Machine-Learning-Coined Noise Induces Energy-Saving Synchrony

Jingdong Zhang,[1, 2, 3] Luan Yang,[2] Qunxi Zhu,[2, 4, 5, *] Celso Grebogi,[3] and Wei Lin[1, 2, 4, 5, †]

[1]*School of Mathematical Sciences, SCMS, and SCAM, Fudan University, Shanghai 200433, China*
[2]*Research Institute of Intelligent Complex Systems, Fudan University, Shanghai 200433, China*
[3]*Institute for Complex Systems and Mathematical Biology, King's College,*
University of Aberdeen, Aberdeen AB24 3UE, United Kingdom
[4]*Shanghai Artificial Intelligence Laboratory, Shanghai 200232, China*
[5]*MOE Frontiers Center for Brain Science and State Key Laboratory*
of Medical Neurobiology, Fudan University, Shanghai 200032, China
(Dated: June 24, 2024)

Noise-induced synchronization is a pervasive phenomenon observed in a multitude of natural and engineering systems. Here, we devise a machine learning framework with the aim of devising noise controllers to achieve synchronization in diverse complex physical systems. We find the implicit energy regularization phenomenon of the formulated framework that engenders energy-saving artificial noise, and we rigorously elucidate the underlying mechanism driving this phenomenon. We substantiate the practical feasibility and efficacy of this framework by testing it across various representative systems of physical and biological significance, each influenced by distinct constraints reflecting real-world scenarios.

Introduction.—Noise-induced synchronization is a widespread phenomenon observed in various physical systems, encompassing from chaotic to limit-cycle oscillators [1–6]. Among the various paradigms of synchronization, complete synchronization (CS) has been extensively discussed in the presence of noise [7, 8]. Since Maritan and Banavar claimed that two chaotic systems subjected to the same and sufficiently strong noise can achieve synchronization [9], the mechanisms of common noise enhancing CS and its variants have been a highly relevant topic [4, 10, 11]. Apart from the common noise, the impact of uncorrelated and correlated noise on synchronization has garnered research attention recently [12, 13].

To reveal the complicated mechanism of noise-induced synchronization, previous work has primarily focused on investigating the local stability of the synchronization manifold [7, 14], or on exploring the global stability via analytically designing the Lyapunov function for the synchronization error dynamics [15, 16]. However, all these methods are system-specific and pose challenges in devising synchronous noise for general networked systems.

In this Letter, we devise and formulate the Artificial-Intelligence Noise (AIN) Synchrony, an inaugural and scalable framework for proficiently devising the noise controllers to attain the CS in diverse physical systems, integrating the stochastic stabilization theory with the machine learning techniques. Indeed, the devised AIN Synchrony is not only of mathematical rigorousness, but also applicable to stabilizing the synchronization manifold in both local and global manners. We illustrate its efficacy and practical feasibility using a wide range of representative systems, including chaotic and networked dynamics of limit-cycle oscillators. The results demonstrate that the machine-learning-coined noise can remarkably achieve energy-saving synchrony, which consumes low energy cost in the control process. We call this phenomenon as implicit energy regularization, and theoretically uncover the universal mechanism that produces it.

Problem setup and notations.—We consider the collective dynamics of coupled oscillators, expressed in a general form as:

$$\frac{\mathrm{d}\boldsymbol{x}_i}{\mathrm{d}t} = \boldsymbol{M}_0(\boldsymbol{x}_i, \boldsymbol{\mu}_0) + \sum_{j=1}^{n} A_{ij}\boldsymbol{M}_1(\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{\mu}_1), \quad (1)$$

where $\boldsymbol{x}_i \in \mathbb{R}^d$ $(i = 1, ..., n)$ is the oscillatory state of the i-th node, $\boldsymbol{M}_0$ represents the self-dynamics of oscillators and exhibits a (unstable) limit cycle or chaotic attractor, denoted by $\boldsymbol{s}$ and satisfying $\mathrm{d}\boldsymbol{s}/\mathrm{d}t = \boldsymbol{M}_0(\boldsymbol{s}, \boldsymbol{\mu}_0)$, $\boldsymbol{M}_1$ describes the i, j pairwise interaction, and $\boldsymbol{A} = (A_{ij}) \in \mathbb{R}^{n \times n}$ captures the interacting structure between the oscillators. Although the systematic parameters $\boldsymbol{\mu}_{0,1}$ characterizing the dynamics $\boldsymbol{M}_{0,1}$ may be potentially distributed across the system's components, in this Letter we focus on the common $\boldsymbol{\mu}_{0,1}$ requiring that there exists a synchronous manifold $\mathcal{M} = \{\boldsymbol{x}_i = \boldsymbol{s}, i = 1, ..., n\}$. We further require the coupling terms to be synchronization non-invasive, i.e., $\sum_{j=1}^{n} A_{ij}\boldsymbol{M}_1(\boldsymbol{s}, \boldsymbol{s}, \boldsymbol{\mu}_1) = \boldsymbol{0}$, which is satisfied when the interacting structure $\boldsymbol{A}$ is the Laplacian or the coupling function vanishes at the synchronous manifold $\boldsymbol{M}_1(\boldsymbol{s}, \boldsymbol{s}, \boldsymbol{\mu}_1) = \boldsymbol{0}$. We aim at designing only *noise controllers* such that: (1) The network achieves the CS physically, that is, $\boldsymbol{x}_i(t) \to \mathcal{M}$ $(t \to +\infty)$ for all i, (2) the noise can act flexibly to any experimentally feasible parts of a system under consideration, including the parameters $\boldsymbol{\mu}_{0,1}$, the interacting structure $\boldsymbol{A}$, and the external forces, (3) and the controller adapts to practical requirements from the real-world scenarios, such as pinning control, communication constraints, and the common or uncorrelated noise.

AIN Synchrony.—For brevity of presentation, we suc-

cinctly denote the collective dynamics (1) as $d\mathbf{x}/dt = \boldsymbol{F}(\mathbf{x}, \boldsymbol{\mu})$ with $\mathbf{x} = (\boldsymbol{x}_i^\top, ..., \boldsymbol{x}_n^\top)^\top$ and $\boldsymbol{\mu} = \{\boldsymbol{\mu}_0, \boldsymbol{\mu}_1\}$. Consider the noise controlled variational dynamics of $\boldsymbol{\xi} = \mathbf{x} - \boldsymbol{E} \otimes \boldsymbol{s}$ with zero solution $\boldsymbol{\xi} \equiv \boldsymbol{0}$ as follows,

$$d\boldsymbol{\xi} = [\boldsymbol{F}(\boldsymbol{\xi} + \boldsymbol{E} \otimes \boldsymbol{s}, \boldsymbol{\mu}) - \boldsymbol{M}_0(\boldsymbol{s}, \boldsymbol{\mu}_0)]\,dt + \boldsymbol{u}(\boldsymbol{\xi}, \boldsymbol{s})d\boldsymbol{B}_t,$$
$$\triangleq \boldsymbol{G}(\boldsymbol{\xi}, \boldsymbol{s})dt + \boldsymbol{u}(\boldsymbol{\xi}, \boldsymbol{s})d\boldsymbol{B}_t, \qquad (2)$$

where $\boldsymbol{E} = (1, ..., 1)^\top \in \mathbb{R}^n$, $\otimes$ is the Kronecker product, $\boldsymbol{u} \in \mathbb{R}^{nd \times r}$ is the state-dependent stochastic controller, and $\boldsymbol{B}_t$ is the r-dimensional (r-D) Brownian motion. The detailed formulation of how the noise controller acts on the parameters $\boldsymbol{\mu}$, the structure $\boldsymbol{A}$, and the external forces are provided in Supplemental Material (SM)-S1. We propose the stability theory for devising the diffusion term $\boldsymbol{u}$ to stabilize the variational dynamics (2). The proof of this theory is included in SM-S2.

Theorem 1.— Suppose that there exists a function $V \in C^{1,2}(\mathbb{R} \times \mathbb{R}^{nd}; \mathbb{R}_{\geq 0})$ such that $V(\boldsymbol{0}, t) = 0$, $V(\boldsymbol{\xi}, t) \geq c\|\boldsymbol{\xi}\|^p$ for constants $c, p > 0$, and

$$\frac{\left[\nabla V(\boldsymbol{\xi}, t)^\top \boldsymbol{u}(\boldsymbol{\xi}, \boldsymbol{s})\right]^2}{V(\boldsymbol{\xi}, t)^2} - b \cdot \frac{\mathcal{L}V(\boldsymbol{\xi}, t)}{V(\boldsymbol{\xi}, t)} \geq 0, \ \boldsymbol{\xi} \neq 0, \quad (3)$$

for some $b > 2$. Here, $\mathcal{L}$ represents Itô's derivative satisfying $\mathcal{L}V = V_t + \nabla V^\top \boldsymbol{G} + \mathrm{Tr}\left[\boldsymbol{u}^\top \nabla^2 V \boldsymbol{u}\right]/2$, ∇ represents the gradient operator with respect to $\boldsymbol{\xi}$, $\mathrm{Tr}[\cdot]$ represents the trace of a given matrix, and the limit $\lim_{\boldsymbol{x} \to \boldsymbol{0}} \frac{\|\nabla V(\boldsymbol{x}, t)^\top \boldsymbol{u}(\boldsymbol{x}, t)\|^2}{V(\boldsymbol{x}, t)^2} > 0$ holds. Then, for $k > 0$, we obtain $\limsup_{t \to +\infty} \frac{1}{t} \log \|\boldsymbol{\xi}(t)\| \leq -k\frac{b-2}{2bp}$ in a physical sense (i.e., with probability one).

Parameterization.—Manually seeking the functions pair V, $\boldsymbol{u}$ satisfying the above conditions including (3) is *impractical* due to the complexity and nonlinearity of the original dynamics $\boldsymbol{G}$. To address this, we introduce the machine learning techniques and devise an algorithm that leverages $\boldsymbol{G}$ and $\boldsymbol{s}$ to identify the expected functions pair $V_{\boldsymbol{\theta}}$ and $\boldsymbol{u}_{\boldsymbol{\phi}}$, where $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ represent the parameters of the neural networks to be trained. To enhance training efficacy, we establish the neural networks such that $V_{\boldsymbol{\theta}}$ and $\boldsymbol{u}_{\boldsymbol{\phi}}$ meet specific prerequisites. Specifically, we construct the function $V_{\boldsymbol{\theta}}$ as $V_{\boldsymbol{\theta}}(\boldsymbol{\xi}, t) = g_{\boldsymbol{\theta}}(\boldsymbol{\xi}, t) + \varepsilon\|\boldsymbol{\xi}\|^2$, where $g_{\boldsymbol{\theta}}$ is a second-order differentiable input convex neural network [17, 19] wherein $g(\boldsymbol{0}, t) = 0$ and ε is a small hyperparameter guaranteeing the positive definite lower bound of $V_{\boldsymbol{\theta}}$. To proceed, we parametrize a synchronous-noninvasive controller with $\boldsymbol{u}_{\boldsymbol{\phi}}(\boldsymbol{0}) = \boldsymbol{0}$ and limit the Lipschitz constant of the controller using the spectral norm regularization method [18, 20]. The detailed formulation of $V_{\boldsymbol{\theta}}$ and $\boldsymbol{u}_{\boldsymbol{\phi}}$ is provided in SM-S6. We theoretically validate the validity of the conditions assumed in Theorem 1 for the parameterized neural networks in SM-S2.

Loss function.—After parameterizing $V_{\boldsymbol{\theta}}$ and $\boldsymbol{u}_{\boldsymbol{\phi}}$, we need to ensure the controlled variational dynamics (2) satisfying the conditions established in (3). As such, we can attain the CS in the original collective dynamics (1) under noise control. To do so,

we heuristically devise the loss function as $L(\boldsymbol{\theta}, \boldsymbol{\phi}) = \frac{1}{m}\sum_{i=1}^m \left\{ \frac{b \cdot \mathcal{L}V_{\boldsymbol{\theta}}(\boldsymbol{\xi}_i, t_i)}{V_{\boldsymbol{\theta}}(\boldsymbol{\xi}_i, t_i)} - \frac{[\nabla V_{\boldsymbol{\theta}}(\boldsymbol{\xi}_i, t_i)^\top \boldsymbol{u}_{\boldsymbol{\phi}}(\boldsymbol{\xi}_i, \boldsymbol{s}(t_i))]^2}{V_{\boldsymbol{\theta}}(\boldsymbol{\xi}_i, t_i)^2} \right\}^+$, where $\{\boldsymbol{\xi}_i, \boldsymbol{s}(t_i), t_i\}_{i=1}^m$ is the training dataset and $\{\cdot\}^+$ denotes the operation of $\max(0, \cdot)$. To circumvent the drawback that training on the finite dataset may not guarantee the validity of the stability condition in Eq. (3) on the whole space, we further provide a stability guarantee theory to endow the stability guarantee to the current framework (More details refer to SM-S2).

Theorem 2.— With the functions specified in Theorem 1, denote by M the maximal Lipschitz constant of $\|\nabla V^\top \boldsymbol{u}\|^2$ and $\mathcal{L}V \cdot V$ on $\mathcal{D}$, where $\mathcal{D}$ is a bounded state space. Further denote by $\tilde{\mathcal{D}}$ the finite discretization of $\mathcal{D}$ with size r such that, for each $\boldsymbol{x} \in \mathcal{D}$, there exists $\tilde{\boldsymbol{x}} \in \tilde{\mathcal{D}}$ with $\|\boldsymbol{x} - \tilde{\boldsymbol{x}}\| < r$. If there exists a constant $0 \leq \delta \leq Mr$ such that $-\|\nabla V(\tilde{\boldsymbol{x}}, t)^\top \boldsymbol{u}(\tilde{\boldsymbol{x}}, t)\|^2 + b \cdot \mathcal{L}V(\tilde{\boldsymbol{x}}, t)V(\tilde{\boldsymbol{x}}, t) + (2 + b)Mr \leq \delta$ for all $\tilde{\boldsymbol{x}} \in \tilde{\mathcal{D}}\backslash\{\boldsymbol{0}\}$. Then, the controller $\boldsymbol{u}$ rigorously satisfies the stability condition in Eq. (3). Therefore, the learned noise controller rigorously satisfies the stability condition through replacing the training dataset by $\tilde{\mathcal{D}}$ and slightly modifying the loss function with the newly added term $(2 + b)Mr$.

Accelerating the training process.—The computational cost for computing the Hessian matrix in $\mathrm{Tr}[\boldsymbol{u}^\top \nabla^2 V_{\boldsymbol{\theta}} \boldsymbol{u}]$ is $\mathcal{O}(d^2)$, which hinders the framework from scaling to higher dimensional tasks. To reduce the computational cost, we establish an unbiased estimator as $\mathrm{Tr}\left[\boldsymbol{u}^\top \nabla^2 V_{\boldsymbol{\theta}} \boldsymbol{u}\right] = \mathbb{E}\left[(\nabla(\boldsymbol{\eta}^\top \nabla V_{\boldsymbol{\theta}}))^\top \boldsymbol{u}\boldsymbol{u}^\top \boldsymbol{\eta}\right]$, where $\boldsymbol{\eta}$ is a d-D noise vector with zero means, commonly referred to as Hutchinson's trace estimator [21]. We thus substitute this expectation representation with the Monte Carlo estimator, so that $\mathrm{Tr}\left[\boldsymbol{u}^\top \nabla^2 V_{\boldsymbol{\theta}} \boldsymbol{u}\right] \approx \frac{1}{m}\sum_{i=1}^m \left(\nabla(\boldsymbol{\eta}_i^\top \nabla V_{\boldsymbol{\theta}})\right)^\top \boldsymbol{u}\boldsymbol{u}^\top \boldsymbol{\eta}_i$ with m noise vectors during our training process. This approach reduces the computational cost from $\mathcal{O}(d^2)$ to $\mathcal{O}(md)$, improving the efficiency of the AIN Synchrony in the higher dimensional task, especially for $m \ll d$.

Except for considering the ground-truth variational dynamics (2), we can also apply the above-established machine learning framework to the traditional linearization equations $d\boldsymbol{\xi}/dt = [\boldsymbol{I}_n \otimes \nabla \boldsymbol{M}_0 + \boldsymbol{A} \otimes \nabla \boldsymbol{M}_1]\boldsymbol{\xi}$, which find extensive application in the master stability function theories for realizing CS [22, 23] (see SM-S3). In addition, as suggested by the stabilization theory, our framework can also be extended to the non-autonomous dynamics, such as the coupling matrix $\boldsymbol{A} = \boldsymbol{A}(t)$ varies temporally (see SM-S4 and SM-S6.4).

To proceed, we numerically validate the efficacy and flexibility of the AIN Synchrony using several representative physical models under the realistic constraints.

AIN Synchrony for driving-response Lorenz systems.— We investigate the unidirectionally-coupled Lorenz systems, where the driving and the response systems, respectively, are $d\boldsymbol{x}_1/dt = \boldsymbol{f}(\boldsymbol{x}_1, \sigma, \rho, \beta)$ and $d\boldsymbol{x}_2 = \boldsymbol{f}(\boldsymbol{x}_2, \sigma, \rho, \beta) + \boldsymbol{u}(\boldsymbol{\xi}, \boldsymbol{x}_1)dB_t$ with $\boldsymbol{x}_i = (x_i, y_i, z_i)^\top$, $\boldsymbol{\xi} =$
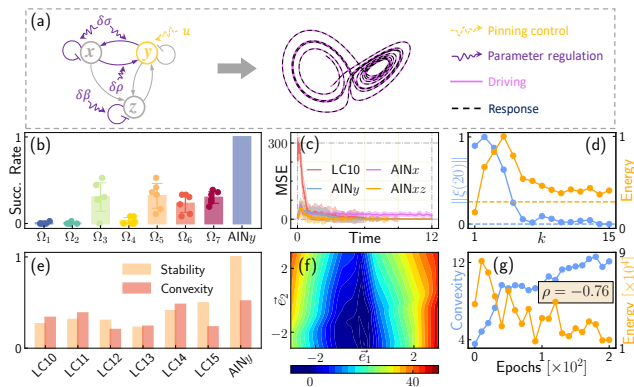
FIG. 1. Synchronizing the driving-response Lorenz systems. (a) Sketch on different noise controlling modes by the parameter regulation and by the external forces. (b) The success rate of 7 combinations of the regulated parameters, compared with the pinning controller AINy. The dots in the left 7 bars represent the success rate with $\boldsymbol{u}_{\max} \in \{1.0, 1.2, 1.4, 1.6, 1.8, 2.0\}$, where $\boldsymbol{u}_{\max} = \max_{\boldsymbol{x}} \|\boldsymbol{u}(\boldsymbol{x})\|_{\infty}$. (c) The MSE between the driving and the response systems under different controllers with the shaded region representing the variance. (d) The synchronization error at time 20 and the energy cost in the controlling process against the strength of the linear controller. The horizontal dashed lines represent the corresponding values of the learned controller AINy. (e) The stability of the controlled trajectories and the convexity of the learned V function of linear controllers compared with that of AINy. (f) The projection of $V_{\text{AIN}y} - V_{\text{LC}_{10}}$ to a random selected 2-D section. (g) The energy cost, the convexity of the V function for the AINy over the training process, and their Pearson correlation coefficient ρ.

$(\xi_1, \xi_2, \xi_3)^{\top} = (x_1 - x_2, y_1 - y_2, z_1 - z_2)^{\top}$, $\boldsymbol{f}(\boldsymbol{x}_i, \sigma, \rho, \beta) = [\sigma(y_i - x_i), \rho x_i - y_i - x_i z_i, x_i y_i - \beta z_i]^{\top}$, and $i = 1, 2$. The Lorenz system is characterized by three parameters σ, ρ, β, being proportional to the Prandtl number, the Rayleigh number, and certain physical dimensions of the layer itself, respectively, which are all experimentally adjustable [24]. In addition, we apply the *pinning control* on this system by partially controlling some variables of the system. As shown in Fig. 1(a), we explore both the parameters control and the pinning control for synchronizing the response system to the attractor produced by the driving Lorenz system. Firstly, we study whether the CS of the driving-response systems is achieved by solely regulating the parameters of the response system using noise. For instance, we examine the impact of adjusting the Prandtl number with noise, employing $\boldsymbol{u} = (\delta\sigma(\boldsymbol{\xi})(y_2 - x_2), 0, 0)^{\top}$ with $\delta\sigma(\boldsymbol{0}) = 0$ as the controller. We employ a loss function after training and the temporal average of the mean square error (MSE) between the driving system and the coupled response system as indicators to evaluate the controller performance. We consider totally 7 different combinations of the controlled parameters, including $\Omega_1 = \{\sigma\}$, $\Omega_2 = \{\rho\}$, $\Omega_3 = \{\beta\}$, $\Omega_4 = \{\sigma, \rho\}$, $\Omega_5 = \{\sigma, \beta\}$, $\Omega_6 = \{\rho, \beta\}$, $\Omega_7 = \{\sigma, \rho, \beta\}$, with a requirement for the scale of the learned noise $\boldsymbol{u}_{\max}$ to be less than 2 for realizing feasible simulations. In Fig. 1(b), we compare $\Omega_{1:7}$ with another pinning controller in the rate of successfully synchronizing the driving-response systems over 50 samples. The results indicate that the parameter β dominates the behavior of the Lorenz system among all these three parameters. Moreover, the findings imply that parameter regulation with noise is not always as effective as external forces, which agrees with the intuition.

Next, we delve into the intricacies of the task of pinning control. Initially, our focus centers on the pinning controllers that act on the nodes x_2 and y_2, denoted by AINx and AINy, respectively. As shown in Fig. 1(c), AINy succeeds in this task while AINx does not, this implies that node y exerts greater influence on the whole dynamics compared to the node x in the Lorenz system in the presence of the noise controller. Additionally, we identify the pinning controller on both x and z, denoted by AINxz, as efficacious. Furthermore, we compare the AINy with the stochastic linear controllers LCk, i.e., $\boldsymbol{u} = (0, k(\xi_1 + \xi_2 + \xi_3), 0)^{\top}$ [25], that only acts on y_2 using different strengths k $(= 1, ..., 15)$, in terms of the MSE and the energy cost. Figure 1(d) reveals that the AINy is significantly energy-saving than the linear LCk for $k \geq 6$ that successfully synchronizes the driving-response systems, hinting an *implicit energy regularization* in the AIN Synchrony. To gain more insights into the observed implicit energy regularization, we focus on the linear controller LCk and train $V_{\boldsymbol{\theta}_k}$ for each LCk within our framework. We consider the spatial convexity $\mathbb{E}_{\boldsymbol{\xi}}\left[\frac{1}{nd}\sum_{i=1}^{nd}\lambda_i(\nabla^2 V)\right]$ defined by the eigenvalues of the Hessian matrix of V. Numerical comparisons of the spatial convexity of $V_{\boldsymbol{\theta}_k}$ and $V_{\boldsymbol{\theta}_{\text{AIN}y}}$ are presented in Fig. 1(e), revealing that $V_{\boldsymbol{\theta}_{\text{AIN}y}}$ function of the AINy boasts the greatest convexity among successful controllers. This finding is further demonstrated in Fig. 1(f), where the random projection onto the 2-D surface of the section of $V_{\boldsymbol{\theta}_{\text{AIN}y}} - V_{\boldsymbol{\theta}_{\text{LC}_{10}}}$ exhibits a pronounced steepness. Figure 1(g) shows the anti-correlation between the energy cost and the convexity of V in the training process, which validates the implicit energy regularization from the numerical perspective. Furthermore, the following theorem elucidates the mechanism of these findings.

Theorem 3.— Consider the controlled dynamics in (2) with the AIN Synchrony controller $\boldsymbol{u}_{\boldsymbol{\phi}}$ learned according to the stability condition in Eq. (3). Then, the loss function in the training process is equivalent to the control energy in the control process by norm, $E = \mathbb{E}\int_0^T \left\|\boldsymbol{u}_{\boldsymbol{\phi}}^{\top}(\boldsymbol{x}(t), \boldsymbol{s}(t))Q\boldsymbol{u}_{\boldsymbol{\phi}}(\boldsymbol{x}(t), \boldsymbol{s}(t))\right\| \mathrm{d}t$, where Q is the variance matrix of the Brownian motion.

Consequently, as we minimize the loss function, the control energy is optimized as well, which thus results in the implicit energy regularization. More detailed demonstrations are included in SM-S5.

AIN Synchrony for subcritical Landau-Stuart oscillator.—Consider coupled Landau-Stuart oscillators governed by the complex-valued differential equations: $\dot{Z}_j =$
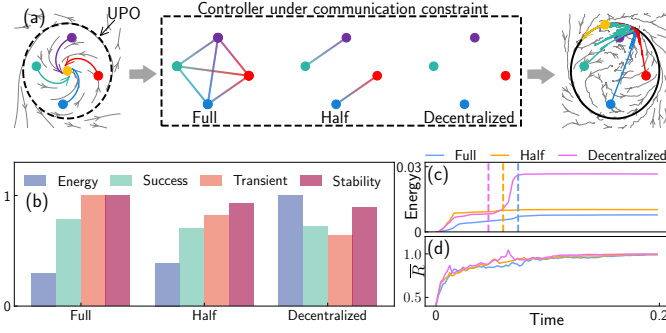
FIG. 2. Synchronizing the subcritical Landau-Stuart oscillators under different communication constraints. (a) Illustration of the CS task. (b) Bar plots of energy cost, succeeding rate, transient time $\tau_{0.1}$ and stability at three different communication modes for 50 sample trajectories, and the mean first passage-time $\tau_{0.1}$ of the controlled process. All the indices are normalized to $[0,1]$. (c) Energy cost and (d) success indicator $\overline{R}$ with time, where the dashed vertical lines represent the corresponding values of $\tau_{0.1}$.

$(\beta + \mathrm{i}\gamma + \mu|Z_j|^2)Z_j + \sum_{k=1}^n A_{jk}Z_k$ for $Z_j \in \mathbb{C}$, $j = 1, ..., n$, $\beta < 0$, and $\mu > 0$. Here, the self-dynamics undergoing the subcritical Andronov-Hopf bifurcation possesses an unstable periodic orbit (UPO) [26]. We focus on synchronizing the coupled oscillators to the UPO under several typical communication constraints, e.g., (i) all oscillators communicate, (ii) half of the oscillators communicate, and (iii) none of the oscillators communicate (resulting in a decentralized controller), as shown in Fig. 2(a).

We examine the impact of communication constraints on the energy cost, the success rate of stabilization and synchronization, the transient time from initial values to the synchronization manifold on the UPO, and the stability of the controllers. We utilize the order parameter $R_1 = \frac{1}{n}\sum_{j=1}^n e^{\mathrm{i}\theta_j}$ as the synchronization indicator, where θ_j is the argument of the complex variable Z_j. Additionally, we employ the distance $R_2 = \frac{1}{n}\sum_{j=1}^n |Z_j|$ between the UPO and the controlled orbits as the stabilization indicator. We set $\overline{R} = (R_1 + R_2)/2$ as the success indicator, and the transient time is assessed by the empirical expectation of the stopping time $\tau_p = \inf\{t \geq 0 : 1 - |\overline{R}(t)| = p\} \vee \inf\{t \geq 0 : |\overline{R}(t)| = 1 - p\}$ over the trajectories with a predefined accuracy p. The stability of the controllers is captured by the reciprocal empirical average of the standard variance of the trajectories. The results are comprehensively presented in Fig. 2(b). Although there is no significant difference in success rate and stability, we observed an interesting trade-off phenomenon that the energy cost increases with the decreased communication capability while the transient time decreases. This property is further demonstrated in Figs. 2(c)-2(d). The reason is that full communication optimally allocates control resources, achieving the task with smaller control costs. As a result, the reduced energy leads to the prolonged transient time since they have significant anti-correlation.

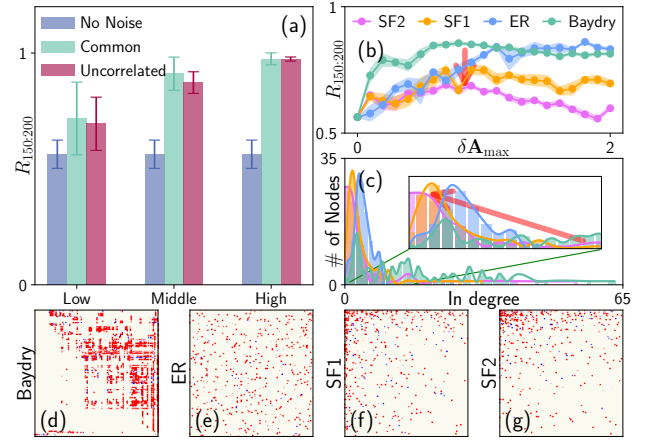*AIN Synchrony for networked neuronal dynamics.*—

FIG. 3. Synchronizing the coupled FitzHigh-Nagumo oscillators on four respective networks. (a) The order parameter R in the time interval $[150, 200]$ under external stochastic forces at three different types of noise intensities: Low ($\boldsymbol{u}_{\max} = 1.0$), middle ($\boldsymbol{u}_{\max} = 1.5$), and high ($\boldsymbol{u}_{\max} = 2.0$), for 10 realizations in each type of noise. Error bars indicate the variance. (b) $R_{150:200}$ of the noise perturbed system using the respective networks over the weight perturbed strength $\boldsymbol{\delta A}_{\max} \in [0, 2]$. (c) The in-degree distribution of the four networks, where the inset panel shows the difference among the peaks and the flatnesses of these networks. (d-g) The heat maps of the learned optimal weighted perturbation structure.

Finally, we synchronize the higher-dimensional networked dynamics using the coupled FitzHugh-Nagumo (FHN) neuronal oscillators with two modes of artificial noise, including direct driving force on each node and perturbing the coupling structure. The networked neuronal dynamics are described by: $\mathrm{d}v_i/\mathrm{d}t = (v_i - v_i^3/3 - w_i) + \sum_{j=1}^n L_{ij}/(1 + e^{-10v_j})$ and $\mathrm{d}w_i/\mathrm{d}t = 0.1(v_i + 0.7 - 0.8w_i)$, where $(L_{ij})_{n \times n} = (\delta_{ij}\sum_{j=1}^n A_{ij} - A_{ij})_{n \times n}$ is the Laplacian matrix of coupling matrix $\boldsymbol{A} = (A_{ij})_{n \times n}$ [27–29]. To quantify the synchronization of all the FHN oscillators, we employ the order parameter introduced in [30] as, $R = \frac{\langle M^2 \rangle - \langle M \rangle^2}{\frac{1}{N}\sum_{i=1}^N (\langle v_i^2 \rangle - \langle v_i \rangle^2)}$, where $M = (\sum_{i=1}^N v_i)/N$ is the mean field and $\langle \cdot \rangle$ represents the time average. As such, $R = 0$ corresponds to the asynchronous regime, while $R = 1$ indicates the CS state. For the external noise forces, we compare the effects of the *common* noise and the *uncorrelated* noise under different noise intensities in achieving the CS, we employ the Baydry, a realistic food-web network [31] with $n = 128$. As shown in Fig. 3(a), common noise gets a larger order parameter while uncorrelated noise exhibits smaller variance over multiple realizations in all noise scales, forming an interesting trade-off between performance and robustness. In the latter control mode, we perturb the structure with $\boldsymbol{\delta A}$ driven by the Brownian motion (see Table S1 in SM-S1 for the detailed formulation). We investigate the influence of network structure using four coupling matrices, viz., the Baydry, the directed Erdös and Rényi (ER) network, and two scale-free networks (denoted by SF1 and SF2) [32]. Fig. 3(b) shows the order parameter of the noise-perturbed dynamics under four networks against

the perturbation intensity $\boldsymbol{\delta A}_{\max} = \|\boldsymbol{\delta A}\|_\infty$. The CS performance of the realistic network significantly exceeds the others in the low intensity. As the intensity increases, the ER network outperforms Baydry foodweb slightly, and both outperform SF1, with SF1 surpassing SF2. To uncover the mechanism behind this phenomenon, we analyze the in-degree distribution plotted in Fig. 3(c). We find that the distribution of the Baydry foodweb is the most flat, having a heavy tail, while the peaks of the distributions ER, SF1, and SF2 tend towards zero. This indicates that the impact of noise decreases gradually as the homogeneity of the networks grows, as further demonstrated in Figs. 3(d)-3(g).

Conclusion.—We have conceived and formulated a machine learning framework to artificially generate noise controllers for synchronizing both general chaotic systems and limit-cycle systems. Harnessing the commutation property of the trace estimator, our framework scales to any higher dimensional systems with a linear-order computational cost. We applied our AIN Synchrony framework to successfully synchronize several representative systems under different constraints conforming to realistic scenarios. The results reveal that the machine-learning-coined noise has an implicit energy regularization phenomenon, inducing the energy-saving synchrony.

———————

[1] H. Nakao, K. Arai, and Y. Kawamura, Phy. Rev. Lett 98, 184101 (2007).

[2] R. F. Galán, N. Fourcaud-Trocmé, G. B. Ermentrout, and N. N. Urban, J. Neur. 26, 3646 (2006).

[3] R. V. Jensen, Phy. Eev. E 58, R6907 (1998).

[4] C.-H. Lai and C. Zhou, Eur. Lett. 43, 376 (1998).

[5] C. Zhou and J. Kurths, Chaos 13, 401 (2003).

[6] J.-n. Teramae and D. Tanaka, Phys. Rev. Lett. 93, 204103 (2004).

[7] C. Zhou and J. Kurths, Phys. Rev. Lett. 88, 230602 (2002).

[8] C. Zhou, J. Kurths, I. Z. Kiss, and J. L. Hudson, Phys. Rev. Lett. 89, 014101 (2002).

[9] A. Maritan and J. R. Banavar, Phys. Rev. Lett. 72, 1451 (1994).

[10] R. Toral, C. R. Mirasso, E. Hernández-García, and O. Piro, Chaos 11, 665 (2001).

[11] A. S. Pikovsky, Phys. Rev. Lett. 73, 2931 (1994).

[12] J. H. Meng and H. Riecke, Sci. Rep. 8, 6949 (2018).

[13] S. Martineau, T. Saffold, T. T. Chang, and H. Ronellenfitsch, Physical Review Letters 128, 098301 (2022).

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

[15] W. Lin and Y. He, Chaos 15, 023705 (2005).

[16] W. Lin and G. Chen, Chaos 16 (2006).

[17] J. Zhang, Q. Zhu, and W. Lin, in *Advances in Neural Information Processing Systems*, Vol. 35 (2022) pp. 9098–9110.

[18] J. Zhang, Q. Zhu, W. Yang, and W. Lin, in *International Conference on Learning Representations* (2023).

[19] B. Amos, L. Xu, and J. Z. Kolter, in *International Conference on Machine Learning* (PMLR, 2017) pp. 146–155.

[20] Y. Yoshida and T. Miyato, arXiv (2017).

[21] M. F. Hutchinson, Commun. Stat. Simul. Comput. 18, 1059 (1989).

[22] L. M. Pecora and T. L. Carroll, Phys. Rev. Lett. 80, 2109 (1998).

[23] A. Ghosh, H. Behl, E. Dupont, P. Torr, and V. Namboodiri, Advances in Neural Information Processing Systems 33 (2020).

[24] M. Gorman, P. Widmann, and K. Robbins, Physica D: Nonlinear Phenomena 19, 255 (1986).

[25] We also tried the linear controller $\boldsymbol{u} = (0, k\xi_2, 0)^\top$, but it failed to synchronize the system for any strength k, see Section S6.1 of SM for more detailed discussion.

[26] Y. A. Kuznetsov, Scholarpedia 1, 1858 (2006).

[27] K. Wang, L. Yang, S. Zhou, and W. Lin, Chaos 33 (2023).

[28] C. Conley and J. Smoller, J. Differ. Equ. 63, 389 (1986).

[29] J. Keener and J. Sneyd, *Mathematical Physiology* (Springer-Verlag, New York, 1998).

[30] J. Garcia-Ojalvo, M. B. Elowitz, and S. H. Strogatz, Proc. Natl. Acad. Sci 101, 10955 (2004).

[31] Y.-Z. Sun, S.-Y. Leng, Y.-C. Lai, C. Grebogi, and W. Lin, Phys. Rev. Lett. 119, 198301 (2017).

[32] K.-I. Goh, B. Kahng, and D. Kim, Phys. Rev. Lett. 87, 278701 (2001).

Supplementary Material for *"Machine-Learning-Coined Noise Induces Energy-Saving Synchrony "*

Jingdong Zhang[1,2,3], Luan Yang[2], Qunxi Zhu[2,4,5,*], Celso Grebogi[3] and Wei Lin[1,2,4,5,*]

[1]*School of Mathematical Sciences, SCMS, and SCAM, Fudan University, Shanghai 200433, China*

[2]*Research Institute of Intelligent Complex Systems, Fudan University, Shanghai 200433, China*

[3]*Institute for Complex Systems and Mathematical Biology, King's College,*
University of Aberdeen, Aberdeen AB24 3UE, United Kingdom

[4]*Shanghai Artificial Intelligence Laboratory, Shanghai 200232, China and*

[5]*MOE Frontiers Center for Brain Science and State Key Laboratory*
of Medical Neurobiology, Fudan University, Shanghai 200032, China

(Dated: May 23, 2024)

CONTENTS

---

* To whom correspondence should be addressed: qxzhu16@fudan.edu.cn; wlin@fudan.edu.cn

S1.   DETAILED FORMULATION OF FLEXIBLE NOISE CONTROL

In this section, we give an explicit description on how the noise controller acts flexibly on the experimentally feasible parts of the original system, including (i) regulation of the systematic parameters, (ii) perturbation to the interacting structure, and (iii) addition of the external forces. Recall the original collective dynamics as follows,

$$\frac{\mathrm{d}\boldsymbol{x}_i}{\mathrm{d}t} = \boldsymbol{M}_0(\boldsymbol{x}_i, \boldsymbol{\mu}_0) + \sum_{j=1}^{n} A_{ij} \boldsymbol{M}_1(\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{\mu}_1). \tag{S1.1}$$

In the main text, the nonlinear controlled variational dynamics (S1.1) can be written as the following compact concatenated vector form:

$$\begin{aligned}
\mathbf{x} &= (\boldsymbol{x}_i^\top, \cdots, \boldsymbol{x}_n^\top)^\top \in \mathbb{R}^{nd}, \quad \boldsymbol{\xi} = \mathbf{x} - \boldsymbol{E} \otimes \boldsymbol{s}, \\
\mathrm{d}\mathbf{x} &= \boldsymbol{F}(\mathbf{x}, \boldsymbol{\mu}, t)\mathrm{d}t, \\
\mathrm{d}\boldsymbol{\xi} &= \left[\boldsymbol{F}(\boldsymbol{\xi} + \boldsymbol{E} \otimes \boldsymbol{s}, \boldsymbol{\mu}, t) - \boldsymbol{M}_0(\boldsymbol{s}, \boldsymbol{\mu}_0) \right] \mathrm{d}t + \boldsymbol{u}(\boldsymbol{\xi}, \boldsymbol{s})\mathrm{d}\boldsymbol{B}_t, \\
&\triangleq \boldsymbol{G}(\boldsymbol{\xi}, \boldsymbol{s}, t)\mathrm{d}t + \boldsymbol{u}(\boldsymbol{\xi}, \boldsymbol{s})\mathrm{d}\boldsymbol{B}_t.
\end{aligned} \tag{S1.2}$$

In order to clarify the controlling formulation, we consider the oscillator-wise controlled variational dynamics as

$$\mathrm{d}\boldsymbol{\xi}_i = \left[\boldsymbol{M}_0(\boldsymbol{\xi}_i + \boldsymbol{s}, \boldsymbol{\mu}_0) - \boldsymbol{M}_0(\boldsymbol{s}, \boldsymbol{\mu}_0) + \sum_{j=1}^{n} A_{ij} \boldsymbol{M}_1(\boldsymbol{\xi}_i + \boldsymbol{s}, \boldsymbol{\xi}_j + \boldsymbol{s}, \boldsymbol{\mu}_1) \right] \mathrm{d}t + \boldsymbol{u}_i(\boldsymbol{\xi}, \boldsymbol{s})\mathrm{d}\boldsymbol{B}_i(t), \tag{S1.3}$$

with $\boldsymbol{u}\mathrm{d}\boldsymbol{B}(t) = \left[(\boldsymbol{u}_1\mathrm{d}\boldsymbol{B}_1(t))^\top, \cdots, (\boldsymbol{u}_n\mathrm{d}\boldsymbol{B}_n(t))^\top \right]^\top$. We now give the specific expression of the diffusion term $\boldsymbol{u}_i$ in all cases mentioned above.

(i) We regulate the systematic parameters $\boldsymbol{\mu}_{0,1}$ of ith oscillator with artificial noise denoted by $(\boldsymbol{\delta\mu}_{0,1})_i$, where $(\boldsymbol{\delta\mu}_{0,1})_i$ represents the scale of the noise. We require the noise control is non-invasive so that the system does not need extra control to sustain the synchronization state, so we let the scale be closed-loop and vanish at zero, that is, $(\boldsymbol{\delta\mu}_{0,1})_i = [(\boldsymbol{\delta\mu}_{0,1})_{\boldsymbol{\phi}}]_i(\boldsymbol{\xi}, \boldsymbol{s})$, $[(\boldsymbol{\delta\mu}_{0,1})_{\boldsymbol{\phi}}]_i(\boldsymbol{0}, \boldsymbol{s}) = \boldsymbol{0}$, parameterized by the neural networks with parameters $\boldsymbol{\phi}$. In this way, the diffusion term in the SDE (S1.3) takes the following form:

$$\boldsymbol{u}_i(\boldsymbol{\xi}, \boldsymbol{s}) = \boldsymbol{M}_0(\boldsymbol{\xi}_i + \boldsymbol{s}, \boldsymbol{\mu}_0 + \boldsymbol{\delta\mu}_0^i(\boldsymbol{\xi}, \boldsymbol{s})) - \boldsymbol{M}_0(\boldsymbol{s}, \boldsymbol{\mu}_0 + \boldsymbol{\delta\mu}_0^i(\boldsymbol{\xi}, \boldsymbol{s})) + \sum_{j=1}^{n} A_{ij} \boldsymbol{M}_1(\boldsymbol{\xi}_i + \boldsymbol{s}, \boldsymbol{\xi}_j + \boldsymbol{s}, \boldsymbol{\mu}_1 + \boldsymbol{\delta\mu}_1^j(\boldsymbol{\xi}, \boldsymbol{s})).$$

(ii) If we perturb the weighted coupling structure $\boldsymbol{A} = (A_{ij})$ with noise scale $\boldsymbol{\delta A} = (\delta A_{ij})$, under the same non-invasive principle, we have two kinds of choices: One is the constant matrix with Laplacian form, i.e., $\sum_j \delta A_{ij} = 0$ for each row i; the other is the closed-loop type control vanishing at zero, i.e., $\delta A_{ij} = \delta A_{ij}(\boldsymbol{\xi})$, $\delta A_{ij}(\boldsymbol{0}) = \boldsymbol{0}$. In reality, changing the scale of the perturbation of each link continuously is computationally infeasible and energy-consuming, so in this work, we focus on the former type of non-invasive perturbation. In this way, the diffusion term is

$$\boldsymbol{u}_i(\boldsymbol{\xi}, \boldsymbol{s}) = \sum_{j=1}^{n} \delta A_{ij} \boldsymbol{M}_1(\boldsymbol{\xi}_i + \boldsymbol{s}, \boldsymbol{\xi}_j + \boldsymbol{s}, \boldsymbol{\mu}_1).$$

Here we set $\boldsymbol{\delta A}$ as a trainable matrix and our framework can find the optimal perturbation structure under limited resources such as $\delta A_{ij} \leq \gamma$ for some prescribed threshold γ, as shown in the Fig. 3(h-k) in the main text.

(iii) Finally, for the direct external forces, the diffusion term has no explicit dependence on the original dynamics and systematic parameters. We simply set $\boldsymbol{u}_i$ as the output of the neural network, to take realistic factors into consideration, we truncate each component of $\boldsymbol{u}_i$ such that it stays in some practically feasible set $[-\gamma, \gamma]$. We give a detailed formulation of this kind of controller in the next section.

After we train the neural networks and find the optimal noise controllers with the variational equation (S1.3), we apply the controllers to the original dynamics as

$$\mathrm{d}\boldsymbol{x}_i = \left[\boldsymbol{M}_0(\boldsymbol{x}_i, \boldsymbol{\mu}_0) + \sum_{j=1}^{n} A_{ij} \boldsymbol{M}_1(\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{\mu}_1) \right] \mathrm{d}t + \boldsymbol{u}_i(\mathbf{x} - \boldsymbol{E} \otimes \boldsymbol{s}, \boldsymbol{s})\mathrm{d}\boldsymbol{B}_i(t). \tag{S1.4}$$

Appendix on formulation of noise controllers.–We present the explicit formulation of how the noise controller acts flexibly to the experimentally feasible parts of the original system, including (i) regulating the systematic parameters, (ii) perturbing the interacting structure, and (iii) directly adding the external forces in Tabs. S1. For reference, the controlled systems obey the following dynamics:

$$d\boldsymbol{x}_i = \left[\boldsymbol{M}_0(\boldsymbol{x}_i, \boldsymbol{\mu}_0) + \sum_{j=1}^{n} A_{ij} \boldsymbol{M}_1(\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{\mu}_1) \right] dt + \boldsymbol{u}_i(\boldsymbol{\xi}, \boldsymbol{s}) d\boldsymbol{B}_i(t).$$

We also present how the controller adapts to practical constraints from the real-world scenarios in Tabs. S2. As an aside, although we only list the conditions considered in this Letter, the AIN Synchrony has the ability to adapt to any other constraints because our framework proposes a general rule for learning noise controllers, which is independent of the realistic constraints.

TABLE S1. Detailed formulation of the noise controller in different cases. Items with subscript $\boldsymbol{\phi}$ represent the output of the neural network controller with trainable parameters $\boldsymbol{\phi}$.

Controlled parts	Controller $\boldsymbol{u}_i(\boldsymbol{\xi}, \boldsymbol{s})$	Learning Target (Type)
Parameter regulation	$\boldsymbol{M}_0(\boldsymbol{\xi}_i + \boldsymbol{s}, \boldsymbol{\mu}_0 + [(\boldsymbol{\delta\mu}_0)_{\boldsymbol{\phi}}]_i(\boldsymbol{\xi}, \boldsymbol{s})) - \boldsymbol{M}_0(\boldsymbol{s}, \boldsymbol{\mu}_0 + [(\boldsymbol{\delta\mu}_0)_{\boldsymbol{\phi}}]_i(\boldsymbol{\xi}, \boldsymbol{s}))$ $+ \sum_{j=1}^{n} A_{ij} \boldsymbol{M}_1(\boldsymbol{\xi}_i + \boldsymbol{s}, \boldsymbol{\xi}_j + \boldsymbol{s}, \boldsymbol{\mu}_1 + [(\boldsymbol{\delta\mu}_1)_{\boldsymbol{\phi}}]_j(\boldsymbol{\xi}, \boldsymbol{s}))$	$(\boldsymbol{\delta\mu}_0)_{\boldsymbol{\phi}}(\boldsymbol{\xi}, \boldsymbol{s})$ (Feedback function) $(\boldsymbol{\delta\mu}_1)_{\boldsymbol{\phi}}(\boldsymbol{\xi}, \boldsymbol{s})$ (Feedback function)
Structure perturbation	$\sum_{j=1}^{n} [(\delta\boldsymbol{A})_{\boldsymbol{\phi}}]_{ij} \boldsymbol{M}_1(\boldsymbol{\xi}_i + \boldsymbol{s}, \boldsymbol{\xi}_j + \boldsymbol{s}, \boldsymbol{\mu}_1)$	$(\delta\boldsymbol{A})_{\boldsymbol{\phi}}$ (Constant matrix)
External forces	$(\boldsymbol{u}_{\boldsymbol{\phi}})_i(\boldsymbol{\xi}, \boldsymbol{s})$	$\boldsymbol{u}_{\boldsymbol{\phi}}(\boldsymbol{\xi}, \boldsymbol{s})$ (Feedback function)

TABLE S2. Introduction of the constraints from the real-world scenarios.

Constraints (reference)	Description	Implementation
Pinning control on intra-structure ([1])	Controller acts on varying component set $N_i \subset \{x_i^1, ..., x_i^d\}$ for node $\boldsymbol{x}_i$	Fig. 1
Pinning control on inter-structure ([2])	Controller acts on partial node set $\mathbf{N} \subset \{\boldsymbol{x}_1, ..., \boldsymbol{x}_n\}$	Fig. S2 in SM
Communication constraints ([3])	Input varying node set $\mathbf{N}_i \subset \{\boldsymbol{x}_1, ..., \boldsymbol{x}_n\}$ to each $\boldsymbol{u}_i$	Fig. 2
Common noise ([4])	$\boldsymbol{u}_i$ is driven by the same Brownian motion $\boldsymbol{B}_i = \boldsymbol{B}$	Fig. 3
Uncorrelated noise ([5])	$\boldsymbol{u}_i$ is driven by the correlated Brownian motion with $\mathrm{Cov}(\boldsymbol{B}_1, ..., \boldsymbol{B}_n) = \mathbf{I}_{n \times n}$	Fig. 3
Varying connection ([6])	Inter-structure $\boldsymbol{A}$ is time varying	Fig. S1 in SM

S2. STOCHASTIC STABILIZATION THEORY

In this section, we prove the main stochastic stabilization theorem used in our framework. The main idea is inspired by the stochastic stabilization theory in [7]. The theorem is formally proposed as follows,

Theorem S2.1 *Consider the following controlled SDE:*

$$\mathrm{d}\boldsymbol{x}(t) = \boldsymbol{G}(\boldsymbol{x}(t), t)\mathrm{d}t + \boldsymbol{u}(\boldsymbol{x}(t), t)\mathrm{d}\boldsymbol{B}_t, \quad t \geq 0, \ \boldsymbol{x}(0) = \boldsymbol{x}_0 \in \mathbb{R}^d. \tag{S2.5}$$

We let $\boldsymbol{G}(\boldsymbol{0}, t) = \boldsymbol{0}$, $\boldsymbol{u}(\boldsymbol{0}, t) = \boldsymbol{0}$ such that $\boldsymbol{0}$ is the equilibrium of both the original system and the controlled system. To guarantee the existence and uniqueness of the SDE (S2.5), we assume that the drift and diffusion terms satisfy the locally Lipschitzian continuity, that is, for every integer $n \geq 1$, there is a number $K_n > 0$ such that

$$\|\boldsymbol{G}(\boldsymbol{x}, t) - \boldsymbol{G}(\boldsymbol{y}, t)\| \leq K_n \|\boldsymbol{x} - \boldsymbol{y}\|, \ \|\boldsymbol{u}(\boldsymbol{x}) - \boldsymbol{u}(\boldsymbol{y})\|_{\mathrm{F}} \leq K_n \|\boldsymbol{x} - \boldsymbol{y}\|,$$

for any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$ with $\|\boldsymbol{x}\| \vee \|\boldsymbol{y}\| \leq n$, where $\|\cdot\|_{\mathrm{F}}$ represents the Frobenius norm. Suppose the following conditions hold simultaneously:

(i) the state space of the SDE (S2.5) is a prescribed bounded region, denoted by Ω,

(ii) there exists a function $V \in C^{1,2}(\mathbb{R} \times \mathbb{R}^d; \mathbb{R})$ such that $V(\boldsymbol{0}, t) = 0$, $V(\boldsymbol{x}, t) \geq c\|\boldsymbol{x}\|^p$ for some constants $c, p > 0$, and $\lim_{\boldsymbol{x} \to \boldsymbol{0}} \dfrac{\|\nabla V(\boldsymbol{x}, t)^\top \boldsymbol{u}(\boldsymbol{x}, t)\|^2}{V(\boldsymbol{x}, t)^2} > 0$, and

(iii) the following inequality holds,

$$\frac{\|\nabla V(\boldsymbol{x}, t)^\top \boldsymbol{u}(\boldsymbol{x}, t)\|^2}{V(\boldsymbol{x}, t)^2} - b \cdot \frac{\mathcal{L}V(\boldsymbol{x}, t)}{V(\boldsymbol{x}, t)} \geq 0, \ \boldsymbol{x} \neq 0, \tag{S2.6}$$

for some $b > 2$.

Then we have

$$\limsup_{t \to \infty} \frac{1}{t} \log \|\boldsymbol{x}(t; \boldsymbol{x}_0)\| \leq -k \frac{b-2}{2bp} \quad a.s., \tag{S2.7}$$

for some constant $k > 0$.

Proof. *By applying Itô's formula [8] to $\log V(\boldsymbol{x}(t), t)$ we have*

$$
\begin{aligned}
\mathrm{d}\log(V(\boldsymbol{x}(t), t)) &= \frac{\mathrm{d}V(\boldsymbol{x}(t), t)}{V(\boldsymbol{x}(t), t)} - \frac{\mathrm{d}V(\boldsymbol{x}(t), t)^\top \mathrm{d}V(\boldsymbol{x}(t), t)}{2V(\boldsymbol{x}(t), t)^2} \\
&= \frac{\nabla V(\boldsymbol{x}(t), t)^\top \mathrm{d}\boldsymbol{x}(t) + \frac{1}{2}\mathrm{Tr}[\mathrm{d}\boldsymbol{x}(t)\nabla^2 V(\boldsymbol{x}(t), t)\mathrm{d}\boldsymbol{x}(t)]}{V(\boldsymbol{x}(t), t)} - \frac{(\nabla V(\boldsymbol{x}(t), t)^\top \mathrm{d}\boldsymbol{x}(t))^2}{2V(\boldsymbol{x}(t), t)^2} \\
&= \left[\frac{\mathcal{L}V(\boldsymbol{x}(t), t)}{V(\boldsymbol{x}(t), t)} - \frac{(\nabla V(\boldsymbol{x}(t), t)^\top \boldsymbol{u}(\boldsymbol{x}(t), t))^2}{2V(\boldsymbol{x}(t), t)^2} \right] \mathrm{d}t + \frac{\nabla V(\boldsymbol{x}(t), t)^\top \boldsymbol{u}(\boldsymbol{x}(t), t)^2}{V(\boldsymbol{x}(t), t)^2} \mathrm{d}\boldsymbol{B}_t \\
&= \left[\frac{\mathcal{L}V(\boldsymbol{x}(t), t)}{V(\boldsymbol{x}(t), t)} - \frac{(\nabla V(\boldsymbol{x}(t), t)^\top \boldsymbol{u}(\boldsymbol{x}(t), t))^2}{bV(\boldsymbol{x}(t), t)^2} + (\frac{1}{b} - \frac{1}{2})\frac{(\nabla V(\boldsymbol{x}(t), t)^\top \boldsymbol{u}(\boldsymbol{x}(t), t))^2}{V(\boldsymbol{x}(t), t)^2} \right] \mathrm{d}t \\
&\quad + \frac{\nabla V(\boldsymbol{x}(t), t)^\top \boldsymbol{u}(\boldsymbol{x}(t), t)^2}{V(\boldsymbol{x}(t), t)^2} \mathrm{d}\boldsymbol{B}_t \\
&\leq -\delta \frac{(\nabla V(\boldsymbol{x}(t), t)^\top \boldsymbol{u}(\boldsymbol{x}(t), t))^2}{V(\boldsymbol{x}(t), t)^2} \mathrm{d}t + \frac{\nabla V(\boldsymbol{x}(t), t)^\top \boldsymbol{u}(\boldsymbol{x}(t), t)^2}{V(\boldsymbol{x}(t), t)^2} \mathrm{d}\boldsymbol{B}_t,
\end{aligned}
\tag{S2.8}
$$

where $\delta = \frac{1}{2} - \frac{1}{b} > 0$. Denote by $M(t)$ the martingale $\int_0^t \frac{\nabla V(\boldsymbol{x}(s), s)^\top \boldsymbol{u}(\boldsymbol{x}(s), s)^2}{V(\boldsymbol{x}(s), s)^2} \mathrm{d}\boldsymbol{B}_s$. Then, by virtue of the exponential martingale inequality [7], we have

$$\mathbb{P}\left\{ \sup_{\leq t \leq n} \left[M(t) - \frac{\varepsilon}{2} \int_0^t \frac{(\nabla V(\boldsymbol{x}(s), s)^\top \boldsymbol{u}(\boldsymbol{x}(s), s))^2}{V(\boldsymbol{x}(s), s)^2} \mathrm{d}s \right] > \frac{2}{\varepsilon} \log n \right\} \leq \frac{1}{n^2},$$

for any $\varepsilon \in (0,1)$ and integer $n \in \mathbb{Z}^+$. According to the Borel-Cantelli lemma [9], we know that there exists an integer n_0 such that the following inequality

$$M(t) \leq \frac{2}{\varepsilon} \log n + \frac{\varepsilon}{2} \int_0^t \frac{(\nabla V(\boldsymbol{x}(s), s)^\top \boldsymbol{u}(\boldsymbol{x}(s), s))^2}{V(\boldsymbol{x}(s), s)^2} \mathrm{d}s, \ n \geq n_0, \ 0 \leq t \leq n$$

holds almost surely. Combining with Eq. (S2.8), we have

$$\log V(\boldsymbol{x}(t), t) \leq -\left(\delta - \frac{\varepsilon}{2}\right) kt + \frac{2}{\varepsilon} \log n + \log V(\boldsymbol{x}_0),$$

where $k = \inf_{\boldsymbol{x} \in \Omega} \frac{\|\nabla V(\boldsymbol{x},t)^\top \boldsymbol{u}(\boldsymbol{x},t)\|^2}{V(\boldsymbol{x},t)^2} \geq 0$. From the condition (ii) and the smoothness of the functions V and $\boldsymbol{u}$, it follows that there exists $\gamma > 0$, such that

$$\inf_{\boldsymbol{x} \in \mathcal{O}(\gamma)} \frac{\|\nabla V(\boldsymbol{x},t)^\top \boldsymbol{u}(\boldsymbol{x},t)\|^2}{V(\boldsymbol{x},t)^2} \geq \frac{1}{2} \lim_{\boldsymbol{x} \to \boldsymbol{0}} \frac{\|\nabla V(\boldsymbol{x},t)^\top \boldsymbol{u}(\boldsymbol{x},t)\|^2}{V(\boldsymbol{x},t)^2} > 0,$$

$$\inf_{\boldsymbol{x} \notin \mathcal{O}(\gamma)} \frac{\|\nabla V(\boldsymbol{x},t)^\top \boldsymbol{u}(\boldsymbol{x},t)\|^2}{V(\boldsymbol{x},t)^2} > 0.$$

Then, we have

$$k = \min \left\{ \inf_{\boldsymbol{x} \in \mathcal{O}(\gamma)} \frac{\|\nabla V(\boldsymbol{x},t)^\top \boldsymbol{u}(\boldsymbol{x},t)\|^2}{V(\boldsymbol{x},t)^2}, \inf_{\boldsymbol{x} \notin \mathcal{O}(\gamma)} \frac{\|\nabla V(\boldsymbol{x},t)^\top \boldsymbol{u}(\boldsymbol{x},t)\|^2}{V(\boldsymbol{x},t)^2} \right\} > 0.$$

This thus implies that

$$\frac{1}{t} \log V(\boldsymbol{x}(t), t) \leq -\left(\delta - \frac{\varepsilon}{2}\right) k + \frac{1}{t} \left(\frac{2}{\varepsilon} \log n + \log V(\boldsymbol{x}_0)\right)$$

$$\leq -\left(\delta - \frac{\varepsilon}{2}\right) k + \frac{1}{n-1} \left(\frac{2}{\varepsilon} \log n + \log V(\boldsymbol{x}_0)\right), \ n - 1 \leq t \leq n.$$

Finally, letting $t \to \infty$ and $\varepsilon \to 0$ yields

$$\lim_{t \to \infty} \frac{1}{t} \log V(\boldsymbol{x}(t), t) \leq -\delta k.$$

This consequently implies that

$$\lim_{t \to \infty} \frac{1}{t} \|\boldsymbol{x}(t)\| \leq -\frac{\delta k}{p} = -\frac{b-2}{2bp} k,$$

which therefore completes the proof.

Now, we explain why the proposed framework in the main text satisfies the conditions (i) and (ii) required in the Theorem S2.1. Firstly, if we consider the dynamics with a chaotic or periodic attractor, the variational dynamics between the original oscillators and the manifold is bounded due to the boundedness of the attractors. If we consider the unstable limit-cycle oscillators, we only focus on synchronizing the oscillators initiating inside the limit-cycle so the state space is bounded by some region taking the limit-cycle as the border. Therefore, the state space of the variations defined in the main text is always bounded, which meets the condition (i). Secondly, we construct the V function as a strictly convex function according to the ICNN [10]. Taking $d = 1$ for an example, by L'Hôpital rule, we have

$$\lim_{x \to 0} \frac{V'(x,t)u(x)}{V(x,t)} = \lim_{x \to 0} \frac{V''(x,t)u(x) + V'(x,t)u'(x)}{V'(x,t)} = u'(0).$$

So we have

$$\lim_{x \to 0} \frac{(V'(x,t)u(x))^2}{V^2(x,t)} = u'(0)^2.$$

The similar relation holds true for $d > 1$, since we use $\mathrm{Tanh}(\cdot)$ and $\mathrm{ReLU}(\cdot)$ as the activation functions in the parametrization of the controller $\boldsymbol{u}$; this controller $\boldsymbol{u}$ has linear order in the vicinity of $\boldsymbol{0}$ and hence $\nabla \boldsymbol{u}(\boldsymbol{0}) \neq \boldsymbol{0}$. These thus guarantee the validity of the condition (ii) in our framework.

Discretization and Stability Guarantee. Next, we provide a stability guarantee theory for our stabilization theory, in a way that we can improve the machine learning framework such that the learned controller satisfies the stability condition (*iii*) in Theorem S2.1 in the whole state space, and therefore ensure the rigorous stability for the controlled system. Our idea is roughly based on the finite covering of state space and the local upper bound estimation ensuing from the Lipschitz condition.

Theorem S2.2 *(Stability Guarantee Theory) With the functions specified in Theorem S2.1, denote by $M = \mathcal{M}(\boldsymbol{G}, \boldsymbol{u}, V, \mathcal{D})$ the maximum of the Lipschitz constants of $\|\nabla V^\top \boldsymbol{u}\|^2$ and $\mathcal{L}VV$ on $\mathcal{D}$, where $\mathcal{D}$ is a bounded state space. Also, denote by $\tilde{\mathcal{D}}$ the mesh grid discretization of $\mathcal{D}$ and by r the mesh size, such that, for each $\boldsymbol{x} \in \mathcal{D}$, there exists $\tilde{\boldsymbol{x}} \in \tilde{\mathcal{D}}$ with $\|\boldsymbol{x} - \tilde{\boldsymbol{x}}\|_2 < r$. Suppose that there exists a non-negative constant $\delta \leq Mr$ such that*

$$-\|\nabla V(\tilde{\boldsymbol{x}}, t)^\top \boldsymbol{u}(\tilde{\boldsymbol{x}}, t)\|^2 + b \cdot \mathcal{L}V(\tilde{\boldsymbol{x}}, t)V(\tilde{\boldsymbol{x}}, t) + (2 + b)Mr \leq \delta, \ \forall \tilde{\boldsymbol{x}} \in \tilde{\mathcal{D}}, \ \tilde{\boldsymbol{x}} \neq 0. \tag{S2.9}$$

Then, the controller $\boldsymbol{u}$ satisfies the stability condition (iii) specified in Theorem S2.1.

Proof. From the condition, we know that $\|\nabla V(\tilde{\boldsymbol{x}}, t)^\top \boldsymbol{u}(\tilde{\boldsymbol{x}}, t)\|^2 - b \cdot \mathcal{L}V(\tilde{\boldsymbol{x}}, t)V(\tilde{\boldsymbol{x}}, t) - (1 + b)Mr \geq 0$. For any $\boldsymbol{x} \in \mathcal{D}$, there exists $\tilde{\boldsymbol{x}}, \in \tilde{\mathcal{D}}$ s.t. $\|\boldsymbol{x} - \tilde{\boldsymbol{x}}\|_2 \leq r$. By interpolation, we have

$$\|\nabla V(\boldsymbol{x}, t)^\top \boldsymbol{u}\|^2 - b \cdot \mathcal{L}V(\boldsymbol{x}, t)V(\boldsymbol{x}, t)$$
$$= - \|\nabla V(\tilde{\boldsymbol{x}}, t)^\top \boldsymbol{u}(\tilde{\boldsymbol{x}}, t)\|^2 + \|\nabla V(\boldsymbol{x}, t)^\top \boldsymbol{u}(\boldsymbol{x}, t)\|^2 + \|\nabla V(\tilde{\boldsymbol{x}}, t)^\top \boldsymbol{u}(\tilde{\boldsymbol{x}}, t)\|^2$$
$$- b \cdot \mathcal{L}V(\tilde{\boldsymbol{x}}, t)V(\tilde{\boldsymbol{x}}, t) - b \cdot \mathcal{L}V(\boldsymbol{x}, t)V(\boldsymbol{x}, t) + b \cdot \mathcal{L}V(\tilde{\boldsymbol{x}}, t)V(\tilde{\boldsymbol{x}}, t)$$
$$\geq (1 + b)Mr - Mr - bMr \geq 0,$$

Since $V > 0$ for any $\boldsymbol{x} \neq 0$, the condition (*iii*) in Theorem S2.1 holds. The proof is complete.

Remark S2.3 *(Discussion of Theorem S2.2) The main difficulty in the original machine learning framework is to guarantee the condition for every point $\boldsymbol{x} \in \mathcal{D}$, as the trained neural network basically guarantee this condition on a finite dataset. With the proposed stability guarantee theory, if we choose the training dataset as the mesh grid discretization $\tilde{\mathcal{D}}$, and terminate the training process until the loss function is smaller than some precision $\delta \leq Mr$, the stability guarantee on the whole $\mathcal{D}$ is obtained. Here the bounded state space is reasonable in most real-world scenarios and hence is a mild condition. For example, in synchronization of the coupled oscillators, each oscillator is located on the attractor and the difference between any two oscillators are bounded by twice size of the attractor.*

Remark S2.4 *(Formula of Lipschitz constant) In Theorem S2.2, we utilize the Lipschitz constants of the functions taking the form as fg, and estimate the Lipschitz constant as $L_{fg} = L_f M_g + L_g M_f$, where $L_{(\cdot)}$ represents the Lipschitz constant and $M_{(\cdot)}$ is the L_∞ norm for the correspondingly given function on the bounded space $\mathcal{D}$.*

Remark S2.5 *(Lipschitz neural network) In our settings, we add the spectral normalization for the neural control function to constrain its Lipschitz constant lower than 1 [11, 12]. Therefore, it is direct to estimate the Lipschitz constant $\mathcal{M}$ is Theorem S2.2. Other Lipschitz regularization methods can be applied in our framework [13, 14] as well.*

Remark S2.6 *(Size of training dataset) In our original version of the AIN Synchrony, we randomly sample the state points as the training data, and the size of the training data is usually independent of system's dimension. For example, we select the size $2,000$ in most of the simulations. However, in the stability guarantee version, since we require the dataset to be the finite mesh grid discretization of the state space, the size of dataset depends on the dimension d and on the mesh size r as well. Specifically, if we need N grids for each variable to obtain the expected mesh size r, the size scales as $\mathcal{O}(N^d)$. Since the mesh grid is not the unique choice of the space discretization, the other discretization methods can be considered depending on the shape of the state space.*

Finally, we define the loss function for the stability guarantee of the controlled system (S2.5) as follows:

$$L_{\tilde{\mathcal{D}}}(\boldsymbol{\theta}, \boldsymbol{\phi}) = \frac{1}{|\tilde{\mathcal{D}}|} \sum_{(\boldsymbol{\xi}, \boldsymbol{s}(t_i), t_i) \in \tilde{\mathcal{D}}} \left(b \cdot \mathcal{L}V_{\boldsymbol{\theta}}(\boldsymbol{\xi}_i, t_i)V_{\boldsymbol{\theta}}(\boldsymbol{\xi}_i, t_i) - \left[\nabla V_{\boldsymbol{\theta}}(\boldsymbol{\xi}_i, t_i)^\top \boldsymbol{u}_{\boldsymbol{\phi}}(\boldsymbol{\xi}_i, \boldsymbol{s}(t_i))\right]^2 + (2 + b)Mr \right)^+. \tag{S2.10}$$

S3. CONTROLLING LINEAR VARIATIONAL EQUATION

In this section, we extend our framework to the linear variational equation near the synchronization manifold which has been extensively studied after the pioneering work [15], known as the master stability function. We consider the coupling structure to be a Laplacian matrix, i.e., $\sum_j A_{ij} = 0$. In this way, the linear expansion around $\boldsymbol{s}$ of the coupling term can be expressed as $\sum_j A_{ij} \left(\boldsymbol{M}_1(\boldsymbol{s}, \boldsymbol{s}, \boldsymbol{\mu}_1) + \nabla_{\boldsymbol{x}} \boldsymbol{M}_1(\boldsymbol{s}, \boldsymbol{s}, \boldsymbol{\mu}_1)\boldsymbol{\xi}_i + \nabla_{\boldsymbol{y}} \boldsymbol{M}_1(\boldsymbol{s}, \boldsymbol{s}, \boldsymbol{\mu}_1)\boldsymbol{\xi}_j \right) = \sum_j A_{ij} \nabla_{\boldsymbol{y}} \boldsymbol{M}_1(\boldsymbol{s}, \boldsymbol{s}, \boldsymbol{\mu}_1)\boldsymbol{\xi}_j$. For the nonlinear controlled dynamics of the variance $\boldsymbol{\xi}$ in (S1.2), the linear expansion around the zero solution is

$$\mathrm{d}\boldsymbol{\xi} = \left[\boldsymbol{I}_n \otimes \nabla_{\boldsymbol{x}} \boldsymbol{M}_0(\boldsymbol{s}, \boldsymbol{\mu}_0) + \boldsymbol{A} \otimes \nabla_{\boldsymbol{y}} \boldsymbol{M}_1(\boldsymbol{s}, \boldsymbol{s}, \boldsymbol{\mu}_1) \right] \boldsymbol{\xi} \mathrm{d}t + \boldsymbol{u}(\boldsymbol{\xi}, \boldsymbol{s})\mathrm{d}\boldsymbol{B}_t. \tag{S3.11}$$

Notice this linear equation still depends on the temporal variable $\boldsymbol{s} = \boldsymbol{s}(t)$, hence it is a non-autonomous dynamical system. Similarly to Section. S1, the detailed formulation of noise controllers under different cases are as follows:

(i) Regulation of the systematic parameters as

$$\boldsymbol{u}(\boldsymbol{\xi}, \boldsymbol{s}) = \left[\boldsymbol{I}_n \otimes \nabla_{\boldsymbol{x}} \boldsymbol{M}_0(\boldsymbol{s}, \boldsymbol{\mu}_0 + \boldsymbol{\delta\mu}_0(\boldsymbol{\xi}, \boldsymbol{s})) + \boldsymbol{A} \otimes \nabla_{\boldsymbol{y}} \boldsymbol{M}_1(\boldsymbol{s}, \boldsymbol{s}, \boldsymbol{\mu}_1 + \boldsymbol{\delta\mu}_1(\boldsymbol{\xi}, \boldsymbol{s})) \right] \boldsymbol{\xi}.$$

(ii) Perturbation to the network structure as

$$\boldsymbol{u}(\boldsymbol{\xi}, \boldsymbol{s}) = \left[\boldsymbol{\delta A} \otimes \nabla_{\boldsymbol{y}} \boldsymbol{M}_1(\boldsymbol{s}, \boldsymbol{s}, \boldsymbol{\mu}_1 + \boldsymbol{\delta\mu}_1(\boldsymbol{\xi}, \boldsymbol{s})) \right] \boldsymbol{\xi}.$$

(iii) Addition of the external forces as

$$\boldsymbol{u}(\boldsymbol{\xi}, \boldsymbol{s}) = \mathcal{T}_\gamma \left(\mathcal{M}(\boldsymbol{M}_0, \boldsymbol{A})(\boldsymbol{u}_{\boldsymbol{\phi}}(\boldsymbol{\xi}, \boldsymbol{s})) \right).$$

Here, $\boldsymbol{u}_{\boldsymbol{\phi}}(\boldsymbol{\xi}, \boldsymbol{s})$ is the output of the neural network, and $\mathcal{T}_\gamma$ is the element-wise truncation operator such that

$$\mathcal{T}_\gamma(x) = \begin{cases} \gamma, & x > \gamma, \\ x, & -\gamma \leq x \leq \gamma, \\ -\gamma, & x < -\gamma, \end{cases}$$

or

$$\mathcal{T}_\gamma(x) = \begin{cases} 0, & x > \gamma, \\ x, & -\gamma \leq x \leq \gamma, \\ 0, & x < -\gamma. \end{cases}$$

Additionally, $\mathcal{M}(\boldsymbol{M}_0, \boldsymbol{A})$ is the mask operator depending on the inner structure of self-dynamics $\boldsymbol{M}_0$ and the coupling structure. This operator is used to determine the controlled components of oscillators or the controlled oscillators such as the pinning control. In practice, we could change the operation order of $\mathcal{T}_\gamma$ and $\mathcal{M}(\boldsymbol{M}_0, \boldsymbol{A})$ according to the specific situation. The specific form of this operator is shown in Section S6. Although our framework is compatible with this non-autonomous linear variational equation, this non-autonomous dynamical system can be further simplified to an autonomous version under mild conditions. To achieve this goal, we first provide the following Comparison Theorem for stochastic differential equations.

Theorem S3.1 *(Comparison Theorems of SDEs) Consider the following two SDEs,*

$$\mathrm{d}\boldsymbol{x} = \boldsymbol{P}(t)\boldsymbol{x}\mathrm{d}t + \boldsymbol{u}(\boldsymbol{x})\mathrm{d}\boldsymbol{B}_t, \ \boldsymbol{x}(0) = \boldsymbol{z}_0 \in \mathbb{R}^d,$$

$$\mathrm{d}\boldsymbol{y} = \boldsymbol{Q}\boldsymbol{y}\mathrm{d}t + \boldsymbol{u}(\boldsymbol{y})\mathrm{d}\boldsymbol{B}_t, \ \boldsymbol{y}(0) = \boldsymbol{z}_0 \in \mathbb{R}^d,$$

where $(\boldsymbol{P}(t)^\top + \boldsymbol{P}(t)) - (\boldsymbol{Q}^\top - \boldsymbol{Q})$ is a temporal negative definite matrix, $\boldsymbol{u}(\boldsymbol{0}) = \boldsymbol{0}$ s.t. $\boldsymbol{0}$ is the equilibrium of the SDEs, and $\boldsymbol{u}$ satisfies the locally linear growth condition such that the strong solutions of both SDEs exist. Then we have $\|\boldsymbol{x}(t)\|_2 \leq \|\boldsymbol{y}(t)\|_2$. Furthermore, if $\lim_{t \to \infty} \boldsymbol{y}(t) = \boldsymbol{0}$, we have $\lim_{t \to \infty} \boldsymbol{x}(t) = \boldsymbol{0}$.

The Comparison Theorem of SDEs has been investigated before in [16]; but our contribution here is to propose a new concise proof based on the constructive auxiliary process.

Proof. *By Itô's formula, we have*

$$\mathrm{d}\boldsymbol{x}^\top \boldsymbol{x} = \left[\boldsymbol{x}^\top(\boldsymbol{P}^\top(t) + \boldsymbol{P}(t))\boldsymbol{x} + \boldsymbol{u}(\boldsymbol{x})^\top \boldsymbol{u}(\boldsymbol{x})\right]\mathrm{d}t + 2\boldsymbol{x}^\top \boldsymbol{u}(\boldsymbol{x})\mathrm{d}\boldsymbol{B}_t,$$

$$\mathrm{d}\boldsymbol{y}^\top \boldsymbol{y} = \left[\boldsymbol{y}^\top(\boldsymbol{Q}^\top + \boldsymbol{Q})\boldsymbol{y} + \boldsymbol{u}(\boldsymbol{y})^\top \boldsymbol{u}(\boldsymbol{y})\right]\mathrm{d}t + 2\boldsymbol{y}^\top \boldsymbol{u}(\boldsymbol{y})\mathrm{d}\boldsymbol{B}_t$$

By setting $z = \boldsymbol{x}^\top \boldsymbol{x} - \boldsymbol{y}^\top \boldsymbol{y}$ and using the following linearization notations,

$$\alpha(t) = \begin{cases} \dfrac{\boldsymbol{x}^\top\left[\boldsymbol{P}(t)^\top + \boldsymbol{P}(t)\right]\boldsymbol{x} - \boldsymbol{y}^\top\left[\boldsymbol{P}(t)^\top + \boldsymbol{P}(t)\right]\boldsymbol{y} + \boldsymbol{u}(\boldsymbol{x})^\top \boldsymbol{u}(\boldsymbol{x}) - \boldsymbol{u}(\boldsymbol{y})^\top \boldsymbol{u}(\boldsymbol{y})}{z}, & z \neq 0 \\ 0, & z = 0, \end{cases}$$

$$\boldsymbol{\beta}(t) = \begin{cases} \dfrac{2\boldsymbol{x}^\top \boldsymbol{u}(\boldsymbol{x}) - 2\boldsymbol{y}^\top \boldsymbol{u}(\boldsymbol{y})}{z}, & z \neq 0 \\ 0, & z = 0, \end{cases}$$

we have the following representation of the dynamics of z,

$$\mathrm{d}z = \left\{\alpha(t)z + \boldsymbol{y}^\top\left[(\boldsymbol{P}(t)^\top + \boldsymbol{P}(t)) - (\boldsymbol{Q}^\top - \boldsymbol{Q})\right]\boldsymbol{y}\right\}\mathrm{d}t + \boldsymbol{\beta}(t)z\mathrm{d}\boldsymbol{B}_t$$

Denote by $\phi(t) = \boldsymbol{y}^\top\left[(\boldsymbol{P}(t)^\top + \boldsymbol{P}(t)) - (\boldsymbol{Q}^\top - \boldsymbol{Q})\right]\boldsymbol{y}$. Hence, we have $\phi(t) \leq 0$ according to the condition. Next, we construct an auxiliary process $M(t)$ as

$$\mathrm{d}M = -\alpha(t)M\mathrm{d}t - \boldsymbol{\beta}(t)M\mathrm{d}\boldsymbol{B}_t, \quad M(0) > 0.$$

Then, the process is an exponential martingale such that $M(t) > 0$. Thus, applying Itô's formula to zM yields

$$\mathrm{d}(zM) = z\mathrm{d}M + M\mathrm{d}z + \mathrm{d}Z\mathrm{d}M$$

$$= z\left[-\alpha(t)M\mathrm{d}t - \boldsymbol{\beta}(t)M\mathrm{d}\boldsymbol{B}_t\right] + M\left\{\left[\alpha(t)z + \boldsymbol{y}^\top((\boldsymbol{P}(t)^\top + \boldsymbol{P}(t)) - (\boldsymbol{Q}^\top - \boldsymbol{Q}))\boldsymbol{y}\right]\mathrm{d}t + \boldsymbol{\beta}(t)z\mathrm{d}\boldsymbol{B}_t\right\}$$

$$+ \left\{-\alpha(t)M\mathrm{d}t - \boldsymbol{\beta}(t)M\mathrm{d}\boldsymbol{B}_t\right\}\left\{\left[\alpha(t)z + \boldsymbol{y}^\top((\boldsymbol{P}(t)^\top + \boldsymbol{P}(t)) - (\boldsymbol{Q}^\top - \boldsymbol{Q}))\boldsymbol{y}\right]\mathrm{d}t + \boldsymbol{\beta}(t)z\mathrm{d}\boldsymbol{B}_t\right\}$$

$$= \left(-\|\boldsymbol{\beta}\|^2 zM + \phi(t)\right)\mathrm{d}t.$$

By the Variation-of-Constants formula [17], we obtain

$$z(t)M(t) = e^{-\int_0^t \|\boldsymbol{\beta}(s)\|_2^2 \mathrm{d}s}\left\{z_0 M_0 + \int_0^t e^{\int_0^s \|\boldsymbol{\beta}(\tau)\|_2^2 \mathrm{d}\tau} M(s)\phi(s)\mathrm{d}s\right\}, \tag{S3.12}$$

Since $z_0 = 0$, $M(t) \geq 0$ and $\phi(t) \leq 0$ on $[0, \infty)$, the right hand side of Eq. (S3.12) is less than 0, we have $z(t) \leq 0$ on $[0, \infty)$, which completes the proof.

We focus on the variational dynamics for each oscillator $\boldsymbol{\xi}_i$ by diagonalizing the coupling structure $\boldsymbol{A}$ as

$$\mathrm{d}\boldsymbol{\xi}_i = \left[\nabla_{\boldsymbol{x}}\boldsymbol{M}_0(\boldsymbol{s}, \boldsymbol{\mu}_0) + \lambda_i \nabla_{\boldsymbol{y}}\boldsymbol{M}_1(\boldsymbol{s}, \boldsymbol{s}, \boldsymbol{\mu}_1)\right]\mathrm{d}t + \boldsymbol{u}_i(\boldsymbol{\xi}, \boldsymbol{s})\mathrm{d}\boldsymbol{B}_t,$$

where $\{\lambda_1, \cdots, \lambda_n\}$ are the eigenvalues of coupling matrix $\boldsymbol{A}$. By applying this Comparison Theorem to equation (S3.11), we obtain the corresponding dominant equation as

$$\mathrm{d}\boldsymbol{\xi}_i = \left[\tilde{\boldsymbol{M}}_0 + \max\{|\lambda_1|, \cdots, |\lambda_n|\}\tilde{\boldsymbol{M}}_1\right]\mathrm{d}t + \boldsymbol{u}_i(\boldsymbol{\xi}, \boldsymbol{s})\mathrm{d}\boldsymbol{B}_t,$$

with $(\tilde{\boldsymbol{M}}_0)_{ij} \geq (\nabla_{\boldsymbol{x}}\boldsymbol{M}_0(\boldsymbol{s}, \boldsymbol{\mu}_0))_{ij}$ and $(\tilde{\boldsymbol{M}}_1)_{ij} \geq (\nabla_{\boldsymbol{y}}\boldsymbol{M}_1(\boldsymbol{s}, \boldsymbol{s}, \boldsymbol{\mu}_1))_{ij}$. We can find such $\tilde{\boldsymbol{M}}_{0,1}$ thanks to the boundedness of the synchronization manifold $\boldsymbol{s}$. Now, we only need to find the appropriate form of the stabilizing noise for this dominant equation with autonomous drift function, and then apply the noise to the original collective dynamics as (S1.4).

We should emphasize that this extension is used to demonstrate the generality of our framework, but not an essential improvement. Since the linear variational equation is only the linear approximation of the variational dynamics near the zero solution, it is not rigorously valid within the whole state space. Thus, we propose to use this simpler equation (S3.11) to find the artificial stabilizing noise only in the case where the original dynamics owns a global attractor, in which the system is going to evolve into the vicinity of the synchronization manifold after a sufficiently long time.

S4.   CONTROLLING COLLECTIVE DYNAMICS WITH TEMPORAL INTERACTING STRUCTURE

In this section, we establish two theorems for extending our framework to synchronize the non-autonomous dynamics with temporal coupling structure $\boldsymbol{A} = \boldsymbol{A}(t)$. Theorem S4.1 addresses global synchronization and Theorem S4.2 addresses local synchronization.

Theorem S4.1 *Consider the nonlinear variational equations in* (S1.3). *If the continuous coupling function* $\boldsymbol{M}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\mu}_1)$ *is positively (resp., negatively) semi-definite and the temporal structure* $\boldsymbol{A} = \boldsymbol{A}(t)$ *is bounded for* $t \in [0, \infty)$, *then the dominant equation of dynamics* (S1.3) *is*

$$
\mathrm{d}\tilde{\boldsymbol{\xi}}_i = \left[\boldsymbol{M}_0(\tilde{\boldsymbol{\xi}}_i + \boldsymbol{s}, \boldsymbol{\mu}_0) - \boldsymbol{M}_0(\boldsymbol{s}, \boldsymbol{\mu}_0) + \sum_{j=1}^n \tilde{A}_{ij} \boldsymbol{M}_1(\tilde{\boldsymbol{\xi}}_i + \boldsymbol{s}, \tilde{\boldsymbol{\xi}}_j + \boldsymbol{s}, \boldsymbol{\mu}_1) \right] \mathrm{d}t + \boldsymbol{u}_i(\tilde{\boldsymbol{\xi}}, \boldsymbol{s}) \mathrm{d}\boldsymbol{B}_t, \qquad \text{(S4.13)}
$$

with $\tilde{A}_{ij} \geq A_{ij}(t)$ *(resp.* $\tilde{A}_{ij} \leq A_{ij}(t)$*), which means if* $\boldsymbol{u}$ *stabilizes the dominant equation* (S4.13), *then it also stabilizes the original nonlinear variational equation* (S1.3).

Proof. *Since* $\boldsymbol{M}_1$ *is positively semi-definite and* $\tilde{A}_{ij} \geq A_{ij}(t)$, *each scalar variable of* $\tilde{A}_{ij} \boldsymbol{M}_1(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\mu}_1)$ *is larger than* $A_{ij}(t) \boldsymbol{M}_1(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\mu}_1)$. *Taking* ξ_1^1 *(the first variable of* ξ_1*) for example, we have*

$$
\begin{aligned}
\mathrm{d}\xi_1^1 &= \left[\boldsymbol{M}_0^1(\boldsymbol{\xi}_1 + \boldsymbol{s}, \boldsymbol{\mu}_0) - \boldsymbol{M}_0^1(\boldsymbol{s}, \boldsymbol{\mu}_0) + \sum_{j=1}^n A_{ij}(t) \boldsymbol{M}_1^1(\boldsymbol{\xi}_1 + \boldsymbol{s}, \boldsymbol{\xi}_j + \boldsymbol{s}, \boldsymbol{\mu}_1) \right] \mathrm{d}t + (\boldsymbol{u}_1(\boldsymbol{\xi}, \boldsymbol{s}) \mathrm{d}\boldsymbol{B}_t)^1 \\
&\triangleq f_1(\xi_1^1, t) \mathrm{d}t + g(\xi_1^1, t) \mathrm{d}\tilde{B}_t,
\end{aligned}
$$

and

$$
\begin{aligned}
\mathrm{d}\tilde{\xi}_1^1 &= \left[\boldsymbol{M}_0^1(\tilde{\boldsymbol{\xi}}_1 + \boldsymbol{s}, \boldsymbol{\mu}_0) - \boldsymbol{M}_0^1(\boldsymbol{s}, \boldsymbol{\mu}_0) + \sum_{j=1}^n \tilde{A}_{ij} \boldsymbol{M}_1^1(\tilde{\boldsymbol{\xi}}_1 + \boldsymbol{s}, \tilde{\boldsymbol{\xi}}_j + \boldsymbol{s}, \boldsymbol{\mu}_1) \right] \mathrm{d}t + (\boldsymbol{u}_1(\tilde{\boldsymbol{\xi}}, \boldsymbol{s}) \mathrm{d}\boldsymbol{B}_t)^1 \\
&\triangleq f_2(\tilde{\xi}_1^1, t) \mathrm{d}t + g(\tilde{\xi}_1^1, t) \mathrm{d}\tilde{B}_t,
\end{aligned}
$$

for some standard Brownian motion $\tilde{B}_t$. *From the condition above we have with* $f_1 \leq f_2$, *based on Comparison Theorems of SDEs in Theorem* S3.1, *if* $\tilde{\xi}_1^1$ *is stabilized by noise term* g, *then* ξ_1^1 *is stabilized by* g. *We could prove it in the same way as we presented in the proof of Theorem* S3.1, *so we omit it here. By induction, we complete the proof.*

If the coupling function $\boldsymbol{M}_1$ is not positive or negative, we could obtain the following comparison theorem for the linear variational equation.

Theorem S4.2 *Consider the linear temporal variational equations in* (S3.11). *If the synchronization manifold* $\boldsymbol{s}$ *is bounded and the Jordan normal form of the temporal structure* $\boldsymbol{A}(t)$ *is* $\boldsymbol{P}^{-1} \boldsymbol{J}(t) \boldsymbol{P}$ *bounded by* $\boldsymbol{J}(t)$, *the dominant equations in* (S3.11) *become*

$$
\mathrm{d}\boldsymbol{\xi}_i = \left[\tilde{\boldsymbol{M}}_0 + \max_{t \in [0, \infty)} \left\{ J_{11}(t), \cdots, J_{nn}(t) \right\} \tilde{\boldsymbol{M}}_1 \right] \boldsymbol{\xi} \mathrm{d}t + \boldsymbol{u}(\boldsymbol{\xi}, \boldsymbol{s}) \mathrm{d}\boldsymbol{B}_t,
$$

where $\tilde{A}_{ij} \geq |A_{ij}(t)|$, $(\tilde{M}_0)_{ij} \geq |\nabla \boldsymbol{M}_0(\boldsymbol{s}, \boldsymbol{\mu}_0)_{ij}|$, *and* $(\tilde{M}_1)_{ij} \geq |\nabla \boldsymbol{M}_1(\boldsymbol{s}, \boldsymbol{s}, \boldsymbol{\mu}_0)_{ij}|$. *Then, the stabilizing noise* $\boldsymbol{u}$ *for the dominant equations also stabilizes the original linear temporal variational equations in* (S3.11).

Proof. *The validity of this theorem follows directly from Theorem* S3.1, *so we omit the proof here.*

The condition of $\boldsymbol{A}(t) = \boldsymbol{P}^{-1} \boldsymbol{J}(t) \boldsymbol{P}$ has been used to construct the temporal network under limit resources [6]. The aforementioned theorems tell us how to extend our framework to synchronize the temporal network with the closed-loop neural noise. In addition, we could directly design open-loop noise to synchronize the system as Theorem S2.1 stands for general non-autonomous dynamics.

S5. IMPLICIT ENERGY REGULARIZATION THEORY

In this section, we theoretically explain the mechanism of implicit energy regularization that induces energy-saving synchrony observed in our results. We summarize the energy-saving property in our machine learning framework in Theorem S5.1, and then detailed analyze the reason.

Theorem S5.1 *Consider the controlled dynamics in* (S2.5) *where the controller* $\boldsymbol{u} = \boldsymbol{u}_{\boldsymbol{\phi}}$ *in the diffusion term is learned through the proposed machine learning framework, under the loss function*

$$L(\boldsymbol{\theta}, \boldsymbol{\phi}) = \frac{1}{m} \sum_{i=1}^{m} \left(\frac{b \cdot \mathcal{L}V_{\boldsymbol{\theta}}(\boldsymbol{\xi}_i, t_i)}{V_{\boldsymbol{\theta}}(\boldsymbol{\xi}_i, t_i)} - \frac{\left[\nabla V_{\boldsymbol{\theta}}(\boldsymbol{\xi}_i, t_i)^\top \boldsymbol{u}_{\boldsymbol{\phi}}(\boldsymbol{\xi}_i, \boldsymbol{s}(t_i)) \right]^2}{V_{\boldsymbol{\theta}}(\boldsymbol{\xi}_i, t_i)^2} \right)^+, \tag{S5.14}$$

where $\{\boldsymbol{\xi}_i, \boldsymbol{s}(t_i), t_i\}_{i=1}^{m}$ *is the training dataset sampled from the state space* $\mathcal{X} \times \boldsymbol{s} \times [0, T]$. *Then, the loss function in the training process is equivalent to the control energy in the control process by norm,*

$$E = \mathbb{E} \int_0^T \|\boldsymbol{u}_{\boldsymbol{\phi}}^\top(\boldsymbol{x}(t), \boldsymbol{s}(t)) Q \boldsymbol{u}_{\boldsymbol{\phi}}(\boldsymbol{x}(t), \boldsymbol{s}(t))\|_{\mathrm{F}} \mathrm{d}t, \tag{S5.15}$$

here Q *is the variance matrix of the Brownian motion. In other words, once we minimize the loss function in the training stage, the control energy in the test stage is optimized at the same time, which leads to the energy-saving phenomenon.*

The neural network tends to minimize the following loss function in the training process,

$$L(\boldsymbol{\theta}, \boldsymbol{\phi}) = \frac{1}{m} \sum_{i=1}^{m} \left(\frac{b \cdot \mathcal{L}V_{\boldsymbol{\theta}}(\boldsymbol{\xi}_i, t_i)}{V_{\boldsymbol{\theta}}(\boldsymbol{\xi}_i, t_i)} - \frac{(\nabla V_{\boldsymbol{\theta}}(\boldsymbol{\xi}_i, t_i)^\top \boldsymbol{u}_{\boldsymbol{\phi}}(\boldsymbol{\xi}_i, \boldsymbol{s}(t_i)))^2}{V_{\boldsymbol{\theta}}(\boldsymbol{\xi}_i, t_i)^2} \right)^+ .$$

Notice that the second part $-\dfrac{\left[\nabla V_{\boldsymbol{\theta}}(\boldsymbol{\xi}_i, t_i)^\top \boldsymbol{u}_{\boldsymbol{\phi}}(\boldsymbol{\xi}_i, \boldsymbol{s}(t_i)) \right]^2}{V_{\boldsymbol{\theta}}(\boldsymbol{\xi}_i, t_i)^2}$ is semi-negative, thus the sufficient condition for minimizing the loss function is to minimize the first part,

$$\tilde{L}(\boldsymbol{\theta}, \boldsymbol{\phi}) = \frac{1}{m} \sum_{i=1}^{m} \left(\frac{b \cdot \mathcal{L}V_{\boldsymbol{\theta}}(\boldsymbol{\xi}_i, t_i)}{V_{\boldsymbol{\theta}}(\boldsymbol{\xi}_i, t_i)} \right)^+ .$$

Next, according to the definition of $\mathcal{L}V$, there is a semi-positive part in the modified loss function $\tilde{L}$:

$$\bar{L}(\boldsymbol{\theta}, \boldsymbol{\phi}) = \frac{1}{m} \sum_{i=1}^{m} \frac{1}{2} \mathrm{Tr} \left[\boldsymbol{u}_{\boldsymbol{\phi}}^\top(\boldsymbol{\xi}_i, \boldsymbol{s}(t_i)) \nabla^2 V_{\boldsymbol{\theta}}(\boldsymbol{\xi}_i, t_i) \boldsymbol{u}_{\boldsymbol{\phi}}(\boldsymbol{\xi}_i, \boldsymbol{s}(t_i)) \right] .$$

This implies that minimizing $\bar{L}$ is necessary for minimizing the modified loss function $\tilde{L}$. Notice that the energy consumed in the control process is defined as

$$E = \mathbb{E} \int_0^T \|\boldsymbol{u}_{\boldsymbol{\phi}}^\top(\boldsymbol{x}(t), \boldsymbol{s}(t)) Q \boldsymbol{u}_{\boldsymbol{\phi}}(\boldsymbol{x}(t), \boldsymbol{s}(t))\|_{\mathrm{F}} \mathrm{d}t,$$

where $\boldsymbol{x}(t)$ is the controlled trajectory, Q is a positive definite matrix related to the specific form of energy consumption, and $\| \cdot \|_{\mathrm{F}}$ is the Frobenius norm. In our experiments, we take $Q = I$ for simplicity. From the equivalence of the matrix norm [18] , we have

$$\|\boldsymbol{u}_{\boldsymbol{\phi}}^\top Q \boldsymbol{u}_{\boldsymbol{\phi}}\|_{\mathrm{F}} \iff \|\boldsymbol{u}_{\boldsymbol{\phi}}^\top Q \boldsymbol{u}_{\boldsymbol{\phi}}\|_2.$$

Here the equivalence relation $A \iff B$ is defined by $\alpha B \le A \le \beta B$ for some $\alpha, \beta > 0$. Since for positive semidefinite matrix $\boldsymbol{u}_{\boldsymbol{\phi}}^\top Q \boldsymbol{u}_{\boldsymbol{\phi}}$, the following inequality holds,

$$\|\boldsymbol{u}_{\boldsymbol{\phi}}^\top Q \boldsymbol{u}_{\boldsymbol{\phi}}\|_2 = \lambda_{\max}(\boldsymbol{u}_{\boldsymbol{\phi}}^\top Q \boldsymbol{u}_{\boldsymbol{\phi}}) \le \mathrm{Tr} \left[\boldsymbol{u}_{\boldsymbol{\phi}}^\top \nabla^2 Q \boldsymbol{u}_{\boldsymbol{\phi}} \right] \le r \lambda_{\max}(\boldsymbol{u}_{\boldsymbol{\phi}}^\top Q \boldsymbol{u}_{\boldsymbol{\phi}})$$

with r being the dimension of the Wiener process, which implies,

$$\|\boldsymbol{u}_{\boldsymbol{\phi}}^{\top}Q\boldsymbol{u}_{\boldsymbol{\phi}}\|_2 \iff \mathrm{Tr}\left[\boldsymbol{u}_{\boldsymbol{\phi}}^{\top}\nabla^2 Q\boldsymbol{u}_{\boldsymbol{\phi}}\right].$$

Furthermore, we can regard positive definite matrices Q and $\nabla^2 V_{\boldsymbol{\theta}}$ as constant matrix since they are independent of parameter $\boldsymbol{\phi}$, leading to the following relation,

$$\|\boldsymbol{u}_{\boldsymbol{\phi}}^{\top}Q\boldsymbol{u}_{\boldsymbol{\phi}}\|_{\mathrm{F}} \iff \mathrm{Tr}\left[\boldsymbol{u}_{\boldsymbol{\phi}}^{\top}\nabla^2 V_{\boldsymbol{\theta}}\boldsymbol{u}_{\boldsymbol{\phi}}\right].$$

Finally, we come to the conclusion that under suitable data distribution in the path space, the necessary loss function $\bar{L}$ is the Monte Carlo realization of the energy E. In this way, the noise controller becomes more energy-saving in the synchronization process by gradient descent with training epochs, leading to the implicit energy regularization.

S6. SIMULATION CONFIGURATIONS AND SUPPLEMENTARY EXPERIMENTS

In this section, we provide detailed descriptions of the experimental configurations of the physical examples in the main text. The first low dimensional models are trained within several minutes on the computational device with a single i7-10870 CPU with 16GB memory, and the left high dimensional tasks are trained within several minutes with a single Nvidia 3070 GPU with 16GB memory, while the neural networks for our artificial noise controllers are trained under PyTorch architecture [19] with Adam optimizer [20]. We fix the number of noise vectors in Hutchinson's trace estimator as $m = 1$ in all the experiments, as proposed in [21, 22]. The source code is released in AIN Synchrony.

1. The input convex neural network (ICNN) V function is constructed as:

$$z_1 = \sigma(\boldsymbol{W}_0\boldsymbol{x} + b_0),$$
$$z_{i+1} = \sigma(\boldsymbol{U}_i\boldsymbol{z}_i + \boldsymbol{W}_i\boldsymbol{x} + b_i), \ i = 1, \cdots, k-1,$$
$$p(\boldsymbol{x}) \equiv \boldsymbol{z}_k,$$
$$V(\boldsymbol{x}) = \sigma(p(\boldsymbol{x}) - p(\boldsymbol{0})) + \varepsilon\|\boldsymbol{x}\|^2,$$

$$\sigma(x) = \begin{cases} 0, & \text{if } x \leq 0, \\ (2dx^3 - x^4)/2d^3, & \text{if } 0 < x \leq d, \\ x - d/2, & \text{otherwise} \end{cases}$$

where σ is the smoothed **ReLU** function for ensuring $V \in C^2(\mathbb{R}^d)$, $\boldsymbol{W}_i \in \mathbb{R}^{h_i \times d}$, $\boldsymbol{U}_i \in (\mathbb{R}_+ \cup \{0\})^{h_i \times h_{i-1}}$, $\boldsymbol{x} \in \mathbb{R}^d$, and, for simplicity, this ICNN function is denoted by $\text{ICNN}(h_0, h_1, \cdots, h_{k-1})$. If the V function depends on time t, we could directly multiply the ICNN function by some parameterized function $h_{\boldsymbol{\theta}_h}(t)$ with $h_{\boldsymbol{\theta}_h}(t) > 0$. The parameters to be trained are summarized as $\boldsymbol{\theta} = \{W_i\}_{i=0}^{k-1} \cup \{U_i\}_{i=1}^{k-1}$ for time independent V function and $\boldsymbol{\theta} = \{W_i\}_{i=0}^{k-1} \cup \{U_i\}_{i=1}^{k-1} \cup \boldsymbol{\theta}_h$ for time dependent V function.

2. The neural control function is constructed as:

$$z_1 = \mathcal{F}(\boldsymbol{W}_0\boldsymbol{x} + B_1),$$
$$z_{i+1} = \mathcal{F}(\boldsymbol{W}_i\boldsymbol{z}_i + b_i), \ i = 1, \cdots, k-1,$$
$$\mathbf{NN}(\boldsymbol{x}) \equiv \boldsymbol{W}_k\boldsymbol{z}_k,$$
$$\boldsymbol{u}(\boldsymbol{x}, \boldsymbol{s}) = \mathcal{N}(\mathbf{NN}(\boldsymbol{x}, \boldsymbol{s}), \boldsymbol{x}), \text{ or } \boldsymbol{u}(\boldsymbol{x}) = \mathcal{N}(\mathbf{NN}(\boldsymbol{x}), \boldsymbol{x}),$$

where $\mathcal{F}(\cdot)$ is the activation function, $W_i \in \mathbb{R}^{h_{i+1} \times h_i}$, and $\mathcal{N}$ is some nonlinear function s.t. $\boldsymbol{u}(\boldsymbol{0}, \boldsymbol{s}) = \boldsymbol{0}$ or $\boldsymbol{u}(\boldsymbol{0}) = \boldsymbol{0}$. The control function is denoted by $\text{Control}(h_0, h_1, \cdots, h_{k+1})$, and the trainable parameters are $\boldsymbol{\phi} = \{\boldsymbol{W}_i\}_{i=0}^k$;

6.1. Driving-response Problem of Lorenz System

The driving system is governed by

$$\dot{x}_1 = \sigma(y_1 - x_1),$$
$$\dot{y}_1 = \rho x_1 - y_1 - x_1 z_1,$$
$$\dot{z}_1 = x_1 y_1 - \beta z_1,$$

and the noise controlled response system is governed by

$$\mathrm{d}x_2 = \sigma(y_2 - x_2)\mathrm{d}t + u_1(\boldsymbol{\xi})\mathrm{d}B_1(t),$$
$$\mathrm{d}y_2 = (\rho x_2 - y_2 - x_2 z_2)\mathrm{d}t + u_2(\boldsymbol{\xi})\mathrm{d}B_2(t),$$
$$\mathrm{d}z_2 = (x_2 y_2 - \beta z_2)\mathrm{d}t + u_3(\boldsymbol{\xi})\mathrm{d}B_3(t),$$

where $\boldsymbol{\xi} = (\xi_1, \xi_2, \xi_3)^\top = (x_2 - x_1, y_2 - y_1, z_2 - z_1)^\top$, and the control goal is finding the diffusion term $\boldsymbol{u}(\boldsymbol{\xi}) = (u_1(\boldsymbol{\xi}), u_2(\boldsymbol{\xi}), u_3(\boldsymbol{\xi}))$ that can completely synchronize the response system to the driving system, that is $\boldsymbol{\xi} = 0$. We fix the parameters as $\sigma = 10$, $\rho = 28$, $\beta = 8/3$. The corresponding variational dynamics of $\boldsymbol{z}$ becomes

$$\mathrm{d}\xi_1 = \sigma(\xi_2 - \xi_1)\mathrm{d}t + u_1(\boldsymbol{\xi})\mathrm{d}B_1(t),$$
$$\mathrm{d}\xi_2 = [\rho\xi_1 - \xi_2 - (x_1 + \xi_1)(z_1 + \xi_3) + x_1 z_1]\mathrm{d}t + u_2(\boldsymbol{\xi})\mathrm{d}B_2(t),$$
$$\mathrm{d}\xi_3 = [(x_1 + \xi_1)(y_1 + \xi_2) - x_1 y_1 - \beta\xi_3]\mathrm{d}t + u_3(\boldsymbol{\xi})\mathrm{d}B_3(t),$$

where $\boldsymbol{s} = (x_1, y_1, z_1)^\top$ is the synchronization manifold in this case. In this case, we focus on the noise control acting on the systematic parameters and the structural control on components of $\{\xi_1, \xi_2, \xi_3\}$. So we only consider the common noise case, that is $B_1(t) = B_2(t) = B_3(t)$. We consider the following cases of noise controllers.

(i) Regulating systematic parameters: We consider the noise taking the form as follows,

$$u_1(\boldsymbol{\xi}) = \mathcal{M}_1\big(\mathcal{T}_\gamma(\delta\sigma_{\boldsymbol{\phi}})(\xi_2 - \xi_1)\big),$$
$$u_2(\boldsymbol{\xi}) = \mathcal{M}_2\big(\mathcal{T}_\gamma(\delta\rho_{\boldsymbol{\phi}})(\xi_1)\big),$$
$$u_3(\boldsymbol{\xi}) = \mathcal{M}_3\big(-\mathcal{T}_\gamma(\delta\beta_{\boldsymbol{\phi}})(\xi_3)\big),$$
$$\mathcal{T}_\gamma(x) = \begin{cases} 0, & x > \gamma, \\ x, & -\gamma \le x \le \gamma, \\ 0, & x < -\gamma, \end{cases}$$

where $\delta\sigma_{\boldsymbol{\phi}}, \delta\rho_{\boldsymbol{\phi}}, \delta\beta_{\boldsymbol{\phi}}$ are parameterized by the neural networks, the mask operator $\mathcal{M} = (\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3)$ corresponds to the specific case among the 7 different regulated parameters combinations $\Omega_{1:7} \in 2^{\{\sigma, \rho, \beta\}}$. Specifically, we have

$$\Omega_1: \mathcal{M}(x) = (x, 0, 0),$$
$$\Omega_2: \mathcal{M}(x) = (0, x, 0),$$
$$\Omega_3: \mathcal{M}(x) = (0, 0, x),$$
$$\Omega_4: \mathcal{M}(x) = (x, x, 0),$$
$$\Omega_5: \mathcal{M}(x) = (x, 0, x),$$
$$\Omega_6: \mathcal{M}(x) = (0, x, x),$$
$$\Omega_7: \mathcal{M}(x) = (x, x, x).$$

To comprehensively investigate the influence of the regulating parameters, we select the scale of the output of neural networks as $\boldsymbol{u}_{\max} = \gamma \in \{1.0, 1.2, 1.4, 1.6, 1.8, 2.0\}$. We parameterize the functions $V(\boldsymbol{x})$ as ICNN(3, 36, 36), the controller as Control(3, 9, 9, 3) with $\mathcal{F} = \mathrm{ReLU}$. We set the hyperparameter as $b = 2.5$, lr = 1e-2 (learning rate), iters = 1e3 (iterations). The synchronization manifold $\boldsymbol{s}$ is pre-generated on 2000 regular time points on time interval $[0, 50]$ by numerically solving the driving system with `dopri5` method in `torchdiffeq` [23] package. For training data of $\boldsymbol{\xi}$, we sample 2000 points from uniform distribution $\mathcal{U}([-5, 5]^3)$. After training, we generate

the controlled trajectories under different combinations of $\Omega_{1:7}$ and γ, and we use Euler–Maruyama method [24] to numerically obtain 50 trajectories in time interval $[0, 10]$ with $\mathrm{d}t = \text{1e-4}$ and random seed from $\{0, 1, \cdots, 49\}$ for each case. To compare the controlled performance, we calculate the success rate as the ratio of the time average L^2 difference $\|(x_2 - x_1, y_2 - y_2, z_2 - z_1)\|_2$ on $[9, 10]$ being less than 1.0 in all the 50 trajectories for each case.

(ii) Adding external forces in the form of pinning control: We consider the following external noise forces:

$$\boldsymbol{u}(\boldsymbol{\xi}) = \mathcal{M}\bigg(\mathcal{T}_\gamma \big(\mathcal{N}(\boldsymbol{u}_{\boldsymbol{\phi}}(\boldsymbol{\xi}), \boldsymbol{\xi}) \big) \bigg),$$

$$\mathcal{N}(\boldsymbol{u}_{\boldsymbol{\phi}}(\boldsymbol{\xi}), \boldsymbol{\xi}) = (u_{\boldsymbol{\phi},1}(\boldsymbol{\xi})\xi_1, u_{\boldsymbol{\phi},2}(\boldsymbol{\xi})\xi_2, u_{\boldsymbol{\phi},3}(\boldsymbol{\xi})\xi_3)^\top, \ \text{ or } \ (\xi_1 + \xi_2 + \xi_3)(u_{\boldsymbol{\phi},1}(\boldsymbol{\xi}), u_{\boldsymbol{\phi},2}(\boldsymbol{\xi}), u_{\boldsymbol{\phi},3}(\boldsymbol{\xi}))^\top$$

where $\mathcal{T}_\gamma$ and $\mathcal{M}$ are the same as defined above. We test the performance under nonlinear operation $\mathcal{N}$ and select the best case in our experiment. To match the scale of the chaotic attractor of the Lorenz dynamics, we set $\gamma = 100$. To take the inner structure of variables x, y, z in Lorenz dynamics into consideration, we select three different operators $\mathcal{M}$ as follows:

$$\begin{aligned} \mathrm{AIN}x &: \ \mathcal{M}(x) = (x, 0, 0), \\ \mathrm{AIN}y &: \ \mathcal{M}(x) = (0, x, 0), \\ \mathrm{AIN}xz &: \ \mathcal{M}(x) = (x, 0, x), \end{aligned}$$

corresponding to the pinning control of the self-dynamics. The training and generating processes are the same as those mentioned above. We generate 50 trajectories on time interval $[0, 20]$ for each case.

(iii) We further compare the learned noise controller $\mathrm{AIN}y$ with the traditional linear noise controllers in terms of energy consumption and stability. We train the V function of the linear noise controller $\boldsymbol{u}(\boldsymbol{\xi}) = (0, \mathcal{T}_{100}(k(\xi_1 + \xi_2 + \xi_3)), 0)^\top$ for $k \in \{1, 2, \cdots, 15\}$, and we generate 50 controlled trajectories on the time interval $[0, 20]$ under these controllers with the same method as mentioned above. We calculate the energy cost as the average of the numerical integration $\int_0^{20} \|\boldsymbol{u}(\boldsymbol{\xi}(t))\|_2^2 \mathrm{d}t$ on 50 trajectories. For computing the convexity of the V function for linear controller and $\mathrm{AIN}y$, we equally select 8000 points an $[-2.5, 2.5]^3$ and use Monto Carlo realization $\frac{1}{8000} \sum_{j=1}^{8000} \frac{1}{3} \sum_{i=1}^{3} \lambda_i(\nabla V(\boldsymbol{\xi}_j))$ as the approximation of the defined convexity $\mathbb{E}\left[\frac{1}{3} \sum_{i=1}^{3} \lambda_i(\nabla V(\boldsymbol{\xi}_j)) \right]$. To numerically reveal the implicit energy regularization phenomenon, we calculate the energy cost and the convexity of trained V function of $\mathrm{AIN}y$ during the training process over 10 trajectories for each 100 iterations with iters $= 2000$ in total.

Actually, we select suitable hyperparameters using the grid search method. For example, we assess the control performance of the machine-learning-coined noise under different learning rates (lr) in $\{0.1, 0.01, 0.001\}$ and fix lr to the parameter under which the noise controller obtains the best performance. For an intuitive illustration, we provide an ablation study of the hyperparameters in pinning control of the driving-response Lorenz systems and summarize the results in Table S3. In our original simulations, we do not observe the training dataset scales with the dimension of the system. We find that the performance of the machine is good enough when the training dataset possesses $2,000$ samples as the dimension of the system ranges from 3 to 256. However, the training dataset in the newly improved version with the stability guarantee does scale with the dimension of the system, as the training data is obtained via the grid discretization of the state space. The size of the dataset is N^d where N is the grid size of each state component and d is the dimension of the system.

TABLE S3. The mean square error (MSE) and the standard variation between the driving and response Lorenz systems under noise pinning controllers trained by different hyperparameters.

Performance	Layers	Lr = 1e-1			Lr = 1e-2			Lr = 1e-3		
		h=6	h=9	h=12	h=6	h=9	h=12	h=6	h=9	h=12
MSE	2	20.52	0.00	0.00	17.53	19.62	17.12	15.16	27.88	11.72
	4	19.33	11.50	16.81	13.86	15.85	17.82	16.22	15.52	18.78
	6	41.51	15.95	0.00	0.00	2.55	16.85	20.31	7.89	13.14
STD	2	12.53	0.00	0.00	7.85	5.95	10.18	11.49	20.71	12.88
	4	6.19	20.27	7.54	5.81	7.87	19.91	10.37	8.11	11.51
	6	25.19	8.94	0.00	0.00	7.64	6.67	18.19	17.19	8.97

6.2. Subcritical Landau-Stuart System

The dynamics of the coupled subcritical Landau-Stuart system is governed by

$$\dot{Z}_j = (\beta + i\gamma + \mu|Z_j|^2)Z_j + \sum_{k=1}^{n} L_{jk}Z_k.$$

Specifically, we set the number of the oscillators as $n = 10$, the dynamical parameters as $\beta = -1$, $\gamma = \mu = 1$ to establish the unit cycle as the UPO of the self-dynamics. In this study, we focus on the impact of the communication constraints on the noise-induced synchronization, and we fix the controller form as external forces on each oscillator $\boldsymbol{u} = (u_1(\boldsymbol{Z}_i), u_2(\boldsymbol{Z}_2), \cdots, u_{10}(\boldsymbol{Z}_{10}))^\top$, $\boldsymbol{Z}_i \subset \{Z_1, \cdots, Z_{10}\}$ driven by common noise. In this case, the communication constraint is reflected in the self-variables of the controller, that is,

$$\begin{aligned}
&\text{Full: } \boldsymbol{Z}_i = \{Z_1, Z_2 \cdots, Z_{10}\}, i = 1, \cdots, 10, \\
&\text{Half: } \boldsymbol{Z}_i = \{Z_1, \cdots, Z_5\}, \ i = 1, \cdots, 5, \ \boldsymbol{Z}_i = \{Z_6, \cdots, Z_{10}\}, \ i = 6, \cdots, 10, \\
&\text{Decentralized: } \boldsymbol{Z}_i = \{Z_i\}, \ i = 1, \cdots, 10.
\end{aligned}$$

In practice, we consider the controlled dynamics in the real vector space, where $Z_j = x_j + iy_j$, as follows,

$$dx_i = \left[-x_i - y_i + x_i(x_i^2 + y_i^2) + \frac{1}{10}\sum_{j=1}^{10}(x_i - x_j)\right]dt + \mathcal{T}_\gamma(u_i^x(\boldsymbol{X}_i, \boldsymbol{Y}_i, \boldsymbol{s}))dB_t,$$

$$dy_i = \left[x_i - y_i + y_i(x_i^2 + y_i^2) + \frac{1}{10}\sum_{j=1}^{10}(y_i - y_j)\right]dt + \mathcal{T}_\gamma(u_i^y(\boldsymbol{X}_i, \boldsymbol{Y}_i, \boldsymbol{s}))dB_t,$$

where $\boldsymbol{X}_i = \{\text{Re}(Z_j) : Z_j \in \boldsymbol{Z}_i\}$, $\boldsymbol{Y}_i = \{\text{Im}(Z_j) : Z_j \in \boldsymbol{Z}_i\}$, $u_i^x = \text{Re}(u_i)$, $u_i^y = \text{Im}(u_i)$. The corresponding controlled variational dynamic is

$$d\xi_i^x = \left[-\xi_i^x - \xi_i^y + (\xi_i^x + s_x)((\xi_i^x + s_x)^2 + (\xi_i^y + s_y)^2) - s_x(s_x^2 + s_y^2) + \frac{1}{10}\sum_{j=1}^{10}(\xi_i^x - \xi_j^x)\right]dt + \mathcal{T}_\gamma(u_i^x(\boldsymbol{\xi}_i^x, \boldsymbol{\xi}_i^y))dB_t,$$

$$d\xi_i^y = \left[\xi_i^x - \xi_i^y + (\xi_i^y + s_y)((\xi_i^x + s_x)^2 + (\xi_i^y + s_y)^2) - s_y(s_x^2 + s_y^2) + \frac{1}{10}\sum_{j=1}^{10}(\xi_i^y - \xi_j^y)\right]dt + \mathcal{T}_\gamma(u_i^y(\boldsymbol{\xi}_i^x, \boldsymbol{\xi}_i^y))dB_t,$$

$$\dot{s}_x = -s_x - s_y + s_x(s_x^2 + s_y^2), \ \dot{s}_y = s_x - s_y + s_y(s_x^2 + s_y^2), \ s_x(0)^2 + s_y(0)^2 = 1,$$

$$\mathcal{T}_\gamma(x) = \begin{cases} 0, & x > \gamma, \\ x, & -\gamma \le x \le \gamma, \\ 0, & x < -\gamma, \end{cases}$$

with $\gamma = 50$. We select a large γ here because a large upper bound of the controller can accelerate the convergence of the training process. Since the input features of these three controllers are different, the total learnable parameters are different under the same neural network architecture. For a comprehensive comparison, we carefully design network architectures for controllers in three distinct cases, ensuring that the corresponding numbers of parameters are nearly the same. We parameterize the functions $V(\boldsymbol{x})$ as ICNN(20, 120, 120), the controller as Control(20, 60, 60, 20) in full communication case, Control(10, 44, 44, 10) for two clusters of oscillators in half communication case, Control(2, 21, 21, 2) for each oscillator in decentralized communication case with $\mathcal{F} = \text{ReLU}$. We set the hyperparameter as $b = 2.1$, lr = 3e-2, iters = 1e3. The synchronization manifold $\boldsymbol{s}$ is pre-generated on 5000 regular time points on time interval $[0, 10]$. For training data of $\boldsymbol{\xi}$, we sample 2000 points from uniform distribution $\mathcal{U}([-3, 3]^{20})$. The efficacy of the learned neural stochastic controllers is rigorously tested across 50 sample trajectories for each case on the time interval $[0, 10]$ with dt = 2e-3, and random seeds range in $\{0, \cdots, 49\}$.

6.3. Networked Dynamics of FitzHugh-Nagumo Systems

The mathematical formulation of the noise-controlled and coupled FitzHugh-Nagumo systems is governed by

$$
\mathrm{d}v_i = \left[\left(v_i - \frac{v_i^3}{3} - w_i \right) + c \sum_{j=1}^{n} A_{ij} \left(h(v_j) - h(v_i) \right) \right] \mathrm{d}t + \mathcal{T}_\gamma(u_i^v(\boldsymbol{\xi}))\mathrm{d}B_i(t),
$$

$$
\mathrm{d}w_i = 0.1(v_i + 0.7 - 0.8w_i)\mathrm{d}t + \mathcal{T}_\gamma(u_i^v(\boldsymbol{\xi}))\mathrm{d}B_i(t),
$$

$$
h(x) = \frac{1}{1 + e^{-v/v_{th}}}, \tag{S6.16}
$$

$$
\mathcal{T}_\gamma(x) = \begin{cases} 0, & x > \gamma, \\ x, & -\gamma \le x \le \gamma, \\ 0, & x < -\gamma, \end{cases}
$$

where h is set as the sigmoid-type coupling function with threshold $v_{th} = 0.1$ [25], $\boldsymbol{A} = (A_{ij} \in \{0,1\})$ is the coupling structure that reveals the realistic interaction of the oscillators, and $c > 0$ is the coupling strength. Here we consider 4 different directed graph structures with 128 nodes, belonging to three major kinds of network structures:

- A realistic food-web network Baydry with a flat degree distribution [26], the data of the network is provided in `Food Webs`;

- A directed Erdös-Rényi (ER) network [27] with average degree $\bar{k} = 4$ generated by the `Networkx` package in Python, this network is a homogeneous network;

- Two directed scale-free (SF) networks with different in and out degree distribution generated by the method in [28], s.t., SF1: $p(k_{\mathrm{in}}) \sim k^{-2.5}$, $p(k_{\mathrm{out}}) \sim k^{-2.5}$; SF2: $p(k_{\mathrm{in}}) \sim k^{-2.1}$, $p(k_{\mathrm{out}}) \sim k^{-2.9}$. From the perspective of the homogeneity, the rank of these networks are Baydry>ER>SF1 >SF2.

In this study, we simply set the trajectory of the first oscillator as the synchronization manifold, i.e., $\boldsymbol{s}(t) = (v_1(t), u_1(t))^\top \triangleq (s_v(t), s_u(t))^\top$. Actually, readers can randomly select any oscillator in the coupled dynamics as a driving system engendering the synchronization manifold, or use the mean field of the oscillators' trajectories as the synchronization manifold. To demonstrate the efficacy of the extended framework on linear variational dynamics as proposed in Section. S3, we consider the following variational dynamic,

$$
\mathrm{d}\boldsymbol{\xi} = \left(I \otimes \nabla \boldsymbol{M}_0 + c\boldsymbol{L} \otimes \nabla \boldsymbol{M}_1 \right) \boldsymbol{\xi}\mathrm{d}t + \mathcal{T}_\gamma(\boldsymbol{u}(\boldsymbol{\xi}))\mathrm{d}\boldsymbol{B}_t,
$$

$$
\boldsymbol{\xi} = (\xi_1, \xi_2, \cdots, \xi_{128})^\top, \ \xi_i = (\xi_i^v, \xi_i^w)^\top = (v_i - s_v, w_i - s_w)^\top, \ i = 1, \cdots, 128,
$$

$$
\mathcal{T}_\gamma(\boldsymbol{u}(\boldsymbol{\xi}))\mathrm{d}\boldsymbol{B}_t = (\mathcal{T}_\gamma(u_1^v(\boldsymbol{\xi}), u_1^w(\boldsymbol{\xi}))^\top \mathrm{d}B_1(t), \mathcal{T}_\gamma(u_2^v(\boldsymbol{\xi}), u_2^w(\boldsymbol{\xi}))^\top \mathrm{d}B_2(t), \cdots, \mathcal{T}_\gamma(u_{128}^v(\boldsymbol{\xi}), u_{128}^w(\boldsymbol{\xi}))^\top \mathrm{d}B_{128}(t)),
$$

where $L_{ij} = \delta_{ij}(\sum_{j=1}^n A_{ij}) - A_{ij}$ is the Laplacian matrix of coupling structure $\boldsymbol{A}$, the coupling strength is fixed as $c = 0.1$. For brevity, we directly use the dominant parts of the Jacobian matrices $\nabla \boldsymbol{M}_{0,1}$ as

$$
\nabla \boldsymbol{M}_0 = \begin{bmatrix} 1 & -1 \\ 0.1 & -0.08 \end{bmatrix}, \quad \nabla \boldsymbol{M}_1 = \begin{bmatrix} \frac{1}{v_{th}} & 0 \\ 0 & 0 \end{bmatrix}.
$$

We note that the second term embracing the Laplacian matrix in the diffusion term is semi-positive $(+cL \otimes \nabla M_1)$ instead of the semi-negative $(-cL \otimes \nabla M_1)$ considered in the previous works [6, 29]. The latter one contributes a negative part in the Lyapunov exponent of the variational dynamics, which makes the CS task more easily realizable. Therefore, we consider a more difficult CS task with the former one to illustrate the capability of our framework. We consider the following cases of the noise controllers.

(i) Adding external forces via common noise and uncorrelated noise: We focus on the realistic Baydry network controlled by the common noise $B_i(t) = B(t)$, $i = 1, \cdots, 128$ and the uncorrelated noise $\mathrm{Cor}(B_i(t), B_j(t)) = \delta_{ij}t$. For comprehensively investigating the impact of the correlation between the driving noise, we consider three different scales of the noise: Low scale with $\gamma = 1.0$, middle scale with $\gamma = 1.5$, and high scale with $\gamma = 2.0$. We parameterize the functions $V(\boldsymbol{x})$ as ICNN(256, 768), the controller as Control(256, 768, 768, 256) with $\mathcal{F} = \mathrm{Tanh}$. We set the hyperparameter as $b = 2.5$, lr = 1e-2, and iters = 5e2. For the training data of $\boldsymbol{\xi}$, we sample 2000 points from the uniform distribution $\mathcal{U}([-4.7, 4.7]^{256})$ (4.7 is a tight upper bound of the difference between the two FHN oscillators). The efficacy of the learned neural stochastic controllers is rigorously tested across 10 sample trajectories for each case on time interval $[0, 200]$ with $\mathrm{d}t = $ 1e-2 and random seeds range in $\{0, \cdots, 9\}$.

(ii) Perturbing the network structure $\boldsymbol{A}$: We aim at finding the optimal noise perturbation structure $\boldsymbol{\delta A} = (\delta A_{ij} \in [-\gamma, \gamma])$ under the common noise $B_i(t) = B(t), \ i = 1, \cdots, 128$. The noise term is

$$(\tilde{\boldsymbol{L}} \otimes \nabla \boldsymbol{M}_1)\boldsymbol{\xi}dB_t, \ \tilde{L}_{ij} = \left(\delta_{ij}\sum_{j=1}^{128}(\boldsymbol{\delta A})_{ij}\right) - (\boldsymbol{\delta A})_{ij}.$$

We consider different upperbounds of the perturbation matrix $\boldsymbol{\delta A}_{\max}$ in $\{0, 0.1, 0.2, \cdots, 2.0\}$ for four networks. We parameterize the perturbation matrix as $\boldsymbol{A} \odot \boldsymbol{W}$, where $\odot$ is the Hadamard product and $\boldsymbol{W} \in \mathbb{R}^{128 \times 128}$ is the trainable matrix. The V function, the other hyperparameters, and the generating process are the same as those specified in (i).

Connection with related work. Sherwood et al discussed the impact of correlated noise on the synchronization of the Kuramoto model, they further provided a method to find the optimal correlation matrix of the driving noise based on their theoretical results [30]. However, their method relies on the stationary distribution of the oscillatory phase deduced from the corresponding Fokker-Planck equation, which is hard to apply to the general oscillators. We now show our framework can find the optimal correlated noise easily by a reparameterization trick. Consider a set of colored noise engendered by the linear combination of uncorrelated standard Brownian motion $[B_1(t), B_2(t), \cdots, B_n(t)]$ as

$$\begin{bmatrix} \tilde{B}_1 \\ \tilde{B}_2 \\ \vdots \\ \tilde{B}_n \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{bmatrix} \begin{bmatrix} B_1 \\ B_2 \\ \vdots \\ B_n \end{bmatrix}.$$

From the property of the Gaussian noise, we have

$$\text{Cov}\left(\begin{bmatrix} B_1 \\ B_2 \\ \vdots \\ B_n \end{bmatrix}\right) = I_{n \times n} \rightarrow \text{Cov}\left(\begin{bmatrix} \tilde{B}_1 \\ \tilde{B}_2 \\ \vdots \\ \tilde{B}_n \end{bmatrix}\right) = \begin{bmatrix} \sum_{i=1}^{n} c_{1i}^2 & \sum_{i=1}^{n} c_{1i}c_{2i} & \cdots & \sum_{i=1}^{n} c_{1i}c_{ni} \\ \sum_{i=1}^{n} c_{1i}c_{2i} & \sum_{i=1}^{n} c_{2i}^2 & \cdots & \sum_{i=1}^{n} c_{2i}c_{ni} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{n} c_{1i}c_{ni} & \sum_{i=1}^{n} c_{2i}c_{ni} & \cdots & \sum_{i=1}^{n} c_{ni}^2 \end{bmatrix}. \quad \text{(S6.17)}$$

To well define the correlated noise, we restrict the rows of the coefficient matrix $\boldsymbol{C} = (c_{ij})$ in the unit ball $\|(c_{i1}, c_{i2}, \cdots, c_{in})\|_2 \leq 1, \ i = 1, \cdots, n$, s.t. the covariance matrix in (S1.2) is semi-positive with spectrum staying in $[-1, 1]$. By setting the elements in the coefficient matrix as the trainable parameters and integrating this restriction into the loss function, we can find the optimal correlated noise that achieves the synchronization. Take finding the optimal perturbation network structure under correlated noise for an example, the noise term is

$$(d\tilde{\boldsymbol{B}}_t \otimes I_{2 \times 2})(\tilde{\boldsymbol{L}} \otimes \nabla \boldsymbol{M}_1)\boldsymbol{\xi} = (\boldsymbol{C}d\boldsymbol{B}_t \otimes I_{2 \times 2})(\tilde{\boldsymbol{L}} \otimes \nabla \boldsymbol{M}_1)\boldsymbol{\xi},$$

$$\tilde{L}_{ij} = \left(\delta_{ij}\sum_{j=1}^{128}(\boldsymbol{\delta A})_{ij}\right) - (\boldsymbol{\delta A})_{ij}, \ \boldsymbol{\delta A} = \boldsymbol{A} \odot \boldsymbol{W},$$

$$\sum_{j=1}^{128} c_{ij}^2 \leq 1, \ i = 1, \cdots, 128,$$

where $\boldsymbol{W}$ and $\boldsymbol{C}$ are trainable matrices, and to guarantee the inequality constraints, we can add them as regularization terms into the loss function as $L_{\text{ineq}} = \sum_{i=1}^{128} \text{ReLU}\left(\sum_{j=1}^{128} c_{ij}^2 - 1\right)$.

6.4. Controlling Non-autonomous Dynamics

We consider the task of synchronizing the coupled FitzHugh-Nagumo systems with a temporal interacting network via external forces, as a validation of the extended framework proposed in Section S4. We apply the temporal network $\boldsymbol{A}(t)$ proposed by Yuanzhao Zhang et al [6]:

$$
A_{ij}(t) = \begin{cases} \dfrac{1 + (6 - \frac{8}{n+1})A\sin(wt)}{n}, & i, j \leq \dfrac{n+1}{2}, \; i \neq j, \\[3mm] \dfrac{1 - 2A\sin(wt)}{n}, & i \text{ or } j > \dfrac{n+1}{2}, \; i \neq j, \end{cases}
$$

where n is odd. The eigenvalues of the corresponding Laplacian matrix are

$$
\lambda_i(t) = \begin{cases} 0, & i = 1, \\[2mm] 1 + 2A\sin(wt), & i = 2, \cdots, \dfrac{n+1}{2}, \\[2mm] 1 - 2A\sin(wt), & i = \dfrac{n+3}{2}, \cdots, n. \end{cases} \tag{S6.18}
$$

The dynamics of noise-controlled and coupled FitzHugh-Nagumo systems is the same in (S6.16) with $\boldsymbol{A} = \boldsymbol{A}(t)$. The coupling strength is set as $c = 1$ because the temporal matrix has already been normalized. Notice that the transition
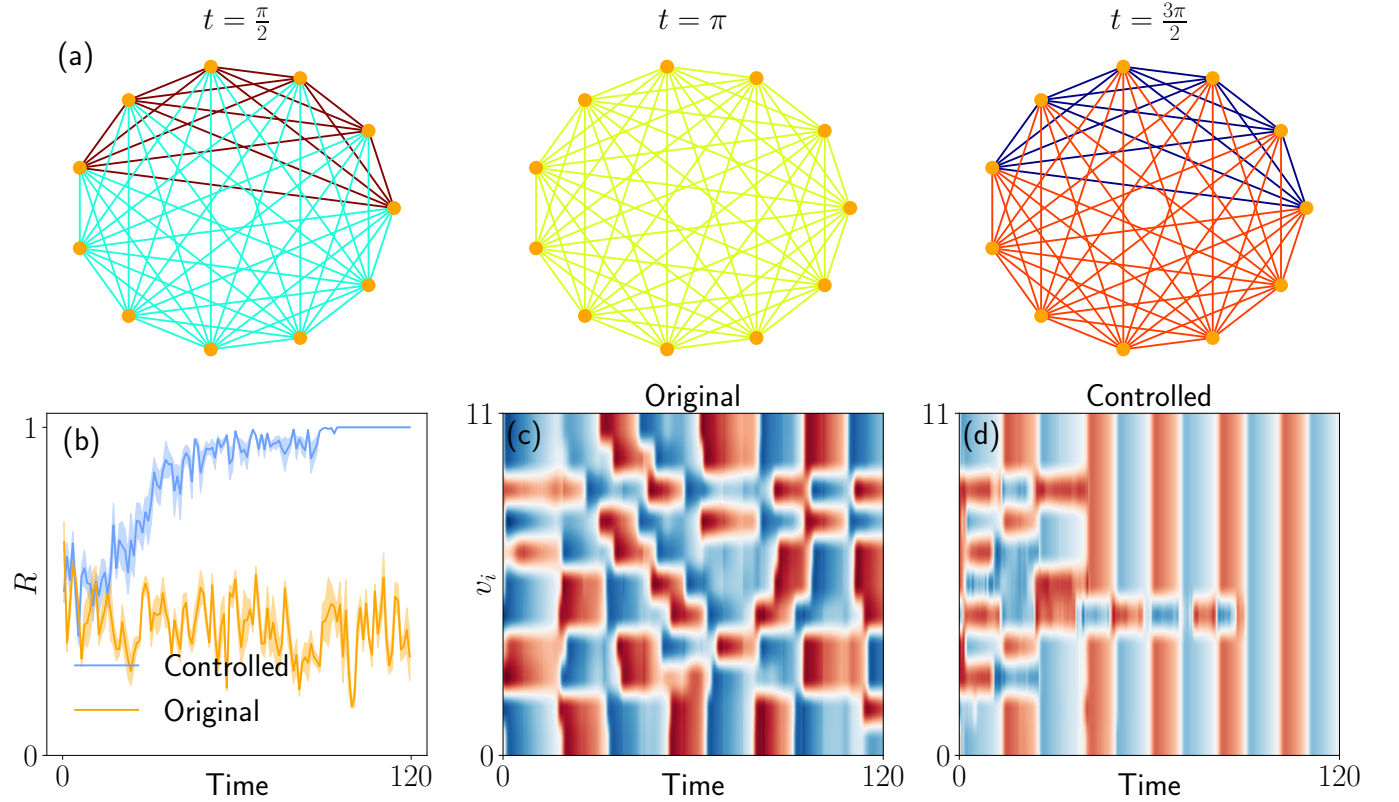

FIG. S1. Synchronizing the coupled FitzHugh-Nagumo systems with the temporal network. (a) The temporal structure. (b) The order parameter of the controlled coupled dynamics, the solid line is the average value, and the shaded region represents the variance. The order parameter is calculated in the time window $\Delta t = 1$ as $[t - 0.5, t + 0.5]$ for $t \in [0.5, 119.5]$. The results are summarized from 10 trajectories. (c) The heatmap of the variables $\{v_i, \; i = 1, \cdots, 11\}$ over time of the original dynamic, (d) and the controlled dynamic.

matrix in the Jordan normal form of $\boldsymbol{A}(t)$ is time invariant. Thus, we set the dominant linear variational dynamic as

$$\mathrm{d}\boldsymbol{\xi} = \left(I \otimes \nabla \boldsymbol{M}_0 + \tilde{\boldsymbol{L}} \otimes \nabla \boldsymbol{M}_1\right)\boldsymbol{\xi}\mathrm{d}t + \mathcal{T}_\gamma(\boldsymbol{u}(\boldsymbol{\xi}))\mathrm{d}\boldsymbol{B}_t,$$

$$\tilde{\boldsymbol{L}} = \sum_{i=2}^{n} \tilde{\lambda}_i \boldsymbol{v}_i \boldsymbol{v}_i^\top, \ \ \boldsymbol{v}_i = \left(\underbrace{\frac{1}{\sqrt{i(i-1)}}, \cdots, \frac{1}{\sqrt{i(i-1)}}}_{i-1}, -\frac{i-1}{\sqrt{i(i-1)}}, \underbrace{0, \cdots, 0}_{n-i} \right)^\top,$$

$$\tilde{\lambda}_i = 1 + 2A \geq \lambda_i, \ i = 2, \cdots, n,$$

$$\nabla \boldsymbol{M}_0 = \begin{bmatrix} 1 & -1 \\ 0.1 & -0.08 \end{bmatrix}, \ \ \nabla \boldsymbol{M}_1 = \begin{bmatrix} \frac{1}{v_{th}} & 0 \\ 0 & 0 \end{bmatrix}. \tag{S6.19}$$

Specifically, we set $A = 1$, $w = 1$, $n = 11$, and $\gamma = 5.0$. We parameterize the functions $V(\boldsymbol{x})$ as ICNN$(22, 66)$, the controller as Control$(22, 66, 66, 22)$ with $\mathcal{F} = $ Tanh. We set the hyperparameter as $b = 2.5$, lr $= $ 1e-2, and iters $= $ 5e2. For the training data of $\boldsymbol{\xi}$, we sample 2000 points from uniform distribution $\mathcal{U}([-4.7, 4.7]^{22})$. The efficacy of the learned neural noise controller is tested across 10 sample trajectories on time interval $[0, 120]$ with $\mathrm{d}t = $ 1e-2 and random seeds range in $\{0, \cdots, 9\}$. Figure S1 shows the artificially designed noise successfully synchronizes the coupled dynamics with temporal networks.

6.5. Pinning Control in the Networked Dynamics of FitzHugh-Nagumo Systems

In the main text, we have investigated the impact of the intrinsic structure of the intra-dynamic to synchronization, by considering pinning control in the driving-response Lorenz system. In addition to this kind of control, controlling a small fraction of oscillators in the collective dynamics has been extensively studied in the last two decades [2, 31–33]. In this section, we consider the CS tasks in coupled FitzHigh-Nagumo systems via pinning control based on the interaction structure. We consider the following controlled FitzHugh-Nagumo dynamics on an ER network with 39 nodes,

$$
\mathrm{d}v_{i_k} = \left[\left(v_{i_k} - \frac{v_{i_k}^3}{3} - w_{i_k} \right) - c \sum_{j=1}^{39} L_{i_k j} v_j \right] \mathrm{d}t + \mathcal{T}_\gamma(u_{i_k}^v(\boldsymbol{\xi})) \mathrm{d}B_i(t),
$$

$$
\mathrm{d}w_{i_k} = 0.1(v_{i_k} + 0.7 - 0.8 w_{i_k}) \mathrm{d}t + \mathcal{T}_\gamma(u_{i_k}^v(\boldsymbol{\xi})) \mathrm{d}B_i(t), \ i_k \in \mathcal{N}_{\mathrm{pin}},
$$

$$
\mathrm{d}v_{\tilde{i}_k} = \left[\left(v_{\tilde{i}_k} - \frac{v_{\tilde{i}_k}^3}{3} - w_{\tilde{i}_k} \right) - c \sum_{j=1}^{39} L_{\tilde{i}_k j} v_j \right] \mathrm{d}t,
$$

$$
\mathrm{d}w_{\tilde{i}_k} = 0.1(v_{\tilde{i}_k} + 0.7 - 0.8 w_{\tilde{i}_k}) \mathrm{d}t, \ \tilde{i}_k \notin \mathcal{N}_{\mathrm{pin}},
$$

$$
L_{ij} = \delta_{ij} \left(\sum_{j=1}^{39} A_{ij} \right) - A_{ij}
$$

$$
\mathcal{T}_\gamma(x) = \begin{cases} 0, \ x > \gamma, \\ x, \ -\gamma \le x \le \gamma, \\ 0, \ x < -\gamma, \end{cases}
$$

where $\mathcal{N}_{\mathrm{pin}}$ represents the set of controlled nodes. Here, we select $\mathcal{N}_{\mathrm{pin}}$ as the root nodes and the feedback vertex set [34] of the network. We fix $c = $ 1e-3, $\gamma = 20$, and we generate the directed ER network with mean degree $\bar{k} = 4$ by the `Networkx` package in Python. We parametrize the functions $V(\boldsymbol{x})$ as ICNN(78, 156), the controller as Control(78, 156, 78) with $\mathcal{F} = $ Tanh. We set the hyperparameter as $b = 2.1$, lr = 1e-2, iters = 5e2. For training data of $\boldsymbol{\xi}$, we sample 3000 points from uniform distribution $\mathcal{U}([-4.7, 4.7]^{22})$. The pinning controller is tested on a rather long time interval $[0, 8000]$ because only a small fraction of the nodes are controlled. The results of the pinning controller are summarized in Fig. S2, the noise-driven pinning controller successfully increases the synchronization degree over time.

S7. LIMITATIONS AND PERSPECTIVE

The proposed framework is applicable only to the CS tasks of identical oscillators. However, the oscillators in the collective dynamics may have different self-dynamics in most realistic scenarios such as the Escherichia coli cells [35]. In this case, researchers delve into the phase synchronization (PS) instead of the CS. The urgent call for cultivating a framework for enhancing PS for general oscillators similarly holds to that for CS as we reported. Although our framework cannot solve this paramount problem directly, our current work is inspiring us to find a way to integrate the PS theory with the machine learning techniques. Frankly, how to find the phase dynamics of the general oscillators is still unsolved presently.

In addition, the closed-loop (also known as the feedback) control we considered in this work has significant advantages over the open-loop control due to its online property and robustness. However, the developed controller is required to change its states continuously in time, which likely causes a huge computational burden as we deploy it into real-life scenarios. One of our future works is to improve the current framework by replacing the closed-loop controller with the piece-wise constant feedback controller.
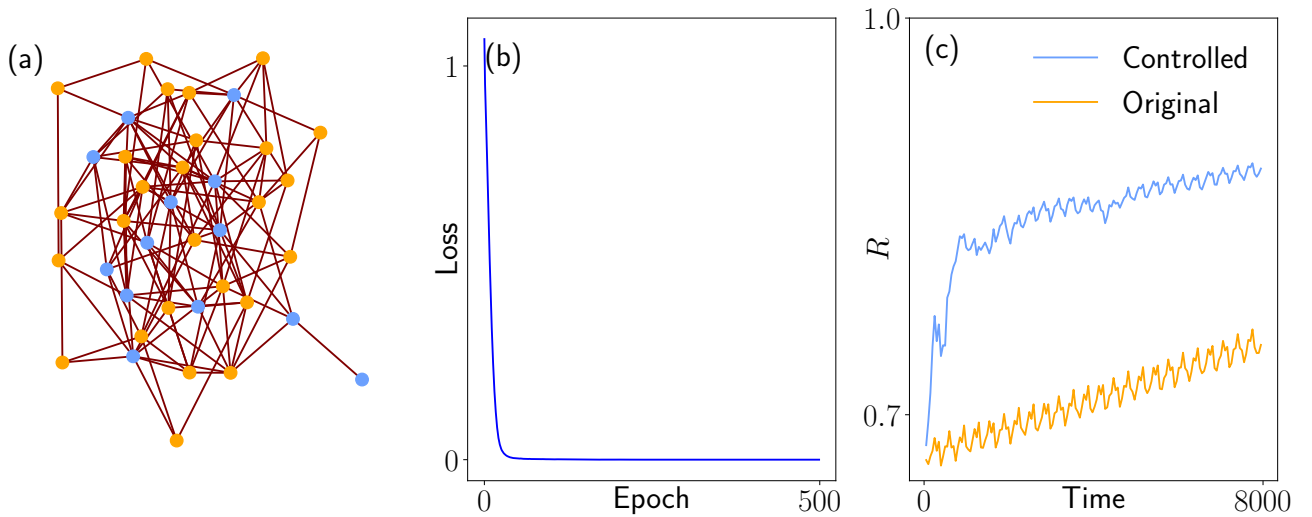

FIG. S2.   Synchronizing the coupled FitzHugh-Nagumo models with pinning control. (a) The ER network with pinning controller acting on the blue nodes. (b) The training loss in the training process. (c) The order parameter of the controlled and the original dynamics, calculated in the time window $\Delta t = 100$ as $[t - 50, t + 50]$ for $t \in [50, 7950]$.

[1] J. Zhang, Q. Zhu, and W. Lin, in Advances in Neural Information Processing Systems, Vol. 35 (2022) pp. 9098–9110.

[2] W. Yu, G. Chen, J. Lu, and J. Kurths, SIAM J. Control Optim. **51**, 1395 (2013).

[3] W. Cui, Y. Jiang, B. Zhang, and Y. Shi, in Advances in Neural Information Processing Systems, Vol. 36 (Curran Associates, Inc., 2024).

[4] C. Zhou and J. Kurths, Phys. Rev. Lett. **88**, 230602 (2002).

[5] C. Zhou, J. Kurths, I. Z. Kiss, and J. L. Hudson, Phys. Rev. Lett. **89**, 014101 (2002).

[6] Y. Zhang and S. H. Strogatz, Nat. Commun. **12**, 3273 (2021).

[7] X. Mao, Stochastic Differential Equations And Applications (Elsevier, 2007).

[8] B. Oksendal, Stochastic Differential Equations: An Introduction With Applications (Springer Science & Business Media, 2013).

[9] K. L. Chung and P. Erdös, Trans. Am. Math. Soc. **72**, 179 (1952).

[10] B. Amos, L. Xu, and J. Z. Kolter, in International Conference on Machine Learning (PMLR, 2017) pp. 146–155.

[11] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida, in International Conference on Learning Representations (2018).

[12] Y. Yoshida and T. Miyato, arXiv preprint arXiv:1705.10941 (2017).

[13] H. Gouk, E. Frank, B. Pfahringer, and M. J. Cree, Machine Learning **110**, 393 (2021).

[14] H.-T. D. Liu, F. Williams, A. Jacobson, S. Fidler, and O. Litany, arXiv preprint arXiv:2202.08345 (2022).

[15] L. M. Pecora and T. L. Carroll, Phys. Rev. Lett. **80**, 2109 (1998).

[16] T. Yamada, J. Math. Kyot. Univ. **13**, 497 (1973).

[17] V. I. Arnold, Ordinary Differential Equations (Springer Science & Business Media, 1992).

[18] S. Lang, Algebra, Vol. 211 (Springer Science & Business Media, 2012).

[19] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, et al., in Advances in neural information processing systems, Vol. 32 (2019).

[20] D. P. Kingma and J. Ba, arXiv:1412.6980 (2014).

[21] W. Grathwohl, R. T. Chen, J. Bettencourt, I. Sutskever, and D. Duvenaud, in International Conference on Learning Representations (2018).

[22] Y. Song, S. Garg, J. Shi, and S. Ermon, in Uncertainty in Artificial Intelligence (PMLR, 2020) pp. 574–584.

[23] R. T. Chen, Y. Rubanova, J. Bettencourt, and D. K. Duvenaud, in Advances in neural information processing systems, Vol. 31 (2018).

[24] J. Li, X. Li, and L. Shijun, Research Square [https://doi.org/10.21203/rs.3.rs-49069/v1] (2020).

[25] K. Wang, L. Yang, S. Zhou, and W. Lin, Chaos **33** (2023).

[26] Y.-Z. Sun, S.-Y. Leng, Y.-C. Lai, C. Grebogi, and W. Lin, Phys. Rev. Lett. **119**, 198301 (2017).

[27] P. Erdős, A. Rényi, et al., Publ. Math. Inst. Hung. Acad. Sci. **5**, 17 (1960).

[28] K.-I. Goh, B. Kahng, and D. Kim, Phys. Rev. Lett. **87**, 278701 (2001).

[29] P. De Lellis, M. Di Bernardo, and F. Garofalo, Chaos **18** (2008).

[30] S. Martineau, T. Saffold, T. T. Chang, and H. Ronellenfitsch, Phys. Rev. Lett. **128**, 098301 (2022).

[31] X. F. Wang and G. Chen, Phys. A **310**, 521 (2002).

[32] J. Lu, J. Kurths, J. Cao, N. Mahdavi, and C. Huang, IEEE Trans. Neural Netw. Learn. Syst. **23**, 285 (2011).

[33] T. Chen, X. Liu, and W. Lu, IEEE Trans. Circuits Syst. **54**, 1317 (2007).

[34] C.-C. Wang, E. L. Lloyd, and M. L. Soffa, J. ACM **32**, 296 (1985).

[35] J. Garcia-Ojalvo, M. B. Elowitz, and S. H. Strogatz, Proc. Natl. Acad. Sci **101**, 10955 (2004).