APPROXIMATION OF DYNAMICAL SYSTEMS BY THE KURAMOTO MODEL

Anonymous authors
Paper under double-blind review

ABSTRACT

Analog computing leverages physics to perform computation energy efficiently in hardware. These systems typically have restrictive dynamics, and their expressive power is poorly understood. We address this question of expressivity for the Kuramoto model, an analog compute model that leverages oscillator coupling behavior. We establish a universal dynamical system approximation theorem: An autonomous Kuramoto network can approximate the flow of any continuously differentiable vector field on a torus over a finite time interval, uniformly over initial conditions, without a learned output map or external forcing. The result extends to smooth periodic coupling functions with a nonzero first Fourier harmonic. Experiments on seven dynamical systems validate our trajectory approximation approach, and flow-matching experiments show that Kuramoto networks can approximate learned flows, suggesting a promising use of the Kuramoto model in generative modeling.

1 INTRODUCTION

Analog computing leverages the evolution of physical phenomena over time to perform computation. Machine learning models that use analog computation are similar to other dynamics-based neural network architectures such as RNNs, SSMs, and neural ODEs, but can be implemented in hardware very energy efficiently, achieving orders of magnitude lower energy consumption than digital (Marković et al., 2020). In contrast to dynamics-based NN architectures which use neural networks to define vector fields, the vector fields for analog compute models are typically governed by a highly restrictive master differential equation that limits the degree and shape of the interactions between variables. These restrictions enable efficient hardware implementation at the cost of expressivity.

In this work, we focus on the expressivity of the *Kuramoto model*, an analog compute model based on oscillator coupling dynamics that has recently become popular in the hardware community. The Kuramoto model has been incorporated into machine learning architectures that use recurrent, graph, and attention-like modules and used for tasks such as classification and generative modeling (Rusch et al., 2022; Nguyen et al., 2024; Miyato et al., 2025; Unconventional AI, 2026). In these settings, the autonomous dynamics of the system is leveraged to perform computation. All of these usages leverage the *transient behavior* of the system, typically over a short time scale.

The approximation capabilities of the autonomous Kuramoto model dynamics are not well-understood. Prior works that have proven universality of the Kuramoto model have augmented the system dynamics, either by requiring a readout transform on the measured phases or introducing a forcing function, rendering the system non-autonomous and requiring additional hardware (Chiba et al., 2024). Empirical studies have shown the autonomous dynamics alone have difficulty approximating arbitrary vector fields (Wang & Achour, 2026). So this leaves an open question: is the base physics model just being trained incorrectly, or is its expressivity fundamentally limited?

In this work, we prove that the autonomous dynamics of Kuramoto model alone is a universal approximator of dynamical systems, capable of approximating vector fields to within $\zeta > 0$ error bound, and provide an instantiation that works well as an approximator. This approximation result holds under two practical assumptions over the target dynamical system: (1) integration time is finite, (2) the dynamical system's flow-field is bounded or periodic. In the pursuit of this universality result, we make several contributions that together enable empirical evaluation of approximation power:

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

- ▷ We the introduce lagged fast–slow (LFS) model which uses both fast-slow dynamics and phase lags and prove its universality. The hidden state of the system and the timescale separation together increase approximation power of the system. We map the LFS model to the Kuramoto model, proving the Kuramoto model's universality.
- ▷ In our proof strategy, we develop analytic mappings that translate the LFS model to the Kuramoto model and to a complex linear combination of normalized affine vectors (NAVs). These mappings are used in our empirical study to translate between representations, enabling ergonomic training on NAVs and deployment on the Kuramoto model.
- ▷ We conduct experiments to explore the approximation capabilities of our universal Kuramoto model instantiation. We test the role of hidden state size and timescale separation on Kuramoto model approximation power. We demonstrate that a cascading network of Kuramoto oscillators can emulate a flow field trained with flow matching for synthetic distribution matching tasks.

2 Background and Related Work

The *Kuramoto Model* leverages the physics of coupled oscillators to perform computation. Given a population of N coupled oscillators with phases $s_1 \dots s_i \dots s_N$, the evolution of the phase over time is governed by the following master differential equation:

$$\dot{s}_i = \omega_i + \sum_{j=1}^{N} C_{ij} \sin(s_j - s_i), \qquad i = 1, \dots, N. \tag{2.1}$$

Each state is an oscillator phase s_i, where the phase space is $\mathbb{T} = \mathbb{R}/2\pi\mathbb{Z}$ and the phase $s \in \mathbb{T}^N$. Each oscillator evolves at its own intrinsic frequency ω_i and is pulled toward or away from the other oscillators by a sinusoidal interaction (Kuramoto, 1975). Depending on the spread of the frequencies and the strength of the couplings C_{ij}, the population drifts incoherently, locks into partial coherence, or synchronizes. The model is widely used to describe synchronization in physical, chemical, and biological systems (Acebrón et al., 2005). Observe the above master differential equation is restrictive, supporting only two-body interactions and a sine phase-difference non-linearity.

The Sakaguchi-Kuramoto model (Sakaguchi & Kuramoto, 1986) is a generalization of the Kuramoto model that supports *phase lags*, and is briefly used in the proof:

$$\dot{s}_i = \omega_i + \sum_{j=1}^{N} C_{ij} \sin(s_j - s_i - \tau), \qquad i = 1, \dots, N. \tag{2.2}$$

The phase lag parameter τ shifts the synchronization frequency, and is related to frustration in physics.

2.1 Related Work on Oscillator Universality

Table 1 summarizes the closest universality results under this distinction between the dynamics and a learned output transformation. Continuous-time recurrent networks can approximate finite-time trajectories of finite-dimensional systems by augmenting the visible state with hidden units (Funahashi & Nakamura, 1993; Kimura & Nakano, 1998). Low-rank recurrent populations can similarly shape arbitrary low-dimensional vector fields on bounded domains (Beiran et al., 2021), while recent Neural ODE results obtain infinite-horizon guarantees for structured dynamical classes (Sagodi & Park, 2026) or flows that preserve prescribed fixed points (Pacifico et al., 2026). These models use flexible neural dynamics or population-level coordinates, unlike the Kuramoto model, whose dynamics are much more restrictive: each state variable is a phase and every interaction is a periodic pairwise coupling.

Universality with oscillatory architectures. Oscillators have been used as approximation primitives. Banks of oscillators approximate a prescribed time signal through a learned output transformation that combines their oscillations (Baldi & Hornik, 1995). In Lanthaler et al. (2023), the second-order neural oscillators are externally forced by an input signal and use learned affine transformations both in their state interactions and at their output to approximate the requested causal operator. Oscillator-inspired recurrent and graph architectures likewise wrap an oscillatory or synchronization-based update in learned neural transformations of the state (Rusch & Mishra, 2021a;b; Rusch et al., 2022; Nguyen et al., 2024; Miyato et al., 2025; Ceni et al., 2024), augmenting the Kuramoto dynamics to obtain a

Table 1: Summary of universality results for oscillator networks and related dynamical systems.

Work	Architecture and Target	Output Transformation	Distinction from this work
Funahashi & Nakamura (1993)	Continuous-time RNN approximating finite-time trajectories.	None.	Uses general affine mixing and nonlinear activations rather than phase states and periodic pairwise interactions.
Lanthaler et al. (2023)	Forced second-order neural oscillators approximating causal operators.	Fitted affine readout.	Allows external inputs and general neural nonlinearities; the oscillator state is not an autonomous target flow.
Baldi & Hornik (1995)	Oscillator modules approximating a prescribed time curve.	Fitted oscillator combination.	Approximates one temporal signal rather than a state-dependent vector field uniformly over initial conditions.
Chiba et al. (2024)	Forced infinite-population Kuramoto reservoir with complete temporal features.	Fitted order-parameter readout.	Establishes temporal readout universality, not universality of a finite autonomous Kuramoto vector field.
This work	Finite autonomous Kuramoto model with extension to periodic couplings.	None.	Approximates arbitrary C^1 vector fields and finite-time flows uniformly over initial conditions.

good approximator. In contrast, we show the Kuramoto model approximates an autonomous target flow without external forcing, a learned output transformation, or general neural couplings.

Reservoir computing and learning with the Kuramoto model. Reservoir methods share the same limitation as the oscillator architectures above: an input transformation drives the reservoir and a learned output transformation reads out the target, whether for generic reservoirs (Grigoryeva & Ortega, 2018) or Kuramoto reservoirs specifically, where a forcing signal drives the oscillators and a learned readout of their order parameters approximates the driven trajectory (Chiba et al., 2024). Feeding such a readout back as input can yield an autonomous predictor empirically (de Jong et al., 2025; de Jong, 2025), but the result still depends on that learned transformation rather than on the unmodified phase dynamics. Adaptive synchronization instead tracks one prescribed Kuramoto model with a second, adapted instance (Alanazi et al., 2025). This is a tracking result, since the target is itself Kuramoto-generated, not an approximation result for arbitrary vector fields.

Analysis and computation with Kuramoto-related dynamics. Muller et al. (2021) derive an exact complex-valued representation of attractive Kuramoto dynamics, while Budzinski et al. (2023) predict when synchronized and unsynchronized groups coexist in Sakaguchi–Kuramoto networks. These works analyze properties of the Kuramoto model but do not concern approximation of arbitrary target flows. Budzinski et al. (2024) derive an exact input–output expression for a linear complex-valued network and demonstrate logic and memory tasks by decoding its transient synchronization patterns, but require a nonlinear decoder and do not show phase trajectories approximate a prescribed flow. Our theorem guarantees that approximation in the phases of an autonomous Kuramoto network.

3 The Kuramoto Model Is Universal

Consider the Kuramoto model equation 2.1 on $\mathbb{T}^N$, whose intrinsic frequencies ω_i and couplings C_{ij} are allowed to be arbitrary real numbers. In our universality result, an oscillator network is constructed that partitions the model oscillator state as $s=(v_1,\dots,v_n,h_1,\dots,h_{N-n})$, where $v=(v_1,\dots,v_n)$ is the *visible state* and $h=(h_1,\dots,h_{N-n})$ is the *hidden state*. The universality result is as follows:

Theorem 3.1 (Universal approximation of dynamical systems with the Kuramoto model)**.** *Let $F:\mathbb{T}^n\to\mathbb{R}^n$ be continuously differentiable, and let $x(t)$ denote the solution of $\dot{x}=F(x)$ with $x(0)=x_0$. For every horizon $t_1>0$ and tolerance $\zeta>0$, there exist $N\geq n$, parameters $\omega\in\mathbb{R}^N$ and $C\in\mathbb{R}^{N\times N}$ of the Kuramoto model*

$$\dot{s}_i=\omega_i+\sum_{j=1}^{N}C_{ij}\sin(s_j-s_i),\qquad i=1,\dots,N,$$

and a continuous initialization map $\mathcal{I}:\mathbb{T}^n\to\mathbb{T}^{N-n}$ with the following property. Write $s=(v,h)$, where the visible state v consists of the first n oscillators and the hidden state h of the remaining $N-n$. For every $x_0\in\mathbb{T}^n$, the solution initialized by $v(0)=x_0$ and $h(0)=\mathcal{I}(x_0)$ satisfies

$$\max_{t\in[0,t_1]}|v(t)-x(t)|<\zeta.$$

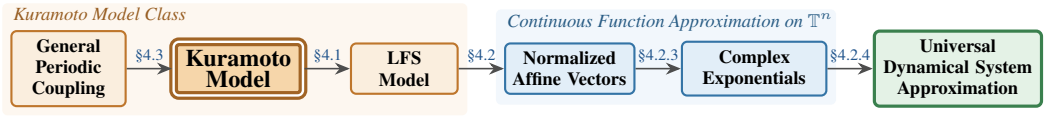

Figure 1: Roadmap of the proof.

This universality result shows that a single finite-dimensional instance of the Kuramoto model approximates the entire finite-time flow of $x(t)$ for any continuously differentiable vector field $F : \mathbb{T}^n \to \mathbb{R}^n$. For a target state $x = (x_1, \ldots, x_n)$, the visible phase v_i tracks the target coordinate x_i, and the theorem compares $v(t)$ itself with the target trajectory. The hidden state h contains auxiliary oscillator phases that shape the velocity of v. Note that no phase lags, external forcing, or learned output map occur in the realizing Kuramoto system.

4 Proof of Universality

The proof has the overall structure and corresponding sections shown in Figure 1. We first introduce a lagged fast–slow (LFS) model and show that each of its lagged couplings can be approximated by ordinary Kuramoto oscillators. We then prove that the LFS model can approximate any continuously differentiable vector field on a torus. The LFS model's fast hidden oscillators converge to normalized affine vector (NAV) features at equilibrium; the visible features assemble linear combinations of NAV features approximating the target field. Finally, a standard trajectory comparison is used to extend vector field error into trajectory error. Combining the LFS approximation result with the lagged coupling approximation proves universality of the original Kuramoto model. Section 4.3 then extends the construction beyond sinusoidal coupling, further generalizing the result. The proof strategy is similar to Funahashi & Nakamura (1993): hidden units support visible-state trajectory approximation, with a vector-field approximation converted into a finite-time trajectory bound.

4.1 The LFS Model

The LFS model is an instantiation of the Sakaguchi–Kuramoto model (equation 2.2) comprised of m fast-evolving hidden oscillators $h_1 \ldots h_a \ldots h_m$, n slow-evolving visible oscillators $v_1 \ldots v_i \ldots v_n$, and fixed oscillators $b_1 \ldots b_a \ldots b_m$. The visible and hidden couplings have lags $\chi_{i,a}$ and $\psi_{a,j}$:

$$\dot{h}_a = \varepsilon^{-1}\Lambda_a \sin(h_a - b_a) + \sum_{j=1}^{n} \varepsilon^{-1} k_{a,j} \sin(h_a - v_j - \psi_{a,j}), \tag{4.1}$$

$$\dot{v}_i = \sum_{a=1}^{m} w_{i,a} \sin(v_i - h_a - \chi_{i,a}). \tag{4.2}$$

The parameters $\Lambda_a, k_{a,j}, w_{i,a}$ are all in $\mathbb{R}$, the hidden oscillator timescale is $\varepsilon > 0$, and the oscillators h_a, b_a, v_i and lags $\chi_{i,a}, \psi_{a,j}$ are all in $\mathbb{T}$. The system evolves at two timescales, with the hidden state evolving ε^{-1} times faster than the visible state. The oscillators in the system are asymmetrically coupled. Each visible oscillator i is coupled to all hidden oscillators a with coupling strengths $w_{i,a}$, and each hidden oscillator a is coupled to all visible oscillators j with strength $\varepsilon^{-1}k_{a,j}$. Finally, each hidden oscillator a is coupled with the fixed oscillator b_a with a coupling strength that exceeds the sum of the hidden-visible coupling strengths, which we call the *dominance condition*:

$$|\Lambda_a| > \sum_{j=1}^{n} |k_{a,j}|, \qquad \mu_a := |\Lambda_a| - \sum_{j=1}^{n} |k_{a,j}| > 0. \tag{4.3}$$

4.1.1 The LFS Model maps to the Kuramoto Model

We map the LFS model to the Kuramoto model (equation 2.1) by replacing the lagged couplings in the LFS model with networks of intermediate oscillators that approximate the lagged coupling behavior, recovering the original Kuramoto dynamics. As with the LFS model, the recovered Kuramoto model has fast hidden and slow visible states.

Lemma 4.1 (Lagged Couplings are Approximable with Oscillators)**.** *Consider a lagged coupling of the form* $\sin(s_i - s_j - \psi_{ij})$ *whose source oscillator has bounded speed* $|\dot{s}_j(t)| \le M_j$ *for all* $t \in [0, T]$.

There exist absolute constants $c_{\mathrm{syn}}, C_{\mathrm{syn}} > 0$ such that, for every lag approximation timescale $\eta > 0$ with $\eta M_j \le c_{\mathrm{syn}}$, a four-oscillator Kuramoto network with three added oscillators reproduces the lagged coupling with error at most $C_{\mathrm{syn}} \eta M_j$ for $t \in [0,T]$, for any initial phases $(s_i(0), s_j(0)) \in \mathbb{T}^2$, with the added oscillators initialized at phases determined by $s_j(0)$.

Proof Sketch. The network consists of the source oscillator s_j and three added oscillators r_1, r_2, r_3:

$$\dot{r}_k = \eta^{-1} \sin(\frac{\psi_{ij}}{3}) + \eta^{-1} \sin(r_{k-1} - r_k), \qquad r_k(0) = s_j(0) + k\frac{\psi_{ij}}{3}, \qquad k = 1,2,3. \tag{4.4}$$

$\eta > 0$ sets the timescale of the added oscillators. For each r_{k-1}, the equilibrium of the added oscillator is when $\dot{r}_k = 0$, that is, when $\sin(r_{k-1} - r_k) = -\sin(\frac{\psi_{ij}}{3})$. The phase $r_k = r_{k-1} + \frac{\psi_{ij}}{3}$ is the attracting equilibrium of equation 4.4 and the convergence speed can be controlled by η. The equilibrium of r_3 tracks $s_j + \psi_{ij}$, and coupling from r_3 reproduces the lagged coupling:

$$-\sin(r_3 - s_i) = -\sin(s_j + \psi_{ij} - s_i) = \sin(s_i - s_j - \psi_{ij}). \tag{4.5}$$

The approximation error comes from the chain's finite response time. For a fixed source-speed bound M_j, it decreases linearly with η; for a fixed η, a faster source is harder to track. The proof in Appendix B.1 gives the valid choices $c_{\mathrm{syn}} = 4.5 \times 10^{-3}$ and $C_{\mathrm{syn}} = 77$.

4.2 Proving the LFS Model is Universal with NAVs

To prove the universality of the LFS model and therefore universality of the Kuramoto model, we demonstrate that each hidden oscillator h settles at an equilibrium phase whose unit complex representation e^{ih} is a *normalized affine vector* (NAV), denoted by $N_a(v)$. The vector field of the visible state can then be rewritten as a complex linear combination of NAVs.

$$\dot{v}_i = \sum_{a=1}^{m} w_{i,a} \sin(v_i - h_a - \chi_{i,a}) = \mathrm{Im}\left(e^{iv_i} \overline{\sum_{a=1}^{m} w_{i,a} e^{i\chi_{i,a}} e^{ih_a}} \right). \tag{4.6}$$

When all hidden oscillators are at their equilibria, $e^{ih_a} = N_a(v)$, yielding the ***reduced field***

$$\widetilde{F}_i(v) := \mathrm{Im}\left(e^{iv_i} \overline{\sum_{a=1}^{m} w_{i,a} e^{i\chi_{i,a}} N_a(v)} \right). \tag{4.7}$$

The NAVs are then shown to be a universal function approximator of functions in the continuous complex-valued function space on the n-torus $C(\mathbb{T}^n, \mathbb{C})$ (Section 4.2.3). A NAV of hidden state h_a is defined over the visible oscillator states v and the hidden oscillator parameters:

$$N_a(v) := u\left(\gamma_a + \sum_j c_{a,j} e^{iv_j} \right) = u(W_a(v)) \tag{4.8}$$

where the complex-valued expressions γ_a, $c_{a,j}$, and $W_a(v)$ are derived from the hidden oscillator parameters (equation 4.1), and u is a normalizing function:

$$\gamma_a := -\Lambda_a e^{ib_a}, \qquad c_{a,j} := -k_{a,j} e^{i\psi_{a,j}}, \qquad W_a(v) := \gamma_a + \sum_{j=1}^{n} c_{a,j} e^{iv_j} \qquad u(z) = z/|z| \tag{4.9}$$

4.2.1 The Steady-State Solution of an LFS Model Hidden Oscillator is a NAV

Lemma 4.2 (Hidden Oscillator Steady-State is a NAV)**.** *For every fixed visible state $v \in \mathbb{T}^n$, the hidden oscillators with dynamics equation 4.1 have a single attracting equilibrium $e^{ih_a^*(v)} = N_a(v)$. The map $v \mapsto N_a(v)$ is smooth on $\mathbb{T}^n$, and perturbations from this equilibrium decay locally at rate at least μ_a/ε, uniformly over v.*

Proof Sketch. Substituting equation 4.9 into the hidden oscillator equation 4.1 gives

$$\varepsilon \dot{h}_a = -\mathrm{Im}\left(e^{ih_a} \overline{W_a(v)} \right) = -|W_a(v)| \sin\big(h_a - \arg W_a(v)\big). \tag{4.10}$$

The hidden oscillator has two equilibria, $h_a = \arg W_a(v)$ and $h_a = \arg(W_a(v)) + \pi$.

The phase $h_a^*(v) = \arg(W_a(v))$ is the attracting equilibrium of equation 4.10, and the other one is repelling. At the attracting equilibrium, $e^{ih_a^*(v)} = W_a(v)/|W_a(v)| = N_a(v)$. The dominance condition equation 4.3 gives $|W_a(v)| \ge \mu_a > 0$ for every v, so $W_a(v)$ never vanishes. Hence its direction $\arg(W_a(v))$ is defined and varies smoothly with v, and the convergence rate μ_a/ε does not degrade to zero. Full proof is presented in Appendix B.2.

4.2.2 MATCHING A VISIBLE VECTOR FIELD OF AN LFS MODEL WITH NAVS

Once the hidden oscillators are at equilibrium, equation 4.7 gives the visible vector field which is the projection of the complex linear combination of NAVs onto the direction tangent to the unit circle at the phase v_i. Producing a target velocity field $F : \mathbb{T}^n \to \mathbb{R}^n$ for the visible states therefore only requires a complex function whose projection onto this tangent direction equals the target.

$$T_i(v) := -ie^{iv_i} F_i(v), \qquad \mathrm{Im}\Big(e^{iv_i}\,\overline{T_i(v)}\Big) = F_i(v). \tag{4.11}$$

Therefore, approximating a target field F reduces to approximating the n functions $T_i \in C(\mathbb{T}^n, \mathbb{C})$ by finite complex linear combinations of NAVs, which Section 4.2.3 shows is always possible.

4.2.3 COMPLEX LINEAR COMBINATIONS OF NAVS ARE UNIVERSAL FUNCTION APPROXIMATORS IN $C(\mathbb{T}^n, \mathbb{C})$

The functions T_i in equation 4.11 are continuous complex-valued functions on the torus. The following proposition shows that finite complex combinations of NAVs approximate every such function, and hence every T_i.

Proposition 4.1 (Density of NAVs). *Finite complex combinations of the NAVs in*

$$\left\{ N(v) = u\left(\gamma + \sum_{j=1}^{n} c_j e^{iv_j}\right) : \gamma \in \mathbb{C},\ c \in \mathbb{C}^n,\ |\gamma| > \sum_j |c_j| \right\}$$

are dense in $C(\mathbb{T}^n, \mathbb{C})$ in the uniform norm. Intuitively, given any $T \in C(\mathbb{T}^n, \mathbb{C})$ and $\delta > 0$, there exist finitely many elements $N_1, \ldots, N_k$ of the above set and coefficients $\alpha_1, \ldots, \alpha_k \in \mathbb{C}$ such that $\left|\sum_{j=1}^{k} \alpha_j N_j(v) - T(v)\right| < \delta$ for all $v \in \mathbb{T}^n$.

Proof Sketch. Since finite combinations of complex exponentials are dense in $C(\mathbb{T}^n, \mathbb{C})$ by the Stone–Weierstrass theorem (Rudin, 1976), it suffices to prove the proposition by showing that every element of the complex exponentials can be approximated uniformly by finite combinations of NAVs. We write $N(v)$ as an equivalent infinite sum of complex exponentials, and show that a weighted sum of NAVs can cancel every exponential except the one wanted. Then, we show that there exists a parametrization of c that guarantees a non-zero coefficient of the isolated exponential term. Since this construction works for every element of the complex exponential basis, combinations of NAVs are dense in $C(\mathbb{T}^n, \mathbb{C})$. Full proof is presented in Appendix B.4.

4.2.4 FROM FUNCTION APPROXIMATION TO TRAJECTORY APPROXIMATION

After we can match the target field uniformly, we would like to convert the field error into a trajectory error. This step is proven and used in previous dynamical system approximation literature (Funahashi & Nakamura, 1993). We present the lemma for completeness:

Lemma 4.3 (Trajectory comparison). *Let $F, \widetilde{F} : D \to \mathbb{R}^n$ be Lipschitz continuous on an open set $D \subseteq \mathbb{R}^n$, let L be a Lipschitz constant of F, and suppose that $|F(x) - \widetilde{F}(x)| \le \delta$ for all $x \in D$. If $x(t)$ and $\tilde{x}(t)$ are solutions of $\dot{x} = F(x)$ and $\dot{\tilde{x}} = \widetilde{F}(\tilde{x})$ on an interval J with $x(t_0) = \tilde{x}(t_0)$, then*

$$|x(t) - \tilde{x}(t)| \le \frac{\delta}{L}\left(e^{L|t-t_0|} - 1\right) \qquad \textit{for all } t \in J.$$

When $L = 0$, the factor $(e^{L|t-t_0|} - 1)/L$ is interpreted as $|t - t_0|$. The same bound holds if $\tilde{x}$ instead solves the forced equation $\dot{\tilde{x}} = F(\tilde{x}) + r(t)$ with $|r(t)| \le \delta$ on J.

The lemma shows that a uniformly bounded field error is amplified over a finite horizon only by $Q_L(t)$, which is defined as follows:

$$Q_L(t) := \frac{e^{Lt} - 1}{L}\ \ (L > 0), \qquad Q_0(t) := t. \tag{4.12}$$

With the lemma, we can now establish the universality of the LFS model:

Proposition 4.2 (Universality of LFS). *Let $F : \mathbb{T}^n \to \mathbb{R}^n$ be C^1, let $t_1 > 0$, and let $\rho > 0$. There exist $m \in \mathbb{N}$, an LFS model equation 4.1–equation 4.2 with m hidden oscillators satisfying the dominance*

Table 2: Oscillator and parameter counts for the LFS model and the Kuramoto model, as functions of the number of hidden (m) and visible (n) states. Number of trainable parameters refers to parameters that can be freely chosen in the NAV parametrization.

Model	# NAVs	Total # Oscillators	# Trainable parameters
LFS	m	$2m+n$	$4mn+2m$
Kuramoto	m	$2m+n+6mn$	$4mn+2m$

condition equation 4.3, and a timescale $\varepsilon > 0$, such that the following holds. For every $x_0 \in \mathbb{T}^n$, the visible trajectory initialized by $v(0) = x_0$ and $e^{ih_a(0)} = N_a(x_0)$ for $a = 1, \ldots, m$ satisfies the following inequality, given $x(t)$ is the solution to $\dot{x} = F(x)$ with $x(0) = x_0$:

$$\sup_{t \in [0, t_1]} |v(t) - x(t)| < \rho.$$

Proof Sketch. The target field F is C^1 on the compact torus and hence has a Lipschitz constant L. Choose the NAVs so that their reduced field $\widetilde{F}$ (equation 4.7) satisfies $\|\widetilde{F} - F\|_\infty \leq \delta$. This gives a fixed parametrized model with m hidden states. For this model, the hidden oscillators track moving NAV equilibria with an error proportional to ε. Therefore, the visible field error is bounded by

$$\left|\dot{v}(t) - F(v(t))\right| \leq \underbrace{\delta}_{\text{density}} + \underbrace{C_h \varepsilon}_{\text{hidden state convergence}}, \tag{4.13}$$

provided ε is below a model-dependent threshold. Here C_h is a uniform sensitivity constant determined by the fixed NAV coefficients; it is independent of ε. Applying Lemma 4.3 gives

$$\sup_{0 \leq t \leq t_1} |v(t) - x(t)| \leq (\delta + C_h \varepsilon) Q_L(t_1). \tag{4.14}$$

We first choose δ, which fixes the finite model and C_h, and only then choose ε small enough. Detailed bounds are given in Appendix B.6.

Proposition 4.2 establishes universality for the LFS model, whose couplings still carry lags. Removing these lags with Lemma 4.1 transfers the result to the Kuramoto model and proves Theorem 3.1.

Proof Sketch of Theorem 3.1. Fix the LFS model chosen by Proposition 4.2 within an error bound. Replacing every lagged coupling by the chain of added oscillators in Lemma 4.1, and every fixed oscillator b_a by a Kuramoto oscillator with zero intrinsic frequency and no incoming coupling, yields a Kuramoto model. The lag approximation and hidden oscillator convergence errors bound the error of the visible vector field. Applying Lemma 4.3 gives

$$\sup_{0 \leq t \leq t_1} |v_{\varepsilon,\eta}(t) - x(t)| \leq \big(\underbrace{\delta}_{\text{density}} + \underbrace{C_h \varepsilon}_{\text{hidden state convergence}} + \underbrace{C_\eta \eta}_{\text{lag approximation}}\big) Q_L(t_1). \tag{4.15}$$

The constant C_η, like C_h, is determined by the fixed LFS model and is independent of ε and η. We choose δ first, then ε, and finally η (with $\eta \leq \varepsilon$ as required by the construction) so that the three terms are below the desired tolerance. Appendix B.7 contains the complete bounds.

4.2.5 Interpretation of the Approximation Error.

The error bound in equation 4.15 has practical implications. First, the number of hidden states affects the approximation power of the Kuramoto system. For reference, Table 2 shows how the oscillator and parameter counts scale with the number of hidden and visible states. For a system with m hidden oscillators, the reduced field is a finite linear combination of m NAVs. Allowing more NAV terms introduces more trainable parameters and gives this finite sum more freedom to approximate the target field, and Proposition 4.1 guarantees that some finite m attains any prescribed error $\delta > 0$. After training the NAV expression to match the reduced field within some error bound, the C_h, C_η values are analytically derived from learned parameters. The hidden state convergence error $C_h \varepsilon$ and lag approximation error $C_\eta \eta$ thus decrease linearly with respect to timescales ε and η, respectively. This reduction increases the timescale separation in the fast–slow dynamics, yielding a stiffer system.

4.3 Extension of Universality Result to general periodic coupling functions

We show any smooth periodic coupling with a nonzero first harmonic can replace the sine coupling.

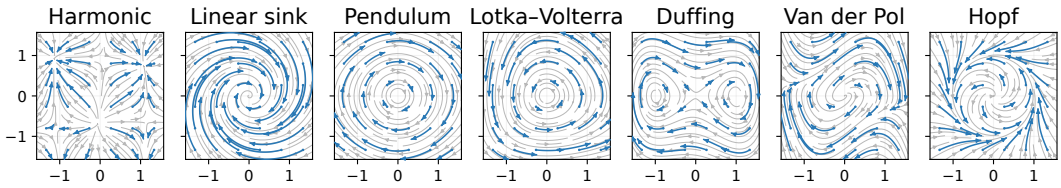

Figure 2: Fitting results with the Kuramoto model. Columns are the seven target dynamical systems. The ■ lines are the target vector field, ■ lines are Kuramoto trajectories (130 oscillators, $\varepsilon = 0.01$).

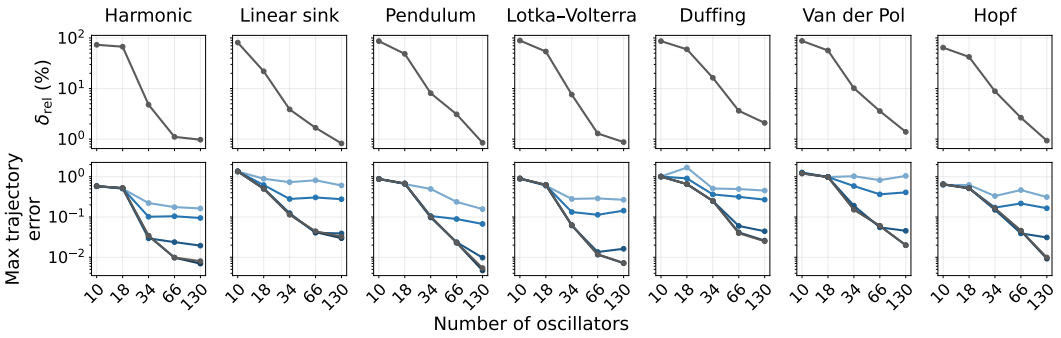

Figure 3: Dynamical system fitting ablation with oscillator count and timescale separation ε with Kuramoto. Columns are the seven target systems; top row is the fitted reduced field error, and bottom is the trajectory error. Lines: ■ reduced field fit; ■, ■, ■, ■ use timescale $\varepsilon = 1, 0.1, 0.01, 0.001$.

Theorem 4.1 (Universality for general periodic coupling)**.** *Let $\Gamma \in C^2(\mathbb{T}, \mathbb{R})$ be a 2π-periodic coupling function whose first Fourier coefficient is nonzero,*

$$\widehat{\Gamma}_1 := \frac{1}{2\pi} \int_0^{2\pi} \Gamma(x) e^{-ix}\, dx \neq 0. \tag{4.16}$$

Then the conclusion of Theorem 3.1 holds with the Kuramoto model replaced by the following equation, where $N \in \mathbb{N}$, $\omega \in \mathbb{R}^N$, and $K \in \mathbb{R}^{N \times N}$:

$$\dot{q}_i = \omega_i + \sum_{j=1}^{N} K_{ij} \Gamma(q_j - q_i). \qquad i = 1, \ldots, N, \tag{4.17}$$

Proof Sketch. Start with the LFS model selected in the proof of Proposition 4.2 and replace each of its sine couplings in two steps. First, a sine is exactly a continuously weighted average of shifted copies of Γ: the sinusoidal weights cancel all Fourier coefficients except those at the first frequency, and $\widehat{\Gamma}_1 \neq 0$ allows the surviving harmonic to be rescaled to the desired sine. Replacing this integral by a finite grid of shifts produces a finite network whose approximation improves as the grid is refined.

Second, each shifted copy is still a lagged coupling. The condition $\widehat{\Gamma}_1 \neq 0$ implies that Γ is nonconstant, so there is an interval on which Γ' has a fixed nonzero sign. Any desired phase lag can be split into finitely many increments within this interval. A chain of added oscillators realizes these increments as attracting phase differences, as in our lagged coupling oscillator approximation. Decreasing the lag approximation timescale causes its final oscillator to track the desired shifted phase more closely.

Taking the shift grid sufficiently fine and the lag approximation and hidden timescales sufficiently small makes both replacements uniformly accurate. The resulting network uses only unlagged copies of Γ, and Lemma 4.3 transfers the visible-field approximation to the required trajectory approximation. The complete construction and bounds are in Appendix B.8.

5 Numerical Experiments

We demonstrate the two approximation steps behind Theorem 3.1: fitting a target vector field with the reduced field comprised of NAVs, and mapping to a Kuramoto network. We also apply the

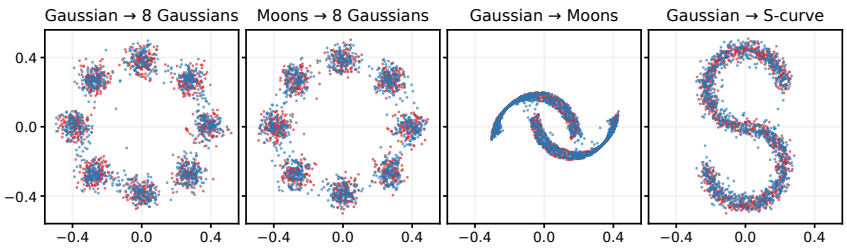

Figure 4: Flow-matching endpoints from Kuramoto (■) and target (■) for four source–target pairs.

construction to a velocity field learned from data. Experimental setup and detailed numerical results are in Appendix C.

Dynamical System Fitting. We fit seven two-dimensional systems spanning sinks, oscillations, nonlinear cycles, and multiple equilibria. Figure 2 overlays trajectories of 130-oscillator Kuramoto networks (ε =0.01) on the target vector fields. Observe the fitted trajectories preserve the qualitative geometry of each system. Figure 3 then varies the number of oscillators and the hidden timescale ε. The top row reports δ_{rel}, the relative error of the reduced field against the target field, with larger networks having better fit quality. The bottom row reports the maximum difference between the target trajectories and the fitted trajectories while the target lies in the fitted box, for both the Kuramoto network and the reduced field fit. Faster hidden dynamics (smaller ε) lower the trajectory error until it becomes indistinguishable from that of the reduced-field fit.

Flow Matching. Recent generative modeling work has used physics-constrained dynamical systems, including Kuramoto networks, to generate samples with endpoint-based objectives such as distribution matching, attractor convergence, and reconstruction (Unconventional AI, 2026; Gagliani et al., 2026). These training methods are much more computationally intensive than flow matching methods. The success of diffusion and flow matching suggests a promising next step: supervising the dynamics throughout generation, rather than only the endpoint distribution (Lipman et al., 2023). Theorem 3.1 shows that Kuramoto networks have the expressivity to approximate such learned flows.

We test this by fitting an MLP flow-matching teacher on four source–target distribution pairs. Time is a third visible phase, and a cascade of 16 Kuramoto networks fits the teacher's vector field, one network per subinterval of $[0, 1]$. Figure 4 shows that the generated samples recover the target clusters and curved distributions across all four pairs. Setup and quantitative results are in Appendix C.2.

6 Conclusion, Limitations, and Future Work

This paper establishes a universal dynamical system approximation theorem using the Kuramoto model. We prove that an autonomous Kuramoto network can approximate the flow of any continuously differentiable vector field on a torus over a finite time interval, uniformly over initial conditions, without a learned output map or external forcing, and we extend the results to smooth periodic coupling functions with a nonzero first Fourier harmonic. We validate this approach by fitting dynamical systems and running a flow matching experiment.

The theorem assumes a finite horizon and a smooth bounded or periodic target field. This is a natural setting for analog systems: the hardware runs for finite time with bounded signal amplitudes. We also note the presented universality theorems are asymptotic existence results. They give neither sample-complexity bounds nor bounds on the number of oscillators needed for a given accuracy, both important for assessing practical expressivity. We have also not addressed optimization or trainability of the oscillator parameters. For example, in the flow-matching experiment, a single large network was difficult to train on the full velocity field, motivating the 16 cascading networks. As with single-hidden-layer MLPs, universality does not guarantee trainability. Developing efficient training methods and architectures for oscillator networks is a promising direction for future work.

AI USE STATEMENT

Generative AI tools assisted with literature organization, proof review, software development, and manuscript editing. The authors checked the mathematical arguments, citations, numerical procedures, generated code, and final prose, and take responsibility for all claims and artifacts in this work.

REPRODUCIBILITY STATEMENT

Theorems 3.1 and 4.1 state all of their hypotheses, and Appendix B proves them in full, including every intermediate lemma and proposition. The proofs give an explicit formula for every constant in the error bounds, such as equation B.33, so each bound can be evaluated for a given model. Appendix C specifies every numerical experiment: the target systems and their coordinate embeddings, the fitted regions, the training data, the optimizers, learning rates, step counts, and random seed, the hidden widths and the rule used to select them, the timescales ε and η, the ODE solvers and their tolerances, and the evaluation metrics. The experiments are implemented in JAX with the Diffrax ODE solvers. Source code is not included with this submission; we will release the code and the fitted models with the final version of the paper.

ETHICS STATEMENT

This work is theoretical and uses only synthetic dynamical systems; it involves no human subjects, personal data, or deployed decision-making system.

REFERENCES

Juan A. Acebrón, Luis L. Bonilla, Conrad J. Pérez Vicente, Félix Ritort, and Renato Spigler. The Kuramoto model: A simple paradigm for synchronization phenomena. *Reviews of Modern Physics*, 77(1):137–185, 2005.

Faizah J. Alanazi, Markus Mueller, and Stuart Townley. Learning and phase tracking by frequency and weight adaptation for coupled networks of Kuramoto oscillators. *SIAM Journal on Applied Dynamical Systems*, 24(1):68–93, 2025. doi: 10.1137/23M1590639.

Pierre Baldi and Kurt Hornik. Universal approximation and learning of trajectories using oscillators. In *Advances in Neural Information Processing Systems*, volume 8, pp. 451–457, 1995.

Manuel Beiran, Alexis Dubreuil, Adrian Valente, Francesca Mastrogiuseppe, and Srdjan Ostojic. Shaping dynamics with multiple populations in low-rank recurrent networks. *Neural Computation*, 33(6):1572–1615, 2021. doi: 10.1162/neco_a_01381.

Roberto C. Budzinski, James W. C. Graham, Ján Mináč, and Lyle E. Muller. Theory of transient chimeras in finite Sakaguchi–Kuramoto networks. *arXiv preprint arXiv:2311.01382*, 2023. URL https://arxiv.org/abs/2311.01382.

Roberto C. Budzinski, Alexandra N. Busch, Samuel Mestern, Erwan Martin, Luisa H. B. Liboni, Federico W. Pasini, Ján Mináč, Todd Coleman, Wataru Inoue, and Lyle E. Muller. An exact mathematical description of computation with transient spatiotemporal dynamics in a complex-valued neural network. *Communications Physics*, 7:239, 2024. doi: 10.1038/s42005-024-01728-0.

Andrea Ceni, Andrea Cossu, Maximilian W. Stölzle, Jingyue Liu, Cosimo Della Santina, Davide Bacciu, and Claudio Gallicchio. Random oscillators network for time series processing. In *Proceedings of the 27th International Conference on Artificial Intelligence and Statistics*, volume 238 of *Proceedings of Machine Learning Research*, pp. 4807–4815, 2024. URL https://proceedings.mlr.press/v238/ceni24a.html.

Hayato Chiba, Koichi Taniguchi, and Takuma Sumi. Reservoir computing with the Kuramoto model. *arXiv preprint arXiv:2407.16172*, 2024.

Thomas Geert de Jong. Designing learning in high dimensional oscillator networks with low dimensional read-out. *arXiv preprint arXiv:2509.00848*, 2025.

Thomas Geert de Jong, Hirofumi Notsu, and Kohei Nakajima. Harnessing omnipresent oscillator networks as computational resource. *arXiv preprint arXiv:2502.04818*, 2025.

Ken-ichi Funahashi and Yuichi Nakamura. Approximation of dynamical systems by continuous time recurrent neural networks. *Neural Networks*, 6(6):801–806, 1993. doi: 10.1016/S0893-6080(05)80125-X.

Stefano Gagliani, Feliciano Giuseppe Pacifico, Lorenzo Chicchi, Duccio Fanelli, Diego Febbe, Lorenzo Buffoni, and Raffaele Marino. Train stochastic non linear coupled ODEs to classify and generate. *Machine Learning: Science and Technology*, 7(2):025020, 2026. doi: 10.1088/2632-2153/ae493a.

Lyudmila Grigoryeva and Juan-Pablo Ortega. Echo state networks are universal. *Neural Networks*, 108:495–508, 2018. doi: 10.1016/j.neunet.2018.08.025.

Masahiro Kimura and Ryohei Nakano. Learning dynamical systems by recurrent neural networks from orbits. *Neural Networks*, 11(9):1589–1599, 1998. doi: 10.1016/S0893-6080(98)00098-7.

Yoshiki Kuramoto. Self-entrainment of a population of coupled non-linear oscillators. In *International Symposium on Mathematical Problems in Theoretical Physics*, pp. 420–422. Springer, 1975.

Samuel Lanthaler, T. Konstantin Rusch, and Siddhartha Mishra. Neural oscillators are universal. In *Advances in Neural Information Processing Systems*, volume 36, 2023.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=PqvMRDCJT9t.

Danijela Marković, Alice Mizrahi, Damien Querlioz, and Julie Grollier. Physics for neuromorphic computing. *Nature Reviews Physics*, 2(9):499–510, 2020.

Takeru Miyato, Sindy Löwe, Andreas Geiger, and Max Welling. Artificial Kuramoto oscillatory neurons. In *International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=nwDRD4AMoN.

Lyle Muller, Ján Mináč, and Tung T. Nguyen. Algebraic approach to the Kuramoto model. *Physical Review E*, 104:L022201, 2021. doi: 10.1103/PhysRevE.104.L022201.

Tuan Nguyen, Hirotada Honda, Takashi Sano, Vinh Nguyen, Shugo Nakamura, and Tan Minh Nguyen. From coupled oscillators to graph neural networks: Reducing over-smoothing via a Kuramoto model-based approach. In *Proceedings of the 27th International Conference on Artificial Intelligence and Statistics*, volume 238 of *Proceedings of Machine Learning Research*, pp. 2710–2718, 2024. URL https://proceedings.mlr.press/v238/nguyen24c.html.

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. URL https://arxiv.org/abs/2605.10613.

Walter Rudin. *Principles of Mathematical Analysis*. McGraw-Hill, 3rd edition, 1976.

T. Konstantin Rusch and Siddhartha Mishra. Coupled oscillatory recurrent neural network (coRNN): An accurate and (gradient) stable architecture for learning long time dependencies. In *International Conference on Learning Representations*, 2021a. URL https://openreview.net/forum?id=F3s69XzWOia.

T. Konstantin Rusch and Siddhartha Mishra. UnICORNN: A recurrent model for learning very long time dependencies. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 9168–9178, 2021b. URL https://proceedings.mlr.press/v139/rusch21a.html.

T. Konstantin Rusch, Ben Chamberlain, James Rowbottom, Siddhartha Mishra, and Michael Bronstein. Graph-coupled oscillator networks. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 18888–18909, 2022. URL https://proceedings.mlr.press/v162/rusch22a.html.

Abel Sagodi and Il Memming Park. Universal approximation theorems for dynamical systems with infinite-time horizon guarantees. *arXiv preprint arXiv:2602.08640*, 2026.

Hidetsugu Sakaguchi and Yoshiki Kuramoto. A soluble active rotator model showing phase transitions via mutual entrainment. *Progress of Theoretical Physics*, 76(3):576–581, 1986.

Alexander Tong, Kilian Fatras, Nikolay Malkin, Guillaume Huguet, Yanlei Zhang, Jarrid Rector-Brooks, Guy Wolf, and Yoshua Bengio. Improving and generalizing flow-based generative models with minibatch optimal transport, 2024. URL https://arxiv.org/abs/2302.00482.

Unconventional AI. Introducing Un-0: Generating images with coupled oscillators. https://unconv.ai/blog/introducing-un-0-generating-images-with-coupled-oscillators/, 2026.

Yu-Neng Wang and Sara Achour. Generative models on analog hardware with dynamics. *arXiv preprint arXiv:2606.27294*, 2026.

APPENDIX

A NOTATION

Symbol	Meaning
$\mathbb{T}=\mathbb{R}/2\pi\mathbb{Z}$	circle; $\mathbb{T}^n$ is the visible state space
$n, m, N=n+m$	visible, hidden, and total oscillator counts
$s=(v,h)$	full state; v represents target coordinates and h is auxiliary
F, L, t_1, ζ	target field, Lipschitz constant, horizon, and tolerance
$u(W)=W/\lvert W\rvert$	complex normalization onto the unit circle
ε	hidden-to-visible timescale ratio
$\gamma_a, c_{a,j}$	complex hidden bias and coupling weights
$W_a(v)$	complex drive to hidden oscillator a
$h_a^*(v), N_a(v)$	attracting hidden phase and its NAV
$w_{i,a}, \Psi_i, T_i$	hidden-to-visible weight, combination of NAVs, and transformed target
$\mu, \nu \in \mathbb{N}^n$	multi-indices, with $\lvert\mu\rvert_1=\sum_j \mu_j$
$\langle k, v\rangle$	standard pairing on Fourier indices and phases
$e^{i\langle k,v\rangle}$	complex exponential indexed by $k\in\mathbb{Z}^n$
$\widehat{\Gamma}_k$	kth Fourier coefficient of the periodic coupling Γ

All trajectory comparisons on $\mathbb{T}^n$ use compatible lifts to $\mathbb{R}^n$. The resulting Euclidean bounds dominate the flat torus distance.

B PROOF FOR KURAMOTO MODEL UNIVERSALITY

The argument has three stages: realize and approximate the normalized affine vectors (NAVs), use them to approximate LFS trajectories, and replace the LFS couplings by unlagged oscillator networks. The last stage also gives the extension to general periodic couplings. Each principal result is restated beside its proof, after the supporting estimates have been established. All vector norms are Euclidean norms, all function norms are uniform norms, and trajectory differences on a torus are evaluated using compatible lifts to Euclidean space.

B.1 PROOF OF LEMMA 4.1

Lemma 4.1 (Lagged Couplings are Approximable with Oscillators)**.** *Consider a lagged coupling of the form* $\sin(s_i - s_j - \psi_{ij})$ *whose source oscillator has bounded speed* $|\dot{s}_j(t)| \le M_j$ *for all* $t\in[0,T]$. *There exist absolute constants* $c_{\mathrm{syn}}, C_{\mathrm{syn}} > 0$ *such that, for every lag approximation timescale* $\eta > 0$ *with* $\eta M_j \le c_{\mathrm{syn}}$*, a Kuramoto network with three added oscillators reproduces the lagged coupling with error at most* $C_{\mathrm{syn}}\eta M_j$ *for* $t\in[0,T]$*, for any initial phases* $(s_i(0), s_j(0)) \in \mathbb{T}^2$*, with the added oscillators initialized at phases determined by* $s_j(0)$.

We first control one phase-shift chain; Lemma B.3 later applies the same bound to all couplings of a fixed LFS model.

Proof of Lemma 4.1. To realize the desired lag, split it into three increments small enough that each tracking equation is uniformly contracting. Choose the representative $\psi_{ij} \in [-\pi, \pi]$, set $\alpha := \psi_{ij}/3$, and let $r_0 := s_j$. Use the three added oscillators in equation 4.4. They have the Kuramoto form equation 2.1, because each added oscillator has intrinsic frequency $\eta^{-1} \sin\alpha$ and one directed incoming coupling of strength η^{-1}. Initialize $r_k(0) = s_j(0) + k\alpha$.

Whenever the source oscillator belongs to a fixed finite Kuramoto model or LFS model, the speed bound M_j holds uniformly over initial phases, because its right-hand side is a finite sum of bounded terms. We next choose a phase-error interval on which that contraction has a lower bound independent of the lag. Set

$$q_0 := \frac{\pi}{12}, \qquad c_0 := \cos\frac{5\pi}{12}, \qquad \kappa_0 := q_0 c_0^3 \approx 4.5 \times 10^{-3}, \qquad S_3 := c_0^{-1} + c_0^{-2} + c_0^{-3} \approx 76.5. \tag{B.1}$$

These constants do not depend on ψ_{ij}, M_j, η, or the initial phases. For $q_k := r_k - r_{k-1} - \alpha$, the mean-value theorem gives

$$|\sin(\alpha + q_k) - \sin\alpha| \geq c_0 |q_k| \qquad \text{whenever } |q_k| \leq q_0,$$

because $|\alpha| \leq \pi/3$, so every point between α and $\alpha + q_k$ lies in $[-5\pi/12, 5\pi/12]$, where the cosine is at least c_0.

To propagate the source-speed bound through the chain, let $M_0 := M_j$ and $M_k := M_{k-1}/c_0$. If

$$\eta M_j \leq \kappa_0, \quad \text{that is,} \quad \eta M_j c_0^{-3} \leq q_0, \tag{B.2}$$

an induction over k shows

$$|q_k(t)| \leq \frac{\eta M_{k-1}}{c_0} = \eta M_j c_0^{-k}, \qquad |\dot{r}_k(t)| \leq M_k, \qquad t \in [0, T], \quad k = 1, 2, 3. \tag{B.3}$$

Indeed,

$$\dot{q}_k = \eta^{-1}\big[\sin\alpha - \sin(\alpha + q_k)\big] - \dot{r}_{k-1}.$$

The initialization gives $q_k(0) = 0$. The smallness condition equation B.2 places both endpoints of the interval in equation B.3 within $[-q_0, q_0]$. At the positive endpoint the bracket is at most $-\eta M_{k-1}$, and the induction hypothesis $|\dot{r}_{k-1}| \leq M_{k-1}$ makes the right-hand side nonpositive; at the negative endpoint it is nonnegative. The interval is therefore invariant on $[0, T]$, because the right-hand side is Lipschitz in q_k and points into the interval at both endpoints. The equation for r_k together with the Lipschitz constant one of sine gives $|\dot{r}_k| \leq |q_k|/\eta \leq M_k$, completing the induction.

Finally, the output phase error is the sum of the three tracking errors. Since $r_3 - s_j - \psi_{ij} = q_1 + q_2 + q_3$,

$$\sup_{0 \leq t \leq T} |r_3(t) - s_j(t) - \psi_{ij}| \leq S_3 M_j \eta. \tag{B.4}$$

Therefore the lagged coupling in Lemma 4.1 is approximated by the Kuramoto coupling $-\sin(r_3 - s_i)$ with error

$$\sup_{0 \leq t \leq T} |\sin(s_i - s_j - \psi_{ij}) + \sin(r_3 - s_i)| \leq S_3 M_j \eta. \tag{B.5}$$

This proves the lemma, because $\kappa_0 \geq 4.5 \times 10^{-3}$ and $S_3 \leq 77$ in equation B.1. The estimate is independent of the initial phases because M_j is, and independent of ψ_{ij} because q_0 and c_0 are. □

B.2 Proof of Lemma 4.2

Lemma 4.2 (Hidden Oscillator Steady-State is a NAV)**.** *For every fixed visible state $v \in \mathbb{T}^n$, the hidden oscillators with dynamics equation 4.1 have a single attracting equilibrium $e^{ih_a^*(v)} = N_a(v)$. The map $v \mapsto N_a(v)$ is smooth on $\mathbb{T}^n$, and perturbations from this equilibrium decay locally at rate at least μ_a/ε, uniformly over v.*

Proof of Lemma 4.2. We first show that its complex drive never vanishes; this makes normalization smooth and gives a uniform stability rate. Fix a hidden oscillator a. The dominance condition equation 4.3 and the reverse triangle inequality give

$$|W_a(v)| \geq |\gamma_a| - \sum_{j=1}^{n} |c_{a,j}| = \mu_a > 0 \qquad \text{for every } v \in \mathbb{T}^n. \tag{B.6}$$

The normalizing function u is smooth on $\mathbb{C}\setminus\{0\}$, so the NAV $N_a = u\circ W_a$ in equation 4.8 is smooth on $\mathbb{T}^n$.

To identify the attracting equilibrium, freeze v and write $W_a(v) = |W_a(v)|e^{i\vartheta_a(v)}$, where only the value of $e^{i\vartheta_a(v)}$ is needed globally. The hidden equation in the form equation 4.10 becomes

$$\varepsilon\dot{h}_a = -|W_a(v)|\sin\big(h_a - \vartheta_a(v)\big).$$

There are two equilibria modulo 2π. At $h_a^*(v) = \vartheta_a(v)$,

$$e^{ih_a^*(v)} = \frac{W_a(v)}{|W_a(v)|} = N_a(v), \qquad \frac{\partial\dot{h}_a}{\partial h_a} = -\frac{|W_a(v)|}{\varepsilon} \le -\frac{\mu_a}{\varepsilon}.$$

The other equilibrium, $h_a^*(v)+\pi$, has derivative $|W_a(v)|/\varepsilon > 0$ and is repelling. Thus $N_a(v)$ is the unique attracting equilibrium on the circle, and perturbations in its basin decay locally at rate at least μ_a/ε, uniformly in v. □

B.3 The visible vector field at the hidden equilibria

Lemma 4.2 places each hidden oscillator at $e^{ih_a} = N_a(v)$. Substituting these equilibria into the visible vector field equation 4.7 defines

$$\Psi_i(v) := \sum_{a=1}^{m} w_{i,a}e^{i\chi_{i,a}}N_a(v), \qquad \widetilde{F}_i(v) := \mathrm{Im}\Big(e^{iv_i}\overline{\Psi_i(v)}\Big). \tag{B.7}$$

The field $\widetilde{F}$ is the reduced field of the LFS model: the visible vector field when every hidden oscillator is exactly at its equilibrium. For the target functions T_i in equation 4.11,

$$|\widetilde{F}_i(v) - F_i(v)| = \Big|\mathrm{Im}\Big(e^{iv_i}\overline{\Psi_i(v)-T_i(v)}\Big)\Big| \le |\Psi_i(v)-T_i(v)|. \tag{B.8}$$

The Euclidean norm of a vector in $\mathbb{R}^n$ is at most $\sqrt{n}$ times its largest component, so the componentwise approximation condition

$$\max_{1\le i\le n}\|\Psi_i - T_i\|_\infty \le \frac{\delta}{\sqrt{n}} \tag{B.9}$$

implies the field bound

$$\|\widetilde{F} - F\|_\infty \le \delta. \tag{B.10}$$

Approximating the target field F within δ therefore reduces to approximating each T_i within $\delta/\sqrt{n}$ by a finite complex linear combination of NAVs, which Proposition 4.1 provides.

B.4 Proof of Proposition 4.1

To prove Proposition 4.1, we first establish the Fourier projection and finite-grid bounds needed to isolate each character.

Let V be the uniform closure of the finite complex span in Proposition 4.1. The identities $u(\alpha z)=u(z)$ for $\alpha>0$ and $u(e^{i\vartheta}z) = e^{i\vartheta}u(z)$ reduce the admissible family to

$$f_{t,\theta}(v) := u\left(1 + t\sum_{j=1}^{n}\tau_j e^{i(\theta_j+v_j)}\right), \qquad 0<t<1, \quad \tau_j>0, \quad \sum_{j=1}^{n}\tau_j = 1, \quad \theta\in\mathbb{T}^n. \tag{B.11}$$

Each $f_{t,\theta}$ is an admissible NAV because the magnitude of its constant term is 1, while the sum of the magnitudes of its nonconstant coefficients is $t<1$.

For a multi-index $\mu\in\mathbb{N}^n$, write

$$|\mu|_1 := \sum_{j=1}^{n}\mu_j, \qquad \tau^\mu := \prod_{j=1}^{n}\tau_j^{\mu_j}, \qquad \begin{bmatrix} p \\ \mu \end{bmatrix} := \frac{p!}{\mu_1!\cdots\mu_n!} \quad \text{when } |\mu|_1 = p.$$

The following lemma gives the projection identity and the error from replacing its integral by a finite sum.

Lemma B.1 (NAV expansion and projection error)**.** *Let $f_{t,\theta}$ be the NAV in equation B.11, with τ fixed. For every $r < 1$,*

$$f_{t,\theta}(v) = \sum_{\mu,\nu\in\mathbb{N}^n} A_{\mu\nu} t^{|\mu|_1+|\nu|_1} \tau^{\mu+\nu} e^{i\langle \mu-\nu,\theta\rangle} e^{i\langle \mu-\nu,v\rangle}, \tag{B.12}$$

where

$$A_{\mu\nu} := \binom{1/2}{|\mu|_1} \binom{-1/2}{|\nu|_1} \begin{bmatrix} |\mu|_1 \\ \mu \end{bmatrix} \begin{bmatrix} |\nu|_1 \\ \nu \end{bmatrix} \neq 0.$$

The series converges absolutely and uniformly for $0 \le t \le r$ and $(\theta, v) \in \mathbb{T}^{2n}$.

For $k \in \mathbb{Z}^n$, let k_+ and k_- be its componentwise positive and negative parts, and define

$$h_k(t,\tau) := \sum_{\omega\in\mathbb{N}^n} A_{k_++\omega,k_-+\omega} \tau^{k_++k_-+2\omega} t^{|k|_1+2|\omega|_1}, \tag{B.13}$$

$$R_k(t,\tau) := \sum_{\omega\neq 0} |A_{k_++\omega,k_-+\omega}| \tau^{k_++k_-+2\omega} t^{2|\omega|_1}. \tag{B.14}$$

Then, for $0 < t < 1$,

$$\frac{1}{(2\pi)^n} \int_{\mathbb{T}^n} f_{t,\theta}(v) e^{-i\langle k,\theta\rangle}\, d\theta = h_k(t,\tau) e^{i\langle k,v\rangle}. \tag{B.15}$$

If

$$R_k(t,\tau) \le \frac{1}{2} |A_{k_+,k_-}| \tau^{k_++k_-},$$

then

$$|h_k(t,\tau)| \ge \frac{1}{2} |A_{k_+,k_-}| \tau^{k_++k_-} t^{|k|_1} > 0. \tag{B.16}$$

Such $t > 0$ exist because $R_k(t,\tau) \to 0$ as $t \downarrow 0$.

Let $\mathcal{Q}_M$ be the normalized midpoint rule in θ on the uniform M^n-point grid of $\mathbb{T}^n$. Then

$$\left\| \frac{\mathcal{Q}_M\left[f_{t,\theta}(\,\cdot\,) e^{-i\langle k,\theta\rangle}\right]}{h_k(t,\tau)} - e^{i\langle k,\,\cdot\,\rangle} \right\|_\infty \le \frac{\pi\sqrt{n}}{M|h_k(t,\tau)|} \left(|k|_2 + \frac{t|\tau|_2}{1-t} \right). \tag{B.17}$$

Proof. To isolate one Fourier character, expand the normalized drive into powers of its affine perturbation and average over the phase parameter. Set

$$\sigma(v) := t \sum_{j=1}^n \tau_j e^{i(\theta_j+v_j)}.$$

Since $|\sigma(v)| \le t < 1$, the principal square-root branches give

$$f_{t,\theta}(v) = (1+\sigma(v))^{1/2} (1+\overline{\sigma(v)})^{-1/2}.$$

Expanding both factors by the generalized binomial series and then applying the multinomial theorem gives equation B.12.

For $0 \le t \le r < 1$, the sum of the absolute values in equation B.12 is at most

$$\left(\sum_{p\ge 0} \left| \binom{1/2}{p} \right| r^p \right) \left(\sum_{q\ge 0} \left| \binom{-1/2}{q} \right| r^q \right), \tag{B.18}$$

because

$$\sum_{|\mu|_1=p} \begin{bmatrix} p \\ \mu \end{bmatrix} \tau^\mu = \left(\sum_{j=1}^n \tau_j \right)^p = 1.$$

Both scalar series in equation B.18 converge by the ratio test. The Weierstrass M-test proves the asserted absolute and uniform convergence.

Uniform convergence permits termwise integration against $e^{-i\langle k,\theta\rangle}$, which retains precisely the terms satisfying $\mu - \nu = k$. Every such pair has the unique form

$$\mu = k_+ + \omega, \qquad \nu = k_- + \omega, \qquad \omega \in \mathbb{N}^n,$$

which proves equation B.13 and the identity equation B.15. Factoring out $t^{|k|_1}$, applying the reverse triangle inequality, and using equation B.14 give equation B.16. Absolute convergence implies $R_k(t,\tau) \to 0$, so the lower-bound condition holds for all sufficiently small positive t.

To turn the projection integral into a finite NAV combination, we bound the midpoint-rule error uniformly in v. The normalizing function $u(z) = z/|z|$ has real derivative with operator norm at most $1/|z|$. The argument of u in equation B.11 has magnitude at least $1-t$, and therefore

$$\|\nabla_\theta f_{t,\theta}(v)\|_2 \le \frac{t|\tau|_2}{1-t}.$$

Thus the integrand in equation B.15 is Lipschitz in θ, uniformly in v, with constant at most

$$|k|_2 + \frac{t|\tau|_2}{1-t}.$$

Every point of a midpoint-grid cell is within $\pi\sqrt{n}/M$ of its midpoint. The normalized integral and normalized midpoint sum consequently differ by at most the numerator on the right-hand side of equation B.17. Division by $|h_k(t,\tau)|$ proves the claim. □

Proposition 4.1 (Density of NAVs)**.** *Finite complex combinations of the NAVs in*

$$\left\{ N(v) = u\left(\gamma + \sum_{j=1}^{n} c_j e^{iv_j}\right) : \gamma \in \mathbb{C},\ c \in \mathbb{C}^n,\ |\gamma| > \sum_j |c_j| \right\}$$

are dense in $C(\mathbb{T}^n, \mathbb{C})$ in the uniform norm.

In other words, given any $T \in C(\mathbb{T}^n, \mathbb{C})$ and $\delta > 0$, there exist finitely many elements $N_1, \dots, N_k$ of the above set and coefficients $\alpha_1, \dots, \alpha_k \in \mathbb{C}$ such that $\left|\sum_{j=1}^{k} \alpha_j N_j(v) - T(v)\right| < \delta$ for all $v \in \mathbb{T}^n$.

Proof of Proposition 4.1. It suffices to place every Fourier character in V, since their finite span is dense. Fix $k \in \mathbb{Z}^n$. By the projection identity equation B.15 and the lower bound equation B.16 of Lemma B.1, the normalized integral of $f_{t,\theta} e^{-i\langle k,\theta\rangle}$ over θ equals $h_k(t,\tau) e^{i\langle k,v\rangle}$, where $h_k(t,\tau) \neq 0$ for all sufficiently small $t > 0$. The integrand is continuous in θ with values in $C(\mathbb{T}^n, \mathbb{C})$, so the integral is the uniform limit of finite Riemann sums of admissible NAVs, at the rate equation B.17. The integral therefore belongs to V, and division by the nonzero scalar $h_k(t,\tau)$ shows that the complex exponential $e^{i\langle k,v\rangle}$ belongs to V.

Every finite combination of complex exponentials therefore belongs to V. These finite combinations are dense in $C(\mathbb{T}^n, \mathbb{C})$ by the complex Stone–Weierstrass theorem (Rudin, 1976). Since V is closed, $V = C(\mathbb{T}^n, \mathbb{C})$, which proves the proposition.

The construction also gives an error budget for a finite NAV approximation. Let $p(v) = \sum_{k \in K} \widehat{p}_k e^{i\langle k,v\rangle}$ be a finite combination of complex exponentials approximating T, and approximate the integral in equation B.15 by the midpoint grid with M_k points in each parameter coordinate. The finite combination A_T obtained by scaling those Riemann sums satisfies

$$\|T - A_T\|_\infty \le \underbrace{\|T-p\|_\infty}_{E_{\mathrm{poly}}} + \sum_{k \in K} |\widehat{p}_k| \underbrace{\frac{\pi\sqrt{n}}{M_k |h_k(t_k,\tau)|} \left(|k|_2 + \frac{t_k|\tau|_2}{1-t_k}\right)}_{E_{\mathrm{quad},k}}. \tag{B.19}$$

Here each $t_k > 0$ is chosen so that $h_k(t_k,\tau) \neq 0$, and the lower bound equation B.16 makes $|h_k(t_k,\tau)|$ computable. The error of the density step is therefore the sum of the error of p and the finite-grid errors displayed in equation B.19. □

B.5 Proof of Lemma 4.3

Lemma 4.3 (Trajectory comparison)**.** *Let $F, \widetilde{F} : D \to \mathbb{R}^n$ be Lipschitz continuous on an open set $D \subseteq \mathbb{R}^n$, let L be a Lipschitz constant of F, and suppose that $|F(x) - \widetilde{F}(x)| \le \delta$ for all $x \in D$. If $x(t)$ and $\tilde{x}(t)$ are solutions of $\dot{x} = F(x)$ and $\dot{\tilde{x}} = \widetilde{F}(\tilde{x})$ on an interval J with $x(t_0) = \tilde{x}(t_0)$, then*

$$|x(t) - \tilde{x}(t)| \le \frac{\delta}{L}\left(e^{L|t-t_0|} - 1\right) \qquad \text{for all } t \in J.$$

When $L = 0$, the factor $(e^{L|t-t_0|} - 1)/L$ is interpreted as $|t - t_0|$. The same bound holds if $\tilde{x}$ instead solves the forced equation $\dot{\tilde{x}} = F(\tilde{x}) + r(t)$ with $|r(t)| \le \delta$ on J.

Lemma 4.3 converts a uniform field error or a uniformly bounded forcing residual into a trajectory error. For $L \geq 0$ and $t \geq 0$, define

$$Q_L(t) := \begin{cases} (e^{Lt}-1)/L, & L>0, \\ t, & L=0, \end{cases} \tag{B.20}$$

which is the continuous extension of $(e^{Lt}-1)/L$ to $L=0$.

Proof of Lemma 4.3. Choose compatible lifts of the two trajectories to $\mathbb{R}^n$, and write $\Delta(t) := |x(t) - \widetilde{x}(t)|$. Because F is L-Lipschitz and $|F - \widetilde{F}| \leq \delta$, the upper right Dini derivative satisfies

$$D^+\Delta(t) \leq |F(x(t)) - F(\widetilde{x}(t))| + |F(\widetilde{x}(t)) - \widetilde{F}(\widetilde{x}(t))| \leq L\Delta(t) + \delta.$$

Since $\Delta(t_0) = 0$, multiplication by $e^{-L|t-t_0|}$ and integration give

$$\Delta(t) \leq \delta\, Q_L(|t-t_0|),$$

which is the stated estimate. For the forced equation, the same calculation replaces $|F - \widetilde{F}|$ by $|r(t)| \leq \delta$ and gives the same bound. □

We use the forced form below with a field G, a residual bound $|r(t)| \leq \bar{r}$, and initial time zero:

$$\dot{y} = G(y) + r(t), \qquad \dot{z} = G(z), \qquad y(0) = z(0),$$

if G is L_G-Lipschitz, Lemma 4.3 gives

$$\sup_{0 \leq t \leq T} |y(t) - z(t)| \leq \bar{r}\, Q_{L_G}(T). \tag{B.21}$$

The estimate is uniform over every family of initial conditions for which the same constants $\bar{r}$ and L_G apply.

B.6 Proof of Proposition 4.2

Before constructing the LFS approximation, we establish a uniform hidden-phase bound. Its perturbation term will also be used in the two universality theorems.

For an LFS model satisfying the dominance condition equation 4.3, define

$$B_i := \sum_{a=1}^{m} |w_{i,a}|, \qquad B := \left(\sum_{i=1}^{n} B_i^2 \right)^{1/2}, \qquad d_a := \frac{\left(\sum_{j=1}^{n} |c_{a,j}|^2 \right)^{1/2}}{\mu_a}, \qquad D_a := B d_a. \tag{B.22}$$

These constants depend only on the parameters of the LFS model other than ε: the visible equation equation 4.2 does not contain ε, and μ_a and $c_{a,j}$ are the unscaled coefficients of equation 4.9. The quantity B_i bounds the speed of visible oscillator i, and D_a bounds the speed of the phase of $N_a(v)$ along the visible trajectory.

The lemma below allows a bounded perturbation ρ_a of the hidden equation. The proof of Proposition 4.2 uses it with $\rho_a = 0$, the proof of Theorem 3.1 uses it with the perturbation caused by lag approximation, and the proof of Theorem 4.1 uses it with the perturbation caused by the first-harmonic grid and lag approximation together.

Lemma B.2 (Hidden oscillators track their equilibria)**.** *Fix the parameters of the hidden equation equation 4.1 satisfying the dominance condition equation 4.3, with the constants equation B.22. Let $v : [0, t_1] \to \mathbb{T}^n$ be a continuously differentiable path with $|\dot{v}_i(t)| \leq B_i$ for every i, and let h_a solve*

$$\varepsilon \dot{h}_a = -\operatorname{Im}\left(e^{ih_a}\, \overline{W_a(v(t))} \right) + \rho_a(t), \qquad e^{ih_a(0)} = N_a(v(0)), \tag{B.23}$$

where ρ_a is continuous with $|\rho_a(t)| \leq \bar{\rho}_a$. If $\varepsilon D_a + \bar{\rho}_a < \mu_a$, then for every $t \in [0, t_1]$ the angular difference between $e^{ih_a(t)}$ and $N_a(v(t))$ is at most

$$\bar{q}_a := \arcsin \frac{\varepsilon D_a + \bar{\rho}_a}{\mu_a} \leq \frac{\pi}{2} \frac{\varepsilon D_a + \bar{\rho}_a}{\mu_a}. \tag{B.24}$$

The bound is uniform in $v(0)$. For the LFS model, $\rho_a = 0$ and v is its visible trajectory, so the hypothesis reads $\varepsilon < \varepsilon_h$ with the hidden timescale threshold

$$\varepsilon_h := \min_{a: D_a > 0} \frac{\mu_a}{D_a}, \tag{B.25}$$

interpreted as $+\infty$ if all $D_a = 0$, and the angular difference is at most $(\pi/2)\varepsilon D_a/\mu_a$. For the Kuramoto model obtained by lag approximation, ρ_a is the change in the hidden equation caused by replacing its lagged couplings with added oscillators; Section B.7 introduces this model and shows that $\bar{\rho}_a$ is proportional to the lag approximation timescale η.

Proof. To bound the lag behind the moving equilibrium, first bound the speed of that equilibrium. By hypothesis $|\dot{v}| \le B$. Choose a continuous lift $\vartheta_a(t)$ of the phase of $N_a(v(t))$, which is the phase of $W_a(v(t))$. Since $\dot{W}_a = \sum_j c_{a,j} i e^{iv_j} \dot{v}_j$, the Cauchy–Schwarz inequality and the lower bound $|W_a| \ge \mu_a$ from equation B.6 give

$$|\dot{\vartheta}_a(t)| = \left| \operatorname{Im} \frac{\dot{W}_a(v(t))}{W_a(v(t))} \right| \le \frac{\left(\sum_j |c_{a,j}|^2 \right)^{1/2}}{\mu_a} |\dot{v}(t)| \le d_a B = D_a.$$

Let $q_a(t) := h_a(t) - \vartheta_a(t)$, using the lift with $q_a(0) = 0$, which the initialization permits. Writing $W_a = |W_a| e^{i\vartheta_a}$ in equation B.23 gives

$$\dot{q}_a = -\frac{|W_a(v)|}{\varepsilon} \sin q_a + \frac{\rho_a}{\varepsilon} - \dot{\vartheta}_a.$$

If $\bar{q}_a = 0$, then $\rho_a \equiv 0$ and $D_a = 0$, so $\dot{\vartheta}_a \equiv 0$ and $q_a \equiv 0$. Otherwise the condition $\varepsilon D_a + \bar{\rho}_a < \mu_a$ gives $\bar{q}_a \in (0, \pi/2)$, and at the upper boundary $q_a = \bar{q}_a$ the lower bound $|W_a| \ge \mu_a$ from equation B.6 gives

$$\dot{q}_a \le -\frac{\mu_a}{\varepsilon} \sin \bar{q}_a + \frac{\bar{\rho}_a}{\varepsilon} + D_a = 0.$$

At the lower boundary $q_a = -\bar{q}_a$, the reverse inequality gives $\dot{q}_a \ge 0$. The interval $[-\bar{q}_a, \bar{q}_a]$ is therefore forward invariant. The inequality in equation B.24 is $\arcsin z \le (\pi/2) z$ for $0 \le z \le 1$, which holds because $\arcsin$ is convex on $[0, 1]$. The constants are uniform in $v(0)$, which proves the lemma. □

Proposition 4.2 (Universality of LFS Kuramoto)**.** *Let $F : \mathbb{T}^n \to \mathbb{R}^n$ be C^1, let $t_1 > 0$, and let $\rho > 0$. There exist $m \in \mathbb{N}$, an LFS Kuramoto model equation 4.1–equation 4.2 with m hidden oscillators satisfying the dominance condition equation 4.3, and a timescale $\varepsilon > 0$, such that the following holds. For every $x_0 \in \mathbb{T}^n$, the visible trajectory initialized by $v(0) = x_0$ and $e^{ih_a(0)} = N_a(x_0)$ for $a = 1, \dots, m$ satisfies*

$$\sup_{t \in [0, t_1]} |v(t) - x(t)| < \rho,$$

where x solves the target equation $\dot{x} = F(x)$ with $x(0) = x_0$, and the difference is taken between lifts to $\mathbb{R}^n$. The model, the timescale, and the initialization rule do not depend on x_0.

Proof of Proposition 4.2. The construction has two choices: first the NAV coefficients make the reduced field close to F; then the hidden timescale makes the actual field close to that reduced field. Let $L \ge 0$ be a Lipschitz constant of the target field F, which exists because F is C^1 on the compact torus. Fix a field tolerance $\delta > 0$. Apply Proposition 4.1 to each T_i from equation 4.11 with accuracy $\delta/\sqrt{n}$. Taking the union of the finitely many resulting NAVs and assigning zero coefficient to unused outputs gives one number m of hidden oscillators and functions Ψ_i satisfying the componentwise condition equation B.9.

To realize those NAVs as hidden equilibria, choose real parameters in equation 4.1 so that

$$-\Lambda_a e^{ib_a} = \gamma_a, \qquad -k_{a,j} e^{i\psi_{a,j}} = c_{a,j}.$$

The admissibility inequality $|\gamma_a| > \sum_j |c_{a,j}|$ is exactly the dominance condition equation 4.3. Likewise, every complex coefficient in Ψ_i has the form $w_{i,a} e^{i\chi_{i,a}}$, so the visible equation 4.2 realizes the reduced field equation B.7. This fixes all parameters except the hidden timescale ε.

We next bound the difference between the actual visible velocity and the reduced field. Let v_ε be the visible trajectory of the LFS model with hidden timescale ε, initialized by $v_\varepsilon(0) = x_0$ and $e^{ih_a(0)} = N_a(x_0)$, as required in Proposition 4.2, and let x solve the target equation from the same

initial condition. The visible equation 4.2 gives $|\dot{v}_{\varepsilon,i}| \le B_i$. For $0 < \varepsilon < \varepsilon_h$, Lemma B.2 with $\rho_a = 0$ bounds the angular difference between $e^{ih_a(t)}$ and $N_a(v_\varepsilon(t))$ by $(\pi/2)\varepsilon D_a/\mu_a$ on $[0,t_1]$, uniformly in x_0. Since

$$|\sin(v_i - h_a - \chi_{i,a}) - \sin(v_i - \vartheta_a(v) - \chi_{i,a})| \le |h_a - \vartheta_a(v)|,$$

the visible equation 4.2 differs from $\widetilde{F}(v_\varepsilon)$ by at most $C_h\varepsilon$, where

$$C_h := \frac{\pi}{2}\left[\sum_{i=1}^{n}\left(\sum_{a=1}^{m}|w_{i,a}|\frac{D_a}{\mu_a}\right)^2\right]^{1/2}. \tag{B.26}$$

The field bound equation B.10 adds at most δ in passing from $\widetilde{F}$ to F, so the residual $r(t) := \dot{v}_\varepsilon(t) - F(v_\varepsilon(t))$ satisfies

$$|r(t)| \le \underbrace{|\widetilde{F}(v_\varepsilon) - F(v_\varepsilon)|}_{\le\delta} + \underbrace{|\dot{v}_\varepsilon - \widetilde{F}(v_\varepsilon)|}_{\le C_h\varepsilon} \le \delta + C_h\varepsilon. \tag{B.27}$$

The forced estimate equation B.21, with $G = F$, $L_G = L$, $y = v_\varepsilon$, and $z = x$, gives the LFS error bound

$$\sup_{0\le t\le t_1}|v_\varepsilon(t) - x(t)| \le \big(\underbrace{\delta}_{\text{density}} + \underbrace{C_h\varepsilon}_{\text{hidden tracking}}\big)\,Q_L(t_1) \qquad (0 < \varepsilon < \varepsilon_h). \tag{B.28}$$

Both terms are linear in their small parameter and share the Grönwall factor $Q_L(t_1)$, which depends only on the target field and the horizon. The constants C_h and ε_h depend only on the finite model fixed by the density step; they do not depend on ε or on x_0. The parameters of the finite model therefore enter the bound linearly, through C_h, and not through the exponent of the Grönwall factor. To make the bound smaller than ρ, choose its parameters in dependency order. Choose δ so that $\delta Q_L(t_1) < \rho/2$. This choice fixes the finite model and the constants C_h and ε_h. Then choose $\varepsilon < \varepsilon_h$ so that $C_h\varepsilon Q_L(t_1) < \rho/2$. The same choices work for every x_0, and equation B.28 proves the proposition. □

B.7 Proof of Theorem 3.1

To prove Theorem 3.1, we apply Lemma 4.1 to every lagged coupling of a fixed LFS model. The following lemma controls their combined residual before we choose the final approximation parameters.

The LFS model has two kinds of lagged coupling. A hidden-row coupling $\varepsilon^{-1}k_{a,j}\sin(h_a - v_j - \psi_{a,j})$ in equation 4.1 has source v_j and is replaced by $\varepsilon^{-1}k_{a,j}\sin(h_a - r_3^{a,j})$, where $r_1^{a,j}, r_2^{a,j}, r_3^{a,j}$ are added oscillators with source $r_0^{a,j} := v_j$ and lag $\psi_{a,j}$. A visible-row coupling $w_{i,a}\sin(v_i - h_a - \chi_{i,a})$ in equation 4.2 has source h_a and is replaced by $w_{i,a}\sin(v_i - r_3^{i,a})$, where $r_1^{i,a}, r_2^{i,a}, r_3^{i,a}$ are added oscillators with source $r_0^{i,a} := h_a$ and lag $\chi_{i,a}$. The term $\varepsilon^{-1}\Lambda_a\sin(h_a - b_a)$ is not a lagged coupling once the fixed oscillator b_a is a Kuramoto oscillator with zero intrinsic frequency and no incoming coupling, because it is then the Kuramoto coupling $-\varepsilon^{-1}\Lambda_a\sin(b_a - h_a)$.

In addition to the constants equation B.22 and equation B.26, and the absolute constants κ_0, S_3 of equation B.1, define

$$K_a := |\gamma_a| + \sum_{j=1}^{n}|c_{a,j}|, \qquad P_a := S_3\sum_{j=1}^{n}|c_{a,j}|B_j, \qquad H_a := \frac{K_aD_a}{\mu_a} + P_a\left(1 + \frac{K_a}{\mu_a}\right), \tag{B.29}$$

and

$$\begin{aligned}\varepsilon_* &:= \min_a \frac{\mu_a}{D_a + P_a}, \qquad \eta_0 := \frac{\kappa_0}{\max\{B_1,\dots,B_n,H_1,\dots,H_m\}},\\ C_\eta &:= \left[\sum_{i=1}^{n}\left(\sum_{a=1}^{m}|w_{i,a}|\left(\frac{\pi}{2}\frac{P_a}{\mu_a} + S_3H_a\right)\right)^2\right]^{1/2},\end{aligned} \tag{B.30}$$

with a minimum over an empty set, or a division by zero, interpreted as $+\infty$. The quantity K_a bounds $|W_a|$, $P_a\eta$ bounds the perturbation of hidden oscillator a caused by its hidden-row added oscillators, and H_a bounds the speed of hidden oscillator a in the lag-synthesized model. Like C_h, all of these constants depend only on the parameters of the LFS model other than ε, and not on ε, η, or x_0.

Lemma B.3 (lag approximation for a finite LFS model)**.** *Fix an LFS model equation 4.1–equation 4.2 satisfying the dominance condition equation 4.3, with the constants equation B.22, equation B.26, equation B.29, and equation B.30. Replace every lagged coupling by the network of Lemma 4.1 with lag approximation timescale η, using three separate added oscillators for each lagged coupling, and replace every fixed oscillator b_a by a Kuramoto oscillator with zero intrinsic frequency and no incoming coupling. Initialize the resulting Kuramoto model by $v(0)=x_0$, $e^{ih_a(0)}=N_a(x_0)$, the fixed oscillators at b_a, and the added oscillators as in Lemma 4.1, and let $v_{\varepsilon,\eta}$ be its visible trajectory. If*

$$0<\varepsilon<\varepsilon_* \qquad \textit{and} \qquad 0<\eta\le\min\{\varepsilon,\eta_0\},$$

then the visible field of the lag-synthesized model differs from the reduced field $\widetilde{F}$ along the visible trajectory by

$$\sup_{0\le t\le t_1}\left|\dot v_{\varepsilon,\eta}(t)-\widetilde{F}(v_{\varepsilon,\eta}(t))\right|\le C_h\varepsilon+C_\eta\eta. \tag{B.31}$$

Proof. We need a residual bound for the visible equation. We obtain it in the order in which the oscillator bounds become available, using each speed bound to control the next group of added oscillators.

Visible speeds. We first bound the sources of the hidden-row chains. Each visible oscillator in the lag-synthesized model is driven by sines with amplitudes $w_{i,a}$, so $|\dot v_i|\le B_i$.

Hidden-row added oscillators. The source v_j has speed at most B_j, and $\eta B_j\le\eta\max_j B_j\le\kappa_0$ because $\eta\le\eta_0$. The phase error equation B.4 of Lemma 4.1 gives

$$|r_3^{a,j}(t)-v_j(t)-\psi_{a,j}|\le S_3B_j\eta.$$

Hidden oscillators. The hidden-row chain errors perturb the hidden equation; we bound this perturbation so the tracking lemma applies. The equation has the form equation B.23, with

$$\rho_a:=\sum_{j=1}^{n}k_{a,j}\left[\sin(h_a-r_3^{a,j})-\sin(h_a-v_j-\psi_{a,j})\right],\qquad |\rho_a|\le\sum_{j=1}^{n}|k_{a,j}|S_3B_j\eta=P_a\eta,$$

because $|k_{a,j}|=|c_{a,j}|$ and sine has Lipschitz constant one. The conditions $\eta\le\varepsilon<\varepsilon_*$ give $\varepsilon D_a+P_a\eta\le\varepsilon(D_a+P_a)<\mu_a$. Lemma B.2 with $\bar\rho_a=P_a\eta$ therefore bounds the angular difference $h_a-\vartheta_a(v)$ between e^{ih_a} and $N_a(v)$ by

$$\bar q_a\le\frac{\pi}{2}\,\frac{\varepsilon D_a+P_a\eta}{\mu_a}. \tag{B.32}$$

Hidden speeds. To apply lag approximation to the visible-row chains, we need a speed bound for their hidden sources. Writing the hidden equation as $\varepsilon\dot h_a=-|W_a(v)|\sin(h_a-\vartheta_a(v))+\rho_a$ and using $|W_a|\le K_a$, $|\sin(h_a-\vartheta_a)|\le\sin\bar q_a=(\varepsilon D_a+P_a\eta)/\mu_a$, and $|\rho_a|\le P_a\eta$ gives

$$|\dot h_a|\le\frac{K_aD_a}{\mu_a}+\frac{\eta}{\varepsilon}P_a\left(1+\frac{K_a}{\mu_a}\right)\le H_a,$$

where the last step uses $\eta\le\varepsilon$. The factor ε^{-1} of the hidden equation is cancelled by the tracking bound, so H_a does not depend on ε.

Visible-row added oscillators. The source h_a has speed at most H_a, and $\eta H_a\le\kappa_0$ because $\eta\le\eta_0$. The phase error equation B.4 gives

$$|r_3^{i,a}(t)-h_a(t)-\chi_{i,a}|\le S_3H_a\eta.$$

Visible residual. We now combine the visible-row phase error with hidden tracking to compare the actual velocity with the reduced field. The reduced field is $\widetilde{F}_i(v)=\sum_a w_{i,a}\sin(v_i-\vartheta_a(v)-\chi_{i,a})$, and the lag-synthesized visible equation is $\dot v_i=\sum_a w_{i,a}\sin(v_i-r_3^{i,a})$. The two arguments differ by

$$r_3^{i,a}-\vartheta_a(v)-\chi_{i,a}=\big(r_3^{i,a}-h_a-\chi_{i,a}\big)+\big(h_a-\vartheta_a(v)\big),$$

so the previous step and equation B.32 give

$$|\dot v_i-\widetilde{F}_i(v)|\le\sum_{a=1}^{m}|w_{i,a}|\left[\frac{\pi}{2}\frac{D_a}{\mu_a}\varepsilon+\left(\frac{\pi}{2}\frac{P_a}{\mu_a}+S_3H_a\right)\eta\right].$$

The triangle inequality in $\mathbb{R}^n$ bounds the Euclidean norm of this residual by $C_h\varepsilon+C_\eta\eta$, with C_h from equation B.26, which is equation B.31. Every constant used is independent of x_0. □

Lemma B.3 bounds a field error, not a trajectory error, so no Grönwall factor of the LFS model appears in it. A Grönwall estimate on the full state (v,h) would instead use a Lipschitz constant of the whole LFS vector field, which is of order ε^{-1} because of the hidden rows, and would produce a factor of order $e^{ct_1/\varepsilon}$ in front of η. Comparing only the visible state avoids this, because the hidden oscillators contract toward their equilibria at rate μ_a/ε and so absorb the hidden-row lag approximation error at the same rate at which it enters.

Theorem 3.1 (Universal approximation of dynamical systems with the Kuramoto model)**.** *Let $F:\mathbb{T}^n\to\mathbb{R}^n$ be continuously differentiable, and let $x(t)$ denote the solution of $\dot{x}=F(x)$ with $x(0)=x_0$. For every horizon $t_1>0$ and tolerance $\zeta>0$, there exist $N\geq n$, parameters $\omega\in\mathbb{R}^N$ and $C\in\mathbb{R}^{N\times N}$ of the Kuramoto model*

$$\dot{s}_i=\omega_i+\sum_{j=1}^{N}C_{ij}\sin(s_j-s_i),\qquad i=1,\dots,N,$$

and a continuous initialization map $\mathcal{I}:\mathbb{T}^n\to\mathbb{T}^{N-n}$ with the following property. Write $s=(v,h)$, where the visible state v consists of the first n oscillators and the hidden state h of the remaining $N-n$. For every $x_0\in\mathbb{T}^n$, the solution initialized by $v(0)=x_0$ and $h(0)=\mathcal{I}(x_0)$ satisfies

$$\max_{t\in[0,t_1]}|v(t)-x(t)|<\zeta.$$

Proof of Theorem 3.1. We synthesize all lags in the LFS construction and add the resulting field error to the density error. Use the construction in the proof of Proposition 4.2. After δ has been chosen, the LFS model is fixed except for its hidden timescale ε, and so are the constants C_h, C_η, ε_*, and η_0 of Lemma B.3. Apply Lemma B.3 to it. The resulting system has exactly the Kuramoto form equation 2.1. The initialization of every added oscillator is an affine function of the phase of its source oscillator, the hidden oscillators are initialized by the smooth maps $x_0\mapsto N_a(x_0)$, and the fixed oscillators are constant. Together these give the continuous initialization map required by Theorem 3.1.

Let $0<\varepsilon<\varepsilon_*$ and $0<\eta\leq\min\{\varepsilon,\eta_0\}$. The residual $r(t):=\dot{v}_{\varepsilon,\eta}(t)-F(v_{\varepsilon,\eta}(t))$ is at most δ by the field bound equation B.10 plus $C_h\varepsilon+C_\eta\eta$ by the lag approximation residual equation B.31. The forced estimate equation B.21, with $G=F$, $L_G=L$, $y=v_{\varepsilon,\eta}$, and $z=x$, gives

$$\sup_{0\leq t\leq t_1}|v_{\varepsilon,\eta}(t)-x(t)|\leq\big(\underbrace{\delta}_{\text{density}}+\underbrace{C_h\varepsilon}_{\text{hidden tracking}}+\underbrace{C_\eta\eta}_{\text{lag approximation}}\big)Q_L(t_1).\tag{B.33}$$

The bound holds uniformly in x_0. The Grönwall factor $Q_L(t_1)$ depends only on F and t_1. The constants C_h and C_η depend on the finite model fixed by δ, but neither depends on ε or η.

To respect the dependence of the constants on the finite model, choose the three tolerances in order. Choose δ so that $\delta Q_L(t_1)<\zeta/3$, which fixes the NAVs, all LFS parameters other than ε, and the constants of Lemma B.3. Choose $\varepsilon<\varepsilon_*$ so that $C_h\varepsilon Q_L(t_1)<\zeta/3$. Finally choose $0<\eta\leq\min\{\varepsilon,\eta_0\}$ so that $C_\eta\eta Q_L(t_1)<\zeta/3$. The bound equation B.33 is then less than ζ, uniformly over x_0, which proves Theorem 3.1. □

B.8 Proof of Theorem 4.1

The general-coupling theorem needs two further ingredients. Lemma B.4 reconstructs a sine coupling from shifted copies of Γ; Lemma B.5 realizes each shift with an oscillator chain. Write

$$\beta_\Gamma:=\frac{\|\Gamma\|_\infty}{|\widehat{\Gamma}_1|}\geq 1,\qquad g_M:=\frac{\pi\,(\|\Gamma\|_\infty+\|\Gamma'\|_\infty)}{M|\widehat{\Gamma}_1|},\tag{B.34}$$

where $\beta_\Gamma\geq 1$ because $|\widehat{\Gamma}_1|\leq\|\Gamma\|_\infty$, and g_M is the grid error per unit amplitude of Lemma B.4.

Lemma B.4 (First-harmonic extraction with a finite-grid bound)**.** *Let Γ satisfy the hypotheses of Theorem 4.1, let $H(x)=\alpha_1\cos x+\alpha_2\sin x$, and let $|H|:=(\alpha_1^2+\alpha_2^2)^{1/2}$ be its amplitude. Set*

$$\widehat{H}_1:=\frac{\alpha_1-i\alpha_2}{2},\qquad z:=\frac{\widehat{H}_1}{\widehat{\Gamma}_1},\qquad w(\theta):=ze^{i\theta}+\overline{z}e^{-i\theta}.$$

For midpoint nodes $\theta_\ell = 2\pi(\ell - \frac{1}{2})/M$, define

$$H_M(x) := \frac{1}{M}\sum_{\ell=1}^{M} w(\theta_\ell)\Gamma(x-\theta_\ell). \tag{B.35}$$

Then the coefficients $w(\theta_\ell)/M$ are real,

$$\frac{1}{M}\sum_{\ell=1}^{M}|w(\theta_\ell)| \le \frac{|H|}{|\widehat{\Gamma}_1|}, \qquad \textit{and} \qquad \|H - H_M\|_\infty \le g_M|H|, \tag{B.36}$$

with g_M from equation B.34.

Proof. With normalized circular convolution,

$$(\Gamma * w)(x) := \frac{1}{2\pi}\int_0^{2\pi}\Gamma(x-\theta)w(\theta)\,d\theta,$$

the convolution theorem gives

$$\widehat{\Gamma * w}_k = \widehat{\Gamma}_k \widehat{w}_k.$$

The function w has Fourier support only at $k = \pm 1$, and the definition of z, together with the reality of Γ, gives $\Gamma * w = H$.

Since $|\widehat{H}_1| = |H|/2$, both w and w' are bounded by $2|z| = |H|/|\widehat{\Gamma}_1|$, which gives the first inequality in equation B.36. For fixed x, the derivative with respect to θ of the integrand is therefore bounded uniformly by

$$\|\Gamma'\|_\infty\|w\|_\infty + \|\Gamma\|_\infty\|w'\|_\infty \le \frac{\|\Gamma\|_\infty + \|\Gamma'\|_\infty}{|\widehat{\Gamma}_1|}\,|H|.$$

Every point of a midpoint interval is within π/M of its node. The normalized midpoint error is therefore at most π/M times this bound, which is $g_M|H|$. □

Lemma B.5 (Lagged Γ-couplings are approximable with oscillators)**.** *There are an integer $L_\Gamma \ge 1$ and constants $\kappa_\Gamma, S_\Gamma > 0$, depending only on Γ, with the following property. Fix $d \in \mathbb{T}$, a coefficient $a \in \mathbb{R}$, and a source phase q_j with $|\dot{q}_j(t)| \le U$ for $t \in [0, t_1]$. For every $\eta > 0$ with $\eta U \le \kappa_\Gamma$, a chain of L_Γ added oscillators, ending in r_{L_Γ} and initialized at phases determined by $q_j(0)$, satisfies*

$$|r_{L_\Gamma}(t) - q_j(t) - d| \le S_\Gamma U\eta, \qquad t \in [0, t_1], \tag{B.37}$$

so the unlagged coupling $a\Gamma(r_{L_\Gamma} - q_i)$ approximates the lagged coupling $a\Gamma(q_j - q_i + d)$ with error at most

$$|a|\,\|\Gamma'\|_\infty S_\Gamma U\eta. \tag{B.38}$$

The bound holds uniformly in d and the initial phases.

Proof. The nonzero first Fourier coefficient implies that Γ is nonconstant. Choose an open interval J on which $\Gamma'(-\xi)$ has constant nonzero sign $\varsigma \in \{-1, 1\}$. Choose a closed interval $J' = [\xi_1, \xi_2] \subset J$ and $q_* > 0$ such that $J' + [-q_*, q_*] \subset J$, and set

$$\lambda_\Gamma := \min_{\xi \in J' + [-q_*, q_*]} |\Gamma'(-\xi)| > 0,$$

which is positive because the minimum of a continuous nonvanishing function over a compact interval is attained. Choose L_Γ with $L_\Gamma(\xi_2 - \xi_1) \ge 2\pi$. The sums $d_1 + \cdots + d_{L_\Gamma}$ with $d_k \in J'$ fill the interval $[L_\Gamma\xi_1, L_\Gamma\xi_2]$ of length at least 2π, so every $d \in \mathbb{T}$ can be written as

$$d_1 + \cdots + d_{L_\Gamma} = d \pmod{2\pi}, \qquad d_k \in J'.$$

Let $r_0 := q_j$, and introduce

$$\dot{r}_k = -\frac{\varsigma}{\eta}\Gamma(-d_k) + \frac{\varsigma}{\eta}\Gamma(r_{k-1} - r_k), \qquad r_k(0) = q_j(0) + \sum_{p=1}^{k} d_p. \tag{B.39}$$

Each equation has the form equation 4.17. For $q_k := r_k - r_{k-1} - d_k$, the mean-value theorem gives $\varsigma[\Gamma(-d_k - q_k) - \Gamma(-d_k)] = -\lambda_k q_k$ with $\lambda_k \ge \lambda_\Gamma$ whenever $|q_k| \le q_*$. Set $R_0 := U$ and $R_k := (\|\Gamma'\|_\infty/\lambda_\Gamma)R_{k-1}$, and define

$$\kappa_\Gamma := q_*\lambda_\Gamma\left(\frac{\lambda_\Gamma}{\|\Gamma'\|_\infty}\right)^{L_\Gamma - 1}, \qquad S_\Gamma := \frac{1}{\lambda_\Gamma}\sum_{k=0}^{L_\Gamma - 1}\left(\frac{\|\Gamma'\|_\infty}{\lambda_\Gamma}\right)^k.$$

The condition $\eta U \le \kappa_\Gamma$ is $\eta R_{L_\Gamma - 1}/\lambda_\Gamma \le q_*$. The invariant-interval argument that proves equation B.3 then gives

$$|q_k(t)| \le \eta \frac{R_{k-1}}{\lambda_\Gamma}, \qquad |\dot{r}_k(t)| \le R_k, \qquad k = 1, \dots, L_\Gamma.$$

Summing the stage errors gives $|r_{L_\Gamma} - q_j - d| \le \eta \sum_k R_{k-1}/\lambda_\Gamma = S_\Gamma U \eta$, which is equation B.37. Applying the Lipschitz constant $\|\Gamma'\|_\infty$ to the output coupling proves equation B.38. The integer L_Γ and the constants κ_Γ, S_Γ depend only on J, J', and q_*, which are chosen from Γ alone. □

Theorem 4.1 (Universality for general periodic coupling)**.** *Let $\Gamma \in C^2(\mathbb{T}, \mathbb{R})$ be a 2π-periodic coupling function whose first Fourier coefficient is nonzero,*

$$\widehat{\Gamma}_1 := \frac{1}{2\pi} \int_0^{2\pi} \Gamma(x) e^{-ix}\, dx \neq 0.$$

Then the conclusion of Theorem 3.1 holds with the Kuramoto model replaced by

$$\dot{q}_i = \omega_i + \sum_{j=1}^{N} K_{ij} \Gamma(q_j - q_i), \qquad i = 1, \dots, N,$$

where $N \in \mathbb{N}$, $\omega \in \mathbb{R}^N$, and $K \in \mathbb{R}^{N \times N}$.

Proof of Theorem 4.1. Fix δ and the LFS model from the proof of Proposition 4.2, which is then fixed except for its hidden timescale ε, with the constants equation B.22 and equation B.26 and the constant K_a of equation B.29. Build the general periodic coupling model in two steps.

Construction. Every coupling of the LFS model is a first harmonic of a phase difference: $\varepsilon^{-1} k_{a,j} \sin(h_a - v_j - \psi_{a,j})$ is $H(v_j - h_a)$ with $|H| = \varepsilon^{-1}|k_{a,j}|$, the term $\varepsilon^{-1}\Lambda_a \sin(h_a - b_a)$ is $H(b_a - h_a)$ with $|H| = \varepsilon^{-1}|\Lambda_a|$, and $w_{i,a} \sin(v_i - h_a - \chi_{i,a})$ is $H(h_a - v_i)$ with $|H| = |w_{i,a}|$. Replace each of them by the sum H_M of Lemma B.4, with one grid size M for all of them. Each term $\Gamma(s_j - s_i - \theta_\ell)$ of H_M is a lagged Γ-coupling with lag $d = -\theta_\ell$. If its source is a fixed oscillator b_a, replace it by $\Gamma(b_{a,\ell} - h_a)$, where $b_{a,\ell}$ is a new fixed oscillator at the constant phase $b_a - \theta_\ell$; this replacement is exact. Otherwise replace it by the chain of added oscillators of Lemma B.5 with timescale η. Initialize $v(0) = x_0$, $e^{ih_a(0)} = N_a(x_0)$, and every chain as in Lemma B.5, and let v_Γ be the visible trajectory.

Constants. In addition to β_Γ and g_M of equation B.34 and $L_\Gamma, \kappa_\Gamma, S_\Gamma$ of Lemma B.5, define

$$C_h^\Gamma := \beta_\Gamma C_h, \quad P_a^\Gamma := \frac{\|\Gamma'\|_\infty S_\Gamma \beta_\Gamma}{|\widehat{\Gamma}_1|} \sum_{j=1}^{n} |c_{a,j}| B_j, \quad H_a^\Gamma := \frac{K_a \beta_\Gamma D_a}{\mu_a} + \left(1 + \frac{K_a}{\mu_a}\right)\left(K_a + P_a^\Gamma\right), \tag{B.40}$$

and

$$\begin{aligned} \varepsilon_\Gamma &:= \min_a \frac{\mu_a}{\beta_\Gamma D_a + K_a + P_a^\Gamma}, \qquad \eta_0^\Gamma := \frac{\kappa_\Gamma}{\max\{\beta_\Gamma B_1, \dots, \beta_\Gamma B_n, H_1^\Gamma, \dots, H_m^\Gamma\}}, \\ C_{\mathrm{grid}} &:= \left[\sum_{i=1}^{n}\left(\sum_{a=1}^{m} |w_{i,a}| \left(1 + \frac{\pi}{2}\frac{K_a}{\mu_a}\right)\right)^2\right]^{1/2}, \\ C_\eta^\Gamma &:= \left[\sum_{i=1}^{n}\left(\sum_{a=1}^{m} |w_{i,a}| \left(\frac{\pi}{2}\frac{P_a^\Gamma}{\mu_a} + \frac{\|\Gamma'\|_\infty S_\Gamma H_a^\Gamma}{|\widehat{\Gamma}_1|}\right)\right)^2\right]^{1/2}. \end{aligned} \tag{B.41}$$

They play the roles of C_h, P_a, H_a, ε_*, η_0, and C_η in Lemma B.3, and they depend only on Γ and the LFS model fixed by δ, not on ε, M, η, or x_0. Assume

$$0 < \varepsilon < \varepsilon_\Gamma, \qquad g_M \le \varepsilon, \qquad 0 < \eta \le \min\{\varepsilon, \eta_0^\Gamma\}.$$

As in the proof of Lemma B.3, each step below uses only the bounds of the previous steps.

Visible speeds. By the first inequality in equation B.36, the shifted copies that replace $w_{i,a} \sin(v_i - h_a - \chi_{i,a})$ have coefficients whose magnitudes sum to at most $|w_{i,a}|/|\widehat{\Gamma}_1|$. Each copy of Γ is bounded by $\|\Gamma\|_\infty$, so $|\dot{v}_i| \le \beta_\Gamma B_i$.

Hidden-row added oscillators. The source v_j has speed at most $\beta_\Gamma B_j$, and $\eta \beta_\Gamma B_j \le \kappa_\Gamma$ because $\eta \le \eta_0^\Gamma$. Lemma B.5 bounds the coupling error of each shifted copy with coefficient a by $|a| \|\Gamma'\|_\infty S_\Gamma \beta_\Gamma B_j \eta$.

Hidden oscillators. Multiplied by ε, the hidden equation has the form equation B.23 with ρ_a equal to the grid error plus the chain error. The grid error of the terms entering h_a is at most $g_M(|\Lambda_a|+\sum_j|k_{a,j}|)=g_M K_a$ by equation B.36. The chain error is at most $\sum_j(|k_{a,j}|/|\widehat{\Gamma}_1|)\|\Gamma'\|_\infty S_\Gamma\beta_\Gamma B_j\eta=P_a^\Gamma\eta$ by the previous step and the coefficient bound in equation B.36. The proof of Lemma B.2 uses the visible speed bound only through $|\dot v|\le B$, so the lemma applies with B replaced by $\beta_\Gamma B$, hence D_a replaced by $\beta_\Gamma D_a$, and with $\bar\rho_a=g_M K_a+P_a^\Gamma\eta$. Its hypothesis holds because $g_M,\eta\le\varepsilon<\varepsilon_\Gamma$, and it bounds the angular difference between e^{ih_a} and $N_a(v)$ by

$$\bar q_a\le\frac{\pi}{2}\,\frac{\varepsilon\beta_\Gamma D_a+g_M K_a+P_a^\Gamma\eta}{\mu_a}. \tag{B.42}$$

Hidden speeds. As in the proof of Lemma B.3, $|W_a|\le K_a$, $\sin\bar q_a\le(\varepsilon\beta_\Gamma D_a+g_M K_a+P_a^\Gamma\eta)/\mu_a$, and $g_M,\eta\le\varepsilon$ give $|\dot h_a|\le H_a^\Gamma$.

Visible-row added oscillators. The source h_a has speed at most H_a^Γ, and $\eta H_a^\Gamma\le\kappa_\Gamma$. Lemma B.5 bounds the coupling error of each shifted copy with coefficient a by $|a|\|\Gamma'\|_\infty S_\Gamma H_a^\Gamma\eta$.

Visible residual. Passing from the visible equation of the constructed model to the reduced field $\widetilde F_i(v)=\sum_a w_{i,a}\sin(v_i-\vartheta_a(v)-\chi_{i,a})$ takes three steps for each a: replace every chain output by its target phase, which costs at most $(|w_{i,a}|/|\widehat{\Gamma}_1|)\|\Gamma'\|_\infty S_\Gamma H_a^\Gamma\eta$; replace H_M by the sine H, which costs at most $g_M|w_{i,a}|$; and replace h_a by $\vartheta_a(v)$, which costs at most $|w_{i,a}|\bar q_a$. With equation B.42 and the triangle inequality in $\mathbb{R}^n$,

$$|\dot v_\Gamma-\widetilde F(v_\Gamma)|\le C_h^\Gamma\varepsilon+C_{\mathrm{grid}}g_M+C_\eta^\Gamma\eta.$$

Trajectory error. Adding the field bound equation B.10 and applying the forced estimate equation B.21 with $G=F$, $L_G=L$, $y=v_\Gamma$, and $z=x$ gives

$$\boxed{\sup_{0\le t\le t_1}|v_\Gamma(t)-x(t)|\le\big(\underbrace{\delta}_{\text{density}}+\underbrace{C_h^\Gamma\varepsilon}_{\text{hidden tracking}}+\underbrace{C_{\mathrm{grid}}g_M}_{\text{first-harmonic grid}}+\underbrace{C_\eta^\Gamma\eta}_{\text{lag approximation}}\big)Q_L(t_1).} \tag{B.43}$$

Here g_M is proportional to $1/M$ by equation B.34, and the bound holds uniformly in x_0, with the same Grönwall factor $Q_L(t_1)$ as in Theorem 3.1. Compared with equation B.33, the only new term is the grid error, and the constants of the hidden tracking and lag approximation terms change only through Γ.

Choose δ so that $\delta Q_L(t_1)<\zeta/4$, which fixes the LFS model and all constants above. Choose $\varepsilon<\varepsilon_\Gamma$ so that $C_h^\Gamma\varepsilon Q_L(t_1)<\zeta/4$. Then choose M with $g_M\le\varepsilon$ and $C_{\mathrm{grid}}g_M Q_L(t_1)<\zeta/4$, and $\eta\le\min\{\varepsilon,\eta_0^\Gamma\}$ with $C_\eta^\Gamma\eta Q_L(t_1)<\zeta/4$; these two choices do not constrain each other. The fixed oscillators and added oscillators are initialized continuously from the visible initial state, all coefficients are real, and every coupling is an unlagged copy of Γ. Thus the constructed system has the form equation 4.17, and equation B.43 proves Theorem 4.1. □

C Numerical-Experiment Details

C.1 Dynamical System Fitting

Every target is two-dimensional and is evaluated for $t_1=0.5$. When a system is defined in natural coordinates y, we use the affine embedding $v=a(y-y_c)$, so $\dot v=aG(y_c+v/a)$. The common fitted region is $K=[-\pi/2,\pi/2]^2$. The seven targets are:

$$\begin{aligned}
\text{Harmonic:}\quad & F(v)=\big(\sin(3v_1)+0.3\sin(v_1-v_2),\ \cos(2v_2)+0.3\sin(v_2-v_1)\big),\\
\text{Linear sink:}\quad & G(y)=(-0.6y_1+2y_2,\ -2y_1-0.6y_2),\\
\text{Pendulum:}\quad & G(y)=(y_2,-\sin y_1),\\
\text{Lotka–Volterra:}\quad & G(y)=(y_1-y_1y_2,\ y_1y_2-y_2),\\
\text{Duffing:}\quad & G(y)=(y_2,\ y_1-y_1^3-0.15y_2),\\
\text{Van der Pol:}\quad & G(y)=(y_2,\ (1-y_1^2)y_2-y_1),\\
\text{Hopf:}\quad & G(y)=(y_1-y_2-(y_1^2+y_2^2)y_1,\ y_1+y_2-(y_1^2+y_2^2)y_2).
\end{aligned}$$

Table 3: Dynamical system fitting at $m = 16$ ($M = 130$), the rightmost points of Figure 3. The field error column reports δ_{rel} of the reduced field, and trajectory error columns report the maximum in-box trajectory error of the reduced field and of the Kuramoto network at each ε.

	Field error	Max trajectory error				
System	δ_{rel}	Reduced field	$\varepsilon = 1$	$\varepsilon = 0.1$	$\varepsilon = 0.01$	$\varepsilon = 0.001$
Harmonic	0.97%	.00801	.163	.0944	.0195	.00697
Linear sink	0.82%	.0326	.607	.277	.0390	.0298
Pendulum	0.85%	.00539	.157	.0672	.00978	.00471
Lotka–Volterra	0.87%	.00725	.267	.144	.0162	.00713
Duffing	2.1%	.0250	.453	.269	.0443	.0259
Van der Pol	1.4%	.0199	1.05	.409	.0454	.0200
Hopf	0.93%	.00989	.313	.165	.0311	.00924

Table 4: Flow-matching results. Field error columns report the largest reduced-field validation error δ and the relative field error δ_{rel} over the 16 subintervals, endpoint error columns the RMS distance from the MLP's endpoints, and W_2^2 columns compare endpoints with target samples. Kuramoto columns compile the $\varepsilon = 10^{-5}$ fit with $\eta = 10^{-6}$.

	Field error		Endpoint error		W_2^2 ($\times 10^{-3}$)	
Pair	δ	δ_{rel}	LFS	Kuramoto	Kuramoto	MLP
Gaussian → 8 Gaussians	1.87	5.8%	.0151	.0152	3.00	3.51
Moons → 8 Gaussians	2.64	4.0%	.0158	.0157	1.86	1.78
Gaussian → Moons	4.22	10.6%	.0195	.0196	.296	.200
Gaussian → S-curve	.811	2.9%	.0108	.0108	1.02	.889

The linear sink, pendulum, and Duffing targets use $y_c = (0,0)$ and $a = 1$. Lotka–Volterra uses $y_c = (1,1)$ and $a = 2.5$. Van der Pol uses $y_c = (0,0)$ and $a = (\pi/4.8, \pi/6.4)$, applied componentwise. Hopf uses $y_c = (0,0)$ and $a = \pi/3.2$.

The reduced field is fitted with random seed zero on 1000 training and 1000 validation samples drawn uniformly from the fitted box, using Adam with learning rate 3×10^{-2} for 8000 steps and the dominance condition enforced as $\sum_j |c_{a,j}| < 0.97$ with $|\gamma_a| = 1$. We fit $m \in \{1,2,4,8,16\}$ hidden oscillators and compile each fit to a Kuramoto network by lag approximation with $\eta = 0.03\varepsilon$; since only the couplings into the hidden oscillators carry lags, the networks have $M = n + 2m + 3mn \in \{10, 18, 34, 66, 130\}$ oscillators[1]. The relative field error δ_{rel} is evaluated on a 201×201 grid over the fitted box. Trajectories of the target, the reduced field, and the Kuramoto network start from a 10×10 grid spanning 90% of the fitted box, and the Kuramoto network is evaluated at $\varepsilon \in \{1, 0.1, 0.01, 0.001\}$. All trajectories use Tsit5 with relative tolerance 10^{-6}, absolute tolerance 10^{-8}, and 100 evaluation times.

We score only the in-box samples, the times at which the target trajectory lies in K, since the reduced field is fitted only on K and is not expected to match the target outside it. The restriction is on the target, so a fitted trajectory that leaves K while the target stays inside is an approximation failure and remains scored. The reported error is the largest coordinate difference between the fitted and target trajectories over all initial conditions and in-box times.

C.2 Flow Matching

Each task transports a source p_0 to a target p_1 on $[-0.5, 0.5]^2$ over $t \in [0,1]$. With $\xi \sim \mathcal{N}(0, I_2)$, the distributions are:

$$
\begin{aligned}
\text{Gaussian:} \quad & y = 0.17\,\xi, \\
\text{8 Gaussians:} \quad & y = \tfrac{1}{13}\big(5\mu_k + 0.1^{1/4}\xi\big), \quad \mu_k = \big(\cos\tfrac{k\pi}{4}, \sin\tfrac{k\pi}{4}\big),\ k \sim \mathcal{U}\{0,\dots,7\}, \\
\text{Moons:} \quad & y = \tfrac{1}{13}\big(3(\tilde{y} + (\nu,\nu)) - 1\big), \quad \nu \sim \mathcal{U}[0, 0.2), \\
\text{S-curve:} \quad & y = \frac{z - \bar{z}}{2\max_i \|z_i - \bar{z}\|_\infty}, \quad z = \big(\sin\tau,\ \mathrm{sign}(\tau)(\cos\tau - 1)\big) + 0.1\,\xi,\ \tau \sim \mathcal{U}(-\tfrac{3\pi}{2}, \tfrac{3\pi}{2}).
\end{aligned}
$$

[1] It can be shown that removing χ from the equation 4.7 does not affect the universality. We train the reduced field without it to reduce the model size.

For Moons, $\tilde{y}$ is spaced evenly on the arcs $(\cos\alpha, \sin\alpha)$ and $(1-\cos\alpha, \frac{1}{2}-\sin\alpha)$, $\alpha \in [0,\pi]$; for the S-curve, $\bar{z}$ is the batch mean.

Each teacher is a GELU MLP with four hidden layers of width 128, trained with optimal-transport conditional flow matching (Tong et al., 2024). Each data point $y=(y_1,y_2)$ at time t is mapped to the visible phases $v=(ay_1, ay_2, t)$ with $a=4.5$; to fit the domain in the fitted box, we scale the distributions and the vector field accordingly. We set path noise level $\sigma=0$, batch size 256, and Adam (learning rate 10^{-3}, cosine schedule) for 10^5 steps. Nearly all trajectories remain in the fitted box $[-2.475, 2.475]^2 \times [0,1]$.

Then we fit the MLP's velocity field using the reduced-field construction C.1. By treating time as a third visible phase, the MLP's velocity field can be represented as an autonomous system on a bounded domain, the type of system Theorem 3.1 addresses. Each MLP is approximated by 16 reduced fields, each with $m=128$ hidden and $n=3$ visible oscillators. Each is trained on $45{,}000$ states drawn once from 4096 teacher trajectories, with (v_1, v_2) perturbed by $\mathcal{N}(0, 0.05^2 I_2)$, so that it fits a neighborhood of these trajectories. The parameters of the hidden oscillators are trained by Adam (learning rate 3×10^{-2}, cosine schedule) for 2000 steps. The couplings into the visible oscillators have no lag, so the reduced field is linear in their weights, which are solved exactly at each step by ridge regression with penalty 10^{-7}.

Each LFS model at $\varepsilon=10^{-5}$ is compiled to a Kuramoto model by lag approximation (Lemma 4.1) with $\eta=10^{-6}$. Only the mn couplings into the hidden oscillators carry lags, so each subinterval has $M=n+2m+3mn=1411$ oscillators. The 16 networks are composed by switching parameters at each subinterval boundary while carrying all phases over. Trajectories use Tsit5 (relative and absolute tolerances 10^{-7} and 10^{-9}). All endpoints start from the same 1024 source samples, and W_2^2 is computed exactly by optimal transport against the same 1024 target samples. On every pair, the LFS and Kuramoto endpoint errors differ by at most 10^{-4}, as the lag term $C_\eta\eta$ of equation 4.15 predicts for small η.