The Reservoir's Perspective on Generalized Synchronization

Thomas Lymburn,[1, a)] David M. Walker,[1] Michael Small,[1, 2] and Thomas Jüngling[1]

[1)]*Complex Systems Group, Department of Mathematics and Statistics, The University of Western Australia, Crawley, Western Australia 6009, Australia*

[2)]*Mineral Resources, CSIRO, Kensington, Western Australia 6151, Australia*

We employ reservoir computing for a reconstruction task in coupled chaotic systems, across a range of dynamical relationships including generalized synchronization. For a drive-response setup, a temporal representation of the synchronized state is discussed as an alternative to the known instantaneous form. The reservoir has access to both representations through its fading memory property, each with advantages in different dynamical regimes. We also extract signatures of the maximal conditional Lyapunov exponent in the performance of variations of the reservoir topology. Moreover, the reservoir model reproduces different levels of consistency where there is no synchronization. In a bidirectional coupling setup, high reconstruction accuracy is achieved despite poor observability and independent of generalized synchronization.

Generalized synchronization is a nonlinear functional relationship between the states of two or more coupled dynamical systems. This functional dependence can have a complicated shape, even for simple systems. Moreover, the presence or absence of generalized synchronization forms only part of a much richer phenomenology. We use reservoir computing to model general dynamical relationships between coupled chaotic systems. The method itself relies on the synchronized response of a large dynamical system – the reservoir – to an input signal. We observe high modelling accuracy up to fundamental limits in a wide range of different dynamical regimes. The key to this performance is the fading memory property of the reservoir, which allows it to access a simplified representation of the complex relationships. We show through different coupling configurations how the reservoir-computing scheme reconstructs the interacting systems effectively, whereas the classification of synchronization plays a minor role.

I. INTRODUCTION

Synchronization refers to the entrainment of the states of two or more coupled systems such that they evolve together with time[1]. Several forms of synchronization have been studied, ranging from phase synchronization (PS) to complete synchronization (CS), the latter in which the states of the interacting systems are identical after transients[2,3]. Here, we address generalized synchronization (GS), which can be considered an extension of CS to non-identical systems[4]. In its original definition, GS is characterized by a nonlinear function $\mathbf{H}(\mathbf{x}(t), \mathbf{y}(t)) = 0$ linking the states $\mathbf{x}(t)$, $\mathbf{y}(t)$ of the interacting systems.

This function defines the GS manifold, which may be a very complicated subset of phase space. Systems which appear unrelated may in fact be in a state of GS[5]. One distinguishes between weak GS, where the manifold is fractal, and strong GS with a smooth manifold[6]. The distinction can be made based on the Lyapunov spectrum. Due to the presence of a potentially complicated functional relationship, the detection of GS is a nontrivial task, and specific methods are required[7]. These include the auxiliary system method[8], nearest-neighbour methods[4,9] and extracting the conditional Lyapunov exponents (CLE)[10,11]. Alternatively, one can directly estimate the functional dependence $\mathbf{H}[\cdot]$, and thus infer the existence of GS[12]. The method of choice depends on experimental accessibility and the amount and type of available recordings, amongst other factors[13].

The understanding of GS has now moved beyond the original concept, to include phenomena like common-signal-induced synchronization[14,15], strong and weak chaos in delay systems[16], consistency and reliability[17,18], or a general response theory of dynamical systems[19,20]. The common theme is a type of coupling-induced interdependence of dynamical states, within a range of shapes and degrees, even though it often cannot be expressed as a function between state space variables. We aim here to unify the broad spectrum of dynamical dependencies in coupled nonlinear systems with the concept of reconstruction[21–23]. Specifically for GS, such reconstruction was shown in Ref.[12], where Volterra kernels are used to model GS manifolds for several systems. We extend these ideas here by studying a type of statistical modelling of the (functional) relationships between chaotic systems, based on echo state networks (ESN).

Echo state networks form one of the roots of the present field of reservoir computing (RC)[24,25]. They are a type of recurrent neural networks (RNN) with simplified training rules, and as such belong to artificial neural networks[26–28]. Reservoir computing is a computational paradigm whereby a large nonlinear dynamical system – the reservoir – acts as a computational substrate for a driving input signal. The output is trained by linear regression or classification methods to approximate a given

[a)]Electronic mail: thomas.lymburn@research.uwa.edu.au

target signal. As such, the training algorithm can be considered a truncated backpropagation method[29–31]. The basic principles of RC have been laid out in its second root in neuroscience, where the concept of a liquid-state machine was introduced[32]. It is now widely accepted that several hard and soft criteria need to be satisfied by the reservoir to perform its function. These include fading memory[33,34] and consistency[35–37], amongst others. A generic practical guide to setting up ESNs is given in Ref.[38], but design principles may vary strongly among tasks. In particular, a wide range of successful physical reservoir implementations demonstrates the power of the concept[25,31,39–46]. Theoretical research in the field is devoted to the identification of key parameters and construction principles which predict the performance and improve general reservoir design[47–49]. Although nonlinear time series tasks have always been a benchmark in the field[26], reservoir computing has more recently been studied in greater depth on tasks such as prediction and reconstruction of nonlinear dynamics and chaos[21–23,50–56].

In this work, we use ESNs for reconstruction tasks in some low-dimensional coupled chaotic systems, with a focus on the properties of GS states. Previously, the detection of GS has been addressed by using ESNs as a binary classifier to distinguish between synchronous and asynchronous states[57]. In contrast, for the reconstruction task an observable of one system serves as input for the ESN, and the readout is trained to approximate a variable of the other system. Our study of this scheme has a twofold aim. First, RC is itself based on GS, what is usually referred to as the echo state property. In this sense, the task to model GS is a natural choice, and we elaborate on whether this analogy is of advantage. By selected experimentation with the coupled systems, as well as with the properties of the reservoir, we aim to learn both the nature of the GS state and the function of the reservoir itself. We give a particular focus on the role of fading memory, which is inherent to GS and also essential for the operation of RC. Second, the RC scheme resembles a naive approach to the functional relationships which is not biased by the categories derived from the Lyapunov spectrum. The reservoir model synthesizes its output based on correlations, in the broadest sense (see e.g. Ref.[58]), in order to minimize the error to the target signal. The result is the reproduction of a continuum of entrainments, including limited levels of consistency as an extended form of GS[17,19,59,60]. The reconstruction performance can thus be interpreted as a measure of 'generalized correlations', forming a significant part of the phenomenology of coupled systems.

In Section II we explore two different representations of GS, then outline our approach, and define the RC scheme with ESN. In Section III we consider reconstruction in a unidirectionally coupled Rössler-Lorenz system. In Section III B we investigate the relationship between the timescale of the GS function and the ESN memory capacity. In Section IV we consider bidirectional coupling in a pair of slightly dissimilar Rössler systems.

II. MODELLING DYNAMICAL RELATIONSHIPS

We first consider two complementary representations of GS, as well as a simplified description of the generic shape of the functional dependence, which is relevant for understanding the performance of statistical models with memory. We then introduce the RC scheme with ESN which we use throughout this work.

A. Different perspectives on generalized synchronization

In order to illustrate the idea of different representations of a functional dependency between dynamical systems, we first consider an M-dimensional chaotic map $\mathcal{X}$ driving a low-pass filter $\mathcal{Y}$. This joint system reads

$$
\begin{aligned}
\mathbf{x}_{n+1} &= \mathbf{f}(\mathbf{x}_n) \\
y_{n+1} &= \kappa y_n + g(\mathbf{x}_n) .
\end{aligned}
\tag{1}
$$

The map $\mathbf{f} : \mathbb{R}^M \mapsto \mathbb{R}^M$ is invertible, and $g : \mathbb{R}^M \mapsto \mathbb{R}$ is a projection, for instance selecting a single component of the M-component map. The parameter $\kappa \in (0,1)$ is the filter constant. We consider a stationary drive, meaning that the sequence $\mathbf{x}_n$ is a chaotic trajectory on an attractor.

The filter is always in a state of generalized synchronization, meaning that its current state is a function of the current state of the dynamical system

$$
y_n = H(\mathbf{x}_n) .
\tag{2}
$$

This property can be shown by unwrapping Eq. (1) backwards in time

$$
y_n = \sum_{l=1}^{\infty} \kappa^{l-1} g(\mathbf{x}_{n-l}) .
\tag{3}
$$

The initial conditions of $\mathcal{Y}$ are an additional transient term of the form $y_0 \kappa^n$, which is decayed for large n.

Through the inverse map of the drive system, $\mathbf{f}^{-1}$, all points in time can be formally expressed as a function of the present one. This leads to

$$
\begin{aligned}
y_n &= \sum_{l=1}^{\infty} \kappa^{l-1} g\left(\mathbf{f}^{-l}(\mathbf{x}_n)\right) \\
&= H(\mathbf{x}_n) .
\end{aligned}
\tag{4}
$$

Thus the iterates of the dynamical system collect all past time steps in the sum in Eq. (3) and reduce the dependency to the present state only, revealing the function $H : \mathbb{R}^M \mapsto \mathbb{R}$, from Eq. (2). The temporal representation, Eq. (3), is a simpler function in the sense of a lower degree of nonlinearity, while the complicated instantaneous relationship is implicitly contained in the backward iterations of $\mathcal{X}$. The trade-off in this representation lies in the increased dimension. Due to the exponential kernel, however, a finite cutoff after $L \gg 1$ time steps will sufficiently well represent the infinite sum.

From a time-series analysis point of view, the temporal representation, although high-dimensional, may be preferred. For the task of detecting GS, methods which utilise a delay embedding only need to model the simpler representation of the functional relationship, meaning that they can be very effective[12]. Echo state networks naturally perform an embedding on an input signal and so fall into this category. We demonstrate a simplified form of the modelling process on the above coupled map system, using a linear model with a delay-embedding of dimension D. We assume that the input signal, $u_n = g(\mathbf{x}_n)$, and the output signal, $s_n = y_n$, are given. The model is of the same form as Eq. (3), but with a truncated sum

$$\hat{y}_n = \sum_{l=1}^{D} \kappa^{l-1} u_{n-l} \ . \tag{5}$$

For any finite value of D, there is a mismatch between the original response y_n, and the model approximation $\hat{y}_n$, which we measure by the normalized root-mean-square error (NRMSE)

$$\mathcal{E} = \frac{\sqrt{\langle (y_n - \hat{y}_n)^2 \rangle_n}}{\sqrt{\langle (y_n - \langle y_n \rangle_n)^2 \rangle_n}} \ . \tag{6}$$

If the drive signal is uncorrelated noise, $u_n \in \mathcal{N}(0,1)$, one can analytically quantify the remainder of the infinite sum, yielding an error of $\mathcal{E} \propto \kappa^D$. This means, that the model error with increasing embedding dimension follows the same exponential trend as the convolution of the filter. We have verified this convergence of the model for a low-pass filter driven by a component of the chaotic Hénon map $x_{1,n+1} = 1 - 1.4x_{1,n}^2 + x_{2,n}$, $x_{2,n+1} = 0.3x_{1,n}$. There is a clear exponential trend indicating the filter constant (not shown). The extension of the result from a noise drive to a chaotic map is not obvious in view of the characteristic correlations of the chaotic signal. However, an argument in favour of a strong analogy is based on homogeneity in time: The remainder of the infinite sum is always the same, except for a time shift and a scaled magnitude. Correlations of the chaotic signal within the remaining convolution tail depend only on the time difference to a reference point, e.g. the cutoff position D. Thus, each embedding increment leads to an equivalent situation up to a constant scaling factor.

B. Modelling drive-response relationships

Considering the case of general unidirectionally coupled systems,

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \mathbf{f}(\mathbf{x}(t)) \\ \dot{\mathbf{y}}(t) &= \mathbf{g}(\mathbf{y}(t), \mathbf{x}(t)) \ , \end{aligned} \tag{7}$$

we ask the following questions: Can we identify the different representations of a GS state? What do they imply

for modelling approaches based on an extended embedding, as given by a reservoir-computing scheme? What can be said about the dynamical dependencies beyond the presence of a functional relationship?

First, it is not feasible to write down a closed form of a function like Eqs. (4) and (A1) when dealing with chaotic, or even noisy trajectories. We may assume, however, that the property of simplifying the functions in the extended representation still applies. This should be understood at least in the sense that, in the case of weak GS, the function in the temporal representation will be smooth. Further, understanding nonlinearity in terms of the order of an expansion necessary to approximate the functions, this order is expected to be reduced through embedding. In the above examples with low-pass filters as response systems, the embedded representation was entirely linear, while all nonlinearity in the instantaneous representation stems from the backward iterates of the drive system.

Second, as far as observations from the simple examples are representative for the general case, it is reasonable to identify the maximal CLE with the filter constants. It is not clear, however, where to expect the signature of the maximal CLE in the model performance as a function of the embedding depth D, as other effects will overlay the trends. The rest of the conditional Lyapunov spectrum, which contributes only partially to the dimension, if at all, will play a role. We propose a simplified decomposition of the GS manifold, in which the approximately adiabatic part of the response is separated by $\mathbf{y}(t) = \mathbf{H}_1(\mathbf{x}(t)) + \mathbf{H}_2(\mathbf{x}(t))$, where $\mathbf{H}_1(\cdot)$ is a convolution related to the maximal CLE, or the active part of the spectrum contributing to the dimension, and $\mathbf{H}_2(\cdot)$ is a smooth function related to the inactive degrees of freedom. In the performance over D, the first part should dominate the asymptotic behavior, while the second manifests at low dimensions. Despite being oversimplified, the model may account for some of the phenomena beyond the classification by the maximal CLE, and provide guidance for numerical experiments. Another interpretation of the second part of the function is a large-scale nonlinear response, which is not directly related to the conditional Lyapunov spectrum. This coarse structure in phase space can be assigned to the individual geometry of the attracting set in phase space, which may be quantified by the mutual information between instantaneous drive and response variables at low resolution.

Finally, if there is no GS manifold because of a positive CLE, the concept of consistency applies to characterize the conditional probability density[17,19]. From the modelling perspective, this poses no material difference, except for the consistency limit being an upper bound of the performance[37]. Below this limit, similar rules may apply as for the GS case, and the maximal CLE from the negative part of the spectrum may take over the role of maximal CLE when modelling GS. We address such aspects in Sec. III, using ESN to reconstruct the dynamical dependencies between drive and response.

C. Reconstruction by echo state networks

A reservoir computer is a dynamical system $\dot{\mathbf{r}}(t) = \mathbf{f}(\mathbf{r}(t), \mathbf{u}(t))$ driven by a set of M input signals $\mathbf{u}(t)$, together with a projection $\hat{s}(t) = \mathbf{R} \cdot \mathbf{r}(t)$ of its high-dimensional state which is adjusted in a training process to approximate a target signal $s(t)$. The role of the RC in this work is to imitate a response system as in Eq. (7) by mapping the input $u(t) = x_i(t)$ to the output $s(t) = y_j(t)$, for a choice of components i, j. We use for this task ESNs, which are discrete-time RCs. An ESN consists of N degrees of freedom, represented as nodes of a network. The nodes are connected according to an internal connection matrix $\mathrm{W} \in \mathbb{R}^{N \times N}$, whose non-zero entries define the links of the network. An input signal $u(n)$, sampled on discrete time $n \in \mathbb{N}$, is injected into the network according to a vector of incoming connection weights $\mathbf{V} \in \mathbb{R}^N$. We adapt a popular choice of the update equation for the state of the network $\mathbf{r}(n) = [r_1(n), r_2(n), ..., r_N(n)]^\top$ which reads

$$\mathbf{r}(n) = \tanh(\mathrm{W} \cdot \mathbf{r}(n-1) + \mathbf{V} \cdot u(n) + \boldsymbol{\beta}) . \qquad (8)$$

The $\tanh(\cdot)$ is the nonlinear activation function which is applied element-wise. The vector $\boldsymbol{\beta} \in \mathbb{R}^N$ contains biases which set the offset of each node. For the ESN training procedure, W, $\mathbf{V}$, and $\boldsymbol{\beta}$ are fixed, and only a linear readout $\mathbf{R}$ is trained to approximate a given target signal $s(n)$, typically according to linear regression

$$\mathbf{R} = \arg\min_{\mathbf{R}'} \left\langle (\mathbf{R}' \cdot \mathbf{r}(n) - s(n))^2 \right\rangle_n .$$

Once set up in a suitable operating regime, the nodes of the ESN each contain a function of the input with fading memory $r_i(n) = h_i(u(n), u(n-1), u(n-2), ...)$. As such, the nodes of the ESN can be thought of as the elements of a basis for a function space. The readout $\mathbf{R}$ is a set of coordinates in this space which selects and combines the elements of this basis to best approximate the function which maps $u(n)$ to $s(n)$. Throughout this work we split the data into distinct training, validation, and testing sets for model fitting, hyperparameter optimization, and performance evaluation, respectively. This includes Tikhonov regularization to avoid overfitting. The regularization parameter $\Lambda \in [0, 1]$ is set to optimize out-of-sample performance.

There is an ongoing discussion in the community on the best design principles for RC. There are few definitive rules, but mostly general guidelines and principles[27,38]. Typically, the elements of W, $\mathbf{V}$ and $\boldsymbol{\beta}$ are chosen randomly. Here, we choose W to be a sparse matrix with probability $p = \frac{10}{N}$ of a non-zero entry, indicating a connection between two nodes. This ensures an average degree of 10 across networks of different sizes. We then set the weights of those connections to random values from $\mathcal{N}(0, 1)$. The matrix W is scaled by a prefactor to achieve a desired spectral radius $\rho(\mathrm{W})$. The spectral radius is the largest absolute value of the eigenvalues of W and is an important parameter in ESN design. It determines the

gain on the internal dynamics of the ESN, with larger spectral radii causing signal to propagate in the network for a longer time, meaning the ESN has longer memory of past input states. The elements of $\mathbf{V}$ are chosen uniformly from the interval $[-\sigma, \sigma]$, with larger values of σ leading to greater nonlinearity in the transformation of $u(n)$. The bias elements are chosen uniformly, $\beta_i \in [-1, 1]$.

III. UNIDIRECTIONAL COUPLING

We follow the approach by Schumacher et al.[12] and consider a unidirectionally coupled pair of non-identical oscillators. A Rössler system drives a Lorenz system with diffusive coupling according to

$$\begin{aligned}
\dot{X}_1 &= -(Y_1 + Z_1) \\
\dot{Y}_1 &= X_1 + a\,Y_1 \\
\dot{Z}_1 &= b + Z_1(X_1 - c) \\
\theta^{-1}\dot{X}_2 &= \sigma_L(Y_2 - X_2) - K(X_2 - f(X_1, Y_1, Z_1)) \\
\theta^{-1}\dot{Y}_2 &= X_2(\rho_L - Z_2) - Y_2 - K(Y_2 - f(X_1, Y_1, Z_1)) \\
\theta^{-1}\dot{Z}_2 &= X_2 Y_2 - \beta_L Z_2 - K(Z_2 - f(X_1, Y_1, Z_1))
\end{aligned}$$

$$(9)$$

The parameters are $a = 0.45$, $b = 0.2$, $c = 4$, $\sigma_L = 10$, $\rho_L = 28$, $\beta_L = 8/3$. The coupling strength $K > 0$ is varied to obtain different levels of entrainment, including weak and strong GS regimes. The timescale parameter $\theta = 0.2$ ensures that the relatively slow Rössler system does not drive the Lorenz system adiabatically. The coupling function is $f(X_1, Y_1, Z_1) = X_1 + Y_1 + Z_1$, though other configurations would be possible which may lead to different behavior.

We set up an ESN to model the relationships, including GS, between this pair of drive and response. A sampling time step of $\delta t = 1$ is chosen to translate the continuous dynamics to the discrete-time reservoir. The first variable of the Rössler system is used as the input. Normalization to zero mean and unit variance is applied, so that $u(n) = \widetilde{X}_1(t_n)$ is the resulting input signal, with $t_n = n\delta t$, and $\widetilde{X}_1$ denoting the normalized X_1. The size N of the ESN is in the range of hundreds of nodes. Input scaling is set to $\sigma = 1$, and the spectral radius $\rho = 1.5$. With regard to the latter, we always test for the echo state property to discard unstable reservoirs. The output of the ESN is trained to approximate the X_2 variable of the driven Lorenz system.

The accuracy of the approximation of the target $s(n) = X_2(t_n)$ by the actual output $\hat{s}(n) = \mathbf{R} \cdot \mathbf{r}(n)$ is quantified by the NRMSE according to Eq. (6). We perform this procedure for $K \in [0, 100]$, which encompasses regimes of no GS, weak GS and strong GS. In order to track the state of synchronization, we calculate the CLE of the response system, using the QR orthonormalization procedure[10,61]. If the maximal CLE is negative, the driven system is conditionally stable and thus in a state of GS.
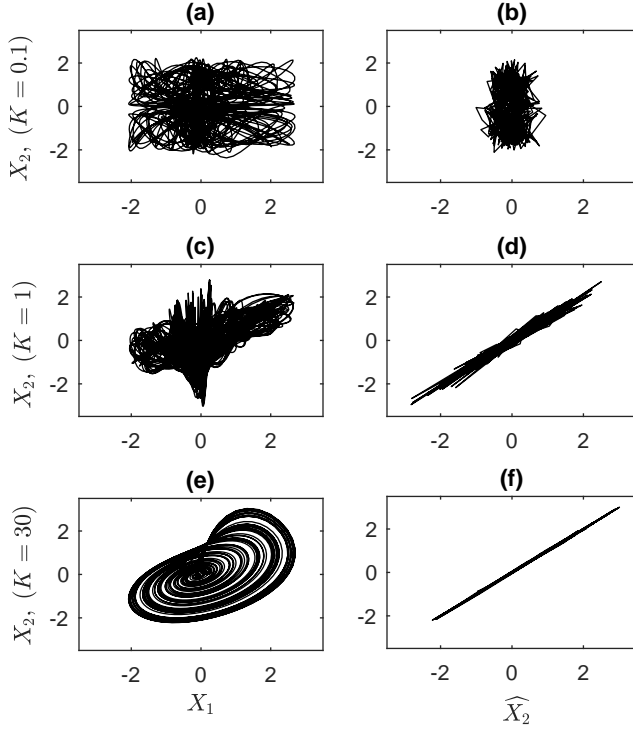
FIG. 1. Demonstration of the ESN's ability to unfold the synchronization manifold for the Rössler-Lorenz system in a state of no GS (a-b), weak GS (c-d), and strong GS (e-f). The synchronization regimes are verified by the maximal CLE, with $\lambda_{\text{MAX}} = 0.19, -0.08, -4.00$ respectively. The left column shows a projection of the trajectory of the combined system to the input-output relationship being modelled. The right column shows the reservoir output with the target signal. The correlations for $K = 0.1, 1, 30$ are $C_{s,\hat{s}} = 0.01, 0.99, 1.00$ respectively.
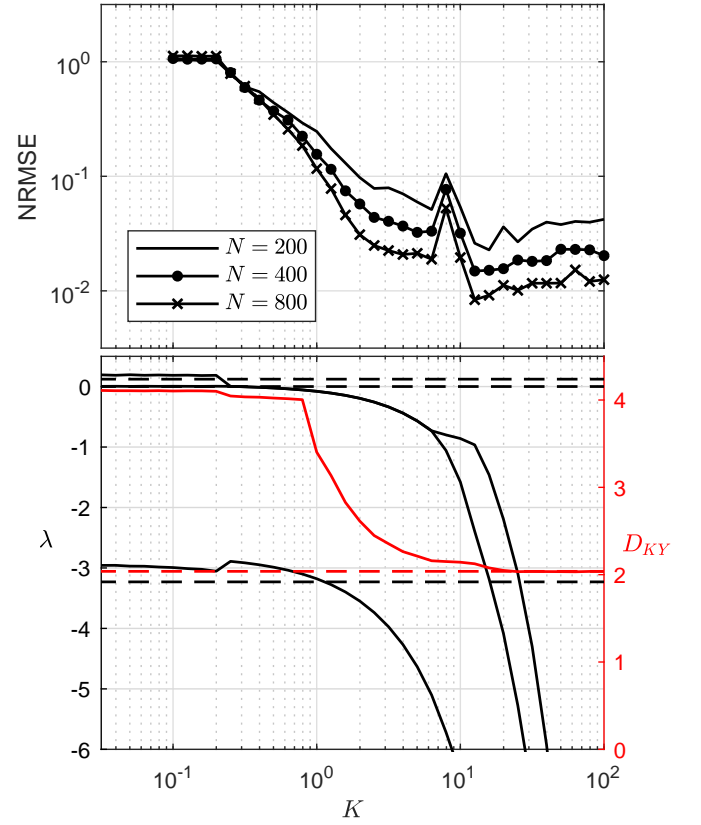

FIG. 2. The reconstruction accuracy for the Rössler-Lorenz system (Eq. 9) by ESNs of various sizes (top panel) with the Lyapunov exponents (black) and Kaplan-Yorke dimension (red) of the combined system (bottom panel). The reconstruction accuracy is averaged over the results of 10 separate ESNs. The dashed lines show the Lyapunov exponents and KY dimension of the Rössler system, which are independent of the coupling strength K.

For weak GS, the functional dependence is extremely complicated in the instantaneous representation (Eq. 4), whereas in the temporal representation (Eq. 3) the relationships may become sufficiently simple according to the model considerations in Sec. II. This means that the ESN, which naturally performs an embedding on the input signal, is expected to approximate the GS state well.

Figure 1 shows portraits of input and output for the coupled Rössler-Lorenz system in each of these three regimes, together with the reservoir approximation. In the first regime of no GS ($K < 0.2$), the ESN is not able to infer the state of the response system reliably, and the NRMSE is of the order of one. For the weak GS regime ($0.2 < K < 20$), some structure appears betwen drive and response coordinates (Fig. 1c). The ESN unfolds this complicated manifold, with a correlation between output and target of $C_{s,\hat{s}} = 0.99$ (Fig. 1d). For the strong GS regime ($K > 20$), the relationship between the two systems is apparently much simpler (Fig. 1e), and the ESN achieves extremely high modelling accuracy with correlations of one (Fig. 1f).

Figure 2 summarizes the performance together with the CLE as a function of the coupling K. We also show the Kaplan-Yorke dimension[62] of the combined system, by which one can distinguish between weak and strong GS. The reconstruction accuracy begins to increase sharply at $K \approx 0.2$, as the maximal CLE becomes negative and the system transitions from no GS to a state of GS, $(X_2(t), Y_2(t), Z_2(t)) = \mathbf{H}(X_1(t), Y_1(t), Z_1(t))$, which persists for all larger couplings. The transition is accompanied by a small decrease in the dimension of the combined system, which however is still larger than that of the drive only, indicating the fractal relationship of weak GS. As K increases, the CLE decrease further, until a state of strong GS is reached at $K \approx 25$. At this point, the dimension collapses to that of the drive, which is a signature of the smooth instantaneous relationship. With the notation of Sec. II B, strong GS is given by $\mathbf{H}(\cdot) = \mathbf{H}_2(\cdot)$. Throughout the three different regimes, the reconstruction accuracy varies over many orders of magnitude, with the ESN able to achieve a very high level of accuracy, especially for large K.

The performance starts to saturate around $K \approx 3$, while an intermediate drop, indicated by the short spike in the NRMSE at $K \approx 8$, appears to slightly shift the saturation level. On the one hand, these patterns are associated with qualitative and quantitative changes of the response behavior of the Lorenz system. Such changes may already be traced back to the Lyapunov spectrum, but we will not further investigate the detailed behavior here. Overall, the GS manifold becomes simpler and more linear with increasing K (see Fig. 3), until in the limit of $K \to \infty$ the response follows the drive signal adiabatically. On the other hand, the performance saturation, despite this increasing simplification, is related to the reservoir becoming ill-matched with the properties of the GS function, in particular with the fast timescale. We investigate this matching in the following with variations of the ESN structure.

A. Memoryless reservoir

The ESNs achieve high modelling accuracy of the GS manifolds through their fading memory, which allows them to access even the nonlinear convolutions of the weak GS regime. We hypothesize that this advantage in weak GS will be detrimental for strong GS, where memory is not essential to approximate the smooth instantaneous functions. In order to investigate the role of memory, we consider first a modified ESN without any memory and compare it to the standard ESN. Such a network can be created by removing all internal connections by setting $W = \mathbf{0}$. This way the ESN effectively reduces to a type of feedforward network with one hidden layer containing randomly fixed activation functions, and a trained linear output. We note that this special architecture is also known as extreme learning machine (ELM)[63,64]. Although ELMs are not a state-of-the-art method among feedforward structures, they still perform well on some non-trivial tasks[65]. The fact that ELMs are a limit case of ESNs makes them ideal candidates for our comparison. For coherence with the paradigm of this study, we will refer to the ELM as a memoryless ESN, or memoryless reservoir.

When applying the memoryless ESN, the full state of the drive system needs to be used as input, as the network does not create the minimal embedding signal to represent the drive state space[66]. We write the memoryless ESN as

$$\mathbf{r}(n) = \tanh(V \cdot \mathbf{u}(n) + \boldsymbol{\beta}).$$

with $V \in \mathbb{R}^{N \times 3}$ to account for the three Rössler coordinates, $\mathbf{u}(n) = (X_1(t_n), Y_1(t_n), Z_1(t_n))^\top$, and the bias $\boldsymbol{\beta}$ remains as before. We repeat the same task, namely approximation of $s(n) = X_2(t_n)$ for $K \in [0, 100]$. A simple linear regression model $\hat{s}_{lin}(n) = s_0 + \mathbf{R}_{lin} \cdot \mathbf{u}(n)$ is also compared to indicate the linear correlation between drive and response. The results are shown in Fig. 3.
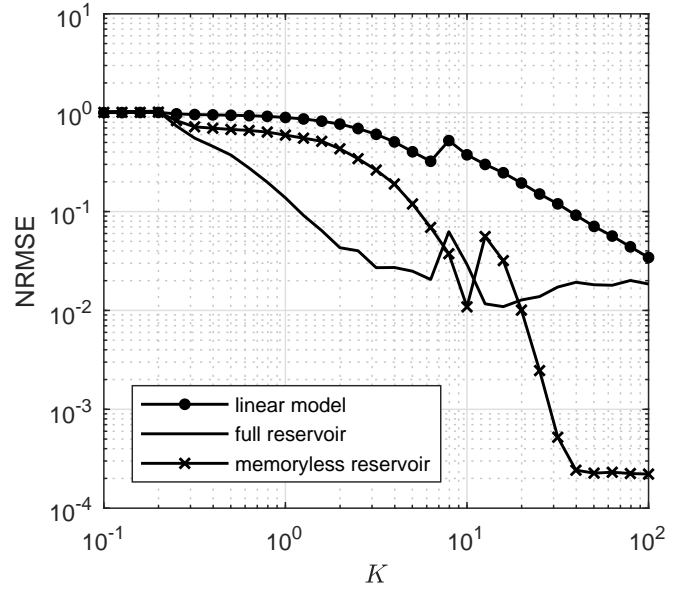

FIG. 3. Reconstruction accuracy for the full reservoir, the memoryless reservoir and a linear model.

For low K, before the onset of synchronization, none of the models are accurate because of the fundamental consistency limit. After the onset of GS at $K \approx 0.2$, the accuracy of both reservoirs increases significantly with K, but the full ESN outperforms the memoryless ESN. This indicates the ability of the full ESN to access the simpler, temporal form of the GS manifold, while the fractal dependence in the instantaneous representation limits the ability of the memoryless ESN. For larger coupling, the accuracy continues to increase for all models, but we recognize the aforementioned plateauing effect of the full ESN. The memoryless ESN overtakes the standard ESN near the transition to strong GS, and reaches a plateau of much better performance around $K \approx 30$. The dominance of each reservoir type can be explained in terms of the capacity of the network. The memoryless ESN reserves its entire capacity for instantaneous functions, which is of advantage in strong GS, whereas the full ESN distributes capacity over its memory tail, allowing for reconstruction of the long-range dependencies in weak GS.

B. Lyapunov time and memory capacity

We investigate further the relationship between timescales of the GS functional and memory of the reservoir. The simple models in Sec. II have shown timescale signatures in the performance when the memory depth of the model is varied. Here, we imitate this variation with controlled modifications of the ESN memory. Fading memory is commonly quantified by the memory capacity[33] (MC), which is calculated by driving the ESN with uncorrelated noise and training it to reproduce the

past input $u(t - \tau)$ for a range of lags τ. The reconstruction accuracy $M(\tau)$ indicates the amount of information held in the network for that lag. The memory capacity is then calculated by summing the square of the correlations between target and approximation over all lags

$$M(\tau) = \langle s_\tau(t) \hat{s}_\tau(t) \rangle_t$$

$$MC = \sum_{\tau=1}^{\infty} M(\tau)^2 \ ,$$

where $s_\tau(t) = u(t - \tau)$ and $\hat{s}_\tau(t)$ is the trained reservoir's approximation of $s_\tau(t)$. The MC cannot be directly controlled in typical ESN, instead it is indirectly affected by the design of the network. In particular, spectral radius ρ and network size N have a strong influence on the MC. We choose to obtain a range of MC by generating networks of different size N while keeping ρ constant.

A second method to adjust memory is related to the unfolded representation of RNN within the backpropagation through time algorithm. A recurrent network can be represented as a layered structure, where each layer represents one time step, and the recurrent connections are replaced by connections between the layers. We generate typical ESN and unfold them in this way. Such an unfolded network leads to an infinite amount of layers, as shown in Fig. 4b. The update equation for layer j reads

$$\mathbf{r}^{(j)}(n) = \tanh\left(W \cdot \mathbf{r}^{(j-1)}(n-1) + \mathbf{V} \cdot u(n) + \boldsymbol{\beta} \right) \ .$$

In this sense, the network can again be considered recurrent, even though there are no recurrent connections within each layer. The time increment is included in the inter-layer connections, with the ESN's dependence on the input history arising from the propagation delay through the layers. This means that the current input $u(n)$ must be injected at each layer, unlike feedforward versions of this unfolding, which incorporate the time increment into the different inputs for each layer.

The result of this unfolding is an extended state vector $\mathbf{r}^*(n) = (\mathbf{r}'(n), \mathbf{r}''(n), ...)^\top$ and modified connection matrices W^* and $\mathbf{V}^*$ as

$$W^* = \begin{bmatrix} 0 & 0 & 0 & \\ W & 0 & 0 & \cdots \\ 0 & W & 0 & \\ & \vdots & & \ddots \end{bmatrix}, \quad \mathbf{V}^* = \begin{bmatrix} \mathbf{V} \\ \mathbf{V} \\ \mathbf{V} \\ \vdots \end{bmatrix},$$

which can be used in the standard ESN update equation (Eq. 8) to implement this unfolded representation. Note that the readout is only trained on the final layer, not the extended state.

Memory can be controlled by truncating the layered structure, see Fig. 4c, which is similar to the model Eq. (5). The layers of this finite chain hold the transient activations of the reservoir. If the echo state property holds, and the number of layers is significantly larger than the MC, the upper layers will hold almost identical
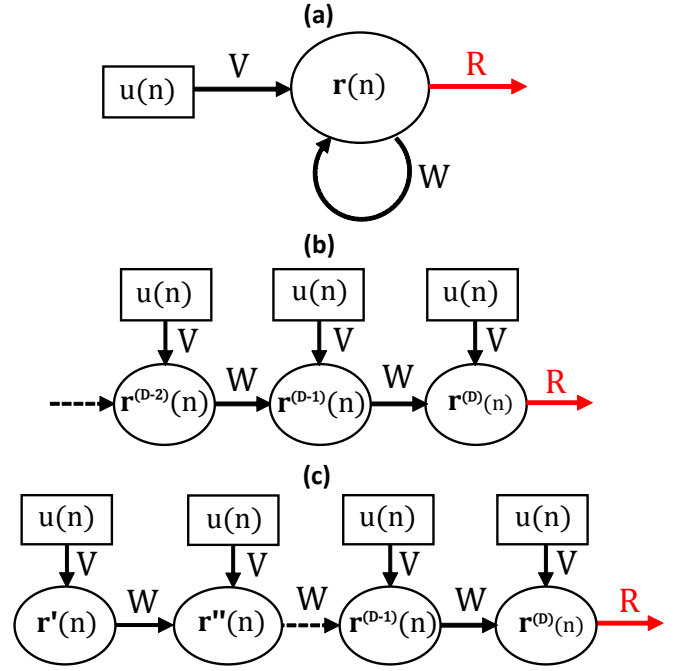

FIG. 4. Recurrent (a), unfolded (b) and truncated (c) representation of an ESN.

activations. Otherwise, by truncating the layered network within the memory tail of the original ESN, the memory can be reduced with minimal effect on the other properties of the network. An example of the memory profile of such a finite stack of layers is shown in Fig. 5a.

We quantify the timescale of the GS interaction by the Lyapunov time $T_\lambda = |\lambda_{\text{MAX}}|^{-1}$, where λ_{MAX} is the maximal CLE. As shown in Fig. 5b, the timescale decreases with the coupling, because $\lambda_{\text{MAX}} \propto K$ for large K.

We repeat the previous GS task with the above two reservoir architectures. Figure 5c shows the results for changing N of the standard reservoir. We observe a broad trend that performance increases with memory. However, by increasing N the reservoir capacity also increases, meaning the set of nonlinear functions available for the readout also increases, which has a strong effect on performance. In contrast, for the truncated unfolded representation, the number of nodes is held constant, thus excluding the effect of capacity. The results for this method are shown in Fig. 5d. Again, we observe the same overall trend between memory and performance. However, for larger values of K, corresponding to shorter timescales, the trend saturates and inverts. Optimum performance for the layered ESN is located at a lower memory than for the fully extended ESN ($D \to \infty$). This is the same effect that was observed in Fig. 3, as functionality of long memory decreases the performance when modelling short timescales.

Following the ideas of Sec. II, we expect to find signatures of the maximal CLE (λ_{MAX}) in the NRMSE over the embedding depth D. Indeed, we identify such a relationship in Fig. 5d. The effect is subject to the limita-

FIG. 6. Decay time $1/\alpha$, extracted from the linear fits in Fig. 5d, against the Lyapunov time $T_\lambda = |\lambda_{\mathrm{MAX}}|^{-1}$. Identity line is shown for reference.
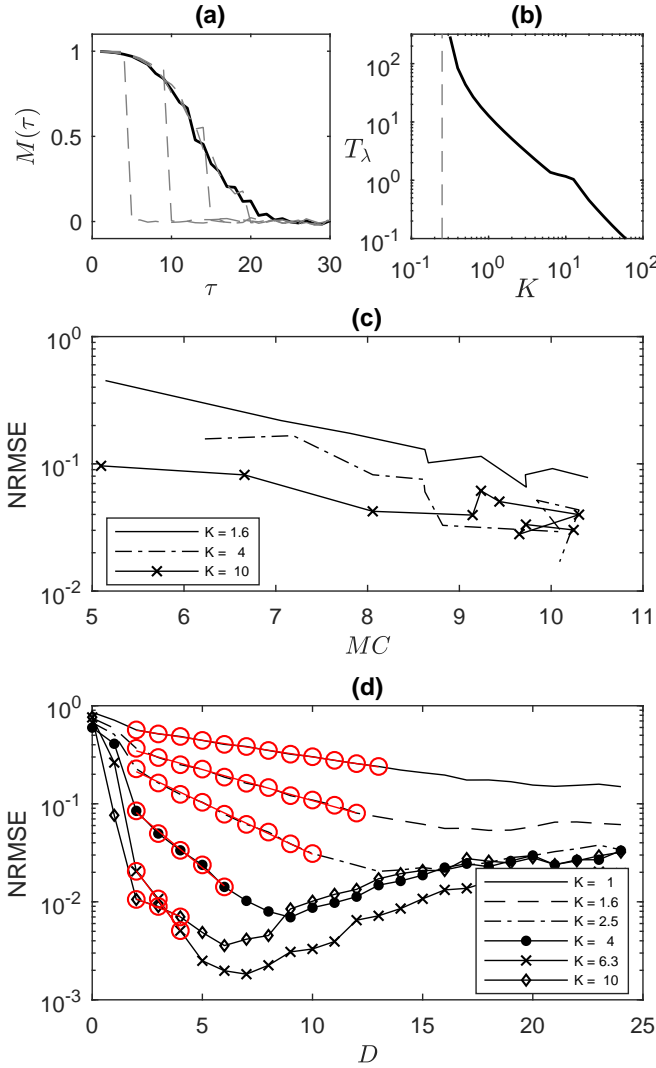

FIG. 5. (a) Fading memory profile of a 500 node ESN, from which the memory capacity is derived. The dashed lines show the typical memory profile for some truncations of the reservoir. (b) Timescale T_λ of the response as a function of coupling strength K. (c) NRMSE for ESN with increasing memory, controlled by changing the network size. (d) NRMSE for ESN with increasing memory, controlled by truncating the unfolded representation. The red circles indicate the data points selected to perform a linear fit.

selections are shown as red circles in Fig. 5d. In general, they lead to appropriate scaling regions which are well fit by a linear fit. However, for large K the timescale of the response becomes small, and the benefits of temporal processing decrease, with the memoryless reservoir performing well relative to reservoirs with memory. This means that the scaling regions for larger K are poorly defined, and may lead to unreliable results.

Gathering the linear fits as shown in Fig. 5d over a range of K, we compare the gradients α of the fits with the Lyapunov time (Fig. 6). We observe that the decay time constant α^{-1} is in excellent agreement with T_λ in the moderate timescale regime, in accordance with the considerations of Sec. II. At the extremes, the identity does not hold. In the short timescale regime, scaling regions are poorly defined because of the vanishing temporal component $\mathbf{H}_1(\cdot)$ of the GS functional. In the long timescale regime, the relationship breaks down near the transition point of GS, where the reservoir is limited by its memory capacity and performance saturates at a finite level. Overall, the signature of T_λ can be traced for almost two orders of magnitude.

C. Consistency

Consistency can be understood as an extension to generalized synchronization, which incorporates cases where the functional dependence is imperfect, in the sense of a 'blur' which arises from noise or instabilities[17,36,67]. Moreover, consistency refers to the degree of dependence of a dynamical node to its input, rather than the type of dependence. That degree can be accessed through the replica scheme, in which either a copy of the system is driven simultaneously together with the original, or the signal is repeatedly injected in a single system[59,68]. For a scalar recording, the measure of consistency is the cross-correlation between the response of original, $s(t)$, and

tions of the ESN's maximum memory capacity, leading to a saturation. Furthermore, this relationship does not apply to the instantaneous component $\mathbf{H}_2(\cdot)$ of the GS manifold, which manifests at the lowest D. We define a method for selecting a scaling region by stipulating the maximum error to be that of the memoryless ESN at the corresponding value of K (Fig. 3). This error level corresponds roughly to that of the layered ESN when the minimum embedding depth for reconstruction of the drive attractor is reached ($D \sim 2, 3$). The lower cutoff of the scaling region is chosen to be $1.5 \times \mathcal{E}_{min}$, where $\mathcal{E}_{min}$ is the minimum error over D at a given K. These
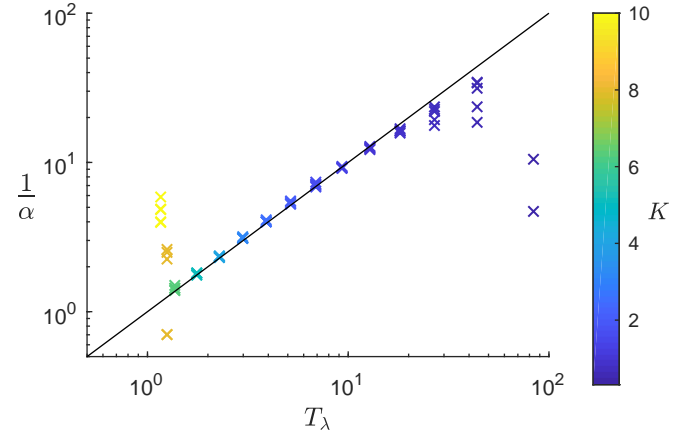
replica, $s'(t)$, as

$$\gamma^2 = \langle \tilde{s}(t) \tilde{s}'(t) \rangle_t \ ,$$

where $\tilde{s}(t)$ and $\tilde{s}'(t)$ denote the time series normalized to zero mean and unit variance. This definition of the consistency correlation γ^2 gives credit to the fact that its square-root γ poses a fundamental limit to any approximation of the response from only the input signal[37,60].

The coupled Rössler-Lorenz system in Eq. (9) illustrates the regimes of no GS, weak GS, and strong GS, where the latter two are accompanied with complete consistency in a noise-free setup. The first regime, in which there is no GS due to positive CLE, has a consistency level $0 < \gamma^2 < 1$ which depends on K. However, except for a narrow region in the vicinity of the transition to GS, γ^2 is very close to zero. We thus repeat the task for a system of directionally coupled Rössler systems which show a wider range of moderate consistency on different levels.

$$\dot{X}_1 = -(Y_1 + Z_1)$$
$$\dot{Y}_1 = X_1 + a\,Y_1$$
$$\dot{Z}_1 = b + Z_1(X_1 - c)$$
$$\dot{X}_2 = -(Y_2 + Z_2) - K(X_2 - Y_1)$$
$$\dot{Y}_2 = X_2 + a\,Y_2$$
$$\dot{Z}_2 = b + Z_2(X_2 - c)$$

where $a = 0.45, b = 2$ and $c = 4$ as before. The cross coupling term ensures that this system displays GS rather than CS.

As well as calculating γ^2 from an exact replica, we also estimate the consistency level by an imperfect replica, which is created by adding a perturbation of 10^{-3} to each of the parameters a, b and c. The approximation is denoted as γ^{*2}. The same reconstruction task as in Section III is performed using an ESN with $N = 500$, $p = \frac{10}{N}$, $\rho = 1.5$ and $\sigma = 1$. Here we quantify the reservoir's accuracy by a cross-correlation coefficient between the reservoir output $\hat{s}(n)$ and target $s(n)$, $C_{s,\hat{s}}$. Consistency levels and performance correlation are shown in Fig. 7.

At $K \approx 0.18$ the system transitions into a state of GS, with complete consistency for stronger coupling and moderate consistency for weaker coupling. Below the transition, there is an almost linear increase of the consistency level $\gamma(K)$. The mismatched consistency γ^* coincides almost perfectly with γ, except for a deviation around the transition point. This is related to the critical properties of the transition[69], due to which the exact repetition is extremely sensitive.

The reservoir performance largely follows the consistency levels. Similarly to the perturbed replica, the vicinity to the critical point is a major limiting factor, which was also evident in the previous setting (Fig. 2). In the inconsistent regime ($K < 0.18$), the approximation of the limit curve $\gamma(K)$ is imperfect, although the consistency correlation $\gamma^2(K)$ of a single replica can be exceeded.
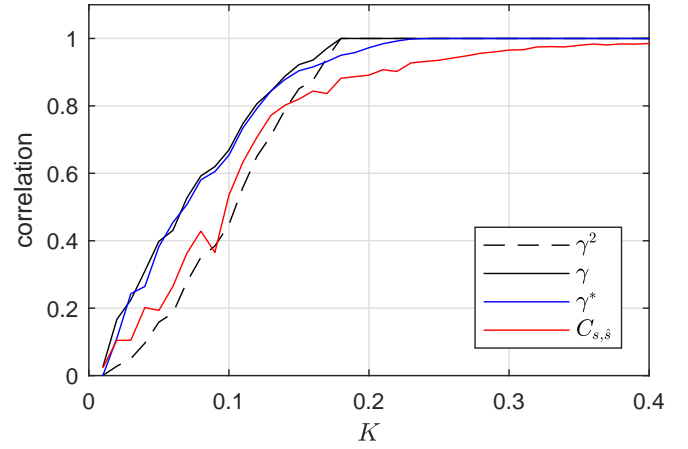

FIG. 7. Consistency correlation (dashed) and limit (solid) for a range of coupling. Estimates of the consistency limit from a mismatched replica are shown in blue. The correlation of an ESN's trained output is in red.

This effect is mostly due to an finite sample length for training. We will not investigate this property further here, but outline the underlying dynamical phenomena. For the reservoir to reach the performance correlation of γ, it needs to not just imitate a single inconsistent replica system, which alone would lead to the level $\gamma^2 < \gamma$. Instead, it needs to infer the ensemble mean $\bar{s}(t) = \langle s_j(t) \rangle_j$, where $s_j(t)$ is the response of the j-th replica system[60], or alternatively the j-th repetition of the task, via the ergodic properties of a long time series. Our present experience indicates that the sample size requirements for the convergence of such an average may be extreme.

IV. BIDIRECTIONAL COUPLING

A reservoir computer is compatible with the drive-response scheme in a similar way to the replica test, in which a perfect copy of the response system is attached to the drive, indicating conditional stability and consistency[8,59]. The reservoir can be considered an imperfect replica emulating the response system with some accuracy. In a bidirectional coupling scheme, or in a connected network of dynamical nodes, the replica approach is in principle not applicable[70]. An attached replica system will indicate conditional stability of the subsystem, but not presence of GS, which is determined by the mutual interaction. What is then the meaning of a reservoir which computes variables of one subsystem from those of another? Already for the unidirectional coupling, it is evident that the reservoir-computing scheme is not a classifier for the GS state. Hence we expect even less specificity for GS manifolds in the case of mutual coupling, because according to embedding theory GS is not a requirement to reconstruct remote degrees of freedom. Instead, reservoir computing with mutually coupled subsystems is a special case of the reconstruction task, which

was investigated recently[21,22,51,55]. The performance will thus depend on observability[23,71].

We apply the ESN toolkit of the previous section to the time series from two detuned Rössler systems with mutual coupling

$$\dot{X}_1 = -(1 - \Delta\omega)\,Y_1 - Z_1$$
$$\dot{Y}_1 = (1 - \Delta\omega)\,X_1 + a\,Y_1 - K(Y_1 - Y_2)$$
$$\dot{Z}_1 = b + Z_1\,(X_1 - c)$$
$$\dot{X}_2 = -(1 + \Delta\omega)\,Y_2 - Z_2 \qquad (10)$$
$$\dot{Y}_2 = (1 + \Delta\omega)\,X_2 + a\,Y_2 - K(Y_2 - Y_1)$$
$$\dot{Z}_2 = b + Z_2\,(X_2 - c)$$

where $a = 0.16$, $b = 0.1$, $c = 8.5$, K is the mutual coupling strength, and $\Delta\omega = 0.02$ is the spiral frequency detuning. This system is known to display phase and generalized synchronization[72,73]. Specifically, the system transitions from an asynchronous state to phase synchronization (PS) at $K \approx 0.04$ and to generalized synchronization at $K \approx 0.09$. These transitions can be observed in the Lyapunov spectrum and Kaplan-Yorke dimension of the full system as shown in Fig. 8a.

In the unidirectional case, the coupling direction determines which subsystem is the input and which is the target, but here either may be used. We choose Y_1 as the input variable for the reservoir, $u(n) = Y_1(t_n)$, with $t_n = n\delta t$, and the sampling time $\delta t = 1$ analog to Sec. III. The target variable for training is $s(n) = Y_2(t_n)$. We apply three different reservoir configurations. The first one is a standard ESN with parameters $N = 500$, $p = \frac{10}{N}$, $\rho = 0.1$, $\sigma = 1$. The second type is the memoryless ESN, which is set up in the same manner as described in Section III with $N = 500$ nodes. As before, the full state of one of the subsystems is required for the memoryless reservoir, as it does not perform an embedding. This allows the memoryless ESN to be sensitive to the presence of GS, under the assumption that the detuning $\Delta\omega$ is small enough such that the manifold does not deviate too much from unity. For the third method, observational noise $\xi(t) \sim \mathcal{N}(0, 0.01)$ is added to the state of the combined system. Finally, we also compare the reservoir methods with a simple linear model, which like the memoryless ESN receives the vector state of one system as input.

The results are shown in Fig. 8b for a range of coupling strengths. Here we use the cross-correlation between the reservoir output and target to quantify the reservoir's performance. In the GS regime, all methods are able to accurately reconstruct the second subsystem, implying the presence of an instantaneous functional dependence. In the PS regime, where the phases of the two subsystems are locked together but the amplitudes are unrelated due to one positive transverse LE, all methods are still highly accurate. We relate the high accuracy to the small detuning and the small relative fluctuations in the amplitudes. Below the PS regime (onset $K \approx 0.04$), where no synchronization is present, the memoryless reservoir and
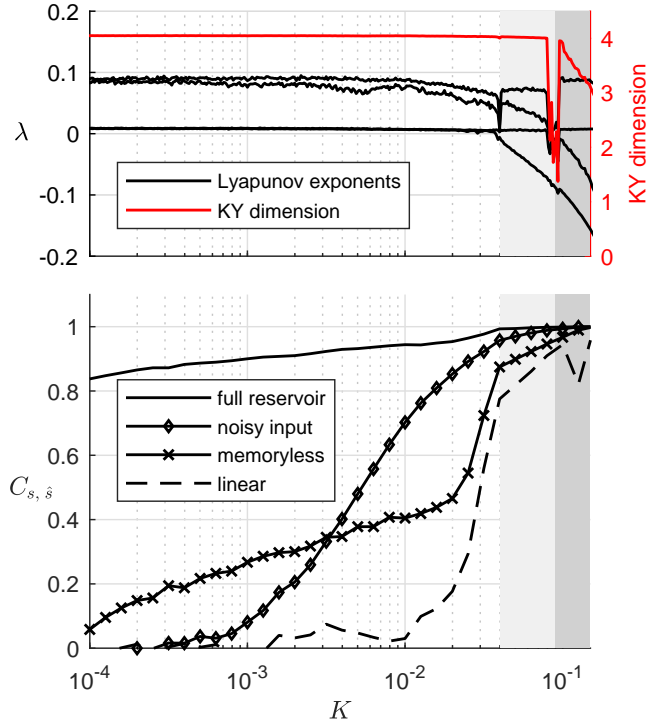

FIG. 8. Top panel: the four largest Lyapunov exponents and Kaplan-Yorke dimension of the bidirectionally coupled Rössler systems (Eq. 10). Bottom panel: Reconstruction accuracy for the three different reservoir configurations and a linear model. The dark grey region marks the GS regime and the light grey regime marks the PS regime. For weaker coupling there is no synchronization present.

the linear model sharply decrease in accuracy. However, it should be noted that the correlation for the memoryless reservoir does not disappear completely, in contrast to the linear model. This shows that the reservoir nonlinearities provide a significant improvement to the linear model for modelling complicated patterns. The full reservoir and the reservoir with noisy input show a much more gradual change of performance below the PS regime, with the full reservoir maintaining a very high level of accuracy even for extremely small coupling.

The performance of the standard reservoir illustrates how the embedding of the reservoir recreates the entire system because of its full connectivity[51], see Fig. 9, regardless of the change of dimension due to a GS manifold. In contrast, when the reservoir is driven by an observable of the drive subsystem in the drive-response configuration, it is only able to embed that particular subsystem. It is remarkable how for small K in Fig. 8b the reservoir accesses the poorly observable signals. When the coupling strength is below $K \approx 10^{-3}$, the gap between the noise-free performance and the performance with noise shows how the remote signal is hidden below the noise level, and thus cannot be recovered.

The memoryless reservoir does not embed the input signal, instead providing a basis of instantaneous nonlin-
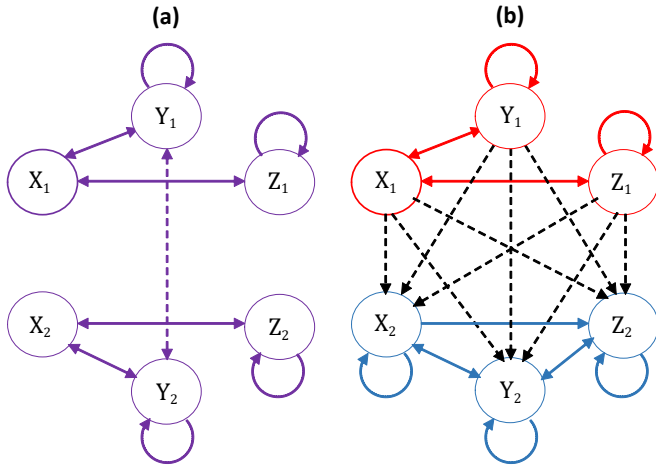
FIG. 9. Different coupling structures for the system with bidirectional coupling (left, Eq. 10) and the system with unidirectional coupling (right, Eq. 9). Each node has incoming connections from all variables which are present in its equation of motion. Dashed lines represent the coupling terms introduced to induce synchronization between the subsystems.

ear transformations. These transformations are sufficient for high correlations in the GS and PS regimes, but in the asynchronous regime, where the degree of correlation decreases significantly, the lack of a representation of the entire system means that the reconstruction accuracy decreases significantly relative to the full reservoir.

V. CONCLUSION

We have studied the reconstruction of coupled chaotic systems by echo state networks, a form of reservoir computer. Our focus lies on a unidirectional coupling scheme, in which the reservoir takes the role of the response system to replicate its output signals. The response can be modelled with high accuracy throughout a wide parameter range.

In weak generalized synchronization, the fading memory of the reservoir allows it to access a simplified representation of the functional dependence. This temporal representation provides a clear advantage over the theoretically equivalent instantaneous representation. Moreover, we have identified a signature of the leading conditional Lyapunov exponent in the convergence of the performance with reservoir memory, where we applied a truncated unfolded reservoir structure to control memory depth. In contrast, any memory emerged to be detrimental in the strong generalized synchronization regime, where a smooth instantaneous relationship is better modelled by a memoryless reservoir. In coupling regimes without generalized synchronization, different levels of consistency can be approximated by the reservoir computer, close to the fundamental limit. The critical behavior at the onset of generalized synchronization impedes reservoir-based modelling, but also poses a challenge to

perfect replicas.

Finally, we have applied the reservoir-computing scheme to mutually coupled chaotic oscillators. While this setup shows to be insensitive to synchronization, the embedding of the reservoir allows for a reconstruction of remote degrees of freedom despite very poor observability. In summary, we have shown how the embedding properties of reservoir computers enable the reconstruction of coupled chaotic systems throughout different dynamical regimes, and we have identified the signatures of generalized synchronization.

Appendix A: Generalized Synchronization in delay systems

A similar situation to that of a drive-response scheme is given in delay systems, from which a basic form reads

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t)) + \mathbf{g}(\mathbf{x}(t - \tau)) \,,$$

with $\tau > 0$ being the delay time, and the functions $\mathbf{f}(\cdot)$ and $\mathbf{g}(\cdot)$ are the vector fields of the nonlinear node and the delayed feedback, respectively. The time-discrete version may also be considered in this context. The delayed feedback acts as an external drive to the undelayed subsystem, thus resembling the drive-response scheme[60]. Two examples of a scalar delay system with a linear local term, $\dot{x}(t) = \mu x(t) + g(x(t - \tau))$, with $\mu < 0$, are given by the Mackey-Glass system and the Ikeda system[74–76]. The variation-of-constants formula, after decay of transients, reveals a type of 'self-GS' as a convolution

$$x(t) = \int_{-\infty}^{t} dt' e^{\mu(t-t')} g(x(t' - \tau)) \,. \qquad (A1)$$

Noteworthy, this functional dependence exists only in the 'embedded' form, and a closed instantaneous expression like Eq. (4) is not defined for the delay system, because the instantaneous response $x(t)$ is itself a component of the state vector of the delay system, which as a whole acts as drive. Modelling, however, the dependence between input $x(t - \tau)$ and output $x(t)$ with an embedding ansatz is naturally reflecting the convolutional form of Eq. (A1). Such an approach is most interesting in cases beyond a simple scalar node, like the Lang-Kobayashi equations for semiconductor lasers, in which non-trivial transitions between negative and positive sub-LE occur, together with different levels of consistency[16,60,77].

ACKNOWLEDGMENTS

[1] A. Pikovsky, M. Rosenblum, and J. Kurths, *Synchronization: A Universal Concept in Nonlinear Sciences*, Cambridge Nonlinear Science Series (Cambridge University Press, 2001).

[2] S. Boccaletti, J. Kurths, G. Osipov, D. Valladares, and C. Zhou, Phys. Rep. **366**, 1 (2002).

[3] A. Arenas, A. Díaz-Guilera, J. Kurths, Y. Moreno, and C. Zhou, Phys. Rep. **469**, 93 (2008).

[4] N. F. Rulkov, M. M. Sushchik, L. S. Tsimring, and H. D. I. Abarbanel, Phys. Rev. E **51**, 980 (1995).

[5] M. C. Soriano, G. Van der Sande, I. Fischer, and C. R. Mirasso, Phys. Rev. Lett. **108**, 134101 (2012).

[6] K. Pyragas, Phys. Rev. E **54**, R4508 (1996).

[7] U. Parlitz, NOLTA, IEICE **3**, 113 (2012).

[8] H. D. Abarbanel, N. F. Rulkov, and M. M. Sushchik, Phys. Rev. E **53**, 4528 (1996).

[9] A. A. Koronovskii, O. I. Moskalenko, and A. E. Hramov, Phys. Rev. E **84**, 037201 (2011).

[10] L. M. Pecora and T. L. Carroll, Phys. Rev. Lett. **64**, 821 (1990).

[11] K. Pyragas, Phys. Rev. E **56**, 5183 (1997).

[12] J. Schumacher, R. Haslinger, and G. Pipa, Phys. Rev. E **85**, 056215 (2012).

[13] H. Kato, M. C. Soriano, E. Pereda, I. Fischer, and C. R. Mirasso, Phys. Rev. E **88**, 062924 (2013).

[14] D. S. Goldobin and A. Pikovsky, Phys. Rev. E **71**, 045201 (2005).

[15] T. Yamamoto, I. Oowada, H. Yip, A. Uchida, S. Yoshimori, K. Yoshimura, J. Muramatsu, S. I. Goto, and P. Davis, Opt. Expr. **15**, 3974 (2007).

[16] S. Heiligenthal, T. Dahms, S. Yanchuk, T. Jüngling, V. Flunkert, I. Kanter, E. Schöll, and W. Kinzel, Phys. Rev. Lett. **107**, 234102 (2011).

[17] A. Uchida, R. McAllister, and R. Roy, Phys. Rev. Lett. **93**, 244102 (2004).

[18] Z. Mainen and T. Sejnowski, Science **268**, 1503 (1995).

[19] G. Giacomelli, S. Barland, M. Giudici, and A. Politi, Phys. Rev. Lett. **104**, 194101 (2010).

[20] P. T. Clemson and A. Stefanovska, Phys. Rep. **542**, 297 (2014).

[21] Z. Lu, J. Pathak, B. Hunt, M. Girvan, R. Brockett, and E. Ott, Chaos **27**, 041102 (2017).

[22] R. S. Zimmermann and U. Parlitz, Chaos **28**, 043118 (2018).

[23] T. Carroll, Chaos **28**, 103117 (2018).

[24] H. Jaeger, German National Research Center for Information Technology GMD Technical Report **148** (2001).

[25] G. Tanaka, T. Yamane, J. B. Héroux, R. Nakane, N. Kanazawa, S. Takeda, H. Numata, D. Nakano, and A. Hirose, Neural Netw. **115**, 100 (2019).

[26] H. Jaeger and H. Haas, Science **304**, 78 (2004).

[27] M. Lukoševičius and H. Jaeger, Comp. Sci. Rev. **3**, 127 (2009).

[28] M. Lukoševičius, H. Jaeger, and B. Schrauwen, KI - Künstliche Intelligenz **26**, 365 (2012).

[29] S. Hochreiter and J. Schmidhuber, Neural Comp. **9**, 1735 (1997).

[30] P. J. Werbos, Neural Netw. **1**, 339 (1988).

[31] M. Hermans, M. Burm, T. Van Vaerenbergh, J. Dambre, and P. Bienstman, Nature Comm. **6**, 6729 (2015).

[32] W. Maass, T. Natschläger, and H. Markram, Neural Comp. **14**, 2531 (2002).

[33] H. Jaeger, German National Research Center for Information Technology GMD Technical Report **152** (2002).

[34] D. Verstraeten, J. Dambre, X. Dutoit, and B. Schrauwen, in *The 2010 International Joint Conference on Neural Networks (IJCNN)* (IEEE, 2010) pp. 1–8.

[35] J. Nakayama, K. Kanno, and A. Uchida, Opt. Expr. **24**, 8679 (2016).

[36] J. Bueno, D. Brunner, M. C. Soriano, and I. Fischer, Opt. Expr. **25**, 2401 (2017).

[37] T. Lymburn, A. Khor, T. Stemler, D. C. Corréa, M. Small, and T. Jüngling, Chaos **29**, 023118 (2019).

[38] M. Lukoševičius, in *Neural networks: Tricks of the trade* (Springer, 2012) pp. 659–686.

[39] L. Appeltant, M. Soriano, G. Van der Sande, J. Danckaert, S. Massar, J. Dambre, B. Schrauwen, C. Mirasso, and I. Fischer, Nature Comm. **2**, 468 (2011).

[40] D. Brunner, M. C. Soriano, and I. Fischer, IEEE Photon. Technol. Lett. **25**, 1680 (2013).

[41] K. Vandoorne, P. Mechet, T. Van Vaerenbergh, M. Fiers, G. Morthier, D. Verstraeten, B. Schrauwen, J. Dambre, and P. Bienstman, Nature Comm. **5**, 3541 (2014).

[42] L. Larger, M. Soriano, D. Brunner, L. Appeltant, J. Gutierrez, L. Pesquera, C. Mirasso, and I. Fischer, Opt. Expr. **20**, 3241 (2012).

[43] L. Larger, A. Baylón-Fuentes, R. Martinenghi, V. S. Udaltsov, Y. K. Chembo, and M. Jacquot, Phys. Rev. X **7**, 011015 (2017).

[44] J. P. Crutchfield, W. L. Ditto, and S. Sinha, Chaos **20**, 037101 (2010).

[45] C. S. Calude, *Unconventional Computing: A Brief Subjective History*, CDMTCS Report 480 (University Auckland, 2015).

[46] A. Adamatzky, ed., *Unconventional Computing*, Encyc. Complex. Syst. Sci. (Springer, New York, 2018).

[47] R. Legenstein and W. Maass, Neural Netw. **20**, 323 (2007).

[48] J. Dambre, D. Verstraeten, B. Schrauwen, and S. Massar, Sci. Rep. **2**, 514 (2012).

[49] L. Grigoryeva, J. Henriques, L. Larger, and J. P. Ortega, Sci. Rep. **5**, 12858 (2015).

[50] J. Pathak, Z. Lu, B. Hunt, M. Girvan, and E. Ott, Chaos **27**, 121102 (2017).

[51] Z. Lu, B. R. Hunt, and E. Ott, Chaos **28**, 061104 (2018).

[52] J. Pathak, B. Hunt, M. Girvan, Z. Lu, and E. Ott, Phys. Rev. Lett. **120**, 024102 (2018).

[53] J. Pathak, A. Wikner, R. Fussell, S. Chandra, B. Hunt, M. Girvan, and E. Ott, Chaos **28**, 041101 (2018).

[54] T. L. Carroll, Phys. Rev. E **98**, 052209 (2018).

[55] T. Weng, H. Yang, C. Gu, J. Zhang, and M. Small, Phys. Rev. E **99**, 042203 (2019).

[56] T. Pyragiene and K. Pyragas, Phys. Lett. A (2019).

[57] D. Ibáñez-Soria, J. Garcia-Ojalvo, A. Soria-Frisch, and G. Ruffini, Chaos **28**, 033118 (2018).

[58] B. A. Marquez, L. Larger, M. Jacquot, Y. K. Chembo, and D. Brunner, Sci. Rep. **8**, 3319 (2018).

[59] N. Oliver, T. Jüngling, and I. Fischer, Phys. Rev. Lett. **114**, 123902 (2015).

[60] T. Jüngling, M. C. Soriano, N. Oliver, X. Porte, and I. Fischer, Phys. Rev. E **97**, 042202 (2018).

[61] J. P. Eckmann and D. Ruelle, in *The Theory of Chaotic Attractors* (Springer, 1985) pp. 273–312.

[62] J. L. Kaplan and J. A. Yorke, in *Functional Differential Equations and Approximation of Fixed Points*, edited by H. O. Peitgen and H. O. Walther (Springer Berlin Heidelberg, Berlin, Heidelberg, 1979) pp. 204–227.

[63] J. Butcher, D. Verstraeten, B. Schrauwen, C. Day, and P. Haycock, Neural Netw. **38**, 76 (2013).

[64] S. Ortín, M. C. Soriano, L. Pesquera, D. Brunner, D. San-Martín, I. Fischer, C. R. Mirasso, and J. M. Gutiérrez, Sci. Rep. **5**, 14945 (2015).

[65] W. F. Schmidt, M. A. Kraaijveld, R. P. Duin, *et al.*, in *International Conference on Pattern Recognition* (IEEE COMPUTER SOCIETY PRESS, 1992) pp. 1–4.

[66] F. Takens, in *Dynamical systems and turbulence, Warwick 1980* (Springer, 1981) pp. 366–381.

[67] A. Uchida, K. Yoshimura, P. Davis, S. Yoshimori, and R. Roy, Phys. Rev. E **78**, 036203 (2008).

[68] T. Jüngling, X. Porte, N. Oliver, M. C. Soriano, and I. Fischer, IEEE J. Sel. Top. Quant. Electron. **25**, 1 (2019).

[69] A. E. Hramov and A. A. Koronovskii, Europhys. Lett. **70**, 169 (2005).

[70] O. I. Moskalenko, A. A. Koronovskii, and A. E. Hramov, Phys. Rev. E **87**, 064901 (2013).

[71] L. A. Aguirre and C. Letellier, Phys. Rev. E **83**, 066209 (2011).

[72] G. V. Osipov, B. Hu, C. Zhou, M. V. Ivanchenko, and J. Kurths, Phys. Rev. Lett. **91**, 024101 (2003).

[73] O. I. Moskalenko, A. A. Koronovskii, A. E. Hramov, and S. Boccaletti, Phys. Rev. E **86**, 036216 (2012).

[74]M. Mackey and L. Glass, Science **197**, 287 (1977).

[75]K. Ikeda, Optics Comm. **30**, 257 (1979).

[76]K. Ikeda, H. Daido, and O. Akimoto, Phys. Rev. Lett. **45**, 709 (1980).

[77]S. Heiligenthal, T. Jüngling, O. D'Huys, D. A. Arroyo-Almanza, M. C. Soriano, I. Fischer, I. Kanter, and W. Kinzel, Phys. Rev. E **88**, 012902 (2013).