Editors' Suggestion

# Lyapunov spectra of chaotic recurrent neural networks

Rainer Engelken*
*Department of Neuroscience, Zuckerman Institute, Columbia University, New York, New York 10027, USA;*
*Max Planck Institute for Dynamics and Self-Organization, 37077 Göttingen, Germany;*
*and Bernstein Center for Computational Neuroscience, 37077 Göttingen, Germany*

Fred Wolf
*Max Planck Institute for Dynamics and Self-Organization, 37077 Göttingen, Germany;*
*Max Planck Institute for Multidisciplinary Sciences, 37075 Göttingen, Germany;*
*Bernstein Center for Computational Neuroscience Göttingen, 37073 Göttingen, Germany;*
*Cluster of Excellence "Multiscale Bioimaging from Molecular Machines to Networks of Excitable Cells" (MBExC), University of Göttingen, 37075 Göttingen, Germany;*
*and Göttingen Campus Institute for Dynamics of Biological Networks, and Faculty of Physics, University of Göttingen, 37073 Göttingen, Germany*

L. F. Abbott
*Department of Neuroscience, Zuckerman Institute, Columbia University, New York, New York 10027, USA*
*and Department of Physiology and Cellular Biophysics, Columbia University, New York, New York 10032, USA*

(Received 5 June 2020; accepted 5 January 2023; published 16 October 2023)

*This article is part of the* Physical Review Research *collection titled Physics of Neuroscience.*

Recurrent networks are widely used as models of biological neural circuits and in artificial intelligence applications. Mean-field theory has been used to uncover key properties of recurrent network models such as the onset of chaos and their largest Lyapunov exponents, but quantities such as attractor dimension and Kolmogorov-Sinai entropy have thus far remained elusive. We calculate the complete Lyapunov spectrum of recurrent neural networks and show that chaos in these networks is extensive with a size-invariant Lyapunov spectrum and attractor dimensions much smaller than the number of phase space dimensions. The attractor dimension and entropy rate increase with coupling strength near the onset of chaos but decrease far from the onset, reflecting a reduction in the number of unstable directions. We analytically approximate the full Lyapunov spectrum using random matrix theory near the onset of chaos for strong coupling and discrete-time dynamics. We show that a generalized time-reversal symmetry of the network dynamics induces a point symmetry of the Lyapunov spectrum reminiscent of the symplectic structure of chaotic Hamiltonian systems. Temporally fluctuating input can drastically reduce both the entropy rate and the attractor dimension. We lay out a comprehensive set of controls for the accuracy and convergence of Lyapunov exponents. For trained recurrent networks, we find that Lyapunov spectrum analysis quantifies error propagation and stability achieved by different learning algorithms. Our methods apply to systems of arbitrary connectivity and highlight the potential of Lyapunov spectrum analysis as a diagnostic for machine learning applications of recurrent networks.

DOI: 10.1103/PhysRevResearch.5.043044

## I. INTRODUCTION

A major challenge in theoretical neuroscience, machine learning, and statistical physics is to develop mathematical concepts to characterize high-dimensional activity and find collective degrees of freedom and information representations of strongly interacting populations of elements, such as neurons. What mechanisms determine the diversity of collective activity in recurrent networks and how can it be quantified? What are the number of stable and unstable directions along the trajectory and how does this relate to trainability of networks? We address these long-standing questions using firing-rate networks.

Theoretical work suggested that complex rate activity in neural systems may originate from chaotic dynamics in recurrent networks. A seminal study showed that in large networks of randomly connected firing-rate units chaotic dynamics does not emerge through a transition scenario consisting of a sequence of increasingly more complex forms of motion but

*To whom correspondence should be addressed: re2365@columbia.edu

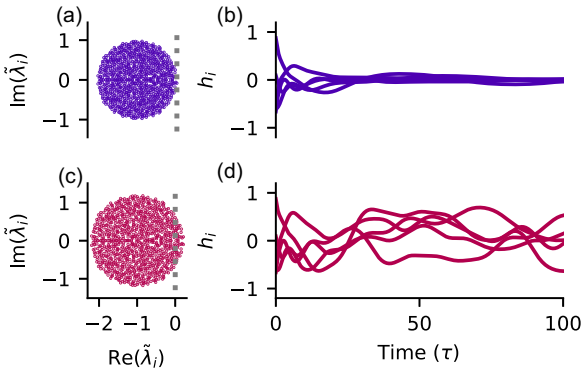


FIG. 1. Transition to chaos for sufficiently strong coupling $g$ in rate networks. (a) Linear stability of rate dynamics near the zero fixed point. Real vs imaginary part of eigenvalues $\hat{\lambda}_i$ of the stability matrix for $g = 0.99$. (b) For subcritical couplings ($g = 0.99$) the trivial fixed point of the system $h_i = 0$ is the only stable solution. (c) In large networks the trivial fixed point loses stability at $g_{\text{crit}} = 1$ and chaos emerges from the nonlinear interaction of rate units where the spectral radius crosses unity (gray dotted line). (d) Rate chaos for $g = 1.2$ (other parameters: network size $N = 1000$, integration step $\Delta t = 10^{-3}\tau$).

through a single sharp transition from an inactive state to a chaotic state [1] (Fig. 1). In this class of models, each rate unit maps its synaptic input $h_i$ smoothly into a firing rate through a hyperbolic tangent input-output transfer function $\phi$. Coupling strengths are drawn independently from a Gaussian distribution with zero mean and standard deviation $g/\sqrt{N}$, where $N$ is the size of the network. Dynamic mean-field theory is applicable in the large network limit $N \to \infty$. In this approach, the recurrent input into a typical unit is modeled by a Gaussian process whose statistics is determined self-consistently. For small coupling $g < 1$, the trivial fixed point $h_i = 0$ for all $i$ is the only stable solution to the mean-field theory [Figs. 1(a) and 1(b)]. For increasing coupling strength, this trivial fixed point loses stability and chaos emerges from the nonlinear interaction of unstable activity modes [Figs. 1(c) and 1(d)]. Sompolinsky, Crisanti, and Sommers showed that in the large network limit $N \to \infty$ above a critical strength $g_{\text{crit}} = 1$, the only stable self-consistent solution is chaotic dynamics [1]. The transition to chaos occurs when the spectral radius $\hat{\lambda}_{\max}$ of the stability matrix obtained from linearizing the rate dynamics around the fixed point $h_i = 0$ crosses unity [Figs. 1(a) and 1(c)].

This classical work has been extended, and the transition has been studied for networks with different subpopulations [2–4], various input-output transfer functions [2,4], bistable units [5], adaptation [6], sparse balanced network architectures [2,4,7], and external stimuli [8–11]. For networks of spiking model neurons, quantitative agreement with a corresponding chaotic rate network in the limit of slow synaptic dynamics was found [4,7] (see also [12]).

The chaotic, heterogeneous state of rate networks possesses high computational capabilities. These arise from its rich internal dynamics that can provide a substrate for complex nonlinear computations, e.g., implementing input/output maps [13–15] and learning temporal sequences [16]; however, it is a challenge to extend this to spiking neural networks [17–21]. Some studies proposed that computational features are favorable near to or even slightly beyond the so-called *edge of chaos* in the chaotic regime [11,15,22–27]. It was claimed and questioned much earlier in dynamical systems research that the edge of chaos is computationally advantageous [28–30].

Recent developments in machine learning, including the renaissance of deep networks, have sparked additional interest in principles of stability and information processing in recurrent rate networks [31,32]. One reason for this is that recurrent networks can be unrolled in time into infinitely deep feed-forward networks with tied weights [33]. To avoid vanishing or exploding gradients during learning, this analogy suggests that learning in deep nonlinear networks is facilitated if the weights are initialized such that the corresponding recurrent networks are close to the edge of chaos ($g_{\text{crit}} = 1$) [31,34–37]. Intriguingly, transient rate chaos yields exponential expressivity in deep networks [32,38,39].

Here we calculate the full set of Lyapunov exponents of classical continuous-time firing-rate networks. Previous analytical studies only considered the largest Lyapunov exponent, but the full Lyapunov spectrum provides valuable additional insights into the collective dynamics of firing-rate networks.

We use concepts from the ergodic theory of dynamical systems to further characterize the complex collective dynamics of rate networks. Often large-scale dissipative systems evolve towards a low-dimensional attractor, but it is a challenge to identify and characterize this lower-dimensional manifold. Ergodic theory provides an estimate of the attractor dimensionality by characterizing the diversity of collective network activity states [40]. It also provides access to the dynamical entropy rate, which measures the amplification of uncertainty due to sensitivity to initial conditions. The dynamical entropy rate constrains the capability of information processing. Given that the initial state is known only with finite precision, the sensitive dependence on initial conditions makes predictions of future states impossible in chaotic systems [41,42]. This corresponds to a dynamical entropy rate because nearby states, which cannot be distinguished by a finite precision readout initially, are pulled apart by the chaotic dynamics and become distinguishable later on. Therefore, the dynamical entropy rate quantifies the speed at which microscopic perturbations affect macroscopic rate fluctuations [41]. Sensitivity to initial conditions in cortical circuits might serve as a dynamical mechanism to pull nearby trajectories apart [43–45]. If the microscopic initial state contains a relevant signal, the dynamical entropy rate measures the rate by which this information becomes accessible. From a neural coding perspective, the dynamical entropy rate can contribute to the so-called noise entropy [46] because the dynamic amplification of microscopic noise by chaotic dynamics can impair coding capacity.

Both the dynamical entropy rate and attractor dimensionality are invariants of dynamical systems, i.e., they do not change under diffeomorphisms of the phase space [47–50] and can be obtained from the set of Lyapunov exponents [51]. This is the only known general way of calculating the entropy of a high-dimensional differentiable dynamical system [40]. Sampling-based methods, such as the Grassberger-Procaccia algorithm [52–54], which estimates the correlation dimension $D_2$, are intractable for high-dimensional systems, because the

amount of data required to estimate the dimensionality $D$ scales exponentially with the dimensionality $D$ itself [55–58].

### Outline

After introducing the model in Sec. II, we demonstrate extensive entropy rate and dimensionality of canonical chaotic rate networks in Sec. III. Moreover, we show that both the dynamical entropy rate and the attractor dimensionality peak with coupling strength $g$, while the largest Lyapunov exponent keeps growing with large $g$. We also find that time discretization increases both the entropy rate and dimensionality. Using random matrix theory, we analytically approximate the full Lyapunov spectrum in several limiting cases in Sec. IV. In Sec. V, we extend the analysis to a balanced network of threshold-linear units. In Sec. VI we show the point symmetry of the Lyapunov spectrum of continuous-time networks around the negative inverse timescale. We demonstrate in Sec. VII that the first covariant Lyapunov vector is extended. In Sec. VIII, we find that time-varying input reduces both the entropy rate and dimensionality. Finally, in Sec. IX, we use Lyapunov spectra to quantify the stability of trained networks, calculate Lyapunov spectra of LSTM networks in Sec. X and describe in Sec. XI a mathematical link between gradients of backpropagation through time and the Lyapunov spectrum.

## II. MODEL

We study the dynamics of a randomly wired network of nonlinear firing-rate units. The dynamics of the state $h_i$ for $i = 1, 2, \ldots, N$, of each firing-unit follows [1]

$$\tau \frac{dh_i}{dt} = F_i = -h_i + \sum_{j=1}^{N} J_{ij}\phi(h_j). \tag{1}$$

Here $h_i$ is the total synaptic current received by firing-rate unit $i$ and $\tau$ is the rate-unit time constant. We draw independent identically distributed entries of the coupling matrix $J_{ij}$ from a Gaussian distribution $J_{ij} \sim \mathcal{N}(0, g^2/N)$, remove self-coupling by setting $J_{ii} = 0$, and choose the transfer function $\phi(x) = \tanh(x)$ [1].

## III. LYAPUNOV SPECTRUM OF CLASSIC RECURRENT NEURAL NETWORKS

To calculate the Lyapunov spectrum, we evaluate the Jacobian of the flow of the dynamics. This measures how infinitesimal perturbations of the network state evolve in the tangent space along the trajectory $h_i$. The instantaneous Jacobian is given by

$$D_{ij}(t_s) = \left.\frac{\partial F_i}{\partial h_j}\right|_{t=t_s} = -\delta_{ij} + J_{ij}\phi'(h_j(t_s)). \tag{2}$$

Thus, in our case, the Jacobian is a negative identity matrix plus the coupling matrix with columns scaled by the squared hyperbolic secant $\phi' = \mathrm{sech}^2$ of the network activity states $h_i$. For strong $g$, the variance of $h_i$ increases proportional to $g^2$ [59] and most rates are in the saturated regime, so $\mathrm{sech}^2(h_i) \approx 0$ for most $i$ and hence most columns of $D_{ij}(t_s)$ are close to zero, aside from the diagonal entries. The full Lyapunov spectrum $\lambda_{\max} \geqslant \lambda_2 \cdots \geqslant \lambda_N$ is obtained by a reorthonormalization procedure [60], which is described in detail in the Supplemental Material [61], including a detailed analysis of the convergence of the Lyapunov spectra.

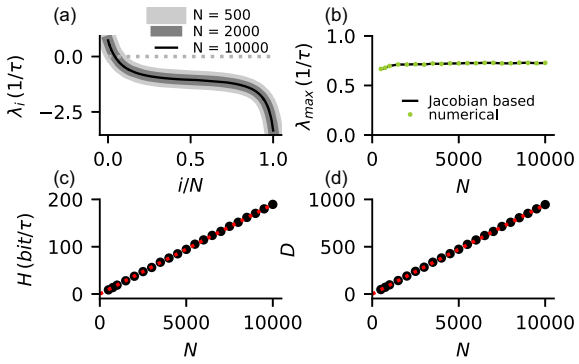


FIG. 2. Extensive chaos revealed by the size invariance of the Lyapunov spectrum. (a) Full Lyapunov spectra for different network sizes $N$ are on top of each other, indicating an identical shape (light gray thick: $N = 500$, dark gray thin: $N = 200$, black: $N = 10\,000$). The Lyapunov spectrum is point symmetric around the mean Lyapunov exponent $\bar{\lambda} = -1/\tau$ (see analytical derivation in Sec. VI and Appendix D). (b) The largest Lyapunov exponent quickly saturates with network size. (c) The Kolmogorov-Sinai entropy rate $H$ grows linearly with $N$ as shown over two orders of magnitude. (d) The same holds for the attractor dimensionality $D$ (black dots: Jacobian-based method, red lines: best fit, other parameters: $g = 10$, $\Delta t = 0.1\tau$, $t_{\mathrm{ONS}} = \tau$, $t_{\mathrm{sim}} = 10^3\tau$).

### A. Extensive spatiotemporal network chaos

In dissipative systems, trajectories converge towards a lower-dimensional attractor. The dimensionality of this attractor can be constant, proportion to the size of the system, or have other more complex dependencies. If it is proportional to the system's size, the system is called extensive. This occurs when the shape of the Lyapunov spectrum is invariant with system size, which also implies an extensive entropy rate.

In the case of the firing-rate networks studied here, we find extensive chaos, indicated by the invariance of the shape of the Lyapunov spectrum to network size $N$ [Fig. 2(a)] for sufficiently large networks (although the structure of the attractor depends on the realization of the connectivity $J_{ij}$). The Lyapunov spectrum is point symmetric around its constant mean value $-1/\tau$. We will investigate the origin of the symmetry of the Lyapunov spectrum in Sec. VI. The largest Lyapunov exponent quickly saturates as a function of network size [Fig. 2(b)]. We investigate the finite-size effect on the largest Lyapunov exponent, its convergence to the value predicted by dynamic mean-field theory, and the finite-size effect in the transition to chaos $g_{\mathrm{crit}}$ in Appendix B. The entropy rate $H$, also called the Kolmogorov-Sinai entropy rate, quantifies the amplification of small state differences by the chaotic dynamics. While formally defined via partitions of the phase space, it is under weak mathematical constraints given by the sum of the positive Lyapunov exponents: $H = \sum_{\lambda_i > 0} \lambda_i$ (see Supplemental Material [61] on Kolmogorov-Sinai entropy rate and Kaplan Yorke attractor dimensionality). For a size-invariant Lyapunov spectrum, it also grows linearly, as demonstrated over two orders of magnitude in Fig. 2(c). The same is true for the attractor dimensionality [Fig. 2(d)], which is given by

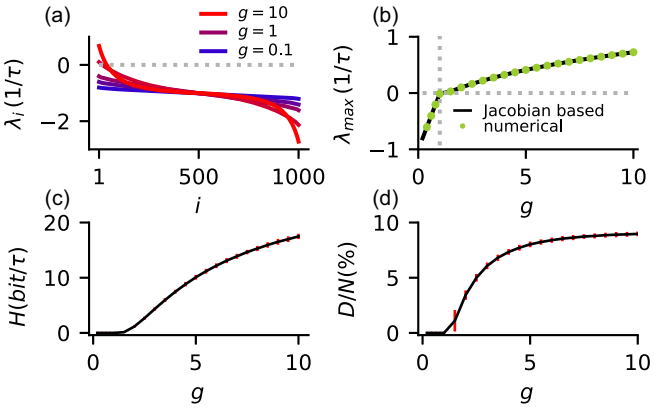


FIG. 3. Entropy rate and attractor dimensionality of tanh firing-rate network dynamics. (a) Full Lyapunov spectra of rate networks with different coupling strengths $g$, color coded from blue (small $g$) to red (large $g$). (b) The largest Lyapunov exponent shows the theoretically predicted linear growth for $g < 1$ and first quadratic and then logarithmic growth for $g > 1$ as a function of $g$ [1]. (Green dots: direct numerical simulations, black line: Jacobian-based method.) (c) The dynamical entropy rate $H$ grows with $g$ for moderate values of $g$ but peaks (see Fig. 4 for large $g$ behavior). (d) Relative attractor dimensionality $D/N$ peaks at $D/N < 10\%$ [see Figs. 4 and 5(a) for large $g$ behavior]. (Averages over 20 network realizations in black, red error bars indicate double std across 20 network realizations, parameters: $N = 1000$, $\Delta t = 10^{-2}\tau$, $t_{\text{sim}} = 10^4\tau$, $t_{\text{ONS}} = \tau$).

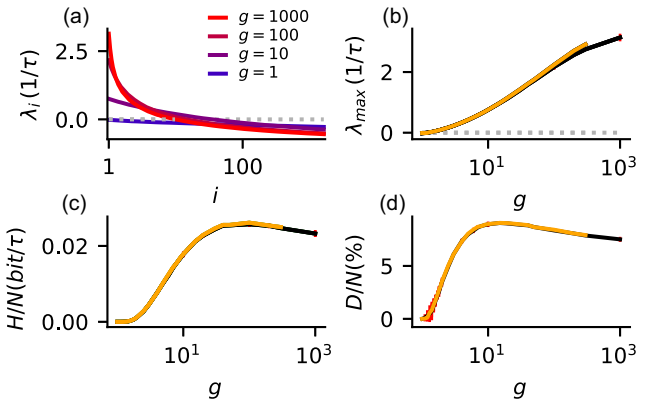


FIG. 4. Peak in dynamical entropy rate and attractor dimensionality for large $g$. (a) For increasing coupling strength $g$, the Lyapunov spectrum is increasingly bent upward with a decreasing fraction of positive Lyapunov exponents ($N = 2000$). (b) The largest Lyapunov exponent grows monotonically for increasing values of $g$ as predicted analytically [see Fig. 21(b)]. For very large $g$, the largest Lyapunov exponents flattens (black lines, $N = 2000$; orange lines, $N = 3000$). Increasing $N$ reduces the flattening, indicating a finite $N$ effect on $\lambda_{\max}$. (c) The dynamical entropy rate $H$ peaks as a function of $g$. (d) The relative attractor dimensionality $D/N$ also peaks as function of $g$. Both positions of the peak in $D$ and $H$ do not shift with $N$ indicating that the peak is not a finite $N$ effect. (Parameters: relative tolerance $= 10^{-10}$, $t_{\text{ONS}} = \tau$, $t_{\text{sim}} = 10^3\tau$, averages across three network realizations).

the interpolated number of Lyapunov exponents that sum to zero,

$$D = k + \frac{\sum_{i=1}^{k} \lambda_i}{|\lambda_{k+1}|} \quad \text{with} \quad k = \max_n \left\{ \sum_{i=1}^{n} \lambda_i \geqslant 0 \right\}. \tag{3}$$

Intuitively, the attractor dimension is the dimensionality of the highest dimensional infinitesimal hypersphere, whose volume does not shrink or grow through the chaotic dynamics. In other words, on the attractor, growth along unstable manifolds is being compensated by shrinking along the stable manifolds. Thus, a $D$-dimensional hypersphere is merely deformed over time, with the volume preserved on average.

The extensivity of the Lyapunov spectrum for continuous-time rate networks was conjectured earlier [1], but never before demonstrated. Extensive chaos is often found in extended systems that are decomposable into locally interacting subsystems, whose number grows linearly with system size [62]. As this is not fulfilled for this fully randomly connected rate network, extensive chaos in our networks is not a trivial property. Globally coupled networks, for instance, can exhibit nonextensive chaos [63].

### B. Entropy rate and dimensionality

Next, we investigate the role of the synaptic coupling strength $g$ (Fig. 3). The full Lyapunov spectrum shows an interesting dependence on $g$. For increasing $g$, the first half of the Lyapunov spectrum is increasingly curved [Fig. 3(a)]. Note that the Lyapunov spectrum is point symmetric for all values of $g$. The largest Lyapunov exponent shows the theoretically predicted $g$-dependence in the stable regime $g < 1$ [Fig. 3(b)]. In the chaotic regime $g > 1$, it grows first quadratically and then logarithmically with $g$ in agreement with previous findings [1,59]. Note that the asymptotic large $g$ behavior $\lambda_{\max} \propto \log(g)$ is only expected when first $N \to \infty$ and then $g \to \infty$. The calculation of the largest Lyapunov exponent is confirmed both by tracking the amplitude of a small perturbation in direct numerical simulations and by using the Jacobian-based method [60] [Fig. 3(b)]. While the exponential separation rate of nearby trajectories increases for growing $g$, the overall dissipation of the system, measured by the mean Lyapunov exponent $\bar{\lambda} = \frac{1}{N}\sum_i \lambda_i$ is independent of $g$ and only depends on the time constant $\tau$. The reasons for this are provided in Appendix D. We now focus first on the entropy rate and attractor dimensionality.

The dynamical entropy rate is zero for $g \leqslant 1$ and first grows for increasing values of $g$ [Fig. 3(c)]. Our numerical results show that for very large $g$, the dynamical entropy rate peaks with $g$ (see Fig. 4) Indeed, we show that $H\tau < D < N$ (Appendix F), when measuring $H$ in units of $nat/\tau$. Again, this asymptotic behavior is only expected when first sending $N$ and then $g \to \infty$. In case the specific initial state of the network does not encode relevant information, the growth of the entropy rate with $g$ can be interpreted as an increasing contribution to noise entropy.

We found that the attractor dimension first increases with $g$ [Fig. 3(d)] in the chaotic regime $g > g_{\text{crit}}$ and peaks as a function of $g$ at less than 10% of the number of phase space dimensions $N$ [Fig. 5(a), see also Appendix F for large $g$ behavior]. This suggests that despite vanishing pairwise correlations [1], rate unit activities are not independent of each other. Even for strongly chaotic networks, the strange attractor of the network dynamics does not fill the entire phase space but only a small but extensive fraction of it. Note that the

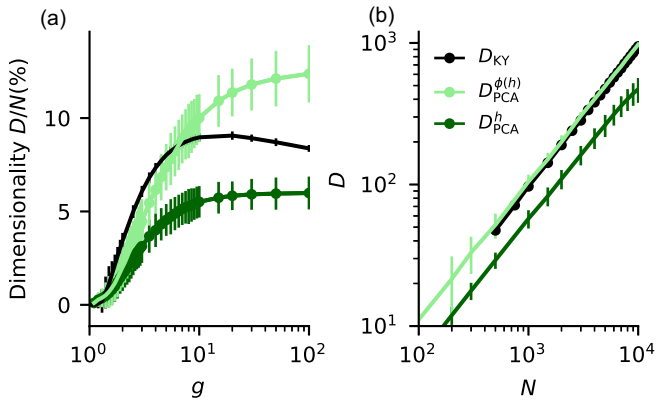


FIG. 5. PCA and attractor dimensions of networks with tanh transfer function. (a) Principal component analysis (PCA)-based dimensionality estimate (green) and attractor dimension based on the Lyapunov spectrum (black) for different values of synaptic strength $g$. Both PCA dimensions saturate for $g \gg 1$, but they saturate at different levels and with distinct exponential rates, while the attractor dimension peaks (error bars are double std across 20 network realizations). PCA dimension estimate of dynamics depends on whether $\tanh(h_i)$ or $h_i$ is considered. (b) Both PCA-based dimensionality estimates are extensive, as indicated by the approximately linear growth with $N$ (other parameters: $N = 1000$ for (a), $g = 10$ for (b), $\Delta t = 0.1\tau$, $t_{\text{ONS}} = \tau$, $t_{\text{sim}} = 10^4\tau$).

geometric structure of the attractor nevertheless changes when $g$ is further increased.

### C. Lyapunov spectra for large coupling $g$

For very large values of $g$, we observe that both attractor dimensionality $D$ and dynamical entropy rate $H$ peak as a function of $g$ (Fig. 4). For increasing network size $N$, this peak does not vanish or shift, indicating that it is not merely a finite-size effect. The peak in both $H$ and $D$ can be explained by a growing fraction of rate units in saturation and, therefore, increasingly sparse Jacobian $D_{ij}(t_s)$, thus fewer active units at each moment and resulting in fewer unstable phase space directions as indicated by the decreasing number of positive Lyapunov exponents. Our seemingly contradictory claim of decreasing dynamical entropy rate $H$ despite a growing largest Lyapunov exponent for large $g$ are consistent for large $N$ and finite $g$, where the relative contribution of the largest Lyapunov exponent to the sum of the positive Lyapunov exponents vanishes.

### D. Comparison of attractor dimension and PCA dimension

In experimental and theoretical neuroscience, dimensionality estimates of the network activity based on principal component analysis (PCA) are commonly used [9,10,64–69], e.g., to quantify the spatiotemporal complexity of neural activity in a data set. We compared the attractor dimensionality with a dimensionality estimate based on second-order statistics of the activity $h_i$ and $\phi(h_i)$ given by the effective number of principal components that account for most of the variance (Supplemental Material [61] for details and definition of PCA-dimension via participation ratio of the normalized eigenvalues of the covariance matrix of $h_i$ or $\phi(h_i)$).

We found that a PCA-based dimension strongly differs depending on whether it is estimated based on the statistics of the firing rates $\phi(h_i)$ or on $h_i$ (Fig. 5). The finding that $D_{\text{PCA}}^{\phi(h)} > D_{\text{PCA}}^{h}$ is explained by the fact that the sigmoid shape of $\phi$ removes variance from the high-variance directions, thus a larger number of PCs is needed to explain the same fraction of the total variance. This is analogous to previous findings in feed-forward networks [66]. Generally, we find for all dimensionality estimates growth of dimension with $g$ in weakly chaotic networks. However, for large $g \gg 1$, we find a peak and subsequent slight decay of the attractor dimension [Figs. 4 and 5(b)]. In contrast, both PCA dimensions saturate for $g \gg 1$, but they saturate at different levels and with distinct rates. The PCA-based dimensionality [both based on $h_i$ and $\phi(h_i)$] grows extensively with network size $N$, as does the attractor dimensionality [Fig. 5(b)].

PCA-based estimates of dimension are generally not invariant with respect to changes of coordinates and can be misleading if applied to limited data sets. Extensivity of the PCA-based dimension does not generally imply extensivity of the attractor dimension, nor vice versa. In addition, PCA analyses, because they are based on a pairwise correlation function, can miss low-dimensional structure hidden in higher-order correlations. In general, the PCA-based dimension can both under- and overestimate the attractor dimension [as in Fig. 5(a)].

### E. Lyapunov spectrum of discrete-time firing-rate network

The dynamics of discrete-time rate networks has attracted much attention because of its mathematically simplicity [8,67,70–73]. We next assess the effect of introducing finite temporal discretization. Here we aim to understand the impact of time discretization on chaotic dynamics. We thus set $\tau = 1$ and study the evolution of the map

$$h_i(t_s + \Delta t) = f_i = (1 - \Delta t)h_i(t_s) + \Delta t \sum_{j=1}^{N} J_{ij}\phi(h_j(t_s)).$$

In the limit $\Delta t \to 0$, a continuous-time dynamics [1,2,11] is recovered. For $\Delta t = 1$, the discrete-time network [8,8,72,73] is obtained.

The Jacobian for the discrete-time map is

$$D_{ij}(t_s) = \left.\frac{\partial f_i}{\partial h_j}\right|_{t=t_s} = (1 - \Delta t)\delta_{ij} + \Delta t \cdot J_{ij}\phi'(h_j(t_s)). \quad (4)$$

The full Lyapunov spectrum is again obtained by a reorthonormalization procedure of the Jacobians along a numerical solution of the map [60]; for details, see Supplemental Material [61].

Time discretization has a drastic effect on the Lyapunov spectrum (Figs. 6 and 7). In discrete-time networks ($\Delta t = 1$), the Lyapunov spectrum is not point symmetric anymore [Fig. 6(a)]. The largest Lyapunov exponent grows slowly as a function of coupling strength $g$ [Fig. 6(b)], as expected from previous analytical results [8]. However, the slow increase of the largest Lyapunov exponent with coupling strength $g$ is overcompensated by a faster decay of the number of positive Lyapunov exponents, which results in a peak of both dynamical entropy rate $H$ [Fig. 6(c)] and attractor dimensionality $D$ [Fig. 6(d)]. For increasing $g$, the Lyapunov spectrum

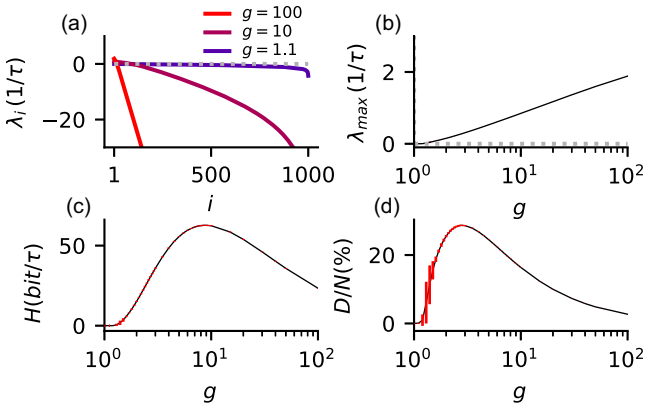


FIG. 6. Lyapunov spectra of discrete-time networks. (a) Full Lyapunov spectra of discrete-time rate networks with different coupling strengths $g$, color coded from blue (small $g$) to red (large $g$). (b) The largest Lyapunov exponent grows as expected monotonically as a function of $g$ [8]. (c) The dynamical entropy rate $H$ peaks with coupling strength $g$. (d) Relative attractor dimensionality $D/N$ also peaks. (Averages over 10 network realizations in black, red error bars indicate double std across 10 network realizations, parameters: $N = 1000$, $\Delta t = \tau$, $t_{\text{sim}} = 10^5\tau$, $t_{\text{ONS}} = \tau$).

bends down to strongly negative values. This is because as $g$ increases, the variance of rate units grows, resulting in a larger fraction of units in saturation. The Jacobian becomes column-sparse and rank-deficient, leading to sparse tangent space dynamics. As described later, this means that an increasing fraction of directions in tangent space can not carry any error signals when training with backpropagation through time (see Appendix E). In the discrete-time case, no intrinsic timescale of decay from a leak term is present, and the

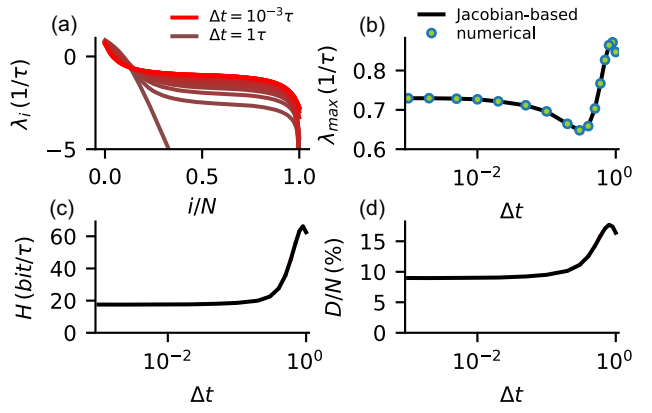


FIG. 7. Full Lyapunov spectrum for different time discretization $\Delta t$. (a) The full Lyapunov spectrum reveals drastic changes for increasing $\Delta t$. For finite $\Delta t$, the Lyapunov spectrum loses its symmetry (see also Fig. 8 and compare with Fig. 3). While the majority of Lyapunov exponents decrease for increasing $\Delta t$, the number of positive exponents increases. (b) The largest Lyapunov exponent converges for small $\Delta t$. For increasing $\Delta t$, it first decreases and then increases moderately. (c) The dynamical entropy rate converges for small $\Delta t$ and increases for large $\Delta t$. (d) The attractor dimensionality behaves similarly to the dynamical entropy rate, (other parameters: $N = 1000$, $g = 10$, $t_{\text{ONS}} = \tau$, $t_{\text{sim}} = 10^4\tau$; averages across 10 network realizations).

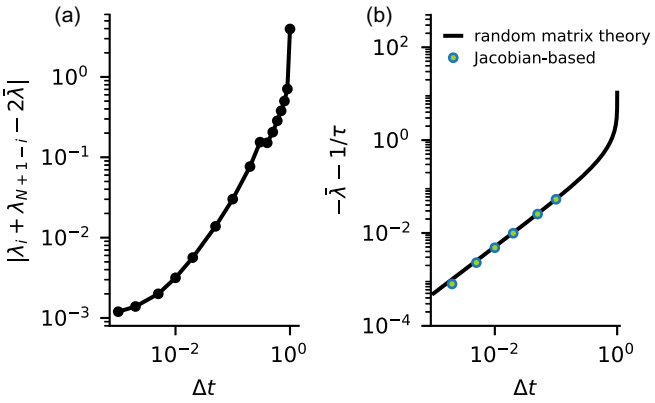


FIG. 8. Point symmetry of Lyapunov spectrum in continuous-time limit $\Delta t \to 0$ and mean Lyapunov exponent $\bar{\lambda}$. (a) For $\Delta t \to 0$, the Lyapunov spectrum approaches point symmetry around $i = N/2$ and $\lambda_i = -\frac{1}{\tau}$, as shown by the convergence of the Lyapunov spectrum towards point symmetry, so $A = \frac{1}{N}\sum_i |\lambda_i + \lambda_{N+1-i} - 2\bar{\lambda}|$ converges towards zero. The Lyapunov spectrum is only point symmetric in the limits $N \to \infty$ and $\Delta t \to 0$. (b) The mean Lyapunov exponent $\bar{\lambda}$ converges for small $\Delta t$ towards $-1/\tau$. For finite $\Delta t$, the mean Lyapunov exponent can be approximated analytically (see Appendix D). (Other parameters: $N = 1000$, $g = 10$, $t_{\text{ONS}} = \tau$, $t_{\text{sim}} = 10^4\tau$; averages across 10 network realizations.)

last Lyapunov exponent becomes progressively more negative [Fig. 6(a)], indicating a rapid divergence of the condition number of the long-term Jacobian $T_t(\mathbf{x}_0)$. From a machine learning perspective, the leak term can be interpreted as a mimicking skip connection that preserves information of the network state across (unrolled) layers even if the rate units are saturated, thus ameliorating the problem of vanishing gradients [74]. Next, we study the effect of gradually decreasing the time-discretization $\Delta t$. At finite $\Delta t$, the Lyapunov spectrum loses its symmetry [Fig. 7(a)]. We demonstrate that the Lyapunov spectrum approaches point symmetry around $\lambda_i = -\frac{1}{\tau}$ and $i = N/2$ for $\Delta t \to 0$ by showing convergence of the Lyapunov spectrum towards its point reflection. We quantify deviations from point symmetry by $\frac{1}{N}\sum_i |\lambda_i + \lambda_{N+1-i} - 2\bar{\lambda}|$; for perfectly point symmetry Lyapunov spectra it would be zero [Fig. 8(a)]. Even for very small $\Delta t$, however, there exists a small asymmetry because of the neutral Lyapunov exponent. Removing the neutral Lyapunov exponent, which is associated with a perturbation in the direction of the flow ($\lambda_{\text{neutral}} = 0$), improves the point symmetry of the Lyapunov spectrum. Note that the symmetry of the Lyapunov spectrum originates in the approximate time-reversal symmetry of the dynamics, which only becomes exact in the limit of large $N$ (see Sec. VI). Thus, the Lyapunov spectrum is only point symmetric in the limits $N \to \infty$ and $\Delta t \to 0$.

While the largest Lyapunov exponent changes only moderately—and nonmonotonously—as the step size increases [Fig. 7(b)], the dynamical entropy rate and attractor dimensionality both strongly grow for large $\Delta t$ [Figs. 7(c) and 7(d)]. This growth of entropy rate and dimensionality is primarily caused by an increasing number of positive Lyapunov exponents [Fig. 7(a)]. At the same time, the negative end of the Lyapunov spectrum decreases drastically [Fig. 7(a)]. This also strongly lowers the mean Lyapunov exponent [Figs. 7(a) and

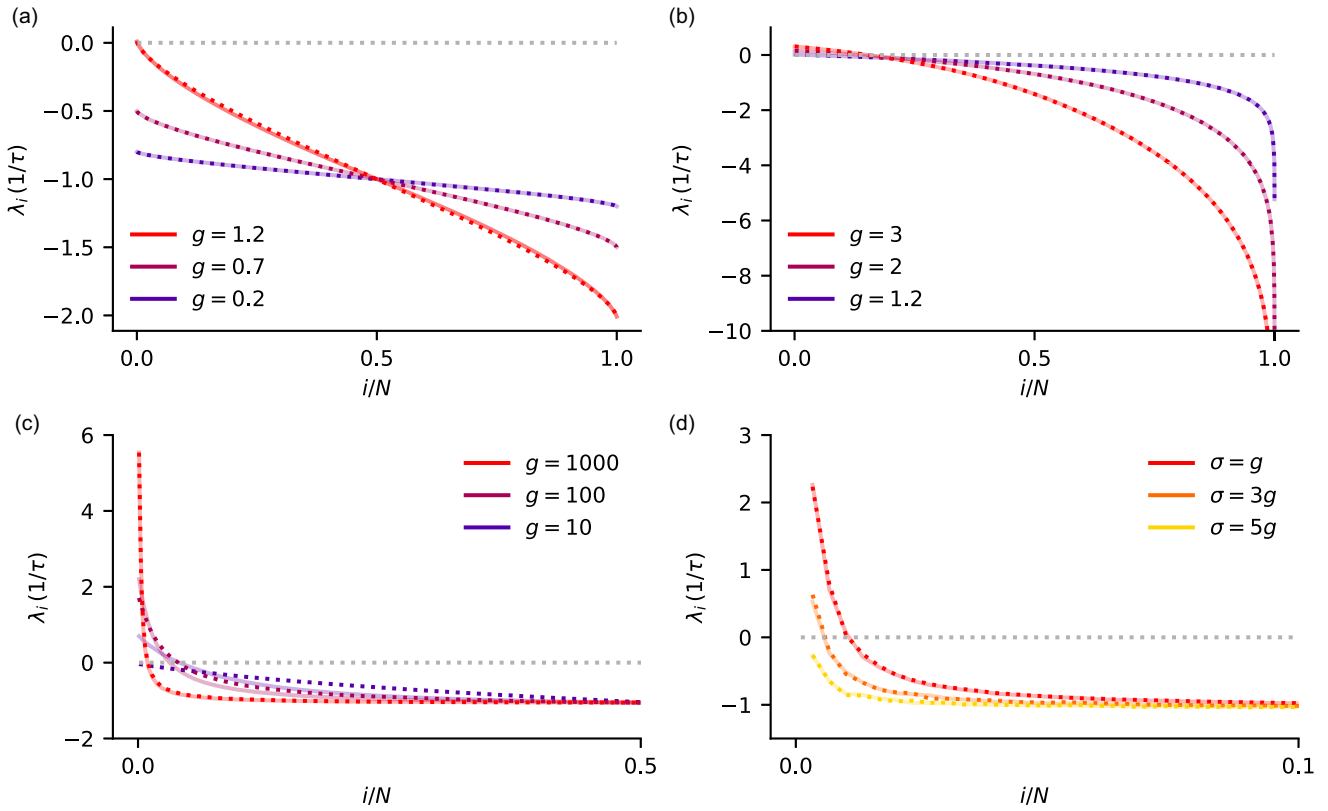


FIG. 9. Analytical approximations of the full Lyapunov spectrum. (a) Lyapunov spectra of autonomous continuous-time rate networks for different coupling strengths $g$, where $g$ is color-coded from purple (small $g$) to red (large $g$), dashed lines are analytical results [Eq. (13)] for the stable case and for $g \to g^+_{\text{crit}}$ [Eq. (17)], full transparent lines are numerical results. (b) In the discrete-time case without leak ($\Delta t = 1$), the Lyapunov spectrum is invariant under shuffling for any $g$ and can be approximated analytically by the triangle law for small $g$ [Eq. (19)]. (c) Lyapunov spectra of autonomous continuous-time rate networks for $g \gg 1$. For large $g$, the Lyapunov spectrum becomes invariant under shuffling the temporal sequence of Jacobians at fixed $\Delta t = 0.1$. (d) Lyapunov spectra of driven continuous-time rate networks for different input strength $\sigma$ at fixed $g = 1000$. Again, for large $g$ and $\sigma$, the Lyapunov spectrum becomes invariant under shuffling the temporal sequence of Jacobians at fixed $\Delta t = 0.1$, and the full Lyapunov spectrum can be approximated by a product of random matrices with entries given by Eq. (10). (Parameters if not stated differently: for $g = 1.2$ in (a) $N = 8000$ and $t_{\text{sim}} = 10^3\tau$, else $N = 1000$, $t_{\text{sim}} = 10^4\tau$, $\Delta t = 0.1\tau$, $t_{\text{ONS}} = \tau$).

8(b)]. The mean Lyapunov exponent $\bar{\lambda}$ converges for small $\Delta t$ towards $-1/\tau$. The dependence of the mean Lyapunov exponent on $\Delta t$ can be approximated analytically using random matrix theory by (Appendix D)

$$\bar{\lambda}(\Delta t) = \frac{\log(1-\Delta t)}{\tau\,\Delta t}. \tag{5}$$

This analytical result agrees well with numerical simulations [Fig. 8(b)].

## IV. ANALYTICAL APPROXIMATIONS FOR THE FULL LYAPUNOV SPECTRUM

The full Lyapunov spectrum is given by the logarithm of the eigenvalues of the Oseledets matrix $\mathbf{\Lambda}$ [75],

$$\mathbf{\Lambda} = \lim_{t\to\infty} \left[\mathbf{T}_t^\top \mathbf{T}_t\right]^{\frac{1}{2t}}. \tag{6}$$

where $\mathbf{T}_t$ is the long-term Jacobian

$$\mathbf{T}_t(\mathbf{h}_0) = \mathbf{D}_{t-1}(\mathbf{h}_{t-1})\ldots\mathbf{D}_1(\mathbf{h}_1)\mathbf{D}_0(\mathbf{h}_0) \tag{7}$$

As $\mathbf{T_t(h_0)}$ is a product of generally noncommuting matrices, it is considered difficult to calculate the full Lyapunov spectrum analytically [76]. However, we identified several limits where temporal correlations between subsequent Jacobians vanish and analytical random matrix approximations are justified [76]. First, we demonstrate an approximation for the stable regime $g < g_{\text{crit}}$, second in the chaotic regime just above the transition $g \to g^+_{\text{crit}}$, third in the limit of large $g \to \infty$, fourth when each rate unit is driven by strong Gaussian white noise process with standard deviation $\sigma$ in the limit $\sigma \to \infty$ (see Sec. VIII for definition), and finally in the discrete-time case without a leak $\tau = \Delta t = 1$. In the discrete-time case, we calculate the full Lyapunov spectrum analytically for hard tanh networks for arbitrary $g$ and give expression for $H$ and $D$. We first numerically confirmed that in the limits mentioned above, the Lyapunov spectrum becomes invariant under shuffling the sequence of Jacobians (Fig. 9) and then found explicit or implicit expressions for the full Lyapunov spectrum in several cases. In the limits we are discussing in the following, the long-term Jacobian can be approximated by a product of random matrices of the form of Eq. (10),

$$\mathbf{T}_t = \prod_{s=0}^{t-1} \mathbf{D}_s = \prod_{s=0}^{t-1} [(1-\Delta t)\mathbb{1} + \Delta t \cdot \mathbf{J}\cdot\mathbf{y}_s]$$

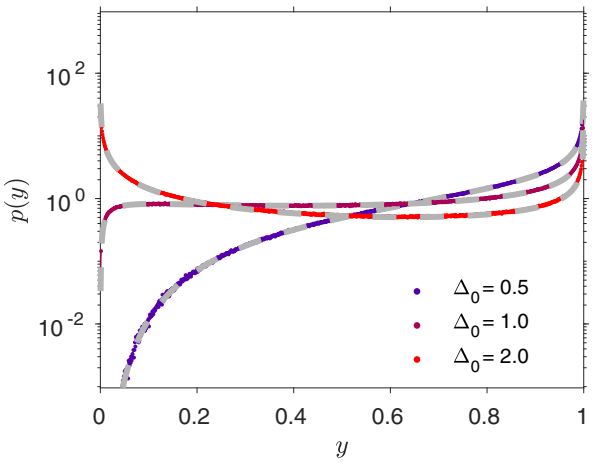


FIG. 10. Distribution of Jacobian factors $\phi(h)'$. Colored dots are from direct numerical simulations and grey dashed lines analytical distributions [Eq. (9)] for different values of variance $\Delta_0$ of the local fields $h$. For small $\Delta_0$, most probability mass is close to 1. For large $\Delta_0$, the distribution becomes bimodal because most units are in saturation of the nonlinearity with corresponding $y$ close to zero and few $y$ near 1. The values of $\Delta_0 \in \{0.5, 1.0, 2.0\}$ correspond to $g \in \{1.36457, 1.62321, 2.0271\}$ and $\lambda_1 \in \{0.0299, 0.0627, 0.116\}$.

We calculated the distribution of entries of the Jacobian analytically in the limit $N \to \infty$, where all $h_i$ follow a Gaussian distribution $h \sim \mathcal{N}(0, \Delta_0)$. First, we calculated the distribution of $y = \phi'$ analytically:

$$p(y) = \int dh\, \delta(y - \phi'(h)) \frac{e^{-\frac{h^2}{2\Delta_0}}}{\sqrt{2\pi\Delta_0}} \tag{8}$$

$$= \frac{\exp\left(-\frac{\ln(1/\sqrt{y} \pm \sqrt{1/y-1})}{2\Delta_0}\right)}{\sqrt{2\pi\Delta_0^2}|2y\sqrt{1-y}|} \tag{9}$$

with support $y \in [0, 1]$, where $\Delta_0$ is obtained analytically from dynamic mean-field theory [1]. We can thus write the Jacobian as

$$D_{ij}(t_s) = (1 - \Delta t)\delta_{ij} + \Delta t \cdot J_{ij} y_j, \tag{10}$$

where $y_j$ are random numbers drawn from the distribution in Eq. (9). The analytically predicted distributions $p(y)$ are in excellent agreement with the results from direct numerical simulations (Fig. 10). We next describe these different analytically tractable cases in more detail starting with $g < 1$. In the stable regime $g \leqslant g_{\text{crit}}$, the Lyapunov spectrum is given by the real parts of the eigenvalue spectrum of the stability matrix [Eq. (2)]. Because the trivial fixed point $h_i = 0$ for all $i$ is the only stable solution for large $N$, Eq. (2) reduces to $D_{ij} = J_{ij} - \delta_{ij}$. For $J_{ij}$ drawn from a Gaussian distribution $J_{ij} \sim \mathcal{N}(0, g^2/N)$, we find that the real parts of the eigenvalues of $D_{ij}$ follow the right-shifted Wigner semicircle distribution [76–78]:

$$p(x) = \frac{2}{2\pi g^2}\sqrt{g^2 - (x+1)^2} \tag{11}$$

with support $x \in [-g - 1,\ g - 1]$. Note that this is not only expected for $J_{ij}$ drawn from a Gaussian distribution, but generally for many random matrix ensembles [79]. The cumulative distribution function is given by

$$\chi(x) = \frac{1}{2} + \frac{\arcsin\left(\frac{x+1}{g}\right)}{\pi} + \frac{(x+1)\sqrt{g^2 - (x+1)^2}}{\pi g^2}. \tag{12}$$

The Lyapunov spectrum in the stable regime follows from the inverse,

$$\lambda_i = \chi^{-1}\left(\frac{N - i + 1}{N}\right). \tag{13}$$

with $\frac{i}{N} \in [0,\ 1]$ and $\lambda_i \in [-g - 1,\ g - 1]$, and with $\lambda_i$ measured in units of $1/\tau$. The analytical Lyapunov spectra are in excellent agreement with the results from direct numerical simulations [Fig. 9(a), purple ($g = 0.2$) and maroon ($g = 0.7$) lines].

Next, we consider the limit $g \to 1^+$ close to the transition to chaos. For that, we need to estimate both the distribution of Jacobian entries and their autocorrelations. The autocorrelations of the activity

$$\Delta_i(t') = \langle \delta h_i(t) \delta h_i(t + t') \rangle \tag{14}$$

can be solved self-consistently [1,59]. Close to the chaotic instability $g \to g^+_{\text{crit}}$, the autocorrelations are approximately [1,59]

$$\Delta(t') = (g-1)\,\text{sech}\left(\frac{t'(g-1)}{\sqrt{3}}\right) + \mathcal{O}((g-1)^2). \tag{15}$$

Thus, the timescale of the autocorrelations of $h_i$ diverges when approaching $g_{\text{crit}}$ with $\tau_h = \sqrt{3}/(g-1)$ [1]. From $\lim_{g\to 1^+} \Delta_0 = g - 1$ follows that $\lim_{g\to 1^+} \phi(h_i(t)) = h_i(t)$. Therefore the autocorrelations of $D_{ij}$ diverge with the same time constant $\tau_D^{-1} = (g-1)/\sqrt{3}$ and subsequent Jacobians are almost identical. Consistent with these analytical considerations, numerical simulations show that for $g \gtrapprox 1$ the Lyapunov spectrum obtained after shuffling the sequence of (almost identical) Jacobians is almost the same. Thus, we conjecture that the Lyapunov spectrum is given by the logarithms of the singular values of a product of almost identical random matrices, which is still approximately given by the Wigner semicircle distribution [76,78]

$$\chi_{\text{chaos}}(x) = \frac{1}{2} + \frac{\arcsin\left(\frac{2(x+1)}{(x-2)x+3}\right)}{\pi} + \frac{2(x+1)\sqrt{((g-1)^2 - 2x)(2x + (g-2)g + 5)}}{\pi((x-2)x+3)^2}$$

with support $x \in [-2\tau - \lambda_{\max},\ \lambda_{\max}]$. Here, we used the analytical knowledge of the largest Lyapunov exponent obtained from dynamic mean-field theory [1,59], which behaves in the limit $g \to g^+_{\text{crit}} = 1^+$ as

$$\lambda_{\max}(g) = \tfrac{1}{2}(g-1)^2 + \mathcal{O}((g-1)^3). \tag{16}$$

The Lyapunov spectrum in the chaotic regime for $g \to g^+_{\text{crit}}$ follows the inverse of $\chi_{\text{chaos}}$:

$$\lambda_i = \chi^{-1}_{\text{chaos}}\left(\frac{N - i + 1}{N}\right). \tag{17}$$

The analytical Lyapunov spectra are in good agreement with the results from direct numerical simulations [Fig. 9(a), red

line for $g = 1.2$]. The approximation breaks down if $g$ is too large and becomes more accurate as $g \to g^+_{\text{crit}}$.

Next, we consider the limit of large $g \to \infty$. Expanding the solution to the self-consistency equation for the autocorrelation of the local fields $h$ in this limit around $t' = 0$ yields in that case

$$\Delta(t') = g^2\Delta_0 - g^2(1-\Delta_0)\frac{t'^2}{2} + \mathcal{O}(t'^4) \tag{18}$$

with $\Delta_0 = 2(1-2/\pi)$ [1,59]. But how can we deal with correlations between subsequent Jacobians? We note that the autocorrelations of the Jacobians become arbitrary short in the limit of large $g$, although the autocorrelations of the activity variables $h$ approach Eq. (18). For large $g$, the model behaves like the fully asymmetric Ising spin glass model [59,80]. Substituting $\Delta(t') = \Delta_0 \exp(-t'/\tau_h)$ into the self-consistency equation and taking the large $t'$ limit yield a relaxation rate for the autocorrelation equal to $\tau_h^{-1} = \sqrt{1-2/\pi}\ \tau^{-1}$ [1,59]. Thus, the autocorrelation of $D_{ij}$ relaxes approximately with $\tau_D \sim \tau_h/g$. Intuitively, for large $g$, the variance $\Delta_0$ grows quadratically with $g$ ($\Delta_0 = 2(1-2/\pi)g^2$) [1,59] [see also Fig. 21(c)]. Therefore, most rate units are in saturation, and rate units cross the nonsaturated regime where they are susceptible to perturbations in shorter time windows. In this limit, most Lyapunov exponents are close to the negative inverse of the characteristic timescale $-1/\tau$. This is in contrast to the case of discrete-time networks where for large $g$ the last Lyapunov exponent becomes progressively more negative [see Fig. 6(a)].

From $\tau_D \sim \tau_h/g$, it follows that for $g \to \infty$ the Jacobian become time uncorrelated justifying the random matrix approximation. As expected, the analytical approximations of the Lyapunov spectra approach the results from direct numerical simulations when the values of $g$ increase [Fig. 9(c)].

We also find that for strong uncorrelated input (see Sec. VIII for numerical results), the Lyapunov spectrum becomes invariant under shuffling the sequence of Jacobians. With increasing input drive $\sigma$ at fixed $g$, all Lyapunov exponents converge towards the negative inverse of the characteristic timescale $-1/\tau$ (not shown). When simultaneously increasing $g$ and $\sigma$, the Lyapunov spectrum also becomes invariant under shuffling the Jacobians [Fig. 9(d)]. In this limit, the first Lyapunov exponent approaches a finite value.

Finally, in the discrete-time case $\Delta t = 1$ without a leak, temporal correlations between subsequent Jacobians can be neglected for large $N$ [8]. For $g \to 1^+$, the Lyapunov spectrum can thus be obtained from a product of uncorrelated Gaussian matrices, whose eigenvalue distribution follows approximately a triangle law [81,82]. As already pointed out earlier [83,84], this triangular law also applies to discrete-time recurrent neural networks and the full Lyapunov spectrum [Fig. 9(b)] in this limit can thus be approximated by

$$\lambda_i \approx \log\left(\exp(\lambda_{\max})\cdot\sqrt{1-\frac{i}{N}}\right) = \lambda_{\max} + \frac{1}{2}\log\left(1-\frac{i}{N}\right), \tag{19}$$

where the largest Lyapunov exponent $\lambda_{\max}$ can be obtained analytically as described earlier both in the discrete and continuous-time case with constant input and additive Gaussian white noise drive [1,2,8,11,59]. However, Eq. (19) only gives a good approximation of the full Lyapunov spectrum for $g \to 1^+$. For the hard tanh input-output transfer function $\phi(x) = \min(1, \max(-1, x))$, we can calculate the full Lyapunov spectrum analytically for arbitrary $g$ (Fig. 11):

$$\lambda_i = \begin{cases} \lambda_{\max} + \frac{1}{2}\log\left(1-\frac{i}{pN}\right), & i \leqslant pN \\ -\infty, & i > pN \end{cases} \tag{20}$$

where $p = \operatorname{erf}(\frac{1}{\sqrt{2}g\Delta_0})$ is the average fraction of unsaturated neurons, obtained from dynamic mean-field theory [8]. The analytical predictions for the Lyapunov spectra agree well with the numerical calculations [Fig. 11(a)].

This results in

$$\frac{H}{N} = p\left(\lambda_1 + \frac{e^{-2\lambda_1}-1}{2}\right) \tag{21}$$

and

$$\frac{D}{N} = p\left(1 - \frac{2\lambda_1 - 1}{W[(2\lambda_1-1)e^{2\lambda_1-1}]}\right) \tag{22}$$

where $W$ is the Lambert $W$ function. Here the origin of the peak in $H$ and $D$ as a function of $g$ and $\sigma$ despite growing $\lambda_1$ becomes obvious: The largest Lyapunov exponent keeps on growing consistently with previous work [8], but $H$ and $D$ start to decrease for large $g$ because of a shrinking fraction of unsaturated units $p$, which makes the Jacobian sparser proportionally to $p$. Thus, for growing $g$, the slow increase of the largest Lyapunov exponent with coupling strength $g$ is overcompensated by a faster decay of the number of positive Lyapunov exponents, which results in a peak of both dynamical entropy rate $H$ and attractor dimensionality $D$. Increasing

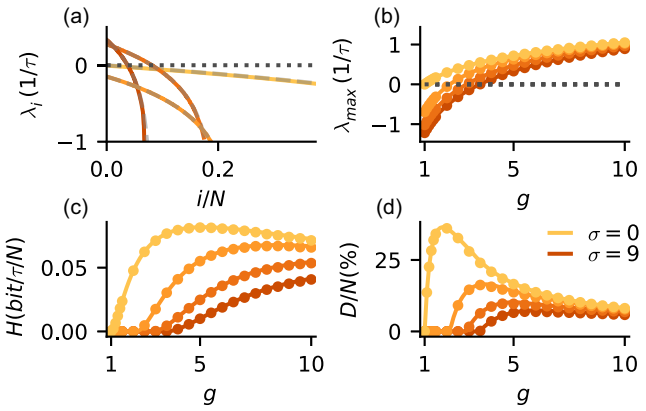


FIG. 11. Analytical Lyapunov spectrum for hard tanh networks. (a) Lyapunov spectra of hard tanh network for coupling strengths $g \in \{1, 1.8, 3, 5\}$ and $\sigma^2 \in \{0, 3, 3, 10\}$ from yellow to red, dashed lines are analytical results [Eq. (20)], full transparent lines are numerical results. (b) The largest Lyapunov exponent grows monotonically as a function of $g$ and decreases with input strength $\sigma^2$, consistently with previous work [8] (dots: Jacobian-based method, lines: analytical solutions). (c) The dynamical entropy rate $H$ peaks at higher $g$ and is reduced by external input. (d) The relative attractor dimensionality $D/N$ decreases for increasing $\sigma$ and peaks at higher $g$ (curves are averages over 10 network realizations, red error bars indicate double std across 10 network realizations, parameters: $N = 4000$, $t_{\text{sim}} = 10^3\tau$, $t_{\text{ONS}} = \tau$).

$\sigma$ shifts the peak of both $H$ and $D$ to larger $g$ [Figs. 11(c) and 11(d)].

For $g \to g_{\text{crit}}^{+}$, this becomes

$$\frac{H}{N} = p\lambda_1^2 = \frac{1}{4}p(g - g_{\text{crit}})^2 \tag{23}$$

and

$$\frac{D}{N} = 2p\lambda_1 = 2p(g - g_{\text{crit}}) \tag{24}$$

Thus, close to the transition, $H$ grows quadratic with $g - g_{\text{crit}}$ while $D$ grows linear, consistent with earlier numerical observations [84]. Equations (20)–(24) also hold for the discrete-time threshold-linear transfer function and in the present of time-dependent external input, but $\lambda_1$, $p$ and $g_{\text{crit}}$ then take different values.

## V. LYAPUNOV SPECTRUM OF BALANCED RATE NETWORK WITH THRESHOLD-LINEAR TRANSFER FUNCTIONS

While odd symmetric saturated sigmoid transfer functions, e.g., $\phi(x) = \tanh(x)$ are popular because of their mathematical tractability [1,59] and because the saturation prevents runaway activity, the firing rate of many cortical neuron types seems in a physiological operating regime not to be limited by intrinsic electrophysiological features. Evidence for this comes from the observation that artificially driven neurons can fire at much higher rates [85] than they actually do in experimental recordings of awake behaving animals [86]. In balanced networks, large externally incoming excitatory currents are dynamically canceled by net inhibitory recurrent currents, which yields a broad parameter regime of asynchronous irregular activity in spiking network models [87–91]. Such balanced state models were recently extended from spiking networks to firing rate networks [2,4,7,92].

Here we extend our Lyapunov spectrum analysis to balanced networks with a threshold-linear transfer function and investigate the role of the synaptic coupling strength $g$ on the Lyapunov spectrum (Fig. 12). Threshold-linear transfer functions are also commonly used in deep learning [93,94]. Another reason to investigate this transfer function is that experimentally measured neural nonlinearities in sensory neurons have been approximated by a power-law threshold nonlinearity [95–99].

Focus on the dynamics of an inhibitory network of $N$ threshold-linear rate units that balance a constant excitatory external input. The dynamics of each firing-rate unit follows

$$\tau \frac{dh_i}{dt} = -h_i + \sum_j J_{ij}\phi(h_j) + I \tag{25}$$

where $I$ is a positive constant and $\phi(x) = \max(x, 0)$. We draw entries of the coupling matrix $J_{ij}$ from a Gaussian distribution $J_{ij} \sim \mathcal{N}(-\bar{g}/N, g^2/N)$ (like [2]).

As in the previously considered tanh networks, the first half of the Lyapunov spectrum is increasingly curved for increasing $g$ [Fig. 12(a)]. For large $g$, the network dynamics turns unstable, and the activities $h_i$ diverge [2], therefore there is no chaotic large $g$-limit for fixed mean coupling strength $J$. The divergence occurs at much larger $g$ than displayed

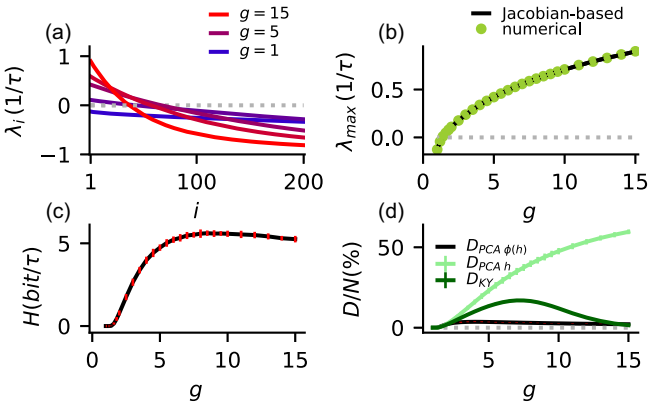


FIG. 12. Entropy rate and dimensionality of a balanced firing-rate network with threshold-linear transfer function. (a) Top 200 Lyapunov exponents of rate networks for different coupling strengths $g$, where $g$ is color-coded from blue (small $g$) to red (large $g$). (b) The largest Lyapunov exponent grows monotonically as a function of $g$ below the divergence. (Green dots: direct numerical simulations, black line: Jacobian-based method). (c) The dynamical entropy rate $H$ peaks with coupling $g$. (d) Relative attractor dimensionality $D/N$ has a peak as a function of $g$, error bars are smaller than line width. (black curves are averages over 20 network realizations, red error bars indicate double std across 20 network realizations, parameters: $N = 4000$, $\bar{g} = 300$, $I = 300$, $t_{\text{sim}} = 10^3\tau$, $t_{\text{ONS}} = \tau$, for (a) $g \in \{1, 2, 5, 7.5, 15\}$).

in Fig. 12. The largest Lyapunov exponent shows the analytically predicted behavior [2]. We confirmed the results obtained from the Jacobian-based method [60] by tracking the amplitude of a small perturbation in direct numerical simulations [Fig. 12(b)]. Similar to tanh networks, the dynamical entropy rate peaks as a function of $g$ because the growth of a small fraction of positive Lyapunov exponents is overcompensated by a decreasing number of positive Lyapunov exponents for large $g$ [Fig. 12(c)]. We find that the peaks in H [Fig. 12(c)] and D [Fig. 12(d)] occur at smaller values of $g$ for balanced threshold-linear networks compared to tanh-networks [Figs. 4, 5(a), and 5(b)]. Note that there is a striking difference between the attractor dimension (black) and both of the PCA-based dimensions in case of threshold-linear transfer functions. Moreover, there is also a vast difference between the PCA-based dimension depending on whether it is estimated based on the statistics of the firing rates $\phi(h_i)$ or on $h_i$ [Fig. 12(d) light green vs dark green].

## VI. POINT-SYMMETRY OF LYAPUNOV SPECTRA FOR CONTINUOUS-TIME DYNAMICS

A symmetry of Lyapunov spectra around zero is usually found in dynamical systems with a symplectic structure [100–104]. This is given, for example, in Hamiltonian systems, where the Lyapunov spectrum is symmetric around zero. Symmetry around a negative value was previously described in a class of dissipative dynamical systems with viscous damping [100]. The recurrent neuronal networks we considered have an asymmetric connectivity $J_{ij} \neq J_{ji}$. Thus, there is no conservation of energy and no time-reversal symmetry. Also, a pseudo-Hamiltonian structure is not given. Moreover, our findings indicate that symmetric Lyapunov

spectra are not a generic feature of recurrent neural networks. But what is the origin of the symmetry in our case of recurrent neural firing-rate networks? The symmetry of the Lyapunov spectrum of recurrent networks in the continuous-time limit originates in the approximate time-reversal symmetry of the dynamics. This can be directly seen by a change of variables into a reference frame that contracts with time. Introducing the new variables $z = e^{\frac{t}{\tau}}h$ turns the original equation of motion [Eq. (1)] to

$$\tau \frac{dz_i}{dt} = e^{\frac{t}{\tau}} \sum_{j=1}^{N} J_{ij}\phi\left(e^{-\frac{t}{\tau}}z_j\right). \tag{26}$$

Making the replacement $z(t) = \tilde{z}(-t)$, gives

$$\tau \frac{d\tilde{z}_i}{dt} = e^{\frac{t}{\tau}} \sum_{j=1}^{N} (-J_{ij})\phi\left(e^{-\frac{t}{\tau}}\tilde{z}_j\right). \tag{27}$$

Thus, with reversed time, one obtains the same dynamics with coupling matrix $\tilde{J}_{ij} = -J_{ij}$. $\tilde{J}_{ij}$ follows the same distribution as $-J_{ij}$. As for large $N$, the Lyapunov spectrum does not depend on the realization of the network connectivity, which is drawn from a Gaussian distribution $J_{ij} \sim \mathcal{N}(0, g^2/N)$, the Lyapunov spectrum is invariant under flipping the sign of $J_{ij}$. Thus, the dynamics is statistically invariant under time-reversal, where "statistically invariant" means under the statistics of the connectivity. The Oseledets matrix in the contracting reference frame is

$$\tilde{\mathbf{\Lambda}}(\mathbf{z}_0) = \lim_{t\to\infty} \left[e^{-\frac{t}{\tau}}\hat{\mathbf{T}}_t(\mathbf{z}_0)^\top \hat{\mathbf{T}}_t(\mathbf{z}_0)e^{-\frac{t}{\tau}}\right]^{\frac{1}{2t}} \tag{28}$$

$$= e^{-\frac{1}{\tau}} \lim_{t\to\infty} [\hat{\mathbf{T}}_t(\mathbf{z}_0)^\top \hat{\mathbf{T}}_t(\mathbf{z}_0)]^{\frac{1}{2t}} \tag{29}$$

$$= e^{-\frac{1}{\tau}} \hat{\mathbf{\Lambda}}(\mathbf{z}_0). \tag{30}$$

Growing tangent space volume elements in forward time correspond to shrinking tangent space volume elements in backward time. Because of the time-reversal symmetry, they are approximately inverse, i.e., the eigenvalues $\mu_i$ of the Oseledets matrix satisfy

$$e^{-\frac{1}{\tau}}\mu_i^+ \approx e^{-\frac{1}{\tau}} \frac{1}{\mu_{N-i+1}^-} \tag{31}$$

where $+(-)$ indicate forward (backward) time direction. Thus, the Lyapunov exponents, given by the logarithm of the eigenvalues of the Oseledets matrix satisfy

$$\lambda_i^+ - \frac{1}{\tau} \approx -\frac{1}{\tau} - \lambda_{N-i+1}^-, \tag{32}$$

where the factor $-\frac{1}{\tau}$ comes from the shrinking reference frame. Note that in contrast to Hamiltonian systems where the symplectic structure of the Hamiltonian implies an exact symmetry of the Lyapunov spectrum around zero, here, the symmetry is only approximate for finite-size networks and around the negative inverse of the characteristic time scale due to the global damping on dynamics through the leak term. Also note that in autonomous systems that are not at a fixed point, there is always a zero Lyapunov exponent $\lambda_{\text{neutral}} = 0$ corresponding to neutral shifts in the direction of time that does not have a symmetric analog at $\lambda = -\frac{2}{\tau}$. Finally, we note that our symmetry argument assumes that there exist statistically analogous backward trajectories, which is generally not correct but justified in our case because of the statistical mirror-symmetry of the connectivity $p(J_{ij})$ around zero. For example, this does not generally hold after training, where the negative connectivity $\tilde{J}_{ij} = -J_{ij}$ can yield statistically very different dynamics.

## VII. EXTENDED FIRST COVARIANT LYAPUNOV VECTOR

To quantify how many rate units contribute to the chaotic dynamics at each moment in time, we investigated properties of the covariant Lyapunov vectors $\vec{v}^{(k)}(t)$. The first covariant Lyapunov vector gives at any point in time the direction in which almost all initial infinitesimal perturbations grow with average rate $\lambda_{\max}$. It corresponds to the first Gram-Schmidt vector and is denoted here as $\vec{v}$ with $\sum_{i=1}^{N} v_i(t)^2 = 1$. The number of rate units contributing to the maximally growing direction at time $t$ can be measured by the participation ratio $P(t) = (\sum_{i=1}^{N} v_i(t)^4)^{-1}$ [105–107]. If all rate units contribute equally to the Lyapunov vector $|v_i(t)| = 1/\sqrt{N}$, the participation ratio is $P(t) = 1/(N/N^2) = N$. If only one rate unit contributes to the Lyapunov vector, the participation ratio is $P(t) = 1$. The Lyapunov vector of firing-rate networks indicates that a temporally varying subset of rate units governs the most unstable direction [Figs. 13(a) and 13(c)]. There is only a moderate temporal fluctuation of the participation ratio [Fig. 13(c)], which declines proportionally to $1/\sqrt{N}$ with network size (not shown). Moreover, the local Lyapunov exponent $\lambda_{\max}^{\text{local}}(t)$, which measures the instantaneous growth rate of the first covariant Lyapunov vector fluctuates in time, indicating different local growth rates along the trajectory [Fig. 13(e)]. Also these temporal fluctuations decrease $\propto 1/\sqrt{N}$ (not shown). The participation ratio $\bar{P} = \langle P(t)\rangle$ is independent of $g$ both for networks with tanh [Fig. 13(b)] and threshold-linear input-output transfer function (not shown).

To further characterize the nature of the chaotic collective network state, we investigated the scaling of the mean participation ratio $\bar{P}$ with network size. Whether the Lyapunov vector is called localized or delocalized depends on how $\bar{P}$ scales as a function of network size $N$. A delocalized state is indicated by a linear scaling $\bar{P} \sim N$, while in the case of a localized state, the participation ratio is independent of $N$. We found a linear scaling $\bar{P} = N/3$ for both tanh [Fig. 13(d)] and threshold-linear transfer function. This results from the fact that the entries of the first covariant Lyapunov vector are independent Gaussian with $v_i \sim \mathcal{N}(0, 1/N)$ [see probability density in Fig. 13(f)], which yields

$$\bar{P} = N\left(\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} x^4 e^{-\frac{1}{2}x^2} dx\right)^{-1} = \frac{N}{3}. \tag{33}$$

This is in contrast to chaos in sparse spiking neural networks, where a sublinear scaling of the participation ratio of the first covariant Lyapunov vector with network size has been reported for sparse networks of quadratic integrate-and-fire neurons in the balanced state [91]. We conclude that the direction of greatest instability in random rate networks is supported by a macroscopic number of rate units, which indicates the existence of collective Lyapunov modes that characterize

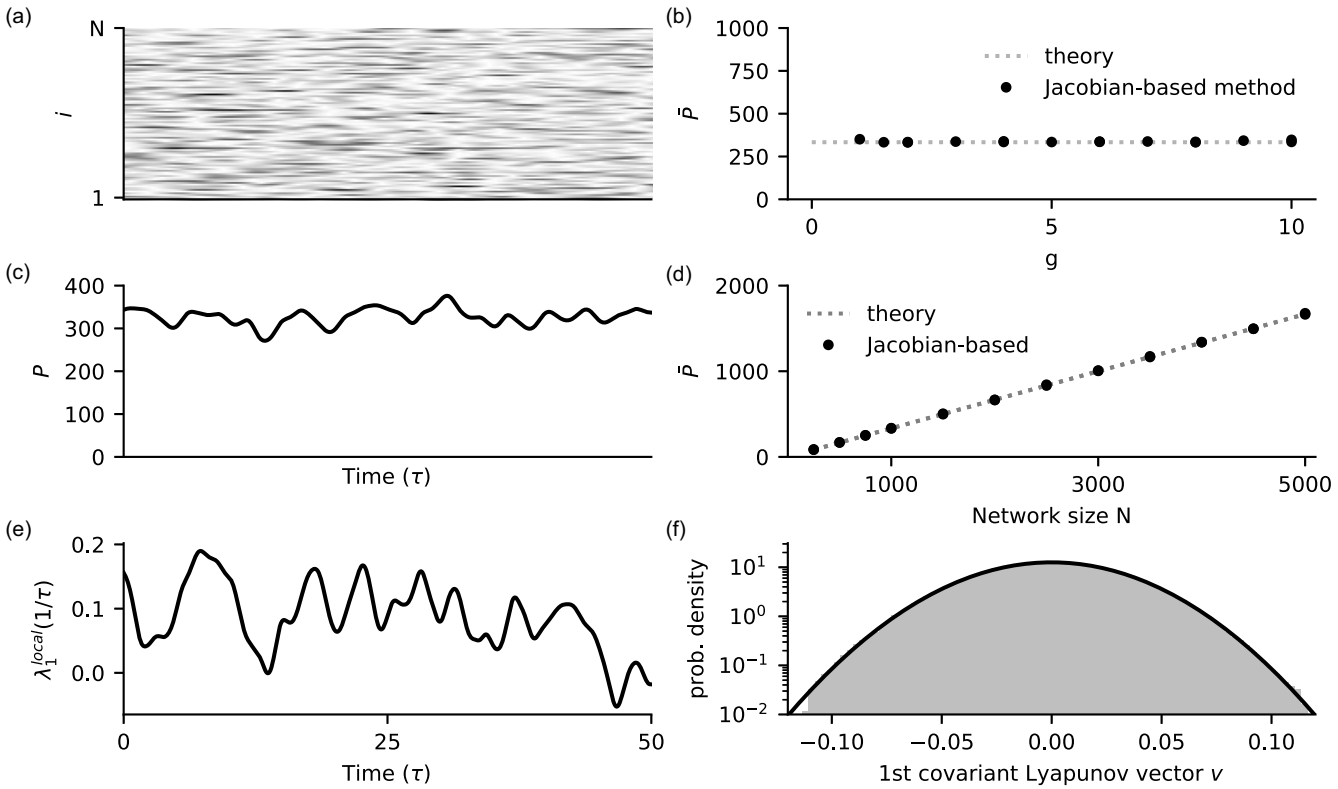


FIG. 13. Spatiotemporal analysis of network chaos and delocalized first covariant Lyapunov vector. (a) First covariant Lyapunov vector $\vec{v}(t)$ (gray scale of $|v_i(t)|$ of a subset of 100 random directions), (b) Average participation ratio $\bar{P}$ vs $g$ stays constant, consistent with prediction of $\bar{P} = N/3$ [Eq. (33)]. (c) Participation ratio $P(t)$ of first covariant Lyapunov vector (CLV). (d) Average participation ratio $\bar{P}$ vs network size $N$ indicates delocalized CLV $\bar{P} = N/3$ (dashed). (e) Local Lyapunov exponent $\lambda_{\max}^{\text{local}}(t)$. (f) Distribution of entries of first covariant Lyapunov vector $v$ [gray simulation, black theory $\mathcal{N}(0, 1/N)$]. (Parameters: $N = 1000$, $g = 2$, $\Delta t = 10^{-2}\tau$, $t_{\text{ONS}} = \tau$, $t_{\text{sim}} = 10^3\tau$, averages across 10 network realizations.)

the instability of the collective dynamic. This is a promising direction for future research that might link the microscopic phase space structure to macroscopic modes of activity [108].

## VIII. LYAPUNOV SPECTRUM OF EXTERNALLY DRIVEN NETWORK

Thus far, we have analyzed the autonomous dynamics of a deterministic firing-rate network, but it is interesting to extend this to a nonautonomous system driven by time-varying input [8,10,11,73,109]. We consider an input-driven network,

$$\tau \frac{dh_i}{dt} = F_i = -h_i + \sum_{j=1}^{N} J_{ij}\phi(h_j) + \xi_i(t), \tag{34}$$

where in the case considered here $\xi_i$ are independent Gaussian white noise processes with autocorrelation function $\langle \xi_i(t)\xi_i(t+t')\rangle = \tau\sigma^2\delta(t')$ that are fixed across initial conditions of $h$.

To assess the dynamic stability of the stochastic differential equation, we employ the theory of random dynamical systems (RDS) [110]. This theory characterizes how reliably different initial states respond to an external input realization fixed across perturbed and unperturbed trajectories. We call a system reliable if different initial conditions converge to the same (time-dependent) trajectory, and unreliable otherwise [111]. More formally, the evolution of a sample measure $\mu_\xi^t$ is studied for a time-dependent noise realization $\xi(t)$ with $t \in (-\infty, \infty)$ that is fixed across initial conditions of the system. This is described in more detail in the Supplemental Material [61].

The mathematical expression for the Jacobian of the flow of the dynamics is the same as in the autonomous case Eq. (2). However, despite this similarity, an external input can have a strong effect both on the distribution of $h_i$ and on the autocorrelations $\Delta_i(\tau) = \langle \delta h_i(t)\delta h_i(t+\tau)\rangle$. First, input fluctuations increase the width of the distribution of $h_i$, meaning that more units are in the saturated regime, and the Jacobian becomes sparser, which suppresses chaos [8]. Second, input fluctuations temporally decorrelate network states, which destroys temporal correlations of subsequent Jacobians resulting in an independent dynamic reduction of chaos [11]. The full Lyapunov spectrum, which is independent of input realization $\xi$ [112], is again obtained by a reorthonormalization procedure of the Jacobians along a numerical solution of the stochastic differential equation integrated with the Euler-Maruyama method [60]; for details, see the Supplemental Material [61].

We explored the effect of increasing input strength $\sigma$ on the Lyapunov spectrum (Fig. 14). For increasing input strength $\sigma$, the Lyapunov spectrum is increasingly pushed towards the

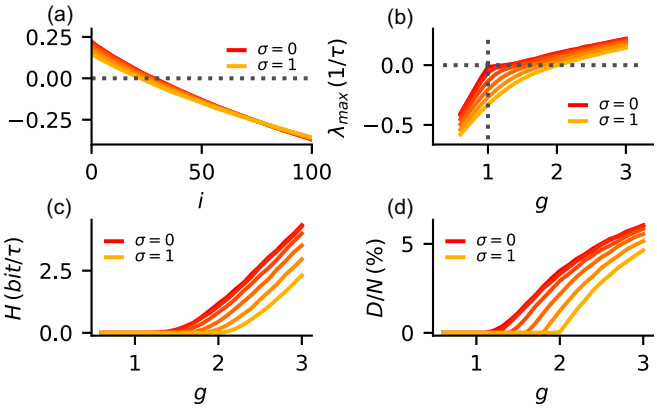


FIG. 14. Time-varying stimuli reduce both the dynamical entropy rate and attractor dimensionality. (a) For increasing input strength $\sigma$, the Lyapunov spectrum is increasingly pushed towards the mean Lyapunov exponent $-1/\tau$. (b) The largest Lyapunov exponent decreases, and the transition is smoothed, consistent with previous work [11]. (c) The dynamical entropy rate $H$ is reduced for increasing $\sigma$. (d) The relative attractor dimensionality $D/N$ decreases for increasing $\sigma$. (Parameters: $N = 1000$, $\Delta t = 10^{-2}\tau$, $t_{\rm ONS} = \tau$, $t_{\rm sim} = 10^3\tau$, averages across 10 network realizations.) $\sigma \in \{0, 0.2, 0.4, 0.6, 0.8, 1\}$.

mean Lyapunov exponent $-1/\tau$ [Fig. 14(a)]. Increasing $\sigma$ monotonically reduces the largest Lyapunov exponent as previously observed in discrete [8,73] and continuous time [11] [Fig. 14(b)]. A similar effect has been observed in rate networks driven by periodic input [9,10] and in one-dimensional mappings [113].

#### Input fluctuations reduce the conditional dynamical entropy rate and conditional attractor dimensionality

The dynamical entropy rate for a given time-dependent external input, which we will call conditional dynamical entropy rate, is calculated from the sum of the positive Lyapunov exponents, decreases for increasing external input strength $\sigma$ [Fig. 14(c)]. For sufficiently strong input, the conditional entropy rate drops to zero. Thus, time-varying input impedes the flow of information from the microscopic states to the macroscopic network states. If the information in the microscopic state is considered to be noise, one can conclude that stronger external input fluctuations reduce the noise entropy arising from sensitivity to initial conditions. The conditional attractor dimensionality also decreases for increasing input strength $\sigma$ [Fig. 14(d)]. Sufficiently strong input suppresses chaos, implying that the sample measure collapses on a (wandering) random sink [114–116]. In other words, almost all initial conditions converge onto a set of measure zero. Thus, while the network dynamics with strong time-varying input might still seem to be high-dimensional because it explores a large fraction of the phase space over time, the conditional attractor dimensionality given the external input can shrink drastically with a time-varying external input that is fixed across initial conditions. Such a transition is relevant for information processing because the network loses its dependence on initial conditions, which could be a desirable feature if the network must reliably generate different output trajectories for different input patterns [14–16]. Note that this perspective considers input that is fixed across initial conditions, which is a very different perspective from considering noise that is different across initial conditions as, e.g., in [117,118], which results in a thickening of the attractor and thus an increased dimension. The same holds for the dynamic entropy rate, which diverges in stochastic systems [119].

## IX. APPLICATIONS TO QUANTIFYING STABILITY OF TRAINED RECURRENT NEURAL NETWORKS

The networks we studied up to this point had random connectivity, but collective network dynamics is strongly shaped by wiring and learning algorithms for training recurrent neural networks in machine learning work by tuning connectivity. We now show that training a recurrent network to perform a task is reflected in the dynamic stability, as quantified by the Lyapunov spectrum, and show in some examples how it can affect the dimensionality and dynamic entropy rate. During training, network dynamics becomes confined to a low-dimensional manifold (Fig. 15). When initializing with a random network structure in the chaotic regime ($g > 1$), the dynamics evolves on a high-dimensional attractor that spans an extensive fraction of the full $N$-dimensional phase space (Fig. 2). A projection of the high-dimensional strange chaotic attractor onto the first two principal components is shown for a network of 50 rate units in Fig. 15(b). After training the network to perform a simple sine oscillation through a linear readout, the network dynamics is confined to a periodic orbit [Fig. 15(d)]. Note that the sine is computed by the coordinated activity of many rate units together that individually have dynamics different from the sine target [Fig. 15(c)]. For this task, the largest Lyapunov exponent becomes zero after training. This is expected for an autonomous network because all but the neutral direction along the flow become stabilized. Therefore, the dynamic entropy rate is trivially zero, and the attractor dimension is unity.

The Lyapunov spectrum can also be used as a quantification of how stable trajectories are after training. In Fig. 16, we compare the result of training a recurrent rate network to output an oscillation with temporally varying frequency in response to a periodic input pulse with three different training algorithms, backpropagation through time (BPTT), FORCE [15], and full-FORCE [120]. In BPTT, the full recurrent weight matrix and a readout vector are iteratively adapted by minimizing an error function using (stochastic) gradient descent [121]. FORCE recursively updates a rank-one perturbation $u_i w_j$ to $J_{ij}$ such that the linear readout $z(t) = \sum_j w_j \phi(x_j(t))$ matches a (potentially time-varying) target output. Full-FORCE does a full-rank recursive update of a task-performing network to match for each unit the activity to a teacher network. For full-FORCE, the Jacobian of the dynamics is

$$D_{ij}^{\rm fF}(t) = -\delta_{ij} + J_{ij}\phi'(x_j(t)), \tag{35}$$

for FORCE, it is

$$D_{ij}^{\rm F}(t) = -\delta_{ij} + (J_{ij} + u_i w_j)\phi'(x_j(t)). \tag{36}$$

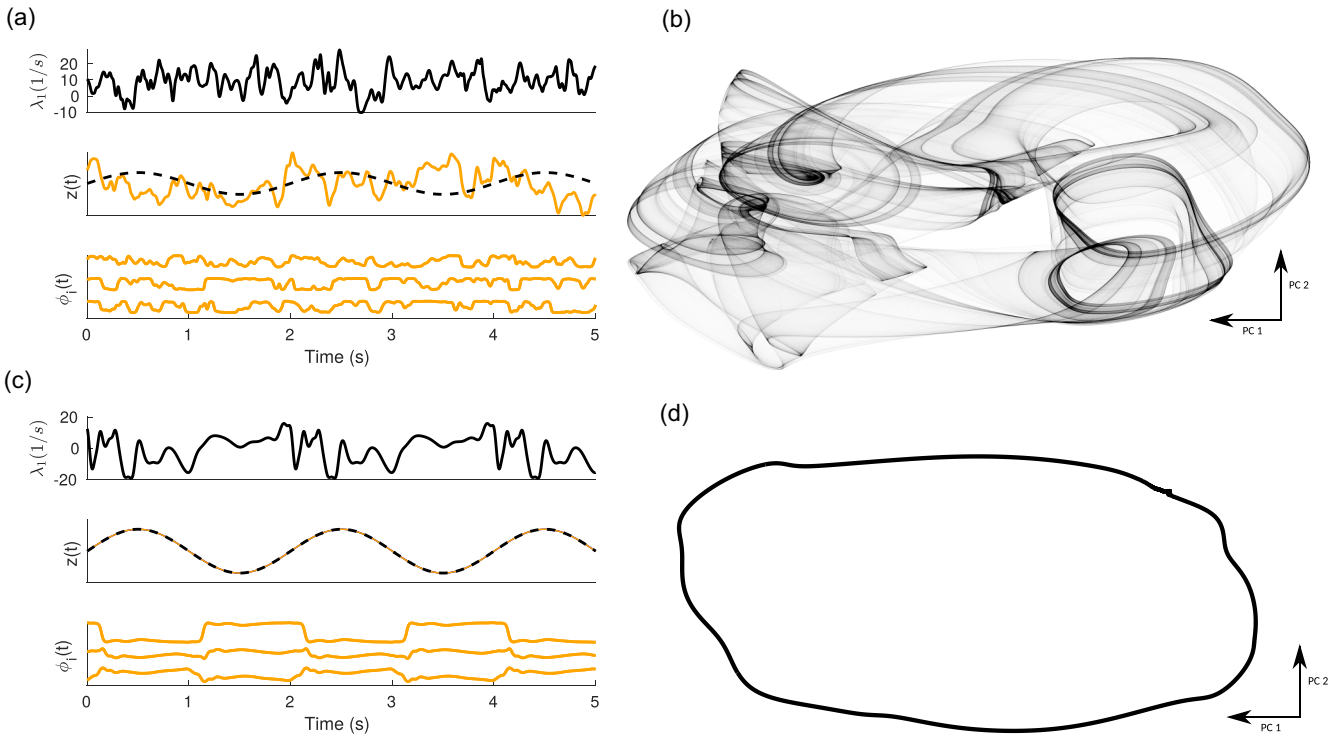


FIG. 15. Reorganization of rate network phase space during learning. (a) Local Lyapunov exponent $[\lambda_i^{\text{local}}(t)]$, output $z(t) = \mathbf{w}^\intercal \phi(\mathbf{x}(t))$, and activity of example rate units $\phi_i(t)$ before learning in the chaotic state. (b) Chaotic network activity $\phi(t)$ before learning projected on the first two principal components. (c) Same as (a) after learning a periodic task using FORCE [15]. (d) Same as (b) after training. (Parameters: $N = 200$, $g = 1.5$, $\Delta t = 10^{-1}\tau$, $t_{\text{ONS}} = 10^{-1}\tau$, $\tau = 10^{-2}$s.)

We obtained Lyapunov exponents before, within, and after training by evolving an orthonormal basis along the trajectory using the analytical Jacobians as described before and in more detail in the Supplemental Material [61].

We find that the full-rank method full-FORCE results in a more negative largest Lyapunov exponent and thus a microscopically more stable dynamics (Fig. 16). Moreover, subsequent Lyapunov exponents drop quicker towards the negative inverse of the characteristic timescale $-1/\tau$. The external periodic input pulses makes the dynamics nonautonomous; therefore, no neutral Lyapunov exponent occurs. Note that convergence of infinitesimally different initial conditions does not necessarily imply stability with respect to finite-size perturbations. For example, in spiking networks there exists multistability [122], and also trained firing-rate networks often exhibit multistability (not shown).

## X. LYAPUNOV SPECTRUM OF RECURRENT LSTM NETWORK

Training recurrent neural networks on tasks that involve long time lags with gradient-based methods is hampered by the loss of gradient information. Long short-term memory (LSTM) units were introduced to ameliorate this problem of vanishing or exploding gradients by adding a latent—potentially slow—additional degree of freedom for each rate unit with dedicated input, output, and forget gates that conspire to retain information over extended time lags [123]. The dynamics of each of the $N$ LSTM units follow the map [123]:

$$f_t = \sigma_g(U_f h_{t-1} + W_f x_t + b_f), \tag{37}$$

$$o_t = \sigma_g(U_o h_{t-1} + W_o x_t + b_o), \tag{38}$$

$$i_t = \sigma_g(U_i h_{t-1} + W_i x_t + b_i), \tag{39}$$

$$\tilde{c}_t = \sigma_h(U_c h_{t-1} + W_c x_t + b_c), \tag{40}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t, \tag{41}$$

$$h_t = o_t \odot \phi(c_t), \tag{42}$$

where $\odot$ denotes the Hadamard product, $\sigma_g(x) = \frac{1}{1+\exp(-x)}$ is the sigmoid function, $\sigma_h(x) = \tanh(x)$, and entries of the matrices $U_x$ are drawn from $U_x \sim \mathcal{N}(0, g_x^2/N)$. The bias terms $b_x$ are scalars for simplicity. Subscripts $f$, $o$, and $i$ denote respectively the forget gate, the output gate, the input gate and $c$ is the cell state. Note that each LSTM unit has two dynamic variables $c$ and $h$ and three gates $f$, $o$, and $i$ that govern signals going in and out of the cell $c$. The full Lyapunov spectrum is again obtained by a reorthonormalization procedure of the Jacobians along a numerical solution of the map [60]; for details, see the Supplemental Material [61]. As a proof-of-concept, we calculate Lyapunov spectra of recurrent LSTM networks both in the autonomous case and for Gaussian white noise input that is fixed across initial conditions (Fig. 17). In the case of external input, the entries of the input coupling matrices $W_x$ are drawn from $W_x \sim \mathcal{N}(0, \tilde{g}_x^2/N)$ and $\xi_i$ independent Gaussian white noise processes with autocorrelation function $\langle \xi_i(t)\xi_i(t+t')\rangle = \tau\sigma^2\delta(t')$ that are fixed across initial conditions. We find that saturating the forget gates by increasing

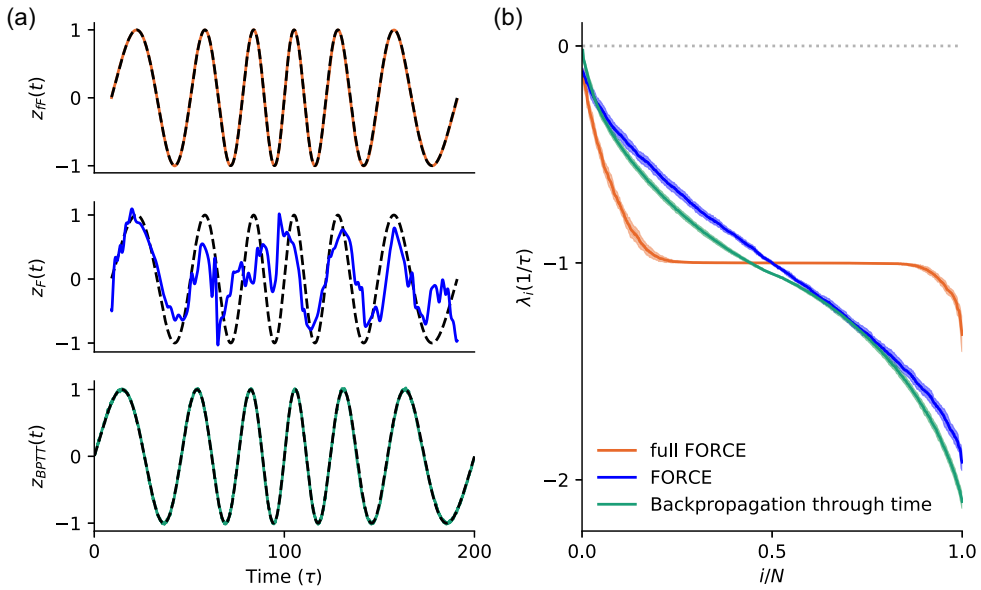


FIG. 16. Quantification of dynamic stability after training rate networks on task using FORCE, full-FORCE, and backpropagation through time. (a) Example output $z(t) = \mathbf{w}^\intercal \phi(\mathbf{x}(t))$ of a network of 500 units trained with FORCE (blue), full-FORCE (orange), and backpropagation through time (green). (b) Lyapunov exponents calculated at the end of the training. We find that the full-rank method full-FORCE results in a more negative largest Lyapunov exponent and thus a microscopically more stable dynamics. Moreover, subsequent Lyapunov exponents drop quicker towards the negative inverse of the characteristic timescale $-1/\tau$. (Other parameters: $N = 500$, $g = 1.5$, $\Delta t = 0.1\tau$, $t_{\text{ONS}} = \tau$, $t_{\text{sim}} = 10^4\tau$, $\sigma = 0$ averages across 10 network realizations.)

$b_f$ of the LSTM network results in slow latent modes and an accumulation of Lyapunov exponents close to 0 [Fig. 17(a)]. As bias current $b_f$ in the forget gate increases, the first half of the Lyapunov spectrum is increasingly pushed to zero. Concomitantly, the autocorrelation of the latent state indicates the emergence of slow latent modes [insets in Fig. 17(c)]. This finding is consistent with a previous stability analysis of the trivial fixed points of the state-to-state Jacobian $\mathbf{D}$ [analogous to Eq. (2) in our case], which suggested an accumulation of eigenvalues of the Jacobian close to 1 for closed forget gates [36,37]. At the same time, the second half of the Lyapunov spectrum sharply drops to very negative values for increasing $b_f$, similarly to the classical tanh rate network with discrete-time dynamics [Fig. 6(a)], with no plateau from an intrinsic characteristic timescale (as the one coming from a leak term [Fig. 9(c)], or a synaptic integration timescale, or adaptation current [4,6,124]). Moreover, our results indicate that LSTM networks can have a high attractor dimensionality $D$ even in a weakly chaotic state when saturating the forget gates [Fig. 17(d)], as a growing number of near-zero Lyapunov exponents is necessary to yield a total sum of zero. In contrast, the dynamical entropy rate $H$ is only governed by positive Lyapunov exponents, which do not reflect the large number of Lyapunov exponents close to zero for increasing $b_f$ [Fig. 17(c)]. Therefore, the attractor dimensionality increases for large $b_f$ while the dynamical entropy rate decreases. We find this phenomenon independent of network size $N$ (not shown), which suggests extensive chaos, as in the classical rate networks (see Fig. 2). Note that the largest Lyapunov exponent, entropy rate $H$, and attractor dimensionality $D$ vary considerably across network realizations (error bars in Fig. 17 indicate double standard dev. across 10 network realizations), consistent with our findings in classical tanh networks (see Appendix B).

Driving each LSTM with independent Gaussian white noise process independent ($W_x \neq 0$) input leads to a reduction of chaos, decreasing dynamical entropy rate $H$ and attractor dimensionality $D$ (Fig. 17 dashed and dotted lines). In contrast to the case of the classical rate networks [11] (Fig. 14), we find that slow modes and near-zero Lyapunov exponents can persist even in the presence of strong input. This might be explained by the multiplicative gating, which is different from the case of classical rate networks [11] (Fig. 14).

## XI. RELATING GRADIENTS IN BACKPROPAGATION THROUGH TIME TO THE FULL LYAPUNOV SPECTRUM

We found a direct mathematical link between the Lyapunov spectrum of the recurrent network dynamics and the problem of vanishing and exploding gradients when training with backpropagation through time. In backpropagation through time, all connection weights are iteratively updated by stochastic gradient descent such that a loss is locally reduced [125–128]. The gradient of the loss with respect to the weights of the recurrent network is evaluated by unrolling the network dynamics in time. The resulting expression for the gradient involves the long-term Jacobian $\mathbf{T}_t$ (more details in Appendix E),

$$\frac{\partial E_t}{\partial \mathbf{J}} = \frac{\partial E_t}{\partial \mathbf{h}_t} \sum_\tau \left( \prod_{\tau'=\tau}^{t-1} \frac{\partial \mathbf{h}_{\tau'+1}}{\partial \mathbf{h}_{\tau'}} \right) \frac{\partial \mathbf{h}_\tau}{\partial \mathbf{J}} \tag{43}$$

$$= \frac{\partial E_t}{\partial \mathbf{h}_t} \sum_\tau \mathbf{T}_t(\mathbf{h}_\tau) \frac{\partial \mathbf{h}_\tau}{\partial \mathbf{J}}. \tag{44}$$

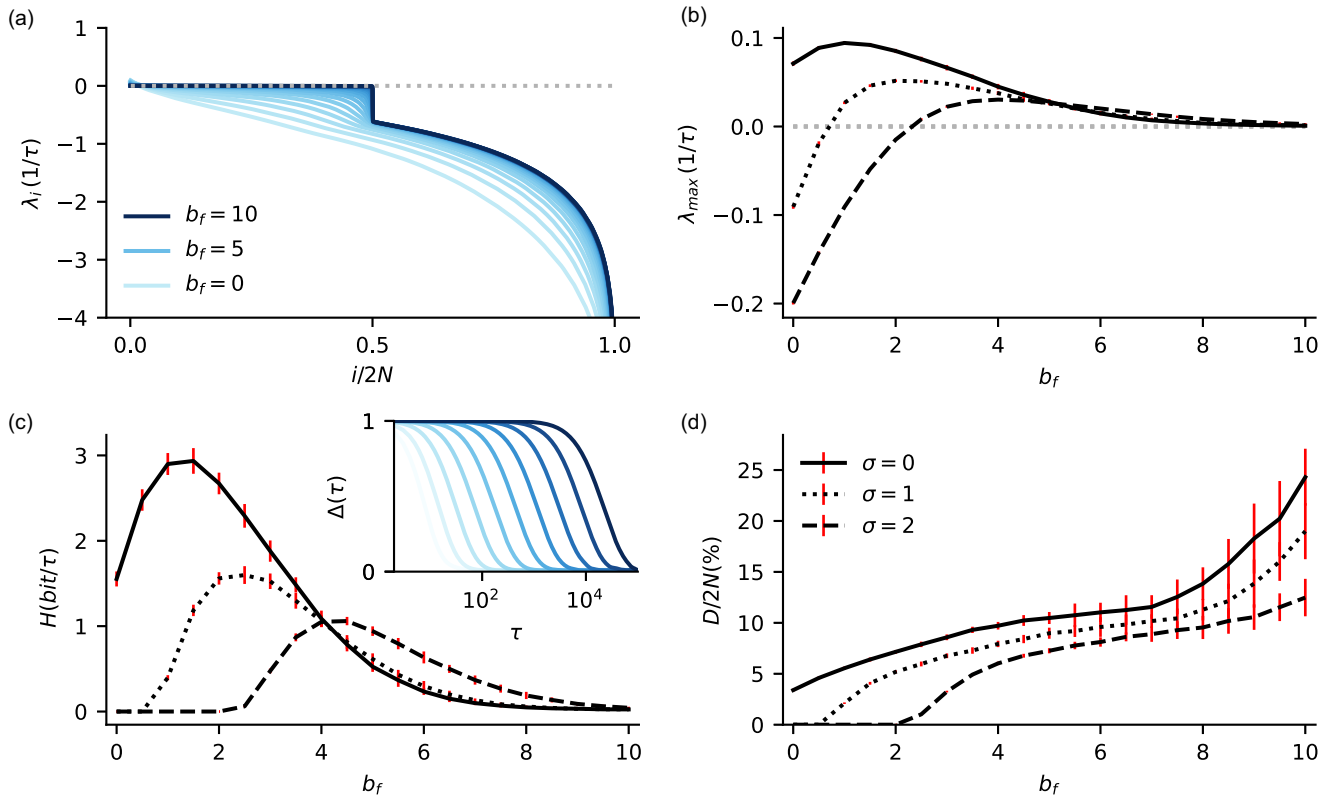


FIG. 17. Lyapunov spectrum, dynamical entropy rate, and dimensionality of recurrent LSTM network. (a) For increasing bias current in the forget gate $b_f$, the first half of the Lyapunov spectrum is increasingly pushed towards zero. Concomitantly, the autocorrelation of the latent state indicates the emergence of slow latent modes. [In inset in (c) the normalized autocorrelation of $h$ averaged over all units in network show emergence of slow timescales for closed gates.] (b) The largest Lyapunov exponent goes for increasing $b_f$ toward zero (full line not external input, dotted and dashed lines show input strengths $\sigma \in \{1, 2\}$). Note that input does not push the Lyapunov exponent below 0. (c) The dynamical entropy rate $H$ is also reduced for growing $b_f$. (d) In contrast, the relative attractor dimensionality $D/2N$ even increases with $b_f$, despite decreasing $\lambda_{\max}$. Dimensionality is reduced by time-varying external input. (Parameters: $N = 2000$, $t_{\text{ONS}} = 1$, $t_{\text{sim}} = 10^4$, $g_f = g_o = g_i = g_c = 3.0$, $b_o = b_i = b_c = 0.0$, median across 10 network realizations, red error bars indicate double std across 10 network realizations.)

A common problem in training recurrent networks is that because of the chain of matrix multiplications in the long-term Jacobian $\mathbf{T}_t$, the gradients tend to vanish or to explode exponentially with time, which impedes training especially in the case of long temporal dependencies [33,121,123,129]. How well error signals can be propagated backward in time is constrained by the dynamics in the tangent space that is spanning a tangle of stable and unstable manifolds nearby the trajectory.

The singular values of the long-term Jacobian, which determine how quickly gradients vanish or explode during backpropagation through time, are directly related to the Lyapunov exponents of the dynamics: The Lyapunov exponents are given by the logarithm of the singular values of the long-term Jacobian (see Appendix E). They also yield a direct estimate of the condition number of the long-term Jacobian [Eq. (45)],

$$\kappa_2\big(\widetilde{\mathbf{Q}}_{s+s_{\text{ONS}}}\big) = \kappa_2\big(\mathbf{R}^{s+s_{\text{ONS}}}\big) = \frac{\sigma_1\big(\mathbf{R}^{s+s_{\text{ONS}}}\big)}{\sigma_m\big(\mathbf{R}^{s+s_{\text{ONS}}}\big)} = \frac{\mathbf{R}_{11}^{s+s_{\text{ONS}}}}{\mathbf{R}_{mm}^{s+s_{\text{ONS}}}}. \quad (45)$$

The condition number $\kappa_2$ of the deformed orthonormal system can be estimated based on the ratio of the largest $\lambda_{\max}$ and the last Lyapunov exponent $\lambda_m$ that is calculated,

$$\kappa_2\big(\widetilde{\mathbf{Q}}_{s+s_{\text{ONS}}}\big) \approx \exp\left((\lambda_{\max} - \lambda_m) s_{\text{ONS}} \Delta t\right).$$

Thus, the Lyapunov spectrum shapes the number of dimensions available for error propagation. To avoid diverging or vanishing gradients, one should initialize recurrent networks such that many singular values of the long-term Jacobian are close to one [31,34–38]. This is equivalent to having many Lyapunov exponents of the forward dynamics close to zero, which means that many directions in tangent space grow and shrink only slowly. As the product of Jacobian is generally numerically ill conditioned, we suggest using the orthonormalization procedure discussed here to quantify and improve the stability of the tangent space. Furthermore, the trainability of RNNs, as quantified by the maximum time difference a recurrent neural network can be trained across using BPTT before running into vanishing/exploding gradients, can be quantified by Lyapunov exponents of the forward dynamics. We can thus use Lyapunov exponents to compare the effect

of different initializations, nonlinearities, and optimizers on trainability. We predict that after learning long-term dependencies, there should be some Lyapunov exponents close to zero reflecting the slow timescales.

## XII. DISCUSSION

### A. Summary

We used canonical measures from the ergodic theory of strange attractors to characterize the chaotic dynamics of randomly wired networks of firing-rate units. In this paper we calculated the full Lyapunov spectrum of a continuous-time random rate network and use it to study dynamical entropy rate and attractor dimensionality.

We showed that, in the classical model, dynamical entropy rate and relative attractor dimensionality first grow and then saturate as a function of coupling strength $g$. Thus, both the intensity and diversity of network activity states saturates for strong coupling, despite a monotonously growing largest Lyapunov exponent. We analytically approximated the full Lyapunov spectrum in several limiting cases using random matrix theory. We found that time-varying input reduces both entropy and dimensionality.

We demonstrated that the shape of the Lyapunov spectrum is size invariant and exhibits a linear growth of attractor dimensionality and entropy rate with network size $N$. This is clear evidence of extensive chaos, which was previously conjectured in [1]. We further found the Lyapunov spectrum to be point symmetric around the mean Lyapunov exponent $-1/\tau$, which we derived analytically (Appendix D). A symmetry of Lyapunov spectra around zero is usually found in dynamical systems with a symplectic structure [100,103]. Symmetry around a negative value was previously described in a class of dissipative dynamical systems with viscous damping [100].

We found a strong effect of time-discretization: Increasing the step size breaks the symmetry of the Lyapunov spectrum and sharply increases the entropy rate and dimensionality. This has methodological implications for further studies: The leak term present in continuous-time networks but absent in discrete-time networks leads to fundamental differences in the properties of their dynamics.

In balanced networks of threshold-linear units, we found that similar to the classical tanh model, both the entropy rate and attractor dimensionality first increase for small values of $g$ and peak for large $g$. Different from the classical tanh model, the Lyapunov spectrum is not point symmetric. Moreover, we observed very different PCA-based dimensions, depending on whether they are calculated based on firing rates $\phi$ or on the local fields $h$.

Time-dependent input reduced both the entropy rate and attractor dimensionality. For strong input, we found that all trajectories collapsed to a time-dependent random sink. If the input is interpreted as an incoming signal, this means that trajectories are reliable across repetitions of the same input realization and do not depend on the initial conditions of the recurrent network.

Finally, we showed that Lyapunov spectra are a useful tool to characterize dynamic stability properties of trained networks and to analyze the solution trajectories without assuming fixed points or "slow points" as done, for instance, in [130–133]. Moreover, we show a direct link between the Lyapunov exponents of the forward dynamics and the gradient stability when training recurrent neural networks with backpropagation through time.

### B. Relation to previous work

Firing-rate networks can generate spontaneous rate fluctuations by recurrent chaotic dynamics [1]. Mechanisms underlying rate chaos have attracted substantial attention in studies of network heterogeneity [3], bistability [5], external stimuli [8–11], and the role of the single unit transfer function [2] and slow synaptic dynamics [4,7] for the collective network state; see also, e.g., [72,73,92,134–138]. Our approach provides a toolkit from dynamical systems theory to analyze how these different factors shape the complex rate dynamics and the structure of the tangent space.

We compared the attractor dimension with a dimensionality estimate based on principal component analysis, which is commonly used in neuroscience [9,10,64,65,69,139]. We often find a different behavior of the PCA-based dimensionality and the attractor dimension: For classical tanh-rate networks, the attractor dimension peaks with $g$, but both PCA-based dimensions monotonously increase and saturate, albeit at different levels and with distinct rates.

Note that Lyapunov exponents and the attractor dimension are invariant under diffeomorphisms of the phase space [50]. In contrast, PCA-based dimensionality estimates are generally not invariant with respect to changes of coordinates and can be misleading for limited data sets [140]. Because the PCA-based dimensionality estimates are based on a two-point correlation function, they miss low-dimensional structure hidden in higher-order correlations. Generally, the PCA-based dimensionality can both under- and overestimate the attractor dimensionality.

There are different dimensionality estimates used in neuroscience that carry different meanings and interpretations: besides quantifying the dimensionality of a single trajectory across time [9,64,65], the dimensionality across different stimuli [141], or within stimulus categories [142] or task-related [143] and others have been considered. Note that these dimensionality estimates often assume a multivariate Gaussian distribution of the activity variable. It is largely an open question, how these relate to the attractor dimensionality discussed here. If the Gaussianity assumption is dropped, there are more agnostic sampling-based estimates of entropy rate and dimensionality available [52–54], but the data required scales exponentially in $D$ [55–58].

Our approach allows interpolation from continuous time to discrete dynamics. Discrete-time dynamics of rate networks has previously been studied in random diluted networks [72], networks with random external fields [84,144], noise-driven networks [8], and on a ring network [83,145]. Extensive Lyapunov spectra were already found in earlier studies in discrete-time rate networks [83,84], as well as in spiking neural networks [91,146–148]. A deeper mathematical understanding of the necessary conditions for extensive chaos is still missing.

Chaotic rate dynamics provide a substrate for complex nonlinear computations, such as learning input-output relations [13,15,17,26,131,149,150] and learning temporal sequences [16]. Intriguingly, transient rate chaos yields exponential expressivity in deep networks, which has been explained by transient chaos across layers [32]. Our tools facilitate the quantification of the reorganization of the collective network dynamics during learning and the underlying mechanisms of different computing strategies. Recently, similar reservoir computing techniques have been used for time-series prediction of spatiotemporal chaotic systems [150,151] and to infer its ergodic properties (e.g., the first few Lyapunov exponents) from data [152]. This was recently also extended to gated networks trained with backpropagation [153].

A suppression of chaos by time-dependent input was studied previously, both with white noise input in discrete-time [8] and continuous-time networks [11] and with sinusoidal input [10]. Such a transition has relevance for information processing because the network loses its dependence on initial conditions, which is expected to affect the ability of a network to generate controlled output trajectories in response to specific input patterns after learning [14–16]. A transition to complete control by an external stimulus and concomitant independence of initial conditions was previously studied in rate networks in the context of echo state networks for reservoir computing and termed the echo state property [154–157]. In FORCE-trained networks, the recurrent network is driven by a low-rank input during training, which is replaced by a feedback loop after training [15,158]. It would be interesting to extend our work to low-rank perturbations of the recurrent weight matrix and to investigate how such perturbations shape the dynamics and the whole Lyapunov spectrum. Recently, it was found that for balanced networks, much larger input amplitude is necessary to suppress chaos if neurons receive common time-varying input, compared to when all neurons receive independent input [159].

It might seem surprising that the entropy and dimensionality are reduced by time-varying input, which increases the variance of the network activity. This can be understood from a perspective of random dynamical systems, where the time evolution of a perturbation is studied with a fixed external input realization. In this perspective, the *conditional* entropy and *conditional* attractor dimension for a given external input are reduced as the variance of the time-varying external input is increased; thus, the network becomes more reliable [160]. In all scenarios studied here (both with tanh and threshold-linear transfer function $\phi$ and both for discrete and continuous-time dynamics), we found that temporally uncorrelated input reduces chaos, and thus makes the network more reliable. This is in contrast to other systems, where time-varying input can also enhance chaos (e.g., in the kicked rotor or in spiking networks [161,162]).

An accumulation of Lyapunov exponents close to zero is consistent with previous theoretical work based on spectra of the state-to-state Jacobian $\mathbf{D}$ [analogous to Eq. (2) in our case], which suggested an accumulation of eigenvalues of the Jacobian close to 1 for closed forget gates [36,37]. This work was recently also extended to a dynamic mean-field theory of gated networks [163].

### C. Outlook

We are only beginning to use ergodic theory to understand neural computation. By employing these concepts in large-scale rate networks, we have laid a foundation for further investigation. Computational ergodic theory of firing-rate networks is currently the only way to measure information-theoretic quantities in large recurrent circuits. It is an important challenge to obtain a more comprehensive understanding of how different factors shape collective network dynamics.

The link between firing-rate networks and spiking neural networks has been studied by investigating networks in the limit of very slow synaptic dynamics. In this limit, the synaptic input current integrates over a long time, and the network dynamics is analogous to a rate network [7] with quantitatively similar activity fluctuations. An interpolation from spiking to rate dynamics with increasing $\tau_s$ and a comparison of the associated Lyapunov spectra of rate and spiking networks might improve our understanding of chaos both in spiking and rate networks.

Collective network dynamics is expected to be strongly shaped by the detailed weight matrix, and learning algorithms operate by modifying the dynamics via changes of the connectivity. Investigating how features of connectivity shape the dynamics is therefore important and can also be investigated with these tools. The role of an excess of bidirectional connections [164], other second-order motifs [165], and strong self-coupling [5] could all be examined. A time-resolved analysis of dynamic stability of the recurrent network dynamics using covariant Lyapunov vectors and local Lyapunov exponents can also help to understand the mechanisms of learning, the geometry of error propagation, and under what conditions the training of recurrent networks fails.

## ACKNOWLEDGMENTS

We thank A. Crisanti, F. Fumarola, S. Goedeke, J. Kadmon, R. Khajeh, G. Lajoie, J. Liedtke, L. Logiaco, U. Parlitz, M. Schottdorf, M. Stern, H. Sompolinsky, and M. Timme for fruitful discussions. Research supported by NSF NeuroNex Award (Grant No. DBI-1707398), the Gatsby Charitable Foundation (Grant No. GAT3708), the Simons Collaboration for the Global Brain (542939SPI), and the Swartz Foundation (2019-5). This work was further supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) 436260547 in relation to NeuroNex (National Science Foundation 2015276) & under Germany's Excellence Strategy - EXC 2067/1- 390729940, DFG - Project-ID 317475864 - SFB 1286, DFG - Project-ID 454648639 - SFB 1528, DFG - Project-ID 273725443 - SPP 1782, DFG - Project-ID 430156276 - SPP 2205, and by the Leibniz Association (project K265/2019) (F.W.).

## APPENDIX: OVERVIEW

We first describe several checks of the convergence of the Lyapunov spectra with various system parameters (A). We then describe finite network-size effects (B) and compare dynamic mean-field theory with the analytical behavior for large $g$ and $g \to g^{+}_{\mathrm{crit}}$ (C). We then give an analytical approximation

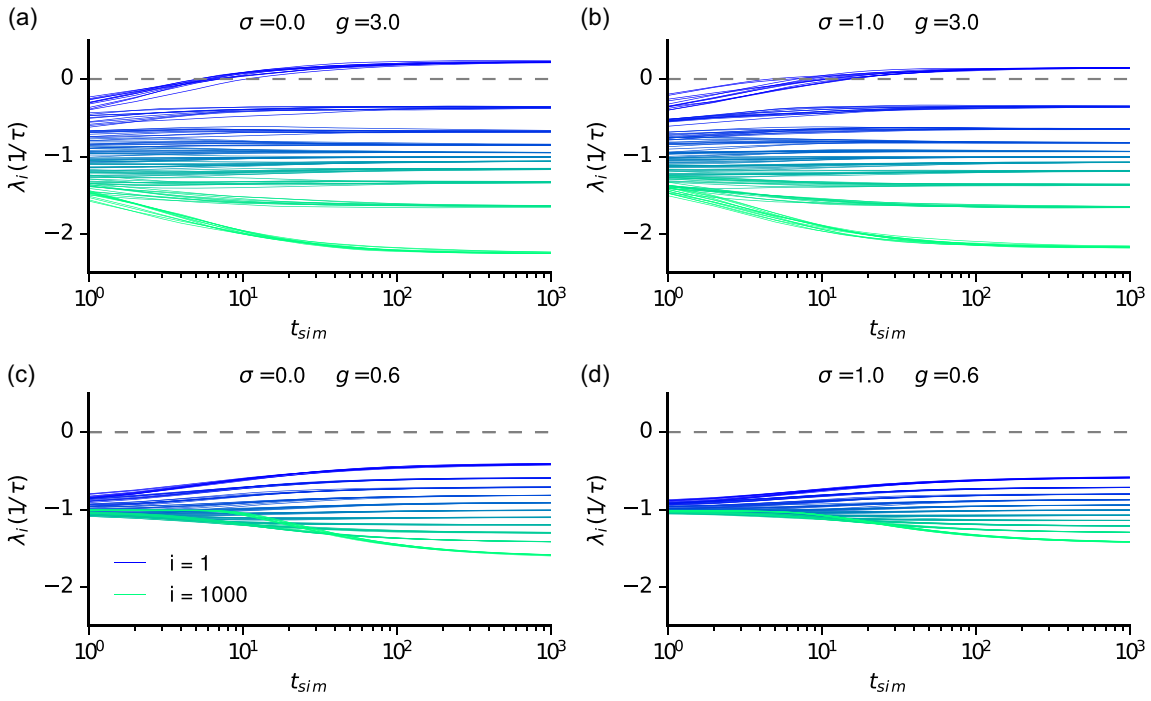


FIG. 18. Convergence of Lyapunov spectrum with simulation time $t_{\rm sim}$. (a) Convergence of selected Lyapunov exponents $\lambda_i$ for ten different network realizations with simulation time (in units of $\tau$)($i = 1, 100, 200, \ldots 1000$) for $\sigma = 0$ and $g = 3$. (b) Same as top left, but for $\sigma = 1$ and $g = 3$. (c) $\sigma = 0$ and $g = 0.6$. (d) $\sigma = 1$ and $g = 0.6$. (Other parameters: $N = 1000$, $\Delta t = 0.01\tau$, $t_{\rm sim} = 10^3\tau$, $t_{\rm ONS} = \tau$).

of the mean Lyapunov exponent (D) and provide additional details the the mathematical link of gradients in BPTT and Lyapunov spectrum (E). Finally, we show that $H$ is bounded as function of $g$ (F).

## APPENDIX A: CONVERGENCE OF THE LYAPUNOV SPECTRUM

We checked the convergence of the Lyapunov spectrum as a function of different simulation parameters. First, the Lyapunov exponents were checked to converge with simulation time $t_{\rm sim}$ (Fig. 18). Figure 12 shows the temporal convergence of selected Lyapunov exponents for ten random network realizations for different values of $g$ and $\sigma$. The Lyapunov spectra were independent of initial conditions but showed some variability across different realizations of the random network structure. There are two main contributions to the variability of numerically calculated Lyapunov spectra, finite-time sampling noise and quenched fluctuations. Indeed, Lyapunov exponents are asymptotic properties numerically estimated from finite-time calculations. Variability also arises from the quenched disorder across different random network realizations. The first contribution would vanish in the limit of long simulations for ergodic systems. The second contribution is expected to vanish in the large network limit due to self-averaging. Quantities that are self-averaging converge in the limit of large system size to the ensemble average.

Second, we confirmed that the orthonormalization interval is chosen sufficiently small [Fig. 19(a)]. If the reorthonormalization is not carried out sufficiently often, the long-term Jacobian $\mathbf{T}_t(\mathbf{x}_0)$ becomes ill conditioned. As a consequence, the orthonormalization becomes numerically unstable, and errors start to accumulate. This results in a flattening of the Lyapunov spectrum beginning at small Lyapunov exponents [Figs. 19(a) and 19(d)]. As described above, a suitable orthonormalization interval inversely scales with the difference between the smallest and largest Lyapunov exponent that is calculated $|\lambda_{\rm max} - \lambda_k|$. Therefore, it is no surprise that for large $\Delta t$, the errors in the Lyapunov spectrum grow faster with $t_{\rm ONS}$ [Figs. 19(c) and 19(d)], because the difference $|\lambda_{\rm max} - \lambda_k|$ is larger [Fig. 19(a)].

Third, we checked convergence with the integration time step $\Delta t$ [Fig. 7(a)]. For large $g$, the integration time step $\Delta t$ has to be chosen smaller, because the autocorrelation of the Jacobians becomes very short ($\tau_{AC} \ll \tau$), despite the finite autocorrelation of the dynamical variables $h_i$ for $g \to \infty$ [1,59].

Fourth, we confirmed the convergence of the shape of the Lyapunov spectrum for large network size $N$ [Fig. 2(b)]. Note that even for very small $\Delta t$, there exists a small asymmetry in the Lyapunov spectrum because of the neutral Lyapunov exponent ($\lambda_i = 0$). Thus, the Lyapunov spectrum is only symmetric in the limits $N \to \infty$ and $\Delta t \to 0$. Fifth, we confirmed numerically that the neutral Lyapunov exponent ($\lambda_i = 0$) associated with a perturbation in the direction of the flow converges towards zero in the limit of small $\Delta t$ (not shown).

Sixth, we confirmed numerically that the Lyapunov spectrum does not depend on the realization of the initially random orthonormal system. Dependence in the realization of the orthonormal system would indicate that the ONS did not converge to the eigenvectors of the Oseledets matrix Eq. (6) [166] (not shown).

Seventh, for large $N$, the numerical estimate of the largest Lyapunov exponent can be compared to one calculated

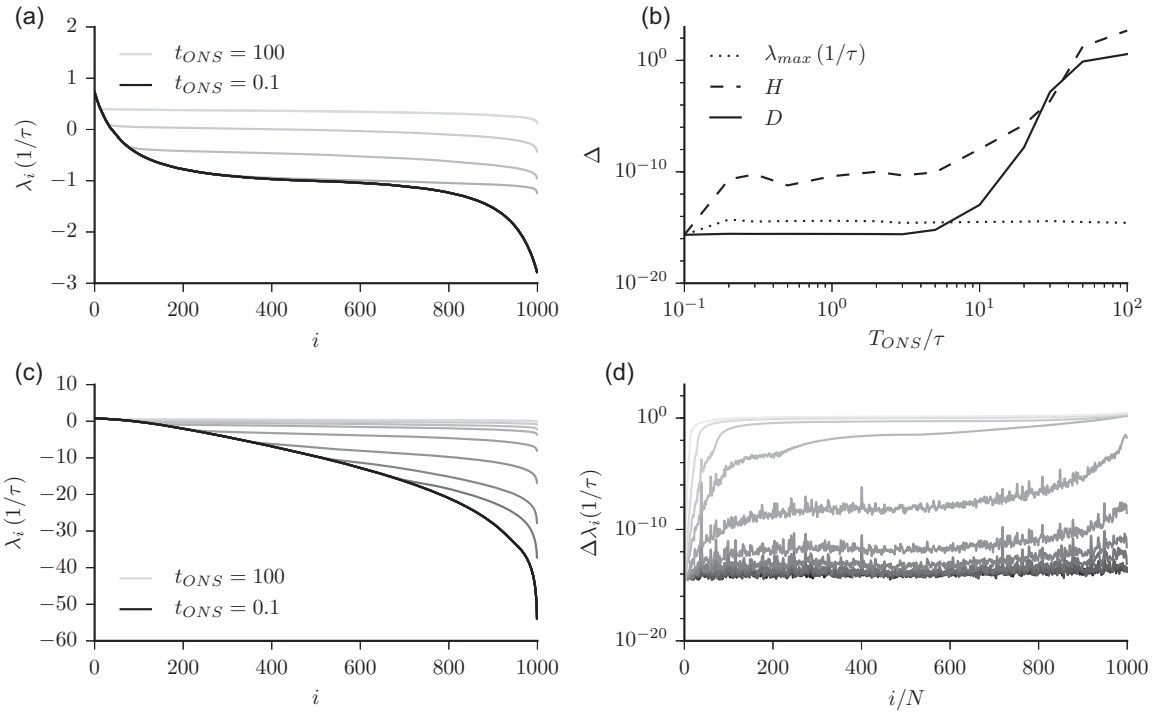


FIG. 19. Convergence of Lyapunov spectrum with reorthonormalization interval $t_{\mathrm{ONS}}$. If the reorthonormalization is not performed sufficiently often, the Lyapunov spectrum is flattening from the end for large $t_{\mathrm{ONS}}$. (a) Lyapunov spectra for $t_{\mathrm{ONS}} \in \{0.1, 0.2, 0.3, 0.5, 1, 2, 5, 10, 20, 50, 100\}\tau$ for $\Delta t = 0.01$. (b) $\Delta\lambda_{\max}$ shows the deviation of the largest Lyapunov exponent for different $t_{\mathrm{ONS}}$ from the smallest $t_{\mathrm{ONS}} = 0.1\tau$. The same is shown for $H$ and $D$. For our typical parameter sets, an orthonormalization interval of $t_{\mathrm{ONS}} = 1\tau$ is sufficient to keep errors in $H$ and $D$ orders of magnitudes smaller than the deviations across network realizations due to quenched fluctuations. (c) Same as (a) for $\Delta t = 1$. (d) Deviations of full Lyapunov spectra for different $t_{\mathrm{ONS}}$ from the smallest $t_{\mathrm{ONS}} = 0.1$ for $\Delta t = 0.01$. (Other parameters: $N = 1000$, $\Delta t = 0.01\tau$, $g = 10$, $t_{\mathrm{sim}} = 10^4\tau$, averages across 10 network realizations.)

analytically using dynamic mean-field theory [1,2,8,11,59] [Figs. 20(a) and 20(c)].

Eighth, we confirmed that the Lyapunov spectrum does not systematically change when increasing the floating-point precision by using arbitrary-precision floating-point arithmetic in spot checks (not shown).

## APPENDIX B: FINITE NETWORK SIZE EFFECTS ON THE TRANSITION TO CHAOS AND LYAPUNOV EXPONENTS

We described so far chaos in large firing-rate networks. Here, we investigated the finite network size effect on the largest Lyapunov exponent and the critical coupling strength $g_{\mathrm{crit}}$, where the transition to chaos occurs. We calculated for the largest Lyapunov exponent of the classical random rate networks with tanh-nonlinearity as a function of network size $N$ for 100 network realizations per size. We found that the largest Lyapunov exponent for small networks exhibits a large diversity across network realizations [Fig. 20(a)]. For increasing network size, the median Lyapunov exponent increases and approaches an asymptotic limit for large $N$. At the same time, the diversity as quantified by the 20% and 80% percentile across the network realizations vanishes. This indicates that for large network size $N$, the variability of Lyapunov exponents coming from the quenched disorder of different network realizations vanishes, and the Lyapunov exponent becomes independent of network realization.

Complementary, we calculated for different realizations and different network size $N$ the critical coupling strength $g_{\mathrm{crit}}$ where the network turns chaotic as indicated by the largest Lyapunov exponent using a noisy bisection method. For small networks, we found a broad diversity of $g_{\mathrm{crit}}$ [Fig. 20(b)]. For many small random network realizations, we could not find a chaotic regime at all. For increasing values of $N$, the median $g_{\mathrm{crit}} - 1$ across 100 realizations decreased $\propto 1/\sqrt{N}$ from a median $g_{\mathrm{crit}} \approx 4$ for $N = 50$ to $g_{\mathrm{crit}} \approx 1.1$ for $N = 10^4$, and the diversity of $g_{\mathrm{crit}}$ as quantified by the 20% and 80% percentile across the network realizations shrank $\propto 1/\sqrt{N}$. This indicates that for large $N$, the coupling strength $g_{\mathrm{crit}}$ converges to 1 and the variability arising from quenched fluctuations disappears. Note that for small networks, there exists not necessarily a unique critical coupling strength $g$, so details of Fig. 20(b) may depend on the bisection scheme utilized.

## APPENDIX C: DYNAMIC MEAN-FIELD THEORY

We used dynamic mean-field theory to obtain the autocorrelations and the largest Lyapunov exponent. Briefly, following [1,11,59], we solved the autocorrelations $\Delta(\tau)$ self-consistently. We first obtained the variance $\Delta_0 = \Delta(0)$, integrated them to obtain $\Delta(\tau)$. We calculated the largest Lyapunov exponent by calculating the ground-state energy via $\lambda_{\max} = -1 + \sqrt{1 - \epsilon_0}$, where the $\epsilon_0$ is obtained from the smallest eigenvalue of the time-independent Schrödinger eigenvalue equation, where the quantum potential $W(\tau) =$

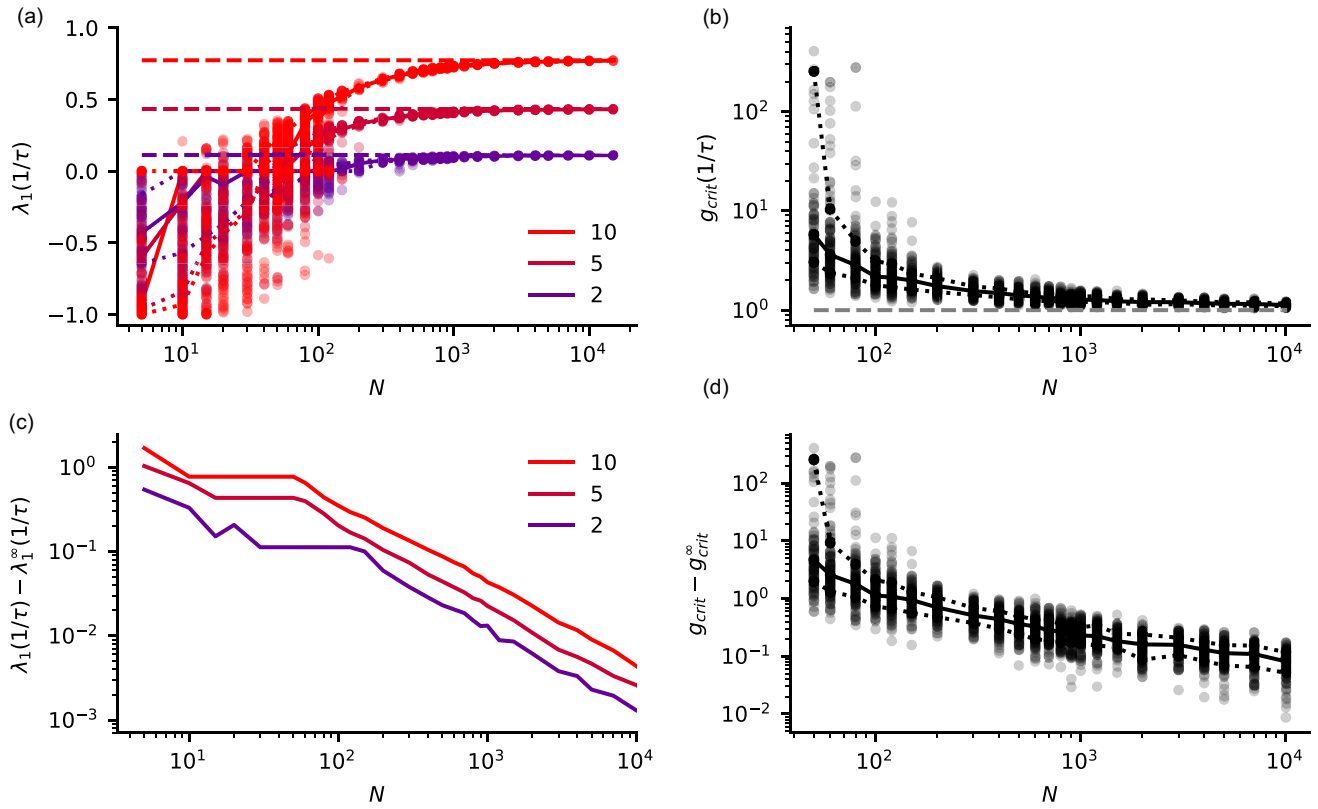


FIG. 20. Finite-size effect on the largest Lyapunov exponent and transition to chaos. (a) Largest Lyapunov exponent across 100 network realizations as a function of network size $N$ for $g \in \{2, 5, 10\}$. Dots indicate individual realizations, full line are median, dotted curves are 20% and 80% percentile, dashed lines are the prediction obtained from dynamic mean-field theory where $g$ is color-coded from blue (small $g$) to red (large $g$). (b) Critical coupling strength $g_{\text{crit}}$ as a function of network size $N$ for 100 network realizations obtained by bisection method. Dots indicate individual realizations, full line are median, dotted curves are 20% and 80% percentile, dashed line the analytical prediction. (c) Difference between mean-field theory prediction and median across 100 realizations as a function of network size. (d) Same data as (b) but $g_{\text{crit}} - 1$ depicted on log-scale (Other parameters: relative tolerance $= 10^{-10}$, $t_{\text{sim}} = 10^4\tau$, $t_{\text{ONS}} = \tau$, median across 100 network realizations).

$-V''(c(\tau)) = 1 - g^2 f_{\phi'}(c(\tau), c_0)$ is evaluated based on the self-consistent solution of the autocorrelation $\Delta(\tau)$ [1,11,59]. We compared the solution of the dynamic mean-field theory with the previously proposed explicit expressions for autocorrelations and the largest Lyapunov exponent in the limits $g \to g_{\text{crit}}^{+}$ and $g \to \infty$ (Fig. 21).

## APPENDIX D: RANDOM MATRIX THEORY OF MEAN LYAPUNOV EXPONENT

From the Jacobian, we derive a random matrix approximation of the mean Lyapunov exponent $\bar{\lambda} = \frac{1}{N}\sum_{i=1}^{N}\lambda_i$. The mean Lyapunov exponent describes the average dissipation rate of phase space compression, captured by the determinant of the long-term Jacobian $\mathbf{T}_t = \mathbf{D}_t \cdots \mathbf{D}_0$. In the discrete-time case, the Jacobian is given by

$$D_{ij}(t_s) = \left.\frac{\partial f(h_i(t))}{\partial h_j(t)}\right|_{t=t_s} = (1-\Delta t)\delta_{ij} + \Delta t \cdot J_{ij}\,\mathrm{sech}^2(h_j(t_s)). \tag{D1}$$

It is known that in the chaotic regime for large $N$, the activity variables $h_i$ approximately follow a Gaussian distribution both in discrete and continuous time, $h \sim \mathcal{N}(0, \Delta_0)$, where for large $N$, $\Delta_0$ solely depends on $g$ [1,2,8,11]. The variance of $h_i$ grows with $g^2$, thus the squared hyperbolic secant of $h_i$ is close to zero for most $i$. For this reason, in the case of strong $g$, most columns of $D_{ij}(t_s)$ are, aside from the diagonal entries, close to zero and $D_{ij}$ becomes sparse.

The long-term Jacobian $\mathbf{T}_t(\mathbf{h_0})$ is

$$\begin{aligned}\mathbf{T}_t(\mathbf{h}_0) &= \mathbf{D}_{t-1}(\mathbf{h}_{t-1})\ldots\mathbf{D}_1(\mathbf{h}_1)\mathbf{D}_0(\mathbf{h}_0)\\ &= \prod_{s=0}^{t-1}\mathbf{D}_s\\ &= \prod_{s=0}^{t-1}((1-\Delta t)\mathbb{1} + \Delta t\cdot\mathbf{J}\cdot\mathrm{sech}^2(\mathbf{h}(t_s))).\end{aligned}$$

Thus, the mean Lyapunov exponent for large $N$ is

$$\begin{aligned}\bar{\lambda} &= \left[\frac{1}{N}\sum_{i=1}^{N}\lambda_i\right] = \left[\frac{1}{N}\sum_{i=1}^{N}\ln\mu_i\right] = \left[\frac{1}{N}\ln\prod_{i=1}^{N}\mu_i\right]\\ &= \left[\frac{1}{N}\ln(\det\mathbf{\Lambda})\right] = \left[\frac{1}{N}\ln\left(\det\lim_{t\to\infty}[\mathbf{T}_t(\mathbf{x}_0)^{\top}\mathbf{T}_t(\mathbf{x}_0)]^{\frac{1}{2t}}\right)\right]\\ &= \left[\frac{1}{N\tau}\lim_{t\to\infty}\frac{1}{t}\ln(\det\mathbf{T}_t)\right]\\ &= \frac{1}{N\tau\Delta t}\left[\lim_{n\to\infty}\frac{1}{n}\sum_{s=0}^{n-1}\ln(\det((1-\Delta t)\mathbb{1}+\Delta t\mathbf{J}\phi'(\mathbf{h}(t_s))))\right]\end{aligned}$$

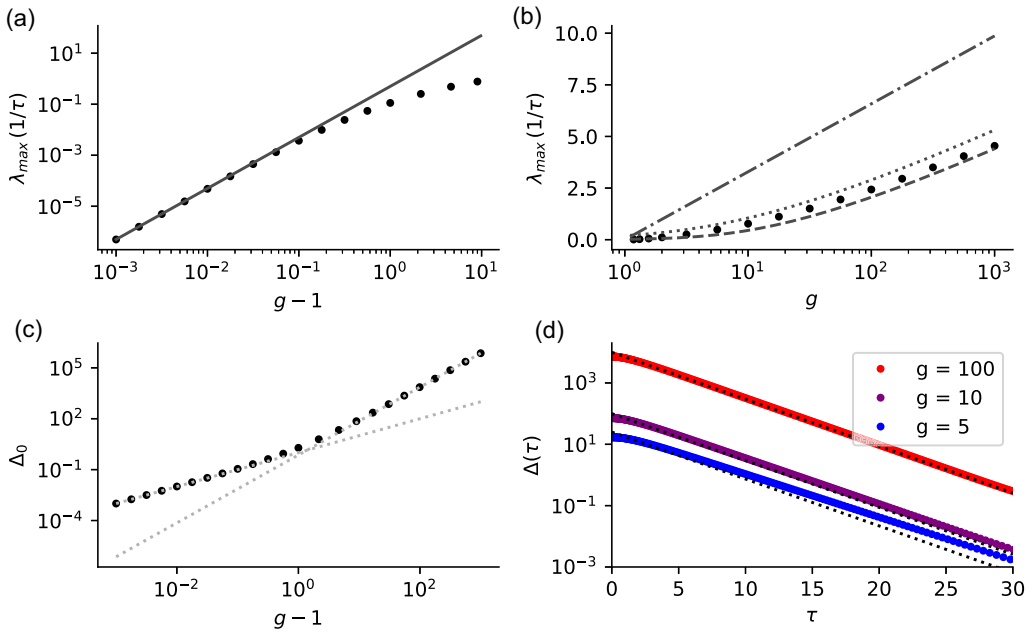


FIG. 21. Dynamic mean-field theory of autocorrelations and the largest Lyapunov exponent in the limits $g \to g^+_{\text{crit}}$ and $g \to \infty$. (a) Largest Lyapunov exponent for $g \to g^+_{\text{crit}}$. Dots are the solution obtained from dynamic mean-field theory [1,11,59]. Full line is $\lambda_{(}g) = \frac{1}{2}(g-1)^2$ [1,59]. (b) Largest Lyapunov exponent for $g \to \infty$. Dots are solution obtained from dynamic mean-field theory, dash-dotted line is the explicit approximation $\lambda_{\max}(g) = C\log(g)$ with $C = \frac{2}{\pi}/\sqrt{\Delta_0(1-\Delta_0)}$ and $\Delta_0 = 2(1-2/\pi)$ [59], dotted line is $\lambda_{\max}(g) = -1 + \sqrt{1 + C^2(W(\frac{g}{d}))^2}$, where $W$ is the Lambert $W$ function and $d = 6/((4-\pi)\sqrt{\pi-2})$ [167]. (c) $\Delta_0$ obtained from dynamic mean-field theory, dashed line are explicit limits for small $g(\Delta_0 = g-1)$ and large $g(\ \Delta_0 = 2(1-2/\pi)g^2)$ [1,59]. (d) Autocorrelations $\Delta(\tau)$ for $g \in \{5,\ 10,\ 100\}$. For large $\tau$ and $g$ the autocorrelations decay exponentially with time constant $\sqrt{(\pi-3)/(\pi-2)}$) (dashed lines). [Other parameters: for (a)–(c) relative tolerance $= 10^{-9}$, for (d) rel. tol. $10^{-11}$.]

$$
\begin{aligned}
&= \frac{1}{N\tau\Delta t}\left[\left\langle \ln\left(\det((1-\Delta t)\mathbb{1})\right.\right.\right.\\
&\quad\left.\left.\left.\times \det\left(\mathbb{1} + \frac{\Delta t}{1-\Delta t}\mathbf{J}\phi'(\mathbf{h}(t_s))\right)\right)\right\rangle\right]\\
&= \frac{1}{N\tau\Delta t}\left[\left\langle \ln\left(\det((1-\Delta t)\mathbb{1})\right.\right.\right.\\
&\quad\left.\left.\left.\times \left(\mathbb{1} + \frac{\Delta t}{1-\Delta t}\mathrm{tr}(\mathbf{J}\phi'(\mathbf{h}(t_s)))\right)\right)\right\rangle\right] + \mathcal{O}((\Delta t)^2)\\
&= \frac{1}{\tau\Delta t}\ln(1-\Delta t) + \frac{1}{N\tau\Delta t}\\
&\quad\times\left[\left\langle \ln\left(1 + \frac{\Delta t}{1-\Delta t}\mathrm{tr}(\mathbf{J}\phi'(\mathbf{y}))\right)\right\rangle\right] + \mathcal{O}((\Delta t)^2)\\
&= \frac{1}{\tau\Delta t}\ln(1-\Delta t) + \mathcal{O}((\Delta t)^2)
\end{aligned}
$$

where $\mathbf{y}$ follows the distribution of Eq. (9), $\langle\dots\rangle$ denotes the time average, and $[\dots]$ denotes the ensemble average. For small $\Delta t$, we find excellent agreement with numerical simulations [see Fig. 8(b)]. In the limit $\Delta t \to 0$, the mean Lyapunov exponent becomes $-\frac{1}{\tau}$.

## APPENDIX E: MATHEMATICAL LINK OF GRADIENTS IN BPTT AND LYAPUNOV SPECTRUM

To train recurrent networks using backpropagation through time, one has to evaluate the gradient of the loss $E$ with respect to all weights of the recurrent network. This is done by unrolling the network dynamics in time [33],

$$\frac{\partial E_t}{\partial \mathbf{J}} = \frac{\partial E_t}{\partial \mathbf{o}_t}\frac{\partial \mathbf{o}_t}{\partial \mathbf{h}_t}\sum_\tau \frac{\partial \mathbf{h}_t}{\partial \mathbf{h}_\tau}\frac{\partial \mathbf{h}_\tau}{\partial \mathbf{J}} \tag{E1}$$

$$= \frac{\partial E_t}{\partial \mathbf{o}_t}\frac{\partial \mathbf{o}_t}{\partial \mathbf{h}_t}\sum_\tau \left(\prod_{\tau'=\tau}^{t-1}\frac{\partial \mathbf{h}_{\tau'+1}}{\partial \mathbf{h}_{\tau'}}\right)\frac{\partial \mathbf{h}_\tau}{\partial \mathbf{J}} \tag{E2}$$

$$= \frac{\partial E_t}{\partial \mathbf{o}_t}\frac{\partial \mathbf{o}_t}{\partial \mathbf{h}_t}\sum_\tau \left(\prod_{\tau'=\tau}^{t-1}\mathbf{J}\,\mathrm{diag}(\phi'(\mathbf{h}_{\tau'}))\right)\frac{\partial \mathbf{h}_\tau}{\partial \mathbf{J}} \tag{E3}$$

$$= \frac{\partial E_t}{\partial \mathbf{o}_t}\frac{\partial \mathbf{o}_t}{\partial \mathbf{h}_t}\sum_\tau \left(\prod_{\tau'=\tau}^{t-1}\mathbf{D}_{\tau'}\right)\frac{\partial \mathbf{h}_\tau}{\partial \mathbf{J}} \tag{E4}$$

$$= \frac{\partial E_t}{\partial \mathbf{o}_t}\frac{\partial \mathbf{o}_t}{\partial \mathbf{h}_t}\sum_\tau \mathbf{T}_t(\mathbf{h}_\tau)\frac{\partial \mathbf{h}_\tau}{\partial \mathbf{J}}, \tag{E5}$$

where $\mathbf{D}_{\tau'}$ is the Jacobian Eq. (2) that we already considered when calculating the Lyapunov spectrum. The recursive dependence of the gradient on the previous network state results in a product of Jacobians, which takes the form of the long-term Jacobian $\mathbf{T}_t(\mathbf{h})$ [Eq. (7)] whose inner product gives the Oseledets matrix [Eq. (6)].

The singular values of the long-term Jacobian $\mathbf{T}_t(\mathbf{h}_\tau)$, which determine how quickly gradients vanish or explode during backpropagation through time, are directly related to the Lyapunov exponents of the forward dynamics: The Lyapunov exponents of the forward dynamics are given by the

logarithm of the singular values of the long-term Jacobian [168]. Thus, our results on how the global coupling strength $g$, simulation parameters (e.g., time-discretization $\Delta t$), time-dependent input, and nonlinearity $\phi$ (e.g., threshold-linear vs tanh) shape the Lyapunov spectrum can directly be translated into predictions on the gradient instability during backpropagation through time. A positive (negative) first Lyapunov exponent would thus be expected to results in exploding (vanishing) gradients. But beyond that, also the full set of Lyapunov exponents is instructive for trainability. It was pointed out previously [35,36,38] that the trainability of recurrent networks is constrained by the condition number $\kappa$ of the long-term Jacobian $\mathbf{T}_t(\mathbf{h}_\tau)$. The condition number can be approximated by the Lyapunov spectrum: $\kappa_2(\mathbf{T}_t(\mathbf{h}_\tau)) = \frac{\sigma_1(\mathbf{T}_t(\mathbf{h}_\tau))}{\sigma_N(\mathbf{T}_t(\mathbf{h}_\tau))} \approx (t-\tau)\exp(\lambda_{\max} - \lambda_N)$, where $t-\tau$ is the time to be bridged by backpropagation through time. More generally, how well $k$ independent error signals $\frac{\partial E_t}{\partial \theta}$ can propagate though the tangent space is limited by the difference between the first and the $k$th Lyapunov exponent multiplied by the time difference $(t-\tau)\exp(\lambda_{\max} - \lambda_k)$. Or the other way round: to have a rank-$k$ update of the parameter space, you need to have at least $k$ Lyapunov exponents sufficiently close to 0.

## APPENDIX F: LARGE $G$ LIMIT

We note that the analytical argument for point symmetry in the Lyapunov spectrum around $i = N/2$ and $\lambda_i = -\frac{1}{\tau}$ together with the fact that $D$ is bounded by $N$ already imply that $H$ has to be bounded: $H\tau < D < N$, as

$$H\tau \leqslant \sum_{i=1}^{p} \lambda_i \tau = -\sum_{i=p+1}^{D} \lambda_i \tau \leqslant -\sum_{i=p+1}^{D} -\frac{1}{\tau} = D - p \leqslant D < N, \tag{F1}$$

where $p$ is the number of positive Lyapunov exponents and $H$ is measured in $nat/\tau$. The left-most less than equal to was proven by Ruelle [169], the next equal sign comes from the definition of the KY dimension [51,170–172], the next less than equal comes from the fact that the Lyapunov spectrum is point symmetric around $-\frac{1}{\tau}$, so for the first half, $-\frac{1}{\tau}$ is a lower bound on the Lyapunov exponents. The next less than equal sign comes from the fact that $p$ is a non-negative number; in the case of $p > 1$, it is an inequality. For notational simplicity, we assumed here an integer dimensionality $D$, but the argument holds generally.

## APPENDIX G: SUPPORTING INFORMATION

*S1 Code - Source code for Lyapunov spectrum of rate networks.* We provide all necessary code to calculate the full Lyapunov spectrum written in Julia [173]. The efficient implementation is parallelized using level-3 matrix-matrix operations from BLAS (Basic Linear Algebra Subprograms) called via LAPACK (Linear Algebra PACKage). The code also provides an alternative estimate of the largest Lyapunov exponents by tracking the evolution of a small but finite initial perturbation and resizing it iteratively [51]. Furthermore, the program provides bootstrapped 95 percentile confidence intervals for the first and the last Lyapunov exponent, the Kolmogorov-Sinai entropy rate, and the attractor dimensionality. Optionally, a principal component-based dimensionality estimate can also be calculated. Finally, the program provides the convergence of the Lyapunov spectrum in time. Input variables are network size $N$, coupling strength $g$, time-discretization $\Delta t$, simulation time $t_{\text{sim}}$, number of Lyapunov exponents to be calculated $nLE$, orthonormalization time interval $t_{\text{ONS}}$, seed for initial conditions $seed_{\text{IC}}$, seed for random network realization $seed_{\text{net}}$, seed for orthonormal system $seed_{\text{ONS}}$, and finally the subdirectory where the results are stored. Code written in MATLAB®/Octave/Python is available on github [174].

*S2 Code - Source code for Lyapunov spectrum of input-driven rate networks.* We also provide Julia code to obtain the full Lyapunov spectrum of a noise-driven rate network by a reorthonormalization procedure [60]. This is done along a numerical solution of the stochastic differential equation obtained with the Euler-Maruyama method [175]. The noise strength $\sigma$ is now an additional input parameter. Code written in MATLAB®/Octave/Python is available on github.


[1] H. Sompolinsky, A. Crisanti, and H. J. Sommers, Chaos in random neural networks, Phys. Rev. Lett. **61**, 259 (1988).

[2] J. Kadmon and H. Sompolinsky, Transition to chaos in random neuronal networks, Phys. Rev. X **5**, 041030 (2015).

[3] J. Aljadeff, M. Stern, and T. Sharpee, Transition to chaos in random networks with cell-type-specific connectivity, Phys. Rev. Lett. **114**, 088101 (2015).

[4] O. Harish and D. Hansel, Asynchronous rate chaos in spiking neuronal circuits, PLoS Comput. Biol. **11**, e1004266 (2015).

[5] M. Stern, H. Sompolinsky, and L. F. Abbott, Dynamics of random neural networks with bistable units, Phys. Rev. E **90**, 062710 (2014).

[6] S. P. Muscinelli, W. Gerstner, and T. Schwalger, How single neuron properties shape chaotic dynamics and signal transmission in random neural networks, PLoS Comput. Biol. **15**, e1007122 (2019).

[7] O. Harish, Network mechanisms of working memory: From persistent dynamics to chaos, Ph.D. thesis, University of Paris V (2013).

[8] L. Molgedey, J. Schuchhardt, and H. G. Schuster, Suppressing chaos in neural networks by noise, Phys. Rev. Lett. **69**, 3717 (1992).

[9] K. Rajan, L. Abbott, and H. Sompolinsky, Inferring stimulus selectivity from the spatial structure of neural network dynamics, in *Advances in Neural Information Processing Systems 23*, edited by J. D. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. S. Zemel, and A. Culotta (Curran Associates, 2010), pp. 1975–1983.

[10] K. Rajan, L. F. Abbott, and H. Sompolinsky, Stimulus-dependent suppression of chaos in recurrent neural networks, Phys. Rev. E **82**, 011903 (2010).

[11] J. Schuecker, S. Goedeke, and M. Helias, Optimal sequence memory in driven random networks, Phys. Rev. X **8**, 041029 (2018).

[12] O. Shriki, D. Hansel, and H. Sompolinsky, Rate models for conductance-based cortical neuronal networks, Neural Comput. **15**, 1809 (2003).

[13] W. Maass, T. Natschläger, and H. Markram, Real-time computing without stable states: A new framework for neural computation based on perturbations, Neural Comput. **14**, 2531 (2002).

[14] H. Jaeger and H. Haas, Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication, Science **304**, 78 (2004).

[15] D. Sussillo and L. F. Abbott, Generating coherent patterns of activity from chaotic neural networks, Neuron **63**, 544 (2009).

[16] R. Laje and D. V. Buonomano, Robust timing and motor patterns by taming chaos in recurrent neural networks, Nat. Neurosci. **16**, 925 (2013).

[17] L. F. Abbott, B. DePasquale, and R.-M. Memmesheimer, Building functional networks of spiking model neurons, Nat. Neurosci. **19**, 350 (2016).

[18] B. DePasquale, D. Sussillo, L. F. Abbott, and M. M. Churchland, The centrality of population-level factors to network computation is demonstrated by a versatile approach for training spiking networks, Neuron **111**, 631 (2023).

[19] D. Thalmeier, M. Uhlmann, H. J. Kappen, and R.-M. Memmesheimer, Learning universal computations with spikes, PLoS Comput. Biol. **12**, e1004895 (2016).

[20] W. Nicola and C. Clopath, Supervised learning in spiking neural networks with FORCE training, Nat. Commun. **8**, 2208 (2017).

[21] A. Ingrosso and L. F. Abbott, Training dynamically balanced excitatory-inhibitory networks, PLoS One **14**, e0220547 (2019).

[22] N. Bertschinger and T. Natschläger, Real-time computation at the edge of chaos in recurrent neural networks, Neural Comput. **16**, 1413 (2004).

[23] N. Schweighofer, K. Doya, H. Fukai, J. V. Chiron, T. Furukawa, and M. Kawato, Chaos may enhance information transmission in the inferior olive, Proc. Natl. Acad. Sci. USA **101**, 4655 (2004).

[24] R. Legenstein and W. Maass, Edge of chaos and prediction of computational performance for neural circuit models, Neural Netw. **20**, 323 (2007).

[25] L. Büsing, B. Schrauwen, and R. Legenstein, Connectivity, dynamics, and memory in reservoir computing with binary and analog neurons, Neural Comput. **22**, 1272 (2010).

[26] T. Toyoizumi and L. F. Abbott, Beyond the edge of chaos: Amplification and temporal integration by recurrent networks in the chaotic regime, Phys. Rev. E **84**, 051908 (2011).

[27] D. Dahmen, S. Grün, M. Diesmann, and M. Helias, Second type of criticality in the brain uncovers rich multiple-neuron dynamics, Proc. Natl. Acad. Sci. USA **116**, 13051 (2019).

[28] J. Kelso, A. Mandell, and M. Shlesinger, Dynamic patterns in complex systems, *Dynamic Patterns in Complex Systems* (World Scientific, Singapore, 1988), pp. 1–432.

[29] C. G. Langton, Computation at the edge of chaos: Phase transitions and emergent computation, Physica D **42**, 12 (1990).

[30] M. Mitchell, P. Hraber, and J. Crutchfield, Revisiting the edge of chaos: Evolving cellular automata to perform computations, Complex Syst. **7**, 89 (1993).

[31] A. M. Saxe, J. L. McClelland, and S. Ganguli, Exact solutions to the nonlinear dynamics of learning in deep linear neural networks, arXiv:1312.6120.

[32] B. Poole, S. Lahiri, M. Raghu, J. Sohl-Dickstein, and S. Ganguli, Exponential expressivity in deep neural networks through transient chaos, arXiv:1606.05340.

[33] R. Pascanu, T. Mikolov, and Y. Bengio, On the difficulty of training recurrent neural networks, arXiv:1211.5063.

[34] B. Hanin and M. Nica, Products of many large random matrices and gradients in deep neural networks, Commun. Math. Phys. **376**, 287 (2020).

[35] M. Chen, J. Pennington, and S. S. Schoenholz, Dynamical isometry and a mean field theory of RNNs: Gating enables signal propagation in recurrent neural networks, arXiv:1806.05394.

[36] D. Gilboa, B. Chang, M. Chen, G. Yang, S. S. Schoenholz, E. H. Chi, and J. Pennington, Dynamical isometry and a mean field theory of LSTMs and GRUs, arXiv:1901.08987.

[37] T. Can, K. Krishnamurthy, and D. J. Schwab, Gating creates slow modes and controls phase-space complexity in GRUs and LSTMs, arXiv:2002.00025.

[38] S. S. Schoenholz, J. Gilmer, S. Ganguli, and J. Sohl-Dickstein, Deep information propagation, in *5th International Conference on Learning Representations*, 2016.

[39] J. Pennington, S. S. Schoenholz, and S. Ganguli, The emergence of spectral universality in deep networks, arXiv:1802.09979.

[40] A. Vulpiani, F. Cecconi, and M. Cencini, *Chaos: From Simple Models to Complex Systems* (World Scientific, Hackensack, NJ, 2009).

[41] R. Shaw, Strange attractors, chaotic behavior, and information flow, Z. Naturforsch. Teil A **36**, 80 (1981).

[42] L.-S. Young, Mathematical theory of Lyapunov exponents, J. Phys. A: Math. Theor. **46**, 254001 (2013).

[43] P. Ashwin and M. Timme, Nonlinear dynamics: When instability makes sense, Nature (London) **436**, 36 (2005).

[44] M. Rabinovich, A. Volkovskii, P. Lecanda, R. Huerta, H. D. I. Abarbanel, and G. Laurent, Dynamical encoding by networks of competing neuron groups: Winnerless competition, Phys. Rev. Lett. **87**, 068102 (2001).

[45] M. Rabinovich, R. Huerta, and G. Laurent, Transient dynamics for neural processing, Science **321**, 48 (2008).

[46] G. Lajoie, K. K. Lin, and E. Shea-Brown, Chaos and reliability in balanced spiking networks with temporal drive, Phys. Rev. E **87**, 052901 (2013).

[47] A. N. Kolmogorov, A new metric invariant of transient dynamical systems and automorphisms in Lebesgue spaces, Dokl. Akad. Nauk. SSSR **119**, 861 (1958).

[48] Y. G. Sinai, On the notion of entropy of a dynamical system, Dokl. Akad. Nauk. SSSR, **124**, 768 (1959).

[49] P. Billingsley, *Ergodic Theory and Information* (Wiley, Hoboken, NJ, 1965).

[50] N. V. Kuznetsov, T. A. Alexeeva, and G. A. Leonov, Invariance of Lyapunov exponents and Lyapunov dimension for regular and irregular linearizations, Nonlinear Dyn. **85**, 195 (2016).

[51] J. P. Eckmann and D. Ruelle, Ergodic theory of chaos and strange attractors, Rev. Mod. Phys. **57**, 617 (1985).

[52] P. Grassberger and I. Procaccia, Characterization of strange attractors, Phys. Rev. Lett. **50**, 346 (1983).

[53] P. Grassberger and I. Procaccia, Estimation of the Kolmogorov entropy from a chaotic signal, Phys. Rev. A **28**, 2591 (1983).

[54] P. Grassberger, Generalized dimensions of strange attractors, Phys. Lett. A **97**, 227 (1983).

[55] J. P. Eckmann and D. Ruelle, Fundamental limitations for estimating dimensions and Lyapunov exponents in dynamical systems, Physica D **56**, 185 (1992).

[56] L. A. Smith, Intrinsic limits on dimension calculations, Phys. Lett. A **133**, 283 (1988).

[57] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*, 2nd ed. (Cambridge University Press, Cambridge, 2004).

[58] A. Pikovsky and A. Politi, *Lyapunov Exponents: A Tool to Explore Complex Dynamics* (Cambridge University Press, Cambridge, 2016).

[59] A. Crisanti and H. Sompolinsky, Path integral approach to random neural networks, Phys. Rev. E **98**, 062120 (2018).

[60] G. Benettin, L. Galgani, A. Giorgilli, and J.-M. Strelcyn, Lyapunov Characteristic Exponents for smooth dynamical systems and for hamiltonian systems; A method for computing all of them. Part 2: Numerical application, Meccanica **15**, 21 (1980).

[61] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevResearch.5.043044 for the mathematical foundations of Lyapunov spectra (II), and our concrete implementation for rate networks (III). Further, we give an introduction to Kolmogorov-Sinai entropy rate and Kaplan Yorke attractor dimensionality (IV). We extend the approach to random dynamical systems and discuss the implementation of Lyapunov spectra for non-autonomous networks with time-dependent input (V). Then we give details about the PCA-based dimensionality estimate (VI).

[62] D. Ruelle, Large volume limit of the distribution of characteristic exponents in turbulence, Commun. Math. Phys. **87**, 287 (1982).

[63] K. A. Takeuchi, F. Ginelli, and H. Chaté, Lyapunov analysis captures the collective dynamics of large chaotic systems, Phys. Rev. Lett. **103**, 154103 (2009).

[64] P. Gao and S. Ganguli, On simplicity and complexity in the brave new world of large-scale neuroscience, Curr. Opin. Neurobiol. **32**, 148 (2015).

[65] J. P. Cunningham and B. M. Yu, Dimensionality reduction for large-scale neural recordings, Nat. Neurosci. **17**, 1500 (2014).

[66] A. Litwin-Kumar, K. D. Harris, R. Axel, H. Sompolinsky, and L. F. Abbott, Optimal degrees of synaptic connectivity, Neuron **93**, 1153 (2017).

[67] M. S. Farrell, S. Recanatesi, G. Lajoie, and E. Shea-Brown, Dynamic compression and expansion in a classifying recurrent network, bioRxiv, 564476 (2019).

[68] S. Recanatesi, G. K. Ocker, M. A. Buice, and E. Shea-Brown, Dimensionality in recurrent spiking networks: Global trends in activity and local origins in connectivity, PLoS Comput. Biol. **15**, e1006446 (2019).

[69] D. G. Clark, L. F. Abbott, and A. Litwin-Kumar, Dimension of activity in random neural networks, Phys. Rev. Lett. **131**, 118401 (2023).

[70] S. I. Amari, Learning patterns and pattern sequences by self-organizing nets of threshold elements, IEEE Trans. Comput. **C-21**, 1197 (1972).

[71] G. Parisi, Asymmetric neural networks and the process of learning, J. Phys. A: Math. Gen. **19**, L675 (1986).

[72] B. Doyon, B. Cessac, M. Quoy, and M. Samuelides, Control of the transition to chaos in neural networks with random connectivity, Int. J. Bifurcation Chaos **03**, 279 (1993).

[73] M. Massar and S. Massar, Mean-field theory of echo state networks, Phys. Rev. E **87**, 042809 (2013).

[74] C. M. Bishop, *Pattern Recognition and Machine Learning*, 1st ed. (Springer, New York, 2007).

[75] V. I. Oseledets, A multiplicative ergodic theorem. Characteristic Ljapunov, exponents of dynamical systems, Trudy Moskovskogo Matematicheskogo Obshchestva **19**, 179 (1968).

[76] A. Crisanti, *Products of Random Matrices: In Statistical Physics* (Springer-Verlag, Berlin, 1993).

[77] E. P. Wigner, On the distribution of the roots of certain symmetric matrices, Ann. Math. **65**, 203 (1957).

[78] F. Götze and A. Tikhomirov, On the asymptotic spectrum of products of independent random matrices, arXiv:1012.2710.

[79] T. Tao, V. Vu, and M. Krishnapur, Random matrices: Universality of ESDs and the circular law, Ann. Probab. **38**, 2023 (2010).

[80] A. Crisanti and H. Sompolinsky, Dynamics of spin systems with randomly asymmetric bonds: Ising spins and Glauber dynamics, Phys. Rev. A **37**, 4865 (1988).

[81] C. M. Newman, The distribution of Lyapanov exponents: Exact results for random matrices, Commun. Math. Phys. **103**, 121 (1986).

[82] M. Isopi and C. M. Newman, The triangle law for Lyapunov exponents of large random matrices, Commun. Math. Phys. **143**, 591 (1992).

[83] M. Bauer and W. Martienssen, Lyapunov exponents and dimensions of chaotic neural networks, J. Phys. A: Math. Gen. **24**, 4557 (1991).

[84] G. Curato and A. Politi, Onset of chaotic dynamics in neural networks, Phys. Rev. E **88**, 042908 (2013).

[85] D. A. McCormick, B. W. Connors, J. W. Lighthall, and D. A. Prince, Comparative electrophysiology of pyramidal and sparsely spiny stellate neurons of the neocortex, J. Neurophysiol. **54**, 782 (1985).

[86] A. L. Barth and J. F. A. Poulet, Experimental evidence for sparse firing in the neocortex, Trends Neurosci. **35**, 345 (2012).

[87] C. van Vreeswijk and H. Sompolinsky, Chaos in neuronal networks with balanced excitatory and inhibitory activity, Science **274**, 1724 (1996).

[88] C. van Vreeswijk and H. Sompolinsky, Chaotic balanced state in a model of cortical circuits, Neural Comput. **10**, 1321 (1998).

[89] C. van Vreeswijk and H. Sompolinsky, Course 9 - Irregular activity in large networks of neurons, in *Les Houches, Methods and Models in Neurophysics*, edited by C. Chow, B. Gutkin, D. Hansel, C. Meunier, and J. Dalibard (Elsevier, Amsterdam, 2005), Vol. 80, pp. 341–406.

[90] N. Brunel, Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons, J. Comput. Neurosci. **8**, 183 (2000).

[91] M. Monteforte and F. Wolf, Dynamical entropy production in spiking neuron networks in the balanced state, Phys. Rev. Lett. **105**, 268104 (2010).

[92] F. Mastrogiuseppe and S. Ostojic, Intrinsically-generated fluctuating activity in excitatory-inhibitory networks, PLoS Comput. Biol. **13**, e1005498 (2017).

[93] X. Glorot, A. Bordes, and Y. Bengio, Deep sparse rectifier neural networks, J. Mach. Learn. Res. **15**, 315 (2011).

[94] A. L. Maas, A. Y. Hannun, and A. Y. Ng, Rectifier nonlinearities improve neural network acoustic models, in *in ICML Workshop on Deep Learning for Audio, Speech and Language Processing* (2013).

[95] D. Hansel and C. v. Vreeswijk, How noise contributes to contrast invariance of orientation tuning in cat visual cortex, J. Neurosci. **22**, 5118 (2002).

[96] N. J. Priebe, F. Mechler, M. Carandini, and D. Ferster, The contribution of spike threshold to the dichotomy of cortical simple and complex cells, Nat. Neurosci. **7**, 1113 (2004).

[97] N. J. Priebe and D. Ferster, Direction selectivity of excitation and inhibition in simple cells of the cat primary visual cortex, Neuron **45**, 133 (2005).

[98] N. J. Priebe and D. Ferster, Mechanisms underlying cross-orientation suppression in cat visual cortex, Nat. Neurosci. **9**, 552 (2006).

[99] I. M. Finn, N. J. Priebe, and D. Ferster, The emergence of contrast-invariant orientation tuning in simple cells of cat visual cortex, Neuron **54**, 137 (2007).

[100] U. Dressler, Symmetry property of the Lyapunov spectra of a class of dissipative dynamical systems with viscous damping, Phys. Rev. A **38**, 2103 (1988).

[101] D. J. Evans, E. G. D. Cohen, and G. P. Morriss, Viscosity of a simple fluid from its maximal Lyapunov exponents, Phys. Rev. A **42**, 5990 (1990).

[102] C. P. Dettmann and G. P. Morriss, Hamiltonian reformulation and pairing of Lyapunov exponents for Nosé-Hoover dynamics, Phys. Rev. E **55**, 3693 (1997).

[103] M. P. Wojtkowski and C. Liverani, Conformally symplectic dynamics and symmetry of the Lyapunov spectrum, Commun. Math. Phys. **194**, 47 (1998).

[104] S. Olmi, Chimera states in coupled Kuramoto oscillators with inertia, Chaos **25**, 123125 (2015).

[105] F. Wegner, Inverse participation ratio in $2+\varepsilon$ dimensions, Z. Phys. B **36**, 209 (1980).

[106] K. Kaneko, Lyapunov analysis and information flow in coupled map lattices, Physica D **23**, 436 (1986).

[107] M. C. Cross and P. C. Hohenberg, Pattern formation outside of equilibrium, Rev. Mod. Phys. **65**, 851 (1993).

[108] K. A. Takeuchi and H. Chaté, Collective Lyapunov modes, J. Phys. A: Math. Theor. **46**, 254007 (2013).

[109] K. Rajan and L. F. Abbott, Eigenvalue spectra of random matrices for neural networks, Phys. Rev. Lett. **97**, 188104 (2006).

[110] L. Arnold, Random dynamical systems, in *Dynamical Systems: Lectures Given at the 2nd Session of the Centro Internazionale Matematico Estivo (C.I.M.E.) held in Montecatini Terme, Italy, June 13–22, 1994*, Lecture Notes in Mathematics, edited by L. Arnold, C. K. R. T. Jones, K. Mischaikow, G. Raugel, and R. Johnson (Springer, Berlin, 1995), pp. 1–43.

[111] K. K. Lin, Stimulus-Response Reliability of Biological Networks, in *Nonautonomous Dynamical Systems in the Life Sciences*, Lecture Notes in Mathematics No. 2102, edited by P. E. Kloeden and C. Pötzsche (Springer International, New York, 2013), pp. 135–161.

[112] Y. Kifer, *Ergodic Theory of Random Transformations* (Springer Science & Business Media, New York, 2012).

[113] K. Matsumoto and I. Tsuda, Noise-induced order, J. Stat. Phys. **31**, 87 (1983).

[114] P. H. Baxendale, Stability and Equilibrium properties of stochastic flows of diffeomorphisms, in *Diffusion Processes and Related Problems in Analysis, Volume II*, Progress in Probability, edited by M. A. Pinsky and V. Wihstutz (Birkhäuser, Boston, 1992), Vol. 27, pp. 3–35.

[115] Y. Le Jan, Equilibre statistique pour les produits de difféomorphismes aléatoires indépendants, Ann. Inst. Henri Poincare, Probab. Stat. **23**, 111 (1987).

[116] L.-S. Young, Generalizations of SRB measures to nonautonomous, random, and infinite dimensional systems, J. Stat. Phys. **166**, 494 (2017).

[117] A. Ben-Mizrachi, I. Procaccia, and P. Grassberger, Characterization of experimental (noisy) strange attractors, Phys. Rev. A **29**, 975 (1984).

[118] E. Ott, E. D. Yorke, and J. A. Yorke, A scaling law: How an attractor's volume depends on noise level, Physica D **16**, 62 (1985).

[119] A. Ostruszka, P. Pakoński, W. Słomczyński, and K. Życzkowski, Dynamical entropy for systems with stochastic perturbation, Phys. Rev. E **62**, 2018 (2000).

[120] B. DePasquale, C. J. Cueva, K. Rajan, G. S. Escola, and L. F. Abbott, full-FORCE: A target-based method for training recurrent networks, PLOS ONE **13**, e0191527 (2018).

[121] Y. Bengio, P. Simard, and P. Frasconi, Learning long-term dependencies with gradient descent is difficult, IEEE Trans. Neural Netw. **5**, 157 (1994).

[122] M. Monteforte and F. Wolf, Dynamic flux tubes form reservoirs of stability in neuronal circuits, Phys. Rev. X **2**, 041007 (2012).

[123] S. Hochreiter and Jürgen Schmidhuber, Long short-term memory, Neural Comput. **9**, 1735 (1997).

[124] M. Puelma Touzel, Cellular dynamics and stable chaos in balanced networks, Ph.D. thesis, Georg-August-University Göttingen, 2016.

[125] P. WERBOS, Beyond regression, Ph. D. dissertation, Harvard University, Cambridge, 1974.

[126] D. B. Parker, Learning-Logic: Casting the Cortex of the Human Brain in Silicon (Center for Computational Research in Economics and Management Science, Alfred P. Sloan School of Management, Massachusetts Institute of Technology, 1985).

[127] Y. LECUN, Une procedure d'apprentissage ponr reseau a seuil asymetrique, Proc. Cognitiva 85, 599 (1985).

[128] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, Learning representations by back-propagating errors, Nature (London) **323**, 533 (1986).

[129] S. Hochreiter, Untersuchungen zu dynamischen neuronalen Netzen, Ph.D. thesis, Technische Universität München, Germany, 1991.

[130] D. Sussillo and O. Barak, Opening the black box: Low-dimensional dynamics in high-dimensional recurrent neural networks, Neural Comput. **25**, 626 (2013).

[131] A. Rivkind and O. Barak, Local dynamics in trained recurrent neural networks, Phys. Rev. Lett. **118**, 258101 (2017).
[132] C. Beer and O. Barak, One step back, two steps forward: Interference and learning in recurrent neural networks, Neural Computation **31**, 1985 (2019).
[133] D. Haviv, A. Rivkind, and O. Barak, Understanding and controlling memory in recurrent neural networks, arXiv:1902.07275.
[134] S. Ostojic, Two types of asynchronous activity in networks of excitatory and inhibitory spiking neurons, Nat. Neurosci. **17**, 594 (2014).
[135] G. Wainrib and J. Touboul, Topological and dynamical complexity of random neural networks, Phys. Rev. Lett. **110**, 118101 (2013).
[136] L. C. García del Molino, K. Pakdaman, J. Touboul, and G. Wainrib, Synchronization in random balanced networks, Phys. Rev. E **88**, 042824 (2013).
[137] T. Cabana and J. Touboul, Large deviations, dynamics and phase transitions in large stochastic and disordered neural networks, J. Stat. Phys. **153**, 211 (2013).
[138] R. Engelken, F. Farkhooi, D. Hansel, C. van Vreeswijk, and F. Wolf, A reanalysis of "Two types of asynchronous activity in networks of excitatory and inhibitory spiking neurons", F1000Research **5**, 2043 (2016).
[139] S. Ganguli and H. Sompolinsky, Compressed sensing, sparsity, and dimensionality in neuronal information processing and data analysis, Annu. Rev. Neurosci. **35**, 485 (2012).
[140] R. Monasson and D. Villamaina, Estimating the principal components of correlation matrices from all their empirical eigenvectors, Europhys. Lett. **112**, 50001 (2015).
[141] C. Stringer, M. Pachitariu, N. Steinmetz, M. Carandini, and K. D. Harris, High-dimensional geometry of population responses in visual cortex, Nature (London) **571**, 361 (2019).
[142] S. Y. Chung, D. D. Lee, and H. Sompolinsky, Classification and geometry of general perceptual manifolds, Phys. Rev. X **8**, 031003 (2018).
[143] M. Rigotti, O. Barak, M. R. Warden, X.-J. Wang, N. D. Daw, E. K. Miller, and S. Fusi, The importance of mixed selectivity in complex cognitive tasks, Nature (London) **497**, 585 (2013).
[144] B. Cessac, B. Doyon, M. Quoy, and M. Samuelides, Mean-field equations, bifurcation map and route to chaos in discrete time neural networks, Physica D **74**, 24 (1994).
[145] J. Aljadeff, D. Renfrew, M. Vegué, and T. O. Sharpee, Low-dimensional dynamics of structured random networks, Phys. Rev. E **93**, 022302 (2016).
[146] M. Monteforte, Chaotic Dynamics in Networks of Spiking Neurons in the Balanced State, Ph.D. thesis, Georg-August-University, Göttingen, 2011.
[147] S. Luccioli, S. Olmi, A. Politi, and A. Torcini, Collective dynamics in sparse networks, Phys. Rev. Lett. **109**, 138103 (2012).
[148] G. Lajoie, J.-P. Thivierge, and E. Shea-Brown, Structured chaos shapes spike-response noise entropy in balanced neural networks, Front. Comput. Neurosci. **8**, 123 (2014).
[149] O. Barak, D. Sussillo, R. Romo, M. Tsodyks, and L. F. Abbott, From fixed points to chaos: Three models of delayed discrimination, Prog. Neurobiol. **103**, 214 (2013).
[150] J. Pathak, B. Hunt, M. Girvan, Z. Lu, and E. Ott, Model-free prediction of large spatiotemporally chaotic systems from data: A reservoir computing approach, Phys. Rev. Lett. **120**, 024102 (2018).
[151] Z. Lu, B. R. Hunt, and E. Ott, Attractor reconstruction by machine learning, Chaos **28**, 061104 (2018).
[152] J. Pathak, Z. Lu, B. R. Hunt, M. Girvan, and E. Ott, Using machine learning to replicate chaotic attractors and calculate Lyapunov exponents from data, Chaos **27**, 121102 (2017).
[153] P. R. Vlachas, J. Pathak, B. R. Hunt, T. P. Sapsis, M. Girvan, E. Ott, and P. Koumoutsakos, Backpropagation algorithms and Reservoir Computing in Recurrent Neural Networks for the forecasting of complex spatiotemporal dynamics, Neural Netw. **126**, 191 (2020).
[154] H. Jaeger, The "echo state" approach to analysing and training recurrent neural networks-with an erratum note, Bonn, Germany: German National Research Center for Information Technology GMD Technical Report 148, 34 (2001).
[155] M. Buehner and P. Young, A tighter bound for the echo state property, IEEE Trans. Neural Netw. **17**, 820 (2006).
[156] H. Jaeger, Echo state network, Scholarpedia **2**, 2330 (2007).
[157] G. Manjunath and H. Jaeger, Echo state property linked to an Iinput: Exploring a fundamental characteristic of recurrent neural networks, Neural Comput. **25**, 671 (2013).
[158] L. Susman, F. Mastrogiuseppe, N. Brenner, and O. Barak, Quality of internal representation shapes learning performance in feedback neural networks, Phys. Rev. Res. **3**, 013176 (2021).
[159] R. Engelken, A. Ingrosso, R. Khajeh, S. Goedeke, and L. F. Abbott, Input correlations impede suppression of chaos and learning in balanced firing-rate networks, PLoS Comput. Biol. **18**, e1010590 (2022).
[160] K. K. Lin, E. Shea-Brown, and L.-S. Young, Reliability of coupled oscillators, J. Nonlinear Sci. **19**, 497 (2009).
[161] G. Casati and J. Ford, Stochastic Behavior in Classical and Quantum Hamiltonian Systems, *Lecture Notes in Physics*, Vol. 93 (Springer, New York, 1979).
[162] K. K. Lin and L.-S. Young, Shear-induced chaos, Nonlinearity **21**, 899 (2008).
[163] K. Krishnamurthy, T. Can, and D. J. Schwab, Theory of gating in recurrent neural networks, Phys. Rev. X **12**, 011011 (2022).
[164] H. J. Sommers, A. Crisanti, H. Sompolinsky, and Y. Stein, Spectrum of large random asymmetric matrices, Phys. Rev. Lett. **60**, 1895 (1988).
[165] L. Zhao, B. B. II, T. Netoff, and D. Q. Nykamp, Synchronization from second order network connectivity statistics, Front. Comput. Neurosci. **5**, 28 (2011).
[166] S. V. Ershov and A. B. Potapov, On the concept of stationary Lyapunov basis, Physica D **118**, 167 (1998).
[167] C. A. Rozzi, D. Varsano, A. Marini, E. K. U. Gross, and A. Rubio, Exact Coulomb cutoff technique for supercell calculations, Phys. Rev. B **73**, 205119 (2006).
[168] K. Geist, U. Parlitz, and W. Lauterborn, Comparison of different methods for computing Lyapunov exponents, Prog. Theor. Phys. **83**, 875 (1990).

[169] D. Ruelle, An inequality for the entropy of differentiable maps, Bol. Soc. Bras. de Mat (1978).

[170] J. L. Kaplan and J. A. Yorke, Preturbulence: A regime observed in a fluid flow model of Lorenz, Commun. Math. Phys. **67**, 93 (1979).

[171] P. Frederickson, J. L. Kaplan, E. D. Yorke, and J. A. Yorke, The Liapunov dimension of strange attractors, J. Diff. Equ. **49**, 185 (1983).

[172] J. C. Alexander and J. A. Yorke, Fat Baker's transformations, Ergod. Theory Dyna. Syst. **4**, 1 (1984).

[173] J. Bezanson, A. Edelman, S. Karpinski, and V. Shah, Julia: A fresh approach to numerical computing, SIAM Rev. **59**, 65 (2017).

[174] https://github.com/RainerEngelken/RNN-LyapunovSpectra.

[175] P. E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations* (Springer, Berlin, 1992).