

Baseline control of optimal performance in recurrent neural networks

Shun Ogawa,^{1,†} Francesco Fumarola,^{1,†} Luca Mazzucato^{2,*}

¹Laboratory for Neural Computation and Adaptation, RIKEN Center for Brain Science,
Hirosawa, Wako, Saitama 351-0198, Japan

² Institute of Neuroscience and Departments of Biology, Mathematics and Physics,
University of Oregon, Eugene, OR 97403, USA

[†]Equal contributors.

*lmazzuca@uoregon.edu

Changes in behavioral state, such as arousal and movements, strongly affect neural activity in sensory areas. Recent evidence suggests that they may be mediated by top-down projections regulating the statistics of baseline input currents to sensory areas, inducing qualitatively different effects across sensory modalities. What are the computational benefits of these baseline modulations? We investigate this question within a brain-inspired framework for reservoir computing, where we vary the quenched baseline inputs to a random neural network. We found that baseline modulations control the dynamical phase of the reservoir network, unlocking a vast repertoire of network phases. We uncover a new zoo of bistable phases exhibiting the simultaneous coexistence of fixed points and chaos, of two fixed points, and of weak and strong chaos. Crucially, we discovered a host of novel phenomena, including noise-driven enhancement of chaos and ergodicity breaking; neural hysteresis, whereby transitions across phase boundary retain the memory of the initial phase. Strikingly, we found that baseline control can achieve optimal performance without any fine tuning of recurrent couplings. In summary, baseline control of network dynamics opens new directions for brain-inspired artificial intelligence and provides a new interpretation for the ubiquitously observed behavioral modulations of cortical activity.

Introduction

The activity of neurons across cortical areas is strongly modulated by changes in behavioral state such as arousal [1, 2], movements [3, 4, 5, 6], and task-engagement [7]. Intracellular recordings showed that these behavioral modulations are mediated by a change of baseline synaptic currents, likely originating from the thalamus and other subcortical areas [8, 9]. Such baseline modulations exert strong effects on neural activity explaining up to 50% of its variance across cortical areas, a much larger effect compared to the task-related modulations [4, 5, 6].

What are the functional effects of these baseline modulations on brain function? Experimental results across different sensory modalities painted a contradictory picture. Locomotion-induced modulations can improve visual processing [3, 10, 11], but degrade auditory processing [12, 13, 14]. Arousal, measured by pupil size, can improve gustatory and auditory processing at low to intermediate levels [2, 15, 16], but degrades performance auditory processing at high levels [17]. This variety of complex and apparently contradictory effects of behavioral modulations on neural activity and task performance poses a challenge to current theories of brain function and cortical circuit dynamics.

We aim to shed light on the effects of baseline modulations on cortical activity within the framework of reservoir computing, a powerful tool based on recurrent neural networks (RNNs) with random couplings. Random RNNs can recapitulate different dynamical phases observed in cortical circuits, such as silent or chaotic activity [18], fixed points [19], and the balanced regime [20]; and provide a simple explanation for task selectivity features [21] and the heterogeneity of timescales [22] observed in cortical neurons. Random RNNs can achieve optimal performance in memory tasks when poised at a critical point either by fine-tuning their random couplings [23] or their noisy input [24].

Following recent theoretical [2, 11] and experimental studies [4, 25], we modeled the effect of changes in an animal's behavioral state as changes in the mean and across-neurons variance of the constant baseline input currents to an RNN. We found that baseline modulations steer the network activity to interpolate between a large zoo of dynamical phases. Beyond known phases, such as fixed points and chaos, baseline modulations unlocked new ergodicity-breaking phases, where the network activity can switch between weak and strong chaos, between a

fixed point and chaos, or between two fixed points, depending on the initial conditions. All these different phases were continuously connected and achieved without any training or fine tuning of synaptic couplings. Strikingly, we found a new effect where an increase in quenched noise can induce chaos. When interpolating adiabatically between phases via baseline modulations, we uncovered the novel phenomenon of neural hysteresis, whereby the network activity retains a memory of the path followed in phase space.

Our theory further revealed that baseline modulations can flexibly control optimal performance in a sequential memory task at the edge of chaos, without any fine-tuning of synaptic weights. More generally, our theory shows that baseline modulations unlock a much richer dynamical phase portrait for RNNs than previously known. Baseline control represents a very simple and efficient way for a reservoir to flexibly toggle its dynamical regime to achieve different computations. Our results thus suggest a novel computational role for behavioral modulations of neural activity, suggesting that they might allow cortical circuits to flexibly adjust the cognitive task they perform to adapt to different contexts.

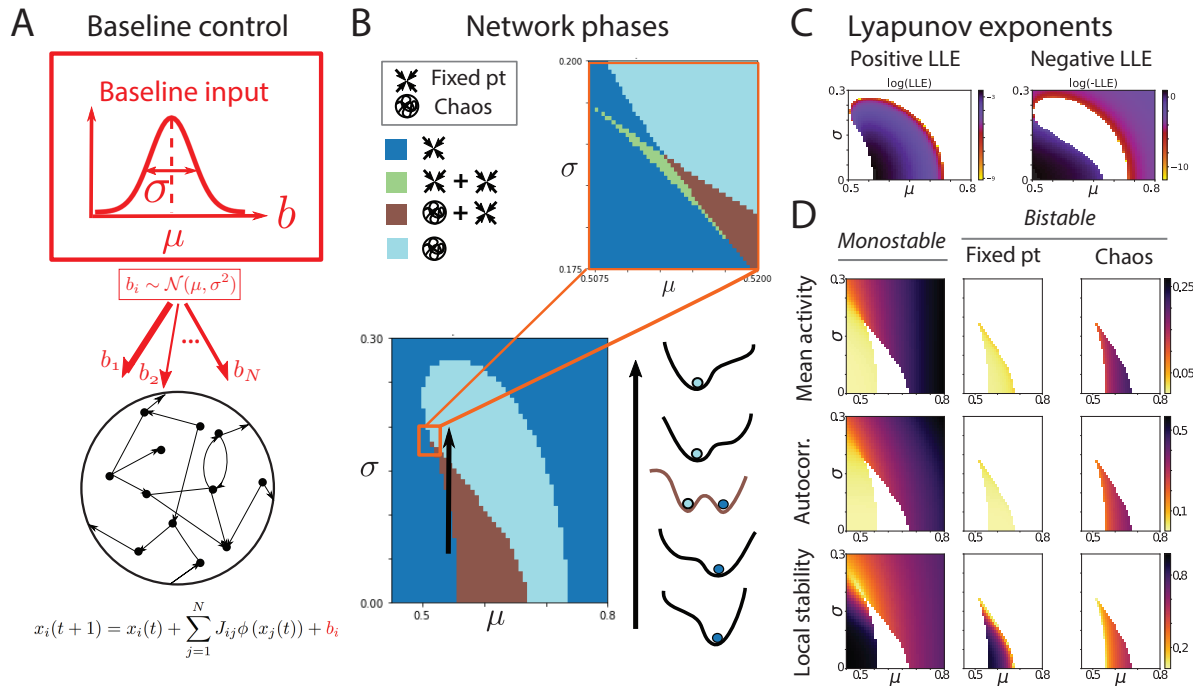


Fig. 1. Baseline control of the network dynamical phase. A) Random neural network where the baseline input current b_i to the i -th neuron is drawn from a normal distribution $\mathcal{N}(\mu, \sigma^2)$ (red). B) Left: Network phase diagram, obtained by varying the mean μ and variance σ^2 of the baseline input, shows four phases: fixed point (blue); chaos (cyan); bistable phase with coexistence of two fixed points (green); bistable phase with coexistence of fixed point and chaos (brown). Top right: multi-critical point. Bottom right: Schematic of in Landau potential along a phase space trajectory (black arrow in inset) from a stable phase with a single fixed point (blue circle), to a bistable phase with coexistence of fixed point and chaos (blue and cyan circles), to a stable phase with chaos (cyan circle). C) Positive (left) and negative (right) largest Lyapunov exponents (in the bistable phases both LLE coexist); D) Order parameters in each phase: Mean network activity (top); autocorrelation (middle); local stability (bottom). Representative network activity in the different phases. Insets: Order parameters (Autocovariance C_0, C_∞ and mean activity M). Network parameters: $J_0 = 0.5$, $g = 5$, $\theta_0 = 1$.

Results

We model our local cortical circuit as a recurrent neuronal network (RNN) of N neurons where the synaptic couplings are randomly drawn from a Gaussian distribution of mean J_0/N and variance g^2/N (Fig. 1A). We choose a positive definite neuronal transfer function $\phi(x) = 1/[1 + \exp(x - \theta_0)]$ with threshold θ_0 . Every neuron in our model receives a constant external synaptic input b_i drawn from a Gaussian distribution with mean μ and variance σ^2 . This baseline represents the afferent projections to the local cortical circuit originating from other areas. Following experimental [8, 9, 25] and theoretical studies [2, 11], we modeled behavioral modulations as a change in the baseline statistics (mean μ and variance σ^2) of synaptic inputs b_i to the local circuit, induced by top-down projections carrying information about the behavioral state of the animal, or other contextual modulations [2, 11]. Because the characteristic timescale of behavioral modulations is typically much slower than a circuit's stimulus responses, we approximate the effects of such modulations as the quenched inputs b_i . Importantly, this baseline modulations are constant, time-independent offsets of the input current to each neurons, and represent *quenched* input noise.

Baseline control of the network dynamical phases

We found that by varying the values of baseline mean and variance μ, σ^2 , one can access a large library of network phases (Fig. 1b-c). The first two phases are generalizations of the fixed point and the chaotic phase which were previously reported in [26]. Strikingly, we found a zoo of new phases including new 'bistable' phases where the network activity can reach two different dynamical branches for the same values of recurrent couplings and baseline input, depending on the initial conditions. In the network of Fig. 1B, the bistable phases are of two different kinds, with coexistence of either a fixed point and chaos (brown) or two fixed points (green). Whereas in the monostable phase the network Landau potential has one global minimum, in the bistable phases it exhibits two local minima, each one defining the basin of attraction of the initial conditions leading to each of the two bistable branches. Depending on the statistics of the random couplings (J_0, g), we found networks with up to five different phases, including a new bistable phase featuring the coexistence of strong and weak chaos (see Supplementary

Material). Each monostable phase and each branch of a bistable phase can be captured in terms of the network order parameters LLE, M and C (Fig. 1C), representing, respectively, the largest Lyapunov exponent LLE and the mean M and variance C of the activity obtained from the self-consistent dynamic mean field equations (see Methods).

The variance of the activity includes a contribution σ^2 from the quenched baseline input and a recurrent contribution. A useful characterization of the network dynamical phase is obtained when considering the population-averaged autocorrelation function $c(t)$ at lag t ; in particular, its zero lag value $c(0) = C$ the network variance, and its asymptotic value for large lag $c(\infty)$. The network is at a fixed point if $c(t)$ does not depend on time (i.e., $c(\infty) = c(0) = C$), while it is in a chaotic phase if $c(0) > c(\infty)$, in which case the LLE is positive. Finally, a value of $c(\infty) > 0$ signals a nonzero mean activity driven by the quenched variance in the baseline input.

Noise-induced enhancement of chaos

Exploring the features of baseline modulations revealed a novel and surprising phenomenon, whereby increasing the variance of the quenched input can enhance chaos. This phenomenon can be understood from a mean field perspective by considering how the baseline and the recurrent synaptic inputs interact with the single cell transfer function to determine the operating point of the network dynamics (Fig. 2). To illustrate this phenomenon, we first revisit the known case of noise-driven suppression of chaos realized in a circuit with quenched inputs and a zero-centered transfer function (Fig. 2A), which can be obtained when the mean baseline is set equal to the threshold $\mu = \theta_0$ (see [27] for a case where they both vanish). On general grounds, one expects the network phase to be chaotic whenever a large fraction of the synaptic input distribution is concentrated in the high gain region of the transfer function, in which $\phi'(x)^2$ is large. The distribution of synaptic inputs has mean M , which is centered at the threshold, and some nonzero variance C , obtained self-consistently from (3) and (4). For zero baseline variance, the network exhibits chaotic activity (case 1), as a large fraction of the synaptic inputs have access to the high gain region of the transfer function. When turning on a quenched baseline variance σ^2 , the synaptic input increases its variance by a value proportional to σ^2 . For larger values of the baseline variance σ^2 , the fraction of synaptic inputs in the high gain region progressively shrinks and for large enough variance chaos is suppressed (case 2).

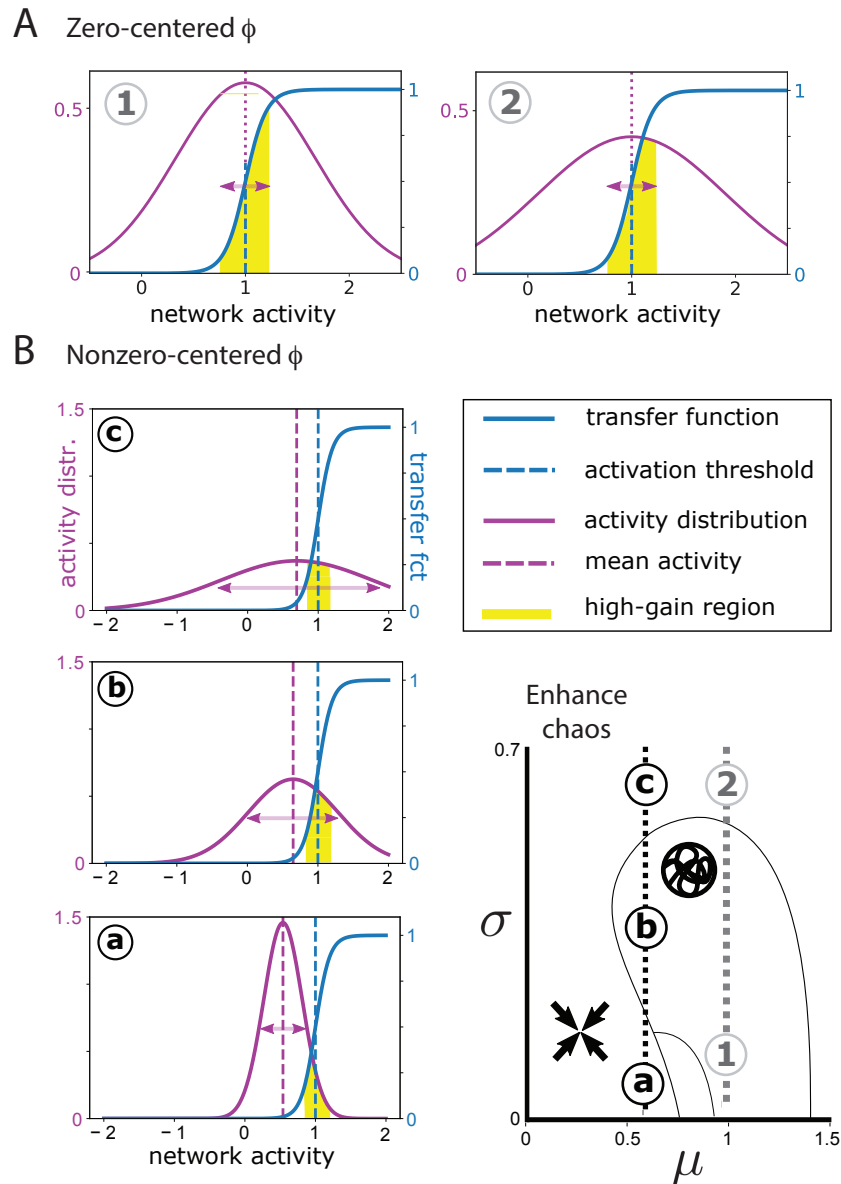


Fig. 2. Noise-driven modulations of chaos. A) When the transfer function input is zero-centered (i.e., the baseline mean μ equals the threshold θ_0). The chaotic phase (1) occurs when a large fraction of the synaptic input distribution (pink curve) lies within the high gain region (yellow shaded area) of the transfer function (blue curve). Increasing the input quenched variance (2) reduces this fraction, suppressing chaos. B) When the transfer function is nonzero-centered ($\mu < \theta_0$), for low (a) and high (c) quenched input variance network activity is at a fixed point, because the high gain region receives a small synaptic input fraction; this fraction is maximized at intermediate quenched input variance (b), enhancing chaos. Network parameters: $g = 5, \theta_0 = 1, J_0 = 0$. Panel A: $\phi(x) = \tanh(x - \theta_0), \mu = 1$; cases (1, 2) $\sigma = (0.1, 0.6)$. Panel B: $\phi = 1/(1 + e^{x - \theta_0}), \mu = 0.5$; case (a, b, c): $\sigma = (0.2, 0.4, 0.9)$.

In the case where $\mu < \theta_0$, the transfer function is not zero-centered, and noise-driven enhancement of chaos can occur (Fig. 2B). For low baseline variance σ^2 , the network is in the fixed point regime as a small fraction of synaptic inputs has access to the high gain region (case a). Increasing the baseline variance σ^2 leads to a transitions into a chaotic phase, as a progressively larger fraction of synaptic inputs has access to the high gain region. At some large enough variance, though, the fraction of synaptic inputs in the high gain region starts decreasing again and eventually this leads to a new transition to the fixed point phase. This chaos enhancement can be achieved either by passing through an intermediate bistable phase (black arrow in Fig. 1B); or by inducing a direct transition from a fixed point to a chaotic phase at lower values of the mean baseline μ (gray arrow). This novel chaos enhancement has a number of striking consequences, such as baseline control of optimal performance and neural hysteresis, which we will examine in the next sections. While previous studies showed that an increase in the *temporal* noise (e.g., white noise inputs) always leads to suppression of chaos [28, 27, 29, 24], we found that quenched noise unlocks a much richer set of phenomena.

Ergodicity breaking in the bistable phases

The network activity in a bistable phase exhibits dynamical breaking of ergodicity. To illustrate this effect, we consider a network with fixed baseline mean μ at different values of σ (Fig. 3). At intermediate values of σ the network is in the bistable phase featuring a coexistence of a fixed point attractor and chaos, while at low and high values the network is in the monostable fixed point phase and the chaotic phase, respectively. In the bistable phase, the network dynamics converge to either a fixed point attractor or to a chaotic attractor, depending on the initial conditions (Fig. 3A). These two branches are characterized by a negative (fixed point) or a positive (chaos) LLE, respectively, and branch-specific values for the network order parameters (C , M , Fig. 3B). We quantified ergodicity breaking in terms of the average distance $\langle d(T) \rangle$ between temporal trajectories (starting from different initial conditions, or between different replicas) over an epoch T (Fig. 3C). In the ergodic monostable phases $\langle d(T) \rangle$ converges to C_∞ at large $T \rightarrow \infty$, since the network activity eventually explores all possible configurations (in the chaotic phase, the decay is typically slower than in a phase with a single attractor). If $\langle d(T) \rangle$ does not decays to C_∞ but it monotonically increases to reach a non-zero late time values larger than C_∞ , the

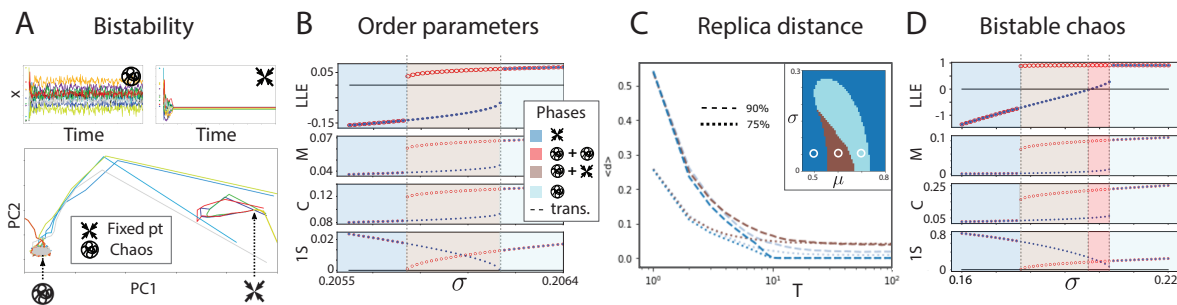


Fig. 3. Ergodicity breaking in the bistable phase. A) Representative trajectories of network activity (top) and their top two principal components (bottom) in the bistable phase obtained from initial conditions within the chaotic (left) and fixed point (right) basins of attraction. B) For increasing values of σ , a crossover from a monostable fixed point phase (left), to a bistable phase fixed point/chaos (middle) to a monostable chaotic phase (right) is revealed by the order parameters (LLE: Largest Lyapunov exponent; M: mean activity; C: mean autocorrelation; 1S: 1-replica stability). In the bistable phase, the fixed point and chaotic branches exhibit different order parameters. C) Average distance between replica trajectories $\langle d \rangle$ reveals ergodicity breaking: in the monostable fixed point (blue) and chaotic (cyan) phases $\langle d \rangle$ asymptotes to C_∞ , but in the bistable phase (brown) it asymptotes to a value larger than C_∞ , representing the average distance between the basins of attraction of the two branches. D) Example of a crossover from a monostable fixed point phase (blue), to a bistable phase fixed point/chaos (brown) to a bistable weak/strong chaos phase (red), to a monostable chaotic phase (cyan), as revealed by the order parameters (same as panel B). Network parameters: panel A: $J_0 = 0.5, \theta_0 = 1, \mu = 0.54, g = 5, \sigma = 0.1$; panel B: $J_0 = 0.5, g = 6, \theta_0 = 1, \mu = 0.5$; panel C: same as Fig. (1B) and $\mu = 0.5, 0.6, 0.7$; panel D: $J_0 = 0.5, \theta = 1, \mu = 0.5, g = 18$.

network breaks ergodicity. This means that, depending on the initial conditions, there are different basins at finite distance from each other. We found that the network is non-ergodic in all the bistable phases, although each one of these phases retains specific values of the order parameters.

The library of bistable phases induced by changes in the baseline statistics include all the phases in Fig. 3 and, remarkably, a novel phase exhibiting the coexistence of two chaotic phases 3D. This double chaos phase features a weak chaotic branch with small positive LLE and slow dynamics, and a strong chaotic branch with large positive LLE and fast dynamics. We found that this double chaos phase occurs for large g and it exhibits important computational properties that we investigate below.

Neural hysteresis retains memory of network phase trajectories

What are the effects of adiabatic changes in baseline statistics on the network dynamics? We sought to elucidate the effects of slow baseline changes, by driving the network with time-varying values of $\sigma(t)$ for fixed μ , describing a closed loop (Fig. 4). We found that the network order parameters C , M changed discontinuously across phase boundaries, signaling a phase transition. When the baseline trajectory crosses the phase boundary from a stable phase (with a single LLE) to a bistable phase (with two branches, each characterized by its own LLE), the network activity in the bistable phase lies on either of the two branches, characterized by two separate basins of attractions (Fig. 3). The rules governing which of the two branches will be reached are determined by a novel hysteresis effect. We found that the network activity in the bistable phase retained a memory of the dynamical branch that it occupied before crossing the phase boundary. In the particular example of Fig. 4, when crossing the boundary from the monostable fixed point to the bistable phase, the activity will persist on the fixed point branch of the bistable phase, whose negative LLE is continuously connected with the fixed point phase. For larger values of $\sigma(t)$, the network will eventually enter the monostable chaos phase, where the LLE discontinuously jumps to a very large value. Vice versa, when inverting the time-varying trajectory in phase space by slowly decreasing the $\sigma(t)$ from the monostable chaotic phase into the bistable phase, the network will persist on the chaotic branch of the latter, whose positive LLE is continuously connected to the monostable chaotic phase. Eventually, for lower $\sigma(t)$ the network falls back into the fixed point phase where the LLE discontinuously jumps from large positive

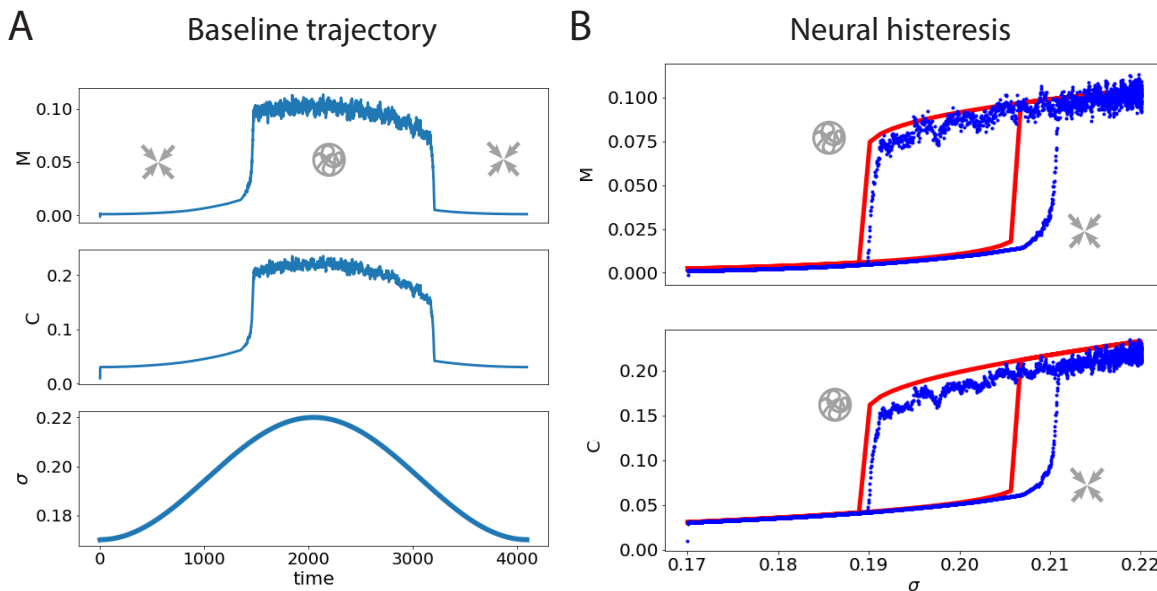


Fig. 4. Neural hysteresis. A) Slow changes in baseline variance $\sigma(t)$ leads to discontinuous transitions in the network order parameters M, C (left: temporal profile of M, C, σ). B) Crossing over phase boundaries by a time-varying $\sigma(t)$ retains memory of the previous phase (blue: network simulations; red: exact DMFT calculation). Network parameters: $J_0 = 0.5, g = 12, \theta_0 = 1, \mu = 0.5, \sigma(t) = \sigma_0 + \sigma_1 \sin(\pi t/T)$ with $T = 2048, \sigma_0 = 0.17, \sigma_1 = 0.025$.

to negative values. Thus, when crossing phase boundaries adiabatically the network will choose the branch of the bistable phase whose LLE is continuously connected to the previous phase.

Neural hysteresis occurs not just in the fixed point/chaos bistable phase, but also in the double fixed point and double chaos bistable phases. When crossing boundaries between two adjacent bistable phases, more complex hysteresis profiles can occur. For example, when crossing into the double chaos phase (with fast/slow chaotic branches, Fig. 3D), from the fixed point branch of the fixed point/chaotic phase, the network dynamics will lie on the slow chaotic branch, whose positive but small LLE is continuously connected to the fixed point branch of the previous bistable phase. However, when crossing into the double chaos phase from the chaotic branch of the fixed point/chaotic bistable phase, the network dynamics will persist on the fast chaotic branch, whose large positive LLE is continuously connected to the chaotic branch of the fixed point/chaotic bistable phase. We then examined

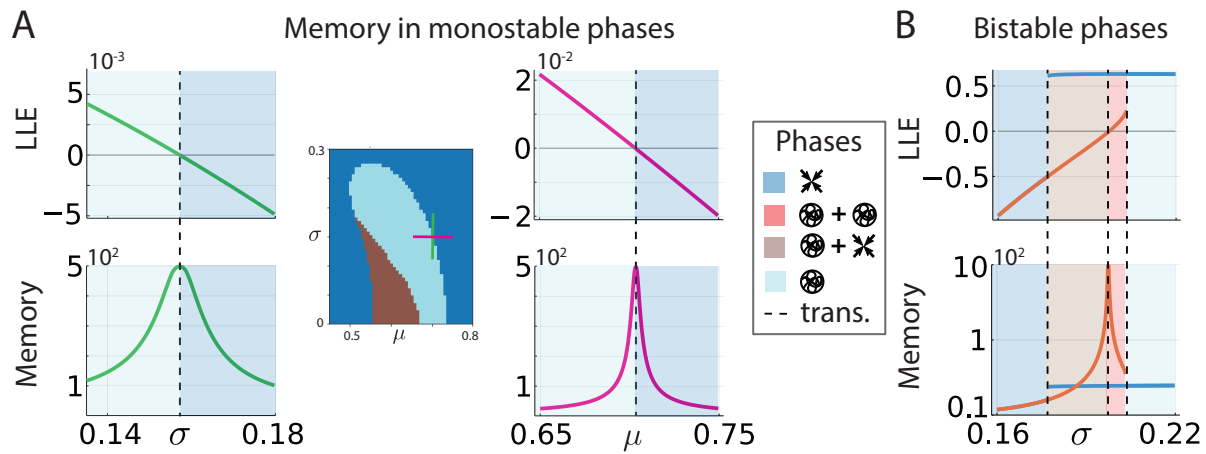


Fig. 5. Baseline control of optimal memory capacity. A) Two representative trajectories in baseline (μ, σ) space (left: green and orange lines) allow to reach a phase transition where the LLE crosses zero (top panels) and memory capacity is optimized (bottom panels). B) In a transition between bistable phases, memory capacity is optimized by a baseline trajectory whose branch exhibits an LLE that crosses zero at the phase boundary (orange curve); the branch with positive LLE (blue curve) does not maximize memory capacity. Network parameters: Panel A, same as Fig. 1A; panel B, same as Fig. 3D.

the relevance of neural hysteresis for controlling the network performance in a memory task.

Baseline control of optimal memory capacity

A classic result in the theory of random neural networks is that, by fine tuning the recurrent couplings at the ‘edge of chaos’, one can achieve optimal performance in a memory task, where the network activity maintains for a very long time a memory of stimuli presented sequentially [23]. This was achieved by fine tuning the network recurrent couplings to values close to the transition between fixed point and chaos, which is a metabolically costly and slow procedure typically requiring synaptic plasticity. Is it possible to achieve optimal memory capacity without changing the recurrent couplings? We found that baseline control can achieve optimal memory capacity by simply adjusting the mean and variance of the baseline input distribution, without requiring any change in the recurrent couplings, (Fig. 5).

We first derived an analytical formula for the memory capacity in the vicinity of a second-order phase transition boundary

$$\mathcal{M} \sim \frac{1}{1 - \langle \phi'^\alpha \phi'^\beta \rangle}, \quad (1)$$

where α, β are replica indices. Optimal memory capacity is achieved close to a phase boundary, and its features are qualitatively different depending on whether the phases separated by the boundary are monostable or bistable. At a boundary between two monostable phases, where the activity transitions between a fixed point and chaotic phase, optimal memory capacity is achieved at the edge of chaos. For fixed values of the recurrent couplings (Fig. 5A), one can easily achieve optimal memory capacity by adiabatically changing either the mean or the variance of the baseline. This external modulation thus sets the network at the edge of chaos, in the region where memory capacity is maximized, via baseline control, without any change in the recurrent couplings. Around a phase boundary involving a bistable phase, the optimal performance region can be reached by making use of the neural hysteresis phenomenon. We illustrate this intriguing scenario in the case of the transition from a bistable fixed point/chaos branch to a bistable double chaos branch (Fig. 5B). Optimal performance is achieved only on the branch of the bistable phase transition which undergoes a second-order phase transition (i.e., the branch whose LLE crosses zero). In this specific case, then, we can reach optimal performance on the lower branch of the LLE curve, describing the transition between the weak chaotic branch of the double chaos phase to the fixed point branch of the fixed point/chaos phase. Because of the neural hysteresis, achieving the optimal performance region requires first initializing the network on the lower LLE branch (on either side of the transition), and then adiabatically controlling the baseline to reach the desired point. The phase boundaries where only first-order phase transitions occur (i.e., no branch exhibits an LLE that crosses zero) do not lead to optimal memory capacity. For example, in Fig. 3B, neither the upper nor lower branch of the transition between a monostable fixed point phase to a bistable fixed point/chaos phase lead to large memory capacity, since no LLE on either branch of the intermediate bistable phase crosses zero. Nevertheless, it is always possible to reach a different second-order phase boundary from any point in (μ, σ) space by following an appropriate adiabatic trajectory in the baseline, where optimal memory capacity can be achieved (see Fig. 5A). Therefore, one can achieve baseline control of optimal performance via

neural hysteresis.

Discussion

We presented a new brain-inspired framework for reservoir computing where we controlled the dynamical phase of a recurrent neural network by modulating the mean and quenched variance of its baseline inputs. Baseline modulations revealed a host of new phenomena. First, we found that they can set the operating point of the network activity by controlling whether synaptic inputs overlap with the high gain region of the transfer function. A manifestation of this effect is a novel noise-induced enhancement of chaos. Second, baseline modulations unlocked access to a large repertoire of network phases. On top of the known fixed point and chaotic ones, we uncovered a new zoo of bistable phases, where the network activity breaks ergodicity and exhibits the simultaneous coexistence of a fixed point and chaos, of two different fixed points, and weak and strong chaos. By driving the network with adiabatic changes in the baseline statistics one can toggle between the different phases, charting a trajectory in phase space. These trajectories exhibited the new phenomenon of neural hysteresis, whereby adiabatic transitions across a phase boundary retain the memory of the adiabatic trajectory. Finally, we showed that baseline control can achieve optimal performance in a memory task at a second-order phase boundary without any fine tuning of the network recurrent couplings.

Noise-induced enhancement of chaos. Previous theoretical work found a noise-induced suppression of chaos in random neural networks driven by time-varying inputs both in discrete time [28] and continuous time [27, 29, 24, 22, 30, 24]. In previous cases, featuring a mean synaptic input centered in the middle of the high-gain region of the transfer function, suppression of chaos occurs because an increase in the variance drives the network away from the chaotic regime. In contrast, we found that, when the baseline statistics sets the mean synaptic input away from the center of the high gain region, one can induce a transition from fixed point to chaos at intermediate values of the variance (Fig. 2). Larger values of the variance eventually suppress chaos, such that a non-monotonic dependence of the Lyapunov exponent on the baseline variance or mean can be realized. To our knowledge this is the first example of noise-induced chaos in a recurrent neural network (although for the logistic map see [31]). We believe

that noise-induced modulation of chaos in discrete time networks is similar for both quenched and dynamical noise [24], since the LLE and the edge of chaos are the same for both cases. We speculate that introducing a leak term and generalizing our results to a continuous time system may induce a dynamical suppression of chaos on general grounds, based on the memory effect. Another interesting direction is to drive the network with dynamical noise at different values of the baseline input and investigate its effect on the different monostable and bistable phases we uncovered via baseline modulation.

Optimal sequential memory. Previous studies showed that optimal performance in random networks can be achieved by either tuning the recurrent couplings at the edge of chaos [23], by driving the network with noisy input tuned to a particular amplitude [24]. The former method represents a metabolically costly and slow process requiring synaptic plasticity. The latter method may lack biological plausibility, since in a spiking circuit the dynamical input noise statistics are self-consistently determined by the spiking dynamics and are not a tunable parameter. We found that optimal sequential memory performance can be achieved by simple regulation of the mean and variance of the baseline current. Achieving optimal performance by changing the across-neuron variance of baseline currents is a simple and biologically plausible mechanism.

Information processing capabilities and bistability. Bistable phases in recurrent networks with random couplings were previously reported in [32]. We generalized this to a new set of bistable phases featuring the coexistence of two fixed points and, remarkably, two chaotic attractors with slow and fast chaos, respectively. To our knowledge, this is the first report of a doubly chaotic phase in recurrent neural networks. Are there any information processing benefits of the double chaos phase? Neural activity unfolding within the weakly chaotic branch of this bistable phase has large sequential memory capacity, as the Fisher information diverges at the edge of chaos. On the other hand, the strongly chaotic branch erases memory fast. In this doubly chaotic phase, the network's information processing ability can be changed drastically by switching between the two branches, for example via an external pulse. It would be tantalizing to explore the computational capabilities of these new bistable phases unlocked by baseline modulation. Here, we only considered homogeneous inputs where the baseline statistics is the same for all network neurons. Although, one may consider a more general set up with heterogeneous inputs, where different

neural populations receive baseline modulations with different statistics. The simplest such possibility would be the ability to perform different tasks by gating in and out specific subpopulations, driving them with negative input. This is a promising new direction for multitasking and we leave it for future work.

Evidence for baseline modulations in brain circuits. In biologically plausible models of cortical circuits based on spiking networks, it was previously shown that increasing the baseline quenched variance leads to improved performance. This mechanism was shown to explain the improvement of sensory processing observed in visual during locomotion [11] and in gustatory cortex with general expectation [2]. In these studies, the effect of locomotion or expectation was modeled as a change in the constant baseline input to each neuron realizing an increase in the input quenched variance. This model was consistent with the physiological observation of the heterogeneous neuronal responses to changes in behavioral state, comprising a mix of enhanced and suppressed firing rate responses (during locomotion [3, 11, 25], movements [4, 6, 5], or expectation [33, 34]. Intracellular recordings showed that these modulations are mediated by a change of baseline synaptic currents, likely originating from subcortical areas [8, 9]. Because the effects of these changes in behavioral state on neural activity unfolded over a slower timescale (a few seconds) compared to the typical information processing speed in neural circuits (sub-second), we modeled them as constant baseline changes, captured by changes in the mean and variance of the distribution of input currents. Our results provide a new interpretation of these phenomena, leading to the hypothesis that they could enable cortical circuits to adapt their operating regimes to changing demands.

Neural hysteresis. A new prediction of our model is that baseline modulations may induce neural hysteresis when crossing a bistable phase boundary. Hysteresis is a universal phenomenon observed in many domains of physics. Our results suggest a potential way to examine the existence of hysteresis in brain circuits, within the assumption that increasing baseline variance represents increasing values of a continuous behavioral modulation such as arousal (e.g., measure by pupil size [17]). A potential signature of hysteresis could be detected if the autocorrelation time of neural activity at a specific arousal level exhibited a strong dependence on whether arousal levels decreased from very high levels or increased from very low levels. We leave this interesting direction for future work.

Methods

Random neural network model

Our discrete time neural network model with top down control, illustrated in Fig. 1, is governed by the dynamical equation

$$x_{i,t+1} = \sum_{j=1}^N J_{ij} \phi(x_{j,t}) + b_i + \eta_t \quad (2)$$

Here b_i is quenched Gaussian noise with mean μ and variance σ^2 , η_t is a possible time-dependent external stimulus (relevant for the sequential memory task below). The mean of the synaptic strength, J_0/N , is not zero and its variance is g^2/N ; the scaling $1/N$ guarantees the existence of the large N limit. We will assume $\mu > 0$ in accordance with the fact that top-down modulation is directly conveyed by long-range pyramidal connections. The activation function $\phi(x) = \frac{1}{2} [\tanh(x - \theta_0) + 1]$ is positive definite and biologically plausible as it incorporates both a soft rectification and thresholding. Indeed the activation function ϕ satisfies $\phi(x) \approx 0$ when $x \ll \theta_0$ and $\phi(x) \approx 1$ when $x \gg \theta_0$.

Order parameters

The order parameters of the model are the population mean and variance at equilibrium of the single neuron activity $\langle x_{i,t} \rangle$. A rigorous derivation of self-consistent equations for these two quantities requires Dynamical Mean Field Theory (see Supplementary Material), a heuristic argument for them can be sketched as follows. Averaging Eq. 2 in the absence of external input yields

$$\langle x_{i,t+1} \rangle = \sum_{j=1}^N \langle J_{ij} \phi(x_{j,t}) \rangle$$

Neglecting correlation between the random variables J_{ij} and $x_{j,t}$ on the right hand side, and using the statistical invariance under permutation of neuron labels to drop cell indices, we obtain $\langle x_{t+1} \rangle = J \langle \phi(x_t) \rangle$. Focusing now on the stationary regime, where the distribution of x_{t+1} and x_t are identical, and assuming them to be gaussian with mean M and variance C , leads to

$$M = J \int \frac{dx}{\sqrt{2\pi}} e^{-x^2/2} \phi(\sqrt{C}x + M) \quad (3)$$

Taking the second moment of Eq. 2, without neglecting the variance of the quenched disorder, term and deploying once again the same assumptions yields

$$C = \sigma^2 + \int \frac{dx}{\sqrt{2\pi}} e^{-x^2/2} \phi(\sqrt{C}x + M)^2 \quad (4)$$

In supplement, the Dynamical Mean-Field Theory approach is rigorously developed to derive two dynamical equations for the mean-field momenta. The stationary limit of those equation is found to correspond to Eqs. 4 and 3, thus confirming the heuristic result.

Distance between replicas

Let us define the mean activity in the replica a (corresponding to some initial conditions $x_i^a(0)$) as

$$\bar{x}_i^a(T) = \frac{1}{T} \int_0^T dt x_i^a(t).$$

We then define the distance between replicas as [35]

$$d_{ab}^2(T) = \frac{1}{N} \sum_{i=1}^N [\bar{x}_i^a(T) - \bar{x}_i^b(T)]^2.$$

as used in the visualization of Fig. 3C.

Memory capacity

Following [36, 37], we define the memory capacity of a dynamical system for an observer in possession of an unbiased estimator for the mean, who can therefore remove the mean values from all the time series he records. Moreover, we would like the resulting memory capacity to be zero when the linear readout is dominated by a constant baseline value, because nothing can be learned from a readout independent on the input. Adopting therefore the mean-removed formula, we find for the memory capacity \mathcal{M} in the neighborhood of the second-order phase transition boundary

$$\mathcal{M} \sim \frac{1}{1 - \langle \phi'^\alpha \phi'^\beta \rangle} \quad (5)$$

The rigorous derivation of this formula is detailed in Supplement.

Largest Lyapunov exponent

The Lyapunov exponent of a dynamical system is a quantity that characterizes the rate of separation of infinitesimally close trajectories. Quantitatively, two trajectories in phase space with an initial separation vector diverge (provided that the divergence can be treated within the linearized approximation) at an exponential rate given, and the Lyapunov exponent governs this exponential growth. The LLE for a discrete-time dynamical system can be written as

$$\lambda_{\max} = \lim_{\tau \rightarrow \infty} \lim_{\|x_t^1 - x_t^2\| \rightarrow 0} \frac{1}{2\tau} \ln \frac{\langle |x_{t+\tau}^1 - x_{t+\tau}^2|^2 \rangle}{\langle |x_t^1 - x_t^2|^2 \rangle}, \quad (6)$$

which indicates how the two orbits get to be far from each other.

Going back to the N body picture, we have

$$\frac{1}{N} \sum_{i=1}^N (x_{i,t}^1 - x_{i,t}^2)^2 \rightarrow \langle |x_t^1 - x_t^2|^2 \rangle = C_{t,t}^{11} + C_{t,t}^{22} - 2C_{t,t}^{12} \quad (7)$$

for $N \rightarrow \infty$. Around the stationary solution, we consider $C_{tt}^{11} = C_{tt}^{22} = C_0$ and $C_{tt}^{12} = C_0 + \delta C_{tt}^{1,2}$. Then, we have the LLE as follows;

$$\begin{aligned} \lambda_{\max} &= \lim_{\tau \rightarrow \infty} \frac{1}{2\tau} \ln \frac{\delta C_{t+\tau,t+\tau}^{12}}{\delta C_{tt}^{12}} \Big|_{C_{tt}^{12}=C_0} \\ &= \lim_{\tau \rightarrow \infty} \frac{1}{2\tau} \sum_{s=0}^{\tau-1} \ln \frac{\delta C_{t+s+1,t+s+1}^{12}}{\delta C_{t+s,t+s}^{12}} \Big|_{C_{t+s,t+s}^{12}=C_0} \\ &\rightarrow \frac{1}{2} \ln \frac{\delta C_{t+1,t+1}^{12}}{\delta C_{tt}^{12}} \Big|_{C_{tt}^{12}=C_0} \quad (t \gg 1). \end{aligned} \quad (8)$$

and the LLE is estimated as [38]

$$\lambda_{\text{LLE}} = \frac{1}{2} \ln \langle \phi'(x_t)^2 \rangle = \frac{1}{2} \ln \int \phi' \left(\sqrt{C} x_t + M \right)^2 D x \quad (9)$$

Here C and M are the stationary solutions M_t and $C_{tt}^{\alpha\alpha}$ to the dynamical mean-field equation in the Supplemental, which are easy to find numerically by iterating substitution. To detect a state of the system (2), what we have to do is just solving Eq. 3 and 4 and check the sign of the LLE (9) for each state. Conceptually, the consequences of Eq. 9 are described in the cartoon of Figure 2. The top-down control can use two levers – mean and variance

of its modulation, and depending on the mean, the variance can have the opposite effects of tuning the controlled network into chaos or out of it.

Contributions and acknowledgments

LM supervised the project; SO worked out the analytics with FF's support; numerical simulations were carried out by FF and SO; all authors wrote the manuscript. We would like to thank Enrico Rinaldi for advice on the numerics and Taro Toyozumi for discussions. SO and FF were partially supported by RIKEN Center for Brain Science; LM was supported by National Institute of Neurological Disorders and Stroke grant R01-NS118461 and by National Institute on Drug Abuse grant R01-DA055439 (CRCNS).

References

- [1] Matthew J McGinley et al. “Waking state: rapid variations modulate neural and behavioral responses”. In: *Neuron* 87.6 (2015), pp. 1143–1161.
- [2] Luca Mazzucato, Giancarlo La Camera, and Alfredo Fontanini. “Expectation-induced modulation of metastable activity underlies faster coding of sensory stimuli”. In: *Nature neuroscience* (2019), p. 1.
- [3] Cristopher M Niell and Michael P Stryker. “Modulation of visual responses by behavioral state in mouse visual cortex”. In: *Neuron* 65.4 (2010), pp. 472–479.
- [4] Carsen Stringer et al. “Spontaneous behaviors drive multidimensional, brain-wide population activity”. In: *BioRxiv* (2018), p. 306019.
- [5] Simon Musall et al. “Single-trial neural dynamics are dominated by richly varied movements”. In: *bioRxiv* (2019), p. 308288.
- [6] David B Salkoff et al. “Movement and performance explain widespread cortical activity in a visual detection task”. In: *Cerebral Cortex* 30.1 (2020), pp. 421–437.
- [7] A. Fontanini and D. B. Katz. “Behavioral states, network states, and sensory response variability”. In: *J Neurophysiol* 100.3 (2008), pp. 1160–8.
- [8] Pierre-Olivier Polack, Jonathan Friedman, and Peyman Golshani. “Cellular mechanisms of brain state-dependent gain modulation in visual cortex”. In: *Nature neuroscience* 16.9 (2013), p. 1331.
- [9] Dennis B Nestvogel and David A McCormick. “Visual thalamocortical mechanisms of waking state-dependent activity and alpha oscillations”. In: *Neuron* 110.1 (2022), pp. 120–138.
- [10] Maria C Dadarlat and Michael P Stryker. “Locomotion enhances neural encoding of visual stimuli in mouse V1”. In: *Journal of Neuroscience* 37.14 (2017), pp. 3764–3775.
- [11] David Wyrick and Luca Mazzucato. “State-dependent regulation of cortical processing speed via gain modulation”. In: *Journal of Neuroscience* 41.18 (2021), pp. 3988–4005.
- [12] David M Schneider, Anders Nelson, and Richard Mooney. “A synaptic and circuit basis for corollary discharge in the auditory cortex”. In: *Nature* 513.7517 (2014), pp. 189–194.
- [13] James Bigelow et al. “Movement and VIP interneuron activation differentially modulate encoding in mouse auditory cortex”. In: *Eneuro* 6.5 (2019).
- [14] Iryna Yavorska and Michael Wehr. “Effects of locomotion in auditory cortex are not mediated by the VIP network”. In: *Frontiers in neural circuits* 15 (2021), p. 24.
- [15] Pei-Ann Lin et al. “Arousal regulates frequency tuning in primary auditory cortex”. In: *Proceedings of the National Academy of Sciences* 116.50 (2019), pp. 25304–25310.
- [16] Charles R Heller et al. “Selective effects of arousal on population coding of natural sounds in auditory cortex”. In: *bioRxiv* (2020).
- [17] Matthew J McGinley, Stephen V David, and David A McCormick. “Cortical membrane potential signature of optimal states for sensory signal detection”. In: *Neuron* (2015).
- [18] A Crisanti and H Sompolinsky. “Path integral approach to random neural networks”. In: *Physical Review E* 98.6 (2018), p. 062120.
- [19] Merav Stern, Haim Sompolinsky, and LF Abbott. “Dynamics of random neural networks with bistable units”. In: *Physical Review E* 90.6 (2014), p. 062710.
- [20] Carl Van Vreeswijk and Haim Sompolinsky. “Chaos in neuronal networks with balanced excitatory and inhibitory activity”. In: *Science* 274.5293 (1996), pp. 1724–1726.

- [21] Audrey Sederberg and Ilya Nemenman. “Randomly connected networks generate emergent selectivity and predict decoding properties of large populations of neurons”. In: PLoS computational biology 16.5 (2020), e1007875.
- [22] Merav Stern, Nicolae Istrate, and Luca Mazzucato. “A reservoir of timescales in random neural networks”. In: arXiv preprint arXiv:2110.09165 (2021).
- [23] Taro Toyozumi and LF Abbott. “Beyond the edge of chaos: Amplification and temporal integration by recurrent networks in the chaotic regime”. In: Physical Review E 84.5 (2011), p. 051908.
- [24] Jannis Schuecker, Sven Goedeke, and Moritz Helias. “Optimal sequence memory in driven random networks”. In: Physical Review X 8.4 (2018), p. 041029.
- [25] Mario Dipoppa et al. “Vision and locomotion shape the interactions between neuron types in mouse visual cortex”. In: Neuron 98.3 (2018), pp. 602–615.
- [26] Haim Sompolinsky, Andrea Crisanti, and Hans-Jurgen Sommers. “Chaos in random neural networks”. In: Physical review letters 61.3 (1988), p. 259.
- [27] Kanaka Rajan, LF Abbott, and Haim Sompolinsky. “Stimulus-dependent suppression of chaos in recurrent neural networks”. In: Physical Review E 82.1 (2010), p. 011903.
- [28] Lutz Molgedey, J Schuchhardt, and Heinz G Schuster. “Suppressing chaos in neural networks by noise”. In: Physical review letters 69.26 (1992), p. 3717.
- [29] L. F. Abbott, K. Rajan, and H. Sompolinsky. “Interactions between Intrinsic and Stimulus-Evoked Activity in Recurrent Neural Networks”. In: The Dynamic Brain: An Exploration of Neuronal Variability and Its Functional Significance. Ed. by M. Ding and D. Glanzman. Oxford University Press, 2011. Chap. 4.
- [30] Rainer Engelken et al. “Input correlations impede suppression of chaos and learning in balanced rate networks”. In: arXiv preprint arXiv:2201.09916 (2022).
- [31] Yuzuru Sato. “Noise-induced phenomena in one-dimensional maps”. In: Plenary Lectures (2010), p. 95.
- [32] Łukasz Kuśmierz, Shun Ogawa, and Taro Toyozumi. “Edge of chaos and avalanches in neural networks with heavy-tailed synaptic weight distribution”. In: Physical Review Letters 125.2 (2020), p. 028101.
- [33] Chad L Samuelsen, Matthew PH Gardner, and Alfredo Fontanini. “Effects of cue-triggered expectation on cortical processing of taste”. In: Neuron 74.2 (2012), pp. 410–422.
- [34] Roberto Vincis and Alfredo Fontanini. “Associative learning changes cross-modal representations in the gustatory cortex”. In: Elife 5 (2016), e16420.
- [35] Kevin Berlemont and Gianluigi Mongillo. “Glassy phase in dynamically-balanced neural networks”. In: bioRxiv (2022).
- [36] Joni Dambre et al. “Information processing capacity of dynamical systems”. In: Scientific reports 2.1 (2012), pp. 1–7.
- [37] Herbert Jaeger. “Short term memory in echo state networks. gmd-report 152”. In: GMD-German National Research Institute for Computer Science (2002), <http://www.faculty.jacobs-university.de/hjaeger/publ/CiteSeer.2002>.
- [38] Bruno Cessac. “Increase in complexity in random neural networks”. In: Journal de Physique I 5.3 (1995), pp. 409–432.
- [39] Moritz Helias and David Dahmen. “Statistical field theory for neural networks”. In: arXiv preprint arXiv:1901.10416 (2019).
- [40] Haim Sompolinsky, Andrea Crisanti, and Hans-Jurgen Sommers. “Chaos in random neural networks”. In: Physical review letters 61.3 (1988), p. 259.

- 411 [41] Jairo RL de Almeida and David J Thouless. “Stability of the Sherrington-Kirkpatrick solution of a spin
412 glass model”. In: Journal of Physics A: Mathematical and General 11.5 (1978), p. 983.
- 413 [42] Moritz Helias and David Dahmen. Statistical Field Theory for Neural Networks. Springer, 2020.

Supplementary Notes

S1 Dynamical Mean Field Theory

We study the model

$$x_{i,t+1} = \sum_{j=1}^N J_{ij} \phi(x_{j,t}) + \zeta_i + \eta_t, \quad (10)$$

where, as stated in the main text, $x_{i,t}$ is the individual neuronal activity at time t , $\phi(x)$ is the transfer function, ζ_i is quenched Gaussian noise with mean μ and variance σ^2 , η_t is a possible time-dependent external stimulus. The synaptic weights J_{ij} are randomly drawn from a Gaussian distribution of mean J_0/N and variance g^2/N .

For this model, the measure of the path integral is

$$\mathcal{D}\mathbf{x} = \prod_{i=1}^N \mathcal{D}x_i, \quad \mathcal{D}x_i = \sum_{t \in \mathbb{Z}} dx_{i,t}.$$

We apply dynamical mean field theory (DMFT) as described in Ref. [39]. The aim of DMFT is to obtain the single body density functional $P_1(x)$ or equivalently its moment generating functional, averaged over the randomness of the synaptic connections and the external noise in the infinite population limit $N \rightarrow \infty$. That is,

$$P_1(x_{1,t}) \equiv \int \langle P_N(\mathbf{x}) \rangle_{\zeta, J} \prod_{i=2}^N \mathcal{D}x_i$$

where $P_N(\mathbf{x}) \mathcal{D}\mathbf{x}$ is the N -body density functional. Calling $X_{i,t}[\zeta]$ the solution to the equations of motion (2) for a given modulation ζ , we have

$$\langle P_N(\mathbf{x}) \rangle = \left\langle \prod_{i=1}^N \left[\delta(x_{i,t} - X_{i,t}[\zeta]) \right] \right\rangle_{\zeta, J} = \left\langle \prod_{i=1}^N \left[\prod_t \delta(x_{i,t+1} - I_{i,t} - \eta_t - \zeta_i) \right] \right\rangle_{\zeta, J}$$

where $I_{i,t} = \sum_{j=1}^N J_{ij} \phi(x_{j,t})$ and we changed variables in the path integral noticing that the relevant Jacobian is equal to unity.

Let us now compute the generating functional $Z_N[l]$ over multiple trials or replicas α , written as a function of a control field l :

$$Z_N[l] = \int \Pi_{\alpha,i,t} dx_{i,t}^{\alpha} e^{\Sigma_{\alpha,i,t} l_{i,t}^{\alpha} x_{i,t}^{\alpha}} \left\langle \delta \left(x_{i,t+1}^{\alpha} - I_{i,t}^{\alpha} - \eta_t^{\alpha} - \zeta_i \right) \right\rangle_{\zeta,J}$$

We express the delta function as a Fourier transform, perform the Gaussian integral over the modulation vectors

ζ , proceed with standard path integral manipulations, and define

$$m_t^{\alpha} = \frac{1}{N} \sum_{j=1}^N \phi_{j,t}^{\alpha}, \quad Q_{ts}^{\alpha\beta} = \frac{1}{N} \sum_{j=1}^N \phi_{j,t}^{\alpha} \phi_{j,s}^{\beta}.$$

Taking the saddle point in the limit $N \rightarrow \infty$, we thus obtain a single body generating functional $Z_N[l] \rightarrow$

$\prod_i Z_1^{\text{MF}}[l_i]$, where MF stands for "mean field":

$$\begin{aligned} Z_1^{\text{MF}}[l] &= \exp \left(-\frac{1}{2} \sum_{\alpha,\beta} \sum_{t,s} l_{t+1}^{\alpha} Q_{ts}^{\alpha\beta}(\eta) l_{s+1}^{\beta} + i \sum_{\alpha} \sum_t l_{t+1}^{\alpha} (m_t^{\alpha}(\eta) + \eta_t^{\alpha}) \right) \\ &= \exp \left(-\frac{1}{2} \sum_{\alpha,\beta} \sum_{t,s} l_{t+1}^{\alpha} \left(Q_{ts}^{\alpha\beta}(0) + Q_{ts,\alpha}^{\alpha\beta}(0) \eta_t^{\alpha} + Q_{ts,\beta}^{\alpha\beta}(0) \eta_t^{\beta} + \frac{Q_{ts,\alpha\alpha}^{\alpha\beta}(0)}{2} (\eta_t^{\alpha})^2 + Q_{ts,\alpha\beta}^{\alpha\beta}(0) \eta_t^{\alpha} \eta_s^{\beta} + \frac{Q_{ts,\beta\beta}^{\alpha\beta}(0)}{2} (\eta_t^{\beta})^2 \right) l_{s+1}^{\beta} \right. \\ &\quad \left. + i \sum_{\alpha} \sum_t l_t^{\alpha} \left(m_t^{\alpha}(0) + m_{0,\alpha}^{\alpha}(0) \eta_t^{\alpha} + \frac{m_{0,\alpha\alpha}^{\alpha}(0)}{2} (\eta_t^{\alpha})^2 + \eta_t^{\alpha} \right) + O(\eta^3) \right), \end{aligned} \quad (11)$$

where the subscripts $(, \alpha)$ and $(, \alpha\beta)$ are respectively $\partial/\partial\eta_t^{\alpha}$ and $\partial^2/\partial\eta_t^{\alpha}\partial\eta_s^{\beta}$; for instance, we have

$$\begin{aligned} Q_{ts}^{\alpha\beta}(0) &= \sigma^2 c_{ts} + \langle \phi_t^{\alpha} \phi_s^{\beta} \rangle|_{\eta=0}, \quad Q_{ts,\alpha}^{\alpha\beta}(0) = \langle \phi_t^{\alpha} \phi_s^{\beta} \rangle|_{\eta=0}, \quad Q_{ts,\beta}^{\alpha\beta}(0) = \langle \phi_t^{\alpha} \phi_s^{\beta} \rangle|_{\eta=0}, \\ m_t^{\alpha}(0) &= J \langle \phi_t^{\alpha} \rangle|_{\eta=0} \quad m_{t,\alpha}^{\alpha}(0) = J \langle \phi_t^{\alpha} \rangle|_{\eta=0}. \end{aligned} \quad (12)$$

In terms of the generating functional, we finally obtain self-consistent equations for the parameters

$$M_t^{\alpha} = \langle x_t^{\alpha} \rangle = -i \frac{\delta Z_1^{\text{MF}}}{\delta l_t^{\alpha}} \Big|_{l=0}, \quad C_{ts}^{\alpha\beta} = \langle x_t^{\alpha} x_s^{\beta} \rangle - M_t^{\alpha} M_s^{\beta} = -\frac{\delta^2 Z_1^{\text{MF}}}{\delta l_t^{\alpha} \delta l_s^{\beta}} \Big|_{l=0} - M_t^{\alpha} M_s^{\beta}, \quad (13)$$

which are explicitly written as follows,

$$M_{t+1}^{\alpha} = J \langle \phi(x_t^{\alpha}) \rangle, \quad C_{t+1,s+1}^{\alpha\beta} = \sigma^2 + \langle \phi(x_t^{\alpha}) \phi(x_s^{\beta}) \rangle, \quad (14)$$

where the indices α, β differentiate the individual replicas.

The terms $\langle \phi(x_t) \rangle$ and $\langle \phi(x_t)\phi(x_s) \rangle$ are explicitly written as

$$\begin{aligned} \langle \phi(x_t^\alpha) \rangle &= \int \phi \left(\sqrt{C_{tt}^{\alpha\alpha}} x + M_t^\alpha + \eta_t^\alpha \right) Dx, \\ \langle \phi(x_t^\alpha)\phi(x_s^\beta) \rangle &= \iint \phi \left(\sqrt{C_{tt}^{\alpha\beta}} x + M_t^\alpha + \eta_t^\alpha \right) \phi \left(\frac{C_{ts}^{\alpha\beta}}{\sqrt{C_{tt}^{\alpha\alpha}}} x + \sqrt{\frac{C_{tt}^{\alpha\alpha}C_{ss}^{\beta\beta} - (C_{ts}^{\alpha\beta})^2}{C_{tt}^{\alpha\alpha}}} y + M_s^\beta + \eta_s^\beta \right) Dx Dy, \end{aligned} \quad (15)$$

where $Dx = \exp(-x^2/2)dx/\sqrt{2\pi}$. This is because $\{x_t\}_{t \in \mathbb{Z}}$ is shown to be a Gaussian random variable whose covariance and mean value is determined self-consistently by use of the generating functional method and by taking mean-field limit $N \rightarrow \infty$.

From Eqs. 14, we derive

$$M = J \int \frac{dx}{\sqrt{2\pi}} e^{-x^2/2} \phi \left(\sqrt{C} x + M \right) \quad (16)$$

for the fixed point $M = \lim_{t \rightarrow \infty} M_t^\alpha$ and

$$C = \sigma^2 + \int \frac{dx}{\sqrt{2\pi}} e^{-x^2/2} \phi \left(\sqrt{C} x + M \right)^2 \quad (17)$$

for $C = \lim_{t \rightarrow \infty} C_{tt}^{\alpha\alpha}$. The inter-replica correlation $C_{tt}^{\alpha\beta}$ can also be written from the above. Finally, the response to the external force η can be computed systematically as

$$\left. \frac{\partial \langle x_{t_1}^{\alpha_1} \cdots x_{t_n}^{\beta_n} \rangle}{\partial \eta_t^\gamma} \right|_{\eta=0} = i(-1)^n \left. \frac{\delta^{n+2} Z_1^{\text{MF}}[l]}{\delta l_{t_0}^{\alpha_0} \cdots \delta l_{t_n}^{\alpha_n} \delta \eta_t^\gamma} \right|_{l=0, \eta=0} \quad (18)$$

in the infinite population limit.

S2 Heuristic derivation of conditions for stability

For arbitrary functions ϕ and ψ , we define

$$\begin{aligned} \langle \phi^\alpha \psi^\beta \rangle_{ts} &\equiv \langle \phi(x_t^\alpha) \psi(x_s^\beta) \rangle = \\ &= \iint \phi \left(\sqrt{C_{tt}^{\alpha\beta}} x + M_t^\alpha + \eta_t^\alpha \right) \psi \left(\frac{C_{ts}^{\alpha\beta}}{\sqrt{C_{tt}^{\alpha\alpha}}} x + \sqrt{\frac{C_{tt}^{\alpha\alpha} C_{ss}^{\beta\beta} - (C_{ts}^{\alpha\beta})^2}{C_{tt}^{\alpha\alpha}}} y + M_s^\beta + \eta_s^\beta \right) Dx Dy, \end{aligned} \quad (19)$$

with $Dx = \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx$

It is easy to see (through integration by parts) that the variation of this quantity under perturbations of M^α and $C^{\alpha\alpha}$ is (omitting time labels for brevity)

$$d\langle \phi^\alpha \psi^\beta \rangle = \langle \phi'^\alpha \psi^\beta \rangle dM^\alpha + \langle \phi^\alpha \psi'^\beta \rangle dM^\beta + \frac{1}{2} \langle \phi''^\alpha \psi^\beta \rangle dC^{\alpha\alpha} + \frac{1}{2} \langle \phi^\alpha \psi''^\beta \rangle dC^{\beta\beta} + \langle \phi'^\alpha \psi'^\beta \rangle dC^{\alpha\beta} \quad (20)$$

The single-replica stability is understood as follows. Using identity 20 for the quantity $\langle \phi \psi \rangle_0 = \langle \phi^\alpha \psi^\alpha \rangle_t$, it is seen that the linearized version of the single-replica equation around the steady state $C_{tt} = C^{\alpha\alpha}$, $M_t = M^\alpha$, becomes

$$\begin{pmatrix} \delta M_{t+1}^\alpha \\ \delta C_{t+1,t+1}^{\alpha\alpha} \end{pmatrix} = A \begin{pmatrix} \delta M_t^\alpha \\ \delta C_{tt}^{\alpha\alpha} \end{pmatrix}, \quad A = \begin{pmatrix} J\langle \phi' \rangle_0 & J\langle \phi'' \rangle_0/2 \\ 2\langle \phi \phi' \rangle_0 & \langle \phi \phi'' \rangle_0 + \langle \phi'^2 \rangle_0 \end{pmatrix} \quad (21)$$

It follows that the steady state is stable if the eigenvalues of A are in the unit circle.

From the above it is also possible to check the stability within one replica, yielding equations for the phase boundaries. The condition of the critical state, $C_{tt} = C$ and $M_t = M$, is indeed

$$\det \begin{pmatrix} J\langle \phi' \rangle_0 - 1 & J\langle \phi'' \rangle_0/2 \\ 2\langle \phi \phi' \rangle_0 & \langle \phi \phi'' \rangle_0 + \langle \phi'^2 \rangle_0 - 1 \end{pmatrix} = 0. \quad (22)$$

This criticality found within single-replica is on the edge of the coexistence region, not at the edge of chaos in general. In the systems dealt with in Refs. [40, 18, 23], this criticality appears on the edge of chaos due to the symmetry and absence of random noise.

We next consider the stability against the inter-replica perturbation. Invoking once again identity 20, the

linearized equation here is found to be

$$\delta C_{t+1,s+1}^{\alpha\beta} = \langle \phi'_\alpha \phi_\beta \rangle \delta M_t^\alpha + \langle \phi_\alpha \phi'_\beta \rangle \delta M_s^\beta + \langle \phi''_\alpha \phi_\beta \rangle \frac{\delta C_{tt}^{\alpha\alpha}}{2} + \langle \phi_\alpha \phi''_\beta \rangle \frac{\delta C_{ss}^{\beta\beta}}{2} + \langle \phi'_\alpha \phi'_\beta \rangle \delta C_{ts}^{\alpha\beta}; \quad (23)$$

assuming that the system is stable against the intra-replica perturbations $\delta M_t^\alpha, \delta C_{tt}^{\alpha\alpha}$ these perturbations converge to 0 so that the linearized equation asymptotically is

$$\delta C_{t+1,s+1}^{\alpha\beta} = \langle \phi'_\alpha \phi'_\beta \rangle \delta C_{ts}^{\alpha\beta}. \quad (24)$$

Summarizing the above discussion, the steady state is stable if and only if the eigenvalues of the matrix A are in unit-circle and the inequality $\langle \phi'_\alpha \phi'_\beta \rangle < 1$ holds.

The single-replica stability is visualized for a range of model parameters in the lowest panels of Fig. 1.

S3 Field theoretical stability analysis

The stability analysis can also be performed by checking the definiteness of the Hessian matrix around the saddle point [18, 41], e.g. along the lines of Ref. [18]. We will use the abbreviation

$$\sum_{t,s,u,v} \sum_{\alpha,\beta,\gamma,\delta} \langle f(x_t^\alpha, x_s^\beta, x_u^\gamma, x_v^\delta) \rangle = \sum_{\alpha,\beta,\gamma,\delta} \langle f(x^\alpha, x^\beta, x^\gamma, x^\delta) \rangle, \quad (25)$$

that is, we do not write time parameters (t, s, \dots) explicitly unless it is necessary, and we only write explicitly the replica parameters as represented by Greek characters. In addition, we will abbreviate $\phi(x_t^\alpha)$ by ϕ^α .

With this notation, the generating functional is

$$\begin{aligned} Z_N[l] = & \int \mathcal{D}Q \mathcal{D}\tilde{Q} \mathcal{D}m \mathcal{D}\tilde{m} \exp \left(iN \sum_{\alpha,\beta} \tilde{Q}^{\alpha\beta} Q^{\alpha\beta} + iN \sum_{\alpha} \tilde{m}^{\alpha} m^{\alpha} \right) \\ & \times \prod_{i=1}^N \int \mathcal{D}\tilde{x} \mathcal{D}x \exp \left(i \sum_{\alpha} \tilde{x}^{\alpha} (Dx^{\alpha} - Jm^{\alpha} - \eta_i^{\alpha}) - \sum_{\alpha,\beta} \frac{Q^{\alpha\beta}}{2} \tilde{x}^{\alpha} \tilde{x}^{\beta} - i \sum_{\alpha} \tilde{m}^{\alpha} \phi^{\alpha} - i \sum_{\alpha,\beta} \tilde{Q}^{\alpha\beta} \phi^{\alpha} \phi^{\beta} + i \sum_{\alpha} l_i^{\alpha} x^{\alpha} \right) \end{aligned} \quad (26)$$

where $Dx_t^{\alpha} = x_{t+1}^{\alpha}$.

Let us expand it around the saddle point $\tilde{Q}_{\text{SP}}, Q_{\text{SP}}, \tilde{m}_{\text{SP}}, m_{\text{SP}}$ with respect to the fluctuations $\tilde{q}, q, \tilde{\mu}, \mu$ and

476 take the 2nd-order variation.

$$\begin{aligned}
 Z_N[l] &= \int \mathcal{D}\tilde{q}\mathcal{D}q\mathcal{D}\tilde{\mu}\mathcal{D}\mu \exp \left(iN \sum_{\alpha\beta} \left(\tilde{q}^{\alpha\beta} Q_{\text{SP}}^{\alpha\beta} + \tilde{q}^{\alpha\beta} q^{\alpha\beta} \right) + iN \sum_{\alpha} \left(\tilde{\mu}^{\alpha} m^{\alpha} + \tilde{\mu}^{\alpha} \mu^{\alpha} \right) \right) \\
 &\times \prod_{i=1}^N \left[\int \mathcal{D}\tilde{x}\mathcal{D}x \exp \left(i \sum_{\alpha} \tilde{x}^{\alpha} (Dx^{\alpha} - Jm_{\text{SP}}^{\alpha} - \eta_i^{\alpha}) - \sum_{\alpha,\beta} \frac{Q^{\alpha\beta}}{2} \tilde{x}^{\alpha} \tilde{x}^{\beta} + i \sum_{\alpha} l_i^{\alpha} x^{\alpha} \right) \right. \\
 &\times \left(1 - i \sum_{\alpha,\beta} \tilde{q}^{\alpha\beta} \phi^{\alpha} \phi^{\beta} - \sum_{\alpha,\beta} \frac{q^{\alpha\beta}}{2} \tilde{x}^{\alpha} \tilde{x}^{\beta} - i \sum_{\alpha} \tilde{\mu}^{\alpha} \phi^{\alpha} - iJ \sum_{\alpha} \mu^{\alpha} \tilde{x}^{\alpha} \right. \\
 &\quad \left. \left. + \frac{1}{2} \left(i \sum_{\alpha,\beta} \tilde{q}^{\alpha\beta} \phi^{\alpha} \phi^{\beta} + \sum_{\alpha,\beta} \frac{q^{\alpha\beta}}{2} \tilde{x}^{\alpha} \tilde{x}^{\beta} + i \sum_{\alpha} \tilde{\mu}^{\alpha} \phi^{\alpha} + iJ \sum_{\alpha} \mu^{\alpha} \tilde{x}^{\alpha} \right)^2 \right) \right] \\
 &= \int \mathcal{D}\tilde{q}\mathcal{D}q\mathcal{D}\tilde{\mu}\mathcal{D}\mu \exp \left(iN \sum_{\alpha\beta} \left(\tilde{q}^{\alpha\beta} Q_{\text{SP}}^{\alpha\beta} + \tilde{q}^{\alpha\beta} q^{\alpha\beta} \right) + iN \sum_{\alpha} \left(\tilde{\mu}^{\alpha} m^{\alpha} + \tilde{\mu}^{\alpha} \mu^{\alpha} \right) \right) \\
 &\times \exp \left(N \ln \left(1 - i \sum_{\alpha,\beta} \tilde{q}^{\alpha\beta} \langle \phi^{\alpha} \phi^{\beta} \rangle - i \sum_{\alpha} \tilde{\mu}^{\alpha} \langle \phi^{\alpha} \rangle \right. \right. \\
 &\quad - \frac{1}{2} \sum_{\alpha,\beta,\gamma,\delta} \tilde{q}^{\alpha\beta} \langle \phi^{\alpha} \phi^{\beta} \phi^{\gamma} \phi^{\delta} \rangle \tilde{q}^{\gamma\delta} - \frac{1}{2} \sum_{\alpha,\beta} \mu^{\alpha} \langle \phi^{\alpha} \phi^{\beta} \rangle \mu^{\beta} - \sum_{\alpha,\beta,\gamma} \tilde{q}^{\alpha\beta} \langle \phi^{\alpha} \phi^{\beta} \phi^{\gamma} \rangle \mu^{\gamma} \\
 &\quad \left. \left. + \frac{i}{2} \sum_{\alpha\beta\gamma\delta} \tilde{q}^{\alpha\beta} \langle \phi^{\alpha} \phi^{\beta} \tilde{x}^{\gamma} \tilde{x}^{\delta} \rangle q^{\gamma\delta} - \sum_{\alpha\beta} \tilde{\mu}^{\alpha} J \langle \phi^{\alpha} \tilde{x}^{\beta} \rangle \mu^{\beta} - \sum_{\alpha,\beta,\gamma} \tilde{q}^{\alpha\beta} \langle \phi^{\alpha} \phi^{\beta} \tilde{x}^{\gamma} \rangle \mu^{\gamma} - i \sum_{\alpha,\beta,\gamma} \tilde{\mu}^{\alpha} \langle \phi^{\alpha} \tilde{x}^{\beta} \tilde{x}^{\gamma} \rangle q^{\beta\gamma} \right) \right] \quad (27)
 \end{aligned}$$

477 It should be noted that i -dependence coming from η_i and l_i is included in the average $\langle \bullet \rangle$, which may otherwise
 478 seems to vanish in the last line of this equation. Using the expansion formula $\ln(1 + \epsilon) = \epsilon - \frac{\epsilon^2}{2} + O(\epsilon^3)$, omitting
 479 the 3rd order of fluctuations, and using the saddle point condition, we have the 2nd order variation around the
 480 saddle point,

$$\begin{aligned}
 Z_N[l] &\propto \int \mathcal{D}\tilde{q}\mathcal{D}q\mathcal{D}\tilde{\mu}\mathcal{D}\mu \exp \left(i \sum_{\alpha,\beta} \tilde{\mu}^{\alpha} (\delta_{\alpha\beta} + iJ \langle \phi^{\alpha} \tilde{x}^{\beta} \rangle) \mu^{\beta} + i \sum_{\alpha,\beta,\gamma,\delta} \tilde{q}^{\alpha\beta} \left(\delta_{\alpha\gamma} \delta_{\beta\delta} + \frac{1}{2} \langle \phi^{\alpha} \phi^{\beta} \tilde{x}^{\gamma} \tilde{x}^{\delta} \rangle \right) \right. \\
 &\quad + \frac{i}{2} \sum_{\alpha,\beta,\gamma} \tilde{\mu}^{\alpha} \langle \phi^{\alpha} \tilde{x}^{\beta} \tilde{x}^{\gamma} \rangle q^{\beta\gamma} - \sum_{\alpha,\beta,\gamma} \tilde{q}^{\alpha\beta} \langle \phi^{\alpha} \phi^{\beta} \tilde{x}^{\gamma} \rangle \mu^{\gamma} \\
 &\quad - \frac{1}{2} \left[\sum_{\alpha,\beta} \tilde{\mu}^{\alpha} (\langle \phi^{\alpha} \phi^{\beta} \rangle - \langle \phi^{\alpha} \rangle \langle \phi^{\beta} \rangle) \tilde{\mu}^{\beta} + \sum_{\alpha,\beta,\gamma,\delta} \tilde{q}^{\alpha\beta} (\langle \phi^{\alpha} \phi^{\beta} \phi^{\gamma} \phi^{\delta} \rangle - \langle \phi^{\alpha} \phi^{\beta} \rangle \langle \phi^{\gamma} \phi^{\delta} \rangle) \tilde{q}^{\gamma\delta} \right. \\
 &\quad \left. \left. + 2 \sum_{\alpha,\beta,\gamma} \tilde{\mu}^{\alpha} (\langle \phi^{\alpha} \phi^{\beta} \phi^{\gamma} \rangle - \langle \phi^{\alpha} \rangle \langle \phi^{\beta} \phi^{\gamma} \rangle) \tilde{q}^{\beta\gamma} \right) \right] \quad (28)
 \end{aligned}$$

481 Let us now define the vectors, $\mathcal{V} = (\mu_t^{\alpha}, q_{su}^{\beta\gamma})_{t,s,u,\alpha,\beta,\gamma}$ and $\tilde{\mathcal{V}} = (\tilde{\mu}_t^{\alpha}, \tilde{q}_{su}^{\beta\gamma})_{t,s,u,\alpha,\beta,\gamma}$. Moreover, let the matrix

482 \mathcal{M} be

$$\mathcal{M} = \begin{pmatrix} (\langle \phi^\alpha \phi^\beta \rangle - \langle \phi^\alpha \rangle \langle \phi^\beta \rangle) & \langle \phi^\alpha \phi^\beta \phi^\gamma \rangle - \langle \phi^\alpha \rangle \langle \phi^\beta \phi^\gamma \rangle \\ \langle \phi^\alpha \phi^\beta \phi^\gamma \rangle - \langle \phi^\alpha \phi^\beta \rangle \langle \phi^\gamma \rangle & \langle \phi^\alpha \phi^\beta \phi^\gamma \phi^\delta \rangle - \langle \phi^\alpha \phi^\beta \rangle \langle \phi^\gamma \phi^\delta \rangle \end{pmatrix} \quad (29)$$

483 and the matrix \mathcal{A} be

$$\mathcal{A} = \begin{pmatrix} \delta_{\alpha\beta} + iJ\langle \phi^\alpha \tilde{x}^\beta \rangle & \frac{\langle \phi^\alpha \tilde{x}^\beta \tilde{x}^\gamma \rangle}{2} \\ i\langle \phi^\alpha \phi^\beta \tilde{x}^\gamma \rangle & \delta_{\alpha\gamma} \delta_{\beta\delta} + \frac{1}{2} \langle \phi^\alpha \phi^\beta \tilde{x}^\gamma \tilde{x}^\delta \rangle \end{pmatrix}. \quad (30)$$

484 By using them Eq. (28) can be written as

$$Z_N[l] = \int d\tilde{\mathcal{V}} \int d\mathcal{V} \exp \left(i\tilde{\mathcal{V}}^\dagger \mathcal{A} \mathcal{V} - \frac{1}{2} \tilde{\mathcal{V}}^\dagger \mathcal{M} \tilde{\mathcal{V}} \right) = \int d\mathcal{V} \exp \left(-\frac{1}{2} \mathcal{V}^\dagger \mathcal{A}^\dagger \mathcal{M}^{-1} \mathcal{A} \mathcal{V} \right) \quad (31)$$

485 The matrix \mathcal{M} is obviously positive definite because it is a covariance matrix. The second variation around the

486 saddle point is thus positive definite if and only if the operator \mathcal{A} has no vanishing eigenvalue. We next derive

487 the stability condition of the steady states by following Ref. [18]. Using the relation $D^\alpha \frac{\partial M^\alpha}{\partial m^\beta} = J\delta_{\alpha\beta}$ and

488 $D^\alpha \frac{\partial C^{\alpha\beta}}{\partial Q^{\gamma\delta}} D^{\beta T} = \delta_{\alpha\gamma} \delta_{\beta\delta}$, each element of $\mathcal{A}\mathcal{V}$ is written as

$$\begin{aligned} \sum_\beta (\delta_{\alpha\beta} + iJ\langle \phi^\alpha \tilde{x}^\beta \rangle) \mu^\beta &= J^{-1} \sum_\beta \left(D^\alpha \frac{\partial M^\alpha}{\partial m^\beta} \mu^\beta - J \frac{\partial \langle \phi^\alpha \rangle}{\partial M^\beta} \frac{\partial M^\beta}{\partial m^\beta} \mu^\beta \right) = J^{-1} (D^\alpha - J\langle \phi^{\alpha'} \rangle) \varphi^\alpha, \\ \frac{1}{2} \sum_{\beta, \gamma} \langle \phi^\alpha \tilde{x}^\beta \tilde{x}^\gamma \rangle q^{\beta\gamma} &= - \sum_{\beta, \gamma} \frac{\partial \langle \phi^\alpha \rangle}{\partial C^{\beta\gamma}} \frac{\partial C^{\gamma\beta}}{\partial Q^{\beta\gamma}} q^{\beta\gamma} = - \frac{\partial \langle \phi^\alpha \rangle}{\partial C^{\alpha\alpha}} \Psi^{\alpha\alpha} = -J^{-1} \frac{J\langle \phi^{\alpha''} \rangle}{2} \Psi^{\alpha\alpha}, \\ i \sum_\gamma J\langle \phi^\alpha \phi^\beta \tilde{x}^\gamma \rangle \mu^\gamma &= - \sum_\gamma \frac{\partial \langle \phi^\alpha \phi^\beta \rangle}{\partial M^\gamma} \frac{\partial M^\gamma}{\partial m^\gamma} \mu^\gamma = -\langle \phi^{\alpha'} \phi^\beta \rangle \varphi^\alpha - \langle \phi^\alpha \phi^{\beta'} \rangle \varphi^\beta, \\ \sum_{\gamma, \delta} \left(\delta_{\alpha\gamma} \delta_{\beta\delta} + \frac{1}{2} \langle \phi^\alpha \phi^\beta \tilde{x}^\gamma \tilde{x}^\delta \rangle \right) q^{\gamma\delta} &= D^\alpha \Psi^{\alpha\beta} D^{\beta T} - \frac{\partial \langle \phi^\alpha \phi^\beta \rangle}{\partial C^{\alpha\alpha}} \Psi^{\alpha\alpha} - \frac{\partial \langle \phi^\alpha \phi^\beta \rangle}{\partial C^{\beta\beta}} \Psi^{\beta\beta} - \frac{\partial \langle \phi^\alpha \phi^\beta \rangle}{\partial C^{\alpha\beta}} \Psi^{\alpha\beta} \\ &= D^\alpha \Psi^{\alpha\beta} D^{\beta T} - \frac{\langle \phi^{\alpha''} \phi^\beta \rangle}{2} \Psi^{\alpha\alpha} - \frac{\langle \phi^\alpha \phi^{\beta''} \rangle}{2} \Psi^{\beta\beta} - \langle \phi^{\alpha'} \phi^{\beta'} \rangle \Psi^{\alpha\beta} \end{aligned} \quad (32)$$

489 where $\varphi^\alpha = \frac{\partial M^\alpha}{\partial m^\alpha} \mu^\alpha$ and $\Psi^{\alpha\beta} = \frac{\partial C^{\alpha\beta}}{\partial Q^{\alpha\beta}} q^{\alpha\beta}$, where the operator $D^\alpha \bullet D^{\beta T}$ acts as $D^\alpha C_{ts}^{\alpha\beta} D^{\beta T} = C_{t+1, s+1}^{\alpha\beta}$ and

490 $D^\alpha M_t^\alpha = M_{t+1}^\alpha$.

491 The steady state is stable if and only if the eigenvalue equation

$$\mathfrak{A} \vec{v} = \Lambda \vec{v} \quad (33)$$

has no solution with the eigenvalue $\Lambda = 0$, where the five-dimensional vector \vec{v} stands for

$$\vec{v} = \left(\varphi_t^\alpha, \Psi_{tt}^{\alpha\alpha}, \varphi_s^\beta, \Psi_{ss}^{\beta\beta}, \Psi_{ts}^{\alpha\beta} \right)$$

and the operator \mathfrak{A} in Eq. (33) is given by

$$\mathfrak{A} = \begin{pmatrix} D^\alpha - J\langle\phi_t^{\alpha'}\rangle & -\frac{J}{2}\langle\phi_t^{\alpha''}\rangle & 0 & 0 & 0 \\ -2\langle\phi_t^\alpha\phi_t^{\alpha'}\rangle & D^\alpha \bullet D^{\alpha T} - (\langle\phi_t^{\alpha'2}\rangle + \langle\phi_t^\alpha\phi_t^{\alpha''}\rangle) & 0 & 0 & 0 \\ 0 & 0 & D^\beta - J\langle\phi_s^{\beta'}\rangle & -\frac{J}{2}\langle\phi_s^{\beta''}\rangle & 0 \\ 0 & 0 & -2\langle\phi_s^\beta\phi_s^{\beta'}\rangle & D^\beta D^\beta - (\langle\phi_s^{\beta'}\phi_s^{\beta'}\rangle + \langle\phi_s^\beta\phi_s^{\beta''}\rangle) & 0 \\ -\langle\phi_t^{\alpha'}\phi_s^\beta\rangle & -\frac{1}{2}\langle\phi_t^{\alpha''}\phi_s^\beta\rangle & -\langle\phi_t^\alpha\phi_s^{\beta'}\rangle & -\frac{1}{2}\langle\phi_t^\alpha\phi_s^{\beta''}\rangle & D^\alpha \bullet D^{\beta T} - \langle\phi_t^{\alpha'}\phi_s^{\beta'}\rangle \end{pmatrix}$$

acting onto the vector $(\delta M_t^\alpha, \delta C_{tt}^{\alpha\alpha}, \delta M_s^\beta, \delta C_{ss}^{\beta\beta}, \delta C_{ts}^{\alpha\beta})^T$.

We first examine the stability of a steady solution, $C_{ts}^{\alpha\beta} = C_0\delta_{ts} + C_\infty(1 - \delta_{ts})$, $M_t^\alpha = M$. We have to check

if there exists a solution to the following equation when $\Lambda = 0$

$$\varphi_{t+1} - J\langle\phi'\rangle\varphi_t - \frac{J\langle\phi''\rangle}{2}\Psi_{tt} = \Lambda\varphi_t \quad (34)$$

$$\Psi_{t+1,t+1} - 2\langle\phi\phi'\rangle\varphi_t - (\langle\phi'^2\rangle + \langle\phi\phi''\rangle)\Psi_{tt} = \Lambda\Psi_{tt}$$

Let the Z -transformation of φ_t and Ψ_{tt} be respectively

$$\tilde{\varphi}_z = \sum_t \varphi_t z^{-t}, \quad \tilde{\Psi}_\zeta = \sum_t \Psi_{tt} \zeta^{-t}, \quad |z|, |\zeta| > 1. \quad (35)$$

In matrix form, the system of equations (34) can be written as

$$\begin{pmatrix} z - J\langle\phi'\rangle & -\frac{J\langle\phi''\rangle}{2} \\ -2\langle\phi\phi'\rangle & \zeta - \langle\phi'^2\rangle - \langle\phi\phi''\rangle \end{pmatrix} \begin{pmatrix} \tilde{\varphi}_z \\ \tilde{\Psi}_\zeta \end{pmatrix} = \Lambda \begin{pmatrix} \tilde{\varphi}_z \\ \tilde{\Psi}_\zeta \end{pmatrix}. \quad (36)$$

If the eigenvalue $\Lambda = 0$ exists, the equation

$$(z - J\langle\phi'\rangle)(\zeta - \langle\phi'^2\rangle - \langle\phi\phi''\rangle) - J\langle\phi\phi'\rangle\langle\phi''\rangle = 0 \quad (37)$$

holds true for some z, ζ satisfying $|z|, |\zeta| > 1$. Now ϕ satisfies $\langle\phi'\rangle > 0$ and $\langle\phi'^2\rangle + \langle\phi\phi''\rangle > 0$; consequently, the

steady state is stable against the intra-replica perturbation if

$$(1 - J\langle\phi'\rangle)(1 - \langle\phi'^2\rangle - \langle\phi\phi''\rangle) - J\langle\phi\phi'\rangle\langle\phi''\rangle > 0. \quad (38)$$

When the steady state is a fixed point (time independent, $C_\infty = C_0$), the stability criterion within a single

replica can be checked by use of Eq. (38). If the steady state is time-dependent $C_\infty < C_0$, we have to consider

the stability against the perturbation $\Psi_{ts}^{\alpha\alpha}$. If the inequality (38) holds true, the possible existence of a vanishing

eigenvalue $\Lambda = 0$ can be brought about by

$$\Psi_{t+1,s+1} - \langle\phi'_t\phi'_s\rangle\Psi_{ts} = \Lambda\Psi_{ts}. \quad (39)$$

As we saw, the matrix \mathfrak{A} appearing in Eq. 33 has the form

$$\begin{pmatrix} \mathfrak{A}^{\alpha\alpha} & 0 & 0 \\ 0 & \mathfrak{A}^{\beta\beta} & 0 \\ \mathfrak{B}^\alpha & \mathfrak{B}^\beta & z\zeta - \langle \phi^\alpha \phi^{\beta'} \rangle \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} \mathfrak{A}_{tt} & 0 & 0 \\ 0 & \mathfrak{A}_{ss} & 0 \\ \mathfrak{B}_t & \mathfrak{B}_s & z\zeta - \langle \phi_t \phi_{s'} \rangle \end{pmatrix}; \quad (40)$$

hence, once the stability against the perturbation $(\delta M_t^\alpha, \delta C_{tt}^{\alpha\alpha})$ is shown, the instability (possibility of the vanishing eigenvalue) can come from only the terms $z\zeta - \langle \phi^\alpha \phi^{\beta'} \rangle$ or $z\zeta - \langle \phi_t \phi_{s'} \rangle$ respectively.

We next consider the stability against the perturbation $\Psi_{ts}^{\alpha\beta}$, that is, the vector $(0, 0, 0, 0, \Psi_{ts}^{\alpha\beta})$. In this case, we have to look at the equation

$$\Psi_{t+1.s+1}^{\alpha\beta} - \langle \phi_t^{\alpha'} \phi_s^{\beta'} \rangle \Psi_{ts}^{\alpha\beta} = \Lambda \Psi_{ts}^{\alpha\beta} \quad (41)$$

Taking the Z -transformation $\hat{\Psi}_{z\zeta} = \sum_{t,s} \Psi_{ts}^{\alpha\beta} z^{-1} \zeta^{-1}$ which is defined when $|z|, |\zeta| > 1$, we have

$$\left(z\zeta - \langle \phi_t^{\alpha'} \phi_s^{\beta'} \rangle - \Lambda \right) \hat{\Psi}_{z\zeta} = 0 \quad (42)$$

We conclude that The steady state is stable against the inter-replica perturbation $\delta C_{ts}^{\alpha\beta}$, if and only if $1 - \langle \phi^{\alpha'} \phi^{\beta'} \rangle > 0$; Hence we derive the stability condition

$$1 - \langle \phi_\infty^{\alpha'} \phi_0^{\beta'} \rangle > 0. \quad (43)$$

S4 Derivation of the Formula for the Critical Memory

The meaning of information processing in dynamical systems has become the subject of a vast literature, well summarized in references [36] and [37].

Within reference [36] two possible definitions are given of the memory capacity of a dynamical system. The first one (Eq. 6) does not include any preliminary shifting of mean levels, while the second one (Eq 2.1 of Supplementary Material in Ref [36]) is equivalent to the definition of Ref. [37] and is more natural from the view point of signal processing. An observer in possession of an unbiased estimator for the mean may remove the mean values from all the time series he records; what matters is the relationships between those mean-removed observations and the mean-removed version of the unobserved underlying process. Moreover, we would like the resulting memory capacity to be zero when the linear readout is dominated by a constant baseline value, because nothing can be learned from a readout independent on the input. Adopting therefore the mean-removed formula, we find for the memory capacity \mathcal{M} in the neighborhood of the second-order phase transition boundary

$$\mathcal{M} \sim \frac{1}{1 - \langle \phi'^\alpha \phi'^\beta \rangle} \quad (44)$$

as given in the main text.

To derive this formula, we proceed along the same lines as in Ref. [24], considering the input signal u_t as $u_t = \frac{1}{N} \sum_t \xi_{i,t}$ and trying to re-construct the input $u(t_0)$ with the sparse linear readout $\sum_{j=1}^K w_j x_{j,t}$ with $O(K) < O(\sqrt{N})$. The memory curve C_τ and capacity C_M are given respectively by the determinant coefficient which measures how well the readout neurons reconstruct the past input $u(t - \tau)$ correctly, and their sum [37]

$$C_\tau = \frac{\sum_{i,j=1}^K \text{Cov}_t(u_t, x_{i,t+\tau}) \text{Cov}_t(x_{i,t}, x_{j,t})^{-1} \text{Cov}_t(u_t, x_{j,t+\tau})}{\text{Var}_t(u_t)}, \quad (45)$$

$$C_M = \sum_{\tau} C_\tau.$$

The read out is sparse, so that the covariance $\text{Cov}_t(x_i(t), x_j(t))$ becomes diagonal in the infinite population limit $N \rightarrow \infty$ [23]. Moreover, we deal with the steady state so that this term is constant with respect to time.

We then have to compute $\sum_{i=1}^K \langle \langle x_{i,t} u_{t-\tau} \rangle_t^2 \rangle_J$. As shown in the Appendix in Ref. [24], when the input signal is a weighted sum of Gaussian random variables, the term $\langle x_{i,t} u_{t-\tau} \rangle$ is given by the linear combination of

533 $\langle x_{i,t} \tilde{x}_{j,t-\tau} \rangle_t$, which is the zero-field susceptibility of the parameter $\langle x_{i,t} \rangle = M_i$, $\chi_{i,\tau} = \left. \frac{\partial M_i}{\partial \eta_{j,t-\tau}} \right|_{\eta_j=0}$
 534 Let the signal be $u_t = \sum_j v_j \xi_{j,t}$. Since we are interested in computing $\langle x_{i,t} u_{t_0} \rangle$, let's proceed throughout
 535 the standard field-theoretical step of inserting an exponential source term for this quantity in side the general
 536 functional, to then differentiate by the relevant parameter. The suitable source term is

$$\exp \left(-i \sum_t k_t \sum_i v_i \xi_{i,t} \right). \quad (46)$$

537 Inserting it into the generating functional, we have

$$Z_N[l, k](J) = \iint \mathcal{D}x \mathcal{D}\tilde{x} \prod_{i=1}^N \exp \left(i \sum_t \tilde{x}_{i,t} (x_{i,t} - I_{i,t} - \zeta_i - \xi_{i,t}) \right) \exp \left(-i \sum_t k_t \sum_i v_i \xi_{i,t} - i \sum_{j,t} l_{j,t} x_{j,t} \right) \quad (47)$$

538 where $I_{i,t}^\alpha = \sum_{j=1}^N J_{ij} \phi(x_{j,t}^\alpha)$, and ζ_i is quenched randomness whose mean and covariance are μ and $\sigma \delta_{ij}$
 539 respectively. Taking average over the dynamical noisy input $\xi_{i,t}$ satisfying $\langle \xi_{i,t} \rangle_\xi = 0$ and $\langle \xi_{i,t} \xi_{j,s} \rangle_\xi = \sigma_{\text{in}} \delta_{ij} \delta_{ts}$
 540 we have

$$\langle Z_N[l, k](J) \rangle_\xi = \iint \mathcal{D}x \mathcal{D}\tilde{x} \prod_{i=1}^N \exp \left(i \sum_t \tilde{x}_{i,t} (x_{i,t} - I_{i,t} - \zeta_i) - \sum_t \frac{\sigma^2}{2} \tilde{x}_{i,t}^2 - \sigma^2 \sum_t k_t v_i \tilde{x}_{i,t} + O(k^2) \right). \quad (48)$$

541 Thus, the term $\langle x_{i,t} u_{t_0} \rangle$ is found to be given by the weighted sum of the linear responses $\langle x_{i,t} \tilde{x}_{j,t_0} \rangle$ as

$$\langle x_{i,t} u_{t_0} \rangle_\xi = (-i)^2 \frac{\delta^2 Z_N[l, k](J)}{\delta l_{i,t} \delta k_{t_0}} \Big|_{l=k=0} = -i \sigma_{\text{in}}^2 \sum_{j=1}^N v_j \langle x_{i,t} \tilde{x}_{j,t_0} \rangle(J). \quad (49)$$

542 The next quantity needed is $\langle h_{i,t,s,t_0} \rangle_J = \langle \langle x_{i,t}^1 u_{t_0}^1 x_{i,s}^2 u_{t_0}^2 \rangle_{\xi, \zeta} \rangle_J$. To compute this, we insert in the generating
 543 functional the single source

$$\exp \left(-\sigma_{\text{in}}^4 \sum_{j,j'} v_j v_{j'} \sum_t r_t \tilde{x}_{j,t}^1 \tilde{x}_{j',t}^2 \right), \quad (50)$$

544 This turns the the generating functional into

$$Z[l, r] = \iint \prod_{\alpha=1,2} \mathcal{D}x^\alpha \mathcal{D}\tilde{x}^\alpha \prod_{i=1}^N \exp \left(i \sum_t \tilde{x}_{i,t} (x_{i,t}^\alpha - I_{i,t}^\alpha - \zeta_i - \xi_{i,t}) \right) \exp \left(-\sigma^4 \sum_{j,j'} v_j v_{j'} \sum_t r_t \tilde{x}_{j,t}^1 \tilde{x}_{j',t}^2 \right), \quad (51)$$

545 which makes additional perturbation to the inter-replica correlation (see Eq. (G11) in Ref. [24]).

Let v_i be $\sim 1/N$. The form of r_t is assumed to be $r_\tau = r_0 \delta_{t,t_0}$. Using it, h_{i,t,s,t_0} is written as

$$\langle h_{i,t,s,t_0} \rangle_J = \frac{\delta}{\delta r_t} \langle x_{i,t}^1 x_{i,s}^2 \rangle_{\xi,\zeta,J} \bigg|_{r_0=0} = \frac{\delta}{\delta r_t} (\langle x_{i,t}^1 x_{i,s}^2 \rangle_{\xi,\zeta,J} - \langle x_{i,t}^1 \rangle_{\xi,\zeta,J} \langle x_{i,s}^2 \rangle_{\xi,\zeta,J}) \bigg|_{r_0=0} \quad (52)$$

The last equality is brought about by $\delta \langle x_{i,t}^\alpha \rangle_{\xi,\zeta,J} / \delta r_t|_{r_0=0} = 0$ (which is derived through Wick's theorem [42]

due to $x_{i,t}$ being Gaussian random variables when we take the infinite population limit $N \rightarrow \infty$) and by use of the

causality or normalization condition which gives $\langle \tilde{x}_{i,t}^\alpha \rangle_J = \langle \tilde{x}_{i,t}^\alpha \tilde{x}_{i,t}^\beta \rangle_J = 0$.

Further, it should be noted that h_{i,t,s,t_0} is a perturbation brought about by the additional source term (50), so

that

$$\frac{\delta}{\delta r_t} \langle \phi(x_{i,t}^1) \phi(x_{i,s}^2) \rangle_J \bigg|_{r_0=0} = \langle \phi'(x_{i,t}^1) \phi'(x_{i,s}^2) \rangle_{\xi,\zeta,J} \frac{\delta C_{ts}^{12}}{\delta r_t} \bigg|_{r_0=0} = \langle \phi'(x_{i,t}^1) \phi'(x_{i,s}^2) \rangle_{\xi,\zeta,J} \langle h_{i,t,s,t_0} \rangle_J. \quad (53)$$

Let h_{t,s,t_0}^M be

$$h_{t,s,t_0}^M = \sum_{i=1}^M \langle h_{i,t,s,t_0} \rangle_J \quad (54)$$

for $M = 1, 2, \dots, N$. What we desire is h_{t,t,t_0}^K , which satisfies

$$h_{t+1,t+1,t_0}^K = \frac{K}{N} \langle \phi'(x_t^1) \phi'(x_t^2) \rangle_J h_{t,t,t_0}^N + \frac{K}{N^2} \sigma_{\text{in}}^4 \delta_{t,t_0}, \quad (55)$$

where the last term is coming from the random inputs

$$\sigma_{\text{in}}^4 r(t) v_i^2 \delta_{ts} \delta_{t,t_0} + \sigma_{\text{in}}^2 \delta_{ts} + \sigma^2, \quad v_i = 1/N \quad (56)$$

The term h_{t,t,t_0}^N in the left hand side evolves as

$$h_{t+1,t+1,t_0}^N = \langle \phi'(x_t^1) \phi'(x_t^2) \rangle_J h_{t,t,t_0}^N + \frac{\sigma^4}{N} \delta_{t,t_0}, \quad (57)$$

so that we have, in the steady state,

$$h_{t_0+\tau,t+\tau,t_0}^N = \frac{\sigma_{\text{in}}^4}{N} \langle \phi(x^1) \phi(x^2) \rangle^{\tau-1} \quad (58)$$

and further, we have

$$h_{t_0+\tau,t_0+\tau,t_0}^K = \frac{K \sigma_{\text{in}}^4}{N^2} \langle \phi(x^1) \phi(x^2) \rangle^{\tau-1} \quad (59)$$

558 from Eq. (55).

559 The memory curve C_τ is proportional to $h_{t_0+\tau, t_0+\tau, t_0}^K$, so that we conclude that the capacity $C = \sum_{\tau=1}^{\infty}$

560 satisfies

$$C_M \propto \sum_{\tau=1}^{\infty} h_{t_0+\tau, t_0+\tau, t_0}^K \propto \frac{1}{1 - \langle \phi'(x^1) \phi'(x^2) \rangle_J} \quad (60)$$