

Data-driven selection of coarse-grained models of coupled oscillatorsJordan Snyder^{1,2,*}, Anatoly Zlotnik³, and Andrey Y. Lokhov³¹*Department of Mathematics, University of California, Davis, California 95616, USA*²*Department of Applied Mathematics, University of Washington, Seattle, Washington 98195, USA*³*Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA*

(Received 28 September 2020; accepted 4 December 2020; published 22 December 2020)

Systematic discovery of reduced-order closure models for multiscale processes remains an important open problem in complex dynamical systems. Even when an effective lower-dimensional representation exists, reduced models are difficult to obtain using solely analytical methods. Rigorous methodologies for finding such coarse-grained representations of multiscale phenomena would enable accelerated computational simulations and provide fundamental insights into the complex dynamics of interest. We focus on a heterogeneous population of oscillators of Kuramoto type as a canonical model of complex dynamics and develop a data-driven approach for inferring its coarse-grained description. Our method is based on a numerical optimization of the coefficients in a general equation of motion informed by analytical derivations in the thermodynamic limit. We show that certain assumptions are required to obtain an autonomous coarse-grained equation of motion. However, optimizing coefficient values enables coarse-grained models with conceptually disparate functional forms, yet comparable quality of representation, to provide accurate reduced-order descriptions of the underlying system.

DOI: [10.1103/PhysRevResearch.2.043402](https://doi.org/10.1103/PhysRevResearch.2.043402)**I. INTRODUCTION**

Numerical simulations of complex multiscale phenomena are fundamental to modern science, for which commonly sought goals involve the development of tractable yet accurate reduced-order models. There are many approaches to this sort of problem throughout many diverse domains. For instance, in turbulence modeling this is known as the closure problem for the Reynolds-averaged Navier-Stokes equation (RANS) [1–4]. In molecular dynamics it is known simply as coarse graining [5–7]. Mean-field approaches to analyzing stochastic dynamics on networks also fall into this description [8,9].

Investigations on model reduction of nonlinear dynamics often consider the Kuramoto model of coupled oscillators [10], which has long been studied as an example of collective behavior [11–15]. The key features of the Kuramoto model are its composition of many oscillatory units with distinct natural frequencies and pairwise coupling that tends to drive phases together. Its popularity as an object of study comes from its tractability in certain special limits [16,17] and its nonetheless rich phenomenology [18–20]. In addition, the Kuramoto model and its variants describe a number of synchronization phenomena in domains of diverse nature such as coupled Josephson junctions [21], neuroscience [22], chemical oscillators [23,24], and the power grid [25].

Substantial work on the Kuramoto model focuses on understanding synchronization under various conditions on the natural frequencies and on the structure of the network that couples oscillators to one another. One main mode of understanding consists of finding simplified mathematical descriptions of the dynamics, and this is where the closure problem arises. Perhaps the best-known example of this type is the seminal work by Ott and Antonsen [17], who showed that in the $N \rightarrow \infty$ limit and under certain conditions on the distribution of natural frequencies and initial phases, the center of mass of a population of Kuramoto oscillators obeys an autonomous ordinary differential equation (ODE). Other studies have focused more on complex coupling topologies, proposing techniques using spectral information to merge nodes together [26] or otherwise systematically discard irrelevant degrees of freedom [27]. Still others take a more strictly data-driven approach and seek, e.g., closed equations of motion for low-order moments of the distribution of phases [28,29] or to identify good coarse-grained variables via manifold learning techniques [30]. Finally there are approaches that employ a “collective coordinate” ansatz governing the phase of each oscillator within a phase-locked cluster and thereby arrive at a closed equation of motion [31–35].

Here we use a collective coordinate ansatz to derive coarse-grained equations of motion at the level of phase-locked clusters that are consistent with arbitrary distributions of natural frequency within each cluster. We focus especially on Gaussian and Cauchy distributions, though our approach is generic. The resulting equations are precisely determined in the $N \rightarrow \infty$ limit in terms of parameters of the distributions in question and the matrix of coupling strengths. Similar to other previously obtained analytical results, the emergence of an explicit closure model crucially depends on a number of

*jsnyd@uw.edu

simplifying assumptions, and no closed-form reduced-order equations of motion are known for finite systems that do not satisfy these assumptions. We aim to move beyond the $N \rightarrow \infty$ limit by treating our derived coarse-grained equations as inductive biases and allowing data from finite- N simulations to determine optimal parameter values. The result is a systematic data-driven procedure for finding physically meaningful coarse-grained models of finite systems of coupled oscillators that provide a more accurate reduced-order description of the system.

This manuscript is organized as follows. In Sec. II we review in detail prior work and precisely state the problem we address. In Sec. III A we state the set of conditions and approximations that let us derive coarse-grained equations of motion for coupled oscillator systems, and in Sec. III B we introduce our data-driven approach for finding optimal coefficient values to use in these coarse-grained equations. In Sec. IV we apply our methods to a concrete example system and evaluate the performance of both our theoretically derived coarse-grained models and the same models optimized to fit training data.

II. BACKGROUND

The Kuramoto model is the ODE system

$$\dot{\theta}_i = \omega_i + \sum_{j=1}^N K_{ij} \sin(\theta_j - \theta_i), \quad i = 1 \dots N, \quad (1)$$

where $\theta_i \in S^1$ is the *phase* of the i th oscillator, $\omega_i \in \mathbb{R}$ is its *natural frequency*, N is the total number of oscillators, and $K \in \mathbb{R}^{N \times N}$ is the *coupling matrix* that defines which oscillators influence each other [10].

We can equivalently formulate the above model in terms of *complex phases* $y_i = \exp(i\theta_i)$. A straightforward calculation yields the equivalent representation

$$\dot{y}_i = i\omega_i y_i + \frac{1}{2} \sum_{j=1}^N K_{ij} (y_j - y_j^* y_i^2), \quad (2)$$

where we have used the fact that $y_i^* y_i = 1$. We subsequently explain why this representation of the Kuramoto dynamics is more convenient for our purposes.

Perhaps the best-studied case of the Kuramoto model is that of *mean-field* coupling, where $K_{ij} = K/N$ for all i, j . In this case it is well known that in the limit $N \rightarrow \infty$, the system (1) exhibits a phase transition with respect to K : if ω_i are sampled from a symmetric, unimodal distribution, then there exists K_c such that for $K < K_c$, the oscillators behave mostly independently, while for $K > K_c$ a subset of oscillators spontaneously locks to a single frequency [36].

Synchronization in the Kuramoto model is typically quantified by the order parameter

$$z := \frac{1}{N} \sum_{j=1}^N e^{i\theta_j} = R e^{i\Phi}, \quad (3)$$

where $R \in [0, 1]$ is the *synchrony* and $\Phi \in [0, 2\pi)$ is the *average phase*. If all phases are equal then $R = 1$, and if the phases

are spread uniformly over the unit circle, then $R \approx 0$. Thus R is a natural measure of synchronization.

The dynamics of z depend on the dynamics of all θ_i , but it is natural to suppose that in some limit there exists a *closed* equation for the dynamics of z . Indeed, there is, as demonstrated by Ott and Antonsen [17]. Assuming that ω_i are Cauchy distributed, i.e., $\omega_i \sim g(\omega)$ where g is the Cauchy probability density function with mode Ω and width δ , and that the coupling is mean field, then z evolves according to a Stuart-Landau equation,

$$\dot{z} = \left(i\Omega - \delta + \frac{K}{2} \right) z - \frac{K}{2} z |z|^2. \quad (4)$$

The form (4) shows clearly that $z = 0$ is always a solution but undergoes a pitchfork bifurcation at $K_c = 2\delta$ when a new solution with $R = \sqrt{1 - 2\delta/K}$ (and $d\Phi/dt = \Omega$) appears, representing partial synchrony that becomes global synchrony (i.e., $R \rightarrow 1$) as $K \rightarrow \infty$.

Interestingly, the same analysis carries over to the case where oscillators are not coupled all to all but are divided into subsets such that the strength of coupling between any two oscillators depends on the subsets to which they belong. Let $\Pi = (P_1, \dots, P_C)$ be a partition of the index set $\{1, \dots, N\}$ and let K be a $C \times C$ matrix of coupling strengths. Following [17], such a modular system can be written as

$$\dot{\theta}_i = \omega_i + \sum_{\sigma'=1}^C \frac{K_{\sigma\sigma'}}{|P_{\sigma'}|} \sum_{j \in P_{\sigma'}} \sin(\theta_j - \theta_i). \quad (5)$$

The same mathematical machinery as before can be applied to show that if for every σ , $\{\omega_i | i \in P_\sigma\}$ are distributed according to a Cauchy distribution with mode Ω_σ and width δ_σ , then the *cluster order parameters* $\{z_\sigma\}$, defined by

$$z_\sigma = \langle e^{i\theta_j} \rangle_{j \in P_\sigma} := \frac{1}{|P_\sigma|} \sum_{j \in P_\sigma} e^{i\theta_j}, \quad (6)$$

obey a coupled Stuart-Landau equation of the form

$$\dot{z}_\sigma = i(\Omega_\sigma + i\delta_\sigma) z_\sigma + \frac{1}{2} \sum_{\sigma'=1}^C K_{\sigma\sigma'} (z_{\sigma'} - z_\sigma^* z_\sigma^2). \quad (7)$$

The consequences of this equation for the existence of mesoscale synchronization were discussed in detail in a prior study [37]. For now, we note that the above representation is mathematically equivalent to the original Kuramoto model written in terms of complex phases [i.e., Eq. (2)], albeit with the linear term containing a nonzero real part. In this sense, this natural emergence of Eq. (7) implies that, in the Kuramoto model, groups of oscillators can behave collectively as a single oscillator. In other words, there is a renormalization procedure for coupled oscillator systems that remains within the same model class [in particular, the one defined by Eq. (7)].

The above analyses are valid in the limit $N \rightarrow \infty$ and require certain regularity assumptions on the initial distribution of phases. As such, the behavior of a finite- N Kuramoto system will in general exhibit fluctuations relative to (4) or (7). On the other hand, the dynamics of any finite- N system in the absence of noise are in fact deterministic and hence its evolution can be computed exactly from its current (N -dimensional)

state. This raises the natural question: Given a Kuramoto system of the form (1), what is the coarsest partition $\Pi = (P_1, \dots, P_C)$ such that the local order parameters (6) evolve according to an autonomous ODE? And what is that ODE?

Alternatively, we suppose that Π is a partition and $z^\Pi = (z_1, \dots, z_C) \in \mathbb{C}^C$ is the vector of local order parameters as defined above. Does the value of z^Π exactly determine the value of \dot{z}^Π ? And if so, what is the functional relationship $\dot{z}^\Pi = f(z^\Pi)$? Our proposal to address this question is to learn the function f from samples $\{(z^\Pi(t), \dot{z}^\Pi(t)) | t \in [0, T]\}$ obtained by coarse-graining experimental or simulated dynamics data of the full system. This learning task becomes a well-posed optimization problem when a finite set of basis functions that can be used to construct f is chosen. Finding f such that $\|\dot{z}^\Pi - f(z^\Pi)\|$ is small means that the dynamics of z^Π (i.e., \dot{z}^Π) can be accurately computed from only the value of z^Π , and we say that the system admits a good coarse-grained model described by $f(z^\Pi)$.

III. METHODS

We now propose criteria to select the coarse-graining partition Π and present analytical calculations that suggest which finite-dimensional function spaces are likely to contain a satisfactory coarse-grained model in the sense of approximating cluster-average phase trajectories. We do this by imposing certain assumptions on the distribution of individual oscillators' phases and arrive at expressions for the coarse-grained model in terms of static parameters of the fine-grained model. These expressions hence suggest a natural set of *basis functions* that can be used to inform our data-driven procedure for inferring coarse-grained dynamics directly from observed time-series data.

A. Conditions and analytical implications

To fix notation, we consider the Kuramoto model in complex form (2) and refer interchangeably to either real-valued phases θ_i or their complex versions $y_i := \exp(i\theta_i)$. It is possible to write exact evolution equations for the coarse variables z_σ (6) if we allow explicit time dependence through the residuals $x_i(t) = z_\sigma(t) - y_i(t)$. Explicitly, we have

$$\dot{z}_\sigma = A_\sigma + B_\sigma z_\sigma + D_\sigma z_\sigma^2 + \sum_{\sigma'=1}^C [E_{\sigma\sigma'}(z_{\sigma'} - z_{\sigma'}^* z_\sigma^2) + F_{\sigma\sigma'} z_{\sigma'}^* + G_{\sigma\sigma'} z_\sigma z_{\sigma'}^*], \quad (8)$$

where each coefficient (A)–(G) is a function of the residuals $x_i(t)$. The terms with coefficients B_σ and $E_{\sigma\sigma'}$ are those that are present in the original Kuramoto equations written in complex form.

The functional form of the coefficients is given by

$$A_\sigma = \left\langle i\omega_i x_i + \sum_{\sigma'=1}^C \sum_{j \in P_{\sigma'}} K_{ij} (x_j - x_j^* x_i^2) \right\rangle_{i \in P_\sigma} \quad (9)$$

$$B_\sigma = \left\langle i\omega_i + \sum_{\sigma'=1}^C \sum_{j \in P_{\sigma'}} K_{ij} (-2x_j^* x_i) \right\rangle_{i \in P_\sigma} \quad (10)$$

$$D_\sigma = \left\langle \sum_{\sigma'=1}^C \sum_{j \in P_{\sigma'}} K_{ij} (-x_j^*) \right\rangle_{i \in P_\sigma} \quad (11)$$

$$E_{\sigma\sigma'} = \left\langle \sum_{j \in P_{\sigma'}} K_{ij} \right\rangle_{i \in P_\sigma} \quad (12)$$

$$F_{\sigma\sigma'} = \left\langle \sum_{j \in P_{\sigma'}} K_{ij} (-x_j^2) \right\rangle_{i \in P_\sigma} \quad (13)$$

$$G_{\sigma\sigma'} = \left\langle \sum_{j \in P_{\sigma'}} K_{ij} (-2x_j) \right\rangle_{i \in P_\sigma} \quad (14)$$

If one could obtain an accurate expression for x_i in terms of microscopic parameters ω_i , K_{ij} , and the coarse variables z_σ , then the explicit time-dependence could be removed, leaving an autonomous ODE for the coarse variables $\{z_\sigma\}$. In order to arrive to simplified expressions, we build on the work of Gottwald concerning collective coordinates [31]. That study proposes that the residual phase of each oscillator within a phase-locked cluster is directly proportional to its natural frequency relative to the cluster-average frequency. The associated constant of proportionality serves as a “collective coordinate,” which enables an expression involving all phase variables in terms of the cluster-average phase and the collective coordinate. Accordingly, we describe how to simplify the above equations based on conditions for a collective coordinate ansatz to be appropriate.

Condition 1. Modularity of the coupling matrix. The coupling matrix K is such that $K_{ij} = K_{\sigma\sigma'}$ whenever $i \in P_\sigma$ and $j \in P_{\sigma'}$.

Through Condition 1, we require that the coupling strength between any two oscillators is a function only of their module memberships. Importantly, Condition 1 is preserved under refinement; if Π_1 refines Π_2 and Π_2 satisfies Condition 1, then Π_1 also satisfies Condition 1. It is also true that there is a unique coarsest partition Π^{struct} that satisfies Condition 1, and any other partition satisfying Condition 1 refines Π^{struct} ; to see this, note that Π^{struct} can be constructed as the set of equivalence classes under the relation $i \equiv j \iff K_{i\cdot} = K_{j\cdot} \wedge K_{\cdot i} = K_{\cdot j}$, i.e., two nodes are equivalent if and only if their in- and out-neighborhoods are identical. We refer to Π^{struct} as the *structural* partition. Note that this definition is more stringent than other partitions that have been used elsewhere, such as (external) equitable partitions [38] or orbit partitions induced by the automorphism group of the coupling matrix [39,40]. Condition 1 greatly simplifies several of the coefficients; we have

$$A_\sigma = \langle i\omega_i x_i \rangle_{i \in P_\sigma} \quad (15)$$

$$B_\sigma = \langle i\omega_i \rangle_{i \in P_\sigma} \quad (16)$$

$$D_\sigma = 0 \quad (17)$$

$$E_{\sigma\sigma'} = K_{\sigma\sigma'} |P_{\sigma'}| \quad (18)$$

$$F_{\sigma\sigma'} = -K_{\sigma\sigma'} |P_{\sigma'}| \langle x_i^2 \rangle_{i \in P_{\sigma'}} \quad (19)$$

$$G_{\sigma\sigma'} = 0. \quad (20)$$

Notice that the dependence on the fine-grained state of the system now appears only in Eqs. (15) and (19), so with an appropriate ansatz for the $\{x_i\}$ in terms of $\{z_\sigma\}$ and parameters, a closed-form system can be obtained.

Condition 2. Phase cohesiveness of clusters. Each cluster is phase cohesive, meaning that $|\theta_i(t) - \theta_j(t)|$ remains bounded for all t , whenever oscillators i and j belong to the same cluster.

As before, Condition 2 is a property preserved under refinement and the set of partitions that satisfy it can be defined by a coarsest partition Π^{dyn} , the *dynamical* partition. Any partition satisfying Condition 2 is a refinement of Π^{dyn} , and one can obtain Π^{dyn} as the set of equivalence classes under the equivalence relation

$$i \equiv j \iff \lim_{T \rightarrow \infty} \frac{\theta_i(T) - \theta_i(0)}{T} = \lim_{T \rightarrow \infty} \frac{\theta_j(T) - \theta_j(0)}{T}. \quad (21)$$

This enables the following approximation, which underpins the present derivation.

Approximation 1 Linear collective coordinate ansatz. We take a simple collective coordinate ansatz, namely, that the phase residual of each oscillator around its cluster mean is directly proportional to its frequency residual about its cluster mean. Formally, we assume

$$\theta_i = \arg(z_\sigma) + \alpha_\sigma \tilde{\omega}_i, \quad (22)$$

where $\tilde{\omega}_i = \omega_i - \langle \omega_i \rangle_{i \in P_\sigma}$ is the residual of oscillator i 's natural frequency relative to the mean in its cluster and α_σ is a cluster-dependent constant of proportionality, to be determined.

Note that this is only one choice of many; another natural choice is that the phase residual depends not linearly on the frequency residual but on its arcsine. We also note that the Conditions 1 and 2 are necessary for such an ansatz to be appropriate, since under these Conditions the only remaining distinction between oscillators within a cluster is in their natural frequencies. Thus we may reasonably expect that the dynamics of different oscillators in a cluster differ in a regular way.

The proportionality constant α_σ must be such that $\langle y_i \rangle_{i \in P_\sigma} = z_\sigma$, and it is this constraint that lets us finally obtain an autonomous ODE for $\{z_\sigma\}$. This leads to a consistency condition that reads

$$\langle y_i \rangle_{i \in P_\sigma} = \frac{z_\sigma}{|z_\sigma|} \chi_\sigma(\alpha_\sigma), \quad (23)$$

where χ_σ is the characteristic function of the distribution of natural frequency residuals $\tilde{\omega}_i = \omega_i - \langle \omega_i \rangle_{i \in P_\sigma}$ within cluster σ . This condition gives the value of α_σ implicitly in terms of z_σ and thereby lets us describe the entire distribution of phases within a given cluster—and hence the expectations in Eqs. (15) and (19)—in terms of the single variable z_σ . The general formulas are

$$\begin{aligned} \langle i\omega_i x_i \rangle_{i \in P_\sigma} &= \frac{z_\sigma}{|z_\sigma|} \chi'_\sigma(\chi_\sigma^{-1}(|z_\sigma|)) \\ \langle x_i^2 \rangle_{i \in P_\sigma} &= z_\sigma^2 \left(\frac{1}{|z_\sigma|^2} \chi_\sigma(2\chi_\sigma^{-1}(|z_\sigma|) - 1) \right). \end{aligned} \quad (24)$$

If, for example, the natural frequency residuals are distributed according to a Gaussian distribution with variance v_σ , then $\chi_\sigma(\alpha_\sigma) = \exp(-v_\sigma \alpha_\sigma^2/2)$, and we have

$$\langle i\omega_i x_i \rangle = -z_\sigma (\sqrt{-2v_\sigma \ln(|z_\sigma|)}) \quad (25)$$

$$\langle x_i^2 \rangle = z_\sigma^2 (|z_\sigma|^2 - 1). \quad (26)$$

Substituting the above expressions into Eq. (8) yields a closed system of equations for the dynamics of the coarse-grained variables, whose parameters are directly computable from the parameters of the fine-grained system.

Closure Model 1. Gaussian collective coordinate (GCC) functional form.

$$\begin{aligned} \dot{z}_\sigma &= (\langle i\omega_i \rangle_{i \in P_\sigma} - \sqrt{-2v_\sigma \ln(|z_\sigma|)}) z_\sigma \\ &+ \sum_{\sigma'=1}^C |P_{\sigma'}| K_{\sigma\sigma'} (z_{\sigma'} - z_{\sigma'}^* z_\sigma^2 |z_\sigma|^2), \end{aligned} \quad (27)$$

is the first class of closed reduced-order system of equations considered here. It was derived by substituting the closure approximations (25) and (26) into Eq. (8). We refer to the functional form on the right-hand side of Eq. (27) as the Gaussian collective coordinate (GCC) form.

On the other hand, it is reasonable to suppose that the equations (7) derived by Ott and Antonsen, proven to be valid in the $N \rightarrow \infty$ limit, are still approximately valid for small N , or at least are worthy of consideration as a candidate model class.

Closure Model 2. Cauchy Ott-Antonsen (COA) functional form.

$$\dot{z}_\sigma = (\langle i\omega_i \rangle_{i \in P_\sigma} - \delta_\sigma) z_\sigma + \sum_{\sigma'=1}^C |P_{\sigma'}| K_{\sigma\sigma'} (z_{\sigma'} - z_{\sigma'}^* z_\sigma^2). \quad (28)$$

is the second class of closed reduced-order system of equations considered here. It is derived by applying the Ott-Antonsen ansatz to the infinite- N limit of Eq. (5) under the assumption of Cauchy-distributed natural frequencies within each module. We refer to the functional form on the right-hand side of Eq. (28) as the Cauchy Ott-Antonsen (COA) form.

It is in fact possible to formally derive Eq. (28) by applying the formulas (24) under the assumption that natural frequencies in each cluster σ are Cauchy distributed with location Ω_σ and scale δ_σ . In this case we have $\chi_\sigma(t) = \exp(-\delta_\sigma |t|)$, so

$$\langle i\omega_i x_i \rangle = -\delta_\sigma z_\sigma \quad (29)$$

$$\langle x_i^2 \rangle = 0. \quad (30)$$

Note, however, that strictly speaking, the collective coordinate approximation (22) is not consistent in this case because it models the phases of oscillators with large natural frequencies to be a large constant greater than the mean phase and implies that their instantaneous frequency is the mean frequency, but this is not the case.

Since in general we consider small numbers of oscillators, it is not necessarily true that the set of natural frequencies

within any given cluster is representative of a Gaussian- or Cauchy-distributed population. To circumvent this, we approximate the characteristic function χ_σ as a Taylor series, and show that when α is small, $\chi_\sigma(\alpha)$ is determined (to a good approximation) by the first and second central moments of the empirical distribution of natural frequencies within a cluster. We have

$$\chi_\sigma(\alpha) = \langle \exp(i\alpha\tilde{\omega}_i) \rangle_{i \in P_\sigma} \tag{31}$$

$$= \left\langle \sum_{k=0}^{\infty} \frac{(i\alpha\tilde{\omega}_i)^k}{k!} \right\rangle_{i \in P_\sigma} \tag{32}$$

$$= \sum_{k=0}^{\infty} \frac{(i\alpha)^k}{k!} \langle \tilde{\omega}_i^k \rangle_{i \in P_\sigma} \tag{33}$$

$$= 1 + i\alpha \langle \tilde{\omega}_i \rangle_{i \in P_\sigma} - \frac{\alpha^2}{2} \langle \tilde{\omega}_i^2 \rangle_{i \in P_\sigma} + \mathcal{O}(\alpha^3). \tag{34}$$

Note that the relevant values of α are necessarily such that the residual phases $\alpha\tilde{\omega}_i$ are not too big and so are inversely related with the largest values of $\tilde{\omega}_i$. Finally, we must impose the consistency condition (23). Assuming that terms of $\mathcal{O}(\alpha^3)$ are negligible and that $\langle \tilde{\omega}_i \rangle = 0$, we can solve the consistency condition $\chi_\sigma(\alpha_\sigma) = |z_\sigma|$ for α_σ to get

$$\alpha_\sigma \approx \sqrt{\frac{2(1 - |z_\sigma|)}{\langle \tilde{\omega}_i^2 \rangle_{i \in P_\sigma}}}. \tag{35}$$

Using this approximate solution to estimate the closure terms in Eq. (24), we obtain

$$\begin{aligned} \dot{z}_\sigma &= \left(\langle i\omega_i \rangle_{i \in P_\sigma} - \frac{1}{|z_\sigma|} \sqrt{2(1 - |z_\sigma|)} \langle \tilde{\omega}_i^2 \rangle \right) z_\sigma \\ &+ \sum_{\sigma'=1}^C |P_{\sigma'}| K_{\sigma\sigma'} \left(z_{\sigma'} - z_{\sigma'}^* z_\sigma^2 \frac{4|z_\sigma| - 3}{|z_\sigma|^2} \right), \end{aligned} \tag{36}$$

which approximates Eq. (27) when $\{|z_\sigma|\}$ are near 1.

B. Data-driven inference of model parameters

The derivations above make several simplifying assumptions and useful approximations, some of which will not be satisfied in practice. To find a coarse-grained description of a given finite-dimensional system of coupled oscillators we take a hybrid approach, using the above derivations as inductive biases and using data to select coefficient values.

We focus especially on the coarse-grained systems in Eq. (27), i.e., GCC, and Eq. (28), i.e., COA. For each coarse-grained system, the right-hand side is a combination of particular functions of $\{z_\sigma\}$, each with a coefficient that depends somehow on the distribution of $\{\omega_i\}$ and the coupling matrix K . In the absence of the $N \rightarrow \infty$ limit, however, it is not clear that simple summary statistics will give a model that accurately reflects the coarse-grained dynamics of the finite- N system.

We now describe a procedure for inferring the parameters of a coarse-grained model, either of the form in Eqs. (27) or (28), assuming that the coarse-graining partition Π is known. To illustrate what we mean explicitly, we restrict our attention momentarily to Eq. (28). Given a solution $\{\theta_i(t)\}$

of the system (5) in a time-series form, we can obtain a coarse-grained time series $\{z_\sigma(t)\}$ according to Eq. (6). Our hypothesis is that these coarse-grained variables evolve according to an equation of the form Eq. (28). If the hypothesis holds, then it should be possible to recover the appropriate coefficients, which we can interpret as effective natural frequencies and coupling parameters, using a *least-squares regression* based on the coarse-grained data $\{z_\sigma(t)\}$. Observe first that the right-hand side of (28) is a linear combination of terms that can be measured directly from data. To clarify this, we rewrite Eq. (28) as

$$\dot{z}_\sigma = \tilde{\omega}_\sigma z_\sigma + \sum_{\sigma'=1}^C B_{\sigma\sigma'} (z_{\sigma'} - z_{\sigma'}^* z_\sigma^2). \tag{37}$$

Let $\mathbf{z}_\sigma = (z_\sigma(t_1), \dots, z_\sigma(t_n))^T$ and $\dot{\mathbf{z}}_\sigma = (\dot{z}_\sigma(t_1), \dots, \dot{z}_\sigma(t_n))^T$ denote the time series of observations of the coarse-grained variable z_σ and its derivative, respectively. Then the parameters $\mathbf{p}_\sigma = (\tilde{\omega}_\sigma, B_{\sigma 1}, \dots, B_{\sigma C})^T$ can be inferred by solving the following least-squares optimization [41,42]:

$$\underset{\mathbf{p}_\sigma}{\operatorname{argmin}} \|\dot{\mathbf{z}}_\sigma - \mathbf{G}^\sigma \mathbf{p}_\sigma\|^2 \tag{38}$$

$$\text{s.t.} \begin{cases} \Im(\tilde{\omega}_\sigma) \geq 0 \\ \Im(B_{\sigma\sigma'}) = 0, \end{cases} \tag{39}$$

where the $n \times C + 1$ matrix \mathbf{G}^σ contains the values of the terms of which the entries of \mathbf{p}_σ are coefficients, and $\Im(\cdot)$ denotes the imaginary part of the argument. Explicitly, we can write the set of basis functions \mathbf{G}^σ as

$$\mathbf{G}^\sigma = [z_\sigma \mid z_1 - z_1^* z_\sigma^2 \mid \dots \mid z_C - z_C^* z_\sigma^2], \tag{40}$$

where complex conjugate, product, and squaring of the vectors \mathbf{z}_σ are understood to be taken element-wise. Minimizing the squared residual (38) for each $\sigma = 1, \dots, C$ gives a complete set of parameters for the COA model in Eq. (37). The constraints in Eq. (39) ensure that the resulting model leaves the C -fold product of the unit disk invariant, i.e., that $|z_\sigma(t)| \leq 1 \forall \sigma \Rightarrow |z_\sigma(s)| \leq 1 \forall \sigma$ for all $s > t$. This property is desirable since $\{z_\sigma\}$ are averages of unit complex numbers and so should always lie inside the unit disk.

We can likewise perform the same analysis using the inductive bias given by GCC in Eq. (27) by setting the set of basis functions as follows:

$$\mathbf{G}^\sigma = \begin{bmatrix} z_\sigma \\ \sqrt{-2 \ln(|z_\sigma|)} \\ |z_1 - z_1^* z_\sigma^2| |z_\sigma^2| \\ \vdots \\ |z_C - z_C^* z_\sigma^2| |z_\sigma^2| \end{bmatrix}^T. \tag{41}$$

In accordance with previous results [17,37], we expect to see a satisfactory fit if N (the number of oscillators) is very large and the coupling is as described in Eq. (5). The numerical procedure described above lets us assess the suitability of GCC in Eq. (27) and COA in Eq. (37) as coarse-grained models for finite N .

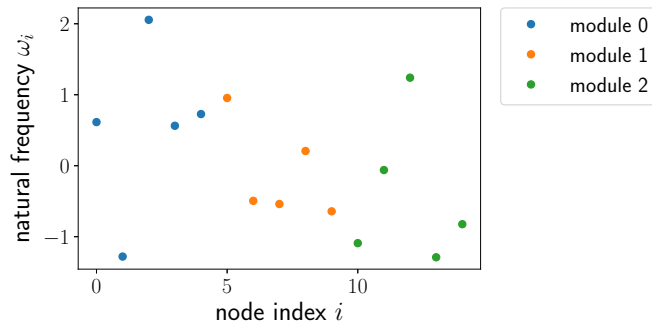


FIG. 1. Natural frequencies used in numerical experiments

IV. RESULTS

We now evaluate the performance of the above-described coarse-grained models by applying them to a range of Kuramoto oscillator systems for which we have fine-grained trajectories. We examine their performance both with theoretically derived coefficient values [as in Eqs. (27) and (28)] as well as those optimized to fit observations. Further, we demonstrate the necessity of both Conditions 1 and 2 by showing that coarse graining in a way that satisfies only one or the other Condition gives a worse model than coarse-graining in a way that satisfies both.

A. Data generation

We perform numerical experiments for a system of $N = 15$ Kuramoto oscillators, i.e., Eq. (1), organized into three modules of five nodes each, and we only consider coupling matrices that respect this organization. This means that we know the *structural partition* Π^{struct} a priori; it is $\{\{1, \dots, 5\}, \{6, \dots, 10\}, \{11, \dots, 15\}\}$. The coupling between nodes is such that each pair of nodes in the same module is coupled with strength $K_{ij} = K_{\text{in}}/N$ and pairs of nodes in different modules are coupled with strength $K_{ij} = K_{\text{out}}/N$. For each oscillator we sample a natural frequency ω_i from a standard normal distribution (i.i.d.) and fix them through all experiments, and their values are shown in Fig. 1. The free parameters in our experiments are $K_{\text{in}} \in [0, 10]$ and $K_{\text{out}} \in [0, 2.5]$; for each parameter we sampled 50 values evenly spaced in the respective interval (inclusive of endpoints). For each pair of $(K_{\text{in}}, K_{\text{out}})$ values, we generate a trajectory of length $T = 1000$ with a time resolution of $\Delta t = 0.01$, starting from uniformly random initial phases.

B. Coarse-grained model selection

For each trajectory we derive a coarse-grained model in each of 12 ($3 \cdot 2 \cdot 2$) different ways, corresponding to three different choices

- (i) Partition: structural [Figs. 5 and 6, (a) and (b)] versus dynamical [Figs. 5 and 6, (e) and (f)] versus meet [Figs. 5 and 6, (c) and (d)]
- (ii) Model class: Cauchy-Ott-Antonsen (COA) [Figs. 5 and 6, (a), (c), and (e)] versus Gaussian-collective coordinate (GCC) [Figs. 5 and 6, (b), (d), and (f)]
- (iii) Method for finding coefficients: theory (Fig. 5) versus inference (Fig. 6)

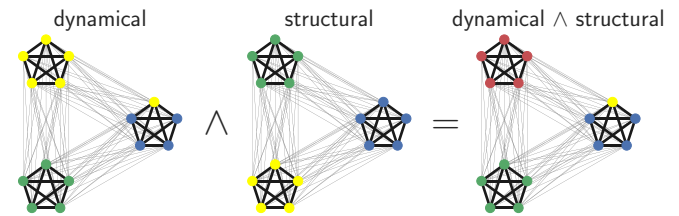


FIG. 2. An example of the meet of the structural and dynamical partitions. In the structural partition, nodes are colored according to the partition that determines the coupling strengths (line thicknesses) between nodes. In the dynamical partition, nodes are colored according to their membership in phase-cohesive groups. The meet (right) is the coarsest partition that refines both the structural and dynamical partitions.

In what follows we clarify the meaning of each of these choices and examine their effect on the quality of coarse-grained model that results.

1. Partition

For each trajectory $\{\theta_i(t)\}$ we must choose a partition to coarse-grain the data. According to Condition 1, the partition should respect the symmetry of the coupling matrix, i.e., it should be a refinement of Π^{struct} , and according to Condition 2 it should be such that all oscillators within a partition element are phase locked, i.e., it should be a refinement of Π^{dyn} . Imposing both conditions means that the partition should refine both Π^{struct} and Π^{dyn} , i.e., it should refine $\Pi^{\text{meet}} := \Pi^{\text{struct}} \wedge \Pi^{\text{dyn}}$ (see Fig. 2). In order to illuminate the importance of each Condition while obtaining the coarsest possible model, we evaluate the performance models coarse-grained according to each of Π^{struct} , Π^{dyn} , and Π^{meet} .

In practice, it is prohibitively expensive to enumerate all partitions of N nodes to find Π^{dyn} that satisfies the condition defined by Eq. (21), so we use a heuristic. Given a trajectory $\{\theta_i(t)\}$, we compute a long-term average frequency $\bar{\omega}_i = (\theta_i(t_0 + T) - \theta_i(t_0)) / (T - t_0)$ and cluster nodes initially by their value of $\bar{\omega}$. We can then check if the resulting clusters are in fact phase cohesive and if they are not, perform a finer clustering of $\bar{\omega}$ and repeat until we have phase cohesive clusters. Other approaches to identifying phase clusters have been considered and could be used in place of the method described above [43–45]. Constructing Π^{meet} from Π^{struct} and Π^{dyn} is straightforward; it is simply the set of all nonempty pairwise intersections of partition elements from Π^{struct} and Π^{dyn} .

2. Model class

Given a partition Π , we coarse-grain the data to obtain $\{z_\sigma(t) | P_\sigma \in \Pi\}$. We now choose what class of model to find, COA or GCC. The COA model is defined by Eq. (28), and the GCC model is defined by Eq. (27).

3. Method for finding coefficients

Finally, we have the choice of using coefficient values predicted from theory (that includes several possibly violated assumptions and approximations) or inferring them from data as described in Sec. III B. Note that the theoretical values

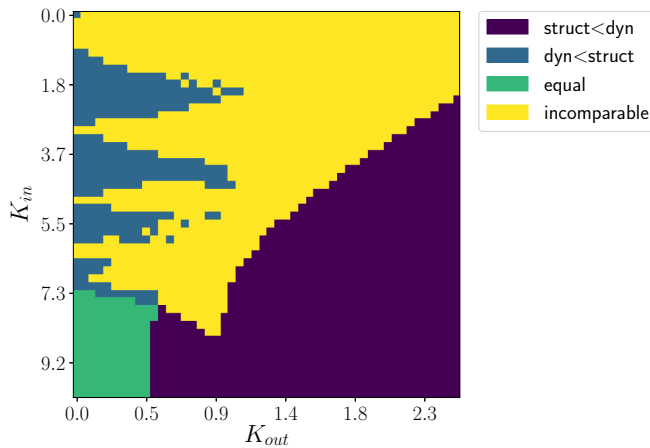


FIG. 3. Parameter space (K_{in} , K_{out}) colored by how the dynamical partition compares to the structural partition. The relation $P < Q$ means that partition P is a refinement of partition Q , and incomparable means that neither $P < Q$ nor $Q < P$. See Fig. 4 for examples of each of these cases.

present in the COA model depend on parameters (Ω_σ , δ_σ) of the Cauchy distributions from which natural frequencies are drawn, which do not actually exist in our case. To circumvent this difficulty we perform a maximum-likelihood inference of these parameters from the data $\{\omega_i | i \in P_\sigma\}$ assuming that they were drawn from a Cauchy distribution (as implemented by `scipy.stats.cauchy.fit` [46]).

Overviews of coarse-grained model quality as a function of coupling parameters (K_{in} , K_{out}) are depicted in Figs. 5 and 6. The metric used is a root-mean-squared error in time derivative, i.e., $\|\hat{f}(z^\Pi) - dz^\Pi/dt\|$, where \hat{f} is the coarse-grained model in question and $\|\cdot\|$ is the root mean square with respect to time t and partition element index σ . Heat-map colors are arranged on a logarithmic scale. Compare to Fig. 3 to relate coarse-grained model performance to characteristics of the dynamical partition relative to the structural partition.

A more systematic set of comparisons illuminating the impact of the three choices (partition, model terms, and method) is given in Figs. 7–10.

C. Numerical results

We now present an overview of the dynamics we simulated and the performance of each of the 12 coarse-grained models described above. While Π^{struct} is known *a priori* and fixed for all experiments, Π^{dyn} changes depending on the values of K_{in} and K_{out} . Figure 3 gives an overview of how Π^{dyn} compares to Π^{struct} as a function of K_{in} and K_{out} . There are four possibilities: either $\Pi^{struct} < \Pi^{dyn}$, $\Pi^{dyn} < \Pi^{struct}$, $\Pi^{struct} = \Pi^{dyn}$, or the two partitions are incomparable (i.e., neither one is a refinement of the other, nor are they equal). Figure 4 depicts an example of each of these four possibilities. As one might expect, the region of parameter space where the two partitions are equal lies where K_{in} is large and K_{out} is small. When both K_{in} and K_{out} are large, global synchronization occurs and thus $\Pi^{struct} < \Pi^{dyn}$ holds. For moderate values of K_{in} and small K_{out} , phase-locked clusters remain restricted to structural clusters and $\Pi^{dyn} < \Pi^{struct}$, while for the remaining parameter space the two partitions are incomparable.

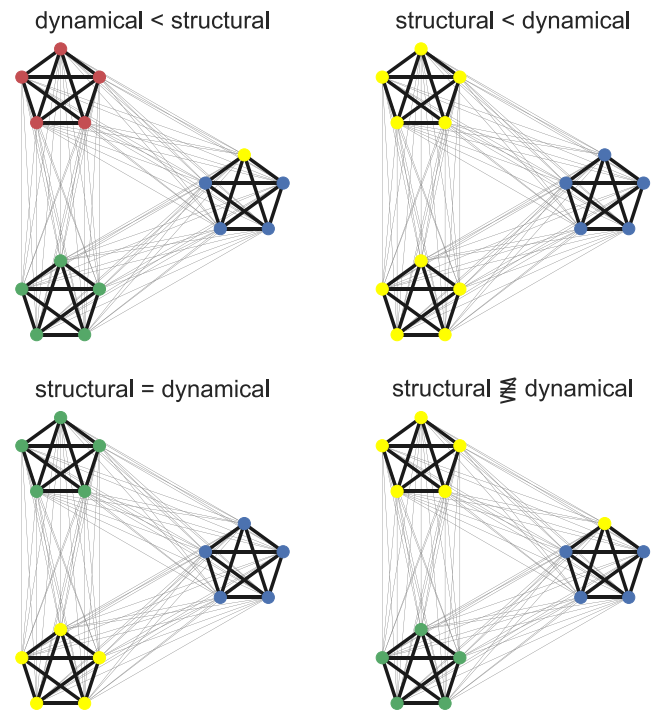


FIG. 4. Depiction of the possible relationships between two partitions. Here the structural partition is the partition into three groups of five nodes indicated by their position and edge thickness, and color indicates the dynamical partition, i.e., nodes of the same color are phase-locked with each other.

1. Theoretically determined coefficients

In Fig. 5 we give an overview of the coarse-grained model quality using theoretically derived coefficient values. We can see that the same subsets of parameter space highlighted in Fig. 3 appear in panels (b), (d), and (f) of Fig. 5, corresponding to the GCC ansatz. Note that applying the GCC ansatz to the meet partition (d) obtains reasonably good performance across the entire parameter space, while under the structural partition (b), good performance is only attained when the structural modules are in fact internally synchronized. Conversely, under the dynamical partition (f), the GCC model performs well mainly in regions of parameter space where the dynamical partition refines or is equal to the structural one (cf. Fig. 3), since in these cases the dynamical partition equals the meet partition. Notice also that when $K_{in} \approx K_{out}$ [red line, panels (e) and (f)], the structural partition is effectively the entire network as a single module, and in this region we also see somewhat better performance of the GCC model. The COA ansatz attains reasonable performance for small values of K_{in} under the meet partition—in this case many dynamical modules are singletons and are therefore trivial to coarse grain. Otherwise, the COA ansatz does not perform very well, indicating that the assumptions underlying its derivation do not hold for the system at hand.

2. Coefficients optimized to fit the data

In Fig. 6 we give an overview of coarse-grained model quality using coefficient values optimized to fit the data. The

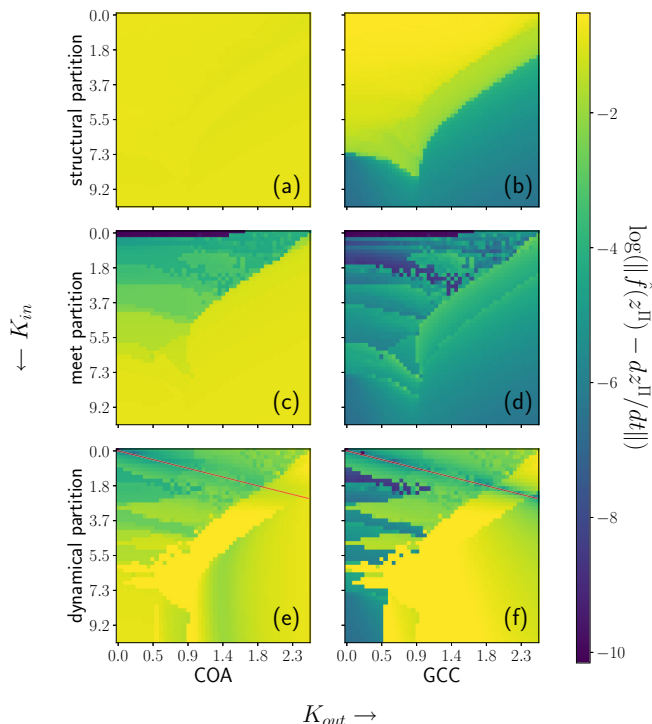


FIG. 5. Coarse-grained model quality as a function of coupling parameters for theoretical coefficient values.

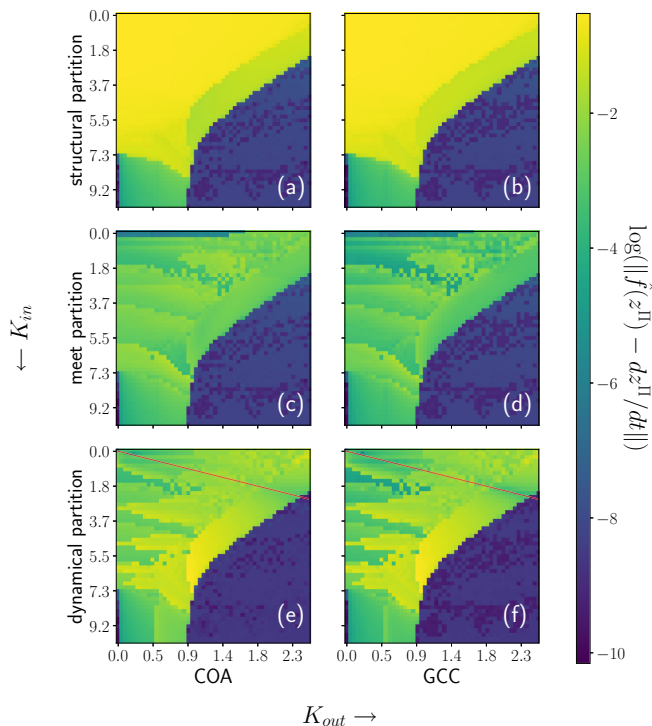


FIG. 6. As in Fig. 5 but for coefficient values optimized to fit the data.

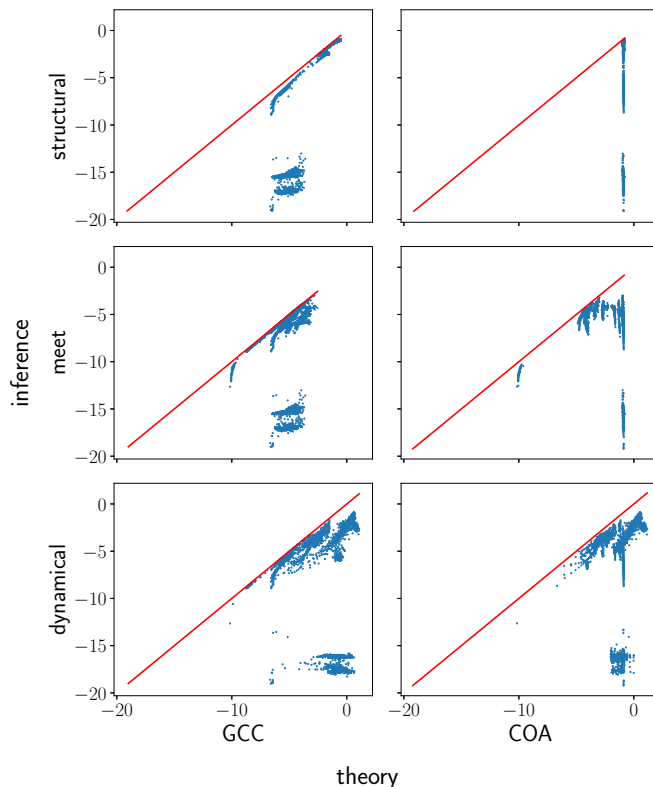


FIG. 7. Log-loss of the coarse-grained model obtained by inferring coefficient values versus using theoretical values. Each point corresponds to a single instance of the coarse-graining problem. In many cases (under the meet partition) the GCC theory performs nearly as well as GCC inference, indicating that the assumptions going into the GCC theory were well supported. Note that COA theory with the structural partition never does particularly well.

most notable contrast with Fig. 5 is that the left and right columns (COA versus GCC) are much more similar to each other in Fig. 6 than in Fig. 5. This indicates that given the possibility of tuning coefficient values, the functional form of the COA ansatz is nearly equivalently expressive to the functional form of the GCC ansatz, relative to the dynamics at hand. Again, note that coarse-graining according to the dynamical (resp. structural) partition attains good performance in regions where it refines the structural (resp. dynamical) partition and hence is equal to the meet partition. However, all partitions are able to achieve good performance in the large region of global synchrony, owing to the simplicity of the resulting dynamics.

3. Quality of inferred models

In Figs. 7–10 we take a more quantitative view on the results just described by forming scatter plots that make pairwise comparisons between coarse-grained models that differ in one design characteristic: partition, model class, and inference method. In each figure, every dot represents one instance of the coarse-graining problem that has been solved in each of 12 ways.

In Fig. 7 we examine the impact of inferring coefficients based on data versus using their theoretical values. We see clearly that in all cases inference affords better performance but that in some cases theoretical coefficients perform nearly

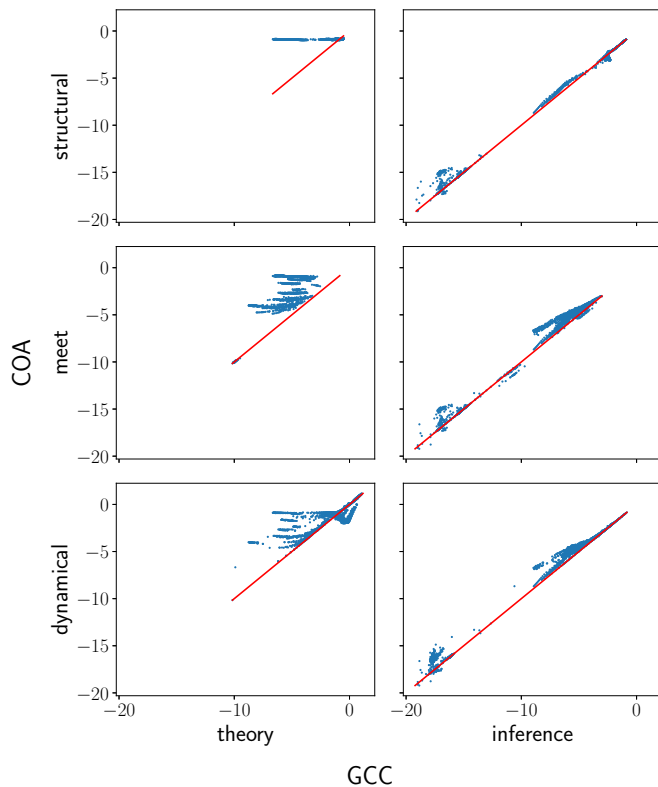


FIG. 8. As in Fig. 7, organized to show the effect of using model terms predicted from GCC instead of COA. Interestingly, when given the freedom to adapt coefficient values to observed data, COA and GCC models perform comparably. This indicates that relative to the behaviors considered in these numerical experiments, the COA and GCC function spaces are somehow comparably expressive.

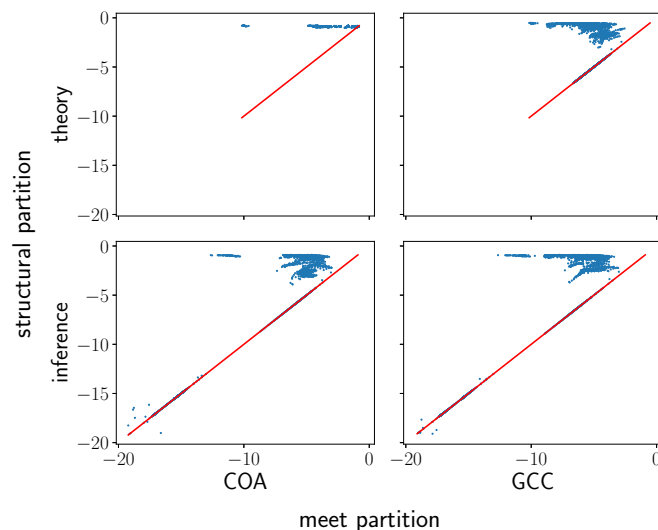


FIG. 9. As in Fig. 7, organized to show the effect of coarse-graining according to the meet partition versus structural partition. Note that the meet partition is never worse than the structural one, and they agree whenever they are equal. This suggests that the added condition of phase cohesiveness within modules (Condition 2) does make a significant difference in one’s ability to find an accurate coarse-grained model.

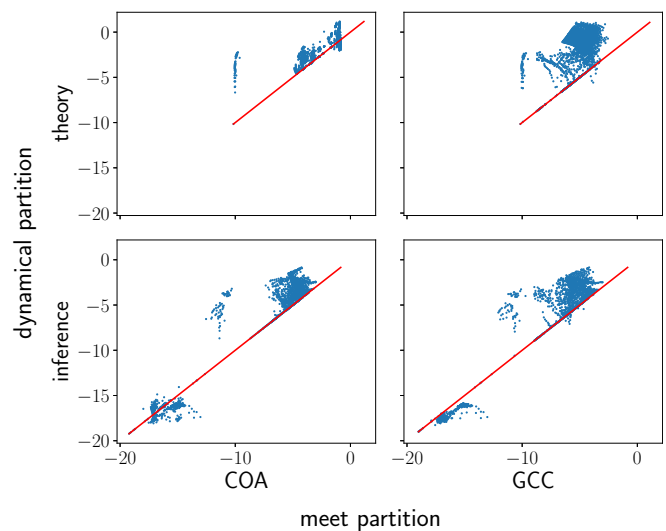


FIG. 10. As in Fig. 7, organized to show the effect of coarse-graining according to the meet partition versus dynamical partition. Note that the meet partition only rarely outperforms the dynamical one, and they agree whenever they are equal. This suggests that the added condition of modularity of coupling strengths (Condition 1) does make a significant difference in one’s ability to find an accurate coarse-grained model.

as well. This is the case for all situations except the COA ansatz using the structural partition, where the model with theoretical coefficients always performs poorly.

In Fig. 8 we examine the impact of using the COA versus the GCC ansatz. As noted qualitatively above, when allowing for inferred coefficient values, the two ansätze attain comparable performance, while GCC is generally better when using theoretical values. This reflects the fact that the two function classes are equivalently expressive for the dynamics at hand, while the assumptions underlying the derivation of GCC are more closely reflective of the system under study than those leading to COA.

In Fig. 9 we investigate the difference in coarse-grained model quality when coarse-graining according to the meet partition versus the structural partition. This amounts to imposing Conditions 1 and 2 versus only imposing Condition 1. Unsurprisingly, the meet partition never gives poorer fit quality than the structural one, but there are cases along the diagonal when the two partitions coincide.

Finally, Fig. 10 compares the meet partition with the dynamical partition. The meet partition generally outperforms the dynamical partition, indicating that Condition 1 is indeed necessary above and beyond Condition 2 to obtain a satisfactory coarse-grained model, even if one is allowed to optimize coefficients to fit the data. This suggests that mere phase cohesiveness is not enough to justify a collective coordinate ansatz such as Eq. (22). Rather, the cluster members should be influenced by other oscillators in the same manner—have the same in-neighborhoods—rendering their natural frequencies the only remaining differences between them to explain their different phases.

D. Discussion and path forward

In this study we introduced and numerically validated a data-driven procedure for discovering the reduced-order equations for the Kuramoto system in the presence of modular synchronization. In particular, we have demonstrated that in a modularly synchronized Kuramoto oscillator system, groups of mutually synchronized oscillators may be treated each as metaoscillators, thereby yielding dynamics that belong to a family that strictly generalizes the Kuramoto model.

A central challenge of the problem considered here is to remove explicit time dependence from the trivially true coarse-grained model (8). Here we accomplished this task by imposing a set of assumptions on the coarse-graining map that justify a collective coordinate ansatz. An alternative approach requiring less advance knowledge of the underlying system would be to search over the space of partitions, either by brute force or greedily, seeking those for which the coefficients (9)–(14) are a single-valued function of the coarse-grained state $\{z_\sigma\}$. If such a function exists, its substitution into Eq. (8) yields an autonomous ODE for the coarse variables.

Stripping away yet another layer of theory, we can remain largely agnostic about the functional form of the equations that should govern \dot{z} and obtain the set basis functions using a procedure such as SINDY [41] or use black-box representations such as neural networks to discover a parsimonious basis. Still, given the enormous complexity of the space of functions $f: \mathbb{C}^C \rightarrow \mathbb{C}^C$, it is prudent to impose certain domain-informed restrictions on the function dictionary we choose.

First, the OA ansatz and its associated dimension reduction shows that the reduced model exhibits on-site and pairwise terms and no higher-order (i.e., three-or-more-particle) terms. Moreover, each coupling term is of the same form. Therefore it is reasonable to suppose that our function basis should contain a set of on-site terms and a set of coupling terms and that these terms should be replicated for each node and edge, respectively.

Next, it is known that coarse graining of dynamics can induce memory effects [47], so it is reasonable to expect that a coarse-grained model including memory or nonlocality (in the form of polyadic coupling terms) would be effective. Data-driven discovery of coarse-grained dynamical models including memory is discussed in [48]. Polyadic coupling terms make no conceptual difference, but combinatorial explosion may make their inclusion computationally costly. A key potential advantage to including memory is the possibility of obtaining a good coarse-grained model with fewer state variables (i.e., a coarser partition).

Finally, we remark on some physical principles that inform our methods for inferring coarse-grained dynamical models. Notice that in Eq. (28) the coefficient of the linear term has a negative real part (because $\delta_\sigma \geq 0$ is the width of a Cauchy distribution and so cannot be negative) and the coupling matrix $K_{\sigma\sigma'}$ is real. These conditions together imply that the C -fold product of the unit disk is invariant; in other words, if all complex phases z_σ initially have magnitude not greater than 1, then they will continue to have magnitude not greater than 1 for all time. This condition breaks down if we allow the coupling matrix to have an imaginary part or the coefficient of

the linear term to have positive real part, and so we impose those constraints during our inference procedure.

Constraints of this type can be thought of in a more general context. We know that the coarse-grained variables z_σ should always remain bounded in the unit disk because they are averages of quantities on the unit circle. Knowing this, we perform *constrained* inference of the right-hand side of an ODE governing those variables, with the constraint being that a function is feasible only if the corresponding ODE leaves the appropriate set in state space invariant. Fortunately, this constraint is convex and is therefore straightforward to impose on the inference procedure.

V. CONCLUSIONS

In this study we present a data-driven and theoretically grounded approach to constructing coarse-grained models for finite-dimensional Kuramoto oscillator systems. To do this we build on prior work that leverages the infinite- N limit, as well as collective coordinate methods that offer the possibility of closing the explicit time-dependence induced by coarse graining. We complement these theoretical approaches with a data-driven inference procedure that selects both the appropriate coarse-graining partition and optimal coefficient values for a given finite Kuramoto system. Our aim is to demonstrate the utility of combining theoretical insight with data-driven inference for finding accurate coarse-grained models of finite heterogeneous systems.

Our results reveal that partitioning nodes in a way that respects the underlying coupling network is vital to the validity of a collective coordinate approach, above and beyond phase cohesiveness. In addition, we find that given the freedom to optimize coefficient values based on data, the functional form predicted by Ott and Antonsen based on the assumption of Cauchy-distributed natural frequencies performs very comparably with the functional form predicted by the more realistic (for the case considered here) assumptions of Gaussian-distributed natural frequencies and a linear collective coordinate ansatz. This shows that there is considerable freedom in learning the right-hand side of a coarse-grained ODE and that other criteria could be brought into play to disambiguate between comparable models. In the present work the derivations given in Sec. III perform this function in their role as inductive biases; other criteria such as parsimony [41] or interpretability might also be applied.

Notably, our method achieves accurate coarse-grained model performance across the whole parameter space we considered, which includes behaviors ranging from complete incoherence to global synchronization. The number of samples required to learn the coarse-grained model can vary according to the state of the system and may drastically change when the system undergoes an abrupt phase transition [49,50]. In future work, it would be interesting to investigate the change of model complexity under phase transitions in original and coarse-grained systems.

Finally, we discuss possible extensions to the work presented here. The derivations presented in Sec. III A rely on partitioning nodes according to mathematical regularities in the underlying equations of motion, in particular regarding the coupling matrix. A similar approach could be applied in

the case that the underlying system exhibits regularities in other properties, such as phase lags or higher harmonics in the coupling function, or in the on-site dynamics. We anticipate that the approach introduced in this work will find applications in other types of complex dynamical systems.

ACKNOWLEDGMENTS

This study was carried out at Los Alamos National Laboratory under the auspices of the National Nuclear Security Administration of the U.S. Department of Energy under Contract No. 89233218CNA000001. J.S. acknowledges sup-

port from Air Force Office of Scientific Research MURI FA9550-19-1-0386, U.S. Army Research Office MURI Award No. W911NF-13-1-0340 and Cooperative Agreement No. W911NF-09-2-0053, and the U.S. Department of Defense, Minerva Grant No. W911NF-15-1-00502. A.Z. acknowledges support from the U.S. Department of Energy National Nuclear Security Administration's Office of Defense Nuclear Nonproliferation Research & Development. A.Y.L. acknowledges support from the Laboratory Directed Research and Development program of Los Alamos National Laboratory under Projects No. 20190059DR, No. 20200121ER, and No. 20210078DR. This manuscript has been approved for unlimited release with control number LA-UR-20-27371.

-
- [1] O. Reynolds, IV, On the dynamical theory of incompressible viscous fluids and the determination of the criterion, *Philos. Trans. R. Soc. London A* **186**, 123 (1895).
- [2] P. Y. Chou, On velocity correlations and the solutions of the equations of turbulent fluctuation, *Q. Appl. Math.* **3**, 38 (1945).
- [3] S. B. Pope, A more general effective-viscosity hypothesis, *J. Fluid Mech.* **72**, 331 (1975).
- [4] K. Duraisamy, G. Iaccarino, and H. Xiao, Turbulence modeling in the age of data, *Annu. Rev. Fluid Mech.* **51**, 357 (2019).
- [5] T. T. Foley, M. S. Shell, and W. G. Noid, The impact of resolution upon entropy and information in coarse-grained models, *J. Chem. Phys.* **143**, 243104 (2015).
- [6] L. Boninsegni, R. Banisch, and C. Clementi, A data-driven perspective on the hierarchical assembly of molecular structures, *J. Chem. Theory Comput.* **14**, 453 (2018).
- [7] J. Wang, S. Olsson, C. Wehmeyer, A. Pérez, N. E. Charron, G. de Fabritiis, F. Noé, and C. Clementi, Machine learning of coarse-grained molecular dynamics force fields, *ACS Cent. Sci.* **5**, 755 (2019).
- [8] J. P. Gleeson, S. Melnik, J. A. Ward, M. A. Porter, and P. J. Mucha, Accuracy of mean-field theory for dynamics on real-world networks, *Phys. Rev. E* **85**, 026106 (2012).
- [9] M. A. Porter and J. Gleeson, *Dynamical Systems on Networks*, Frontiers in Applied Dynamical Systems: Reviews and Tutorials Vol. 4 (Springer International Publishing, Cham, 2016).
- [10] Y. Kuramoto, Self-entrainment of a population of coupled nonlinear oscillators, in *International Symposium on Mathematical Problems in Theoretical Physics* (Springer, Berlin, 1975), Vol. 39, p. 420.
- [11] J. Aschoff and R. Wever, The circadian system of man, in *Biological Rhythms* (Springer, New York, 1981), pp. 311–331.
- [12] S. H. Strogatz, C. M. Marcus, R. M. Westervelt, and R. E. Mirollo, Collective dynamics of coupled oscillators with random pinning, *Physica D* **36**, 23 (1989).
- [13] T. D. Frank, A. Daffertshofer, C. E. Peper, P. J. Beek, and H. Haken, Towards a comprehensive theory of brain activity: Coupled oscillator systems under external forces, *Physica D* **144**, 62 (2000).
- [14] A. Arenas, A. Díaz-Guilera, and C. J. Pérez-Vicente, Synchronization Reveals Topological Scales in Complex Networks, *Phys. Rev. Lett.* **96**, 114102 (2006).
- [15] D. Li, I. Leyva, J. A. Almendral, I. Sendiña-Nadal, J. M. Buldú, S. Havlin, and S. Boccaletti, Synchronization Interfaces and Overlapping Communities in Complex Networks, *Phys. Rev. Lett.* **101**, 168701 (2008).
- [16] S. H. Strogatz, From Kuramoto to Crawford: Exploring the onset of synchronization in populations of coupled oscillators, *Physica D* **143**, 1 (2000).
- [17] E. Ott and T. M. Antonsen, Low dimensional behavior of large systems of globally coupled oscillators, *Chaos* **18**, 037113 (2008).
- [18] M. J. Panaggio and D. M. Abrams, Chimera states: Coexistence of coherence and incoherence in networks of coupled oscillators, *Nonlinearity* **28**, R67 (2015).
- [19] C. Bick, M. J. Panaggio, and E. A. Martens, Chaos in Kuramoto oscillator networks, *Chaos* **28**, 071102 (2018).
- [20] Y. Zhang, Z. G. Nicolaou, J. D. Hart, R. Roy, and A. E. Motter, Critical Switching in Globally Attractive Chimeras, *Phys. Rev. X* **10**, 011044 (2020).
- [21] K. Wiesenfeld, P. Colet, and S. H. Strogatz, Frequency locking in Josephson arrays: Connection with the Kuramoto model, *Phys. Rev. E* **57**, 1563 (1998).
- [22] F. Varela, J.-P. Lachaux, E. Rodriguez, and J. Martinerie, The brainweb: Phase synchronization and large-scale integration, *Nat. Rev. Neurosci.* **2**, 229 (2001).
- [23] I. Z. Kiss, Y. Zhai, and J. L. Hudson, Emerging coherence in a population of chemical oscillators, *Science* **296**, 1676 (2002).
- [24] A. Zlotnik, R. Nagao, I. Z. Kiss, and J.-S. Li, Phase-selective entrainment of nonlinear oscillator ensembles, *Nat. Commun.* **7**, 10788 (2016).
- [25] F. Dörfler, M. Chertkov, and F. Bullo, Synchronization in complex oscillator networks and smart grids, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 2005 (2013).
- [26] D. Gfeller and P. De Los Rios, Spectral Coarse Graining and Synchronization in Oscillator Networks, *Phys. Rev. Lett.* **100**, 174104 (2008).
- [27] Y. Izumida and H. Kori, Coarse-grained description of general oscillator networks, [arXiv:1311.0917](https://arxiv.org/abs/1311.0917).
- [28] S. J. Moon, R. Ghanem, and I. G. Kevrekidis, Coarse Graining the Dynamics of Coupled Oscillators, *Phys. Rev. Lett.* **96**, 144101 (2006).
- [29] K. Rajendran and I. G. Kevrekidis, Coarse graining the dynamics of heterogeneous oscillators in networks with spectral gaps, *Phys. Rev. E* **84**, 036708 (2011).

- [30] T. N. Thiem, M. Kooshkbaghi, T. Bertalan, C. R. Laing, and I. G. Kevrekidis, Emergent spaces for coupled oscillators, *Front. Comput. Neurosci.* **14**, 36 (2020).
- [31] G. A. Gottwald, Model reduction for networks of coupled oscillators, *Chaos* **25**, 053111 (2015).
- [32] E. J. Hancock and G. A. Gottwald, Model reduction for Kuramoto models with complex topologies, *Phys. Rev. E* **98**, 012307 (2018).
- [33] Lachlan D. Smith and G. A. Gottwald, Chaos in networks of coupled oscillators with multimodal natural frequency distributions, *Chaos* **29**, 093127 (2019).
- [34] J. R. Engelbrecht and R. Mirollo, Is the Ott-Antonsen manifold attracting? *Phys. Rev. Res.* **2**, 023057 (2020).
- [35] W. Yue, L. D. Smith, and G. A. Gottwald, Model reduction for the Kuramoto-Sakaguchi model: The importance of non-entrained rogue oscillators, *Phys. Rev. E* **101**, 062213 (2020).
- [36] Y. Kuramoto, Cooperative dynamics of oscillator community, *Prog. Theor. Phys. Suppl.* **79**, 223 (1984).
- [37] P. S. Skardal and J. G. Restrepo, Hierarchical synchrony of phase oscillators in modular networks, *Phys. Rev. E* **85**, 016208 (2012).
- [38] M. T. Schaub, N. O'Clery, Y. N. Billeh, J.-C. Delvenne, R. Lambiotte, and M. Barahona, Graph partitions and cluster synchronization in networks of oscillators, *Chaos* **26**, 094821 (2016).
- [39] L. M. Pecora, F. Sorrentino, A. M. Hagerstrom, T. E. Murphy, and R. Roy, Cluster synchronization and isolated desynchronization in complex networks with symmetries, *Nat. Commun.* **5**, 4079 (2014).
- [40] F. Sorrentino, L. M. Pecora, A. M. Hagerstrom, T. E. Murphy, and R. Roy, Complete characterization of the stability of cluster synchronization in complex dynamical networks, *Sci. Adv.* **2**, e1501737 (2016).
- [41] S. L. Brunton, J. L. Proctor, and J. N. Kutz, Discovering governing equations from data by sparse identification of nonlinear dynamical systems, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 3932 (2016).
- [42] A. Y. Lokhov, M. Vuffray, D. Shemetov, D. Deka, and M. Chertkov, Online learning of power transmission dynamics, in *2018 Power Systems Computation Conference (PSCC)* (IEEE, New York, 2018), pp. 1–7.
- [43] S. Bialonski and K. Lehnertz, Identifying phase synchronization clusters in spatially extended dynamical systems, *Phys. Rev. E* **74**, 051909 (2006).
- [44] A. Groth and M. Ghil, Multivariate singular spectrum analysis and the road to phase synchronization, *Phys. Rev. E* **84**, 036206 (2011).
- [45] C. Allefeld, M. Müller, and J. Kurths, Eigenvalue decomposition as a generalized synchronization cluster analysis, *Int. J. Bifurcation Chaos* **17**, 3493 (2007).
- [46] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey *et al.*, SciPy 1.0: Fundamental algorithms for scientific computing in Python, *Nat. Methods* **17**, 261 (2020).
- [47] R. Zwanzig, Memory effects in irreversible thermodynamics, *Phys. Rev.* **124**, 983 (1961).
- [48] S. Pan and K. Duraisamy, Data-driven discovery of closure models, *SIAM J. Appl. Dyn. Syst.* **17**, 2381 (2018).
- [49] F. Giannuzzi, D. Marinazzo, G. Nardulli, M. Pellicoro, and S. Stramaglia, Phase diagram of a generalized Winfree model, *Phys. Rev. E* **75**, 051104 (2007).
- [50] J. Gómez-Gardeñes, S. Gómez, A. Arenas, and Y. Moreno, Explosive Synchronization Transitions in Scale-Free Networks, *Phys. Rev. Lett.* **106**, 128701 (2011).