, 2016.

[185] Héctor J Sussmann. Single-input observability of continuous-time systems. *Mathematical systems theory*, 12(1):371–393, 1978.

[186] Paulo Tabuada and Bahman Gharesifard. Universal approximation power of deep residual neural networks via nonlinear control theory. *arXiv preprint arXiv:2007.06007*, 2020.

[187] Paulo Tabuada and Bahman Gharesifard. Universal approximation power of deep residual neural networks through the lens of control. *IEEE Transactions on Automatic Control*, 68(5): 2715–2728, 2022.

[188] Shaohua Tan, Jianbin Hao, and Joos Vandewalle. Efficient identification of RBF neural net models for nonlinear discrete-time multivariable dynamical systems. *Neurocomputing*, 9(1): 11–26, 1995.

[189] Takeshi Teshima, Isao Ishikawa, Koichi Tojo, Kenta Oono, Masahiro Ikeda, and Masashi Sugiyama. Coupling-based invertible neural networks are universal diffeomorphism approximators. In *Advances in Neural Information Processing Systems*, 2020. 34th Conference on Neural Information Processing Systems (NeurIPS 2020).

[190] Takeshi Teshima, Koichi Tojo, Masahiro Ikeda, Isao Ishikawa, and Kenta Oono. Universal approximation property of Neural Ordinary Differential Equations, 2020. URL `https://arxiv.org/abs/2012.02414`.

[191] Stuart Townley, Achim Ilchmann, Martin G Weiß, Warren McClements, Antonio C Ruiz, David H Owens, and D Pratzel-Wolters. Existence and learning of oscillations in recurrent neural networks. *IEEE Transactions on Neural Networks*, 11(1):205–214, 2000.

[192] Lloyd N. Trefethen and Mark Embree. *Spectra and Pseudospectra: The Behavior of Nonnormal Matrices and Operators*. Princeton University Press, 2005.

[193] Leslie G Valiant. A theory of the learnable. *Communications of the ACM*, 27(11):1134–1142, 1984.

[194] B. van der Pol. On 'relaxation-oscillations'. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11):978–992, 1926. doi: 10.1080/14786442608564127.

[195] Arjan Van der Schaft. Bisimulation of dynamical systems. In Rajeev Alur and George J. Pappas, editors, *Hybrid Systems: Computation and Control*, volume 2293, pages 555–569. Springer, 2004.

[196] Ramon Van Handel. *Filtering, stability, and robustness*. PhD thesis, California Institute of Technology, 2007.

[197] Christopher Versteeg, Jonathan D McCart, Mitchell Ostrow, David M Zoltowski, Clayton B Washington, Laura Driscoll, Olivier Codol, Jonathan A Michaels, Scott W Linderman, David Sussillo, et al. Computation-through-Dynamics Benchmark: Simulated datasets and quality metrics for dynamical models of neural activity. *bioRxiv*, pages 2025–02, 2025.

[198] Vito Volterra. Sopra le funzioni che dipendono da altre funzioni. *Rendiconti della Reale Accademia dei Lincei*, 3:97–105, 1887.

[199] Saurabh Vyas, Matthew D Golub, David Sussillo, and Krishna V Shenoy. Computation through neural population dynamics. *Annual review of neuroscience*, 43:249–275, 2020.

[200] Shida Wang and Beichen Xue. State-space models with layer-wise nonlinearity are universal approximators with exponential decaying memory. *Advances in Neural Information Processing Systems*, 36, 2024.

[201] Kevin Warwick, George William Irwin, Kenneth J Hunt, et al. *Neural Networks for control and systems*, volume 46. Peter Peregrinus on behalf of the Institution of Electrical Engineers, 1992.

[202] Ezekiel Williams, Alexandre Payeur, Avery Hee-Woon Ryoo, Thomas Jiralerspong, Matthew G Perich, Luca Mazzucato, and Guillaume Lajoie. Expressivity of neural networks with random weights and learned biases. *arXiv preprint arXiv:2407.00957*, 2024.

[203] Matthew O Williams, Ioannis G Kevrekidis, and Clarence W Rowley. A data–driven approximation of the Koopman operator: Extending dynamic mode decomposition. *Journal of Nonlinear Science*, 25:1307–1346, 2015.

[204] Hiroki Yasumoto and Toshiyuki Tanaka. Universality of reservoir systems with recurrent neural networks. *Neural Networks*, 188:107413, 2025.

[205] Koichiro Yawata, Norihisa Namura, Yuzuru Kato, and Hiroya Nakao. Data-driven phase control for limit-cycle oscillators under partial observation. In *2025 IEEE 64th Conference on Decision and Control (CDC)*, pages 381–386. IEEE, 2025.

[206] Zhang Yi, Kok Kiong Tan, and Tong Heng Lee. Multistability analysis for recurrent neural networks with unsaturating piecewise linear transfer functions. *Neural Computation*, 15(3): 639–662, 2003.

[207] Zhen Yi. nmODE: Neural memory ordinary differential equation. *Artificial Intelligence Review*, 56:14403–14438, 2023. doi: 10.1007/s10462-023-10496-2.

[208] Chulhee Yun, Srinadh Bhojanapalli, Ankit Singh Rawat, Sashank J Reddi, and Sanjiv Kumar. Are transformers universal approximators of sequence-to-sequence functions? *arXiv preprint arXiv:1912.10077*, 2019.

[209] Muhammad Zakwan, Massimiliano d'Angelo, and Giancarlo Ferrari-Trecate. Universal approximation property of Hamiltonian deep neural networks. *IEEE Control Systems Letters*, 7:2689–2694, 2023.

[210] Majid Zamani, Peyman Mohajerin Esfahani, Rupak Majumdar, Alessandro Abate, and John Lygeros. Symbolic control of stochastic systems via approximately bisimilar finite abstractions. *IEEE Transactions on Automatic Control*, 59(12):3135–3150, 2014.

[211] Han Zhang, Xi Gao, Jacob Unterman, and Tom Arodz. Approximation capabilities of Neural ODEs and Invertible Residual Networks. In Hal Daumé III and Aarti Singh, editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 11086–11095. PMLR, 13–18 Jul 2020. URL `https://proceedings.mlr.press/v119/zhang20h.html`.

Classical results versus our contributions. We separate what we cite from what we prove. The classical results we use without reproving are Palis-Smale structural stability, the stable manifold theorem, Fenichel persistence of normally hyperbolic manifolds, Kupka-Smale density, Malkin's first variation of the period, Grönwall's inequality, and the C^1 universal approximation property. These supply the combinatorics of the invariant sets, but none of them controls the matched-time error $\sup_t \|\varphi - \hat{\varphi}\|$, because topological equivalence matches orbits only up to a time reparametrization; the gap is sharpest for a limit cycle, where a nearby orbit of period $\hat{T} \neq T$ lets $\sup_t \|\varphi - \hat{\varphi}\|$ grow to the cycle diameter. Our contributions are: (i) simultaneous multi-cycle period matching, via a period Jacobian that is strictly diagonally dominant and hence invertible (Theorem 4), which recovers the time compatibility structural stability discards; (ii) realizability of the localized corrections from the C^1 universal approximation property alone (Lemma 18), so Definition 1 is the only architectural assumption needed; (iii) a uniform transient time together with a Lyapunov-trapping construction that make an infinite-horizon Grönwall bound possible at all (Lemma 11); (iv) the explicit tiling of the structurally unstable continuous-attractor case (Theorem 5); (v) the isochrony obstruction and a phase-coherence bound across tiles; and (vi) the bridge from ε-δ closeness to training metrics (Theorem 6).

A Definitions and Background

A.1 Norms and Function Spaces

Definition 15 (C^1 norm)**.** For $f : \mathcal{X} \to \mathbb{R}^n$ continuously differentiable:

$$\|f\|_{C^1} \coloneqq \sup_{x \in \mathcal{X}} \|f(x)\| + \sup_{x \in \mathcal{X}} \|Df(x)\|_{\mathrm{op}}.$$

Definition 16 (Hausdorff distance)**.** For compact sets $A, B \subset \mathcal{X}$:

$$d_H(A, B) \coloneqq \max\left\{ \sup_{a \in A} \inf_{b \in B} \|a - b\|, \sup_{b \in B} \inf_{a \in A} \|a - b\| \right\}.$$

A.2 Dynamical Systems

Definition 17 (ω-limit set)**.** For a trajectory $\varphi(t, x_0)$, the **ω-limit set** is:

$$\omega(x_0) \coloneqq \bigcap_{T > 0} \overline{\{\varphi(t, x_0) : t \geq T\}}.$$

Equivalently, $y \in \omega(x_0)$ if and only if there exists a sequence $t_n \to \infty$ with $\varphi(t_n, x_0) \to y$.

Definition 18 (Attractor)**.** A compact invariant set $\mathcal{A} \subset \mathcal{X}$ is an **attractor** if it has an open neighborhood $U \supset \mathcal{A}$ such that $\varphi(t, U) \subset U$ for all $t > 0$ and $\bigcap_{t>0} \varphi(t, U) = \mathcal{A}$.

Definition 19 (Basin of attraction)**.** For attractor $\mathcal{A}$:

$$\mathrm{BoA}(\mathcal{A}) \coloneqq \{x_0 \in \mathcal{X} : \lim_{t \to \infty} \mathrm{dist}(\varphi(t, x_0), \mathcal{A}) = 0\}.$$

Definition 20 (Separatrix)**.** A **separatrix** is the boundary between two basins of attraction. For Morse-Smale systems, separatrices are (unions of) stable manifolds of saddle-type equilibria or periodic orbits.

Definition 21 (Non-wandering set)**.** A point $x \in \mathcal{X}$ is **non-wandering** if for every neighborhood U of x and every $T > 0$, there exists $t > T$ such that $\varphi(t, U) \cap U \neq \emptyset$. The **non-wandering set** $\Omega(f)$ is the set of all non-wandering points. Fixed points and periodic orbits are always non-wandering.

Definition 22 (Transversal intersection)**.** Two submanifolds $M, N \subset \mathcal{X}$ intersect **transversally** at $p \in M \cap N$ if $T_pM + T_pN = T_p\mathcal{X}$ (the tangent spaces span the ambient space). The intersection is transversal if this holds at every point of $M \cap N$.

Definition 23 (Tubular neighborhood)**.** For a submanifold $M \subset \mathcal{X}$, a **tubular neighborhood** of radius $r > 0$ is the set $N_r(M) = \{x \in \mathcal{X} : \mathrm{dist}(x, M) < r\}$.

Definition 24 (Poincaré return map)**.** Let γ be a periodic orbit and Σ a local cross-section (codimension-1 submanifold transverse to the flow) intersecting γ at a point p. The **Poincaré return map** $P : U \to \Sigma$ is defined on a neighborhood $U \subset \Sigma$ of p by $P(x) = \varphi(\tau(x), x)$, where $\tau(x) > 0$ is the first return time to Σ. The periodic orbit γ is hyperbolic if and only if $DP(p)$ has no eigenvalues of modulus 1.

Definition 25 (Topological equivalence)**.** Flows φ and ψ are **topologically equivalent** if there exists a homeomorphism $h : \mathcal{X} \to \mathcal{X}$ mapping orbits of φ to orbits of ψ preserving orientation (i.e., the direction of time along orbits).

Structural stability is a fundamental concept in the study of dynamical systems[81,88,132,149,154,162,163].

Definition 26 (Structural stability)**.** Let G be an open domain in $\mathbb{R}^n$ with compact closure and smooth $(n-1)$-dimensional boundary. Consider the space $X^1(G)$ consisting of restrictions to G of C^1 vector fields on $\mathbb{R}^n$ that are transversal to the boundary of G and are inward oriented. This space is endowed with the C^1 metric. A vector field $F \in X^1(G)$ is *weakly structurally stable* if for any sufficiently small perturbation F_1, the corresponding flows are *topologically equivalent* (Definition 25) on G: there exists a homeomorphism $h : G \to G$ which transforms the oriented trajectories of F into the oriented trajectories of F_1. If, moreover, for any $\varepsilon > 0$ the homeomorphism h may be chosen to be C^0 ε-close to the identity map when F_1 belongs to a suitable neighborhood of F depending on ε, then F is called (strongly) *structurally stable*.

Remark 13. Structural stability ensures that the qualitative features of the dynamics (number and type of fixed points, periodic orbits, connections between them) persist under small perturbations. This is essential for robust modeling, as approximation errors act as perturbations to the true system.

A.3 Universal Approximation

Theorem 8 (C^r UAP for vector fields[87])**.** *Let $\mathcal{X} \subset \mathbb{R}^n$ be compact. For any $f \in C^r(\mathcal{X}, \mathbb{R}^n)$ with $r \geq 0$ and any $\varepsilon > 0$, there exists a feedforward neural network $\hat{f}$ with smooth activation (e.g., sigmoid, tanh) such that $\|f - \hat{f}\|_{C^r} < \varepsilon$.*

Remark 14. The C^r approximation follows from the C^0 UAP combined with smoothness of neural network outputs: smooth activations produce smooth outputs, and derivatives can be approximated by approximating the function[87]. For our theorems, $r = 1$ suffices.

Lemma 9 (Levy-Desplanques)**.** *A square matrix $M \in \mathbb{R}^{k \times k}$ is **strictly diagonally dominant** if $|M_{ii}| > \sum_{j \neq i} |M_{ij}|$ for all i. Every strictly diagonally dominant matrix is invertible.*

Proof. Suppose $Mx = 0$ for some $x \neq 0$. Let $i = \arg\max_j |x_j|$, so $|x_i| > 0$. Row i gives $M_{ii}x_i = -\sum_{j \neq i} M_{ij}x_j$, hence $|M_{ii}||x_i| \leq \sum_{j \neq i} |M_{ij}||x_j| \leq |x_i| \sum_{j \neq i} |M_{ij}|$. This contradicts diagonal dominance. □

B Topological structure of the space of dynamical systems

In this section, we rigorously formalize the notion of approximation used in our main result. We demonstrate that our (ε, δ)-closeness condition (Def. 9) generates a valid topology on the space of dynamical systems, specifically the topology of *convergence in measure* regarding the supremum norm of trajectories.

Throughout this appendix, we work with three objects. The state space $\mathcal{X} \subset \mathbb{R}^n$ is the bounded open domain fixed in Section 2. The space of dynamical systems under consideration is denoted $\mathfrak{S} \subseteq \mathfrak{X}^1(\mathcal{X})$; concretely, $\mathfrak{S}$ may be taken as any of the target classes $\mathcal{F}_{\mathrm{FP}}, \mathcal{F}_{\mathrm{LC}}, \mathcal{F}_{\mathrm{CA}}$ of Section 4 together with the hypothesis class $\hat{\mathcal{F}}$, or any larger ambient family. The measure μ is the normalized Lebesgue measure on $\mathcal{X}$, i.e., $\mu(A) \coloneqq \mathrm{vol}(A)/\mathrm{vol}(\mathcal{X})$ for every Borel set $A \subseteq \mathcal{X}$, so that $\mu(\mathcal{X}) = 1$ and μ agrees with the integrand of Eq. (1). We define the trajectory difference between two systems $f, g \in \mathfrak{S}$ at an initial condition $x_0 \in \mathcal{X}$ as:

$$\Delta(f, g, x_0) \coloneqq \sup_{t \geq 0} \|\varphi_f(t, x_0) - \varphi_g(t, x_0)\|. \tag{2}$$

For any reference system $f \in \mathfrak{S}$ and constants $\varepsilon > 0, \delta > 0$, we define the neighborhood basis element $B(f, \varepsilon, \delta)$ as the set of systems that deviate by at most ε on all but a small volume of initial conditions:

$$B(f, \varepsilon, \delta) \coloneqq \{g \in \mathfrak{S} \mid \mu(\{x_0 \in \mathcal{X} : \Delta(f, g, x_0) \geq \varepsilon\}) < \delta\}. \tag{3}$$

Proposition 10. *The collection $\mathcal{B} = \{B(f, \varepsilon, \delta)\}$ satisfies the conditions to be a basis for a topology on $\mathfrak{S}$.*

Proof. We verify the two fundamental properties of a topological basis. First, we verify the covering property. For any $f \in \mathfrak{S}$, it holds that $f \in B(f, \varepsilon, \delta)$ for any $\varepsilon, \delta > 0$. Since $\Delta(f, f, x_0) = 0$ for all $x_0 \in \mathcal{X}$, the set where the error exceeds ε is empty, and $\mu(\emptyset) = 0 < \delta$. Thus, $\mathfrak{S} = \bigcup_{B \in \mathcal{B}} B$.

Next, we verify the intersection property. Let $h \in B(f, \varepsilon_1, \delta_1) \cap B(g, \varepsilon_2, \delta_2)$. We show there exists a neighborhood $B(h, \alpha, \beta)$ such that $B(h, \alpha, \beta) \subseteq B(f, \varepsilon_1, \delta_1)$. The inclusion for g follows by symmetry, and taking the intersection of the two resulting neighborhoods yields the required basis element.

Since $h \in B(f, \varepsilon_1, \delta_1)$, let $E_h = \{x_0 \in \mathcal{X} : \Delta(f, h, x_0) \geq \varepsilon_1\}$. By definition, $\mu(E_h) < \delta_1$. We utilize the continuity of measure to establish a safety margin. Consider the sets $A_n = \{x_0 \in \mathcal{X} : \Delta(f, h, x_0) \geq \varepsilon_1 - 1/n\}$. Note that $\bigcap_{n=1}^{\infty} A_n = E_h$. Since $\mu(E_h) < \delta_1$, there exists sufficiently large N such that $\mu(A_N) < \delta_1$. We define the slack parameters $\alpha \coloneqq 1/N$ and $\beta \coloneqq \delta_1 - \mu(A_N)$.

Consider any $k \in B(h, \alpha, \beta)$. Let $E_k = \{x_0 \in \mathcal{X} : \Delta(h, k, x_0) \geq \alpha\}$ be the exception set for k, noting that $\mu(E_k) < \beta$. We apply the triangle inequality: $\Delta(f, k, x_0) \leq \Delta(f, h, x_0) + \Delta(h, k, x_0)$. For any point $x_0 \notin (A_N \cup E_k)$, we have

$$\Delta(f, k, x_0) < \left(\varepsilon_1 - \frac{1}{N}\right) + \frac{1}{N} = \varepsilon_1.$$

Thus, the new exception set for k relative to f is contained in $A_N \cup E_k$. By the sub-additivity of the measure,

$$\mu(\{x_0 \in \mathcal{X} : \Delta(f, k, x_0) \geq \varepsilon_1\}) \leq \mu(A_N) + \mu(E_k) < \mu(A_N) + (\delta_1 - \mu(A_N)) = \delta_1.$$

Therefore, $k \in B(f, \varepsilon_1, \delta_1)$, which completes the proof. □

Remark 15 (Connection to the Ky Fan metric). Our definition of ε-δ closeness (Eq. 1) is topologically equivalent to the topology induced by the **Ky Fan metric**[48] on the space of trajectories. Specifically, the distance between two flows φ and $\hat{\varphi}$ can be defined as:

$$d_{KF}(\varphi, \hat{\varphi}) = \inf\left\{\varepsilon > 0 \;\middle|\; \mu(\{x \in \mathcal{X} : \sup_{t \geq 0} \|\varphi_t(x) - \hat{\varphi}_t(x)\| \geq \varepsilon\}) < \varepsilon\right\}.$$

Our guarantee essentially states that $d_{KF}(\varphi, \hat{\varphi})$ can be made arbitrarily small.

Remark 16 (Relation to standard error measures). ε-δ closeness takes the supremum of the trajectory error over time and controls it in measure over initial conditions. Relaxing either choice gives a standard measure, and both relaxations are strict. The uniform error $\sup_{x_0} \sup_{t \geq 0} \|\varphi_t(x_0) - \hat{\varphi}_t(x_0)\|$ is stronger and cannot be made small in a multistable system: a C^1-small perturbation shifts the separatrix, so a collar of initial conditions converges to the other attractor and has error of order D whatever the fit. ε-δ closeness concedes exactly this collar, as the set of measure below δ. The time-averaged L^p error of Theorem 6 is weaker: a transient excursion that recontracts has supremum error of order one while its time average tends to zero, so the average cannot constrain long-run behaviour. Topological equivalence is incomparable with all three, since it matches orbits only up to a time reparametrization. The damped Duffing fit of Section 6 shows the first gap: its uniform error is about 2.2, of the order of the distance between the two attractors, while $\delta(0.3) = 0.003$ and basin agreement is 99.96%.

C Proof of Theorem 3

Lemma 11 (Trajectory bounds inside basins)**.** *Let $\mathcal{A}$ be a hyperbolic attractor (stable fixed point) with spectral gap $\lambda > 0$ (slowest eigenvalue decay rate, i.e., $\lambda = \min_i |\mathrm{Re}(\mu_i)|$ for eigenvalues μ_i of $Df(\mathcal{A})$). Let $r > 0$ be a radius such that linearization is valid in $B_r(\mathcal{A})$. Let $\mathcal{V} \subset \mathrm{BoA}(\mathcal{A})$ with $\inf_{x \in \mathcal{V}} \mathrm{dist}(x, \partial\,\mathrm{BoA}(\mathcal{A})) \geq \eta > 0$.*

There exists $\nu_0 = \nu_0(\eta, \lambda, L, r) > 0$ such that for all $\nu < \nu_0$: if $\|f - \hat{f}\|_{C^1} < \nu$, then $\hat{f}$ has a hyperbolic attractor $\hat{\mathcal{A}}$ with $\mathcal{V} \subset \mathrm{BoA}(\hat{\mathcal{A}})$, and for all $x_0 \in \mathcal{V}$:

$$\sup_{t \geq 0} \|\varphi(t, x_0) - \hat{\varphi}(t, x_0)\| \leq \frac{K\nu}{\lambda},$$

where $K = K(\eta, f) = \lambda C K_1 + C$ with $K_1 = (e^{LT^*} - 1)/L$, $T^* = T^*(\eta)$ is the uniform transient time, $L = \mathrm{Lip}(f)$, and C bounds the flow Jacobian (fundamental matrix) in the linearization regime.

Proof. The proof proceeds in two stages.

Stage 1: Transient phase $[0, T^*]$**.** Since $\mathcal{V}$ is bounded away from the separatrix by $\eta > 0$, trajectories from $\mathcal{V}$ have uniform convergence time. To see this, note that $\bar{\mathcal{V}}$ (the closure of $\mathcal{V}$) is compact, and the convergence time function $\tau(x) = \inf\{t : \mathrm{dist}(\varphi(t, x), \mathcal{A}) < r/2\}$ is upper semicontinuous on the basin of attraction. Since $\bar{\mathcal{V}} \subset \mathrm{BoA}(\mathcal{A})$ is compact and bounded away from $\partial\,\mathrm{BoA}(\mathcal{A})$, the supremum $T^* = \sup_{x\in\bar{\mathcal{V}}} \tau(x)$ is finite[30] Thm. 2.1.

For $t \in [0, T^*]$, by Grönwall's inequality:

$$\|\varphi(t, x_0) - \hat{\varphi}(t, x_0)\| \le \nu \cdot \frac{e^{Lt} - 1}{L} \le K_1 \nu,$$

where $L = \mathrm{Lip}(f)$ and $K_1 = (e^{LT^*} - 1)/L$. Choosing $\nu_0 \le r/(2K_1)$ ensures the perturbed trajectory $\hat{\varphi}(T^*, x_0)$ is within distance r of $\hat{\mathcal{A}}$ (since $\|\mathcal{A} - \hat{\mathcal{A}}\| \le C_1\nu$ for some $C_1 > 0$ by structural stability).

Stage 2: Asymptotic phase $[T^*, \infty)$**.** Both trajectories start within the linearization neighborhood at time T^*. For $x \in B_r(\mathcal{A})$, the exponential contraction satisfies $\|D\varphi_{s,t}(x)\| \le Ce^{-\lambda(t-s)}$ for $t > s$.

Claim: Trajectories remain in the linearization neighborhood for all $t \ge T^*$. Because the Jacobian $A = Df(\mathcal{A})$ has eigenvalues with strictly negative real parts, there exists a symmetric positive definite matrix P solving the continuous Lyapunov equation $A^\top P + PA = -I$. We define the adapted Lyapunov function $V(x) = (x-\mathcal{A})^\top P(x-\mathcal{A})$, whose sublevel sets form invariant ellipsoids around $\mathcal{A}$. Along the unperturbed flow f, the derivative satisfies $\dot{V}_f(x) \le -c_1\|x - \mathcal{A}\|^2$ inside B_r for some constant $c_1 > 0$. For the approximated flow $\hat{f}$, the derivative is:

$$\dot{V}_{\hat{f}}(x) = \nabla V(x) \cdot \hat{f}(x) = \dot{V}_f(x) + \nabla V(x) \cdot (\hat{f}(x) - f(x)).$$

Since $\|\nabla V(x)\| \le 2\|P\|\|x - \mathcal{A}\|$ and the uniform perturbation is bounded by $\|f - \hat{f}\|_{C^0} < \nu$, we obtain:

$$\dot{V}_{\hat{f}}(x) \le -c_1\|x - \mathcal{A}\|^2 + 2\|P\|\|x - \mathcal{A}\|\nu.$$

This derivative is strictly negative for $\|x - \mathcal{A}\| > 2\|P\|\nu/c_1$. Thus, for ν sufficiently small, $\dot{V}_{\hat{f}} < 0$ on the boundary of an ellipsoidal sublevel set of V contained entirely within B_r, ensuring that any trajectory entering this set is permanently trapped.

By the variation of constants formula[196], applied with the linearized flow's fundamental matrix bounded in operator norm by C:

$$\|\varphi(t, x_0) - \hat{\varphi}(t, x_0)\| \le CK_1\nu \cdot e^{-\lambda(t-T^*)} + \int_{T^*}^{t} Ce^{-\lambda(t-s)}\nu\, ds \le CK_1\nu + \frac{C\nu}{\lambda}.$$

Setting $K = \lambda CK_1 + C$ gives the bound $K\nu/\lambda$.

Explicit ν_0*:* Set $\nu_0 = \min(r/(2K_1), \lambda r/(2C), \delta_{\mathrm{ss}})$ where δ_{ss} is the structural stability radius. □

Lemma 12 (Basin error volume)**.** *For any* $\delta > 0$*, there exists* $\eta_{\mathrm{vol}} > 0$ *such that* $\|f - \hat{f}\|_{C^1} < \eta_{\mathrm{vol}}$ *implies:* $\mu(\mathcal{E}_{\mathrm{basin}}) < \delta$*, where* $\mathcal{E}_{\mathrm{basin}} = \bigcup_i (\mathrm{BoA}(\mathcal{A}_i^f) \setminus \mathrm{BoA}(\mathcal{A}_i^{\hat{f}}))$.

Proof. By structural stability, $\|f - \hat{f}\|_{C^1} < \eta$ implies topological equivalence via homeomorphism h with $\|h - \mathrm{id}\|_{C^0} \to 0$ as $\eta \to 0$.

The basin error set is contained in the $\|h - \mathrm{id}\|_{C^0}$-neighborhood of the separatrices $S = \bigcup_i \partial\,\mathrm{BoA}(\mathcal{A}_i)$. Since separatrices are stable manifolds of saddles with codimension ≥ 1, they have Lebesgue measure zero. By continuity of measure, $\mu(\mathcal{E}_{\mathrm{basin}}) \to 0$ as $\eta \to 0$. □

Lemma 13 (Forward invariance under approximation)**.** *Let* $f \in \mathfrak{X}^1(\mathcal{X})$ *be strictly inward-pointing at* $\partial\mathcal{X}$ *with constant* $\gamma > 0$ *(Definition 7). If* $\hat{f} : \mathcal{X} \to \mathbb{R}^n$ *satisfies* $\|f - \hat{f}\|_{C^0} < \gamma$*, then* $\hat{f} \in \mathfrak{X}^1(\mathcal{X})$*, i.e., trajectories of* $\hat{f}$ *remain in* $\mathcal{X}$ *for all* $t \ge 0$.

Proof. Let $\nu(x)$ denote the outward unit normal at $x \in \partial\mathcal{X}$. By the strictly inward-pointing condition, $f(x) \cdot \nu(x) < -\gamma$ for all $x \in \partial\mathcal{X}$.

For $\hat{f}$ with $\|f - \hat{f}\|_{C^0} < \gamma$:

$$\hat{f}(x) \cdot \nu(x) = f(x) \cdot \nu(x) + (\hat{f}(x) - f(x)) \cdot \nu(x) < -\gamma + \|\hat{f} - f\|_{C^0} < 0.$$

Thus $\hat{f}$ also points strictly inward at $\partial\mathcal{X}$. By the flow box theorem, trajectories starting in $\mathcal{X}$ cannot exit through $\partial\mathcal{X}$, so $\hat{\varphi}(t, x_0) \in \mathcal{X}$ for all $t \geq 0$ and $x_0 \in \mathcal{X}$. □

Full proof of Theorem 3. Given $\varepsilon, \delta > 0$:

Step 1 (spatial radius): Choose a spatial radius $\eta > 0$ small enough that $\mu(N_\eta(S)) < \delta/2$. This is possible because S has measure zero (Theorem 2) and μ is continuous from above.

Step 2 (perturbation budget for basin error): By Lemma 12, there exists a perturbation budget $\eta_{\text{vol}} = \eta_{\text{vol}}(\eta) > 0$ such that $\|f - \hat{f}\|_{C^1} < \eta_{\text{vol}}$ implies $\mathcal{E}_{\text{basin}} \subset N_\eta(S)$.

Step 3 (perturbation budget for trajectory error): Define $\mathcal{V} \coloneqq \mathcal{X} \setminus N_\eta(S)$, depending on f, η alone. By Lemma 11, the constants $K_1 = (e^{LT^*} - 1)/L$ and $K = K(\eta, f)$ are determined a priori; choose $\eta_{\text{traj}} = \lambda\varepsilon/K$ so the trajectory error is $< \varepsilon$ on $\mathcal{V}$.

Step 4 (combined budget): Let $\gamma > 0$ be the strict inward-pointing constant. Set

$$\eta_0 \coloneqq \min(\eta_{\text{vol}}, \eta_{\text{traj}}, \gamma).$$

All three quantities are now perturbation budgets in the C^1 norm.

Step 5 (apply UAP): By Theorem 8, choose $\hat{f} \in \hat{\mathcal{F}}$ with $\|f - \hat{f}\|_{C^1} < \eta_0$.

Since $\|f - \hat{f}\|_{C^0} < \gamma$, forward invariance holds by Lemma 13. By Step 2, $\mathcal{E}_{\text{basin}} \subset N_\eta(S)$, so for the error set $\mu\big(\{x_0 : \sup_t \|\varphi - \hat{\varphi}\| > \varepsilon\}\big) \leq \mu(N_\eta(S)) < \delta/2 < \delta$. □

D Proof of Theorem 4

Lemma 14 (Structural stability of hyperbolic limit cycles)**.** *Let γ be a hyperbolic periodic orbit with period T. There exist $L_H, L_T, \delta_0 > 0$ such that $\|f - \hat{f}\|_{C^1} < \delta_0$ implies:*

(i) $\hat{f}$ has a unique hyperbolic periodic orbit $\hat{\gamma}$ near γ

(ii) $d_H(\gamma, \hat{\gamma}) \leq L_H \|f - \hat{f}\|_{C^1}$

(iii) $|\hat{T} - T| \leq L_T \|f - \hat{f}\|_{C^1}$

Proof. By Fenichel's persistence theorem for normally hyperbolic invariant manifolds[55]. The period functional is Fréchet differentiable for hyperbolic cycles[30], hence Lipschitz continuous. □

D.1 With multiplicative correction

Lemma 15 (Period correction)**.** *Let $\hat{f}$ have hyperbolic limit cycle $\hat{\gamma}$ with period $\hat{T}$. For target period T with $|\hat{T} - T|$ small, there exists $\alpha^* = (\hat{T} - T)/T$ such that:*

$$\hat{f}_{\alpha^*} \coloneqq (1 + \alpha^*\psi)\hat{f}$$

has periodic orbit $\hat{\gamma}$ with period exactly T, where ψ is a bump function with $\psi \equiv 1$ on $\hat{\gamma}$.

Proof. By time reparametrization, $(1 + \alpha\psi)\hat{f}$ preserves $\hat{\gamma}$ as a periodic orbit with period $P(\alpha) = \hat{T}/(1 + \alpha)$ (since $\psi = 1$ on $\hat{\gamma}$). Setting $P(\alpha^*) = T$ yields $\alpha^* = (\hat{T} - T)/T$. □

Lemma 16 (Simultaneous period correction)**.** *Let the vector field $\hat{f}$ possess N disjoint hyperbolic limit cycles $\{\hat{\gamma}_1, \ldots, \hat{\gamma}_N\}$ with respective periods $\hat{\mathbf{T}} = (\hat{T}_1, \ldots, \hat{T}_N)$. Let $\{\psi_1, \ldots, \psi_N\}$ be a set of*

smooth bump functions where each ψ_i is equal to 1 on $\hat{\gamma}_i$ and decays such that $\sup_{x\in\hat{\gamma}_k}|\psi_i(x)|<\xi$ for $i\neq k$. Consider the parameterized family of vector fields:

$$\hat{f}_{\boldsymbol{\alpha}}(x) \coloneqq \left(1+\sum_{i=1}^{N}\alpha_i\psi_i(x)\right)\hat{f}(x), \quad \boldsymbol{\alpha}\in\mathbb{R}^N.$$

For any target period vector $\mathbf{T}=(T_1,\dots,T_N)$ sufficiently close to $\hat{\mathbf{T}}$, provided the leakage ξ is sufficiently small, there exists a unique parameter vector $\boldsymbol{\alpha}^*$ such that the limit cycles of $\hat{f}_{\boldsymbol{\alpha}^*}$ have periods exactly equal to $\mathbf{T}$.

Proof. Since the limit cycles are hyperbolic, structural stability implies they persist for small $\boldsymbol{\alpha}$. Let $\mathbf{T}(\boldsymbol{\alpha}):\mathbb{R}^N\to\mathbb{R}^N$ be the map taking the parameters to the periods of the resulting cycles. We aim to solve $\mathbf{T}(\boldsymbol{\alpha})=\mathbf{T}_{target}$.We analyze the Jacobian matrix of this map, $\mathbf{J}\in\mathbb{R}^{N\times N}$, evaluated at $\boldsymbol{\alpha}=\mathbf{0}$. The entry $J_{ki}=\frac{\partial T^k}{\partial\alpha_i}$ represents the sensitivity of the k-th cycle's period to the i-th bump function.

The period of a cycle γ subject to a velocity scaling $v(x)\to(1+\eta(x))v(x)$ is given by $T=\oint_\gamma\frac{ds}{\|(1+\eta)v\|}\approx\oint\frac{ds}{\|v\|}-\oint\frac{\eta}{\|v\|}ds$. Thus, the partial derivatives are:

$$\left.\frac{\partial T^k}{\partial\alpha_i}\right|_{\boldsymbol{\alpha}=0} = -\int_0^{\hat{T}_k}\psi_i(\hat{\gamma}_k(t))\,dt.$$

1. Diagonal terms ($i=k$): Since $\psi_k\equiv 1$ on $\hat{\gamma}_k$, the integral is exact:

$$J_{kk}=-\int_0^{\hat{T}_k}1\,dt=-\hat{T}_k.$$

2. Off-diagonal terms ($i\neq k$): Since the support of ψ_i is concentrated on cycle i, its value on cycle k is bounded by the leakage parameter ξ:

$$|J_{ki}|=\left|-\int_0^{\hat{T}_k}\psi_i(\hat{\gamma}_k(t))\,dt\right|\leq\int_0^{\hat{T}_k}\xi\,dt=\xi\hat{T}_k.$$

3. Invertibility via Diagonal Dominance: We construct the Jacobian to be strictly diagonally dominant. The condition $|J_{kk}|>\sum_{i\neq k}|J_{ki}|$ becomes:

$$\hat{T}_k>\sum_{i\neq k}\xi\hat{T}_k=(N-1)\xi\hat{T}_k \implies 1>(N-1)\xi.$$

By choosing the bump functions such that the leakage $\xi<\frac{1}{N-1}$, the matrix $\mathbf{J}$ is strictly diagonally dominant. By the Lévy-Desplanques Theorem (Theorem 9), $\mathbf{J}$ is non-singular (invertible).

Since the Jacobian is invertible at $\mathbf{0}$, the Inverse Function Theorem guarantees that $\mathbf{T}(\boldsymbol{\alpha})$ is a local diffeomorphism mapping a neighborhood of $\mathbf{0}$ to a neighborhood of $\hat{\mathbf{T}}$. Thus, for any target periods $\mathbf{T}$ sufficiently close to $\hat{\mathbf{T}}$, there exists a unique $\boldsymbol{\alpha}^*$ such that $\mathbf{T}(\boldsymbol{\alpha}^*)=\mathbf{T}$. □

Remark 17 (Hyperbolicity preservation). The corrected vector field $\hat{f}_{\alpha^*}$ preserves the hyperbolicity of $\hat{\gamma}$. Since ψ is C^∞ with support in a tubular neighborhood of $\hat{\gamma}$ and $|\alpha^*|\leq L_T\eta/T$ is small, the correction is a small C^1 perturbation. By structural stability of hyperbolic periodic orbits, $\hat{\gamma}$ remains hyperbolic for the corrected system. Moreover, $\|\hat{f}-\hat{f}_{\alpha^*}\|_{C^1}=O(|\alpha^*|\cdot\|\psi\|_{C^1}\cdot\|\hat{f}\|_{C^1})$, which can be made arbitrarily small by choosing η small.

Lemma 17 (Trajectory bounds near limit cycles)**.** *Let $\mathcal{V}\subset\mathrm{BoA}(\gamma)$ with*

$$\inf_{x\in\mathcal{V}}\operatorname{dist}(x,\partial\,\mathrm{BoA}(\gamma))\geq\eta>0.$$

If periods match exactly ($T=\hat{T}$) and $d_H(\gamma,\hat{\gamma})<\varepsilon_{\mathrm{geom}}$, then:

$$\limsup_{t\to\infty}\|\varphi(t,x_0)-\hat{\varphi}(t,x_0)\|\leq\varepsilon_{\mathrm{geom}}+O(\|f-\hat{f}\|_{C^1}).$$

Proof. The argument mirrors the two-stage analysis of Lemma 11, with the cycle's asymptotic phase function (Definition 28) replacing the role of the fixed point.

Transient phase $[0, T^*]$. Since $\mathcal{V}$ is bounded away from the basin boundary, there is a uniform transient time $T^* < \infty$ before trajectories from $\mathcal{V}$ enter a tubular neighborhood of γ and $\hat{\gamma}$. On this interval, Grönwall's inequality applied to the difference $e(t) = \varphi(t, x_0) - \hat{\varphi}(t, x_0)$ yields

$$\|e(t)\| \;\le\; \|f - \hat{f}\|_{C^0} \cdot \frac{e^{Lt} - 1}{L}, \qquad t \in [0, T^*],$$

where $L = \mathrm{Lip}(f)$.

Asymptotic phase $[T^*, \infty)$. After T^*, $\varphi(t, x_0)$ converges exponentially to γ at asymptotic phase $\Psi^f(x_0)$, and $\hat{\varphi}(t, x_0)$ converges exponentially to $\hat{\gamma}$ at $\Psi^{\hat{f}}(x_0)$. With exact period matching ($T = \hat{T}$), the phase difference $\Delta\Psi = \Psi^f(x_0) - \Psi^{\hat{f}}(x_0)$ is constant in t. The asymptotic trajectory error decomposes as

$$\limsup_{t\to\infty} \|e(t)\| \;\le\; d_H(\gamma, \hat{\gamma}) + L_\gamma |\Delta\Psi|,$$

where $L_\gamma = \sup \|\dot{\gamma}\|$. By Floquet smoothness of the phase function on C^1-close vector fields (the same bound used in the proof of Lemma 26), $|\Delta\Psi| \le C_\Psi \|f - \hat{f}\|_{C^1}$, giving the claim. □

Proof of Theorem 4 (multiplicative correction). **Step 1 (Base approximation):** By Theorem 8, choose $\tilde{f} \in \hat{\mathcal{F}}$ with $\|f - \tilde{f}\|_{C^1} < \eta$. By Lemma 14, $\tilde{f}$ has limit cycles $\tilde{\gamma}_i$ with periods $\tilde{T}_i$.

Step 2 (Period correction): For single cycle: $\hat{f} = c^* \tilde{f}$ with $c^* = \tilde{T}/T$ (Lemma 15). For multiple cycles with disjoint tubular neighborhoods: $\hat{f} = (1 + \sum_i \alpha_i^* \psi_i)\tilde{f}$ (Lemma 16).

Step 3 (Error control): By Lemma 12 (adapted for limit cycles), choose η so $\mu(\mathcal{E}_{\mathrm{basin}}) < \delta/2$. By Lemma 17, exact period matching bounds asymptotic error. Choose η so geometric and transient errors sum to $< \varepsilon$.

The error set has measure $< \delta$, completing the proof. □

D.2 With additive correction

Lemma 18 (Bump Function Realizability)**.** *Let $\tilde{f}$ be an approximation of f with limit cycles $\tilde{\gamma}_i \subset N_i$. Let $\nu, \zeta > 0$ be tolerance parameters. There exist functions $\{\hat{\mathbf{\Phi}}_i\}_{i=1}^N \subset \hat{\mathcal{F}}$ such that for each i:*

1. *Alignment: $\|\hat{\Phi}_i(x) - \tilde{f}(x)\| < \nu$ for all $x \in \tilde{\gamma}_i$.*
2. *Support Decay (C^1 Leakage): $\|\hat{\mathbf{\Phi}}_i(x)\|_{C^1} < \zeta$ for all $x \notin M_i$.*
3. *Smoothness: $\|\hat{\mathbf{\Phi}}_i\|_{C^1}$ is bounded.*

Proof. Let $\mathbf{\Phi}_i^*$ be a smooth "ideal" bump function that equals $\tilde{f}$ on N_i and vanishes outside M_i. Since $\hat{\mathcal{F}}$ possesses the C^1 UAP, there exists $\hat{\mathbf{\Phi}}_i \in \hat{\mathcal{F}}$ such that $\|\hat{\mathbf{\Phi}}_i - \mathbf{\Phi}_i^*\|_{C^1} < \min(\nu, \zeta)$. The condition on $x \in \tilde{\gamma}_i$ is satisfied by ν-closeness. The condition on $x \notin M_i$ (where $\mathbf{\Phi}_i^* = 0$) is satisfied because the UAP yields ζ-closeness in the C^1 norm. □

Definition 27 (Adjoint Solution)**.** Let $\tilde{\varphi}_t$ denote the flow generated by the vector field $\tilde{f}$, and let $\tilde{\gamma}(t)$ be a hyperbolic periodic orbit with period $\tilde{T}$. The **linear variational equation** along the orbit is given by $\dot{u} = D\tilde{f}(\tilde{\gamma}(t))u$. The corresponding **adjoint equation** is defined as:

$$\dot{z} = -[D\tilde{f}(\tilde{\gamma}(t))]^\top z.$$

There exists a unique $\tilde{T}$-periodic solution $Z(t)$ to this adjoint equation, normalized such that:

$$\langle Z(t), \tilde{f}(\tilde{\gamma}(t))\rangle \equiv 1 \quad \text{for all } t \in [0, \tilde{T}].$$

The vector $Z(t)$ represents the sensitivity of the period to instantaneous perturbations of the vector field.

Lemma 19 (First Variation of the Period[131])**.** *Consider the perturbed vector field $\hat{f}_\alpha = \tilde{f} + \alpha g$. The derivative of the period $T(\alpha)$ with respect to the parameter α at $\alpha = 0$ is given exactly by:*

$$\left.\frac{dT}{d\alpha}\right|_{\alpha=0} = -\int_0^{\tilde{T}} \langle Z(t), g(\tilde{\gamma}(t))\rangle \, dt.$$

See also Guckenheimer and Holmes[72] and Ermentrout and Terman[54].

Lemma 20 (Exact Jacobian Form)**.** *Consider the multi-cycle approximation $\hat{f}_{\boldsymbol{\alpha}} = \tilde{f} + \sum_{j=1}^{N} \alpha_j \hat{\boldsymbol{\Phi}}_j$. Let $Z_i(t)$ be the normalized adjoint solution associated with the i-th limit cycle $\tilde{\gamma}_i$. Applying Lemma 19 with perturbation $g = \hat{\boldsymbol{\Phi}}_j$, the entries of the period Jacobian matrix $J = D_{\boldsymbol{\alpha}}\mathbf{T}(\mathbf{0})$ are:*

$$J_{ij} = \left.\frac{\partial T_i}{\partial \alpha_j}\right|_{\boldsymbol{\alpha}=\mathbf{0}} = -\int_0^{\tilde{T}_i} \langle Z_i(t), \hat{\boldsymbol{\Phi}}_j(\tilde{\gamma}_i(t))\rangle \, dt.$$

Lemma 21 (Robust Invertibility of Period Jacobian)**.** *Let $C_Z = \max_i \sup_{t\in[0,\tilde{T}_i]} \|Z_i(t)\|$ be the maximum norm of the adjoint solutions. Assume the approximator bump functions $\hat{\boldsymbol{\Phi}}_j$ satisfy the following proximity conditions:*

- ***Alignment (on cycle):*** *$\|\hat{\boldsymbol{\Phi}}_i(x) - \tilde{f}(x)\| < \nu$ for all $x \in \tilde{\gamma}_i$.*
- ***Leakage (off cycle):*** *$\|\hat{\boldsymbol{\Phi}}_j(x)\| < \zeta$ for all $x \in \tilde{\gamma}_i$ where $i \neq j$.*

If the approximation parameters satisfy the condition:

$$\nu + (N-1)\zeta < \frac{1}{C_Z},$$

then the Jacobian matrix $J = D_{\boldsymbol{\alpha}}\mathbf{T}(\mathbf{0})$ is strictly diagonally dominant and therefore non-singular.

Proof. We analyze the entries of the Jacobian J_{ij} defined in Lemma 20.

1. Diagonal Terms ($i = j$): On the cycle $\tilde{\gamma}_i$, we write $\hat{\boldsymbol{\Phi}}_i = \tilde{f} + \boldsymbol{\Delta}_i$, where $\|\boldsymbol{\Delta}_i\| < \nu$. Substituting this into the integral:

$$J_{ii} = -\int_0^{\tilde{T}_i} \langle Z_i(t), \tilde{f}(\tilde{\gamma}_i(t)) + \boldsymbol{\Delta}_i(t)\rangle \, dt.$$

Using linearity and the normalization property $\langle Z_i(t), \tilde{f}(\tilde{\gamma}_i(t))\rangle \equiv 1$:

$$J_{ii} = \underbrace{-\int_0^{\tilde{T}_i} 1 \, dt}_{=\tilde{T}_i} \underbrace{-\int_0^{\tilde{T}_i} \langle Z_i(t), \boldsymbol{\Delta}_i(t)\rangle \, dt}_{E_{ii}}.$$

We bound the error term E_{ii}:

$$|E_{ii}| \leq \int_0^{\tilde{T}_i} \|Z_i(t)\|\|\boldsymbol{\Delta}_i(t)\| \, dt \leq \tilde{T}_i C_Z \nu.$$

Thus, the magnitude of the diagonal entry is bounded from below:

$$|J_{ii}| \geq \tilde{T}_i(1 - C_Z\nu).$$

2. Off-Diagonal Terms ($i \neq j$): For $i \neq j$, the function $\hat{\boldsymbol{\Phi}}_j$ acts on the cycle $\tilde{\gamma}_i$. By the leakage assumption, $\|\hat{\boldsymbol{\Phi}}_j\| < \zeta$ on $\tilde{\gamma}_i$.

$$|J_{ij}| = \left| -\int_0^{\tilde{T}_i} \langle Z_i(t), \hat{\boldsymbol{\Phi}}_j(\tilde{\gamma}_i(t))\rangle \, dt \right| \leq \int_0^{\tilde{T}_i} \|Z_i(t)\|\zeta \, dt \leq \tilde{T}_i C_Z \zeta.$$

3. Diagonal Dominance: For strict diagonal dominance, we require $|J_{ii}| > \sum_{j\neq i} |J_{ij}|$. Substituting our bounds:

$$\tilde{T}_i(1 - C_Z\nu) > (N-1)\tilde{T}_i C_Z \zeta.$$

Dividing by $\tilde{T}_i$ and rearranging:

$$1 > C_Z\nu + C_Z(N-1)\zeta \implies \nu + (N-1)\zeta < \frac{1}{C_Z}.$$

This condition holds by hypothesis, ensuring J is invertible. □

Proposition 22 (Correction Existence)**.** *Let $\Delta T = \mathbf{T}_{target} - \tilde{\mathbf{T}}$. Since $J(\mathbf{0})$ is non-singular (Lemma 21), by the Inverse Function Theorem, there exists a radius $R_{sol} > 0$ and a constant $\kappa > 0$ such that if $\|\Delta T\| < R_{sol}$, there exists a unique parameter vector $\boldsymbol{\alpha}^*$ satisfying $\mathbf{T}(\boldsymbol{\alpha}^*) = \mathbf{T}_{target}$ with the bound:*

$$\|\boldsymbol{\alpha}^*\| \leq \kappa \|\Delta T\|.$$

Lemma 23 (Geometric Stability)**.** *Let $\hat{f} = \tilde{f} + \sum \alpha_i^* \hat{\mathbf{\Phi}}_i$. Then:*

1. ***Cycle Shift:*** $d_H(\tilde{\gamma}, \hat{\gamma}) \leq L_H \, NB_\Phi \, \|\boldsymbol{\alpha}^*\|$.
2. ***Basin Error:*** $\mu(\mathcal{B}(\tilde{f}) \triangle \mathcal{B}(\hat{f})) \leq K_{sep} \, NB_\Phi \, \|\boldsymbol{\alpha}^*\|$.

where L_H is the Lipschitz constant from Lemma 14(ii) and K_{sep} is the basin-boundary Lipschitz constant.

Proof. (i) The cycle shift follows from Lemma 14(ii) applied to the perturbation $\hat{f} - \tilde{f} = \sum_i \alpha_i^* \hat{\mathbf{\Phi}}_i$, which has C^1 norm $\leq NB_\Phi \|\boldsymbol{\alpha}^*\|$ by Lemma 18. (ii) Basin boundaries between adjacent attractors are stable manifolds of intermediate saddles. Their position depends C^1-Lipschitz on the vector field[30] Theorem 5.7.5, so the symmetric-difference measure is bounded by K_{sep} times the C^1 perturbation magnitude $NB_\Phi \|\boldsymbol{\alpha}^*\|$. □

Proof of Theorem 4 (additive correction). **Step 1: Configuration (Fixing Constants).** Set bump parameters $\nu = \frac{1}{3C_Z}$ and $\zeta = \frac{1}{3NC_Z}$. By Lemma 21, this ensures the period Jacobian is invertible. By Lemma 18, such bumps exist in $\hat{\mathcal{F}}$ with bound B_Φ.

Step 2: Base Approximation. We choose a base approximation $\tilde{f} \in \hat{\mathcal{F}}$ such that $\|f - \tilde{f}\|_{C^1} < \eta_{\text{base}}$. By Lemma 14(ii), the initial period error is bounded by $\|\Delta T\| \leq L_T \eta_{\text{base}}$. From Proposition 22, if a solution exists, the correction magnitude is bounded by $\|\boldsymbol{\alpha}^*\| \leq \kappa L_T \eta_{\text{base}}$. We now determine the required η_{base} by enforcing four simultaneous constraints:

Constraint A: Solvability. To invoke Proposition 22, we require $\|\Delta T\| < R_{sol}$. Using the structural stability bound:

$$\eta_{\text{base}} < \frac{R_{sol}}{L_T}.$$

Constraint B: Basin Stability (δ). The total basin error comes from the base approximation (structural stability) plus the separatrix shift due to the bump-function correction. Since separatrix shifts scale with the C^1 norm of the perturbation, the relevant bound on the corrective term is $\|\sum_j \alpha_j^* \hat{\mathbf{\Phi}}_j\|_{C^1} \leq NB_\Phi \|\boldsymbol{\alpha}^*\|$ (Lemma 18, in conjunction with Lemma 23(ii)). Combining,

$$K_{\text{sep}}\eta_{\text{base}} + K_{\text{sep}} NB_\Phi(\kappa L_T \eta_{\text{base}}) < \delta \implies \eta_{\text{base}} < \frac{\delta}{K_{\text{sep}}(1 + NB_\Phi \kappa L_T)}.$$

Constraint C: Trajectory Accuracy (ε). We require the asymptotic trajectory error to satisfy $\limsup_{t\to\infty} \|\varphi(t,x) - \hat{\varphi}(t,x)\| < \varepsilon$. By the triangle inequality, this error is bounded by the sum of three contributions: the geometric distance between the limit cycles, the residual phase offset that gets frozen once periods match, and the asymptotic phase mismatch coming from non-equal isochron foliations.

1. **Geometric Error:** The Hausdorff distance between the true cycle γ and the corrected cycle $\hat{\gamma}$ satisfies:

$$d_H(\gamma, \hat{\gamma}) \leq \underbrace{d_H(\gamma, \tilde{\gamma})}_{\text{Base Error}} + \underbrace{d_H(\tilde{\gamma}, \hat{\gamma})}_{\text{Correction Shift}} \leq L_H \eta_{\text{base}} + L_H NB_\Phi \|\boldsymbol{\alpha}^*\|.$$

2. **Transient trajectory divergence (Grönwall).** On the transient interval $[0, T_{\text{trans}}]$ before trajectories lock onto their respective cycles, the difference $e(t) := \varphi(t, x_0) - \hat{\varphi}(t, x_0)$ satisfies $\dot{e} = f(\varphi) - \hat{f}(\hat{\varphi})$, which decomposes as

$$\dot{e} = \big(f(\varphi) - f(\hat{\varphi})\big) + \big(f(\hat{\varphi}) - \hat{f}(\hat{\varphi})\big).$$

The first term is bounded in norm by $L\|e\|$ where $L = \text{Lip}(f)$; the second by $\|f - \hat{f}\|_{C^0}$. Grönwall's inequality[30] then yields

$$\|\varphi(t, x_0) - \hat{\varphi}(t, x_0)\| \;\le\; \|f - \hat{f}\|_{C^0} \cdot \frac{e^{Lt} - 1}{L}, \qquad t \in [0, T_{\text{trans}}], \tag{4}$$

where $\|f - \hat{f}\|_{C^0} \le \eta_{\text{base}} + NB_\Phi\|\boldsymbol{\alpha}^*\|$. Once exact period matching takes effect (Lemma 16 or Proposition 22), this transient offset is frozen (it neither amplifies nor decays), and it bounds the asymptotic phase contribution. The Grönwall amplification $G_T := (e^{LT_{\text{trans}}} - 1)/L$ is finite for each fixed transient horizon, so this term is $O(\eta_{\text{base}} + \|\boldsymbol{\alpha}^*\|)$ but with a Lipschitz-controlled multiplicative constant.

3. **Asymptotic phase mismatch (isochron Floquet bound).** After the transient, trajectories track their respective cycles at asymptotic phases $\Psi^f(x_0)$ and $\Psi^{\hat{f}}(x_0)$ (Definition 28). With exact period matching, the phase difference $\Delta\Psi := \Psi^f(x_0) - \Psi^{\hat{f}}(x_0)$ is constant in t and contributes to the trajectory error by $L_\gamma|\Delta\Psi|$, where $L_\gamma = \sup\|\dot{\gamma}\|$. By the same Floquet-smoothness argument used in Lemma 26 (Stage 2, applied directly to the isolated cycle pair $\gamma, \hat{\gamma}$),

$$|\Delta\Psi| \le C_\Psi\|f - \hat{f}\|_{C^1} \le C_\Psi(\eta_{\text{base}} + NB_\Phi\|\boldsymbol{\alpha}^*\|).$$

Substituting the correction bound $\|\boldsymbol{\alpha}^*\| \le \kappa L_T \eta_{\text{base}}$ and summing:

$$\text{Total Error} \;\le\; \eta_{\text{base}} \cdot (1 + NB_\Phi\kappa L_T) \cdot \big(L_H + G_T + L_\gamma C_\Psi\big).$$

To ensure the total error is strictly less than ε, we impose:

$$\eta_{\text{base}} \;<\; \frac{\varepsilon}{(1 + NB_\Phi\kappa L_T)\big(L_H + G_T + L_\gamma C_\Psi\big)}. \tag{5}$$

The Lipschitz amplification $G_T = (e^{LT_{\text{trans}}} - 1)/L$ enters multiplicatively; for systems with large L or long transients, this constant can dominate the budget, but it is finite and yields a positive lower bound on the achievable η_{base}.

Constraint D: Forward Invariance. To satisfy Lemma 13, the total error must be less than γ:

$$\|f - \hat{f}\| \le \|f - \tilde{f}\| + \|\hat{f} - \tilde{f}\| \le \eta_{\text{base}}(1 + NB_\Phi\kappa L_T) < \gamma.$$

Conclusion. Let η^* be the minimum of the bounds derived in Constraints A–D. By the C^1 UAP of $\hat{\mathcal{F}}$, there exists $\tilde{f}$ satisfying $\|f - \tilde{f}\|_{C^1} < \eta^*$. This $\tilde{f}$ is sufficiently accurate that a period correction $\boldsymbol{\alpha}^*$ exists (A), maintains basin structure (B), ensures trajectory tracking (C), and preserves the domain (D). □

E Proof of Theorem 5

Theorem 24 (Fenichel persistence[55])**.** *Let $\mathcal{M}$ be a compact normally hyperbolic invariant manifold (NHIM) for f. There exist $\delta_0 > 0$ and $L_F > 0$ (depending on $\mathcal{M}$) such that for $\|f - \hat{f}\|_{C^1} < \delta_0$:*

(i) $\hat{f}$ has a unique NHIM $\hat{\mathcal{M}}$ diffeomorphic to $\mathcal{M}$;

(ii) $d_H(\mathcal{M}, \hat{\mathcal{M}}) \le L_F\|f - \hat{f}\|_{C^1}$.

Theorem 25 (Kupka-Smale Density Theorem[116])**.** *Let $\mathcal{X}$ be a compact smooth manifold and $\mathfrak{X}^r(\mathcal{X})$ be the space of C^r vector fields ($r \ge 1$) endowed with the C^r topology. The set of vector fields $f \in \mathfrak{X}^r(\mathcal{X})$ whose periodic orbits are all hyperbolic and whose stable/unstable manifolds intersect transversally is residual (and thus dense) in $\mathfrak{X}^r(\mathcal{X})$.*

Remark 18 (Intuition: Genericity and Discretization). The Kupka-Smale theorem essentially states that dynamical features (such as non-hyperbolic cycles or coincidental manifold intersections) are rare and fragile. If you randomly perturb a dynamical system, these degenerate features break apart into robust, hyperbolic structures. For our proof, this provides a rigorous justification for **discretization**. While the original system f may possess a continuum of non-isolated periodic orbits (a degenerate feature common in isochronous systems), there exists an arbitrarily close system $f^\dagger$ where this continuum has broken down into a finite skeleton of isolated, stable limit cycles. This allows us to approximate the continuous manifold problem with a finite number of fixed points or limit cycle problem, knowing the error between them is negligible (η_{model}).

Before proving Case 2, we record two ingredients that the existing argument tacitly assumed but never bounded: the *asymptotic phase function* of a hyperbolic limit cycle, and the closeness of the target's and approximator's phase functions on each tile.

Definition 28 (Asymptotic phase function)**.** Let $g \in \mathfrak{X}^1(\mathcal{X})$ have a hyperbolic limit cycle γ of period T with basin $\mathrm{BoA}(\gamma)$. Fix a base point $p_0 \in \gamma$ and write $\gamma(\tau)$ for the orbit through p_0 at $\tau = 0$. The **asymptotic phase function** $\Psi^g : \mathrm{BoA}(\gamma) \to \mathbb{R}/T\mathbb{R}$ is the unique continuous map satisfying

$$\lim_{t\to\infty} \big\|\varphi^g(t, x_0) - \gamma(t + \Psi^g(x_0))\big\| = 0 \qquad \text{for all } x_0 \in \mathrm{BoA}(\gamma).$$

The level sets of Ψ^g are the *isochrons* of γ. For an isochronous NHIM $\mathcal{M}$ foliated by orbits of common period T, the phase function Ψ^g extends continuously to all of $\mathcal{M}$ and to a tubular neighborhood thereof where the normal contraction is well-defined.

Lemma 26 (Phase coherence under tiling)**.** *Let $f \in \mathcal{F}_{\mathrm{CA}}$ have isochronous NHIM $\mathcal{M}$ with global phase function Ψ^f (Definition 28), and let $\hat{f}$ be the corrected approximator constructed in Steps 1–3 of the proof of Theorem 5 Case 2, with discrete cycles $\{\hat{\gamma}_i\}_{i=1}^N$ forming an ϵ_{tile}-net of $\mathcal{M}$ and per-cycle phase functions $\Psi_i^{\hat{f}}$ defined on each basin* $\mathrm{BoA}(\hat{\gamma}_i)$.

There exist constants $C_\Psi, \rho_0 > 0$ depending only on f such that for every x_0 lying in some $\mathrm{BoA}(\hat{\gamma}_i)$ *at distance $\geq \rho_0$ from $\partial\,\mathrm{BoA}(\hat{\gamma}_i)$,*

$$\big|\Psi^f(x_0) - \Psi_i^{\hat{f}}(x_0)\big| \;\leq\; C_\Psi \cdot \big(\|\hat{f} - f\|_{C^1} + \epsilon_{\text{tile}}\big).$$

Proof. The phase function for a hyperbolic limit cycle depends C^1-Lipschitz on the generating vector field at points bounded away from the cycle's basin boundary: this is a standard consequence of Floquet theory and the smooth dependence of the stable foliation on parameters[72] § III.7. Quantitatively, if g_0, g_1 are C^1-close vector fields with hyperbolic cycles γ_0, γ_1 sharing a common basin region $\mathcal{V}$ at distance $\geq \rho_0$ from any separatrix, and with phase functions Ψ^{g_0}, Ψ^{g_1}, then

$$\sup_{x\in\mathcal{V}} |\Psi^{g_0}(x) - \Psi^{g_1}(x)| \;\leq\; C(\rho_0) \cdot \|g_0 - g_1\|_{C^1}. \tag{6}$$

We apply (6) in two stages, comparing f to $\hat{f}$ via the proxy $f^\dagger$ and the base approximator $\tilde{f}$.

Stage 1: $f \to f^\dagger$. By Fenichel persistence (Theorem 24) applied to $\mathcal{M}$, the perturbed manifold $\mathcal{M}^\dagger$ satisfies $d_H(\mathcal{M}, \mathcal{M}^\dagger) \leq L_F \eta_{\text{model}}$. Let $\theta_i^* \in \Theta$ index the original target cycle $\gamma_{\theta_i^*}$ chosen as the cycle on $\mathcal{M}$ closest to the projection of $\gamma_i^\dagger$ onto $\mathcal{M}$. Since $\gamma_i^\dagger \subset \mathcal{M}^\dagger$, we have $d_H(\gamma_i^\dagger, \gamma_{\theta_i^*}) = O(\eta_{\text{model}})$: the transverse component is bounded by $L_F\eta_{\text{model}}$ via the Fenichel bound, and the in-manifold component is absorbed into the same $O(\eta_{\text{model}})$ since $\gamma_{\theta_i^*}$ is chosen optimally. Applying (6) on the basin region $\mathcal{V}_i$,

$$\sup_{x\in\mathcal{V}_i} \big|\Psi^f_{\theta_i^*}(x) - \Psi_i^{f^\dagger}(x)\big| \;\leq\; C(\rho_0) \cdot \eta_{\text{model}},$$

where $\Psi^f_{\theta_i^*}$ denotes the restriction of the global Ψ^f to the foliation basin of $\gamma_{\theta_i^*}$. Since Ψ^f is Lipschitz on $\mathcal{M}$ in the in-manifold direction,

$$|\Psi^f(x) - \Psi^f_{\theta_i^*}(x)| \;\leq\; \mathrm{Lip}_{\mathcal{M}}(\Psi^f) \cdot \mathrm{dist}_{\mathcal{M}}(x, \gamma_{\theta_i^*}) \;\leq\; C_1\epsilon_{\text{tile}}.$$

Combining,

$$\sup_{x\in\mathcal{V}_i} \big|\Psi^f(x) - \Psi_i^{f^\dagger}(x)\big| \;\leq\; C(\rho_0)\eta_{\text{model}} + C_1\epsilon_{\text{tile}}.$$

Stage 2: $f^\dagger \to \hat{f}$. The corrected approximator $\hat{f} = \tilde{f} + \sum_j \alpha_j^* \hat{\mathbf{\Phi}}_j$ satisfies $\|\hat{f} - f^\dagger\|_{C^1} \le \eta_{\text{base}} + N\|\boldsymbol{\alpha}^*\|B_\Phi$ by Lemma 18 and Step 3. Applying (6) once more on each basin,

$$\sup_{x \in \mathcal{V}_i} \left|\Psi_i^{f^\dagger}(x) - \Psi_i^{\hat{f}}(x)\right| \;\le\; C(\rho_0) \cdot (\eta_{\text{base}} + N\|\boldsymbol{\alpha}^*\|B_\Phi).$$

Triangle inequality gives the claim with $C_\Psi = \max\big(C(\rho_0)(1+\kappa N L_T B_\Phi),\, C_1\big)$, recalling $\|\boldsymbol{\alpha}^*\| \le \kappa L_T \eta_{\text{base}}$ from Proposition 22. □

The proof of Theorem 5 splits into two cases.

Proof of Case 1: Manifold of Fixed Points. Let $\mathcal{M}$ be a NHIM consisting entirely of fixed points, i.e., $f(x) = 0$ for all $x \in \mathcal{M}$.

Step 1: Target Selection and Constructive Discretization. The continuum of fixed points is structurally unstable. While the Kupka-Smale theorem guarantees the density of vector fields with hyperbolic critical elements and transversal invariant manifolds, it does not explicitly guarantee that a generic perturbation will yield a spatially dense ϵ-net of attractors. Therefore, we explicitly construct a structurally stable proxy field $f^\dagger$. For any compact continuous attractor $\mathcal{M}$, there exists a Morse function $V : \mathcal{X} \to \mathbb{R}$ whose local minima on $\mathcal{M}$ form an $\varepsilon/6$-net.

We construct the proxy field by applying a global gradient perturbation: $f^\dagger(x) = f(x) - \eta_{\text{tile}} \nabla V(x)$. Generating a dense $\varepsilon/6$-net requires V to have high-frequency spatial oscillations, which increases the C^1 norm of ∇V proportionally to $1/\varepsilon^2$. However, because the spatial configuration is fixed by ε independently of the scaling amplitude η_{tile}, we can choose $\eta_{\text{tile}} > 0$ to be arbitrarily small. This ensures that the global perturbation remains strictly bounded by $\|\eta_{\text{tile}} \nabla V\|_{C^1} < \eta_{model}$.

Because this perturbation is globally C^1-small, it preserves the macroscopic dynamics: the transient approach to the manifold remains largely unperturbed, and global basin boundaries (separatrices) are structurally maintained. Thus, for any $\eta_{model} > 0$ and $\varepsilon > 0$, we construct $f^\dagger$ such that:

1. **Closeness:** $\|f - f^\dagger\|_{C^1} < \eta_{model}$.
2. **Persistence:** The manifold $\mathcal{M}$ persists as a perturbed invariant manifold $\mathcal{M}^\dagger$.
3. **Discretization:** The flow on $\mathcal{M}^\dagger$ possesses a finite set of hyperbolic fixed points $\{\hat{x}_i\}_{i=1}^N$ forming an $\varepsilon/6$-net of $\mathcal{M}$. The remaining $\epsilon/6$ in our spatial budget safely accounts for the structural shift of these equilibria when the final approximation $\tilde{f} \in \hat{\mathcal{F}}$ targets $f^\dagger$.

Step 2: Robust Approximation. We target the stable proxy $f^\dagger$. By the C^1 UAP, choose $\tilde{f} \in \hat{\mathcal{F}}$ such that $\|\tilde{f} - f^\dagger\|_{C^1} < \eta_{\text{base}}$. Since the fixed points $\{\hat{x}_i\}$ of $f^\dagger$ are hyperbolic, they are structurally stable. By the Implicit Function Theorem, $\tilde{f}$ possesses exactly N fixed points $\{\tilde{x}_i\}$ close to the proxy points.

Step 3: Global Error Decomposition. For any initial condition x_0 in the basin of attraction, the trajectory error $\|\varphi(t) - \hat{\varphi}(t)\|$ is bounded by three components, allocated a budget of $\varepsilon/3$ each:

1. **Manifold Error (Transverse):** The physical distance between the true manifold $\mathcal{M}$ and the approximated manifold $\hat{\mathcal{M}}$. Controlled by Fenichel persistence (24):
$$d_H(\mathcal{M}, \hat{\mathcal{M}}) \le L_{\text{NHIM}}(\eta_{\text{model}} + \eta_{\text{base}}) < \frac{\varepsilon}{3}.$$
2. **Tiling Error (Longitudinal):** The error of approximating a stationary point $p \in \mathcal{M}$ with the nearest discrete attractor $\tilde{x}_i \in \hat{\mathcal{M}}$. Since the proxy points form an $\varepsilon/6$-net, and the approximation shift is bounded by $O(\eta_{base}) < \epsilon/6$, any trajectory on the manifold will converge to a sink within distance $\epsilon/3$:
$$\lim_{t\to\infty} \|\varphi(t) - \hat{\varphi}(t)\| \le \max_i \operatorname{diam}(\text{basin}(\tilde{x}_i) \cap \mathcal{M}) < \frac{\varepsilon}{3}.$$

3. **Transient and asymptotic trajectory error (Grönwall + variation of constants).** This component controls the deviation accrued while the trajectories *approach* the invariant manifold transversally; the subsequent *in-manifold* drift from the landing point to the nearest tile equilibrium is not treated here, since it is already charged to the Longitudinal component above, where it is bounded statically by the tile spacing ($< \varepsilon/3$) with no exponential factor. Separating the two is essential: the transverse approach is fast and amplitude-independent, whereas the in-manifold settling is slow, its rate being the $O(\eta_{\text{tile}})$ tile drift.

   Transient phase $[0, T_\perp^*]$. Let $T_\perp^*$ be the uniform time for trajectories from the valid set to enter a fixed-radius Fenichel neighborhood of $\hat{\mathcal{M}}$. This transverse approach is governed by the normal-hyperbolicity contraction rate $\lambda_\perp > 0$ of $\mathcal{M}$, which persists under the $O(\eta_{\text{model}} + \eta_{\text{base}})$ perturbation (Theorem 24) and is bounded below by a constant depending only on $\mathcal{M}$, *independent of the tiling amplitude* η_{tile} *and the resolution* ε. Hence $T_\perp^*$ is finite and $O(1)$, uniformly in η_{tile} and ε. On this phase, Grönwall gives

$$\|\varphi(t, x_0) - \hat{\varphi}(t, x_0)\| \;\le\; \|f - \hat{f}\|_{C^0} \cdot \frac{e^{Lt} - 1}{L}, \qquad t \in [0, T_\perp^*],$$

   with $L = \text{Lip}(f)$.

   Asymptotic phase $[T_\perp^*, \infty)$. After $T_\perp^*$ the trajectory lies in the Fenichel neighborhood, where the normal contraction acts at the rate $\lambda_\perp$ (the normal-hyperbolicity rate, *not* the $O(\eta_{\text{tile}})$ in-manifold relaxation rate of an individual tile). By the variation-of-constants formula [196],

$$\|\varphi(t, x_0) - \hat{\varphi}(t, x_0)\| \;\le\; C\|\varphi(T_\perp^*) - \hat{\varphi}(T_\perp^*)\| \cdot e^{-\lambda_\perp (t - T_\perp^*)} + \int_{T_\perp^*}^{t} C\, e^{-\lambda_\perp (t-s)} \|f - \hat{f}\|_{C^0}\, ds,$$

   where C bounds the flow Jacobian in the linearization regime. The second term is uniformly bounded in t by $C\|f - \hat{f}\|_{C^0}/\lambda_\perp$, so

$$\sup_{t \ge T_\perp^*} \|\varphi(t, x_0) - \hat{\varphi}(t, x_0)\| \;\le\; C\|f - \hat{f}\|_{C^0} \cdot \frac{e^{LT_\perp^*} - 1}{L} + \frac{C\|f - \hat{f}\|_{C^0}}{\lambda_\perp}.$$

   Since $T_\perp^*$ and $\lambda_\perp$ are fixed $O(1)$ constants independent of η_{tile} and ε, and $\|f - \hat{f}\|_{C^0} \le \eta_{\text{model}} + \eta_{\text{base}}$, this transverse contribution satisfies

$$\sup_{t \ge 0} \|\varphi(t, x_0) - \hat{\varphi}(t, x_0)\|_\perp \;\le\; C(\eta_{\text{model}} + \eta_{\text{base}}) \cdot \left(\frac{e^{LT_\perp^*} - 1}{L} + \frac{1}{\lambda_\perp} \right) \;<\; \frac{\varepsilon}{3},$$

   attainable at every resolution ε once $\eta_{\text{model}} + \eta_{\text{base}}$ is small enough, since the bracketed factor is now a fixed constant rather than a quantity that grows (as $e^{L/\eta_{\text{tile}}}$) as the tiling is refined. Combining this transverse bound with the Manifold and Longitudinal components ($\varepsilon/3$ each) by the triangle inequality yields $\sup_{t \ge 0} \|\varphi - \hat{\varphi}\| < \varepsilon$.

Conclusion. By choosing η_{model} and η_{base} sufficiently small, the total uniform error is bounded by ε. The basin mismatch volume is bounded by δ due to the structural stability of the domain boundaries (separatrices) of the hyperbolic proxy $f^\dagger$. □

Proof of Case 2: Isochronous Limit Cycle Manifold. Let $\mathcal{M}$ be a Normally Hyperbolic Invariant Manifold (NHIM) of f foliated by a continuum of periodic orbits $\{\gamma_\theta\}_{\theta \in S^1}$, all having a common period T.

Step 1: Target Selection and Constructive Discretization. The continuum of orbits in f is structurally unstable and cannot be approximated robustly directly. While the Kupka-Smale theorem guarantees that periodic hyperbolicity is dense, we must explicitly construct a perturbation to ensure the resulting isolated limit cycles form a dense ϵ-net without destroying the global basin structure. Therefore, we explicitly construct a structurally stable proxy field $f^\dagger$.

For any compact normally hyperbolic manifold $\mathcal{M}$ foliated by isochronous limit cycles, we introduce a weak transverse drift to break the continuum into discrete orbits. Let $V : \mathcal{X} \to \mathbb{R}$

be a smooth function that oscillates transversally to the limit cycles such that its local transverse minima form an $\epsilon/16$-net of $\mathcal{M}$. We construct the proxy field by applying a global perturbation: $f^\dagger(x) = f(x) - \eta_{\text{tile}} \nabla V(x)$. Generating a dense $\epsilon/16$-net requires V to have high-frequency spatial oscillations, scaling the C^1 norm of ∇V proportionally to $1/\epsilon^2$. Because this spatial configuration is fixed by ϵ independently of the amplitude η_{tile}, we can choose $\eta_{\text{tile}} > 0$ arbitrarily small. This ensures the global perturbation remains strictly bounded by $\|\eta_{\text{tile}} \nabla V\|_{C^1} < \eta_{model}$, preserving the macroscopic transient approach and global separatrices.

Thus, for any $\eta_{model} > 0$ and $\epsilon > 0$, we construct $f^\dagger$ such that:

1. **Closeness:** $\|f - f^\dagger\|_{C^1} < \eta_{model}$.
2. **Persistence:** The manifold $\mathcal{M}$ persists as $\mathcal{M}^\dagger$.
3. **Discretization:** The flow on $\mathcal{M}^\dagger$ possesses a finite number of hyperbolic limit cycles $\{\gamma_i^\dagger\}_{i=1}^N$ forming an $\epsilon/16$-net of $\mathcal{M}$. The remaining $\epsilon/16$ in our spatial budget safely accounts for the structural geometric shift of these cycles when the final approximation $\tilde{f} \in \hat{\mathcal{F}}$ targets $f^\dagger$.

Step 2: Robust Base Approximation. We treat the stable proxy $f^\dagger$ as our target. By the C^1 UAP, there exists $\tilde{f} \in \hat{\mathcal{F}}$ such that $\|\tilde{f} - f^\dagger\|_{C^1} < \eta_{\text{base}}$. Since the limit cycles $\{\gamma_i^\dagger\}$ of $f^\dagger$ are hyperbolic, they are structurally stable. By Lemma 14, for sufficiently small η_{base}, $\tilde{f}$ is guaranteed to possess exactly N limit cycles $\{\tilde{\gamma}_i\}$ close to the proxy cycles.

Step 3: Additive Period Correction. We apply the additive correction strategy (Theorem 4) to $\tilde{f}$. Since the original system f is isochronous, we define the target period for all N discrete cycles to be exactly T.

$$\hat{f} = \tilde{f} + \sum_{i=1}^{N} \alpha_i^* \hat{\mathbf{\Phi}}_i.$$

This creates a phase-locked skeleton on the approximated manifold $\hat{\mathcal{M}}$, where every cycle has period T.

Step 4: Global Error Decomposition. Define the **good set**

$$G \coloneqq \mathcal{X} \setminus \big(S \cup N_{\rho_0}(\partial\mathcal{B})\big),$$

where S is the union of separatrices of $\hat{f}$, $\partial\mathcal{B}$ denotes the basin boundaries between adjacent tiles $\hat{\gamma}_i, \hat{\gamma}_j$, and N_{ρ_0} is the ρ_0-tubular neighborhood from Lemma 26. Since separatrices have codimension ≥ 1, $\mu(\mathcal{X} \setminus G) \to 0$ as $\rho_0 \to 0$, and we choose ρ_0 so that $\mu(\mathcal{X} \setminus G) < \delta$.

For $x_0 \in G$, let $\theta = \theta(x_0) \in \Theta$ be the foliation parameter of x_0 in the target and $i = i(x_0)$ the basin index in the approximator. The trajectory error decomposes as

$$\|\varphi(t, x_0) - \hat{\varphi}(t, x_0)\| \leq \underbrace{\big\|\varphi(t, x_0) - \gamma_\theta(t + \Psi^f(x_0))\big\|}_{\text{(A) target attraction}} + \underbrace{\big\|\gamma_\theta(t + \Psi^f(x_0)) - \hat{\gamma}_i(t + \Psi_i^{\hat{f}}(x_0))\big\|}_{\text{(B) cycle–cycle}}$$
$$+ \underbrace{\big\|\hat{\gamma}_i(t + \Psi_i^{\hat{f}}(x_0)) - \hat{\varphi}(t, x_0)\big\|}_{\text{(C) approximator attraction}}.$$

We allocate $\varepsilon/4$ to each of four contributions: the attraction terms (A) and (C) jointly; the geometric part of (B); the phase part of (B); and the transverse manifold error.

(A) and (C): Exponential attraction to respective cycles. By normal hyperbolicity of $\mathcal{M}$ for f and hyperbolicity of $\hat{\gamma}_i$ for $\hat{f}$, both attraction errors decay exponentially with rates $\lambda_f, \lambda_{\hat{f}} > 0$: (A) $\leq K_f e^{-\lambda_f t}$ and (C) $\leq K_{\hat{f}} e^{-\lambda_{\hat{f}} t}$. For $t \geq T_{\text{trans}} \coloneqq \max(K_f, K_{\hat{f}}) / \min(\lambda_f, \lambda_{\hat{f}}) \cdot \log(8/\varepsilon)$, both terms are $\leq \varepsilon/16$. For $t < T_{\text{trans}}$, the standard Grönwall bound yields $\|\varphi - \hat{\varphi}\| \leq (e^{LT_{\text{trans}}} - 1) \cdot \|f - \hat{f}\|_{C^0}/L$, which can be made $\leq \varepsilon/4$ by choosing $\eta_{\text{base}} + \eta_{\text{model}} + \|\boldsymbol{\alpha}^*\| B_\Phi$ small.

(B-geom) Geometric tile-to-target distance. By Lemma 18, Lemma 23, and the ϵ_{tile}-net property,

$$d_H(\gamma_\theta, \hat{\gamma}_i) \;\le\; \underbrace{d_H(\gamma_\theta, \gamma_{\theta_i^*})}_{\le \epsilon_{\text{tile}}} + \underbrace{d_H(\gamma_{\theta_i^*}, \gamma_i^\dagger)}_{\le L_F \eta_{\text{model}}} + \underbrace{d_H(\gamma_i^\dagger, \tilde{\gamma}_i)}_{\le L_H \eta_{\text{base}}} + \underbrace{d_H(\tilde{\gamma}_i, \hat{\gamma}_i)}_{\le L_H \|\boldsymbol{\alpha}^*\| B_\Phi}.$$

Choosing the budget so the right-hand side is $\le \varepsilon/4$ controls the geometric part of (B).

(B-phase) Phase mismatch. Equal periods alone do not imply that $\gamma_\theta(t + \Psi^f(x_0))$ and $\hat{\gamma}_i(t + \Psi_i^{\hat{f}}(x_0))$ are at corresponding positions on their respective cycles. The phase difference $\Delta\Psi(x_0) \coloneqq \Psi^f(x_0) - \Psi_i^{\hat{f}}(x_0)$ contributes

$$\big\|\gamma_\theta(t + \Psi^f(x_0)) - \hat{\gamma}_i(t + \Psi_i^{\hat{f}}(x_0))\big\| \;\le\; d_H(\gamma_\theta, \hat{\gamma}_i) + L_\gamma \cdot |\Delta\Psi(x_0)|,$$

where $L_\gamma = \sup \|\dot{\gamma}\|$ on the cycles. By Lemma 26, on the good set G,

$$|\Delta\Psi(x_0)| \;\le\; C_\Psi\big(\|\hat{f} - f\|_{C^1} + \epsilon_{\text{tile}}\big) \;\le\; C_\Psi\big(\eta_{\text{model}} + \eta_{\text{base}} + \|\boldsymbol{\alpha}^*\| B_\Phi + \epsilon_{\text{tile}}\big).$$

Choose the budget so $L_\gamma C_\Psi \cdot (\cdots) \le \varepsilon/4$.

Transverse manifold error. By Fenichel persistence (Theorem 24), $d_H(\mathcal{M}, \hat{\mathcal{M}}) \le L_{\text{NHIM}} \cdot \|\hat{f} - f\|_{C^1} \le \varepsilon/4$ for sufficiently small $\eta_{\text{model}} + \eta_{\text{base}}$.

Total. Summing the four budgets, $\sup_{t \ge 0} \|\varphi(t, x_0) - \hat{\varphi}(t, x_0)\| < \varepsilon$ for every $x_0 \in G$, with $\mu(\mathcal{X} \setminus G) < \delta$. This is precisely $\|\varphi - \hat{\varphi}\|_\varepsilon < \delta$. □

Remark 19 (Consistency with Constraint C of Theorem 4). The phase-coherence term in (B-phase) and the Grönwall amplification in (A)+(C) above are precisely the two contributions formalized in Constraint C of the additive-correction proof of Theorem 4 (Eq. 5). The bound (5) on η_{base} derived there propagates directly into the budget cascade for Case 2, so no separate budget needs to be tracked here: a choice of $\eta_{\text{base}}, \eta_{\text{model}}, \epsilon_{\text{tile}}$ satisfying (5) (with ρ_0 chosen small enough that $\mu(\mathcal{X} \setminus G) < \delta$) closes both proofs simultaneously.

F Numerical Experiments

The following experiments illustrate the three theorems and the B/P/D taxonomy of Section 1 on trained Neural ODEs and on real task-trained recurrent networks. They demonstrate the mechanisms used in the proofs; they are not part of them. Unless noted otherwise, we fit a Neural ODE $\dot{x} = \hat{f}_\theta(x)$ with a smooth-activation MLP field (hence C^1) by regressing the vector field in C^1, minimizing $\text{MSE}(\hat{f}, f) + \text{MSE}(\nabla\hat{f}, \nabla f)$, integrate both flows with a fixed-step solver, and report the empirical ε-δ profile

$$\hat{\delta}(\varepsilon) \;=\; \frac{1}{M}\,\#\Big\{\, x_0^{(i)} \;:\; \sup_t \big\|\varphi(t, x_0^{(i)}) - \hat{\varphi}(t, x_0^{(i)})\big\| \ge \varepsilon \,\Big\}, \tag{7}$$

the sample form of Definition 9 over M sampled initial conditions. All Neural ODE fits use double precision; the task-trained flip-flop network is evaluated in its native single precision.

F.1 The three failure modes on planar targets

We fit one Neural ODE per error mode (Figure 4).

B-type (basin): damped Duffing. For $\dot{x} = y,\ \dot{y} = x - x^3 - \frac{1}{2}y$ (two stable spirals at $(\pm 1, 0)$, a saddle at the origin, a curved separatrix), a width-64 field with $\|f - \hat{f}\|_{C^0} = 9.4 \times 10^{-3}$ reproduces both basins with 99.96% agreement (3 of 8100 initial conditions switch basin). The trajectory sup-error is of order 10^{-3} throughout both basins and rises to the basin diameter $D \approx 1.8$ only on a thin collar around the separatrix, precisely the localization of Lemma 12 and Remark 8. At the scale of the wells ($|x| = 1$), $\hat{\delta}(0.05) = 0.014$, $\hat{\delta}(0.1) = 0.007$, $\hat{\delta}(0.2) = 0.004$, falling to the pure basin-flip fraction $3/8100 = 4 \times 10^{-4}$ as ε approaches the basin diameter.

P-type (phase): Van der Pol. For $\dot{x} = y,\ \dot{y} = \mu(1 - x^2)y - x$ with $\mu = 1$[194], the learned cycle matches the true one in shape, but a residual period error $|T - \hat{T}| = 1.25 \times 10^{-3}$ drives a phase-drift beat: the running sup-error grows from 0.038 at $t = 5$ to the full cycle diameter 5.66

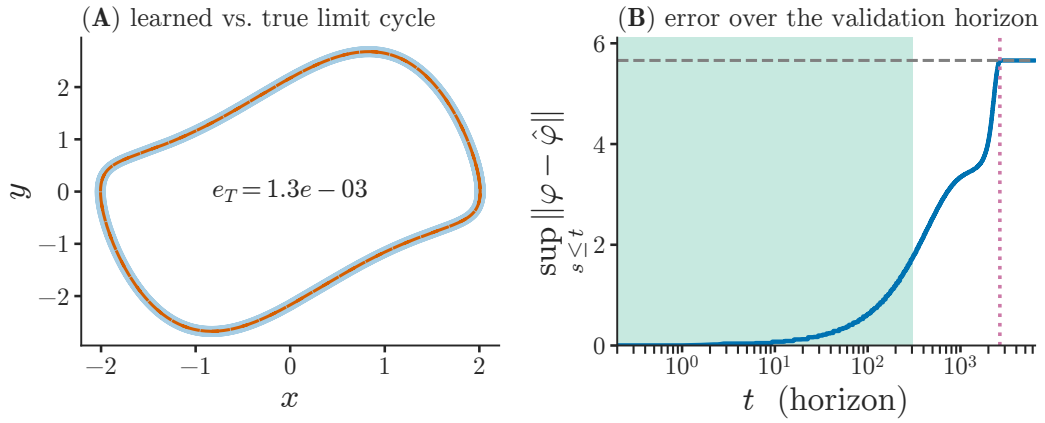


Figure 5: P-type on the Van der Pol oscillator. Left: the learned limit cycle (dashed) coincides in shape with the true one (solid band) almost exactly; the period error is $|T - \hat{T}| = 1.25 \times 10^{-3}$. Right: the running supremum of the trajectory error over a horizon of length t climbs to the cycle diameter D (dashed) only near the detection horizon $t^* \approx 2657$ (dotted); a 45-cycle validation window (shaded, $t \le 300$) captures only a fraction of it, so finite-horizon validation underestimates the drift.

near $t \approx 6.4 \times 10^3$. The detection horizon predicted by the $1/|T - \hat{T}|$ law, $t^* = 2657$, matches the empirical worst-case time 2641. A 45-cycle validation window ($T = 300$) detects essentially nothing, so finite-horizon validation is blind to P-type error; this is the gap that exact period matching (Theorem 4) removes (Figure 5).

D-type (discretization): ring attractor. For $\dot{z} = (1 - \|z\|^2)z$ (the unit circle is a continuum of equilibria), the ring persists (normal contraction $\approx -2 \times 10^{-4}$, so the flow is ε-δ close), but the continuum is discretized into 10 isolated fixed points (5 sinks alternating with 5 saddles), and the non-wandering set is never homeomorphic to S^1. The time-averaged L^2 error stays small (8×10^{-2}) despite the changed topology, which is the tiling regime of Theorem 5 (Figure 8A,B, top row). Figure 4 in the main text shows the B-type fit and its $\delta(\varepsilon)$ curve.

Decomposition: two coaxial limit cycles. To separate the phase mode from the basin mode, we fit a Neural ODE to a system with two stable limit cycles (radii 1 and 2) separated by an unstable cycle at $r = 1.5$ (Figure 6). Without period correction, phase drift fills both basins (P-type), and $\delta(\varepsilon) \approx 1$ up to $\varepsilon \approx 0.5$. The exact period correction of Theorem 4 collapses the bad set to a thin B-type collar at the unstable cycle $r = 1.5$, and $\delta(\varepsilon)$ drops to 0.026, 0.015, 0.004 at $\varepsilon = 0.05, 0.10, 0.30$: phase drift accounts for $\approx 99\%$ of the ε-exceeding initial conditions, the basin collar the rest.

F.2 $\mathcal{F}_{\mathrm{FP}}$ holds across C^1-dense architectures

Theorem 3 assumes only the C^1 universal approximation property (Definition 1), and no specific architecture. We verify this on a three-dimensional target with $2^3 = 8$ hyperbolic fixed points, $\dot{x}_i = x_i - x_i^3$ for $i = 1, 2, 3$, whose attractors are the corners of the cube (Figure 7). Three C^1-dense hypothesis classes, a smooth-activation MLP, a C^1 RNN field $\dot{x} = -x + W \tanh(Vx + b)$, and a radial-basis-function field, each recover all eight attractors with basin agreement above 99.8% and $\hat{\delta}(0.3) \le 6 \times 10^{-3}$. The guarantee follows from C^1 density and holds for every such parameterization, as Theorem 3 states.

F.3 Real task-trained networks lie in the target classes

The theorems concern targets in $\mathcal{F}_{\mathrm{FP}}$, $\mathcal{F}_{\mathrm{LC}}$, and $\mathcal{F}_{\mathrm{CA}}$. We now show that networks trained on standard tasks fall into these classes and are ε-δ approximable.

Working memory as $\mathcal{F}_{\mathrm{FP}}$: the 3-bit flip-flop. We take a gated recurrent network trained on the 3-bit flip-flop task[183] from the Computation-through-Dynamics toolkit[168,197] (latent dimension 64). Its linear readout reconstructs the task output with $R^2 = 0.9999$. FixedPointFinder[65] with zero

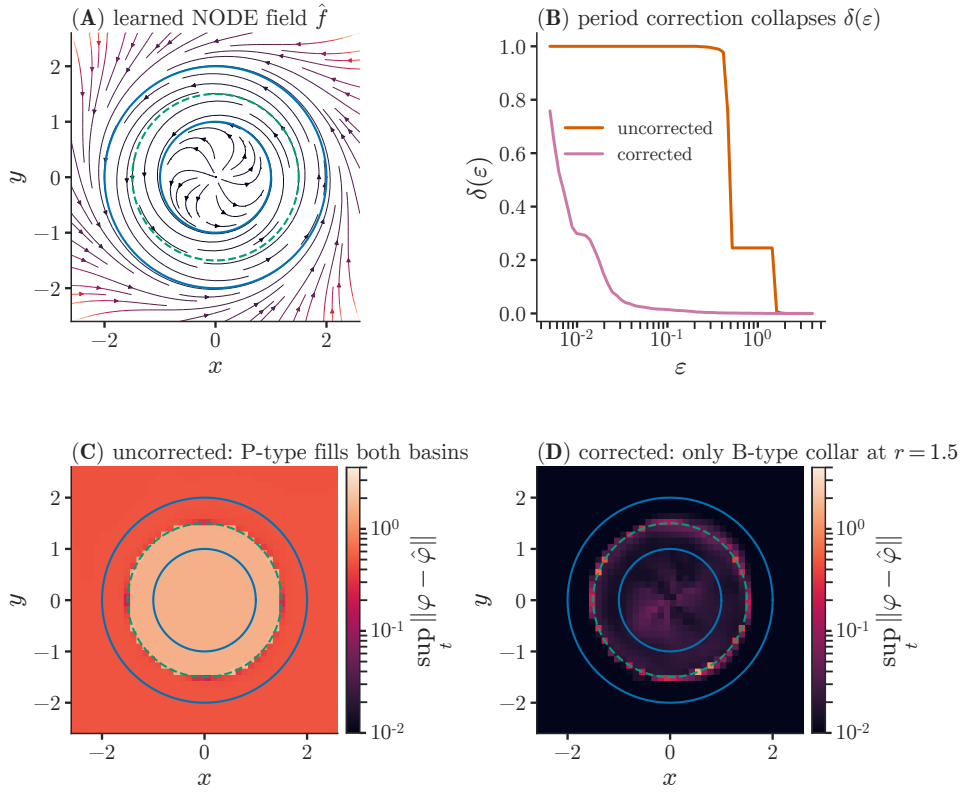


Figure 6: Isolating the phase mode with two coaxial limit cycles (stable at $r = 1, 2$; unstable at $r = 1.5$). **(A)** The learned Neural ODE field recovers both cycles. **(B)** The exact period correction collapses $\delta(\varepsilon)$ from ≈ 1 (uncorrected, phase-dominated) to the basin floor (corrected). **(C)** Uncorrected, phase drift (P-type) fills both basins. **(D)** Corrected, only a thin B-type collar remains at the unstable cycle $r = 1.5$.

input recovers stable fixed points at all 8 memory states (the cube corners), so the trained network is a member of $\mathcal{F}_{\mathrm{FP}}$, a real instance of the multistable target class (Figure 4C, main text; the linear readout has $R^2 = 0.9999$).

A Neural ODE approximates the trained network. Writing F for the network's gated recurrent cell, the discrete update $h_{t+1} = F(u_t, h_t)$ is exactly forward Euler (step $\Delta t = 1$) of the ODE $\dot{x} = g^*(x, u)$ with $g^*(x, u) = F(u, x) - x$; the trained network is thus the $\Delta t = 1$ Euler discretization of that ODE, not the ODE itself. Setting $u = 0$ gives the autonomous field, whose eight hyperbolic fixed points are the memory states, and this is the object $\mathcal{F}_{\mathrm{FP}}$ addresses. We fit a separate MLP field $\hat{f}_\theta \approx g^*$ (a different architecture, carrying its own nonzero approximation error) to the residual field over a tube around the trajectories, fine-tuned through the rollout, and ask whether it ε-δ approximates the network. Autonomously ($u = 0$), the approximation preserves the whole fixed-point structure over the infinite horizon: rolling out the network and the Neural ODE from the same states, they settle on the same memory state for 99.8% of initial conditions with $\hat{\delta}(0.3) = 6 \times 10^{-3}$, unchanged out to 5000 steps, and all eight memory states are inherited (Figure 7H). Under the task inputs the Neural ODE tracks the network's own (imperfect) readout at 99.2% per-bit accuracy (Figure 7F,G); there the trajectory sup-error is larger ($\hat{\delta}(0.3) = 0.32$, median 0.25) because a bit flip is a nearly discontinuous one-step transition, so a continuous flow has a brief transient mismatch there. This is the discrete-to-continuous analogue of the fast-transient region, where ε-δ closeness holds even though C^1 closeness does not.

Continuous attractors as $\mathcal{F}_{\mathrm{CA}}$: head-direction integration. A 64-unit rectified-tanh recurrent network trained to integrate angular velocity on S^1 [166] holds heading on a ring attractor, the canonical continuous-attractor model of the head-direction circuit, whose ring topology has been recovered

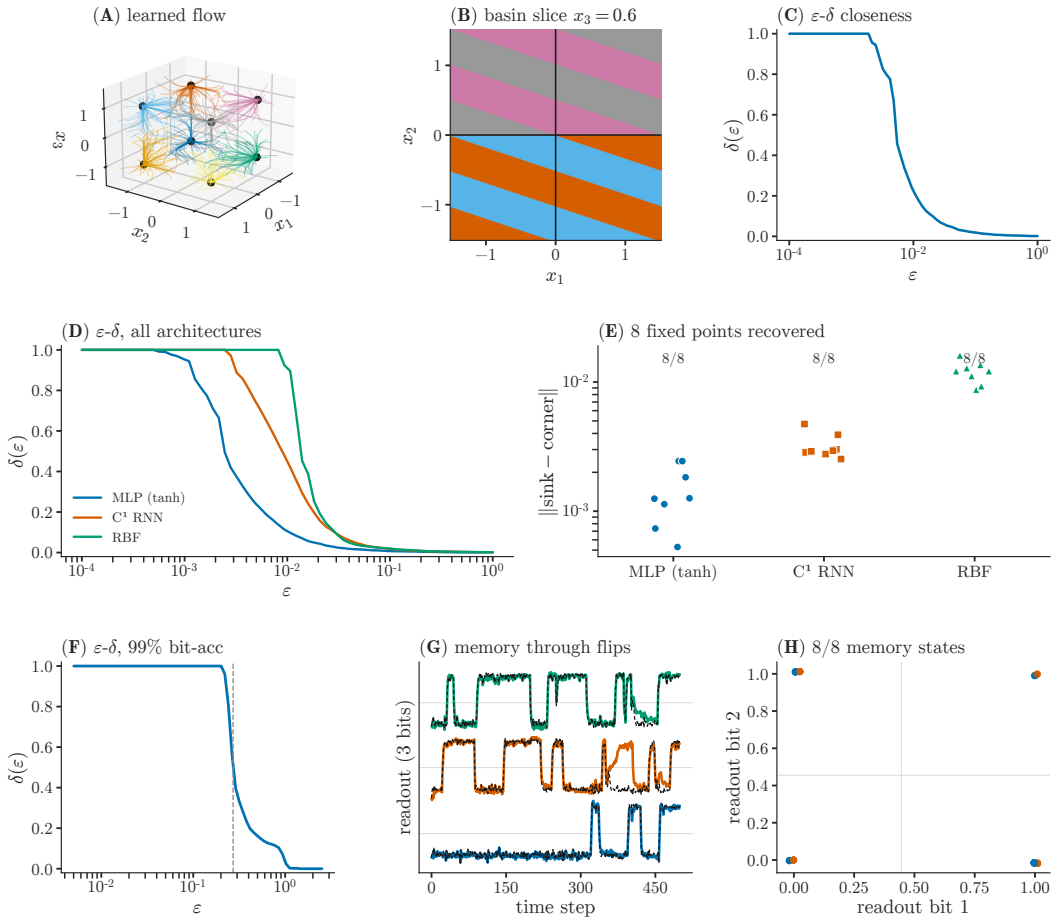


Figure 7: $\mathcal{F}_{\mathrm{FP}}$ across architectures and on a real trained network. **(A–C)** A 3D eight-fixed-point target $\dot{x}_i = x_i - x_i^3$: trajectories from many initial conditions settle onto the eight cube-corner attractors, colored by basin (A); a basin slice (B); and the ε-δ curve (C). **(D, E)** Three C^1-dense hypothesis classes, a smooth-activation MLP, a C^1 RNN field, and a radial-basis-function field, each reach ε-δ closeness (D) and recover all 8 fixed points (E), with $\hat{\delta}(0.3) \leq 6 \times 10^{-3}$. **(F–H)** A Neural ODE ε-δ approximates a gated recurrent network trained on the 3-bit flip-flop task, scored on the network's own readout: the ε-δ curve at 99.2% per-bit accuracy (F); the three bits tracked through many memory flips, network (solid) versus Neural ODE (dashed) (G); and the Neural ODE's autonomous field, with stable fixed points at all 8 memory states (H).

from neural population recordings[22]. A finite network cannot hold a true continuum; it instead realizes a structurally stable approximation of it[166], a normally attracting ring carrying isolated hyperbolic fixed points. These are the zeros of the along-ring drift (the autonomous field projected onto the ring tangent), which as the sign changes of a scalar on S^1 alternate sinks and saddles in equal number; the network we use tiles the ring into 12 sinks alternating with 12 saddles, all clearly hyperbolic (median leading eigenvalue -0.20), and a dense rollout of held headings converges onto exactly these discrete sinks. Because the fixed points are hyperbolic they are structurally stable, so any C^1 field sufficiently close to the network's must inherit them. We verify this by fitting an independent Neural ODE $\hat{g}(x) = -x + \mathrm{MLP}(x)$ (a continuous-time recurrent form, whose $-x$ leak guarantees global stability) to the network's autonomous field in a neighbourhood of the ring: its fixed points coincide with the network's (22 of the 24 within 0.15 rad; Figure 8C), and its held-heading drift $\delta(\varepsilon)$ matches the network's (Figure 8D). This is exactly the setting of Theorem 5: a finite model discretizes the continuum into isolated headings, the D-type mechanism of head-direction drift, and the discrete fixed-point structure transfers to the approximating flow.

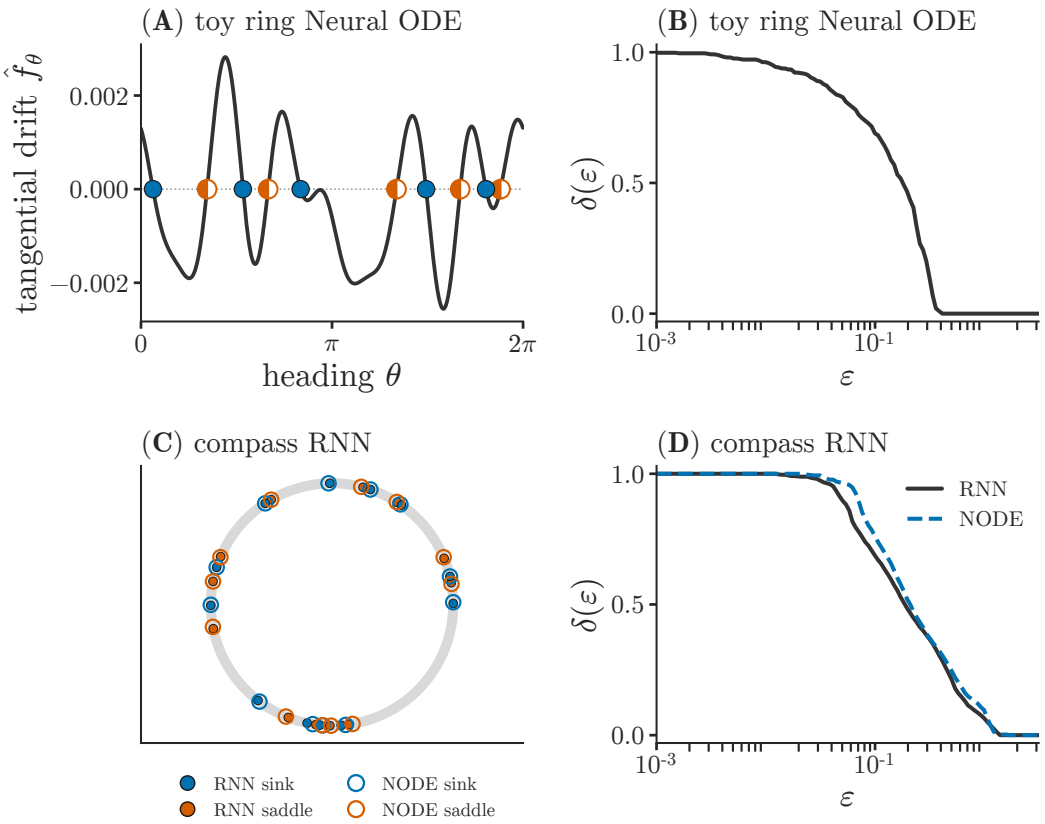


Figure 8: Continuous attractors are discretized (D-type), and the discretization transfers to a Neural ODE. Top: a Neural ODE fit to the analytic ring attractor $\dot{z} = (1 - \|z\|^2)z$. Bottom: a 64-unit rectified-tanh network trained to integrate angular velocity[166], and an independent Neural ODE fitted to it. **(A)** The toy ring's along-ring tangential drift; its zeros are isolated fixed points (filled circle: sink; half-filled: saddle), so the continuum is tiled by discrete equilibria, 5 sinks alternating with 5 saddles. **(B)** ε-δ closeness of the toy Neural ODE to the analytic continuum: $\delta(\varepsilon)$ is the fraction of held headings whose trajectory error exceeds ε. **(C)** Fixed points of the trained network on the ring, in canonical (readout) coordinates: 12 sinks (blue) alternating with 12 saddles (orange), all hyperbolic. Filled markers are the network's fixed points, open markers those of the fitted Neural ODE; they coincide (22 of 24), the structurally-stable transfer of the discrete fixed-point set. **(D)** Held-heading drift $\delta(\varepsilon)$ for the network (solid) and the Neural ODE (dashed): both discretize the continuum, and the Neural ODE reproduces the network's drift.

F.4 Budget and width scaling

Finally, we check the budget cascade of the proofs quantitatively. On a one-parameter bistable family we fit the field accuracy η_0 required to reach a target (ε, δ) to a power law $\eta_0 \propto \varepsilon^a \delta^b$. Across saddle rates spanning $L/\mu_u \in [1, 18]$ we measure $a = 0.94$ and $b = 1.02$ on average, so $\eta_0 = \Theta(\varepsilon\delta)$: the required vector-field accuracy is the symmetric product of the tolerance and the reliability, consistent with the linear basin-error scaling of Remark 8. For the planar targets the trajectory error is linear in the field error through the origin (slope 1.18 for Duffing), matching $\eta_{\mathrm{traj}} = \lambda\varepsilon/K$ of Lemma 11, and the approximation-limited field error scales as $\eta(W) \sim CW^{-p}$ with $p \in [1.1, 1.35]$, giving a width requirement $W \sim (\varepsilon\delta)^{-1/p}$ for these finite-width fields. These are composed sufficiency bounds and involve the problem constants L, λ, μ_u, so they bound the required width from above; a matching lower bound is left open.

G Extended Literature Review

This appendix provides a comprehensive review of universal approximation results for dynamical systems, establishing the context and novelty of our contributions.

G.1 Approximation on Finite Time Intervals

Universal approximation theorems for dynamical systems primarily address two regimes: finite-time simulation via high-dimensional embedding and direct trajectory approximation in the native state space.

Foundational results. Seminal works by Li [119], Funahashi and Nakamura [60] and Doya [44] established that RNNs approximate general dynamical systems by embedding dynamics into high-dimensional hidden states. This lineage was extended to the approximation of continuous functionals and nonlinear operators by Chen and Chen [26] and Chen and Chen [27], broadening the scope of universality to map input signals to outputs. This lineage of results extends back to homogeneous state-affine systems, which were shown to possess similar universality properties in the discrete-time setting for compact time intervals [56,178,179]. Crucially, these guarantees are *strictly limited to bounded time intervals*, leaving long-term asymptotic behavior unconstrained.

Control-theoretic universality. A complementary route to universality goes through geometric control theory: viewing a deep residual network or Neural ODE as a control system, its flow can be steered to approximate an arbitrary target map precisely when the constituent vector fields are bracket-generating, which is the nonlinear controllability criterion of Hermann and Krener [83]. This yields universal approximation and interpolation guarantees for residual networks and Neural ODEs [1,37,164,187], and it is what underwrites the C^1 universal approximation property we assume of the hypothesis class (Definition 1). As with the embedding results above, however, these guarantees are finite-time: they control the flow map at a terminal time (for classification, interpolation, or transport), not the asymptotic behavior as $t \to \infty$ that our ε-δ conditions constrain. The dual notion of observability [83,185] governs whether such a vector field can in turn be identified from partial observations, a prerequisite for state-space interpretability that is complementary to, and outside the scope of, our approximation results.

Grönwall-based bounds. Most finite-time results rely on Grönwall's inequality to control trajectory differences, incurring exponential degradation of accuracy over time [32,60,122,180]. For time horizon T, if $\|f - \hat{f}\|_{C^0} < \eta$, then $\|\varphi(t) - \hat{\varphi}(t)\| \leq \eta(e^{Lt} - 1)/L$, which grows unboundedly as $t \to \infty$.

Discrete-time sequences. RNNs act as universal approximators for various temporal mappings, though often constrained to finite-time horizons. They can uniformly approximate the dynamics of continuous finite-memory systems on compact domains [75] and state-space trajectories over closed discrete-time intervals [100,153]. More recent work has explicitly established the finite-time universality of discrete-time RNNs [2] and linear recurrences with nonlinear projections [147], alongside their ability to map continuous past-dependent sequence-to-sequence functions [177]. Furthermore, CNNs have emerged as effective alternatives for modeling sequences with finite memory [5,99].

Continuous-time trajectories. The universal approximation properties of static neural networks [59] initially established them as powerful tools for modeling nonlinear systems [145,201], often employing multilayer perceptrons or radial basis functions [23,25,31,62,188]. Expanding this to dynamic contexts, the capabilities of continuous-time recurrent networks for autonomous systems were demonstrated by Funahashi and Nakamura [60] and extended by Kimura and Nakano [105] and Chow and Li [32]. For non-autonomous systems, any trajectory can be approximately realized on finite intervals [144], a property also held by Neural ODEs for input-driven systems [120,209]. Furthermore, even networks with randomly initialized, fixed weights can achieve universal approximation of continuous functions and dynamical systems over compact intervals by learning only the bias parameters [202]. Recent Neural ODE variants add frequency-domain [3] and energy-behavior-preserving [20] parameterizations, apply universal differential equations [158] to neural population data [53], and enable data-driven phase control of learned limit-cycle oscillators [205].

For non-autonomous systems, where the dynamics explicitly depend on external inputs, it has been shown that the states of the output units of a continuous-time RNN, can approximate the solution of ODEs on compact time intervals[62] and linear, continuous and regular functionals[124]. A subset of an RNN's units can approximate any smooth (C^1) dynamical system (input-affine) with arbitrary precision for finite time[102]. Neural oscillators have the universal approximation property on compact time intervals (approximation of causal and continuous operators)[90,117]. Any finite trajectory of an n-dimensional continuous dynamical system can be approximated by the internal state of the hidden units and n output units of a Liquid time-constant (LTC) network[80], since extended with attractor-based memory for sequence modelling[115]. Finally, Maass et al.[130] extended these results to biological feedback models, proving that neural networks with dynamic synapses are universal approximators for dynamical systems, particularly in closed-loop feedback configurations.

Flow approximation. Flow-based approximations for finite time have been established for architectures including Recurrent High-Order Neural Networks[112]. Any continuous dynamical system can be approximately realized on finite intervals by an RNN[32], extended to time-variant systems with fixed initial states[122]. More recently, Tabuada and Gharesifard[186] established that Neural ODEs are universal approximators for monotone analytic functions homotopic to the identity, proving that the flow of a single-layer system with time-varying parameters can approximate such mappings over compact sets. Despite these advances in representing complex mappings through flows, these results remain confined to finite-time horizons and do not address the long-term topological stability of multistable systems.

G.2 Fading Memory Systems

For infinite time horizons, existing results predominantly require the **fading memory property** (FMP).

Definition 29 (Fading Memory Property[14])**.** An operator H mapping input histories to outputs possesses the FMP if it is continuous with respect to a weighted norm that discounts the past: there exists a weighting function $w(z) \to 0$ as $z \to \infty$ such that closeness in the weighted past implies closeness in output.

Topological constraint. The FMP imposes a severe restriction: systems with fading memory "forget" initial conditions, implying global asymptotic stability (monostability). This *inherently excludes multistable systems*, those with multiple fixed points or limit cycles where long-term behavior depends on initial conditions. A simple bistable system has no fading memory: the final state depends entirely on the basin of attraction.

Reservoir computing. Echo State Networks (ESNs) were shown to be universal for fading memory systems by Jaeger[96], formalized by Grigoryeva and Ortega[70,71]. Liquid State Machines achieve universal computation for filters with FMP[129]. Various reservoir architectures approximate FMP systems: linear systems with polynomial readouts[14], state-affine systems[68], and Simple Cycle Reservoirs[118].

Other stability conditions. Various results establish universality for systems with decaying memory, including linear RNNs for regular functionals[123,125], State Space Models for exponentially decaying memory[200], and fixed reservoirs with linear readouts[204]. Extending this to infinite time, Hanson and Raginsky[76] and Bishop[10], Bishop et al.[12] provide guarantees for deterministic and stochastic RNNs, respectively. However, these results strictly require incremental stability or exponential contraction to bound error propagation. While effective for filtering, these conditions mathematically enforce a single global attractor, thereby precluding the history-dependent bifurcations and multistability essential for cognitive tasks.

FMP trivializes approximation. Every fading-memory system[169] can be uniformly approximated by linear state dynamics with a nonlinear readout[136]. Typically, the FMP is used to reduce the approximation problem to one over a finite, bounded index set, and then appeal to the density of fully connected neural network to obtain approximation[67]. This constraint is so restrictive that Boyd and Chua[14] proved FMP is the necessary and sufficient condition for a system to be uniformly approximated by a **finite Volterra series**[58,198]. The FMP implies "asymptotic independence" of

the state from initial conditions[33,133,140], making multistability impossible. Our work addresses this fundamental gap.

This result was extended to include approximation of systems with an input[134]. More recently, these results were rigorously formalized by Grigoryeva and Ortega[70,71], who proved that ESNs are universal uniform approximators for discrete-time fading memory filters, subject to uniformly bounded inputs. In the domain of spiking neural networks, a parallel result holds: Liquid State Machines (LSMs) achieve universal computational power for all time-invariant filters that exhibit the FMP[129].

Generalizing these results, we observe that universality is not unique to standard reservoir architectures. If the domain of the functional H is restricted to a space of uniformly bounded sequences with the fading memory property, various families of state-space transformations can approximate it uniformly. These include linear systems with polynomial or neural network readouts[14,36,66,71], state-affine systems with linear readouts[68,71], and the previously mentioned echo state networks[66,67,69,70]. More recently, Simple Cycle Reservoirs were also shown to be universal for this set of systems[118]. Departing from these state-space approaches, a new method has been developed to approximate FMP systems using a kernel representation of the model[91].

Dehghani et al.[41] introduced the recurrently-stacked Universal Transformer, and Yun et al.[208] showed that transformer models are universal approximators of continuous permutation-equivariant sequence-to-sequence functions with compact support. Independently, Bai et al.[6] proposed the Deep Equilibrium Model (DEQ), an implicit-depth architecture based on root-finding for steady states of weight-tied deep networks; DEQs can approximate any sequence whose underlying dynamics converge to a fixed point. The DEQ, which can be seen as an implicit-depth model similar to neural ordinary differential equations, was later extended to capture systems with arbitrary invariant sets; however, its universal approximation property has not yet been established[108]. Finally, assuming a global attractor property, infinite time approximation can be guaranteed[137,207].

G.3 Approximation of Diffeomorphisms and Flows

Flow maps. Neural ODEs and invertible neural networks are universal approximators for diffeomorphisms[89,95,98,114,135,189,190]. For dynamical systems, this guarantees only finite-time approximation: the flow φ_t is a diffeomorphism for each fixed t, but approximating φ_T does not control behavior for $t > T$.

Topological conjugacy. ESNs can be trained to have topologically conjugate dynamics to structurally stable systems[77]. For NODEs, this follows immediately from structural stability (Theorem 1): C^1-close vector fields yield topologically equivalent flows. Our contribution is making this observation precise with quantitative ε-δ bounds. Finally, we would like to mention the possibility to construct dynamics with Cohen-Grossberg networks for arbitrary sets of fixed point and limit cycles through embedding of dynamics[35]. Finally, models can be constructed to realize attractors[39], though general infinite-horizon guarantees for these results are lacking.

G.4 Why Multistability Requires New Theory

Fundamental obstructions. Multistable systems exhibit two failure modes absent in FMP systems:

1. **B-type error**: Near separatrices, arbitrarily small perturbations cause trajectories to converge to different attractors (Figure 2).
2. **P-type error**: For limit cycles, any period mismatch causes unbounded phase drift as $t \to \infty$.

Structural stability as the key. Morse-Smale systems are the natural target class: they are structurally stable (small perturbations preserve qualitative dynamics) yet allow multistability. This paper provides universal approximation results exploiting this property.

Outside the Morse-Smale class. Chaotic attractors (C-type error) require fundamentally different metrics, for example, invariant measures or attractor reconstruction[42,78], and remain outside our framework.

G.5 Computational Universality

For completeness, we note that RNNs are Turing universal[104,142,143,156,171,172]. This computational universality mirrors that of smooth ODEs[15], Hopfield nets[173], and finite automata simulations[94,111,113,182]. However, computational universality is orthogonal to approximation: it concerns symbolic computation, not trajectory tracking. Our results address the latter.

G.6 Overview

Table 1: Comprehensive Landscape of Universal Approximation for Dynamical Systems

Category	Citations	Hor.	Metric / Guarantee	Stability / Constraint	Asymptotic Property
Computation	15,94,104,111,113,142,143,156,171–173,182	Inf.	Turing universality	Analog computation	Recursive equivalence
RNN	44,60,119	Fin.	L^∞ Trajectory tracking	Compact state space	Local existence
Operator	23,25–27,31,62,80,90,102,117,145,188,201	Fin.	Operator norm	Continuous signals	Signal mapping
Fin. seq.	2,5,41,75,99,100,147,153,177,208	Fin.	Uniform convergence	Finite time window	Temporal mapping
Grönwall	32,60,122,180	Fin.	Exp. error bound	Lipschitz continuity	Vacuous ($t \to \infty$)
Flow	120,121,144,151,157,186,209	Fin.	Trajectory realization		Flow approximation
Diffeo.	89,95,98,114,135,189,190	Fin.	Diffeo. mapping	Homotopy to identity	Coordinate warping
FMP	14,66–71,96,118,129,130,136,169,207	Inf.	Decaying history	Unique global eq.	Global monostability
Contractive	6,10,12,108,123,125,200	Inf.	Stationary L^p error	Exp. p-contraction	Unique measure
Incr. stab.	33,76,133,140	Inf.	Uniform tracking	Asymptotic incr. stab.	Trajectory sync.
Top. equiv.	35,77	Inf.	Qualitative tracking	Structural stability	Topological conjugacy
Formal Sim.	63,64,155,195,210	Inf.	Bisimulation distance	Metric / Symbolic	Simulation equivalence
This Work	**Thrm. 3, 4, 5**	**Inf.**	**ε-δ closeness**	**Morse-Smale / NHIM**	**Multistable attractors**

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: We provide full proofs for each theorem, and Section 6 illustrates the guarantees and the failure-mode taxonomy on trained Neural ODEs and real networks.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: Section 7 discusses the limitations: the theorems guarantee existence, not learnability; exact period matching is not structurally stable; chaotic targets are excluded; and stochastic and input-driven systems remain open.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes]

Justification: We fully characterize the class of systems for which we then show the universal approximation property of Neural ODEs for infinite time horizon.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: The experiments are demonstrations of the theory. Section 6 and Appendix F specify each target system, the hypothesis-class architectures, the C^1 vector-field fitting objective, the integration scheme, and the reported metrics ($\hat{\delta}(\varepsilon)$, basin agreement, period error), which suffice to reproduce the results.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in

some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: Code reproducing all experiments, with the trained checkpoints, is public at `https://github.com/Asagodi/uap-infinite-time-code` (Section 6). The flip-flop experiments additionally use the public Computation-Through-Dynamics toolkit[45].

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: Section 6 and Appendix F give the target systems, network architectures, the fitting objective, the optimizer, and the evaluation protocol.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [No]

   Justification: The experiments are deterministic demonstrations; the reported $\hat{\delta}(\varepsilon)$ is itself a measure over thousands of sampled initial conditions (Eq. 7), and basin agreement is computed over the same population, so the sample size characterizes variability. We do not report error bars over random seeds.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: All experiments run on a laptop CPU (Intel Core i7-1165G7, 4 cores, 16 GB RAM), with no GPU; all Neural ODE fits are in double precision (Appendix F). Fitting each planar Neural ODE takes 3–9 minutes and each of the three eight-fixed-point architectures 1–11 minutes; the flip-flop Neural ODE, which adds trajectory fine-tuning at widths 448 and 640, takes about 5.5 hours.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: We complied with the NeurIPS Code of Ethics.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [N/A]

    Justification: For such a theoretical paper, we believe that such potential impacts are too far removed.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
    - The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
    - If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: This is a theoretical paper.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

Justification: The trained-network experiments use the open-source Computation-Through-Dynamics toolkit[45] and FixedPointFinder[65], and the head-direction network of Ságodi et al.[166]; all are cited. FixedPointFinder is released under the Apache 2.0 license; the Computation-Through-Dynamics toolkit repository states no license, and we use its published code and trained network for research only.

Guidelines:

- The answer [N/A] means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [Yes]

Justification: The released code and trained checkpoints (`https://github.com/Asagodi/uap-infinite-time-code`, MIT license) are documented in the repository README, which gives the setup, the command that produces each figure, and the source of each checkpoint.

Guidelines:

- The answer [N/A] means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [N/A]

Justification: No human subjects were used.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: No human subjects were used.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

    Answer: [N/A]

    Justification: We did not make any usage of LLMs apart from writing and editing.

    Guidelines:

    - The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.