# Theory of Coupled Neuronal-Synaptic Dynamics and Chaos

**A Review of Theory of Coupled Neuronal-Synaptic Dynamics and Chaos**

**David G. Clark and L. F. Abbott**

## Individual Study

Narges Khorshidi

Dr. Mohamad Reza Rahimi Tabar

Department of Physics at Sharif University of Technology

Summar 2025

**Topics:**

- Neuron dynamic equation
- Dynamical Mean Field Theory
- Jacobian spectrum
- Lyapunov spectrum
- Freezable Chaos

**General Description:**

In this lesson, we reviewed the article (David G. Clark and L. F. Abbott). The point of view of this article is to consider the synapse and neuron as coupled variables and to consider the synaptic weights consisting of two components, random and variable. With these definitions, the author tries to extract the dynamic equation of the neuron and synapse and then their coupling and, by analyzing the dynamic system, arrive at the definition of Freezable Chaos and examine its meaning with respect to the two concepts of learning and memory.

## Introduction

In neural circuits, it is not just the dynamics of neurons that matter; synaptic weights also change dynamically due to plasticity, creating independent degrees of freedom themselves. When we introduce “ongoing” Hebbian plasticity into a recurrent network, the behavior of the network changes. A positive Hebbian can create chaos in quiet networks or slow down activity in chaotic networks.
Synaptic dynamics are often divided into short-term plasticity, which operates on short timescales and depends on presynaptic activity , and long-term plasticity, which acts on much longer timescales and depends on both pre- and postsynaptic activity. However, short-term forms of Hebbian plasticity exist, suggesting that the timescale distinction is little more than a con- vention . 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. 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, intro- ducing presynaptic plasticity to this model simply adds an effective constant input to each neuron.
These claims are calculated and demonstrated with tools such as DMFT, Lyapunov analysis, Jacobian spectroscopy, and random matrices.

# 1.Neuron dynamic equation

## Hebbian Theory

Hebbian theory is a neuropsychological theory claiming that an increase in synaptic efficacy arises from a presynaptic cell's repeated and persistent stimulation of a postsynaptic cell. It is an attempt to explain synaptic plasticity, the adaptation of neurons during the learning process.
Consider a presynaptic neuron like i and a postsynaptic neuron like j. If we denote the weight of ij by Wij, then according to Hebb's law, we have the following relationship:

$$\Delta w_{ij} = \eta \, x_i \, y_j$$

that $\eta$ is learning rate. Hebb's law states that if neuron i is active and neuron j is also active, the synapse Wij also becomes positive and strengthened.
If the activity of neuron $i$ typically precedes that of neuron $j$, the pathway $i \to j$ plays a causal role in transmitting excitation. Strengthening $w_{ij}$ makes this pathway easier to trigger ("more clickable"), so similar patterns are more likely to become active in the future.
So The strength of the connection (synapse) between two neurons increases if they are repeatedly and simultaneously activated.
Now that we have an intuition about Hebb's law, we move on to defining the dynamic equations of neurons and synapses.

## Neuron and Synapse Dynamics

The neuron dynamic equation can be defined as follows:

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

Where X is the preactivation of neuron i, or we can say:

$$\dot{x}_i(t) = -x_i(t) + \sum_j W_{ij}(t) \, \phi_j(t).$$

We can say that the pre-activity of neuron i at time t or is all the inputs it receives. Also, $\phi$j is the output of neuron i, which is defined as the tangent, normalizing the value between 1 and -1.

$$\phi_j(t) = \phi\big(x_j(t)\big) = \tanh\big(x_j(t)\big).$$

This equation says The rate of change of the activity of neuron i plus its current value is equal to the weighted sum of the input activities it receives. This is a standard dynamics for neurons in recurrent network models.

Now divide synaptic connections into two parts: random connections and plastic connections.

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

J is a random constant from a distribution $J_{ij} \sim \mathcal{N}\big(0,\, g^2/N\big)$ is applied to it, that g scales the variance or strength of these random couplings and, in effect, adjusts the intensity of the network's random feedback. What does that mean? In fact, it is impossible to study all the neural connections exactly. So we assume that we have a basic initial state that comes from this distribution and is updated with a dynamic component like A.
From a biological perspective, slow-only weights maintain long-term memory and overall structure.
Fast activity-dependent weights temporarily write transient information into the weights themselves.
g controls the intensity of the disorder; the larger it is, the more disorder it causes. Dividing by N makes the function well-behaved in the limit towards infinity for taking the mean field.

Now we turn to the dynamic equation of synapses, which is a type of dynamic Hebbian equation!

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

Since A(t) is formed by averaging these outer products over time and decays with a time constant p, at any instant it keeps a memory of about p previous time steps.
In equation 3, positive k shows the system being ”Hebbian” meaning that two neurons firing at the same time will strengthen the synapse.
On the contrary systems with k < 0 are ”non-Hebbian”, meaning their asynchronous firing will strengthen the synapse! Also p is the time constant of the system's ”memory”, as the synapse will forget 63% of the activity of the neurons after a time p. There's a restriction on this constant that comes from biological reasons: neurons tend to forget faster than synapses. So since we took time constant of forgetting in neurons to be 1 (equation 1), then we have p>1.
Therefore the maximum rank of the matrix A(t) will be about p.
Now if we look at the neuronal traces xi(t) in different parameters (k, g) and in constant p (figure 1.b), we will have some intuitions about the evolution of the system in different phases. When the plasticity is off, like in point (i) in parameter space with k=0, the system shows a chaotic behavior in time. Ultimately if the system will be non-Hebbian, as in point (iii) with k<0, this chaotic behavior would strengthen! But for (k>0, g) the system shows a noticable characteristics: When we halt the system, meaning we freeze all synapses and won't let them

evolve anymore, the neuronal behavior can totally change.
For small k, if the system still shows a chaotic behavior, we call it a nonfreezable chaos. Then for larger k, If the activity would decrease significantly, but not have reached a fixed point, then it's a semifreezable chaos as in the diagram of point (vii). Finally for significantly large k, if the neuronal activity would also freeze quickly after the halt in synapses activity, then we have freezable chaos (point (vii)). In all the three cases the neuronal activity would come back to normal after we "release" the synopsis dynamics again.
There is also a phase shift around g=1. For g<1 the system need a very high plasticity (k»0) so we can see the signatures of chaos (point (iv)). So for small k you only see the trivial fixed point of the system (point (v)). For g being slightly higher than 1 there exists an interesting dynamics in which the system go through a short chaotic period before collapsing into it's fixed points.

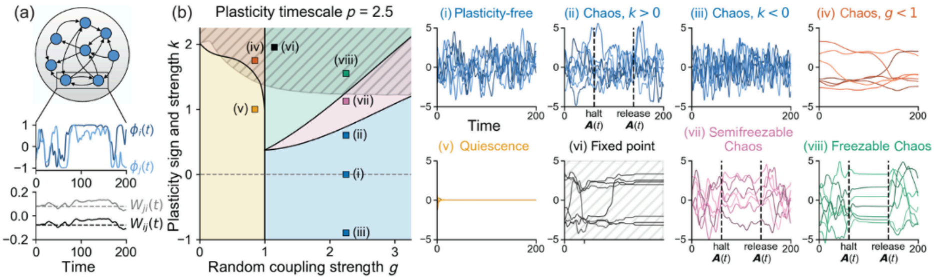

Figure 1: (a) Dynamics of a pair of neurons (top) and of the synapses through which they are reciprocally coupled (bottom). Synapses fluctuate about quenched random strengths (dashed lines) in response to pre- and postsynaptic activity according to a Hebbian rule.(b) Left: phase diagram of the plastic network for p=2.5. Right: example neuronal traces xi(t) from simulations of each phase-diagram region, with parameters given by the location of the associated square marker.

## 2. Dynamical Mean Field Theory

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:

$$C(\tau) = \langle \phi_i(t)\, \phi_i(t+\tau) \rangle_J. \tag{4}$$

C($\tau$) measures how well a neuron's activity at time t correlates with its own activity at time t+$\tau$.
Integrating the synaptic dynamics, Eq. (3), and inserting this into the neuronal dynamics, Eq. (1), gives.
So this is where the idea of the synchrony of neurons and synapses becomes tangible!

At first we solve the synaptic equation(Eq3):

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

Which is a linear differential equation and by substituting the Green's function we have:

$$e^{-(t-t')/p} = G(t-t').$$

Then

$$A_{ij}(t) = \frac{k}{N}\int_{-\infty}^{t} dt'\; e^{-(t-t')/p}\,\phi_i(t')\,\phi_j(t').$$

tprime is the past time.
Now we substitute Aij into equation 1:

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

$$W_{ij}(t) = J_{ij} + A_{ij}(t).$$

$$(1+\partial_t)\,x_i(t) = \sum_j J_{ij}\,\phi_j(t) + \sum_j A_{ij}(t)\,\phi_j(t)$$

We replace Aij.

$$= \sum_j J_{ij}\,\phi_j(t) + \sum_j \left[\frac{k}{N}\int_{-\infty}^{t} dt'\; e^{-(t-t')/p}\,\phi_i(t')\,\phi_j(t')\right]\phi_j(t). \qquad (5)$$

The Green's function term here is the decay kernel. That is, it shows that memory can decrease over time as the constant p increases.
$\sum_j J_{ij}\,\phi_j(t)$ The instantaneous input from the connections is fixed and
$\frac{1}{N}\sum_j\,\phi_j(t)\,\phi_j(t')$ This is the average correlation of the activity of the entire network between time t' (past) and t (present). It essentially tells us how similar the network is to itself. If the network is similar in the past (t') and the present (t), this value is large.
$\phi_j(t')$This is the interesting part. the past activity of neuron i itself means that plasticity creates a feedback connection from the neuron itself to itself.
When $N \to \infty$, according to the law of large numbers and the central limit theorem:

$$\sum_j J_{ij}\,\phi_j(t) = \eta(t).$$

This is an effective Gaussian field that has replaced all the complex inputs from other neurons.

If we look at this term of the equation:

$$\Big[\frac{1}{N}\sum_j \phi_j(t)\,\phi_j(t')\Big]\;\phi_i(t')$$

We find that according to the definition of equation 4, this is the correlation between the past time t' and the present time t. so we have:

$$\frac{1}{N}\sum_j \phi_j(t)\,\phi_j(t')\;\to\;\langle\phi(t)\,\phi(t')\rangle\;=\;C(t-t').$$

By substituting this expression and the Gaussian field expression in $N\to\infty$ in Equation 5, we have:

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

And as for the discovery of the statistical property of $\eta$, we will have:

$$=\Big\langle\Big(\sum_j J_{ij}\,\phi_j(t)\Big)\Big(\sum_k J_{ik}\,\phi_k(t{+}\tau)\Big)\Big\rangle=\sum_j\sum_k\langle J_{ij}J_{ik}\rangle\,\langle\phi_j(t)\,\phi_k(t{+}\tau)\rangle=\langle\eta(t)\,\eta(t{+}\tau)\rangle$$

$$\langle\eta(t)\eta(t+\tau)\rangle=\sum_j\frac{g^2}{N}\,\langle\phi_j(t)\phi_j(t+\tau)\rangle,$$

when $N\to\infty$:

$$\frac{1}{N}\sum_j\langle\phi_j(t)\phi_j(t+\tau)\rangle\;\to\;C(\tau),$$

$$\Rightarrow\qquad\langle\eta(t)\eta(t+\tau)\rangle=g^2\,C(\tau). \tag{7}$$

The DMFT is closed by the self-consistency condition:

$$C(\tau)=\big\langle\phi(t)\,\phi(t+\tau)\big\rangle_{\eta}. \tag{8}$$

## 3. Jacobian spectrum

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 dynam- ics, 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 tran- sition 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 cou- plings, with neurons serving as the connections.

I write another form of equations 1-3 with the vector definition

$$\partial_t x(t) = F\big(x(t), a(t)\big), \tag{8}$$

that x(t), The state vector of the neurons at time t.

$$x(t) = \big(x_1(t), \, x_2(t), \, \ldots, \, x_N(t)\big)^{\mathsf{T}}.$$

where xi(t) can represent the membrane potential or the activity rate of neuron i.
a(t)=vecA(t): vectorize all $S = N^2$ elements of the synaptic connection matrix A(t). This means we dump all the synaptic weights into a long column vector.
This function models how the current state of neurons (x) and synaptic weights (a) affect the dynamics of neurons.

$$\big(1 + p\,\partial_t\big)a(t) = k\,G\big(x(t)\big). \tag{9}$$

$\mathbf{a}(t) = \operatorname{vec} A(t)$:
vector containing all synaptic connection entries; $S = N^2$ is the number of elements of the connectivity matrix (i.e., the length of a).
$p$: timescale of synaptic plasticity.
$\partial_t \mathbf{a}(t)$: time derivative of the synaptic weights. $k$:plasticity gain (strength of plasticity).
$G(x(t))$: vector function specifying the rate of synaptic weight change as a function of neuronal activity.
This function specifies how the activity of neurons drives the change in synapses.
In equation (3):

$$\partial_t A_{ij}(t) = \frac{1}{N}\,\phi\big(x_i(t)\big)\,\phi\big(x_j(t)\big) - A_{ij}(t)$$

so we can wtire:

$$G_{ij}\big(x(t)\big) = \frac{1}{N}\,\phi\big(x_i(t)\big)\,\phi\big(x_j(t)\big)$$

Which is a Hebbian rule: synapses are strengthened based on the co-activity of the pre- and post-synaptic neuron.
Solving equation (9) shows that synapses:
1.Relax with time constant p towards the target value kG(x(t))
2. The target value is determined by the instantaneous activity of the neurons G(x(t)))
3. Apply a low-pass filter to the neuronal activity.
Now we can see the entire system as a single dynamical system:

$$\partial_t \begin{pmatrix} x(t) \\ a(t) \end{pmatrix} = \begin{pmatrix} F\big(x(t), a(t)\big) \\ \frac{k}{p}\,G\big(x(t)\big) - \frac{1}{p}\,a(t) \end{pmatrix}.$$

We can calculate the Jacobian matrix of the entire system:

$$M = \begin{pmatrix} \frac{\partial \dot{x}}{\partial x} & \frac{\partial \dot{x}}{\partial a} \\ \frac{\partial \dot{a}}{\partial x} & \frac{\partial \dot{a}}{\partial a} \end{pmatrix} = \begin{pmatrix} \frac{\partial F}{\partial x} & \frac{\partial F}{\partial a} \\ \frac{k}{p}\,\frac{\partial G}{\partial x} & -\frac{1}{p}\,I_S \end{pmatrix}. \tag{10}$$

Neuronal Front (F) Wants to keep the system chaotic via J (asymmetric random matrix)
The synaptic front (G) seeks to stabilize the system through memory formation and self-coupling.
When we halt the synapses ($\partial_t a = 0$):

$$(1 + p\,\partial_t)a(t) = k\,G\big(x(t)\big) \;\Rightarrow\; a(t) = \text{constant}$$

$$\partial_t x(t) = F\big(x(t), a(t)\big)$$

That is: fixed synapses can create a stable fixed point. so, In the normal regime, $a(t)$ keeps changing $\Rightarrow$ fixed points destabilize.
When $a(t)$ is held constant: fixed points remain stable $\Rightarrow$ chaos "freezes".
Term 22 The M matrix shows the effect of synapses on synapses. This block asks: "If I slightly perturb one synapse, how do the other synapses react?" Simple answer: "No effect!"
$I_S$ (identity matrix) $\Rightarrow$ synapses are independent.
All synapses relax back to rest with rate $1/p$ (eigenvalue $-1/p$); they simply decay—i.e., synapses are independent of each other.
The M matrix is important in distinguishing between chaos and stability.
If $\Re\lambda > 0$ for some eigenvalue of $M$, perturbations grow (chaos); if $\Re\lambda < 0$ for all eigenvalues, perturbations decay (stability).
Now let's find fixed points of the system. First let's define this parameter as a measure to see whether the neuronal system is chaotic or fixed:

$$\mathrm{PR}_{A(t)} = \frac{\big[\sum_i \lambda_i(t)\big]^2}{\sum_i \lambda_i^2(t)} = \frac{\big[\operatorname{tr} A(t)\big]^2}{\|A(t)\|_F^2}. \tag{11}$$

This quantity can also be seen as an average from DMFT plus fluctuations around it. For $N \to \infty$ we can find:

$$\operatorname{tr} A(t) = \frac{k}{p}\int_0^\infty e^{-\tau/p}\,\frac{1}{N}\sum_i \phi_i(t-\tau)^2\,d\tau = \frac{k}{p}\int_0^\infty e^{-\tau/p} C(0)\,d\tau = k\,C(0).$$

and

$$\begin{aligned}
\|A(t)\|_F^2 &= \sum_{ij} A_{ij}(t)^2 \\
&= \Big(\frac{k}{p}\Big)^2 \int_0^\infty\int_0^\infty e^{-(\tau+\tau')/p}\Big[\frac{1}{N}\sum_i \phi_i(t-\tau)\phi_i(t-\tau')\Big]^2 d\tau d\tau' \\
&= \Big(\frac{k}{p}\Big)^2 \int_0^\infty\int_0^\infty e^{-(\tau+\tau')/p} C(\tau-\tau')^2 d\tau d\tau' \\
&= \Big(\frac{k}{p}\Big)^2 p\int_0^\infty e^{-s/p} C(s)^2 ds.
\end{aligned} \tag{12}$$

now by combining everything

$$\mathrm{PRA}(t) = \frac{[k\,C(0)]^2}{\big(\frac{k}{p}\big)^2 p\int_0^\infty e^{-s/p} C(s)^2 ds} = \frac{p\,C(0)^2}{\int_0^\infty e^{-s/p} C(s)^2 ds} = \frac{p}{\displaystyle\int_0^\infty e^{-s/p}\Big(\frac{C(s)}{C(0)}\Big)^2 ds}.$$

or equivalently

$$\text{PRA}(t) = \frac{p}{\mathcal{T}} \qquad \text{with } \mathcal{T} \equiv \int_0^\infty e^{-s/p} \left(\frac{C(s)}{C(0)}\right)^2 ds, \tag{13}$$

As it is shown in the result, $PR_A(t)$ is independent as we go to N→∞ and is equal to the time constant p divided by $\mathcal{T}$. $\mathcal{T}$ can be something between 0 and p, so $PR_{A(t)}$ will always be greater than 1.

Also as the plasticity increase, i.e as k increase, the fluctuations will be reduced and $PR_A(t)$ again get close to 1. These all show that $PR_A(t)$ is a measure to see whether we have reached a fixed point or not. When it is high (for any reason: small N or small k) it means that the neuronal system is evolving chaotically, but when it approaches to 1, it means that we are reaching a fixed point. We should also notice that there is not a single fixed point in the system. For example as k→∞, any configuration with only on/off neurons would be a steady state, counting $2^N$ fixed points.

## 4. Lyapunov spectrum

Rigorously characterizing the nonlinear dynamics requires a calculation of the spectrum of Lyapunov exponents.
A positive maximal exponent $\lambda$max >0 signals exponential sensitivity and hence chaos. Because the Jacobian is non-normal and its eigenvectors evolve in time, the Lyapunov spectrum cannot be inferred directly from eigenvalues of M; it must be measured dynamically.

For g < 0, the authors realized dynamic network states using the deformation method described in Sec. For each setting of (g, k) , authors simulated 200 random networks. The output of this analysis is displayed as a heat map in (g, k) parameter space in Fig. 2(a). Parameter values that resulted in convergence to nonzero fixed points 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 [Fig.2(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.2(a), solid-to-hatched boundary]. This cross- over occurs in a parameter regime for which phase space is densely filled with stable fixed points . Using the present finite-N analysis, we are unable to determinewhether  max in the hatched region in Fig.

1(a) is small and positive, or negative. Additionally, solving the DMFT in the hatched region is numerically difficult due to the diverging dynamic timescale . As N is increased over a decade, the boundary marking this crossover shifts slightly upward (Fig. 3).

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.
Generally, 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 or vanishing and to extract their growth or decay rates. The plastic network has N + N$^{2}$ variables, making propagating a complete basis prohibitively computationally expensive for large N. Fortunately, O(N$^{2}$) Lyapunov exponents concentrate at -1/p; therefore, it suffices to compute the O(N) largest and smallest exponents. The largest exponents are computed using the aforementioned method with an undercomplete set of tangent-space vectors. The smallest exponents are found by applying the same method to the time-reversed dynamics, noting that the smallest exponents of the original system are the largest exponents of the time-reversed system. A complication is that the time-reversed dynamics are unstable; therefore, the time-reversed tangent-space dynamics—tamed by orthonormalization—are run on a time-reversed trajectory generated by the forward-time dynamics. This method's accuracy has been verified for non-plastic networks.

Histograms of the Lyapunov spectra show that for k=0, the spectrum is identical to that of a non-plastic network with a spike at -1/p. The measurement of $\lambda$max obtained using this method also matches the perturbation measurement displayed in the heatmap. For large k, the Lyapunov spectra recapitulate the topological transition into two bands—similar to the Jacobian spectra—further confirming the existence of a synapse-driven dynamic state. By analogy with the Jacobian spectra, the slow, destabilizing Lyapunov band spans from -1/p to $\lambda$max, and the fast, relaxational band has an upper limit near -1.

Calculating the Lyapunov spectrum enables the computation of diffeomorphic-invariant properties of the strange attractor, including its dimension. The attractor dimension is defined by the Kaplan-Yorke conjecture as the number of exponents that must be summed in descending order to achieve a cumulative sum of zero and is displayed as an intensive quantity (divided by N). The eventual decrease in the attractor dimension is reminiscent of the behavior of the participation ratio-based activity dimension in networks with partially symmetric connectivity, which was recently computed analytically. The decline in $\lambda$max and the attractor dimension at large k likely reflects the proliferation of stable fixed points throughout phase space.

(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 $g$.

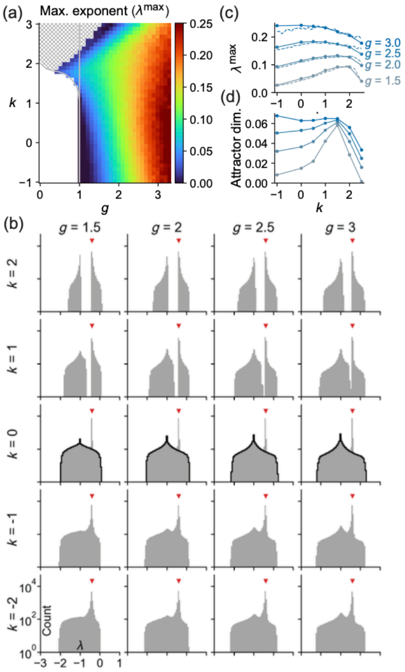

Figure 2: a: Maximum Lyapunov exponent and b: convergence

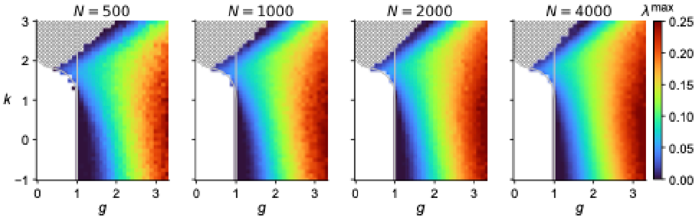

Figure 3: Same as Fig. 1(a) with different network sizes N

# 5. Freezable Chaos

The 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.

Freezable chaos is a dynamical state in a neural network with synaptic plasticity where a stable fixed point of the neuronal activity can be destabilized by the ongoing synaptic dynamics, leading to chaotic behavior.

The core idea is to classify chaotic states based on what happens when synaptic plasticity is artificially halted at a specific time (t=0):

Nonfreezable Chaos: After halting synapses, the neuronal activity remains chaotic.

Semifreezable/Freezable Chaos: After halting synapses, the neurons settle into a stable fixed point that retains a memory of the neuronal state at the halt time.

A crucial factor enabling this is the alignment between the static, random part of the connectivity (J) and the plastic, low-rank part (A(0)). This alignment arises naturally because the plasticity rule (e.g., Hebbian) causes A(t) to be shaped by the neuronal activity, which is itself driven by J.

The importance of this alignment is demonstrated by a key numerical experiment:

Synaptic plasticity is halted, creating a stable neuronal fixed point.

This shuffling destroys the fixed point, and the network reverts to chaos, proving that the specific correlation between J and A(0) is essential for the existence of the memory.

When plasticity is restored, A(t) adapts to the new (shuffled) J, and halting it again at a later time creates a new stable fixed point.

In essence, freezable chaos is a state where the continuous interaction between neuronal activity and synaptic plasticity hides stable ”memory” states. These memories are only revealed and become stable when the plastic synapses are frozen, and their existence relies on a specific correlation structure between the static and plastic components of the network’s connectivity.

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

$$W_{ij}(0) = J_{ij} + \frac{k}{p}\int_0^\infty dte^{-t/p}\phi_i^A(-t)\phi_j^A(-t) \tag{14}$$

This equation describes how the final connection matrix W(0) is formed when we stop plastic synapses at time t=0.

This integral shows how much each pair of neurons has been active simultaneously. On the other hand, more recent experiences have more weight and the entire history of activity plays a role in the final connections.

For example, if two neurons i and j are always active together: The integral and connections Wij become strong.

Now we will have two equations that can tell us how much memory is retained and how it is retained:

$$Q(t) = \lim_{t'\to\infty} \left\langle \phi_i^A(t)\phi_i^B(t') \right\rangle_J . \tag{28}$$

$$D(\tau) = \lim_{t\to\infty} \left\langle \phi_i^B(t)\phi_i^B(t+\tau) \right\rangle_J , \tag{29}$$

$Q$ represents the average product of corresponding activities in two networks. If they are completely similar, $Q \approx 1$, and if they are completely independent, $Q \approx 0$.
If D($\tau$) = constant, Chaos is stabilizable and the network reaches a fixed point.
If $D(\tau) \rightarrow$ 0 constant for big $\tau$: The chaotic network remains, and the memory is gone and the chaos is unfreezeable.
On the other hand, to examine the system with fixed points, we have an equation that calculates the radius of the spectrum of the eigenvalues of the Jacobian matrix:

$$r^2 = g^2 \langle \phi'(x_i^B)^2 \rangle_J \tag{15}$$

If r < 1: System is stable

If r > 1: System is unstable

If r = 1: Marginal stability.

and The stronger the g, the greater the tendency to instability.
In non-freezing chaos, r is undefined. In semi-freezing chaos, Q > 0 and $D(\tau) \rightarrow 0$ and $r > 1$ *to unstablefixedpoint.*
and in freezable chaos, Q>0 and D($\tau$) = constant and $r < 1 \rightarrow$ *stablefixedpoint.*

Now we examine the network dynamics in two cases with dynamic synapses and stalled synapses.

$$(1 + )x^A(t) = \eta^A(t) + \frac{k}{p}\int_{-\infty}^{t} dt' e^{-(t-t')/p} C(t-t')^A(t') \tag{16}$$

This equation shows that each neuron is influenced by random input from a synaptic feedback network that remembers its own activity history.

$$(1 + )x^B(t) = \eta^B(t) + \frac{k}{p}\int_{0}^{t} dt' e^{-(t-t')/p} Q(t-t')^B(t') \tag{17}$$

Then

$$\left\langle \begin{pmatrix} \eta^A(t) \\ \eta^B(t) \end{pmatrix} \begin{pmatrix} \eta^A(t') & \eta^B(t') \end{pmatrix} \right\rangle = g^2 \begin{pmatrix} C(t-t') & Q(t) \\ Q(t') & D(t-t') \end{pmatrix} \tag{18}$$

The definition of the correlation structure between the effective noises in two networks A and B is what encodes the "alignment" between the networks.
Network B is only affected by random input (independent of ongoing learning) and embedded memory in stalled synapses.

In mean field theory, $\eta$A(t) and $\eta$B(t) are not just simple noise - they are collective representations of all network interactions.
The non-diagonal element does not depend on t - t', but only on one of the times (e.g. t or t'). This usually occurs in non-stationary systems or in models with asymmetric time memory, because:

$$\langle \eta^A(t)\, \eta^B(t')\rangle = g^2 Q(t), \langle \eta^B(t)\, \eta^A(t')\rangle = g^2 Q(t').$$

$Q(t) \equiv \langle \phi^A(t)\, \phi^B(t)\rangle$ is the *overlap* between the neuronal states of the two replicas (Eq. 28). Its value at the halting time, $Q(0)$, measures *how much of the halted neuronal state is remembered.* $D(\tau) \equiv \langle \phi^B(0)\, \phi^B(\tau)\rangle$ is the autocovariance in replica **B**; in the *fixed–point* ansatz we set $D(\tau) \equiv D$ (constant). $\varepsilon \equiv g - 1 > 0$ quantifies how far we are above the nonplastic stability threshold $g = 1$. $k > 0$ controls Hebbian strength.

The reduced single-site equations near $g \to 1^+$ are

$$(1 + \partial_t)\, x^A(t) = \eta^A(t) + k\, C(0)\, x^A(t), \tag{19}$$

$$x^B(t) = \eta^B(t) + k\, Q(0)\, x^A(0). \tag{20}$$

Here $\eta^{A,B}$ are zero-mean Gaussian fields with covariances given by $\mathrm{Var}(\eta^B) = g^2 C(0)$.
average over $\eta^A, \eta^B$.

$$\underbrace{\big\langle (x^B)^2 \big\rangle}_{=D} = \underbrace{\big\langle (\eta^B)^2 \big\rangle}_{=g^2 C(0)} + k^2 Q(0)^2 \underbrace{\big\langle \big(x^A(0)\big)^2 \big\rangle}_{=C(0)}.$$

Near $g \to 1^+$ one can consistently express $C(0)$ in terms of $D$ and $\varepsilon$; to leading order this gives

$$g^2 C(0) = D - \varepsilon\, C(0) \qquad \Longrightarrow \qquad D\big(D - \varepsilon\big) = k^2\, Q(0)^2.$$

Solving for $Q(0)$ yields the order-parameter equation

$$Q(0) = \pm \frac{\sqrt{D\,\big[D - \varepsilon\big]}}{k}. \tag{21}$$

At the freezable-chaos threshold one finds $r \to 1+$ from the dynamic side, while the fixed-point and dynamic solutions merge at r=1.
So Eq. (15) is the *memory–overlap law*: it ties the retained overlap at the halting time $Q(0)$ to the fixed-point variance $D$, the distance from criticality $\varepsilon = g - 1$, and the Hebbian gain $k$.

By multiplying equations 13 and 14, then taking multiple limits and solving the finite differential equation, we will arrive at an equation that has interesting results:

$$(1 + \partial_t)\, x^A(t) = \eta^A(t) + k\, C(0)\, x^A(t)$$

$$x^B(t) = \eta^B(t) + k\, Q(0)\, x^A(0)$$

After averaging and applying the limit $g \to 1+$ we have:

$$\left\langle k\, C(0)\, x^A(t) \cdot k\, Q(0)\, x^A(0) \right\rangle = k^2 C(0) Q(0) \left\langle x^A(t)\, x^A(0) \right\rangle = k^2 C(0) Q(0)\, C(t)$$

By adding the principal terms and ignoring the higher order terms, we arrive at the following differential equation:

$$(1 + \partial_t)\, Q(t) = Q(t) + k\, C(0)\, Q(t) + k^2 C(0)\, Q(0)\, C(t)$$

in limit D $\approx \epsilon$ and with:

$$f(t) = k\, Q(0)\, C(t)$$

And by solving the differential equation:

$$Q(t) = e^{-\lambda t} \left[ \int_{-\infty}^{t} e^{\lambda t'} f(t')\, dt' + \text{constant} \right]$$

By substituting f(t) and considering the boundary condition that Q(t) is bounded as $t \to\ \infty$, we arrive at the final solution

$$Q(t) = k\, Q(0) \int_{-\infty}^{t} dt'\ e^{-(D-\varepsilon)(t-t')}\, C(t'). \tag{22}$$

1. **Memory as a causal filter (Causal Filter).**

   - Eq. (16) shows that $Q(t)$ is a causal filter of the network's autocovariance $C(t)$.
   - The integral runs only over past times ($t' \leq t$); memory is built solely from past information.

2. **Peak at $t > 0$ (future prediction).**

   - If $C(t)$ is an even function around zero (e.g., a decaying even kernel), the causal filter causes $Q(t)$ to peak at positive lags $t > 0$.
   - This is a surprising, counterintuitive result: memory predicts the network's future state better than the state at the halting instant.
   - In practice this is evident in the paper: the curves of $Q(t)$ for various $k$ exhibit a clear maximum at $t > 0$

3. **Physical interpretation—network "momentum".**

   - The effect implies the network has momentum/inertia. When synapses are halted at $t = 0$, the network tends to continue along its prior trajectory for a short time before becoming trapped at a fixed point.
   - Consequently, the eventual fixed point is closer to the state the live network (B) was heading toward than to the exact halted state of network A at $t = 0$.

By setting $t = 0$ and using the analytical form of $C(t)$ near $g \to 1^+$ (from DMFT), the above integral can be evaluated explicitly. As a result, after substituting $Q(0)$ from (15) and eliminating $Q(0)$, one obtains a scalar equation relating $D$, $\varepsilon$, and $k$, which is written as this:

$$h(u,k) = \frac{2}{\sqrt{3}} \left[ \psi\left( \frac{\sqrt{3}(1-k)(u-1)+3}{4} \right) - \psi\left( \frac{\sqrt{3}(1-k)(u-1)+1}{4} \right) \right]^{-1}. \tag{23}$$

- Using $D$ from (17) and $Q(0)$ from (15), the spectral radius of the fixed-point solution (for network B), to first order in $\varepsilon$, is $r = \varepsilon + D - 1$.
- At the threshold, since $D = \varepsilon$ (i.e., $u = 1$), we have $r = 1$: the dynamical solution and the fixed–point solution meet at $r = 1$. For $k < k_c$, Eq. (17) gives $\varepsilon < D$, and therefore $r > 1 \Rightarrow$ the fixed point is stable and "freezable chaos" is realized.
- From a physical viewpoint, (17) is the self–consistency condition for memory: the Hebbian gain $k$ must exactly balance $D - \varepsilon$ in the causal filter (Eq. 16) so that $Q(0)$ remains nonzero and the fixed–point statistics (i.e., $D$) are preserved.

$\frac{2}{\sqrt{3}} \approx 0.37$ :

Quantitatively, this equation and the function $h(u,k)$ predict the boundaries between different phases:
For $k < 0.37$: only the trivial solution $Q(0) = 0$ exists, $\Rightarrow$ unfreezable chaos.

For $k > 0.37$: a nontrivial solution $Q(0) \neq 0$ appears, $\Rightarrow$ semi-freezable or freezable chaos.

When $k$ crosses 0.37, the system undergoes a continuous phase transition to memory-holding states.

This equation predicts the critical threshold for memory formation (k  0.37) and provides a quantitative description of how memory emerges from disorder by providing a self-consistency relation for calculating the order parameters and determining the stability of different solutions.

This analytical analysis in $g \to 1+$ provides deep insight into the fundamental mechanisms of memory formation in plastic neural networks and provides a theoretical basis for understanding numerical results and simulations.

## Discussion

This study characterized the dynamics of N neurons coupled to N2 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. 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 , 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 . 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 or backpropagation.

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 under- mined by others showing“activity-silent” WM. 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. 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. 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 PHYS. REV. X 14, 021001 (2024) 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 experimen- tally . 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”. Freezable chaos pro- vides 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. 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 pharma- cological manipulations. 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 pres- ence or absence of a resetting signal could enable exper- imental disambiguation of our proposal. The model offers a mechanism for WM, but lacks a mechanism for long-term memory. This could be addressed by introducing slow Hebbian dynamics into J. For exam- ple, prior DMFT studies have taken J to be a static result of associative plasticity, resulting in various combinations of chaos and long-term memory retrieval. 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 plas- ticity of A(t) was disabled, the frozen state could be consolidated into J.