Featured in Physics

Theory of Coupled Neuronal-Synaptic Dynamics

David G. Clark [*] and L. F. Abbott [†]

Zuckerman Institute, Department of Neuroscience, Columbia University, New York, New York 10027, USA

In neural circuits, synaptic strengths influence neuronal activity by shaping network dynamics, and neuronal activity influences synaptic strengths through activity-dependent plasticity. Motivated by this fact, we study a recurrent-network model in which neuronal units and synaptic couplings are interacting dynamic variables, with couplings subject to Hebbian modification with decay around quenched random strengths. Rather than assigning a specific role to the plasticity, we use dynamical mean-field theory and other techniques to systematically characterize the neuronal-synaptic dynamics, revealing a rich phase diagram. Adding Hebbian plasticity slows activity in already chaotic networks and can induce chaos in otherwise quiescent networks. Anti-Hebbian plasticity quickens activity and produces an oscillatory component. Analysis of the Jacobian shows that Hebbian and anti-Hebbian plasticity push locally unstable modes toward the real and imaginary axes, respectively, explaining these behaviors. Both random-matrix and Lyapunov analysis show that strong Hebbian plasticity segregates network timescales into two bands, with a slow, synapse-dominated band driving the dynamics, suggesting a flipped view of the network as synapses connected by neurons. For increasing strength, Hebbian plasticity initially raises the complexity of the dynamics, measured by the maximum Lyapunov exponent and attractor dimension, but then decreases these metrics, likely due to the proliferation of stable fixed points. We compute the marginally stable spectra of such fixed points as well as their number, showing exponential growth with network size. Finally, in chaotic states with strong Hebbian plasticity, a stable fixed point of neuronal dynamics is destabilized by synaptic dynamics, allowing any neuronal state to be stored as a stable fixed point by halting the plasticity. This phase of freezable chaos offers a new mechanism for working memory.

I. INTRODUCTION

Computations in neural circuits are commonly thought to be implemented through the coordinated dynamics of neurons [1–3]. Under this view, the role of synaptic connectivity is to sculpt neuronal dynamics to implement computations. In actuality, synapses undergo plasticity over diverse timescales in response to neuronal activity and thus constitute dynamic degrees of freedom in their own right [4]. A more accurate picture of computation in neural circuits should involve the coupled dynamics of neurons and synapses. Indeed, it is possible that a network is better described by the states of its synapses than of its neurons [5]. Here, we study the consequences of treating neurons and synapses as mutually coupled dynamic variables on equal footing.

Synaptic dynamics are often divided into short-term plasticity, which operates on short timescales and depends on presynaptic activity [6–11], and long-term plasticity, which acts on much longer timescales and depends on both pre- and postsynaptic activity [12–15]. However, short-term forms of Hebbian plasticity exist, suggesting that the timescale distinction is little more than a convention [16–24] (see Ref. [25] for a review). Hebbian plasticity is more powerful than the presynaptic variety due to its ability to create attractor states of neuronal dynamics, the basis of Hopfield networks [26]. We are therefore motivated to introduce ongoing Hebbian plasticity in a recurrent network, without necessarily imposing a separation of timescales between neuronal and synaptic dynamics. This has unexpected, computationally useful consequences, a key example being *freezable chaos*, a phase in which a stable fixed point of neuronal dynamics is destabilized through synaptic dynamics. By contrast, introducing presynaptic plasticity to this model simply adds an effective constant input to each neuron (Appendix A).

[*]dgc2138@cumc.columbia.edu, he/him/his

[†]lfa2103@columbia.edu

Our work thus provides a theoretical impetus for further experimental investigation of ongoing Hebbian plasticity mechanisms.

The view that synapses serve solely to sculpt neuronal dynamics is mirrored in machine learning. In artificial neural networks, weights are trained via gradient descent, then fixed. However, allowing weights to be modulated by the activity of the units has been shown to confer computational advantages [27–31], particularly in tasks requiring short-term memory storage [32]. For example, Ba *et al.* [33] showed that recurrent networks benefit from a combination of "slow weights" trained via backpropagation and "fast weights" that undergo activity-dependent updates (the model we study is essentially the continuous-time counterpart of this proposal). In practice, recurrent networks have been superseded by transformers [34]. While these models were not neuroscientifically motivated, Schlag *et al.* [35] showed that linearized transformers [36] are equivalent to fast weight programmers [37], a recurrent network with activity-dependent weight updates [38].

A major impediment to studying neuronal-synaptic dynamics, in both neuroscience and machine learning, is that the analytical methods developed for nonplastic networks often do not translate to plastic networks, particularly when the neuronal and synaptic timescales are not well separated. For example, a common simplification is to study nonplastic networks with linear neuronal dynamics; however, such networks become highly nonlinear when synaptic plasticity is introduced [39]. Moreover, synaptic degrees of freedom increase the dimension of phase space from $\mathcal{O}(N)$ to $\mathcal{O}(N^2)$. In studying and training nonplastic networks, the theory of random networks has played a crucial role [40]. In seminal work, Sompolinsky *et al.* [41] showed that random recurrent networks exhibit a phase transition to high-dimensional chaotic activity at a critical coupling variance [42]. While this analysis was in firing-rate (nonspiking) networks with fully unstructured couplings, key phenomena, such as the transition to chaos, generalize to spiking networks with more realistic distributions over couplings [43]. Here, we extend this approach to plastic networks—that is, we develop a theory for such a model in the thermodynamic limit, compute its phase diagram, and characterize its dynamics—thereby providing a foundation for understanding how coupled neuronal-synaptic dynamics could underlie computation.

II. MODEL

We augment the random-network model of Sompolinsky *et al.* [41] with dynamic couplings. There are N neuronal units with preactivations $x_i(t)$ and activations $\phi_i(t) = \phi(x_i(t))$, where $\phi(\cdot)$ is a nonlinearity. Throughout, we use $\phi(\cdot) = \tanh(\cdot)$. Neurons interact through all-to-all time-dependent couplings $W_{ij}(t)$ according to the neuronal dynamics:

$$(1 + \partial_t)x_i(t) = \sum_j W_{ij}(t)\phi_j(t). \tag{1}$$

Concurrently, $W_{ij}(t)$ displays synaptic dynamics. We first express these couplings as a sum of quenched and fluctuating terms,

$$W_{ij}(t) = J_{ij} + A_{ij}(t), \tag{2}$$

where $J_{ij} \sim \mathcal{N}(0, g^2/N)$ provides quenched disorder. The fluctuating term $A_{ij}(t)$ follows a local plasticity rule,

$$(1 + p\partial_t)A_{ij}(t) = \frac{k}{N}\phi_i(t)\phi_j(t), \tag{3}$$

where k is the sign and strength of the plasticity, which is Hebbian or anti-Hebbian for $k > 0$ or $k < 0$, respectively, and p is the synaptic decay timescale in units in which the neuronal decay timescale is unity. We do not require $p \gg 1$, though a reasonable constraint from biology is $p > 1$ since the synaptic timescale is unlikely to be shorter than the neuronal timescale. The couplings $W(t)$ include self-connections (on diagonals) with the same dynamics as non-self-connections (off diagonals). However, the effect of these self-connections on each neuron is $\sim 1/\sqrt{N}$, and thus is negligible as $N \to \infty$. The full set of dynamic variables comprises the N neurons $x_i(t)$ and N^2 fluctuating couplings $A_{ij}(t)$. We study their collective dynamics as a function of g, k, and p as $N \to \infty$ [Fig. 1(a)].

Because $A(t)$ is an average over outer products for each order-one time step and decays with timescale p, it has rank of order p or smaller. Given that p is order one (not order N), the approximate rank of $A(t)$ is intensive. Note that $J_{ij} \sim 1/\sqrt{N}$ and $A_{ij}(t) \sim 1/N$, so plasticity is vanishingly weak at the single-synapse level as $N \to \infty$. Nevertheless, because $A(t)$ is approximately intensive rank, the random and fluctuating couplings both have order-one macroscopic effects. This is because when neuronal activity exhibits alignment to the states encoded in $A(t)$, the J and $A(t)$ terms each contribute an order-one input to neurons, as made clear by the mean-field analysis in Sec. IV. Further intuition for this scaling can be obtained from the spectra of J and $A(t)$; the chosen scaling implies that the eigenvalues of both matrices are order one, allowing the modes they drive in the network to compete on equal footing. This scaling of low-rank structure relative to randomness at individual synapses is generic in models in which the couplings are the sum of random and low-rank terms [44–46] (but see Ref. [47] where a $\sim 1/\sqrt{N}$ rank-one term given by an outer product of orthogonal vectors was used). The scaling also appears in spiked matrix or tensor models in statistical physics [48]. In experiments that measure changes in synaptic strengths, plasticity on this weak scale could go unnoticed, but nevertheless exert dramatic influence at the network level.
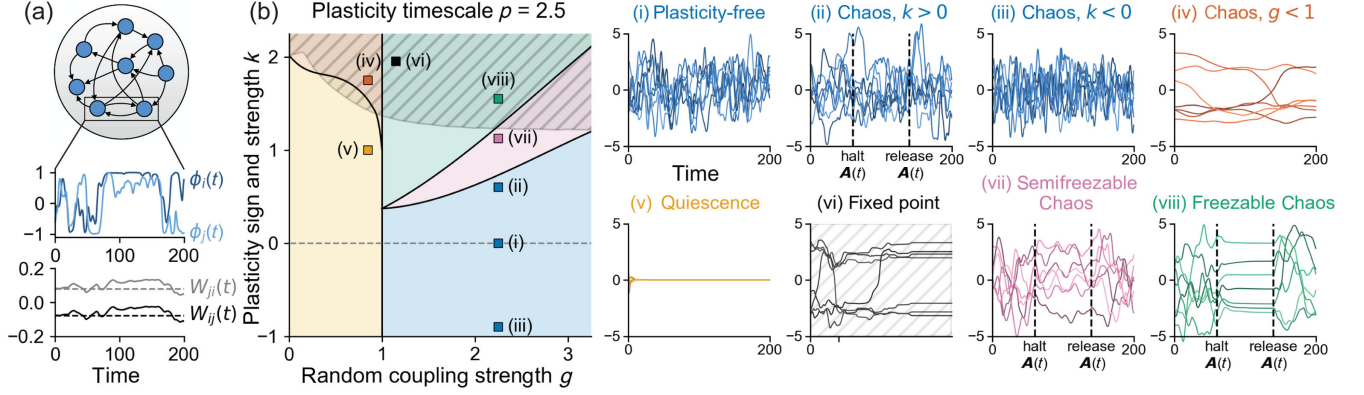
FIG. 1. (a) Dynamics of a pair of neurons (top) and of the synapses through which they are reciprocally coupled (bottom). Synapses fluctuate about quenched random strengths (dashed lines) in response to pre- and postsynaptic activity according to a Hebbian rule. (b) Left: phase diagram of the plastic network for $p = 2.5$. Right: example neuronal traces $x_i(t)$ from simulations of each phase-diagram region, with parameters given by the location of the associated square marker.

III. PHASE DIAGRAM

We summarize the behavior of the model with a phase diagram in (g, k) parameter space for $p = 2.5$ [Fig. 1(b)], noting that constant-p slices of the full (g, k, p) diagram are similar for order-one values of p. We refer to this phase diagram for the rest of this section.

For $k = 0$, the model reduces to that of Sompolinsky *et al.* [41] [dashed horizontal line in Fig. 1(b)]. For $g > 1$, this nonplastic network is chaotic (i). For $g < 1$, the trivial neuronal fixed point of the nonplastic network, $x(t) = 0_N$, is globally stable and the network is quiescent. In analogy with the nonplastic network, the plastic network can produce chaotic activity for $g > 1$, and the activity is further shaped by synaptic plasticity. Hebbian plasticity, $k > 0$, slows activity (ii), while anti-Hebbian plasticity, $k < 0$, quickens activity and generates an oscillatory component (iii).

For $g < 1$, the trivial neuronal-synaptic fixed point, $(x(t), A(t)) = (0_N, 0_{N \times N})$, is stable. However, in contrast to the nonplastic network, this fixed point is not necessarily globally stable, but coexists with chaotic states for large k (iv). Thus, Hebbian plasticity can induce dynamic activity in an otherwise quiescent network. For $g < 1$, if k is not large enough to induce dynamic activity, the network is globally quiescent (v).

In dynamic states, network activity is shaped by a proliferation of stable fixed points throughout phase space. In particular, if Hebbian plasticity is strong (hatched region), there exist stable fixed points to which finite-size networks settle following transient chaos, accompanied by the rank of $A(t)$ collapsing to unity (vi). Such fixed points are stable with respect to the combined neuronal-synaptic dynamics. Their number, which we compute, is exponential in N.

Two scenarios can lead to transient chaotic states. First, when $g < 1$ and k is large enough to induce chaotic activity, finite-size networks may collapse to the trivial fixed point.

Additionally, when (g, k) is in the hatched region, finite-size networks may collapse to nonzero fixed points.

Finally, we describe freezable chaos. Consider a chaotic state with Hebbian plasticity. Suppose we abruptly halt synaptic dynamics. If k is small, the halted-synapse neuronal dynamics are chaotic, with no trace of the halt-time neuronal state in the activity (ii). If k is larger, the halted-synapse neuronal dynamics are chaotic, but neurons fluctuate around the halt-time state, so a memory of this state is retained (vii). If k is sufficiently large, the halted-synapse neuronal dynamics are nonchaotic; neurons flow to a stable fixed point near the halt-time state (viii). In all cases, releasing the synapses returns the network to neuronal-synaptic chaos. As $N \to \infty$, these three cases, which we label nonfreezable, semifreezable, and freezable chaos, respectively, are distinct phases. In freezable chaos, there is, at any instant, a stable fixed point of neuronal dynamics that is destabilized by synaptic dynamics. Thus, a stable fixed point can be created near any neuronal state by halting synaptic plasticity, enabling a new form of short-term memory storage.

IV. DYNAMICAL MEAN-FIELD THEORY (DMFT)

The temporal structure of network activity is described in the limit $N \to \infty$ by a dynamical mean-field theory (DMFT) whose main order parameter is the single-unit autocovariance (two-point) function,

$$C(\tau) = \langle \phi_i(t) \phi_i(t + \tau) \rangle_J, \qquad (4)$$

where we assume statistical stationarity in time. Integrating the synaptic dynamics, Eq. (3), and inserting this into the neuronal dynamics, Eq. (1), gives

$$(1 + \partial_t) x_i(t) = \sum_j J_{ij} \phi_j(t) + \frac{k}{p} \int_{-\infty}^{t} dt' \, e^{-(t-t')/p}$$

$$\times \left[\frac{1}{N} \sum_j \phi_j(t) \phi_j(t') \right] \phi_i(t'), \qquad (5)$$

where we have separated terms arising from J (first term on the rhs) and $A(t)$ (second term on the rhs). Taking the limit $N \to \infty$ yields the single-site picture,

$$(1+\partial_t)x(t) = \eta(t) + \frac{k}{p}\int_{-\infty}^{t} dt'\, e^{-(t-t')/p} C(t-t')\phi(t'), \quad (6)$$

where $\eta(t)$ is an effective Gaussian field with zero mean and second-order statistics:

$$\langle \eta(t)\eta(t+\tau)\rangle_\eta = g^2 C(\tau). \quad (7)$$

The DMFT is closed by the self-consistency condition:

$$C(\tau) = \langle \phi(t)\phi(t+\tau)\rangle_\eta. \quad (8)$$

In the single-site dynamics of Eq. (6), synaptic plasticity introduces a convolutional self-coupling with a kernel that depends self-consistently on $C(\tau)$. For $k = 0$, the self-coupling vanishes and the DMFT reduces to that of Ref. [41], which can be solved analytically. For $k \neq 0$, the nonlinearity of the self-coupling induces a non-Gaussian distribution over $x(t)$—in particular, a distribution that becomes increasingly bimodal with larger k due to saturation of the tanh function—so we solve the DMFT equations using standard numerical techniques [49–54] (Appendix C 1). The DMFT agrees closely with simulations [Fig. 2(a)].
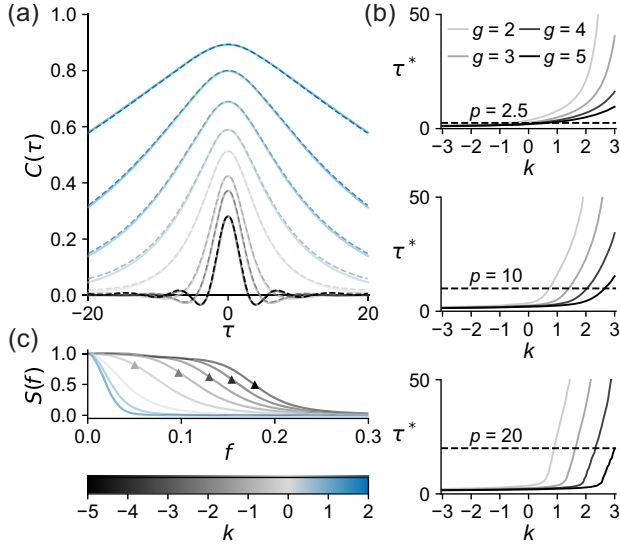

FIG. 2. Chaotic states with $g > 1$. (a) $C(\tau)$ from the DMFT (solid lines) and in simulations (dashed lines) for $g = 2$, $p = 2.5$, and various values of k (indicated by the lower color bar). (b) Dynamic timescale τ^* [Eq. (9)] as a function of k for various values of g. Dotted line indicates p. (c) Power spectra [normalized such that $S(0) = 1$] for the autocovariance functions shown in (a). For anti-Hebbian ($k < 0$) power spectra, triangular markers indicate an oscillatory frequency computed from the zero crossings of $C(\tau)$.

A. Chaotic states with $g > 1$

We now examine the solutions $C(\tau)$ of the DMFT. In the regime where the nonplastic network is chaotic, $g > 1$, Hebbian plasticity slows the activity, broadening $C(\tau)$ [Fig. 2(a), $k > 0$ solutions]. A network with static, symmetric couplings (e.g., the Hopfield network) admits a Lyapunov function that guarantees convergence to fixed points [26]. The slow activity generated by Hebbian plasticity results from competition between J, a random, asymmetric matrix that promotes dissipative, chaotic dynamics, and $A(t)$, a symmetric matrix that promotes convergence to a drifting fixed point that trails the neuronal state.

This competition has an interesting dependence on the model parameters. For small k and large p, neurons fluctuate rapidly relative to the synaptic decay timescale, and the effect of plasticity is averaged out. As k is increased, synapses slow neurons by attracting the neuronal state toward its history, permitting a stronger synaptic signal to be encoded. This stronger signal causes further neuronal slowing. The interaction of neurons and synapses in this positive-feedback process causes the timescale of neuronal fluctuations to diverge as k increases. For large p and k, finite-size networks can show bistability between a fast state in which plasticity is averaged out and a slow state in which synapses drag neurons (Appendix B).

This slowing behavior can also be understood through the DMFT. We quantify the speed of neuronal fluctuations by defining the dynamic timescale,

$$\tau^* = \int_0^\infty d\tau \left[\frac{C(\tau)}{C(0)}\right]^2, \quad (9)$$

whose dependence on g, k, and p is illustrated in Fig. 2(b). The size of the integral term in the single-site dynamics [Eq. (6)] is $\sim k\tau^*/p$ for $\tau^* \ll p$ (and $\sim k$ for $\tau^* \gg p$). Increasing this term slows $x(t)$, increasing τ^*. That τ^* depends on the size of this term, which itself depends on τ^*, produces a positive-feedback loop. Once k is large enough so that the integral term competes with $\eta(t)$, which occurs when $k\tau^*/p \sim g$, τ^* grows rapidly. The inflection of τ^* in k is sharpest for large p or g, in which case $k\tau^*/p$ and g are well separated at small k [Fig. 2(b), $p = 20$].

Under anti-Hebbian plasticity, rather than synapses attracting the neuronal state to its history, this effect is repulsive. This quickens the dynamics and adds an oscillatory component to the activity, tightening $C(\tau)$ and creating oscillations during its decay to zero [Fig. 2(a), $k < 0$ solutions]. In the single-site picture [Eq. (6)], plasticity modifies the dynamics of $x(t)$ by introducing time-delayed negative feedback, which generically induces oscillations [55,56]. While finite-size simulations of the model of [41] show limit cycles, our calculation of $C(\tau)$ for $N \to \infty$ demonstrates that this anti-Hebbian oscillatory component is not merely a finite-size effect.

These oscillations are further characterized by the (normalized) power spectrum $S(f) = |\hat{C}(f)|^2 / |\hat{C}(0)|^2$, where $\hat{C}(f)$ denotes the Fourier transform of $C(\tau)$ [Fig. 2(c)]. Rather than containing a peak at a nonzero frequency, the power spectra corresponding to the anti-Hebbian autocovariance functions in Fig. 2(a) possess a range of large frequencies [though $S(f)$ develops a nonzero peak for smaller values of g, not shown]. Point estimates of the dominant oscillatory frequency computed from the first three zero crossings of $C(\tau)$ roughly capture the frequency scale at which $S(f)$ decays [Fig. 2(c), triangular markers].

B. Chaotic states for $g < 1$

We next consider the regime $g < 1$ in which the nonplastic network is globally quiescent. The plastic network has a trivial fixed point, $(x(t), A(t)) = (\mathbf{0}_N, \mathbf{0}_{N \times N})$, that is stable when all eigenvalues of J have real part less than unity. For $N \to \infty$, Girko's circular law implies that this requires $g < 1$ [57]. In contrast to the nonplastic network, the trivial neuronal-synaptic fixed point coexists with dynamic states for large k as indicated by DMFT solutions [Fig. 3(a), solid lines]. In Sec. V B, we confirm over a restricted parameter regime that these solutions are chaotic (i.e., have positive maximum Lyapunov exponent).
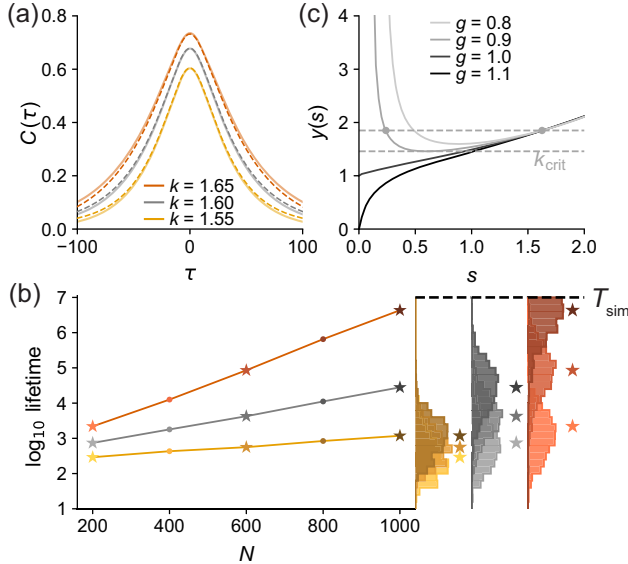

FIG. 3. Chaotic states for $g < 1$. (a) $C(\tau)$ from the DMFT (lines) and in simulations for $g = 0.9$, $p = 2.5$, and various values of k. (b) Left: median log-lifetime of transient activity, before collapsing to the trivial fixed point, as a function of N for $g = 0.9$ and values of k from (a). Right: histograms of the log-lifetime of transient chaos, corresponding to stars in the left-hand plot. Simulations were terminated at time $T_{\text{sim}} = 10^7$. (c) Curves $y(s)$ for solutions of the model of Stern *et al.* [53] for various values of g. Dashed horizontal lines correspond to different values of k, intersecting $y(s)$ at self-consistent solutions of Eq. (10).

Dynamic states for $g < 1$ eventually collapse to the trivial fixed point or, if k is large enough, to a stable nonzero fixed point (Secs. V B and VI). In this section, we consider values of $g < 1$ and k leading to chaotic states that are prone to collapsing to the trivial fixed point. Realizing these states in simulations is nontrivial because random initial conditions typically miss the dynamic attractor and decay to zero. A work-around is to deform a chaotic state with $g > 1$, for which the trivial fixed point is unstable, to a chaotic state for $g < 1$ by reducing g in the spirit of annealing. Using this method, we verified that simulations agree with the DMFT solutions [Fig. 3(a), dashed lines].

After realizing plasticity-induced chaotic states via this procedure, we observe transient activity with a lifetime that is approximately log-normally distributed [Fig. 3(b), right]. The median log-lifetime before collapsing is linear in N over several decades, consistent with the typical lifetime diverging exponentially and becoming infinite in the limit $N \to \infty$ in which the DMFT applies [Fig. 3(b), left]. This divergence is faster for larger k.

C. First-order transition to nontrivial DMFT solutions for $g < 1$

For $g < 1$, if k is not large enough to induce dynamic activity, the plastic network is globally quiescent. We now analyze the boundary between these phases. For order-one values of p, we find $\tau^* \gg p$ for dynamic states for $g < 1$, which reduces the single-site dynamics to the "slow" form:

$$(1 + \partial_t) x(t) = \eta(t) + k C(0) \phi(t). \tag{10}$$

This single-site picture is related to that of a nonplastic network with order-one self-coupling parameter s for which the single-site dynamics read

$$(1 + \partial_t) x(t) = \eta(t) + s \phi(t), \tag{11}$$

a model studied by Stern *et al.* [53]. This network has a continuous transition from quiescence to dynamic activity at $g + s = 1$. We map solutions of the Stern network onto solutions of the plastic network by enforcing $y(s) = s / C(0) = k$. For a given $g < 1$, $C(0)$ becomes nonzero at $s = 1 - g$, grows with s, and saturates at unity; thus, $y(s)$ descends from infinity at $s = 1 - g$ and grows as s for large s, tracing a U shape [Fig. 3(c)]. Each k draws a horizontal line intersecting $y(s)$ at self-consistent solutions of Eq. (10). The critical k, defining the boundary between phase-diagram regions (iv) and (v), is the minimum of $y(s)$ [Fig. 3(c), lower dashed line]. As $C(0)$ is finite here, this is a discontinuous, first-order transition. Physically, this is because sustaining dynamic activity for $g < 1$ requires finite self-coupling. Since the self-coupling depends on $C(0)$, time-dependent solutions with arbitrarily small $C(0)$ are not possible. As $g \to 0$, the dynamic timescale diverges,

leaving only fixed points. We show in Sec. VI that fixed points exist for $k > 2.02$ at $g = 0$.

Because of the ∪ shape of $y(s)$ for $g < 1$, $y(s) = k$ has two solutions for k larger than its critical value: one with large s and $C(0)$, the other with small s and $C(0)$ [Fig. 3(c), upper dashed line]. As $g \to 1^-$, the small-s solution vanishes (Sec. IV D) while the large-s solution remains finite. For $g > 1$, $y(s) = k$ has a unique solution whose deformation to $g < 1$ gives the large-s solution [Fig. 3(c), $g = 1.1$ curve].

D. Second-order transition at $g = 1$

Finally, we solve the DMFT for $g \to 1^+$, $k < 1$. This limit marks a continuous transition from dynamic activity to global quiescence in which $C(0)$ vanishes and τ^* diverges. Because of the vanishing variance, the self-coupling in Eq. (10) can be linearized:

$$(1 - kC(0) + \partial_t)x(t) = \eta(t). \tag{12}$$

This is equivalent to the single-site dynamics of the non-plastic network with $g_{\text{eff}} = g/[1 - kC(0)]$ and time constant $\tau_{\text{eff}} = 1/[1 - kC(0)]$. In the nonplastic network, to leading order in $\epsilon = g - 1 \ll 1$, $C(\tau) = \epsilon \operatorname{sech}(\epsilon\tau/\sqrt{3})$ [41]. Enforcing $C(0) = g_{\text{eff}} - 1$ gives, for the plastic network,

$$C(\tau) = \gamma \operatorname{sech}\left(\frac{\gamma\tau}{\sqrt{3}}\right), \qquad \gamma = \frac{\epsilon}{1 - k}, \tag{13}$$

to leading order in ϵ. In contrast to the behavior away from this transition, activity becomes faster with increasing k. $C(0)$ diverges as $k \to 1^-$, indicating that solutions with $g = 1$ for $k > 1$ have finite variance. Validity of the solution Eq. (13) requires $k < 1$, lest we obtain a nonphysical negative variance. On the other hand, as $g \to 1^-$ for $k > 1$, a positive variance is obtained; this is the small-s solution for $g < 1$ described in Sec. IV C.

V. HIGH-DIMENSIONAL ANALYSIS

The DMFT describes the temporal structure of network activity through an effective single-site picture. Importantly, the network dynamics result from a complex interaction of high-dimensional neuronal-synaptic modes. We now probe the high-dimensional origin of the dynamics, first through an analytical study of the spectrum of the Jacobian describing the local, linear dynamics, and then through a numerical study of the Lyapunov spectrum describing the global, nonlinear dynamics. Both the Jacobian and Lyapunov spectra show a topological transition at large k to a form with a slow, synapse-dominated band and a fast, neuron-dominated band, with the former driving network activity. This suggests a flipped view of the network dynamics as being driven by the synaptic couplings, with neurons serving as the connections.

A. Jacobian spectrum

Let us represent Eqs. (1)–(3) defining the model in the form

$$\partial_t \boldsymbol{x}(t) = \boldsymbol{F}(\boldsymbol{x}(t), \boldsymbol{a}(t)), \tag{14a}$$

$$(1 + p\partial_t)\boldsymbol{a}(t) = k\boldsymbol{G}(\boldsymbol{x}(t)), \tag{14b}$$

where $\boldsymbol{a}(t) = \operatorname{vec}\boldsymbol{A}(t)$ contains all $S = N^2$ elements of $\boldsymbol{A}(t)$. We use the notation $[\partial \boldsymbol{x}/\partial \boldsymbol{y}]_{ij} = \partial x_i/\partial y_j$ for vectors $\boldsymbol{x}, \boldsymbol{y}$. The Jacobian is a $(N + S)$-dimensional block matrix:

$$\boldsymbol{M} = \begin{pmatrix} \frac{\partial \dot{\boldsymbol{x}}}{\partial \boldsymbol{x}} & \frac{\partial \dot{\boldsymbol{x}}}{\partial \boldsymbol{a}} \\ \frac{\partial \dot{\boldsymbol{a}}}{\partial \boldsymbol{x}} & \frac{\partial \dot{\boldsymbol{a}}}{\partial \boldsymbol{a}} \end{pmatrix} = \begin{pmatrix} \frac{\partial \boldsymbol{F}}{\partial \boldsymbol{x}} & \frac{\partial \boldsymbol{F}}{\partial \boldsymbol{a}} \\ \frac{k}{p}\frac{\partial \boldsymbol{G}}{\partial \boldsymbol{x}} & -\frac{1}{p}\boldsymbol{I}_S \end{pmatrix}. \tag{15}$$

At any instant, the neurons provide N-dimensional input to the S synaptic variables, inducing a low-dimensional structure of the Jacobian. Rather than consider $\boldsymbol{M}$ directly, it is convenient to study the linear dynamics that it generates:

$$\partial_t \delta\boldsymbol{x}(t) = \frac{\partial \boldsymbol{F}}{\partial \boldsymbol{x}}\delta\boldsymbol{x}(t) + \frac{\partial \boldsymbol{F}}{\partial \boldsymbol{a}}\delta\boldsymbol{a}(t), \tag{16a}$$

$$(1 + p\partial_t)\delta\boldsymbol{a}(t) = k\frac{\partial \boldsymbol{G}}{\partial \boldsymbol{x}}\delta\boldsymbol{x}(t). \tag{16b}$$

The input that $\delta\boldsymbol{a}$ receives from $\delta\boldsymbol{x}$ is confined to an N-dimensional subspace spanned by the columns of $\partial\boldsymbol{G}/\partial\boldsymbol{x}$. Perturbations to $\delta\boldsymbol{a}$ in the $(S - N)$-dimensional orthogonal complement subspace relax with timescale p. Because of these relaxational modes, $\boldsymbol{M}$ has eigenvalue $-1/p$ with multiplicity $S - N$. The remaining $2N$ eigenvalues result from the interaction of $\delta\boldsymbol{x}$ and the component of $\delta\boldsymbol{a}$ in the N-dimensional subspace that receives input from $\delta\boldsymbol{x}$. Projecting $\delta\boldsymbol{a}$ into this subspace via $\delta\tilde{\boldsymbol{a}} = (\partial\boldsymbol{G}/\partial\boldsymbol{x})^+\delta\boldsymbol{a}$, where $(\cdot)^+$ denotes the pseudoinverse, the $2N$-dimensional dynamics of interest are

$$\partial_t\delta\boldsymbol{x}(t) = \frac{\partial \boldsymbol{F}}{\partial \boldsymbol{x}}\delta\boldsymbol{x}(t) + \frac{\partial \boldsymbol{F}}{\partial \boldsymbol{a}}\frac{\partial \boldsymbol{G}}{\partial \boldsymbol{x}}\delta\tilde{\boldsymbol{a}}(t), \tag{17a}$$

$$(1 + p\partial_t)\delta\tilde{\boldsymbol{a}}(t) = k\delta\boldsymbol{x}(t). \tag{17b}$$

The associated $2N$-dimensional dynamics matrix is

$$\tilde{\boldsymbol{M}} = \begin{pmatrix} \frac{\partial \boldsymbol{F}}{\partial \boldsymbol{x}} & \frac{\partial \boldsymbol{F}}{\partial \boldsymbol{a}}\frac{\partial \boldsymbol{G}}{\partial \boldsymbol{x}} \\ \frac{k}{p}\boldsymbol{I}_N & -\frac{1}{p}\boldsymbol{I}_N \end{pmatrix}. \tag{18}$$

In summary, the eigenvalues of $\boldsymbol{M}$ are $-1/p$ with multiplicity $S - N$ together with the $2N$ eigenvalues of $\tilde{\boldsymbol{M}}$. We refer to $\tilde{\boldsymbol{M}}$ as the *reduced Jacobian* [58].

Each eigenvector of $\tilde{\boldsymbol{M}}$ can be written in terms of its $\boldsymbol{x}$ and $\boldsymbol{a}$ components, $\boldsymbol{v} = (\boldsymbol{v}_x, \boldsymbol{v}_a)$. From the lower block row

of $\tilde{M}$, we obtain the relation $kv_x = (1 + p\lambda)v_a$. Letting $f_a(\lambda) = \|v_a\|^2/\|v\|^2$ denote the relative weight on the synaptic component, this relation implies

$$f_a(\lambda) = \frac{k^2}{k^2 + p^2|\lambda + 1/p|^2}, \qquad (19)$$

which falls off radially with distance from $-1/p$. Modes become synapse dominated as $\lambda \to -1/p$, giving way to a delta function of $S - N$ purely synaptic modes at this point.

The Jacobian analysis thus far holds for any $F(\cdot, \cdot)$, $G(\cdot)$, and $S > N$. In particular, the simplification to the structure of the Jacobian does not depend on the specific plasticity rule posited in Eq. (3). We now substitute the forms of $F(\cdot, \cdot)$ and $G(\cdot)$ from Eqs. (1)–(3), yielding $\tilde{M} = \tilde{M}_{\text{bulk}} + \tilde{M}_{\text{low-rank}}$, where

$$\tilde{M}_{\text{bulk}} = \begin{pmatrix} -I_N + J\text{diag}[\phi'] & C(0)\text{diag}[\phi'] \\ \frac{k}{p}I_N & -\frac{1}{p}I_N \end{pmatrix}, \quad (20a)$$

$$\tilde{M}_{\text{low-rank}} = \begin{pmatrix} A\text{diag}[\phi'] & \frac{1}{N}\phi\phi^T\text{diag}[\phi'] \\ 0 & 0 \end{pmatrix}. \quad (20b)$$

Here, (x, A) is a point in phase space, $\phi = \phi(x)$, $\phi' = \phi'(x)$, and $C(0) = \|\phi\|^2/N$. $\tilde{M}_{\text{bulk}}$ generates the bulk of the spectrum of $\tilde{M}$ while $\tilde{M}_{\text{low-rank}}$ can, in principle, contribute an intensive number of outlier eigenvalues. However, both in chaotic states and at fixed points, we find that the reduced Jacobian does not have outliers [60]. We therefore focus on the spectrum of $\tilde{M}_{\text{bulk}}$.

We compute the boundary curve encompassing the compact spectrum of $\tilde{M}_{\text{bulk}}$ for $N \to \infty$ using a theorem of Ahmadian et al. [61] concerning random matrices expressible as a linear reparametrization of an elementwise independent and identically distributed random matrix (Appendix E). This shows that the limiting spectral density of $\tilde{M}_{\text{bulk}}$ has support at $\lambda \in \mathbb{C}$ if

$$\left\langle \left| \frac{g(1 + p\lambda)}{(1 + p\lambda)(1 + \lambda)\frac{1}{\phi'(x)} - kC(0)} \right|^2 \right\rangle_x \geq 1, \quad (21)$$

where $\langle \cdot \rangle_x$ is an average over the components of x.

We use this result to probe the high-dimensional origin of the dynamics in the plastic network, noting that the spectral density of the Jacobian computed at any point on a dynamic trajectory is time independent and self-averaging as $N \to \infty$. We use the DMFT to obtain a Monte Carlo estimate of $\langle \cdot \rangle_x$. The predicted boundary tightly hugs the Jacobian spectra evaluated from simulations in dynamic states (Fig. 4). These simulation spectra do not contain outliers, justifying our focus on $\tilde{M}_{\text{bulk}}$. We place a dot at $-1/p$ to indicate the delta function of $N^2 - N$ synaptic relaxational modes present in the spectrum of M. Modes
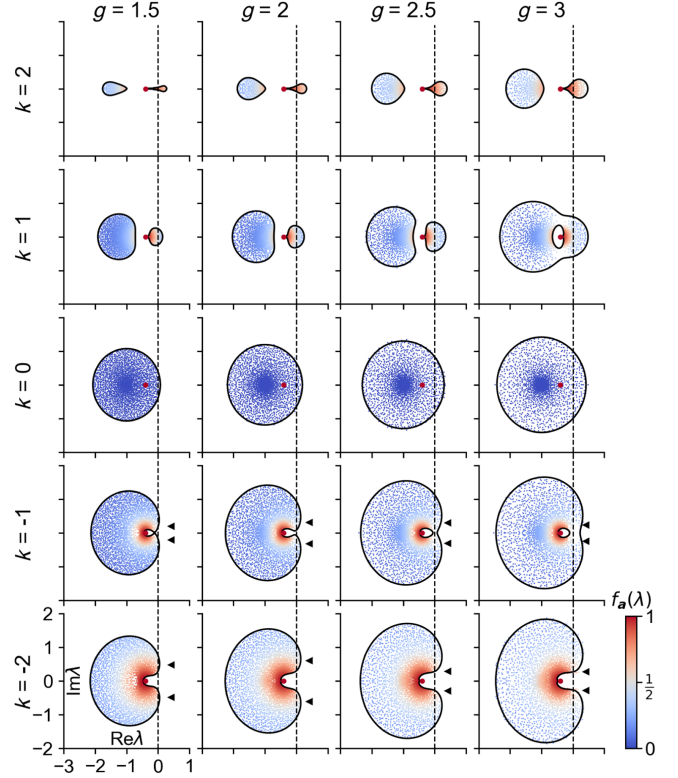

FIG. 4. Spectra of the Jacobian for $p = 2.5$ and various values of g (running horizontally) and k (running vertically). Lines: predicted boundary curves from random matrix theory and DMFT. Dots: spectra measured in simulations of chaotic plastic networks. Modes are colored by $f_a(\lambda)$, the weight on the synaptic part of the corresponding eigenvector of the reduced Jacobian [Eq. (19)]. The red dot at $\lambda = -1/p$ indicates a delta function of $N^2 - N$ synaptic modes. For anti-Hebbian ($k < 0$) spectra, triangular markers indicate an oscillatory frequency computed from the zero crossings of $C(\tau)$.

past the stability line, $\text{Re}(\lambda) = 0$, locally destabilize the network and drive the dynamics.

Setting $k = 0$ recovers the circularly symmetric spectrum of the nonplastic network, enclosed by $|1 + \lambda| \leq g\sqrt{\langle \phi'(x)^2 \rangle_x}$, as predicted by Eq. (21) (Fig. 4, $k = 0$ row). As k is increased, the delta function at $-1/p$ repels eigenvalues leftward, creating a hole (Fig. 4, $g = 3$, $k = 1$ spectrum). Further increasing k produces a topological transition to a spectrum with two bands and no holes (Fig. 4, all other spectra with $k > 0$). Unstable modes are increasingly focused along the real axis, leading to slow activity as seen via the DMFT. The slow, destabilizing band is dominantly synaptic, and the fast, relaxational band is dominantly neuronal. The slowest relaxing of the fast modes have real part close to -1, the neuronal decay timescale. This two-band topology therefore reflects a dynamic state in which slow network activity is synapse driven rather than neuron driven.

As k is decreased from zero into the anti-Hebbian regime, the delta function repels eigenvalues rightward,

creating a hole to its right (Fig. 4, $k = -1$ row). Further decreasing k produces a topological transition to a form with a single band and no holes (Fig. 4, $k = -2$ row). In contrast to the Hebbian spectra, there are pronounced lobes of dominantly imaginary unstable modes, generating fast, oscillatory activity also seen via the DMFT. The oscillatory frequency computed from zero crossings of $C(\tau)$, as in Fig. 2(c), corresponds roughly to the locations of the imaginary lobes of the spectra, particularly for smaller values of g (Fig. 4, triangular markers for $k < 0$ spectra).

B. Lyapunov spectrum

While analytically accessible, the Jacobian spectrum describes only the locally linear dynamics. Rigorously characterizing the nonlinear dynamics requires a calculation of the spectrum of Lyapunov exponents, which are defined as follows. Suppose a ball of radius ϵ is tossed into the flow. As the dynamics unfold, the ball expands and contracts into an ellipsoid. The Lyapunov exponents are the exponential growth or decay rates of the principal axes of this ellipsoid as $t \to \infty$ (simultaneously, $\epsilon \to 0$ is taken so that the ellipsoid stays small). A positive Lyapunov exponent implies exponential sensitivity to initial conditions, indicating chaos. The Lyapunov spectrum cannot be derived from the Jacobian spectrum due to both the non-normality of the Jacobian and the time dependence of its eigenvectors.

We first study the maximum Lyapunov exponent $\lambda^{\max}$, which dominates trajectory divergence as $t \to \infty$. We measured $\lambda^{\max}$ in simulations by injecting a small perturbation to the system and measuring the slope (versus t) of the log of the norm of the difference between the perturbed and unperturbed trajectories. For $g < 0$, we realized dynamic network states using the deformation method described in Sec. IV B. For each setting of (g, k), we simulated 200 random networks. The output of this analysis is displayed as a heat map in (g, k) parameter space in Fig. 5(a). Parameter values that resulted in convergence to nonzero fixed points (Sec. VI) within the simulation time for at least 80% of networks are hatched, values that resulted in quiescence of all networks are white.

In regions of parameter space that do not converge to fixed points over the simulation time, $\lambda^{\max}$ is positive, indicating chaos. This includes part of the region $g < 1$, confirming that plasticity can induce chaos in an otherwise quiescent network. This analysis provides a simulation-based confirmation of the boundary marking the first-order transition to nontrivial DMFT solutions for $g < 1$ derived in Sec. IV D [Fig. 5(a), gray lines].

As k is increased and/or g is decreased, we observe a smaller and smaller Lyapunov exponent that eventually results in simulations reliably collapsing to nonzero fixed points [Fig. 5(a), solid-to-hatched boundary]. This crossover occurs in a parameter regime for which phase space is densely filled with stable fixed points (Sec. VI). Using the present finite-N analysis, we are unable to determine
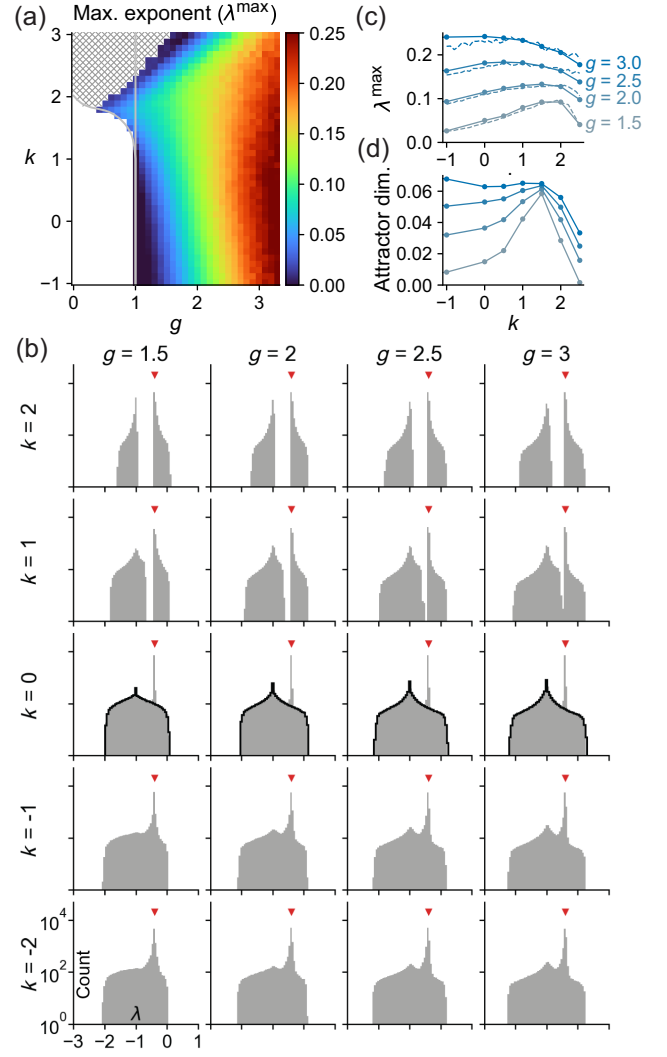

FIG. 5. (a) Maximum Lyapunov exponent $\lambda_{\max}$, computed by a perturbation method, throughout (g, k) parameter space with $p = 2.5$, $N = 4000$. White: quiescence. Hatched: convergence to nonzero fixed points. (b) Histograms of Lyapunov spectra, computed using tangent-vector propagation with $N = 900$, for $p = 2.5$ and various values of g (running horizontally) and k (running vertically). Black outline for $k = 0$ histograms: spectra of nonplastic network. The red triangle marks $-1/p$, where there are $\mathcal{O}(N^2)$ exponents in the full spectrum. (c) $\lambda_{\max}$ as a function of k for various values of g. Solid lines: estimate from tangent-vector propagation. Dashed lines: estimate from perturbation method. (d) Attractor dimension divided by N as a function of k for various values of g.

whether $\lambda^{\max}$ in the hatched region in Fig. 5(a) is small and positive, or negative. Additionally, solving the DMFT in the hatched region is numerically difficult due to the diverging dynamic timescale (Sec. IV A). As N is increased over a decade, the boundary marking this crossover shifts slightly upward (Fig. 9).

We next study the full Lyapunov spectrum for chaotic states. In general, Lyapunov spectra can be computed numerically by propagating a set of vectors in the tangent

space of the flow and periodically orthonormalizing them to prevent their explosion (vanishing) and to extract their growth (decay) rates, as explained in detail in prior works [49,62,63]. The plastic network has $N + N^2$ variables, so propagating a complete basis is prohibitively computationally expensive for large N. Fortunately, $\mathcal{O}(N^2)$ exponents concentrate at $-1/p$, so it suffices to compute the $\mathcal{O}(N)$ largest and smallest exponents. We compute the largest exponents using the aforementioned procedure with an undercomplete set of $\mathcal{O}(N)$ tangent-space vectors. We find the smallest exponents by doing the same for the time-reversed dynamics, noting that the smallest exponents are the largest of the time-reversed system. A complication is that the time-reversed dynamics are unstable. We therefore run time-reversed tangent-space dynamics, tamed by orthonormalization, atop a time-reversed trajectory produced by the forward-time dynamics. We verified that this method accurately computes the largest and smallest Lyapunov exponents of the nonplastic network [41,63].

Histograms of the Lyapunov spectra are shown in Fig. 5(b). For $k = 0$, the spectrum is the same as that of a nonplastic network with a spike at $-1/p$ [64]. We verified that the measurement of λ^{max} obtained using this method matches the perturbation measurement displayed in the heat map [Fig. 5(c)].

For large k, the Lyapunov spectra recapitulate the topological transition to two bands of the Jacobian spectra, further demonstrating a synapse-driven dynamic state [Fig. 5(b)]. In analogy with the Jacobian spectra, the slow, destabilizing Lyapunov band spans $-1/p$ to λ^{max}, and the fast, relaxational band has an upper limit near -1.

Our calculation of the Lyapunov spectrum allows for further calculation of diffeomorphic-invariant properties of the strange attractor [63]. We focus on its dimension, shown in Fig. 5(d), defined by the Kaplan-Yorke conjecture as the number of exponents that must be summed, ranked in descending order, to achieve a cumulative sum of zero [65]. We display an intensive version of this quantity that has been divided by N [42,63]. Both λ^{max} and the attractor dimension vary nonmonotonically in k [Figs. 5(c) and 5(d)]. This contrasts with the dynamic timescale τ^*, which increases monotonically [Fig. 2(b)]. The eventual decrease of the attractor dimension is reminiscent of the behavior of the participation ratio-based dimension of activity in networks with partially symmetric connectivity, which was recently computed analytically [42]. The decline of λ^{max} and the attractor dimension at large k likely reflects the proliferation of stable fixed points throughout phase space, the subject of the next section.

VI. FIXED POINTS

For large k, the dynamics of the plastic network are shaped by a proliferation of stable fixed points, and finite-size networks settle to these fixed points after

transient chaos [Fig. 1(b), hatched region and (vi)]. We first probe this settling process by analyzing how $A(t)$ collapses to a rank-1 state, measuring the approximate rank of $A(t)$ as the participation ratio of its spectrum $\{\lambda_i(t)\}$:

$$\mathrm{PR}_{A(t)} = \frac{[\sum_i \lambda_i(t)]^2}{\sum_i \lambda_i^2(t)} = \frac{[\mathrm{tr} A(t)]^2}{\|A(t)\|_F^2}. \qquad (22)$$

Note that if $A(t)$ encodes P decorrelated neuronal states with equal magnitude [i.e., $\lambda_1(t), \ldots, \lambda_P(t) = \mathrm{const}$ and $\lambda_{P+1}(t), \ldots, \lambda_N(t) = 0$], then $\mathrm{PR}_{A(t)} = P$. Evaluating Eq. (22) in the limit $N \to \infty$ gives

$$\mathrm{PR}_{A(t)} = \frac{p}{\mathcal{T}}, \quad \mathcal{T} = \int_0^\infty d\tau e^{-\tau/p} \left[\frac{C(\tau)}{C(0)}\right]^2. \qquad (23)$$

Thus, $\mathrm{PR}_{A(t)}$ is intensive and, as $N \to \infty$, time independent. If $\tau^* \ll p$, then $\mathcal{T} \approx \tau^*$, so $\mathrm{PR}_{A(t)}$ is the ratio of these timescales. On the other hand, if $\tau^* \gg p$, then $\mathcal{T}$ is slightly less than p, so $\mathrm{PR}_{A(t)}$ is slightly larger than unity.

Hebbian plasticity tends to slow chaos. As k is increased, $\mathrm{PR}_{A(t)}$ therefore drops closer to unity, with temporal fluctuations about the mean-field value in finite-size networks [Fig. 6(b)]. During chaos, neurons continuously escape the synaptic drag. However, in finite-size networks with sufficiently large k, synapses can "win"; namely, the network fluctuates into fixed point of the form $(x(t), A(t)) = (x, \phi(x)\phi(x)^T)$ with $\mathrm{PR}_{A(t)} = 1$ [Fig. 6(b)]. These fixed points are stable with respect to the combined neuronal-synaptic dynamics.

We now compute the number of stable fixed points in phase space. As $k \to \infty$, there is a stable fixed point associated with each of the 2^N binary neuronal states. Thus, any initial condition is pulled to a fixed point with high overlap with the initial neuronal state, consistent with the diverging dynamic timescale τ^* in this regime (Sec. IV A). For finite k, we expect exponentially many stable fixed points. This is in contrast to the nonplastic network [41], for which there exist exponentially many fixed points for $g > 1$, but they are all unstable [66,67]. Because of the exponential scaling, we will study the log-number of stable fixed points per neuron, an intensive, self-averaging quantity. Fluctuations around fixed points are governed by the Jacobian spectrum analyzed in Sec. V. Our derivation follows that of Stern et al. [53], differing in its initial steps that enforce stability using this neuronal-synaptic spectrum.

At a fixed point, the DMFT equations reduce to

$$x = \eta + kC(0)\phi(x), \qquad (24)$$

where $\eta \sim \mathcal{N}(0, g^2 C(0))$ and $C(0) = \langle \phi^2(x) \rangle_\eta$. Equation (24) can be written $g(x) = x - kC(0)\phi(x) = \eta$. The Gaussian
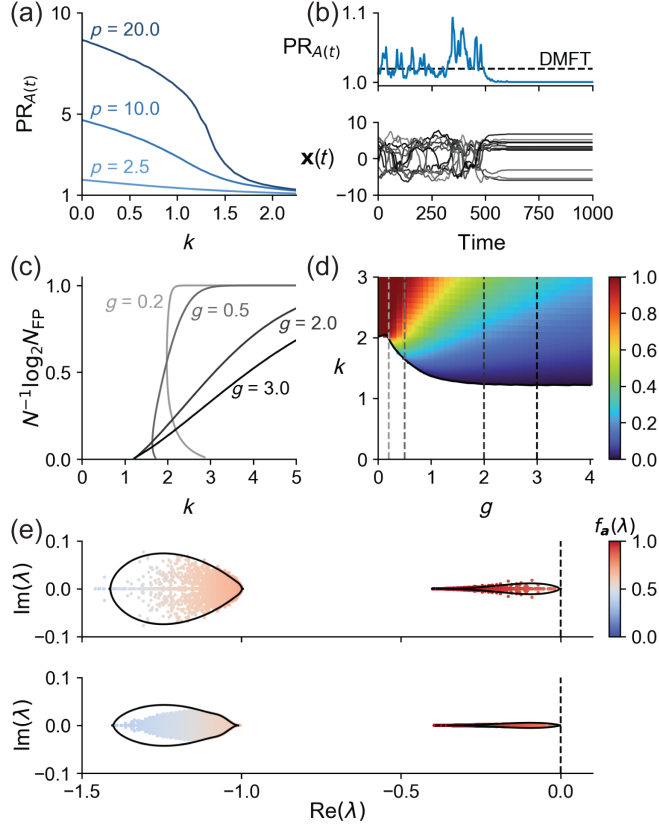
FIG. 6. Stable fixed points. (a) DMFT participation ratio [Eq. (23)] of $A(t)$ as a function of k for $g = 3$ and various values of p. (b) Top: empirical participation ratio [Eq. (22)] of $A(t)$ with $g = 2$, $k = 2.25$, and $p = 2.5$. The network settles to a stable fixed point. Dashed line: DMFT value. Bottom: neuronal traces during the same settling event. (c) Log-number of stable fixed points per neuron as a function of k for various values of g. (d) Log-number of stable fixed points per neuron throughout (g, k) parameter space. (e) Jacobian spectra at fixed points. Lines: predicted boundary curve from mean-field analysis. Dots: spectra after settling to a fixed point in many simulations. Top: $g = 2$, $k = 2.25$, $p = 2.5$ [as in (b)]. Bottom: $g = 0.5$, $k = 1.9$, $p = 2.5$.

distribution over η induces a non-Gaussian distribution over x due to the nonlinearity of $g(x)$. If $|\eta| < \eta_m$ for some η_m, there are three solutions to $g(x) = \eta$. The two outer solutions are at points of positive slope of $g(x)$; the central solution has negative slope. The negative-slope solution renders a fixed point unstable for the following reason. The denominator of the averaged quantity in the Jacobian boundary [Eq. (21)] is

$$\frac{1}{\phi'(x)} [g'(x) + \lambda + p\lambda(1 + \lambda)]. \tag{25}$$

Because $\lambda + p\lambda(1 + \lambda)$ varies from zero to infinity as λ varies from zero to infinity, if $g'(x) < 0$, the denominator vanishes for some $\lambda > 0$, precluding stability. Two solutions remain for $|\eta| < \eta_m$. Stern *et al.* [53] observed that typical fixed points maximize the combinatorial number of solutions

subject to stability and, moreover, are marginally stable, meaning that the spectral boundary sits at $\lambda = 0$. Setting $\lambda = 0$ in Eq. (21), marginal stability requires

$$\left\langle \left(\frac{g}{\cosh^2 x - kC(0)} \right)^2 \right\rangle_x = 1. \tag{26}$$

Since p appears in neither Eq. (24) nor Eq. (26), it drops out of the calculation, indicating that the number of stable fixed points is independent of this timescale [note, however, that the shape of the Jacobian spectrum is p dependent; see Fig. 6(e)]. Following Sec. IV, we map solutions of the nonplastic network of Ref. [53] onto solutions of the plastic network by enforcing $s = kC(0)$.

The structure the solutions is as follows. For each g, there is an onset of fixed points at a critical k [Fig. 6(c)]. The log-number grows monotonically with k and saturates at $\log_2 2^N/N = 1$. The onset of fixed points is discontinuous for $g < 0.76$, in which case there are two solutions of the fixed-point mean-field theory for a given (g, k): one with large s and $C(0)$, the other with small s and $C(0)$. The former are exponentially dominant. In Fig. 6(d), we display the log-number of fixed points per neuron as a heat map in (g, k) parameter space.

We now study the spectra of fluctuations around fixed points. We used the analytical form of the distribution over x from the mean-field analysis to evaluate $\langle \cdot \rangle_x$ in Eq. (21), yielding a prediction for the Jacobian spectrum at fixed points. We also allowed many simulations of chaotic networks to settle to fixed points, then measured their Jacobian spectra. We find that the simulation fixed points are indeed marginal, with spectra described accurately by the theoretical prediction [Fig. 6(e)].

For $g = 0$, fixed points satisfy $x_i = kN^{-1}\|\phi(\boldsymbol{x})\|^2 \phi(x_i)$ and thus $x_i = \pm\chi$, where $\chi = k\phi^3(\chi)$. For fixed points to exist, k must be large enough for this constraint to have a nontrivial solution in χ. The smallest such k satisfies $1 = 3k\phi^2(\chi)\phi'(\chi)$ (upon differentiating both sides in χ). Combining these constraints gives $\phi(\chi)/\phi'(\chi) = 3\chi$ with solution $\chi = \pm 1.42$, and thus $k = 2.02$. The stability of fixed points with this critical value of k is checked by direct evaluation of the eigenvalues of the reduced Jacobian, which are $\lambda = -[1 + p \pm \sqrt{(1 + p)^2 - 8p/3}]/2p < 0$ with multiplicity $N - 1$ for each sign, comprising the bulk; $\lambda = -1/p - 2/3 < 0$ with multiplicity one; and $\lambda = 0$ with multiplicity one, indicating marginal stability. Further increasing k gives proper stability. Here, the marginal eigenvalue is an outlier, while the bulk has finite negative real part. This suggests that the mean-field fixed-point calculation, which assumed marginal stability of the bulk, breaks down for small g. Simulations and theory agree down to $g = 0.5$, $k = 1.9$, so such a breakdown would have to occur for smaller values of these parameters.

VII. FREEZABLE CHAOS

The previous section characterized stable fixed points of the dynamics, as is typical dynamical-systems studies. Another question is whether a subsystem can have a stable fixed point that is unstable in the context of full system. We now study *freezable chaos*, a state where a stable fixed point of neuronal dynamics is continuously destabilized through synaptic dynamics, generating chaos.

As described in Sec. III, for chaotic states with Hebbian plasticity, we define nonfreezable, semifreezable, and freezable chaos depending on the neuronal dynamics that ensue after halting synaptic dynamics. In (semi)freezable chaos, neurons retain a stable memory of the halt-time state as we illustrate in Fig. 7(a).

Picking the halt time to be $t = 0$, the couplings remain at $W(0) = J + A(0)$. Networks with such "random-plus-low-rank" couplings have been studied in the context of nonplastic networks by Mastrogiuseppe and Ostojic [44], who found chaotic, structured-chaotic, and structured-homogeneous phases that are qualitatively similar to non-freezable, semifreezable, and freezable chaos in plastic networks. Crucially, our analysis in plastic networks is complicated by $A(0)$ arising due to synaptic plasticity dependent upon neuronal activity driven by J. More specifically, the neuronal states comprising $A(0)$ are largely confined to a subspace spanned by dominant eigenvectors of J [42], inducing alignment between J and $A(0)$. We demonstrate the importance of this alignment by running a simulation in which we halt $A(t)$, storing the halt-time neuronal state as a stable fixed point [Fig. 7(b)]. We then shuffle J while keeping the halted $A(t)$ fixed. This preserves the statistics of J, but destroys correlations between J and the halted $A(t)$. If these correlations were negligible, the neuronal fixed point would reorganize to a different fixed point. Instead, shuffling destroys the fixed point, causing the network to switch to neuronal chaos. Upon releasing $A(t)$, the network returns to neuronal-synaptic chaos, and $A(t)$ adapts to the permuted version of J. At a later time, we halt synaptic dynamics again. As $A(t)$ has adapted to the shuffled J, a stable neuronal fixed point is created.

We handle this alignment through a replica mean-field analysis involving two networks, A and B, with neuronal
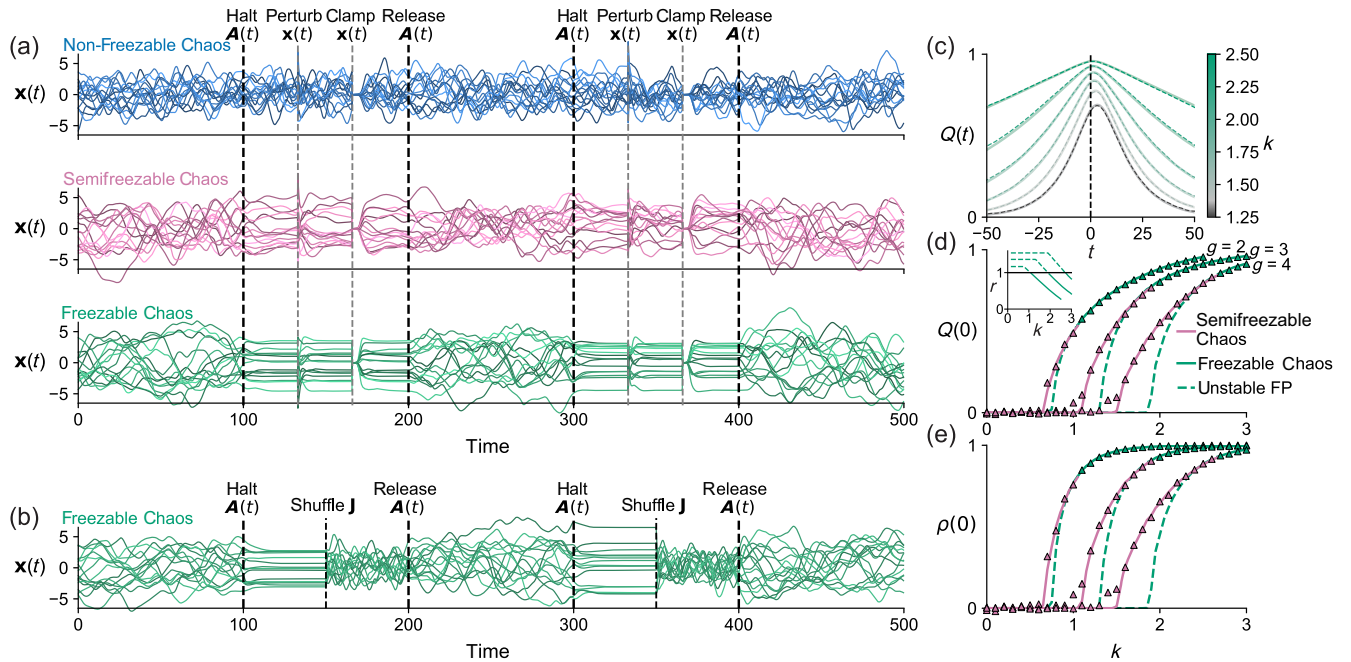

FIG. 7. Freezable chaos. (a) Neuronal traces in nonfreezable (top, $k = 0.6$), semifreezable (middle, $k = 1.2$), and freezable (bottom, $k = 1.55$) chaos for $g = 2.25$ and $p = 2.5$. At the first vertical line, $A(t)$ is halted. In (semi)freezable chaos, a memory of the halt-time neuronal state persists. Stability is demonstrated by perturbing the neurons (second line) and clamping them near zero (third line). In (semi)freezable chaos, neurons relax back to the stored pattern after these manipulations. At the fourth line, $A(t)$ is released. This process is repeated at the fifth line. (b) Demonstration of the alignment between J and $A(t)$. At the first vertical line, $A(t)$ is halted. At the second line, J, but not the halted $A(t)$, is shuffled, destroying the neuronal fixed point. At the third line, $A(t)$ is released. This process is repeated at the fourth line. $g = 2.5$, $k = 2.55$, and $p = 2.5$ [as in the freezable-chaos plot in (a)]. (c) Time-dependent overlap $Q(t)$ from the two-replica DMFT (solid lines) and in simulations (dashed lines) for $g = 2$, $p = 2.5$, and various values of k. (d) Halt-time overlap $Q(0)$ from the two-replica DMFT (lines) and in simulations (triangular markers) as a function of k for $p = 2.5$ and various values of g. Markers are colored according to whether simulations show semifreezable or freezable activity. (e) Same as (d), but using the zero-time correlation coefficient $\rho(0)$. In (c)–(e), we show only the positive solutions for $Q(t)$ and $\rho(t)$, which are realized upon halting synapses; there is also a symmetric negative solution.

states $\boldsymbol{\phi}^A(t) = \phi(\boldsymbol{x}^A(t))$ and $\boldsymbol{\phi}^B(t) = \phi(\boldsymbol{x}^B(t))$. Network A has neuronal-synaptic dynamics for all t with quenched random couplings $\boldsymbol{J}$. Network B has neuronal dynamics for all t with halt-time couplings $\boldsymbol{W}(0)$ constructed from the same $\boldsymbol{J}$ and neuronal states $\boldsymbol{\phi}^A(t)$ as network A:

$$W_{ij}(0) = J_{ij} + \frac{k}{p}\int_0^\infty dt\, e^{-t/p}\phi_i^A(-t)\phi_j^A(-t). \quad (27)$$

We define order parameters to characterize the phases of interest, starting with the overlap:

$$Q(t) = \lim_{t'\to\infty}\langle\phi_i^A(t)\phi_i^B(t')\rangle_{\boldsymbol{J}}. \quad (28)$$

Evaluated at $t = 0$, this parameter indicates how accurately the halt-time neuronal state is retained as a memory. We also define the autocovariance function of network B,

$$D(\tau) = \lim_{t\to\infty}\langle\phi_i^B(t)\phi_i^B(t+\tau)\rangle_{\boldsymbol{J}}, \quad (29)$$

where we assume statistical stationary in time. The autocovariance function of network A is $C(\tau)$, derived previously. Finally, assessing stability of the neuronal fixed point, when one exists, requires the stability matrix $-\boldsymbol{I}_N + \boldsymbol{W}(0)\mathrm{diag}[\phi'(\boldsymbol{x}^B)]$, where $\boldsymbol{x}^B$ is a fixed point in network B. Stability requires that the spectrum has negative

real part. Instability is dominated by the circularly symmetric bulk, so it suffices to compute its radius r,

$$r^2 = g^2\langle\phi'(x_i^B)^2\rangle_{\boldsymbol{J}}. \quad (30)$$

which follows from random matrix theory [61].

Having established these order parameters, we can define the phases of interest quantitatively. Chaos is nonfreezable when $Q(t) = 0$ is the only solution. Chaos is semifreezable when there is a nontrivial solution for $Q(t)$ associated with a solution for $D(\tau)$ that decays in τ (to a nonzero value). In this case, there may or may not be a fixed-point solution in which there is a distinct nontrivial solution for $Q(t)$ associated with $D(\tau) = \mathrm{const}$. If there is a fixed point, it is unstable, $r > 1$. Finally, chaos is freezable when only a fixed-point solution for $Q(t)$ and $D(\tau)$ exists. In this case, it is stable, $r < 1$. Because the fixed point of the halted-synapse system is stable, the lifetime of the memory is infinite. Because of the $\boldsymbol{x} \to -\boldsymbol{x}$ symmetry of the network, if $Q(t)$ is a solution, so is $-Q(t)$. Upon halting synapses, the positive solution is realized, barring "flips" that occur in finite-size networks near the onset of semifreezable chaos.

We derive a two-replica DMFT that permits calculation of these order parameters. The high-dimensional equations describing the two replicas are

$$(1 + \partial_t)x_i^A(t) = \sum_j J_{ij}\phi_j^A(t) + \frac{k}{p}\int_{-\infty}^t dt'\, e^{-(t-t')/p}\left[\frac{1}{N}\sum_j\phi_j^A(t)\phi_j^A(t')\right]\phi_i^A(t'), \quad (31a)$$

$$(1 + \partial_t)x_i^B(t) = \sum_j J_{ij}\phi_j^B(t) + \frac{k}{p}\int_0^\infty dt'\, e^{-t'/p}\left[\frac{1}{N}\sum_j\phi_j^A(-t')\phi_j^B(t)\right]\phi_i^A(-t'). \quad (31b)$$

Sending $N \to \infty$ and taking the limit where the time coordinate of network B is much greater than zero, we obtain the single-site picture,

$$(1 + \partial_t)x^A(t) = \eta^A(t) + \frac{k}{p}\int_{-\infty}^t dt'\, e^{-(t-t')/p}C(t-t')\phi^A(t'), \quad (32a)$$

$$(1 + \partial_t)x^B(t) = \eta^B(t) + \frac{k}{p}\int_0^\infty dt\, e^{-t/p}Q(-t)\phi^A(-t), \quad (32b)$$

where $\eta^A(t)$ and $\eta^B(t)$ are Gaussian fields with zero mean and second-order statistics:

$$\left\langle\begin{pmatrix}\eta^A(t)\\\eta^B(t)\end{pmatrix}\begin{pmatrix}\eta^A(t') & \eta^B(t')\end{pmatrix}\right\rangle_{\eta^A,\eta^B}$$
$$= g^2\begin{pmatrix}C(t-t') & Q(t)\\Q(t') & D(t-t')\end{pmatrix}. \quad (33)$$

The off-diagonal covariances effectively encode the alignment between $\boldsymbol{J}$ and $\boldsymbol{A}(0)$. These off diagonals do not depend on the time coordinates of network B in accordance with the limits taken in Eqs. (28) and (29). The system is closed by the self-consistency conditions:

$$Q(t) = \langle\phi^A(t)\phi^B(t')\rangle_{\eta^A,\eta^B}, \quad (34a)$$

$$D(\tau) = \langle\phi^B(t)\phi^B(t+\tau)\rangle_{\eta^A,\eta^B}. \quad (34b)$$

We solve the DMFT numerically; imposing $\partial_t x^B(t) = 0$ and $D(\tau) = \mathrm{const}$ gives the fixed-point solution (Appendix C 2). We find excellent agreement between theory and simulations [Figs. 7(c)–7(e)].

We now examine the solutions of the two-replica DMFT. In the freezable-chaos regime, $Q(t)$ peaks at $t > 0$, indicating that the fixed point is more aligned with neuronal states that would have unfolded after the halt time than with the halt-time state itself [Fig. 7(c)]. This reflects a tendency of

neurons to continue with "momentum" before becoming trapped in a fixed point [Fig. 7(a)].

Increasing k from zero yields a sequence of continuous phase transitions that we analyze by plotting $Q(0)$ and r against k [Fig. 7(d)]. For small k, $Q(0) = 0$, indicating nonfreezable chaos. As k is increased, $Q(0)$ develops a nonzero solution associated with a decaying $D(\tau)$, marking the onset of semifreezable chaos. As k is increased further, $Q(0)$ develops an additional nonzero solution associated with $D(\tau) = $ const, marking the onset of an unstable fixed point. Instability is signaled by $r > 0$, with r computed under the fixed-point solution. Continuing to increase k causes r to decrease and drop below unity, marking the onset of freezable chaos, at which point the dynamic and fixed-point solutions converge. The convergence of the dynamic and fixed-point solutions at $r = 1$ is expected on physical grounds and can also be derived through the DMFT: when r is computed under the dynamic solution, the decay timescale of $D(\tau)$ diverges as $\sim 1/\sqrt{r-1}$ as $r \to 1^+$, giving $D(\tau) = $ const at $r = 1$ (Appendix D). The quality of memory retention can be measured by the correlation coefficient,

$$\rho(t) = \frac{Q(t)}{\sqrt{D(0)C(0)}}, \tag{35}$$

which varies between zero and unity [Fig. 7(e)].

The two-replica DMFT can be solved analytically as $g \to 1^+$. In this limit, $C(t), D(t), Q(t) \sim g - 1$ and the decay timescales of $C(t)$ and $Q(t)$ diverge. We consider fixed-point solutions, $D(\tau) = D$. At leading order in $\epsilon = g - 1$, the single-site equations reduce to

$$(1 + \partial_t)x^A(t) = \eta^A(t) + kC(0)x^A(t), \tag{36a}$$

$$x^B = \eta^B + kQ(0)x^A(0). \tag{36b}$$

To determine $Q(0)$, we square Eq. (36b) and average over $\eta^A(t)$ and η^B. This gives, to order ϵ^2, $D^2 - \epsilon D - kQ^2(0) = 0$, with the solution

$$Q(0) = \pm\sqrt{\frac{D(D-\epsilon)}{k}}, \tag{37}$$

implying $D \geq \epsilon$, with strict equality when $Q(0) = 0$. Next, we multiply Eqs. (36a) and (36b) and average over $\eta^A(t)$ and η^B. This gives, to order ϵ^2, $(D - \epsilon + \partial_t)Q(t) = kQ(0)C(t)$, with the solution

$$Q(t) = kQ(0) \int_{-\infty}^{t} dt' e^{-(D-\epsilon)(t-t')}C(t'). \tag{38}$$

This causal filtering makes $Q(t)$ peaked at $t > 0$. Setting $t = 0$ and doing the integral using Eq. (13) gives a

condition for determining D, namely, $h(D/\epsilon, k) = k$, where

$$h(u, k) = \frac{2}{\sqrt{3}}\left[\psi\left(\frac{\sqrt{3}(1-k)(u-1)+3}{4}\right) - \psi\left(\frac{\sqrt{3}(1-k)(u-1)+1}{4}\right)\right]^{-1}, \tag{39}$$

and $\psi(\cdot)$ is the digamma function. Because $h(u, k)$ is monotonically increasing in u and, in our case, $u = D/\epsilon \geq 1$, a nontrivial solution exists only for $k > h(1, k) = 2/\sqrt{3}\pi \approx 0.37$ [$h(1, k)$ does not depend on k]. Stability is checked by computing the fixed-point spectral radius to first order in ϵ, $r = 1 - D + \epsilon \leq 1$, implying that the fixed-point and dynamic solutions emerge together at $k = 0.37$ as $g \to 1^+$ [Fig. 1(b)].

VIII. DISCUSSION

We characterized the dynamics of N neurons coupled to N^2 dynamic synapses. Strong Hebbian plasticity causes the timescales of the system, measured through the Jacobian or Lyapunov spectra, to segregate into a slow, synapse-dominated band and and a fast, neuron-dominated band. The synapse-dominated band drives the dynamics. It is possible that this two-band structure could be detected through *in vivo* recordings of neuronal activity. Takens's embedding theorem implies that it is possible, in principle, to extract the spectrum of neuronal-synaptic timescales from neuronal activity alone [68]. This segregation of timescales could also be examined during task execution. If the dynamics are synapse driven, neurons may revert to their trial-average trajectories upon optogenetic or electro-physiological perturbation. Prior studies have attributed such robustness to neuronal mechanisms such as excita-tory-inhibitory balance [69], but our study invites reeval-uation of such data with an emphasis on synaptic dynamics. Indeed, Hebbian plasticity can enhance the robustness of an attractor manifold against distractors [70].

Increasing the strength of Hebbian plasticity initially enriches network dynamics, indicated by an increased maximum Lyapunov exponent and attractor dimension. Beyond a certain plasticity strength, these metrics decrease, likely due to the increased presence of stable fixed points throughout phase space. This implies that there may be an optimal level of plasticity for task performance—one that is robust enough to enrich the dynamics compared to a nonplastic network but not so overpowering that it simplifies the dynamics through the overabundance of fixed points. This could be investigated by training plastic networks to solve tasks, e.g., using the FORCE learning algorithm [40] or backpropagation.

Our analyses point to a region of parameter space with large k and/or small g where it is possible that the behaviors

of the fixed-point density and maximum Lyapunov exponent are more interesting than what we have explored. In particular, as $(g, k) \to (0, 2.02)$, fixed points are marginally stable by virtue of outliers, not the bulk (Sec. VI). Additionally, our Lyapunov analysis leaves open the possibility that the maximum exponent is negative for sufficiently large k even as $N \to \infty$ (Sec. V B and Appendix F; see Fig. 9). Given that phase space is densely filled with stable fixed points in this parameter regime, it is possible that these features signal a novel glassy phase of the model. Such phases in dynamic networks remain poorly understood and are an important direction for future research [71,72].

Our study addresses chaotic networks that either generate activity autonomously or produce complex responses to inputs [73]. A potentially desirable alternative property is stability, defined by the ability of a network to generate input-driven trajectories that are robust against perturbations; however, such networks cannot generate rich activity autonomously. Kozachkov *et al.* [74] analyzed a plastic network with the same governing equations as our model, but without quenched disorder, establishing conditions for stable dynamics [75]. A key finding was that anti-Hebbian plasticity can promote stable dynamics, pointing to a possible function for ongoing plasticity in input-driven computations.

We considered plastic synapses with strengths weaker than random synapses by a factor of $1/\sqrt{N}$. In experiments measuring synaptic strength changes, such plasticity may easily be overlooked despite its order-one network-level impact. Our model assumes all-to-all connectivity; if neurons receive $K < N$ inputs, the structure-to-randomness scaling is $1/\sqrt{K}$, making detection of plasticity more feasible if K is not too large.

Humans and animals can remember a stimulus over a delay period, implying a form of rapid information storage in neural circuits, i.e., working memory (WM). Freezable chaos provides a new WM mechanism that we now compare to prior models. Most WM models rely on either cell-intrinsic or network-level mechanisms that support self-sustained activity. These "persistent activity" models are supported by some experimental studies, but undermined by others showing "activity-silent" WM [76]. These latter studies suggest that information can be rapidly stored in synapses, requiring fast synaptic plasticity. Synaptic WM models typically use short-term facilitation (STF) due to the convention that fast plasticity is presynaptic [77,78]. Because such plasticity cannot create attractor states of neuronal dynamics, STF models require existing symmetric structure in the synapses, potentially formed through prior Hebbian plasticity. A prototypical example is the model of Mongillo *et al.* [9] in which clusters of excitatory neurons with broad inhibition prime a network to function in a metastable regime. Because of STF, an activity pattern can be selectively sustained by providing transient external input to one of the clusters. A key requirement of this class of proposals is that the possible neuronal states to be stored are known in advance.

The inability of STF models to store novel patterns suggests the existence of fast Hebbian plasticity. This is at odds with conventional wisdom, but supported experimentally [16–24]. For examples of Hebbian WM models, see Refs. [70,79–86]. Because of its Hebbian nature and ability to store novel patterns, freezable chaos aligns more with these proposals than with STF models. A crucial feature distinguishing freezable chaos from both STF and Hebbian WM models is that plasticity is deactivated, rather than activated, to store a pattern. Whereas other models require an external input carrying the pattern to be stored, this feature allows our model to store the neuronal state while it is engaged in strongly recurrent dynamics (in our random-network model, chaos).

Hinton and collaborators considered the possibility of a network performing a computation, saving its state in synapses, using neurons to perform a subroutine, and resuming computation from the saved state. This was termed "true recursion" [29,33,87]. Freezable chaos provides a minimal example of this: the neuronal state can be saved by halting plasticity, allowing neurons to engage in arbitrary dynamics before returning to the saved state. In our model, halting plasticity leaves a globally stable fixed point, so neuronal dynamics during the subroutine must be driven by external inputs. An interesting question is whether halting plasticity can leave the network with a fixed point that coexists with a dynamic regime that can be used for recurrent computation. This could be implemented in an *ad hoc* manner by turning on feedback loops upon halting plasticity.

This feature of freezable chaos suggests a method of detecting it *in vivo*, namely, by "interrupting" a task requiring strong recurrent dynamics, such as evidence integration, for variable periods of time [88]. Finding that neurons involved in the computation show continuity in their activities at the beginning and end of the interruption period would be suggestive of freezable dynamics. This conclusion would be further supported if task performance degrades when synaptic plasticity is disabled by genetic or pharmacological manipulations [22]. The activity expressed by the neurons during the interruption period would depend on whether and how they are recruited in this interval.

In other Hebbian WM models, an external neuronal or neuromodulatory signal is typically required to erase information stored in the synapses and return the network to a dynamic state. Freezable chaos avoids this requirement by leveraging fixed points that are stable with respect to neuronal, but not neuronal-synaptic dynamics. The presence or absence of a resetting signal could enable experimental disambiguation of our proposal.

Our model offers a mechanism for WM, but lacks a mechanism for long-term memory. This could be addressed

by introducing slow Hebbian dynamics into $\boldsymbol{J}$. For example, prior DMFT studies have taken $\boldsymbol{J}$ to be a static result of associative plasticity, resulting in various combinations of chaos and long-term memory retrieval [44,89–91]. Models like these could be extended by incorporating fast Hebbian synaptic dynamics atop this static structure. It would be particularly interesting if the short- and long-term dynamics could be made to interact, e.g., to implement memory consolidation. For example, during frozen chaos, if long-term plasticity was activated while the shorter-term plasticity of $\boldsymbol{A}(t)$ was disabled, the frozen state could be consolidated into $\boldsymbol{J}$.

Humans have a WM capacity of ~7 items, but freezable chaos can store just one item because synaptic plasticity is halted once a pattern is stored. One way of overcoming this limitation would be to use multiple $\boldsymbol{A}(t)$ matrices that can be independently modulated, corresponding to either different biophysical plasticity mechanisms or disjoint sets of synapses. Organisms might benefit from these different sets of synaptic variables possessing a hierarchy of timescales.

Large language models display a remarkable capacity for *in-context learning*: producing output that incorporates information or algorithms contained in the input [92]. This is surprising because, in both neuroscience and machine learning, such learning is generally thought to require weight updates. One possibility is that these models have such capabilities because they emulate weight dynamics through attention layers [35]. Nevertheless, explicit weight dynamics could benefit machine-learning models, for example, by enabling in-context learning with fewer model parameters [33] or on longer sequences [36]. An impediment to work in this direction is that weight dynamics are computationally costly. An important direction of work therefore pertains to ameliorating this burden, e.g., by formulating new forms of weight dynamics that are both expressive and have low computational complexity, or by exploring low-power neuromorphic architectures with dynamic weights.

APPENDIX A: EXTENSIONS AND LIMITS

Here, we derive how the single-site picture of Eq. (6) changes under various modifications of the synaptic dynamics of Eq. (3). We begin with two modifications that change the single-site dynamics simply by modifying the plasticity kernel:

$$K(\tau) = \frac{k}{p} e^{-\tau/p} C(\tau). \tag{A1}$$

(1) First, we take the limit $k, p \to \infty$ in Eq. (3) while keeping $k/p = \tilde{k}$ constant. This is equivalent to eliminating the decay term in the synaptic dynamics:

$$\partial_t A_{ij}(t) = \tilde{k} \phi_i(t) \phi_j(t). \tag{A2}$$

In this case, the kernel Eq. (A1) becomes

$$K(\tau) = \tilde{k} C(\tau). \tag{A3}$$

Thus, the decay timescale of the kernel is determined entirely though self-consistency.

(2) We next analyze the case where each synapse has a different time constant p that follows a heavy-tailed distribution with a power-law decay. We assume an inverse-gamma distribution, $f(p; \alpha, \beta) = N(\alpha, \beta) \times \exp(-\beta/p) p^{-(\alpha+1)}$, where $N(\alpha, \beta) = \beta^\alpha / \Gamma(\alpha)$ is a normalization constant with $\Gamma(\cdot)$ the gamma function. This distribution decays as $\sim 1/p^{\alpha+1}$ for large p and is exponentially suppressed as $p \to 0$ according to $\sim \exp(-\beta/p)$. In this case, the kernel of Eq. (A1) becomes

$$K(\tau) = \frac{k\alpha}{\beta} \left(\frac{\beta}{\beta + \tau}\right)^{\alpha+1} C(\tau). \tag{A4}$$

Thus, the kernel inherits the power-law decay of the distribution of time constants.

(3) We next consider a form of plasticity with a presynaptic dependence:

$$(1 + p\partial_t) A_{ij}(t) = \frac{k}{N} \phi_j(t). \tag{A5}$$

For this plasticity rule, the single-site dynamics are

$$(1 + \partial_t) x(t) = \eta(t) + I, \tag{A6}$$

where I is a time-independent input given self-consistently by

$$I = \frac{k}{p} \int_0^\infty d\tau e^{-\tau/p} C(\tau). \tag{A7}$$

(4) Next, by introducing a time delay d, we consider a form of temporally asymmetric Hebbian plasticity:

$$(1 + p\partial_t) A_{ij}(t)$$
$$= \frac{k}{N} [\phi_i(t) \phi_j(t - d) - \phi_i(t - d) \phi_j(t)]. \tag{A8}$$

This single-site problem for this plasticity rule is

$$
(1 + \partial_t)x(t)
$$
$$
= \eta(t) + \frac{k}{p}\int_{-\infty}^{t} dt' e^{-(t-t')/p}[C(t - t' + d)\phi(t')
$$
$$
- C(t - t')\phi(t' - d)]. \tag{A9}
$$

(5) Finally, we consider the case in which plasticity depends on arbitrary functions of pre- and post-synaptic activity, $f_{\text{pre}}(\cdot)$ and $f_{\text{post}}(\cdot)$,

$$
(1 + p\partial_t)A_{ij}(t) = \frac{k}{N}f_{\text{post}}(\phi_i(t))f_{\text{pre}}(\phi_j(t)). \tag{A10}
$$

In this case, the single-site dynamics are

$$
(1 + \partial_t)x(t)
$$
$$
= \eta(t) + \frac{k}{p}\int_{-\infty}^{t} e^{-(t-t')/p}C_{\text{pre},\phi}(t - t')f_{\text{post}}(\phi(t')), \tag{A11a}
$$

where $C_{\text{pre},\phi}(\tau) = \langle f_{\text{pre}}(\phi(t))\phi(t + \tau)\rangle_\eta.$ (A11b)

APPENDIX B: BISTABLE ATTRACTORS FOR FINITE N

For large p and k, the plastic network has two approximate dynamic solutions: a fast solution in which plasticity is averaged out, with dynamic timescale $\tau^* = \mathcal{O}(1)$, and a slow solution in which synapses drag neurons, with $\tau^* \gg p$. In finite-size networks, both behaviors can be
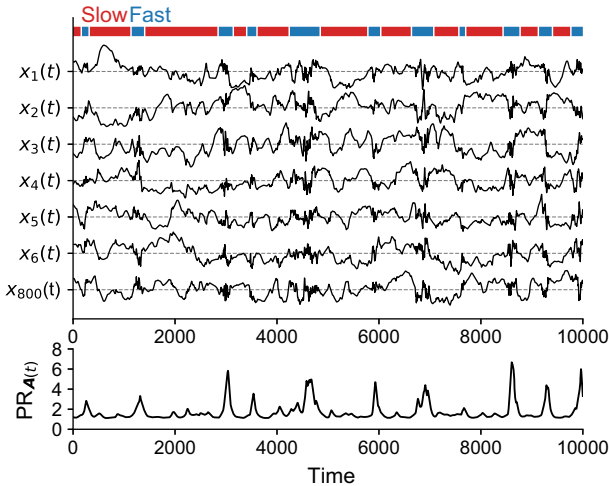

FIG. 8. Bistable chaotic attractors in a simulation with $g = 2$, $k = 1.1$, $p = 45$, and $N = 800$. Top: example neuronal traces $x_i(t)$. Bottom: participation ratio of $A(t)$ (as discussed in Sec. VI). "Fast" and "slow" states are defined by thresholding a low-pass filtered version of $\|\dot{x}\|$.

realized depending on initial conditions. Moreover, for appropriately tuned values of the parameters, the system can switch between these behaviors in a bistable manner (Fig. 8).

APPENDIX C: DYNAMICAL MEAN-FIELD THEORY NUMERICS

We solve the DMFT equations in this paper using iterative Monte Carlo methods [49–54].

1. Solution for $C(\tau)$

We first describe solving for $C(\tau)$. At each iteration, we sample a field $\eta(t)$ from a Gaussian process with zero mean and autocovariance $C(\tau)$. This is achieved by independently drawing Fourier coefficients from Gaussian distributions with the appropriate variances, while ensuring Hermitian symmetry. The sampled field $\eta(t)$ is then processed using the single-site dynamics given by Eq. (6) to produce $x(t)$ and consequently $\phi(t)$. An updated estimate of $C(\tau)$ is obtained from the empirical autocovariance of $\phi(t)$. This procedure is repeated until $C(\tau)$ converges. Multiple fields $\eta(t)$ are drawn and processed in parallel at each iteration.

2. Solution for $Q(t)$ and $D(\tau)$

The replica DFMT for freezable chaos follows a similar procedure, but is more complicated due to the requirement to sample correlated fields $\eta^A(t)$ and $\eta^B(t)$. Given that $C(\tau)$ is known, our task is to determine $Q(t)$ and $D(\tau)$. First, we sample $\eta^A(t)$ with the correct marginal statistics, namely, zero mean and autocovariance $C(\tau)$. Given $\eta^A(t)$, the conditional distribution for $\eta^B(t)$ is Gaussian with mean $\mu^B[\eta^A]$ and autocovariance $\Sigma^B(\tau)$, given by

$$
\mu^B[\eta^A] = \int dt \int dt' C^{-1}(t - t')Q(t)\eta^A(t'), \tag{C1a}
$$

$$
\Sigma^B(\tau) = g^2[D(\tau) - \delta], \tag{C1b}
$$

where $\delta = \int dt \int dt' C^{-1}(t - t')Q(t)Q(t').$ (C1c)

These integrals are straightforwardly evaluated in Fourier space. To find the dynamic solution where $D(\tau)$ decays as a function of τ, we sample $\eta^B(t)$ from this conditional distribution and process it through the single-site dynamics given by Eq. (32b) to obtain $x^B(t)$ and thus $\phi^B(t)$. We then update the estimates of $Q(t)$ and $D(\tau)$ based on the empirical statistics of $\phi^A(t)$ and $\phi^B(t)$. For fixed-point solutions with $D(\tau) = D$, low-variance estimates of $Q(t)$ and D can be efficiently obtained via numerical evaluation of Gaussian integrals using the conditional mean and variance. In both the fixed-point and dynamic cases, an effective increase in the number of samples is achieved by

shifting $\eta^A(t)$, which multiplies the conditional mean by a phase factor and leaves the conditional variance unchanged. Finally, similar to the solution for $C(\tau)$, multiple fields $\eta^A(t)$ are sampled during each iteration of the solver.

APPENDIX D: RELATING r AND THE DECAY OF $D(\tau)$

Given $\eta^A(t)$, the conditional distribution of $\eta^B(t)$ is Gaussian with mean Eq. (C1a) and autocovariance Eq. (C1b). Denote the integral term in Eq. (32b) by $I^B[\eta^A]$. Then, Eq. (32b) can be expressed as $x^B(t) = z(t) + \mu^B[\eta^A] + I^B[\eta^A]$, where $z(t)$ is a zero-mean Gaussian field with autocovariance $\tilde\Sigma^B(\tau) = g^2[\tilde{D}(\tau) - \delta]$. Here, $\tilde{f}(\tau)$ is related to $f(\tau)$ via

$$(1 - \partial_\tau^2)\tilde{f}(\tau) = f(\tau). \tag{D1}$$

The self-consistent condition for $D(\tau)$ becomes

$$D(\tau) = \left\langle \int \mathcal{D}u \left[\int \mathcal{D}x \phi \left(\sqrt{\tilde\Sigma^B(0) - \tilde\Sigma^B(\tau)}x \right.\right.\right.$$
$$\left.\left.\left. + \sqrt{\tilde\Sigma^B(\tau)}u + \mu^B[\eta^A] + I^B[\eta^A] \right) \right]^2 \right\rangle_{\eta^A}. \tag{D2}$$

The right-hand side depends on τ only through $\tilde\Sigma^B(\tau)$, which in turn depends on τ only via $\tilde{D}(\tau)$. Differentiating both sides of Eq. (D2) twice in τ and applying Price's theorem yields

$$\partial_\tau^2 D(\tau) = g^4 D^{\phi''}(\tau)[\partial_\tau \tilde{D}(\tau)]^2 + g^2 D^{\phi'}(\tau)\partial_\tau^2 \tilde{D}(\tau), \tag{D3}$$

where $D^{\phi^{(n)}}(\tau)$ is given by Eq. (D2) with $\phi(\cdot) \to \phi^{(n)}(\cdot)$. Setting $\tau = 0$ and using Eq. (D1) gives

$$\partial_\tau^2 D(\tau)|_{\tau=0} = r^2[\tilde{D}(0) - D(0)], \tag{D4}$$

noting that $\partial_\tau \tilde{D}(\tau)|_{\tau=0} = 0$ and $r^2 = g^2 D^{\phi'}(0)$. Taylor expanding $D(\tau)$ about $\tau = 0$, this can be expressed as

$$[(r^2 - 1)\partial_\tau^2 + r^2(\partial_\tau^4 + \partial_\tau^6 + \cdots)]D(\tau)|_{\tau=0} = 0. \tag{D5}$$

Assuming that $D(\tau)$ decays on a timescale $T \gg 1$, dimensional analysis implies that $\partial_\tau^n D(\tau)|_{\tau=0} = c_n/T^n$, where c_n are order-one coefficients. Keeping terms up to $1/T^4$ in Eq. (D5) results in

$$T = \sqrt{-\frac{c_4}{c_2}\frac{r}{\sqrt{r^2 - 1}}}, \tag{D6}$$

where $c_2 < 0$ and $c_4 > 0$ for a generic decaying autocovariance function with smoothness at $\tau = 0$. Thus, as $r \to 1^+$, T diverges as $T \sim 1/\sqrt{r - 1}$.

APPENDIX E: RANDOM-MATRIX THEORY

We study the spectrum of the Jacobian by applying a result of Ahmadian *et al.* [61] concerning a D-dimensional random matrix of the form $\mu + LNR$, where μ, L, and R are deterministic square matrices with L and R invertible. N is a random square matrix with entries drawn from $\mathcal{N}(0, 1/D)$. It was shown that the limiting spectral density of this matrix has support at a point $\lambda \in \mathbb{C}$ when $\lim_{D\to\infty}D^{-1}\|R(\mu - \lambda I_N)^{-1}L\|_F^2 \geq 1$. We can write $\tilde{M}_{\text{bulk}} = \mu + LNR$ by setting $D = 2N$ and choosing

$$\mu = \begin{pmatrix} -I_N & C(0)\text{diag}[\phi'(x)] \\ \frac{k}{p}I_N & -\frac{1}{p}I_N \end{pmatrix},$$

$$L = \begin{pmatrix} I_N & 0_N \\ 0_N & 0_N \end{pmatrix}, \qquad R = \sqrt{2}\begin{pmatrix} \text{diag}[\phi'(x)] & 0_N \\ 0_N & 0_N \end{pmatrix}.$$

In this parametrization, three of the four $N \times N$ blocks of the random matrix N have zero contribution to the result, reflecting the fact that the randomness of $\tilde{M}_{\text{bulk}}$, a $2N \times 2N$ matrix, is generated through J, an $N \times N$ matrix. Additionally, while L and R are singular, a violation of the assumptions of the theorem, one can add ϵI_N to each of L and R and safely compute the resulting spectral boundary curve, then take $\epsilon \to 0$ at the end. We proceed to compute the normalized Frobenius norm, yielding Eq. (21).

APPENDIX F: LYAPUNOV NUMERICS

For the maximum Lyapunov exponent heat maps in Figs. 5(a) and 9, the dynamics were run for time $T_{\text{sim}} = 850$ with a random perturbation of magnitude 10^{-2} applied at time 250.

Figure 9 shows how the heat map in Fig. 5(a) changes for different values of N. We find empirically that the boundary between regions of parameter space producing small, positive λ^{max} (dark blue) and convergence to stable nonzero fixed points (in at least 80% of simulations; hatched) is well fit by an N-dependent isocontour of the log-number of fixed points [Fig. 6(d)].

To compute the Lyapunov spectra in Fig. 5(b), we use the forward- and backward-pass method described in Sec. V B. In both the forward and backward directions, we computed 1600 exponents for networks of size $N = 900$. Dynamics were run for time $T_{\text{sim}} = 800$ with a burn-in period of duration 100. We used orthonormalization intervals of
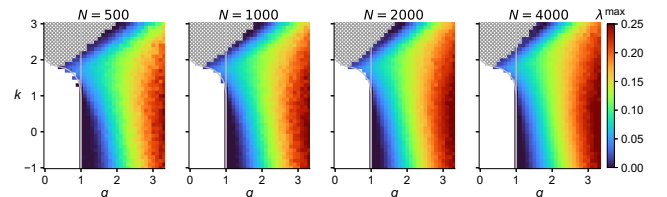

FIG. 9.　Same as Fig. 5(a) with different network sizes N.

20 and 5 for the forward and backward directions, respectively. For each setting of the parameters, we computed the spectra for five different network realizations and combined the results.

———————

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

[2] C. Pandarinath, D. J. O'Shea, J. Collins, R. Jozefowicz, S. D. Stavisky, J. C. Kao, E. M. Trautmann, M. T. Kaufman, S. I. Ryu, L. R. Hochberg *et al.*, *Inferring single-trial neural population dynamics using sequential auto-encoders*, Nat. Methods **15,** 805 (2018).

[3] S. Vyas, M. D. Golub, D. Sussillo, and K. V. Shenoy, *Computation through neural population dynamics*, Annu. Rev. Neurosci. **43,** 249 (2020).

[4] M. K. Benna and S. Fusi, *Computational principles of synaptic memory consolidation*, Nat. Neurosci. **19,** 1697 (2016).

[5] L. F. Abbott and W. G. Regehr, *Synaptic computation*, Nature (London) **431,** 796 (2004).

[6] L. F. Abbott, J. Varela, K. Sen, and S. Nelson, *Synaptic depression and cortical gain control*, Science **275,** 221 (1997).

[7] W. Maass and A. Zador, *Dynamic stochastic synapses as computational units*, in *Advances in Neural Information Processing Systems 10, NIPS 1997* (1997), https://papers.nips.cc/paper_files/paper/1997/hash/05311655a15b75fab86956663e1819cd-Abstract.html.

[8] O. Barak and M. Tsodyks, *Persistent activity in neural networks with dynamic synapses*, PLoS Comput. Biol. **3,** e35 (2007).

[9] G. Mongillo, O. Barak, and M. Tsodyks, *Synaptic theory of working memory*, Science **319,** 1543 (2008).

[10] D. V. Buonomano and W. Maass, *State-dependent computations: Spatiotemporal processing in cortical networks*, Nat. Rev. Neurosci. **10,** 113 (2009).

[11] Y. Mi, M. Katkov, and M. Tsodyks, *Synaptic correlates of working memory capacity*, Neuron **93,** 323 (2017).

[12] R. Penney, A. Coolen, and D. Sherrington, *Coupled dynamics of fast spins and slow interactions in neural networks and spin systems*, J. Phys. A **26,** 3681 (1993).

[13] R. Penney and D. Sherrington, *Slow interaction dynamics in spin-glass models*, J. Phys. A **27,** 4027 (1994).

[14] P. D. Roberts, *Dynamics of temporal learning rules*, Phys. Rev. E **62,** 4077 (2000).

[15] G. K. Ocker, A. Litwin-Kumar, and B. Doiron, *Self-organization of microcircuits in networks of spiking neurons with plastic synapses*, PLoS Comput. Biol. **11,** e1004458 (2015).

[16] B. Gustafsson, F. Asztely, E. Hanse, and H. Wigström, *Onset characteristics of long-term potentiation in the guinea-pig hippocampal CA1 region in vitro*, Eur. J. Neurosci. **1,** 382 (1989).

[17] R. C. Malenka, *Postsynaptic factors control the duration of synaptic enhancement in area ca1 of the hippocampus*, Neuron **6,** 53 (1991).

[18] R. C. Malenka and R. A. Nicoll, *NMDA-receptor-dependent synaptic plasticity: Multiple forms and mechanisms*, Trends Neurosci. **16,** 521 (1993).

[19] A. Volianskis and M. S. Jensen, *Transient and sustained types of long-term potentiation in the CA1 area of the rat hippocampus*, J. Physiol. **550,** 459 (2003).

[20] M. A. Erickson, L. A. Maramara, and J. Lisman, *A single brief burst induces GluR1-dependent associative short-term potentiation: A potential mechanism for short-term memory*, J. Cogn. Neurosci. **22,** 2530 (2010).

[21] A. Volianskis, N. Bannister, V. J. Collett, M. W. Irvine, D. T. Monaghan, S. M. Fitzjohn, M. S. Jensen, D. E. Jane, and G. L. Collingridge, *Different NMDA receptor subtypes mediate induction of long-term potentiation and two forms of short-term potentiation at CA1 synapses in rat hippocampus in vitro*, J. Physiol. **591,** 955 (2013).

[22] N. R. Driesen, G. McCarthy, Z. Bhagwagar, M. H. Bloch, V. D. Calhoun, D. C. D'souza, R. Gueorguieva, G. He, H.-C. Leung, R. Ramani *et al.*, *The impact of NMDA receptor blockade on human working memory-related prefrontal function and connectivity*, Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology **38,** 2613 (2013).

[23] P. Park, A. Volianskis, T. M. Sanderson, Z. A. Bortolotto, D. E. Jane, M. Zhuo, B.-K. Kaang, and G. L. Collingridge, *NMDA receptor-dependent long-term potentiation comprises a family of temporally overlapping forms of synaptic plasticity that are induced by different patterns of stimulation*, Phil. Trans. R. Soc. B **369,** 20130131 (2014).

[24] J. Lisman, *Glutamatergic synapses are structurally and biochemically complex because of multiple plasticity processes: Long-term potentiation, long-term depression, short-term potentiation and scaling*, Phil. Trans. R. Soc. B **372,** 20160260 (2017).

[25] A. Lansner, F. Fiebig, and P. Herman, *Fast Hebbian plasticity and working memory*, Curr. Opin. Neurobiol. **83,** 102809 (2023).

[26] J. J. Hopfield, *Neural networks and physical systems with emergent collective computational abilities*, Proc. Natl. Acad. Sci. U.S.A. **79,** 2554 (1982).

[27] T. Miconi, A. Rawal, J. Clune, and K. O. Stanley, *Backpropamine: Training self-modifying neural networks with differentiable neuromodulated plasticity*, in *Proceedings of the International Conference on Learning Representations, New Orleans, Louisiana, 2018*.

[28] T. Miconi, K. Stanley, and J. Clune, *Differentiable plasticity: Training plastic neural networks with backpropagation*, in *Proceedings of the International Conference on Machine Learning* (PMLR, Stockholmsmässan, 2018), pp. 3559–3568.

[29] G. E. Hinton and D. C. Plaut, *Using fast weights to deblur old memories*, in *Proceedings of the 9th Annual Conference of the Cognitive Science Society, 1987* (Cognitive Science Society, Seattle, 1987), pp. 177–186.

[30] T. Miconi, *Learning to acquire novel cognitive tasks with evolution, plasticity and meta-meta-learning*, in *Proceedings of the 40th International Conference on Machine Learning*, edited by A. Krause, E. Brunskill, K. Cho, B. Engelhardt, S. Sabato, and J. Scarlett (PMLR, 2023), Vol. 202, pp. 24756–24774, https://proceedings.mlr.press/v202/miconi23a/miconi23a.pdf.

[31] D. Tyulmankov, G. R. Yang, and L. F. Abbott, *Meta-learning synaptic plasticity and memory addressing for continual familiarity detection*, Neuron **110**, 544 (2022).

[32] N. Y. Masse, G. R. Yang, H. F. Song, X.-J. Wang, and D. J. Freedman, *Circuit mechanisms for the maintenance and manipulation of information in working memory*, Nat. Neurosc. **22**, 1159 (2019).

[33] J. Ba, G. E. Hinton, V. Mnih, J. Z. Leibo, and C. Ionescu, *Using fast weights to attend to the recent past*, Adv. Neural Inf. Process. Syst. **29**, 4331 (2016).

[34] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, *Attention is all you need*, Adv. Neural Inf. Process. Syst. **30**, 5998 (2017).

[35] I. Schlag, K. Irie, and J. Schmidhuber, *Linear transformers are secretly fast weight programmers*, in *Proceedings of the International Conference on Machine Learning, PMLR, 2021*, pp. 9355–9366.

[36] A. Katharopoulos, A. Vyas, N. Pappas, and F. Fleuret, *Transformers are RNNs: Fast autoregressive transformers with linear attention*, in *Proceedings of the International Conference on Machine Learning, PMLR, 2020*, pp. 5156–5165.

[37] J. Schmidhuber, *Learning to control fast-weight memories: An alternative to dynamic recurrent networks*, Neural Comput. **4**, 131 (1992).

[38] This equivalence arises due to the fact that *attention*, the core component of transformers, is based on forming linear combinations of neuronal states weighted by inner products between pairs of neuronal states; this can be mimicked by multiplying a neuronal state by a coupling matrix that has been shaped through activity-dependent plasticity to encode outer products of neuronal states.

[39] M. O. Magnasco, O. Piro, and G. A. Cecchi, *Self-tuned critical anti-Hebbian networks*, Phys. Rev. Lett. **102**, 258102 (2009).

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

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

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

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

[44] F. Mastrogiuseppe and S. Ostojic, *Linking connectivity, dynamics, and computations in low-rank recurrent neural networks*, Neuron **99**, 609 (2018).

[45] F. Schuessler, A. Dubreuil, F. Mastrogiuseppe, S. Ostojic, and O. Barak, *Dynamics of random recurrent networks with correlated low-rank structure*, Phys. Rev. Res. **2**, 013111 (2020).

[46] F. Schuessler, F. Mastrogiuseppe, A. Dubreuil, S. Ostojic, and O. Barak, *The interplay between randomness and structure during learning in RNNs*, Adv. Neural Inf. Process. Syst. **33**, 13352 (2020).

[47] I. D. Landau and H. Sompolinsky, *Coherent chaos in a recurrent neural network with structured connectivity*, PLoS Comput. Biol. **14**, e1006309 (2018).

[48] J. Baik, G. Ben Arous, and S. Péché, *Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices*, Ann. Probab. **33**, 1643 (2005).

[49] K. Krishnamurthy, T. Can, and D. J. Schwab, *Theory of gating in recurrent neural networks*, Phys. Rev. X **12**, 011011 (2022).

[50] F. Roy, G. Biroli, G. Bunin, and C. Cammarota, *Numerical implementation of dynamical mean field theory for disordered systems: Application to the Lotka–Volterra model of ecosystems*, J. Phys. A **52**, 484001 (2019).

[51] H. Eissfeller and M. Opper, *New method for studying the dynamics of disordered spin systems without finite-size effects*, Phys. Rev. Lett. **68**, 2094 (1992).

[52] H. Eissfeller and M. Opper, *Mean-field Monte Carlo approach to the Sherrington-Kirkpatrick model with asymmetric couplings*, Phys. Rev. E **50**, 709 (1994).

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

[54] F. Mignacco, F. Krzakala, P. Urbani, and L. Zdeborová, *Dynamical mean-field theory for stochastic gradient descent in Gaussian mixture classification*, Adv. Neural Inf. Process. Syst. **33**, 9540 (2020), https://proceedings .neurips.cc/paper_files/paper/2020/hash/6c81c83c4bd0b58 850495f603ab45a93-Abstract.html.

[55] M. Beiran and S. Ostojic, *Contrasting the effects of adaptation and synaptic filtering on the timescales of dynamics in recurrent networks*, PLoS Comput. Biol. **15**, e1006893 (2019).

[56] 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).

[57] V. L. Girko, *Circular law*, Theory Probab. Appl. **29**, 694 (1985).

[58] There is a more direct derivation of the spectrum of $\boldsymbol{M}$, though it does not provide an interpretation of the reduced $2N$-dimensional dynamics. We use the Schur-complement identity $|{}^{A\,B}_{C\,D}| = |\boldsymbol{D}||\boldsymbol{A} - \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C}|$ to write the characteristic polynomial of $\boldsymbol{M}$ as $p_M(\lambda) \propto (1 + p\lambda)^{S-N}|\lambda^2 p\boldsymbol{I}_N + \lambda[\boldsymbol{I}_N - p(\partial \boldsymbol{F}/\partial \boldsymbol{x})] - [(\partial \boldsymbol{F}/\partial \boldsymbol{x}) + k(\partial \boldsymbol{F}/\partial \boldsymbol{a})(\partial \boldsymbol{G}/\partial \boldsymbol{x})]|$. The first term, $(1 + p\lambda)^{S-N}$, encodes the eigenvalue $-1/p$ with multiplicity $S - N$. The second term encodes an N-dimensional quadratic eigenvalue problem (QEP). A QEP can be solved by linearization, which entails forming a $2N$-dimensional linear eigenvalue problem whose eigenvalues coincide with those of the QEP [59]. Applying the aforementioned identity to the determinant defining the characteristic polynomial of the reduced Jacobian $\tilde{\boldsymbol{M}}$ shows that $\tilde{\boldsymbol{M}}$ is one such linearization of the QEP. Thus, we recover the originally derived spectrum. There are infinitely many linearizations given by matrices related to $\tilde{\boldsymbol{M}}$ by similarity transformation, so in this sense, the reduced Jacobian is not unique.

[59] F. Tisseur and K. Meerbergen, *The quadratic eigenvalue problem*, SIAM Rev. **43**, 235 (2001).

[60] In chaotic states, this is a result of the network being locally destabilized by a continuous band of eigenvalues in analogy with the nonplastic network [41]. At fixed points, this is a result of the marginal stability of typical fixed points, which

have a continuous band of Jacobian eigenvalues brushing against the stability line, $\text{Re}(\lambda) = 0$.

[61] Y. Ahmadian, F. Fumarola, and K. D. Miller, *Properties of networks with partially structured and partially random connectivity*, Phys. Rev. E **91**, 012820 (2015).

[62] K. Geist, U. Parlitz, and W. Lauterborn, *Comparison of different methods for computing Lyapunov exponents*, Prog. Theor. Phys. **83**, 875 (1990).

[63] R. Engelken, F. Wolf, and L. F. Abbott, *Lyapunov spectra of chaotic recurrent neural networks*, Phys. Rev. Res. **5**, 043044 (2023).

[64] If the whole spectrum were computed, this spike would have $\mathcal{O}(N^2)$ exponents; here, there are fewer since only the ends of the spectrum are computed. Moreover, in principle it should be a delta-function spike, but in the numerics there is some "leakage" across modes such that the spike has finite width.

[65] J. L. Kaplan and J. A. Yorke, *Chaotic behavior of multidimensional difference equations*, in *Proceedings of the Conference on Functional Differential Equations and Approximation of Fixed Points, Bonn, 1978* (Springer, New York, 2006), pp. 204–227.

[66] G. Wainrib and J. Touboul, *Topological and dynamical complexity of random neural networks*, Phys. Rev. Lett. **110**, 118101 (2013).

[67] J. Stubenrauch, C. Keup, A. C. Kurth, M. Helias, and A. van Meegen, *The distribution of unstable fixed points in chaotic neural networks*, arXiv:2210.07877.

[68] S. L. Brunton, B. W. Brunton, J. L. Proctor, E. Kaiser, and J. N. Kutz, *Chaos as an intermittently forced linear system*, Nat. Commun. **8**, 19 (2017).

[69] D. J. O'Shea, L. Duncker, W. Goo, X. Sun, S. Vyas, E. M. Trautmann, I. Diester, C. Ramakrishnan, K. Deisseroth, M. Sahani, and K. V. Shenoy, *Direct neural perturbations reveal a dynamical mechanism for robust computation*, 10.1101/2022.12.16.520768.

[70] A. Sandberg, J. Tegnér, and A. Lansner, *A working memory model based on fast Hebbian learning*, Network **14**, 789 (2003).

[71] D. Martí, N. Brunel, and S. Ostojic, *Correlations between synapses in pairs of neurons slow down dynamics in randomly connected neural networks*, Phys. Rev. E **97**, 062314 (2018).

[72] K. Berlemont and G. Mongillo, *Glassy phase in dynamically-balanced neuronal networks*, 10.1101/2022.03.14.484348.

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

[74] L. Kozachkov, M. Lundqvist, J.-J. Slotine, and E. K. Miller, *Achieving stable dynamics in neural circuits*, PLoS Comput. Biol. **16**, e1007659 (2020).

[75] L. Kozachkov and J.-J. Slotine, *A note on matrix measure flows, with applications to the contraction analysis of plastic neural networks*, arXiv:2212.12639.

[76] M. G. Stokes, *'Activity-silent' working memory in prefrontal cortex: A dynamic coding framework*, Trends Cognit. Sci. **19**, 394 (2015).

[77] R. S. Zucker and W. G. Regehr, *Short-term synaptic plasticity*, Annu. Rev. Physiol. **64**, 355 (2002).

[78] Y. Wang, H. Markram, P. H. Goodman, T. K. Berger, J. Ma, and P. S. Goldman-Rakic, *Heterogeneity in the pyramidal network of the medial prefrontal cortex*, Nat. Neurosci. **9**, 534 (2006).

[79] C. Von Der Malsburg and W. Schneider, *A neural cocktail-party processor*, Biol. Cybern. **54**, 29 (1986).

[80] S. M. Polyn, K. A. Norman, and M. J. Kahana, *A context maintenance and retrieval model of organizational processes in free recall.*, Psychol. Rev. **116**, 129 (2009).

[81] F. Fiebig and A. Lansner, *A spiking working memory model based on Hebbian short-term potentiation*, J. Neurosci. **37**, 83 (2017).

[82] S. G. Manohar, N. Zokaei, S. J. Fallon, T. P. Vogels, and M. Husain, *Neural mechanisms of attending to items in working memory*, Neurosci. Biobehav. Rev. **101**, 1 (2019).

[83] F. Fiebig, P. Herman, and A. Lansner, *An indexing theory for working memory based on fast Hebbian plasticity*, eNeuro **7** (2020).

[84] Q.-S. Huang and H. Wei, *A computational model of working memory based on spike-timing-dependent plasticity*, Front. Comput. Neurosci. **15**, 630999 (2021).

[85] A. Bocincova, T. J. Buschman, M. G. Stokes, and S. G. Manohar, *Neural signature of flexible coding in prefrontal cortex*, Proc. Natl. Acad. Sci. U.S.A. **119**, e2200400119 (2022).

[86] L. Kozachkov, J. Tauber, M. Lundqvist, S. L. Brincat, J.-J. Slotine, and E. K. Miller, *Robust and brain-like working memory through short-term synaptic plasticity*, PLoS Comput. Biol. **18**, e1010776 (2022).

[87] Y. Bengio, Y. Lecun, and G. Hinton, *Deep learning for AI*, Commun. ACM **64**, 58 (2021).

[88] V. Mante, D. Sussillo, K. V. Shenoy, and W. T. Newsome, *Context-dependent computation by recurrent dynamics in prefrontal cortex*, Nature (London) **503**, 78 (2013).

[89] B. Tirozzi and M. Tsodyks, *Chaos in highly diluted neural networks*, Europhys. Lett. **14**, 727 (1991).

[90] U. Pereira and N. Brunel, *Attractor dynamics in networks with learning rules inferred from in vivo data*, Neuron **99**, 227 (2018).

[91] U. Pereira-Obilinovic, J. Aljadeff, and N. Brunel, *Forgetting leads to chaos in attractor networks*, Phys. Rev. X **13**, 011009 (2023).

[92] E. Akyürek, D. Schuurmans, J. Andreas, T. Ma, and D. Zhou, *What learning algorithm is in-context learning? Investigations with linear models*, in *Proceedings of the Eleventh International Conference on Learning Representations, Kigali Rwanda* (2023), https://openreview.net/forum?id=0g0X4H8yN4I.