

## Spontaneous and stimulus-induced coherent states of critically balanced neuronal networks

Takashi Hayakawa<sup>1,2,\*</sup> and Tomoki Fukai<sup>1,3</sup><sup>1</sup>RIKEN Center for Brain Science, Wako, Saitama 3510198, Japan<sup>2</sup>School of Medicine, Nihon University, Itabashi, Tokyo 1738610, Japan<sup>3</sup>Okinawa Institute of Science and Technology, Onna-son, Okinawa 9040495, Japan

(Received 20 February 2018; revised manuscript received 13 December 2019; accepted 4 February 2020; published 4 March 2020)

How the information microscopically processed by individual neurons is integrated and used in organizing the behavior of an animal is a central question in neuroscience. The coherence of neuronal dynamics over different scales has been suggested as a clue to the mechanisms underlying this integration. Balanced strong excitation and inhibition may amplify microscopic fluctuations to a macroscopic level, thus providing a mechanism for generating coherent multiscale neuronal dynamics. Previous theories of brain dynamics, however, were restricted to cases in which inhibition dominated excitation and suppressed fluctuations in the macroscopic population activity. In the present study, we investigate the dynamics of neuronal networks at a critical point between excitation-dominant and inhibition-dominant states. In these networks, the microscopic fluctuations in neuronal activities are amplified by the strong excitation and inhibition to drive the macroscopic dynamics, while the macroscopic dynamics determine the statistics of the microscopic fluctuations. Developing a novel type of mean-field theory applicable to this class of interscale interactions, for which an analytical approach has previously been unknown, we show that the amplification mechanism generates spontaneous, irregular macroscopic rhythms similar to those observed in the brain. Through the same mechanism, microscopic inputs to a small number of neurons effectively entrain the dynamics of the whole network. These network dynamics undergo a probabilistic transition to a coherent state, as the magnitude of either the balanced excitation and inhibition or the external inputs is increased. Our mean-field theory successfully predicts the behavior of this model. Furthermore, we numerically demonstrate that the coherent dynamics can be used for state-dependent read-out of information from the network. These results show a novel form of neuronal information processing that connects neuronal dynamics on different scales, advancing our understanding of the circuit mechanisms of brain computing.

DOI: [10.1103/PhysRevResearch.2.013253](https://doi.org/10.1103/PhysRevResearch.2.013253)

## I. INTRODUCTION

The cerebral cortex and hippocampus, the areas believed to be the origin of the versatile intelligent functionality of the mammalian brain, exhibit characteristic activities on two different scales. On the microscopic scale, neurons in these areas display various temporal patterns of firing activities in response to external stimuli or to being driven internally. These activities are correlated with fine features of the information the animal is processing [1–5]. On the macroscopic scale, electroencephalograms (EEGs) and measurements of local-field potentials (LFPs) have revealed a diverse range of rhythmic activities. These vary in both frequency and amplitude, but they are clearly correlated with the behavioral states of the animal, such as its attention and arousal levels [6–9]. Furthermore, in recent years, increasing numbers of experimental results have suggested that *coherence* of activities

on these two scales—namely, the degree of temporal cross-correlation among the activities—is finely controlled, reflecting the mechanisms underlying the binding of sensory stimuli, sensori-motor coordination, and learning in behavioral tasks [10–18].

How patterns emerge in multiscale dynamics in highly nonlinear and nonequilibrium regimes has been a subject of active research in statistical physics. From this perspective, understanding the multiscale dynamics in the brain and their coherence can be considered as a challenge in statistical physics. Physicists have thus far constructed various models of the activities in the brain and have investigated those models both numerically and theoretically. In particular, a mean-field theory (MFT) of randomly connected neuronal networks (RNNs) has provided a solid theoretical foundation that allows us to investigate neuronal dynamics using analytical methods similar to those employed for spin-glass systems [19]. To enhance its applicability, different versions of the theory have been developed for different models, ranging from simple networks of neurons described by one-dimensional firing-rate variables to structured networks of neurons described by binary spike variables or more realistic kinetic variables of biological membranes [20–34]. These studies have theoretically shown that their RNNs have dynamical phases with different characteristics, such as chaotic fluctuations in

\*hayakawa.takashi@nihon-u.ac.jp

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

firing-rates, asynchronous irregular firing, and regular and irregular synchronized firing [19,25,27,35]. Efficient computation that takes advantage of the dynamical properties of RNNs has also been investigated recently, in both biological and engineering contexts [36–44].

A primary constraint in modeling neuronal dynamics of cortical areas is the fact that neurons in a local circuit densely form synapses on one another, and that these synapses obey “Dale’s law,” a principle that prohibits neurons from forming both excitatory and inhibitory synapses. Despite the knowledge about RNNs and MFTs accumulated over decades, only recently have researchers successfully begun to develop MFTs for networks under these two constraints. In RNNs comprising  $N (\gg 1)$  neurons with these two constraints,  $O(N)$  excitatory and inhibitory neurons are connected with a fixed  $O(1)$  probability. As is usually the case for physical systems with random couplings, nontrivial dynamics are observed for synaptic strengths with  $O(1/\sqrt{N})$  standard deviation. Then, Dale’s law requires us to determine the means of the excitatory and inhibitory synaptic strengths to be  $\pm O(1/\sqrt{N})$ , respectively. As a result, neurons receive very strong excitatory and inhibitory recurrent inputs from other neurons. By extending a previous theory [24], recent work showed that as observed experimentally [45–50], feedback inhibition balances strong excitatory recurrent inputs to neurons and stably produces an asynchronous and irregular firing state in such a model [27]. Although this model did not have nontrivial population dynamics, two very recent studies [51,52] investigated spatially extended versions of such balanced RNNs in the presence of external inputs, reporting multiscale dynamics in which macroscopic spatiotemporal patterns and microscopic irregular firing of individual neurons coexisted.

Although these studies have successfully demonstrated the relationship between spatial structures and multiscale dynamics of balanced neuronal networks, there remains a fundamental issue that stems from a limitation inherent in these models. In the previous models, a change in the population activity caused by a small number of constituent neurons is quickly counter-balanced by the feedback-inhibition mechanism, resulting in only vanishingly small responses in the population dynamics [see further discussion in Sec. V C]. Presumably, this is the reason why the previous studies required a network-wide application of external inputs—that is, an extrinsic origin—to induce population dynamics. The vanishingly small responses of their models, however, contrast with recent experimental results that a moderate stimulation of a single excitatory neuron effectively evokes a population response of nearby neurons in the local circuit, suggesting an intrinsic origin of population dynamics [53,54]. This discrepancy may imply a fundamental difference between the nature of the dynamics of the previous models and those of the neuronal circuits investigated experimentally. Theoretically, the weak effects of the stimulation of neurons in the previous models originate from the fact that a set of population statistics of neuronal activities follows equations that are closed among themselves [52]. This implies that the time evolution of those population statistics are independent of the microscopic fluctuations in the activities of individual neurons; namely, that there exists a separation of scales. From a general point of view in statistical mechanics, finding such

a separation of scales is a common step in constructing an MFT. The description of intrinsically generated population dynamics, however, requires a theory without such a separation of scales. Although microscopic fluctuations are known to evoke very large responses in systems in critical states, the previous theories of critical phenomena are not of immediate use for this purpose, because the average of critical fluctuations over time and population are still vanishingly small in those theories [55,56]. Thus, regardless of the observation of critical responses in the brain both on the microscopic and macroscopic scales [57], it remains theoretically unclear how the critical responses of neurons are reflected in the population dynamics.

In this study, we present a solution to this fundamental issue by constructing a novel type of MFT for densely connected RNNs with Dale’s law, for which mean synaptic weights are set to critical values between those for excitation-dominant and inhibition-dominant states. In this theory, unlike the previous theories of critical dynamics, we show that fluctuations in individual neuronal activities are amplified by the strong excitation and inhibition to provide stochastic driving forces for the population dynamics. We also show that external inputs to a  $O(\sqrt{N})$  number of neurons effectively entrain the whole network, comprising  $N$  excitatory and  $N$  inhibitory neurons, through the same amplification mechanism. Then, we observe that the network dynamics undergo a transition from irregular dynamics to coherent dynamics as the magnitude of either the excitation and inhibition or the external inputs is increased. The transition to a coherent state is found to be strongly dependent on the configuration of the random connectivity. These phenomena are predicted by our MFT, which yields good quantitative agreement with direct numerical results. Numerical results further suggest that such coherent dynamics can be used for reading out information from the network in a state-dependent manner. Although, for the sake of mathematical clarity, our theory is derived for a network of simplified neurons described by firing-rate variables, we confirm numerically that similar multiscale dynamics arise in a network of leaky integrate-and-fire (LIF) neurons.

## II. MODEL

Our theory is formulated for a single pair of excitatory and inhibitory neuronal populations (denoted by index  $k = E$  and  $I$ , respectively), each of which consists of  $N$  neurons [Fig. 1(a)]. The  $i$ th neuron of population  $k$  is described by a single, real-valued dynamical variable  $h_i^{(k)}$  that obeys the following dynamical equation:

$$\frac{d}{dt} h_i^{(k)}(t) = -h_i^{(k)}(t) + \sum_{\ell=E,I, 1 \leq j \leq N} J_{k\ell}^{ij} \phi(h_j^{(\ell)}(t)) + I_i^{(k)}(t), \quad (1)$$

$$J_{k\ell}^{ij} = \frac{\sigma_0 \mathcal{J}_{k\ell}^{ij}}{\sqrt{N}} + \frac{g_{k\ell}}{\sqrt{N}}. \quad (2)$$

In these equations, the function  $\phi$  is the hyperbolic tangent function, which describes the sigmoidal response of the neurons. The quantities  $h_i^{(k)}$  and  $\phi(h_i^{(k)})$  represent the internal

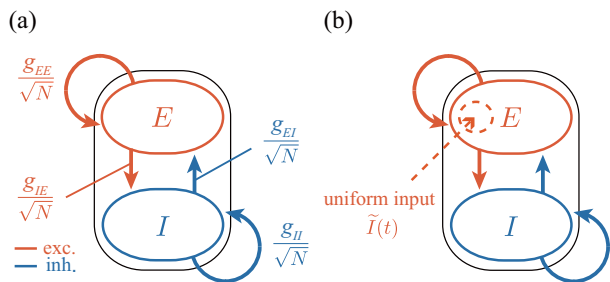


FIG. 1. A schematic illustration of the model. (a) The model network consists of a pair of excitatory and inhibitory populations (denoted by  $E$  and  $I$ , respectively), each of which contains  $N$  neurons. The neurons interact with one another through synaptic connections with (almost) *i.i.d.* quenched random weights. The mean strengths of the synaptic weights depend on the populations to which the presynaptic and postsynaptic neurons belong. They are parameterized as  $g_{k\ell}/\sqrt{N}$  ( $k, \ell = E, I$ ) as indicated in the panel. The variance of the synaptic weights is  $\sigma_0^2/N$ , with parameter  $\sigma_0$ . (b) External inputs are applied uniformly to only  $\sqrt{N}$  neurons in the  $E$  population [Eq. (4)].

state and firing rate of the neuron. The variables  $J_{k\ell}^{ij}$  and  $I_i^{(k)}(t)$  denote the strengths of the recurrent synapses on, and the external input to, the neuron. The synaptic strengths are independently and identically distributed (*i.i.d.*) quenched random variables, and their means and standard deviations are parameterized as  $g_{k\ell}/\sqrt{N}$  and  $\sigma_0/\sqrt{N}$ , respectively. Note that the means, but not the standard deviations, depend on the populations to which the presynaptic and postsynaptic neurons belong [Fig. 1(a)]. To describe this randomness, we have used the *i.i.d.* quenched random variables  $\mathcal{J}_{k\ell}^{ij}$  with zero mean and unit variance in Eq. (2). Unless otherwise stated, we consider the following case throughout this study:

$$g_{EE} = g_{IE} = -g_{EI} = -g_{II} = g_0, \quad g_0 \geq 0. \quad (3)$$

Equation (3) tunes the model to the critical point at which neuronal fluctuations evoke large responses in the population dynamics. With suitable distributions for  $\mathcal{J}_{k\ell}^{ij}$  (see Appendix A), the model constrained by Eq. (3) describes a densely connected network of excitatory and inhibitory neurons obeying the Dale's law. In the case with  $g_0 = 0$  and  $I_i^{(k)}(t) \equiv 0$ , this model is equivalent to the classical model investigated by a previous study [19]. For the case with  $I_i^{(k)}(t) \neq 0$ , we apply external inputs of the same strength,  $\tilde{I}(t)$ , to only  $\sqrt{N}$  neurons in the excitatory population [Fig. 1(b)] as

$$I_i^{(k)}(t) = \begin{cases} \tilde{I}(t), & (k, i) \in \mathcal{S}, \\ 0, & (k, i) \notin \mathcal{S}, \end{cases} \quad (4)$$

where we define  $\mathcal{S} \stackrel{\text{def}}{=} \{(k, i) | k = E, 1 \leq i \leq \sqrt{N}\}$ . Neurons in cortical areas have been thought to receive such sparse inputs [58,59]. Previous experiments showed that inputs to a small number of cortical neurons can drive the whole local circuit [53,54]. We model these neuronal responses.

In addition to the above model, we also study a model obeying the same dynamical equation as Eq. (1) with the

following additional tuning of the synaptic weights:

$$J_{k\ell}^{ij} = \frac{\sigma_0 \tilde{\mathcal{J}}_{k\ell}^{ij}}{\sqrt{N}} + \frac{g_{k\ell}}{\sqrt{N}}, \quad \tilde{\mathcal{J}}_{k\ell}^{ij} = \mathcal{J}_{k\ell}^{ij} - \frac{1}{N} \sum_{1 \leq j' \leq N} \mathcal{J}_{k\ell}^{ij'}. \quad (5)$$

Note that the sum of the random variations in the weights of the excitatory and inhibitory synapses on each neuron is finely tuned to zero; namely,  $\sum_{1 \leq j \leq N} \tilde{\mathcal{J}}_{k\ell}^{ij} = 0$ . The weight matrix  $J_{k\ell}^{ij}$  is known to have a finite number of random outlier eigenvalues for the untuned synaptic weights obeying Eq. (2), even in the limit of infinitely large  $N$ , but not for the finely tuned synaptic weights obeying Eq. (5) [60,61]. We observe how this qualitative difference is reflected in the dynamics of the model.

### III. MEAN-FIELD THEORY

#### A. Model with finely tuned synaptic weights

In this section, we formulate an MFT for the models described above. The detailed derivation is given in Appendix E. The MFT is slightly simpler for the finely tuned model obeying Eqs. (1), (3), and (5) with  $I_{k,i}(t) \equiv 0$  than for the untuned model. Therefore we first consider the dynamics of the finely tuned model.

Following a similar analysis to the previous one [32], we divide the dynamics of the model into macroscopic and microscopic parts. Let  $m_k(t)$  be certain averages of neuronal variables  $h_i^{(k)}(t)$  for  $k = E, I$ . The microscopic deviations from these averages are defined as

$$\delta h_i^{(k)}(t) \stackrel{\text{def}}{=} h_i^{(k)}(t) - m_k(t). \quad (6)$$

Similarly, we decompose the outputs of the neurons into macroscopic and microscopic parts as

$$\delta \phi_i^{(k)}(t) \stackrel{\text{def}}{=} \phi(h_i^{(k)}(t)) - \bar{\phi}_k(t). \quad (7)$$

The precise definitions of  $\bar{m}_k$  and  $\bar{\phi}_k$  will be stated below. Here, we note that  $m_k$  and  $\bar{\phi}_k$  coincide with the averages of  $h_i^{(k)}$  and  $\phi(h_i^{(k)})$ , respectively, over the population  $k$  up to the leading order in  $N$ .

With these decompositions, we can rearrange the model equations into the following form in the large  $N$  limit:

$$\frac{d}{dt} m_k(t) = -m_k(t) + \sum_{\ell, j} \frac{g_{k\ell}}{\sqrt{N}} \phi(h_j^{(\ell)}(t)), \quad (8)$$

$$\frac{d}{dt} \delta h_i^{(k)}(t) = -\delta h_i^{(k)}(t) + \sigma_0 \sum_{\ell, j} \frac{\tilde{\mathcal{J}}_{k\ell}^{ij}}{\sqrt{N}} \delta \phi_j^{(\ell)}(t). \quad (9)$$

The first of the above equations with a suitable initial condition defines  $m_k$ .

The configuration of the random synaptic weights does not change during time evolution. In the framework of MFT, however, we consider the distributions of the microscopic variables,  $\delta h_i^{(k)}$  and  $\delta \phi_j^{(\ell)}$ , over a large ensemble of networks with different weight configurations. Therefore the time evolution of these variables is stochastic. In the stochastic dynamics, the

driving-force term in the right-hand side of Eq. (9),

$$\eta_i^{(k)}(t) \stackrel{\text{def}}{=} \sigma_0 \sum_{\ell,j} \frac{\tilde{J}_{k\ell}^{ij}}{\sqrt{N}} \delta\phi_j^{(\ell)}(t), \quad (10)$$

has the following property: given the entire time evolution of the mean activity  $m_E$  and  $m_I$ , the conditional probability distribution for  $(\eta_i^{(k)}(t_1), \eta_i^{(k)}(t_2), \dots)$ , for any finite set of time points,  $t_1, t_2, \dots$ , is a zero-mean Gaussian that is *i.i.d.* with respect to the index  $i$ . This is intuitively justified by the central limit theorem applied to Eq. (10), under the assumption that the random synaptic weights and fluctuations in the neuronal outputs are almost independent (see Appendix E for further justification). Furthermore, since the driving-force term in the linear equation (9) has a conditionally *i.i.d.* Gaussian distribution with zero mean, so also do fluctuations in  $\{\delta h_i^{(k)}\}_{i,k}$ . Then, the following first- and second-order moments fully characterize these Gaussian fluctuations:

$$\begin{aligned} C_k(t, s) &\stackrel{\text{def}}{=} \langle \delta\phi_i^{(k)}(t) \delta\phi_i^{(k)}(s) \rangle, & \bar{\phi}_k(t) &\stackrel{\text{def}}{=} \langle \phi(h_i^{(k)}(t)) \rangle, \\ D_k(t, s) &\stackrel{\text{def}}{=} \langle \delta h_i^{(k)}(t) \delta h_i^{(k)}(s) \rangle, & m_k(t) &= \langle h_i^{(k)}(t) \rangle, \end{aligned} \quad (11)$$

where the brackets denote averages over the Gaussian fluctuations. Note that  $\bar{\phi}_k$  is defined in the above. The correlation function for  $\eta_i^{(k)}$  is obtained from  $C_\ell(t, s)$ :

$$\langle \eta_i^{(k)}(t) \eta_i^{(k)}(s) \rangle = \sigma_0^2 \sum_{\ell=E,I} C_\ell(t, s). \quad (12)$$

Below, we omit the population index  $k$  because the excitatory and inhibitory populations have the same statistical properties in our model setting. Thus we have

$$\begin{aligned} \bar{\phi} &\stackrel{\text{def}}{=} \bar{\phi}_E = \bar{\phi}_I, & m &\stackrel{\text{def}}{=} m_E = m_I, \\ C &\stackrel{\text{def}}{=} C_E = C_I, & D &\stackrel{\text{def}}{=} D_E = D_I. \end{aligned} \quad (13)$$

The statistics of the fluctuations defined above must satisfy certain consistency conditions. First,  $m$  and  $\bar{\phi}$  have been defined from dynamical variables that evolve under the influences of  $m$  and  $\bar{\phi}$  themselves [Eqs. (8), (9), and (11)]. Therefore their values need to be determined in a self-consistent manner. Second, the two Gaussian fluctuations characterized by  $(m, D)$  and  $(\bar{\phi}, C)$  are related to each other because they originate from the same dynamical variables. Consistency among these statistics gives rise to self-consistent equations that determine the time evolution of  $\bar{\phi}$ ,  $C$ , and  $D$  for given values of  $m$  and boundary conditions. Firstly,  $\bar{\phi}$  and  $C$  are represented as nonlinear functions of  $m$  and  $D$  [see Eqs. (F5) and (F6) in Appendix F for the details]:

$$\begin{aligned} \bar{\phi}(t) &= G_1(m(t), D(t, t)), \\ C(t, s) &= G_2(m(t), m(s), D(t, t), D(s, s), D(t, s)). \end{aligned} \quad (14)$$

Secondly, the relation between  $\eta_i^{(k)}$  and  $\delta h_i^{(k)}$  results in the following dynamical equation:

$$(1 + \partial_t)(1 + \partial_s)D(t, s) = 2\sigma_0^2 C(t, s). \quad (15)$$

An important difference between our MFT and conventional MFTs lies in the macroscopic driving force in Eq. (8). The right-hand side of this equation involves the summation of fluctuations in the outputs of individual neurons which can be

considered as independent random quantities with correlation  $C(t, s)$ . Then, the sum of these quantities in Eq. (8),

$$\begin{aligned} \eta(t) &\stackrel{\text{def}}{=} \frac{1}{\sqrt{2}g_0} \sum_{\ell,j} \frac{g_{k\ell}}{\sqrt{N}} \phi(h_j^{(\ell)}(t)) \\ &= \frac{1}{\sqrt{2N}} \left\{ \sum_i \delta\phi_i^{(E)}(t) - \delta\phi_i^{(I)}(t) \right\}, \end{aligned} \quad (16)$$

is also a random quantity (which is equal for  $k = E, I$ ). From the first line to the second line, we have used  $\bar{\phi}_E(t) = \bar{\phi}_I(t)$ . This stochasticity contrasts starkly with the deterministic macroscopic dynamics in conventional MFTs. The central limit theorem implies that  $\eta$  obeys the following probability distribution:

$$p(\eta) \propto \exp\left(-\frac{1}{2}\eta^T C^{-1}\eta - F\right), \quad (17)$$

where the normalization term  $F$  is given by

$$F = \ln \det \left[ \left( 1 - \frac{\partial C^{1/2}}{\partial \eta} C^{-1/2} \eta \right)^{-1} C^{1/2} \right]. \quad (18)$$

Here, we have regarded  $\eta$  and  $C$  as a vector and a matrix, respectively, that consist of their values for infinitesimally discretized time steps. With these stochastic dynamics for  $\eta$ , the macroscopic dynamics in Eq. (8) reads

$$\frac{d}{dt}m(t) = -m(t) + \sqrt{2}g_0\eta(t). \quad (19)$$

We note that, for a given history of  $\eta$  and  $m$  up to time  $t$ , the conditional distribution of  $\eta(t + \Delta t)$  determined by Eq. (17) with small  $\Delta t$  is approximately Gaussian. In this case, in the conditional distribution up to time  $t + \Delta t$ ,

$$p(\eta(t + \Delta t) | \{\eta(s)\}_{s \leq t}) \propto \exp\left(-\frac{1}{2}\eta^T C^{-1}\eta - F\right), \quad (20)$$

the correlation matrix  $C$  and normalization term  $F$  are independent of  $\eta(t + \Delta t)$  up to the leading order in  $\Delta t$ . Therefore the deviation of the conditional probability distribution from a Gaussian distribution is negligibly small. This fact enables us to solve the stochastic dynamics numerically for  $\eta$ ,  $m$ ,  $\bar{\phi}$ ,  $C$ , and  $D$  by iteratively updating their values with the Euler method (see Appendix F for the details).

In sum, the microscopic fluctuations in the neuronal activities obey a Gaussian distribution that depends on the mean activity  $m$  [Eqs. (14) and (15)], and the probability of realizing the mean activity  $m$  depends on the correlation matrix  $C$  of the microscopic fluctuations [Eqs. (17) and (19)]. Due to this strong link between the microscopic and macroscopic dynamics, the entire dynamics are, in general, non-Gaussian, even though the distribution of  $\eta$  resembles a Gaussian distribution [Eq. (17)]. This means that—unlike in conventional MFTs—a solution for the model cannot be completely determined by the first- and second-order moments.

## B. Balance equations

In Eq. (16), we removed the population-averaged part by using  $\bar{\phi}_E(t) = \bar{\phi}_I(t)$ . Before using this relation, the macroscopic part of the mean-field equations reads, to leading



order,

$$\begin{aligned} \frac{d}{dt}m_E(t) &\approx -m_E(t) + \sqrt{N}(g_{EE}\bar{\phi}_E(t) + g_{EI}\bar{\phi}_I(t)), \\ \frac{d}{dt}m_I(t) &\approx -m_I(t) + \sqrt{N}(g_{IE}\bar{\phi}_E(t) + g_{II}\bar{\phi}_I(t)). \end{aligned} \quad (21)$$

In these equations, the driving-force terms on the right-hand side are  $O(\sqrt{N})$  and hence may diverge. Thus the following condition must hold for stable dynamics: except for  $O(1/\sqrt{N})$  residuals,

$$\begin{aligned} g_{EE}\bar{\phi}_E(t) + g_{EI}\bar{\phi}_I(t) &\approx 0, \\ g_{IE}\bar{\phi}_E(t) + g_{II}\bar{\phi}_I(t) &\approx 0. \end{aligned} \quad (22)$$

Equations (22) are called “*balance equations*.” In previous theories, the balance equations were often nondegenerate and determined unique values for  $\bar{\phi}_E(t)$  and  $\bar{\phi}_I(t)$ . This implies that population-averaged activities exhibit only vanishingly small fluctuations if the entire dynamics are stable. In contrast, in our theory—for the values of  $g_{kl}$  satisfying Eq. (3)—Eqs. (22) are degenerate, and they are satisfied as long as equality  $\bar{\phi}_E(t) = \bar{\phi}_I(t)$  holds. Furthermore, this equality is always ensured to hold as a consequence of the fact that the macroscopic equation, Eq. (8), has exactly the same driving-force term for the excitatory and inhibitory populations, and hence  $m_E(t) = m_I(t)$ . As a result, the average of the neuronal outputs  $\bar{\phi}_k(t)$  is allowed to fluctuate strongly.

### C. Model with untuned synaptic weights

Next, we describe how the above theory is modified for the untuned model [Eqs. (1) and (2)]. In this case, the dynamical equation is divided into microscopic and macroscopic parts in a slightly different manner:

$$\frac{d}{dt}m_k(t) = -m_k(t) + \sum_{\ell,j} \frac{g_{k\ell}}{\sqrt{N}} \phi(h_j^{(\ell)}(t)), \quad (23)$$

$$\frac{d}{dt}\delta h_i^{(k)}(t) = -\delta h_i^{(k)}(t) + \sigma_0 \sum_{\ell,j} \frac{\mathcal{J}_{k\ell}^{ij}}{\sqrt{N}} \phi(h_j^{(\ell)}(t)). \quad (24)$$

Note that, in Eq. (24),  $\delta\phi_j^{(\ell)}(t)$  in Eq. (9) is replaced by the uncentered quantity  $\phi(h_j^{(\ell)}(t))$ . As a result, the microscopic driving-force terms,

$$\tilde{\eta}_i^{(k)}(t) \stackrel{\text{def}}{=} \sigma_0 \sum_{\ell,j} \frac{\mathcal{J}_{k\ell}^{ij}}{\sqrt{N}} \phi(h_j^{(\ell)}(t)), \quad (25)$$

have the following correlation function:

$$\begin{aligned} \langle \tilde{\eta}_i^{(k)}(t) \tilde{\eta}_i^{(k)}(s) \rangle &= 2\sigma_0^2 \tilde{C}(t, s), \\ \tilde{C}(t, s) &\stackrel{\text{def}}{=} \langle \phi(h_i^{(k)}(t)) \phi(h_i^{(k)}(s)) \rangle \\ &= C(t, s) + \bar{\phi}(t)\bar{\phi}(s). \end{aligned} \quad (26)$$

These equations imply that individual neurons receive additional synchronous inputs with random amplitudes, which can be represented as

$$\tilde{\eta}_i^{(k)} = \eta_i^{(k)}(t) + \sqrt{2}\sigma_0\xi_i^{(k)}\bar{\phi}(t), \quad (27)$$

where  $\{\xi_i^{(k)}\}_{i,k}$  are *i.i.d.* quenched Gaussian variables with zero mean and unit variance. Then, the modified self-consistent equation,

$$(1 + \partial_t)(1 + \partial_s)D(t, s) = 2\sigma_0^2 \tilde{C}(t, s), \quad (28)$$

together with Eqs. (14) and (26) determines the time evolution of  $\bar{\phi}$ ,  $\tilde{C}$ , and  $D$  for a given orbit  $m$ . On the other hand, the macroscopic dynamics of  $m$  are described by Eqs. (17) and (19).

### D. Application of external inputs

In Sec. IV C, we apply external inputs of strength  $\tilde{I}(t)$  of  $\sqrt{N}$  neurons in the  $E$  population of the finely tuned model. The stimulus-driven dynamics are analyzed by an MFT that is slightly modified from the one introduced above. It is described by

$$\begin{aligned} \frac{d}{dt}m(t) &= -m(t) + \sqrt{2}g_0\eta(t) + g_0\tilde{\phi}(t), \\ \frac{d}{dt}\delta h_i^{(k)}(t) &= -\delta h_i^{(k)}(t) + \eta_i^{(k)}(t) + I_i^{(k)}(t), \\ \tilde{\phi}(t) &\stackrel{\text{def}}{=} \langle \phi(h_i^{(k)}(t)) \rangle_{(k,i) \in S} - \langle \phi(h_i^{(k)}(t)) \rangle_{(k,i) \notin S}, \end{aligned} \quad (29)$$

where the microscopic and macroscopic stochastic driving-force terms,  $\eta_i^{(k)}$  and  $\eta$ , are distributed according to the same equations as those for the autonomous case, namely, Eqs. (12) and (17), with the same self-consistent equations, Eqs. (14) and (15). The difference,  $\tilde{\phi}(t)$ , between the averages of the stimulus-driven and -undriven neuronal variables over the Gaussian fluctuations gives an additional driving force term for the mean activity in Eq. (29).

## IV. RESULTS

### A. Dynamics of the finely tuned model

#### 1. Fluctuations in the mean activity

We first examine the finely tuned model described by Eqs. (1) and (5) without external inputs, for which the MFT takes the simplest form (Sec. III A). For  $g_0 = 0$ , the MFT of this model is essentially equivalent to that studied previously [19]. The previous theory showed that the present model with  $g_0 = 0$  undergoes a transition from a trivial fixed point to a chaotic state at  $\sigma_0 = 1/\sqrt{2}$ , in which the mean activity,  $m$ , is constantly zero and individual neuronal activities,  $h_i^{(k)}$ , exhibit Gaussian fluctuations around it.

We are particularly interested in cases with nonzero values of  $g_0$ . We study these cases both by numerically solving the mean-field equations and by directly simulating the model for a large value of  $N$ . In these numerical simulations, we find only a trivial fixed-point solution for  $\sigma_0 \leq 1/\sqrt{2}$ . In contrast, for  $\sigma_0 > 1/\sqrt{2}$ , we obtain nontrivial solutions. Since the repertoire of solutions is qualitatively the same for different values of  $\sigma_0$ , we show typical activity patterns only for  $\sigma_0 = 1.2$  in Figs. 2(a)–2(d). In our MFT, the excitatory and inhibitory populations obey the same dynamical equations. Therefore we plot only a single representative solution from MFT for each value of  $g_0$ . In fact, in the plots from the direct simulations, the mean activities of the excitatory and

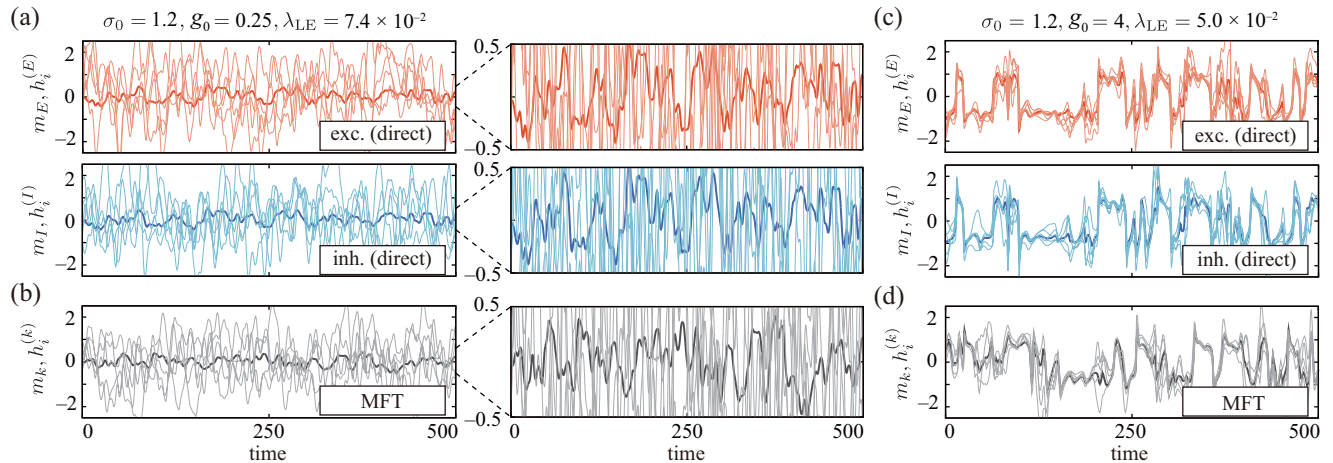


FIG. 2. Activity patterns of the finely tuned model for  $\sigma_0 = 1.2$  and for two different values  $g_0 = 0.25$  and  $4$ . In each plot, the thick and thin lines represent the mean activity  $m_k$  and the activities  $h_i^{(k)}$  of five representative neurons, respectively. For the direct simulations, the mean activity  $m_k$  is approximated by the average of  $h_i^{(k)}$  over the population  $k$  comprising  $N = 10240$  neurons (see Appendix A for the details of the simulations). Solutions for the  $E$  and  $I$  populations from direct simulations (indicated by “direct”) and solutions of the mean-field equations (indicated by “MFT”) are depicted in red, blue, and gray, respectively. The numerically determined value of the largest Lyapunov exponent,  $\lambda_{LE}$ , is shown above each panel.

inhibitory populations are almost equal, and the individual neuronal activities in the two populations exhibit similar temporal patterns. Comparing the plots from the MFT and direct simulations, we observe similar amplitudes and temporal patterns for the mean activities and the microscopic fluctuations around them. These results suggest that our theory successfully predicts the behavior of the model. Below, we further evaluate this point quantitatively.

As the value of  $g_0$  is increased from zero, the mean activity of the model starts to fluctuate with nonzero amplitudes. For relatively small values of  $g_0$ , the temporal profiles of the fluctuations, both in the mean and in the individual neuronal activities, are similar to the Gaussian fluctuations in individual neuronal activities at  $g_0 = 0$  [Figs. 2(a) and 2(b)]. This is expected from the MFT, which shows that the driving force for the mean activity is the summation of individual neuronal fluctuations scaled by  $g_0/\sqrt{N}$  [Eq. (8)]. With a further increase in the value of  $g_0$ , the model starts to show irregular, intermittent dynamics, varying between positive and negative values close to  $\pm 1$ , with patterns that are reminiscent of the UP-DOWN states observed in the brain [6,15] [Figs. 2(c) and 2(d)]. This bimodality in the mean activity indicates the non-Gaussianity of the dynamics and contrasts with the dynamics for small values of  $g_0$ . Numerically determined largest Lyapunov exponent [Fig. 2] indicates that the both types of solutions described above are chaotic.

Increasing the value of  $g_0$  still further, we occasionally observe stable fixed-points and regularly oscillating solutions, as well as irregular, chaotic solutions. Although these nonchaotic solutions are observed for networks with a fairly large number of neurons (Appendix C), further theoretical analyses suggest that these solutions are due to finite-size effects and not stable in the thermodynamic limit (see the discussion in Sec. IV B 3 and Appendix H).

To examine the extent to which the description provided by our MFT is accurate, we calculate statistics of the dynamics from numerical solutions of the mean-field equations and of

the original model equations. We calculate the autocorrelation functions for the mean activity and for the individual neurons as

$$\mu(\tau) \stackrel{\text{def}}{=} \langle m(t)m(t-\tau) \rangle, \quad (30)$$

$$D(\tau) \stackrel{\text{def}}{=} \langle \delta h_i^{(k)}(t)\delta h_i^{(k)}(t-\tau) \rangle. \quad (31)$$

Since we expect the dynamics to be non-Gaussian, we also calculate the fourth-order statistics defined by

$$\kappa(\tau) \stackrel{\text{def}}{=} \langle m(t)^2 m(t-\tau)^2 \rangle - \langle m(t)^2 \rangle^2 - 2\langle m(t)m(t-\tau) \rangle^2, \quad (32)$$

$$q(\tau) \stackrel{\text{def}}{=} \langle \delta h_i^{(k)}(t)^2 \delta h_i^{(k)}(t-\tau)^2 \rangle - \langle \delta h_i^{(k)}(t)^2 \rangle^2 - 2\langle \delta h_i^{(k)}(t)\delta h_i^{(k)}(t-\tau) \rangle^2. \quad (33)$$

In these equations, bracketing indicates averaging over both time and configurations of the random connectivity. In Eqs. (31) and (33), we also take averages over population  $k$  in direct simulations and averages over microscopic Gaussian fluctuations in the corresponding MFT. The fourth-order statistics defined above vanish if the dynamics are Gaussian. For the direct simulations, we show only the statistics of the  $E$  population, because those of the  $I$  population are essentially the same.

The panels in Fig. 3 compare these calculated statistics, and they show good agreement between the theory and direct simulations. This indicates that our theory predicts the behavior of the model quantitatively, at least statistically. In this figure, we also observe large fourth-order statistics for networks with large values of  $g_0$ , which implies the highly non-Gaussian nature of the dynamics.

For the autocorrelation functions defined above, one would expect a perturbative expansion to provide a good analytical approximation, as it does for many physical systems. We can actually formulate such a method by expanding the dynamics

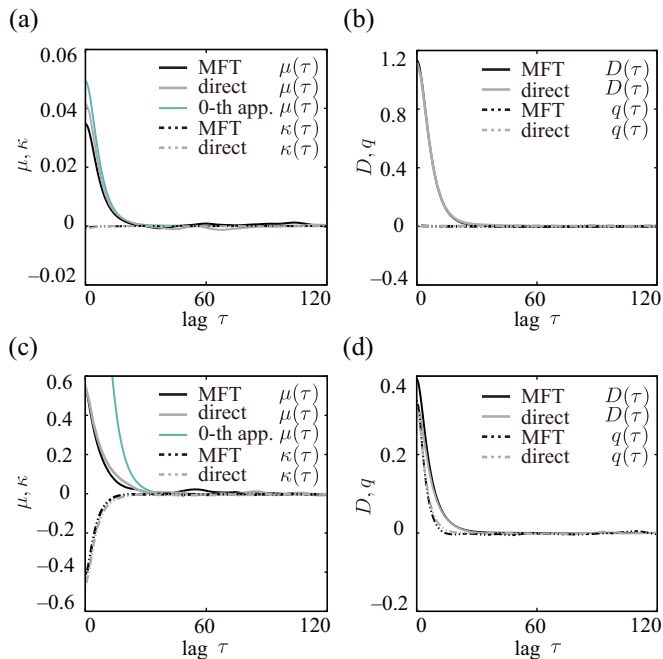


FIG. 3. Statistics of the network dynamics with the following values of the parameters: [(a) and (b)]  $(\sigma_0, g_0) = (1.2, 0.25)$  and [(c) and (d)]  $(\sigma_0, g_0) = (1.2, 4)$ . (a) and (c) show the values of the autocorrelation functions and the fourth-order statistics of the mean activities,  $\mu(\tau)$  and  $\kappa(\tau)$ . (b) and (d) show the values of those statistics of the neuronal activities,  $D(\tau)$  and  $q(\tau)$ . These values are determined either from the direct simulations (labeled “direct”), the simulations of the mean-field equations (labeled “MFT”), or the zeroth-order perturbative approximation (labeled “0-th app.”), and they are indicated by different types of lines. In the simulations, these quantities are averaged over the time period  $1000 \leq t \leq 2200$  and over 15 random configurations of directly simulated networks with population size,  $N = 40\,960$  [(a) and (b)] and  $20\,480$  [(c) and (d)]; or averaged over the same time period and 15 sequences of random numbers used for the simulations of the mean-field equations.

around  $g_0 = 0$ . However, it is numerically intractable to carry out the calculation of even the first-order expansion (see Appendix G). Thus we restrict ourselves to showing only the zeroth-order term,  $\mu(\tau) \approx g_0^2 D_0(\tau) / \sigma_0^2$ , of this perturbative expansion [Figs. 3(a) and 3(c)]. Here,  $D_0(\tau)$  is the autocorrelation function of the microscopic variables  $\delta h_i^{(k)}$  for  $g_0 = 0$ . This zeroth-order approximation gives vanishing fourth-order statistics:  $\kappa(\tau) = q(\tau) = 0$ . For small values of  $g_0$ , this solution shows relatively good agreement with the estimate obtained from direct simulations, while it does not do so for large values of  $g_0$ .

## 2. Waveforms of the mean activity and the signature of time-reversal symmetry breaking

For larger values of  $g_0$ , the analytical approach encounters another difficulty in addition to the computational problems mentioned above. Figures 2(c) and 2(d) show that the trajectories of the mean activity are observed with frequencies that are obviously asymmetric with respect to the time reversal of the trajectories. Note that the mean activity overshoots immediately after it makes an intermittent transition

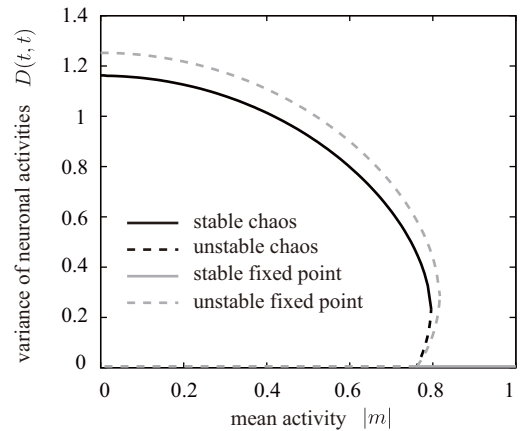


FIG. 4. Variances of neuronal activities in fixed-point and chaotic solutions and their stability, depicted for different absolute values of the constant mean activity  $|m|$ . Connected pairs of branches of nontrivial fixed-point and chaotic solutions, as well as trivial fixed-point solutions, are shown.

between positive and negative values, and that the temporal order of the transition and the overshooting is never reversed. Analytical approaches—such as a perturbative expansion around  $g_0 = 0$ —however, yield only symmetric solutions (see Appendix G). This inconsistency suggests the possibility of symmetry breaking with respect to time reversal. If a symmetry is broken, one cannot expect a symmetry-breaking solution to be obtained from a series expansion around the symmetric solution. In the following, we use a heuristic approach to seek clues to the occurrence and mechanism of such symmetry breaking and to an understanding of the waveform of the mean activities for large  $g_0$ .

In our MFT, the correlation function of the microscopic fluctuations is determined by Eqs. (14) and (15) for a given trajectory of the mean activity, which in turn determines the realization probability of the mean activity. Since this dependence is complicated, we first focus on the case with constant mean activities of different values, expecting the results to provide some clue to the dynamics with time-varying mean activities. Applying the previous theory [19,32] to this analysis, we find multiple fixed-point solutions and chaotic solutions. Figure 4 shows the variance of the neuronal activities of these solutions for different constant values of the mean activity. A branch of chaotic solutions (the black solid line in Fig. 4) coincides with the solution examined in a previous study [19] for  $m(t) \equiv 0$ . As the absolute value of the mean activity increases, the neuronal fluctuations in these solutions decrease. Another branch of chaotic solutions with smaller neuronal fluctuations (the black dotted line in Fig. 4) emerges at the value satisfying  $2\sigma_0^2 \phi'(m) = 1$ ,  $|m| \approx 0.76$ . The neuronal fluctuations in these solutions increase as the mean activity increases, and this branch eventually connects to that with larger neuronal fluctuations. From the numerical simulations, we find that the branch of chaotic solutions with larger fluctuations is stable, while that with smaller fluctuations is unstable. Fixed-point solutions and their stability can also be examined by applying the previous theory [32] (or the method presented in Appendix H), and we find two connected

branches of unstable fixed-point solutions as well as trivial fixed-point solutions [Fig. 4]. The trivial fixed-point solution is stable for  $|m|$  larger than the bifurcation point given by  $2\sigma_0^2\phi'(|m|) = 1$ , while it is unstable below this point.

Figure 4 suggests the following explanation for the waveform of the mean activity and its time-reversal asymmetry observed in Figs. 2(c) and 2(d). Let us assume that, for a time-varying mean activity, the instantaneous behavior of the neuronal fluctuations is the same as the above solution for the corresponding value of the constant mean activity. When the mean activity remains small for some time, the neuronal fluctuations increase. Since the neuronal fluctuations serve as a driving force for the mean activity, the mean activity is stochastically pushed to larger values. For larger values of the mean activity, the neuronal fluctuations decrease (along the black solid line in Fig. 4), while still remaining chaotic. When the mean activity reaches a value for which there are no stable chaotic solutions, the neuronal fluctuations start to decay to the trivial fixed point. Then, the mean activity loses its driving force and decays to smaller values. In this descending part of the mean activity, the network state passes through the region with the unstable chaotic and fixed-point solutions (the lower branches of the nontrivial solutions in Fig. 4). The profile of the neuronal fluctuations in this descending part is therefore different from that of the ascending part. Because of this passage through the region with unstable fixed points, both the neuronal fluctuations and the mean activity slow down, as we observe in Figs. 2(c) and 2(d). We suggest that this hysteresis in the multiscale dynamics is the mechanism for the observed waveform of the mean activity and its time-reversal asymmetry.

### 3. Ferromagnetic effects and critical fluctuations

Thus far, we have examined balanced networks with parameter values satisfying the condition in Eq. (3). In this section, we briefly mention what happens if this condition is not satisfied. As in a previous study [32], if the balance equation is not degenerate, the mean activities of the neuronal populations take a set of constant values uniquely determined by the balance equations, or else diverge. The remaining cases are described with two parameters  $\alpha$  and  $\beta$  as

$$\begin{pmatrix} g_{EE} & g_{EI} \\ g_{IE} & g_{II} \end{pmatrix} = g_0 \begin{pmatrix} 1 + \alpha & -1 + \alpha \\ \beta(1 + \alpha) & \beta(-1 + \alpha) \end{pmatrix}. \quad (34)$$

By definition, the parameter  $\alpha$  is interpreted as the magnitude of ferromagnetic interaction, while  $\beta$  is interpreted as the relative gain of the synaptic input to inhibitory neurons. The case we have examined in the previous sections corresponds to  $(\alpha, \beta) = (0, 1)$ . We examine the fluctuations in the mean activity by calculating their mean and variance averaged over a long period of simulations and by observing how they change as the value of  $\alpha$  or  $\beta$  deviates from  $(\alpha, \beta) = (0, 1)$  [Figs. 5(a) and 5(b)]. We find that the mean activity diverges as  $\alpha$  increases or  $\beta$  decreases, while the variance of the fluctuations decays to zero as  $\alpha$  decreases or  $\beta$  increases. As shown in Figs. 5(a) and 5(b), the rate of this divergence and decay is proportional to  $\sqrt{N}$ , which indicates that in the  $N \rightarrow \infty$  limit, the macroscopic dynamics of the network are divergent or trivial for  $\alpha \neq 0, \beta = 1$  and  $\alpha = 0, \beta \neq 1$ .

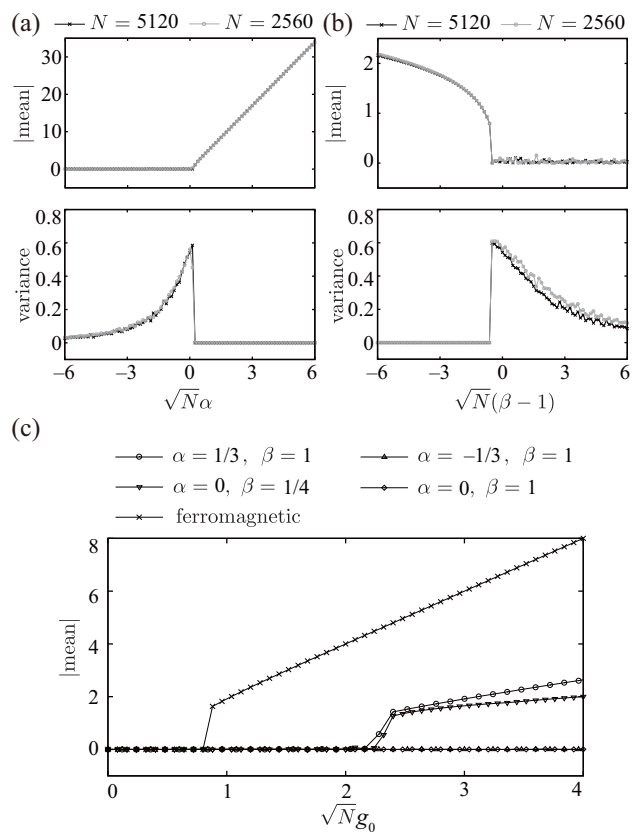


FIG. 5. Absolute values of the mean and variance of fluctuations in the mean activity of the excitatory population, calculated by taking long-time averages in numerical simulations with  $(\sigma_0, g_0) = (1.2, 4)$  and different values of (a)  $\alpha$  and (b)  $\beta$ , respectively. We performed simulations with the two different system sizes indicated in the panels. (c) The absolute value of the long-time average of the mean activity of the excitatory population, calculated by taking long-time averages in numerical simulations with  $\sigma_0 = 1.2$ , different values of  $g_0$ , and the following values of  $\alpha$  and  $\beta$ :  $(\alpha, \beta) = (\pm 1/3, 0)$ ,  $(0, 1/4)$ ,  $(0, 0)$ . Results for  $g_{EE} = g_{IE} = g_{EI} = g_{II} = 3g_0/2$  (indicated as “ferromagnetic”) are also shown.

This behavior can be understood by first examining  $O(1/\sqrt{N})$  values of  $g_0$  and then taking the  $g_0 \rightarrow \infty$  limit. As shown in a previous study [32], for  $g_{kl} \propto 1/\sqrt{N}$ , the dynamics of the mean activity are no longer subject to the balance between strong excitation and inhibition but instead are described by a simpler MFT:

$$\frac{d}{dt}m_k(t) = -m_k(t) + \sum_{\ell=E,I} \sqrt{N}g_{k\ell}\check{\phi}_\ell(t), \quad (35)$$

for  $k = E$  and  $I$ , where  $\check{\phi}_\ell(t)$  is the population average of  $\phi(h_j^{(\ell)}(t))$ . Figure 5(c) shows how the mean activity changes as  $g_0$  increases for fixed values of  $\alpha$  and  $\beta$ . As  $g_0$  increases for  $\alpha > 0, \beta = 1$  or  $\alpha = 0, \beta < 1$ , the second-term on the right-hand side of Eq. (35) starts to dominate, causing the trivial solution to bifurcate in a similar manner to a ferromagnetic transition [the result with  $(\alpha, \beta) = (1/3, 1)$ ,  $(0, 1/4)$  and the result for a purely ferromagnetic interaction shown in Fig. 5(c)]. For  $\alpha \leq 0, \beta = 1$  or  $\alpha = 0, \beta \geq 1$ , we find that the mean activity remains at zero. For this case of strict



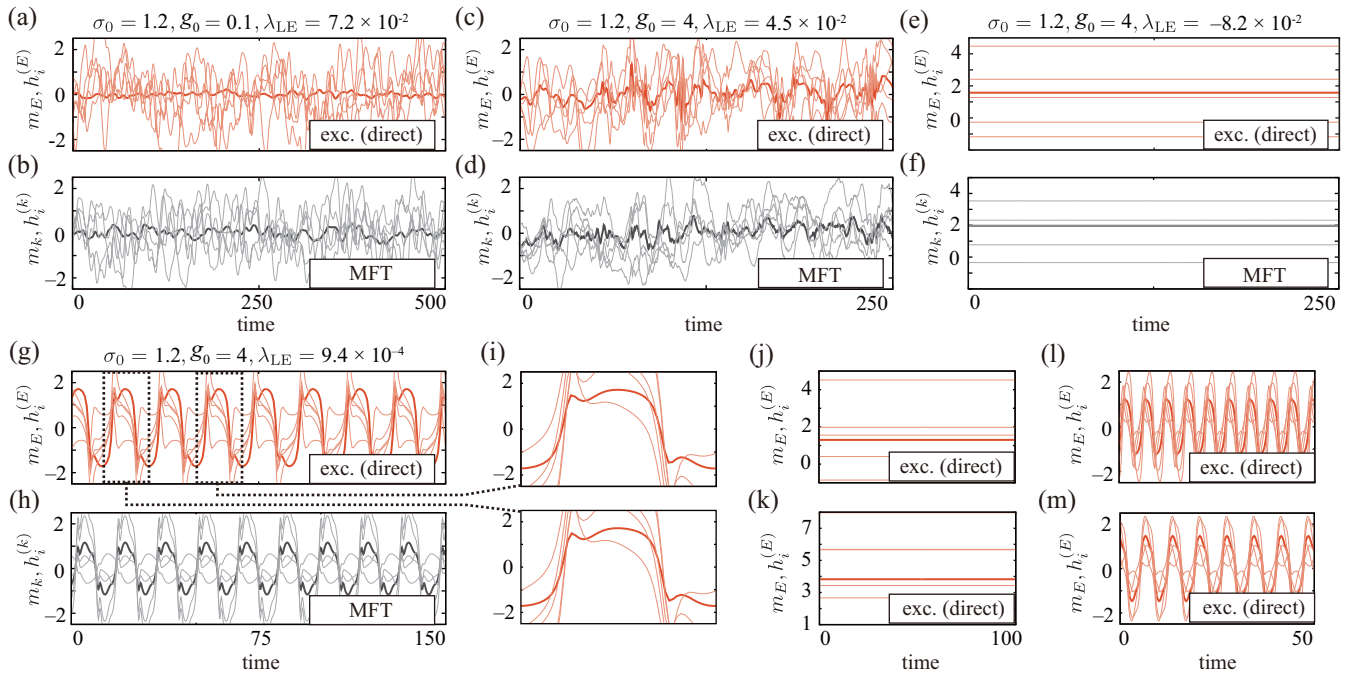


FIG. 6. Activity patterns of the untuned model. [(a)–(h)] Typical activity patterns for networks with the indicated parameter values are shown. In each plot, the thick and thin lines represent the mean activity  $m_k$  and the activities  $h_i^{(k)}$  of five representative neurons, respectively. For the direct simulations, the mean activity  $m_k$  is approximated by the average of  $h_i^{(k)}$  over the population  $k$  comprising  $N = 10\,240$  neurons (see Appendix A for the details of the simulations). Solutions for the  $E$  populations from direct simulations (indicated by “direct”) and solutions of the mean-field equations (indicated by “MFT”) are depicted in red and gray, respectively. The plots for the  $I$  population are omitted. For  $(\sigma_0, g_0) = (1.2, 4)$ , we find three qualitatively different solutions: irregular solutions [(c) and (d)], fixed-point solutions [(e) and (f)], and regularly oscillating solutions [(g) and (h)]. (i) shows magnified images of the regular oscillations, where the coherence between the mean activity and the activities of individual neurons can be seen. [(j)–(m)] For the fixed-point and regularly oscillating solutions, additional plots of the activity patterns illustrate the diversity of their values, waveforms, and frequencies. Which of the diverse solutions is observed for a given set of parameter values depends on the configuration of the random connectivity of the directly simulated networks or on the sequence of the random numbers used for the simulations of the mean-field equations. The numerically determined value of the largest Lyapunov exponent,  $\lambda_{LE}$ , is shown above each panel.

inequality, the second term on the right-hand side of Eq. (35) supplies feedback suppression to changes in the mean activity in a similar manner to antiferromagnetic effects. For  $\alpha = 0$ ,  $\beta = 1$ , however, such a feedback mechanism does not work. These behaviors of the model with  $O(1/\sqrt{N})$  values of  $g_0$  account for the divergent or suppressed dynamics observed for  $O(1)$  values of  $g_0$  as the limit of the former. These results also suggest that the present model under the condition given by Eq. (3) is at the critical point between the two states governed by extremely strong ferromagnetic and antiferromagnetic interactions.

## B. Dynamics of the untuned model

### 1. Qualitatively different solutions

The behavior of the untuned model, described by Eqs. (1) and (2), is different from the results discussed above. We show plots of its activity patterns in Figs. 6(a)–6(m). Simulations based on our MFT yield solutions with profiles similar to those from the direct simulations in this case, too. For smaller values of  $g_0$ , the network exhibits nearly Gaussian dynamics [Figs. 6(a) and 6(b)]. For larger values of  $g_0$ , it exhibits not only irregular dynamics [Figs. 6(c) and 6(d)] but also constant

activities (fixed-point solutions) [Figs. 6(e) and 6(f)] and regularly oscillating dynamics (limit-cycle solutions) [Figs. 6(g) and 6(h)]. The values of the mean activities of the observed fixed-point solutions are widely distributed over positive and negative values [Figs. 6(j) and 6(k)]. Note that because of the symmetry of the model equations, fixed points are necessarily located symmetrically at two points with positive and negative mean activities of the same absolute value. The waveforms and frequencies of the observed regular oscillations are also diverse [Figs. 6(l) and 6(m)]. Which of these diverse solutions is observed for a given set of parameter values depends on the configuration of the random connectivity of the directly simulated networks or on the sequence of the random numbers used for the simulations of the mean-field equations. In the regularly oscillating solutions, we also observe that both the mean activity and the activities of individual neurons are coherent, which means that the activities of individual neurons have various waveforms but are all phase-locked to the same rhythm [Fig. 6(i)].

### 2. Strong dependence on the synaptic-weight configuration

Next, we examine in further detail the three qualitatively different solutions for larger values of  $g_0$ . Here, we

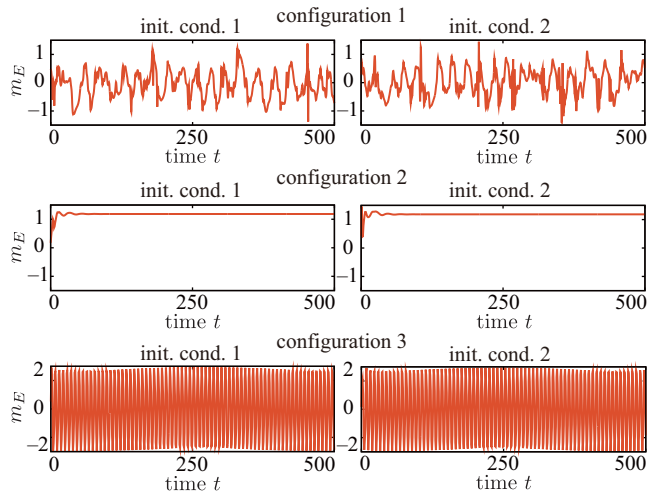


FIG. 7. Direct simulations of the untuned model with three different random weight configurations for the same parameter values,  $(\sigma_0, g_0) = (1.2, 4)$ , for two different initial conditions. The mean activities of the excitatory populations of the three networks are plotted. The three networks consistently showed irregular, static, and regularly oscillating dynamics, respectively, in simulations from different initial conditions.

emphasize that the type of the observed dynamics depends on the configuration of the random connectivity but not on the initial condition of the simulations. Figure 7 plots the activity patterns of three networks with the same parameter values but different configurations. We find that a network with the same configuration shows dynamics convergent to the same attractor when it is simulated from different initial conditions, while those with the same parameter values but different configurations show various dynamics. Such individuality among networks with different configurations was expected from the configuration-dependent outlier eigenvalues of the synaptic weight matrices [60–62], which we also confirm numerically (Appendix B). The outlier eigenvalues in the synaptic weight matrices indicate that the untuned model has a strong configuration dependence at the level of its dynamical equation.

Despite this obvious configuration dependence, our mean-field equations reproduce activities similar to those of the directly simulated networks. We therefore expect the MFT to give us further insights into the configuration-dependent dynamics, and we examine this point below.

### 3. Fixed-point solutions and their stability

We first examine the observed fixed-point solutions. Suppose that the activity of the entire network is constant, with mean activity  $m(t) \equiv m_f$ . Recall that the dynamics of the mean activity are described by Eq. (19), rewritten here as

$$\frac{d}{dt}m(t) = -m(t) + \sqrt{2}g_0\eta(t), \quad (36)$$

where the fluctuation term  $\eta(t)$  is generated according to Eq. (17). If the network state remains at a fixed point, the

correlation function of the microscopic fluctuations,  $\tilde{C}(t, s)$ , is constant, and therefore, neuronal activities take normally distributed values that do not change temporally. Then, the fluctuation term  $\eta(t)$  does not change temporally either, because it is the sum of the microscopic neuronal fluctuations [Eq. (17)]. From Eq. (36), we find that

$$m(t) \equiv \sqrt{2}g_0\eta(t) \equiv m_f, \quad (37)$$

must hold in order for the network state to remain at the fixed point without requiring an external input. Applying the MFT, we find that there is a continuous band of values for  $m_f$  for which the stable solution for  $\tilde{C}(t, s)$  is constant. From that analysis, we expect that a solution satisfying Eq. (37) exists with a nonzero probability (also see the discussion at the end of Appendix H). The existence of this band suggests the mechanism for the appearance of fixed-point solutions as follows. When the mean activity stays in this band,  $\tilde{C}(t, s)$  tends to be constant, and hence, the microscopic fluctuations slow down. As a result, the mean activity receiving driving forces from the microscopic fluctuations also slows down. This leads to the convergence of the entire dynamics to an equilibrium point.

This scenario is justified by a perturbative stability analysis. In this analysis, we examine the response of the system around a fixed-point solution to a temporary external perturbative input. Suppose that Eq. (37) holds and that the network state is set to a fixed-point solution with mean activity  $m(t) = m_f$  for a long time prior to  $t = 0$ . Then, suppose that temporary external inputs, collectively denoted by  $\mathbf{p}$ , are applied in  $t > 0$ . For  $t \leq 0$ , the self-consistent equation, Eq. (15), reads

$$D_0 = 2\sigma_0^2\tilde{C}_0, \quad \tilde{C}_0 = \int d\mathcal{N}(z)\phi(\sqrt{D_0}z + m_f)^2, \quad (38)$$

with the variance of  $\delta h_i^{(k)}$ , denoted by  $D_0$ , and with the mean square of  $\phi(h_i^{(k)})$ , denoted by  $\tilde{C}_0$ . Here,  $\mathcal{N}(z)$  denotes a unit Gaussian distribution. The condition for the stability of a static solution to Eq. (38) is given by

$$\begin{aligned} 1 - a_1 - 2a_2 &> 0, \quad a_1 < 1, \quad \text{with} \\ a_1 &= 2\sigma_0^2\langle\phi'(h_i^{(k)})^2\rangle_0, \\ a_2 &= \sigma_0^2\langle\phi''(h_i^{(k)})\phi(h_i^{(k)})\rangle_0, \end{aligned} \quad (39)$$

(see Appendix H for the derivation). Here, the angle brackets with subscript 0 denote averaging over the unperturbed dynamics with  $m(t) \equiv m_f$ .

For  $t \geq 0$ , we perturbatively expand the dynamics around the fixed-point solution. We calculate how a change in the mean activity,  $\delta m(t) = m(t) - m_f$ , evokes a response in the correlation  $D(t, s)$  and how the evoked response in the correlation generates additional fluctuations in  $\eta(t)$ . We refer readers interested in the details of this analysis to Appendix H. From this analysis, we find that up to the first order, a self-consistent equation of the following form—with *i.i.d.* unit-Gaussian coefficients  $\xi_{j\ell}$  and  $\xi'_{j\ell}$  determined by the configuration of the

random connectivity—must be satisfied:

$$\begin{aligned} \delta m(t) = & (1 + \partial_t)^{-1} p_0(t) + m_f d_1[\delta m](t) \\ & + \frac{g_0}{\sigma_0} \left\{ \sum_{1 \leq j \leq 3, 1 \leq \ell < \infty} \xi_{j\ell} d_{2,j\ell}[\delta m](t) \right. \\ & \left. + \sum_{j, 1 \leq \ell < \infty} \xi'_{j\ell} d_{3,j\ell}[\mathbf{p}](t) \right\} + O(|\mathbf{p}|^2). \quad (40) \end{aligned}$$

Here, the term  $p_0$  is the component of the perturbative input that is uniformly applied to all neurons, and the terms  $d_1[\delta m]$ ,  $d_{2,j\ell}[\delta m]$ , and  $d_{3,j\ell}[\mathbf{p}]$  are certain linear transformations of  $\delta m$  or  $\mathbf{p}$ , respectively. The operation denoted by  $(1 + \partial_t)^{-1}$  is defined as

$$(1 + \partial_t)^{-1} x(t) \stackrel{\text{def}}{=} \int_{-\infty}^t e^{-(t-\tau)} x(\tau) d\tau. \quad (41)$$

The solution of this self-consistent equation can be obtained explicitly. From this solution, we find that if we have

$$m_f \theta_1 < 1, \quad a_1 < 1, \quad (42)$$

for constant  $\theta_1$  calculated from the unperturbed dynamics (see Eq. (H26) for the details),  $\delta m(t) \rightarrow 0$  holds with a nonzero probability as  $t \rightarrow \infty$ , depending on the values of  $\xi_{j\ell}$ , and hence, on the random weight configuration. This convergence of the mean activity, together with the microscopic stability given by Eq. (39), implies the stability of the entire dynamics around the fixed point, and hence, justifies the scenario with the slowing down of both the mean activity and the microscopic fluctuations. From the same analysis, we also see that, depending on the values of  $\xi_{j\ell}$ , the obtained solution diverges at some  $t > 0$ .

In summary, we find that the model has fixed-point solutions for which the mean and variance of the neuronal activities are determined in a configuration-dependent manner [Eqs. (17), (37), and (38)]. Then, even for fixed-point solutions with the same statistics of neuronal activities, we find that their stability depends on the individual configurations. For the fixed-point solution for which Eqs. (39) and (42) are satisfied, configurations that yield stable fixed points exist with a nonzero probability. To check the stability of the numerically observed fixed points, we compute the values of  $m_f \theta_1$ ,  $D_0$ ,  $a_1$ , and  $1 - a_1 - 2a_2$  for different values of  $m_f$  [Fig. 8(a)]. We find that the conditions for the stability are actually satisfied for a certain range of  $m_f$  [orange arrows in Fig. 8(a)]. We find that all fixed points observed in numerical simulations, including those shown in Fig. 6, fall in this range of  $m_f$  [Fig. 8(b)]. This contrasts with fixed points that are occasionally observed for the finely tuned model with large  $g_0$ . Fixed points of that model never satisfy the corresponding stability condition. This implies that fixed points do not exist in the thermodynamic limit. In the above analysis, we find that the condition given by Eqs. (39) and (42) itself does not depend on the values of  $g_0$ , while the probability of realizing stable fixed points does depend on it. For small  $g_0$ , the values of  $\eta(t)$  that satisfy Eq. (37) with a value of  $m_f$  within the range of stability is very large. The realization probability of such a large value of  $\eta(t)$  is expected to be small from Eq. (17). This explains the reason that we do not observe fixed points for a

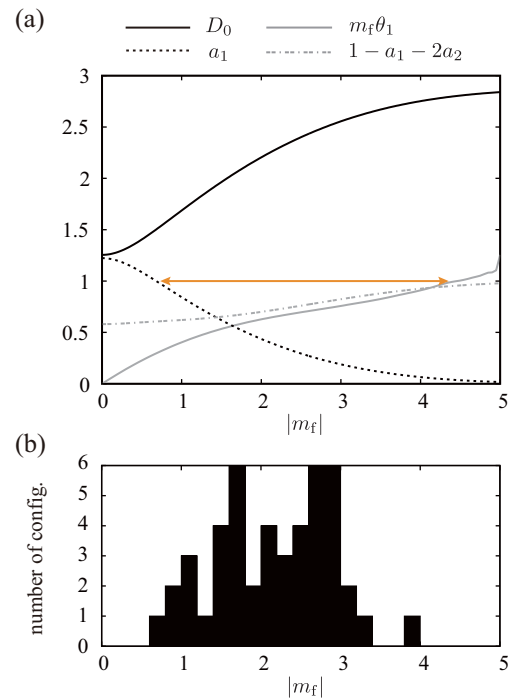


FIG. 8. (a) Values of the constants that appear in the stability condition for a fixed-point solution with  $m(t) = m_f$  and  $\sigma_0 = 1.2$ . The values of the constants do not depend on the value of  $g_0$ . The stability condition is satisfied for the range of  $m_f$  indicated by the double-ended orange arrow. (b) The histogram of (the absolute values of) the mean activity of the fixed-point solutions obtained by directly simulating one hundred networks in the same setting as Fig. 6(e) except for different weight configurations.

very small  $g_0$  and also suggests that stable fixed points still exist, albeit with a very small probability, for such small  $g_0$ .

#### 4. Regularly oscillating solutions and their stability

Next, we examine the regularly oscillating solutions. If the entire network dynamics have stable oscillations, so do the microscopic parts of the dynamics. To find such microscopic oscillations for a given oscillatory orbit of the mean activity,  $m(t) = m_o(t)$ , we solve the self-consistent equation, Eq. (28), which we rewrite as

$$(1 + \partial_t)(1 + \partial_s)D(t, s) = 2\sigma_0^2 \tilde{\mathcal{C}}(t, s). \quad (43)$$

This equation can be solved iteratively in the frequency domain (see Appendix I). Using the mean activity observed in Fig. 6(g) for  $m_o(t)$ , we compute the autocorrelation in the frequency domain,

$$\widehat{D}(\omega_1, \omega_2) \stackrel{\text{def}}{=} \frac{1}{2\pi} \int_{\mathbf{R}^2} e^{-i\omega_1 t + i\omega_2 s} D(t, s) dt ds, \quad (44)$$

[Fig. 9(a)]. We find that the solution has a nonzero value only for multiples of the basic frequency  $\omega_0$  of the mean activity, which indicates that the microscopic dynamics are completely entrained by the oscillatory mean activity. We also confirm this by checking the following averaged autocorrelation function calculated from  $D(t, s)$ : with  $T_0 = 2\pi/\Delta\omega$  for the

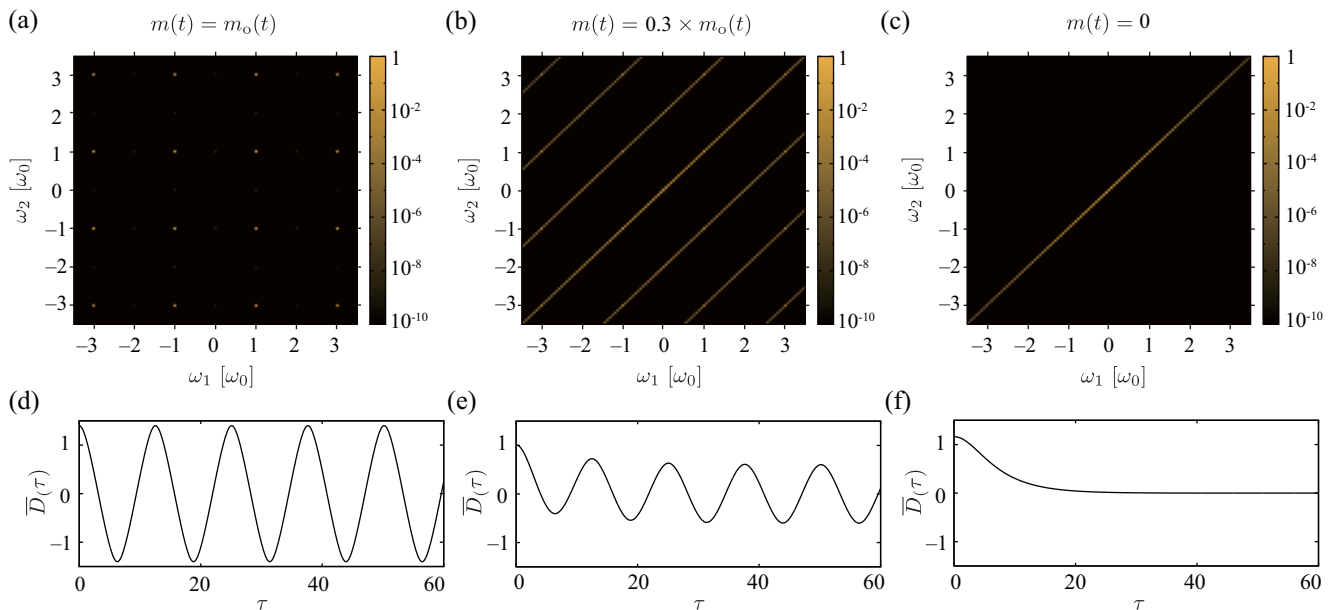


FIG. 9. [(a)–(c)] Fourier-transformed autocorrelation function  $\widehat{D}(\omega_1, \omega_2)$  for the mean activity indicated above each panel. The time series  $m_o(t)$  is the mean activity shown in Fig. 6(g), which oscillates regularly with the basic frequency  $\omega_0$ . The values of  $\widehat{D}(\omega_1, \omega_2)$  are calculated over the frequency domain  $-256\omega_0 \leq \omega_1, \omega_2 \leq 256\omega_0$  discretized into bins for which the frequency values are represented by multiples of  $\omega_0/16$ . We plot the absolute values of the calculated  $\widehat{D}(\omega_1, \omega_2)$  for  $-3.5\omega_0 \leq \omega_1, \omega_2 \leq 3.5\omega_0$  in a logarithmic color scale. The two argument frequencies,  $\omega_1$  and  $\omega_2$ , are shown in units of the basic frequency  $\omega_0$ . [(d)–(f)] The averaged autocorrelation function  $\overline{D}(\tau)$  for the three cases in the upper panels.

bin width,  $\Delta\omega = \omega_0/16$ , of the frequency discretization,

$$\overline{D}(\tau) \stackrel{\text{def}}{=} \frac{1}{T_0} \int_0^{T_0} D(t, t + \tau) dt. \quad (45)$$

The autocorrelation function in Fig. 9(d) shows that the neuronal variables are periodic. In contrast, as is well known from a previous study [19], the autocorrelation function  $D(t, s)$  for zero mean activity,  $m(t) \equiv 0$ , has the frequency representation,  $\widehat{D}(\omega_1, \omega_2) = \widehat{D}_0(\omega_1)\delta(\omega_1 - \omega_2)$ , with a continuous function  $\widehat{D}_0(\omega_1)$  [Fig. 9(c)]. The averaged autocorrelation function for this case is unimodal and tends to zero as  $\tau \rightarrow \infty$  [Fig. 9(f)]. The qualitative difference between these two autocorrelation functions suggests the occurrence of a phase transition from one to the other. To check this, we decrease the amplitude of the mean activity,  $m_o(t)$ , without changing its waveform, and we find that the solution starts to have a continuous spectrum extending over frequencies other than the multiples of  $\omega_0$ . After the transition, the autocorrelation function has the form of  $\widehat{D}(\omega_1, \omega_2) = \sum_k \widehat{D}_k(\omega_1)\delta(\omega_1 - \omega_2 - k\omega_0)$  [Fig. 9(b)] and the averaged autocorrelation function has both a periodic component and a component that vanishes at infinity [Fig. 9(e)]. As the amplitude of the mean activity decreases, the periodic component in the autocorrelation function gradually decays, disappearing at  $m(t) = 0$ . The analysis we perform in Appendix I actually shows that these observed entrained dynamics are stably realized. The transition behavior observed above is qualitatively the same as that observed in a previous study [29], although that study used periodic inputs with random phases to induce the transition.

This microscopic transition gives an intuitive explanation for the mechanism of the observed oscillations. Recall that the dynamics of the mean activity is described by Eq. (19),

rewritten here as

$$\frac{d}{dt}m(t) = -m(t) + \sqrt{2}g_0\eta(t). \quad (46)$$

Because the driving-force term,  $\eta(t)$ , is the sum of the microscopic fluctuations, it is entrained to the oscillation of the mean activity itself, if the mean activity oscillates with a sufficiently large amplitude, to entrain the microscopic fluctuations. We suggest this reverberation of entrainment between the mean activity and the microscopic fluctuations as the mechanism underlying the coherent oscillations we observe in Figs. 6(g) and 6(h).

The stability of this reverberation mechanism can be examined by using a perturbative method similar to that employed for fixed points: we derive a self-consistent equation with random coefficients that determines linear responses to external inputs, and construct its solution from which the condition for the stability of the reverberation can be examined numerically. From this analysis, we draw the same conclusion about the stability as that for the fixed points: regularly oscillating solutions for the untuned model, such as that observed in Fig. 6(g), are linearly stable with a nonzero probability; occasionally observed regular oscillations of the finely tuned model [Fig. 15(c) in Appendix C], however, turn out to be unstable. These conclusions are consistent with the tendency observed in the results of direct simulations of networks with different system sizes [Figs. 17(a) and 17(b) in Appendix C]. Since this analysis is complicated and essentially the same as that for fixed points, we omit its presentation here and refer interested readers to Appendix I. We only note that we cannot examine exhaustively oscillatory orbits, and therefore



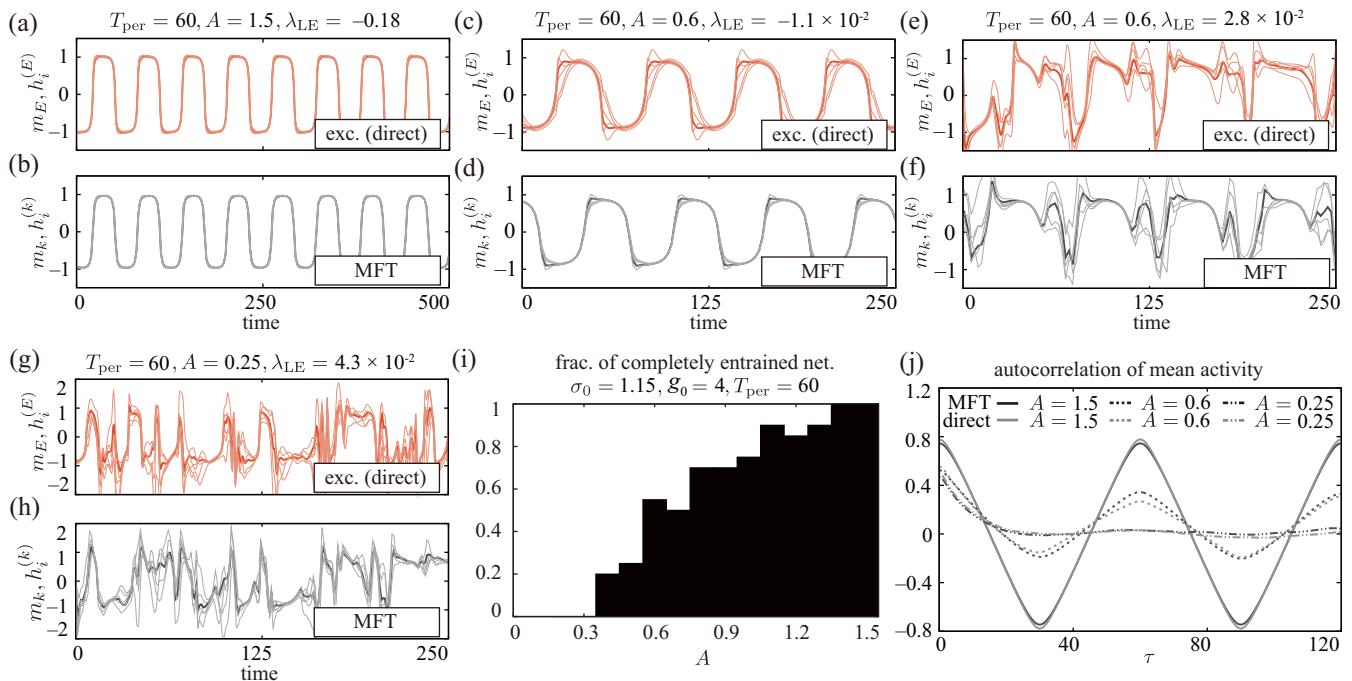


FIG. 10. [(a)–(h)] Typical activity patterns for the finely tuned model with  $(\sigma_0, g_0) = (1.15, 4)$  under sinusoidal external inputs with the indicated values of  $A$  and  $T_{\text{per}}$ . In each plot, the thick and thin lines represent the mean activity  $m_k$  and the activities  $h_i^{(k)}$  of five representative neurons, respectively. For the direct simulations, the mean activity  $m_k$  is approximated by the average of  $h_i^{(k)}$  over the population  $k$  comprising  $N = 10240$  neurons (see Appendix A for the details of the simulations). Solutions for the  $E$  population in direct simulations and solutions of the mean-field equations are depicted in red and gray, respectively. The plots for the  $I$  population are omitted. The numerically determined value of the largest Lyapunov exponent,  $\lambda_{\text{LE}}$ , is shown above each panel. Qualitatively different dynamics are observed for the same parameter values [(c)–(f)], depending on the configuration of the random connectivity of the directly simulated networks or on the sequence of random numbers used for the simulations of the mean-field equations. (i) For each value of  $A$ , the percentage of twenty directly simulated networks with random weight configurations that synchronize with the sinusoidal inputs is plotted as a histogram. (j) Autocorrelation function of the mean activity  $\mu(\tau)$  for  $(\sigma_0, g_0) = (1.15, 4)$  and for the indicated values of the amplitude  $A$  averaged over 15 different configurations or 15 series of random numbers and over the period  $1000 \leq t \leq 2500$  for both direct simulations and solutions of the mean-field equations.

we cannot completely exclude the existence of stable limit-cycle solutions for the finely tuned model.

### C. External inputs to $\sqrt{N}$ excitatory neurons

#### 1. Coherent states induced by sinusoidal inputs

In this section, we apply sinusoidal inputs of amplitude  $A$  and period  $T_{\text{per}}$  to  $\sqrt{N}$  neurons in the  $E$  population of the finely tuned model, namely,

$$\tilde{I}(t) = A \sin(2\pi t/T_{\text{per}}) \quad (47)$$

in Eq. (4). In this model setting, we observe two qualitatively different types of behavior [Figs. 10(a)–10(h)]. We find that the activity patterns obtained from the direct simulations and from the MFT are quite similar, suggesting that our theory successfully predicts the behavior of the model for the case with external inputs as well. As we increase the amplitude  $A$  for a fixed value of  $g_0$ , the solutions undergo a transition from irregular chaotic dynamics partially entrained by the input to regular nonchaotic dynamics synchronous with the input. This indicates that inputs to an  $O(\sqrt{N})$  number of neurons can effectively entrain the whole network in this model.

We further observe that this transition occurs at different values of  $A$ , depending on the configuration, but not on the

initial condition, similarly to Fig. 7 (not shown). Figure 10(i) shows a histogram depicting the percentage of twenty networks with random configurations that synchronize with the inputs for each value of  $A$ . We see that the transition point is highly variable among networks with different configurations. Nevertheless, in the autocorrelation function  $\mu(\tau)$  of the mean activity averaged over configurations and time according to the definition in Eq. (30), we observe good agreement between direct simulations and MFT [Fig. 10(j)].

#### 2. Stability of the entrained dynamics

The stability of the regular, entrained dynamics can be examined using the same perturbative method as that for the regularly oscillating solutions of the model without external inputs. In this case, the results of the analysis indicate that the numerically observed oscillations of the model are linearly stable (Appendix D). Consistent with this finding, in the direct numerical simulations, the induced coherent states are robustly observed for networks with large system sizes (Appendix C). In contrast with the untuned model, in the present case, the synaptic weight matrix of the model does not have configuration-dependent outlier eigenvalues. We observe, however, that the linear variational equation determin-

ing the stability of the attractors does have coefficient matrices with outlier eigenvalues (see Appendix B for details). These results suggest that the finely tuned model still shows strong configuration dependence in its stimulus-driven dynamics.

### 3. Reading out information from coherent dynamics

Figures 10(a)–10(d) show that the mean activities and individual neuronal activities are coherent in dynamics synchronous with the inputs. This has a computational implication. Suppose that we read out microscopic fluctuations of these networks by taking a weighted average with  $O(1/N)$  weighting coefficients, as described by the following equation:

$$\gamma(t) = \sum_{k,i} r_k^i \frac{r_k^i}{N} (\phi(h_i^{(k)}(t)) - \check{\phi}_k(t)). \quad (48)$$

Here,  $\check{\phi}_k$  denotes the population average of  $\phi(h_i^{(k)}(t))$  over the population  $k$ . In the above equation, the contribution of this population average is subtracted. This is because the population-averaged activity simply replicates the external inputs, and reading out this component does not have much computational value.

If the dynamics are coherent, we expect the read-out values,  $\gamma(t)$ , to be  $O(1)$ . In contrast, if the dynamics are chaotic, the ensemble of neuronal activities can be regarded as an incoherent Gaussian fluctuations, and therefore, values read out from them are expected to be  $O(1/\sqrt{N})$ . We numerically test this hypothesis by examining the values read out with the following weighting coefficients:

$$r_k^i = \begin{cases} 1 & \text{if } \phi(h_i^{(k)}(t_{\text{init}})) - \check{\phi}_k(t_{\text{init}}) \geq 0 \\ -1 & \text{otherwise} \end{cases}. \quad (49)$$

In this equation, we set the coefficients to such values that the initial value of  $\gamma(t)$  at time  $t_{\text{init}}$  is  $O(1)$ . We show the network activities and typical read-out values obtained from them in Figs. 11(a) and 11(b). The values read out from the coherent regular oscillation show a regular pattern of magnitudes comparable to the initial value,  $\gamma(t_{\text{init}})$ , [Fig. 11(a)], while those read out from the irregular activity decay rapidly from the initial value [Fig. 11(b)]. This observation is consistent with the above argument. To evaluate the magnitudes of the read-out values further, we calculate the following normalized standard deviation,  $\widehat{S}_\gamma$ , of  $\gamma(t)$  for networks of different system sizes:

$$\begin{aligned} S_\gamma &\stackrel{\text{def}}{=} \left\{ \frac{1}{T_{\text{avg}}} \int_{t_0}^{t_0+T_{\text{avg}}} (\gamma(t) - \bar{\gamma})^2 dt \right\}^{1/2}, \\ \bar{\gamma} &\stackrel{\text{def}}{=} \frac{1}{T_{\text{avg}}} \int_{t_0}^{t_0+T_{\text{avg}}} \gamma(t) dt, \quad \widehat{S}_\gamma \stackrel{\text{def}}{=} \frac{S_\gamma}{S_\phi}, \\ S_\phi &\stackrel{\text{def}}{=} \left\{ \frac{1}{T_{\text{avg}}} \int_{t_0}^{t_0+T_{\text{avg}}} ((\phi(h_i^{(k)}(t)) - \check{\phi}_k)^2) dt \right\}^{1/2}. \end{aligned} \quad (50)$$

The bracket in this equation indicates the average over the entire network. We also take the average over a long time, of length  $T_{\text{avg}}$ , starting from a suitably chosen initial time  $t_0$ , for a simulation that starts from a random initial condition different from that of the simulation for which we have determined the weighting coefficient  $r_k^i$ . We also show the activity patterns

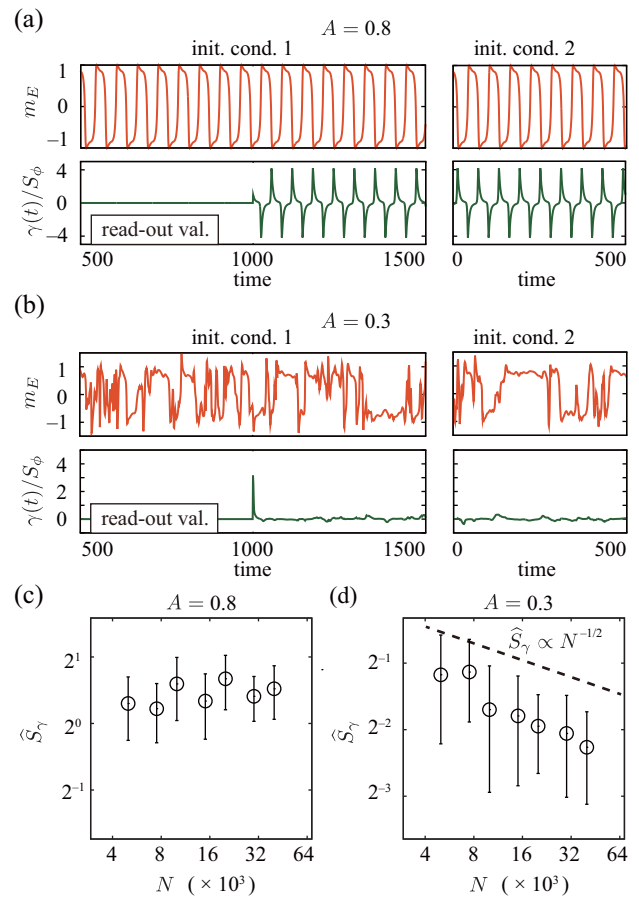


FIG. 11. [(a) and (b)] Mean activities of the excitatory population  $m_E$  and normalized read-out values  $\gamma(t)/S_\phi$ , as calculated for networks with  $(\sigma_0, g_0) = (1.15, 4)$  that receive sinusoidal inputs with the indicated values of  $A$  and  $T_{\text{per}} = 60$ . The weighting coefficients  $\{r_k^i\}_{k,i}$  are determined at  $t_{\text{init}} = 1000$  in the simulations from initial condition 1. The mean activities and read-out values of networks from another random initial condition (initial condition 2) are also shown. [(c) and (d)] Scaling of the normalized standard deviation  $\widehat{S}_\gamma$  of values read out from networks with the same parameter values as in (a) and (b). The averages of  $\widehat{S}_\gamma$  over fifteen different configurations, together with their standard errors, are plotted on logarithmic scales. Only networks with configurations giving rise to synchronous dynamics (c) or irregular dynamics (d) are analyzed. In (d), a straight line with slope  $-1/2$ , representing  $\widehat{S}_\gamma \propto N^{-1/2}$ , is also shown.

obtained from this initial condition in Figs. 11(a) and 11(b). In Figs. 11(c) and 11(d), we show the calculated values of the normalized standard deviations on logarithmic scales, and we find that the values read out from regular oscillations do not depend much on the system size, while those read out from irregular dynamics are roughly proportional to  $1/\sqrt{N}$ , as we expect. These results suggest that the above mechanism for reading out  $O(1)$  values only when the network dynamics are coherent enables neuronal networks to transmit information in a state-dependent manner. Note that we have repeatedly read out the same pattern from the coherent dynamics, regardless of the independent initial conditions [Fig. 11(a)]. Regardless of the symmetry among the neurons that receive inputs through statistically the same set of synaptic weights, identical

coherent dynamics—not coherent dynamics randomly reshuffled with respect to the neuronal indices—are always realized. Although the above coherent states are induced by artificial sinusoidal input, similar results are obtained for the case with irregular input (see Appendix D for details).

#### 4. Remarks on the untuned model under external inputs

The untuned model behaves in a qualitatively similar manner to the finely tuned model when both are driven by external inputs. The untuned model also shows transitions from irregular, partially entrained dynamics to regular dynamics that are completely synchronous with external inputs in a configuration-dependent manner. Thus, to avoid redundancy, we do not present the results for this model setting in this paper. From a quantitative viewpoint, we note that the entrainment in this model setting is more complicated than that for the finely tuned model, presumably because the untuned model has inherent configuration-dependent rhythms, as observed in Fig. 6.

#### D. Multiscale dynamics of critically balanced networks of LIF neurons

Thus far, we have focused on a highly simplified model with firing-rate variables. However, from way we have constructed our MFT, we expect that in principle a similar theoretical framework will hold for networks of spiking neurons (and hence we expect similar multiscale dynamics to those observed in the rate model). To demonstrate this, we numerically examine a commonly investigated network of leaky-integrate-and-fire (LIF) neurons (see, e.g., Ref. [63]) described by the following equation: for  $k = E$  and  $I$ ,

$$\tau_m \frac{d}{dt} V_i^{(k)}(t) = -V_i^{(k)}(t) + \sum_{\ell=E,I, 1 \leq j \leq N,n} J_{k\ell}^{ij} \delta(t - t_{j,n}^{(\ell)} - t_\Delta), \quad (51)$$

In this equation, the variables  $V_i^{(k)}$  and  $t_{i,n}^{(k)}$  denote the membrane potential and the  $n$ -th spike time of the  $i$ th neuron in the  $k$  population, respectively. If  $V_i^{(k)}$  exceeds a threshold potential  $V_{\text{th}}$ , a spike is emitted from the neuron, and  $V_i^{(k)}$  is reset to  $V_{\text{reset}}$  and held for a time of length  $t_{\text{ref}}$ . The constants  $\tau_m$  and  $t_\Delta$  represent the membrane time constant and the delay of synaptic transmissions, respectively. The synaptic weights  $J_{k\ell}^{ij}$  are given by Eq. (2) or (5) subject to the condition in Eq. (3). For this model, following the same argument as that for the simplified model (Sec. III), fluctuations in the inputs to the neurons in the network are considered to be conditionally Gaussian, given the orbit of the population-averaged input to the neurons. Since the population-averaged input is given by the sum of the fluctuating inputs to individual neurons divided by  $\sqrt{N}$ , the population-averaged input is stochastic and its realization probability is determined through the correlation of the microscopic fluctuations. The neuron model used in the above, however, is highly nonlinear, and thus solving the mean-field equations demands much more intensive numerics than those we presented in the previous sections. Thus we restrict ourselves to numerically simulating the model

and examining whether similar multiscale dynamics arise intrinsically.

As the value of  $g_0$  is increased from zero with the above model settings, we actually observe increasingly large fluctuations in the population activity [Figs. 12(a)–12(c)]. Examining the autocorrelation and cross-correlation functions of individual and population-averaged inputs to the neurons, we observe that the population-averaged input is strongly correlated with the excitatory, inhibitory, and overall inputs to individual neurons [Figs. 12(d)–12(f)], which can account for the experimentally observed large cross-correlations between EEG/LFPs and neuronal activities [15] (see Sec. VC for further discussion). As we increase the value of  $g_0$ , these dynamics undergo a transition to coherent dynamics in a configuration-dependent manner [Figs. 12(c) and 12(f)]. These transitions are observed robustly for different values of  $N (\gg 1)$  (Fig. 17 in Appendix C). Similar transitions are induced by external inputs to  $\sqrt{N}$  neurons (Fig. 16 in Appendix C). The successful prediction of the occurrence of intrinsically generated population dynamics indicates that the multiscale dynamics revealed by our MFT generally emerge in a variety of neuronal networks at critical parameter values.

#### V. DISCUSSION

In the present study, we developed a novel type of MFT for densely-connected RNNs consisting of a pair of excitatory and inhibitory populations of simplified neurons with finely tuned or untuned synaptic weights that obey Dale's law. The mean strengths of the synaptic weights were assumed to take a set of critical values. In this theory, microscopic fluctuations in the neuronal activities that are amplified by the strong excitation and inhibition serve as driving forces for the macroscopic dynamics of the population activity, while the population activity constrains the statistics of the microscopic fluctuations. The investigated RNNs exhibited nonvanishing fluctuations in their population activity. When the magnitudes of excitation and inhibition were large, we found interesting dynamical properties in these fluctuations, such as high non-Gaussianity and asymmetry with respect to time reversal for the finely tuned model, and strongly configuration-dependent transition to a static or oscillatory nonchaotic state for the untuned model. In the oscillatory state, neuronal activities have various waveforms while they are all phase locked to the rhythm of the population activity. Our theory successfully predicts these dynamical properties. We found that these multiscale dynamics occur at a critical point between extremely strong ferromagnetic and antiferromagnetic states.

In networks with external inputs, periodic inputs to an  $O(\sqrt{N})$  number of excitatory neurons effectively entrained the whole network. As the amplitude of the inputs was increased, the networks underwent a transition from irregular, partially entrained dynamics to regular dynamics synchronous with the input; the transition point again depended on the configuration. Unlike the autonomous case, the application of external inputs induced a transition to a coherent oscillation in the finely tuned model, indicating that the fine-tuning of the synaptic weights reduces but does not remove the configuration dependence from the network dynamics. We also showed numerically that the induced coherent dynamics can be used

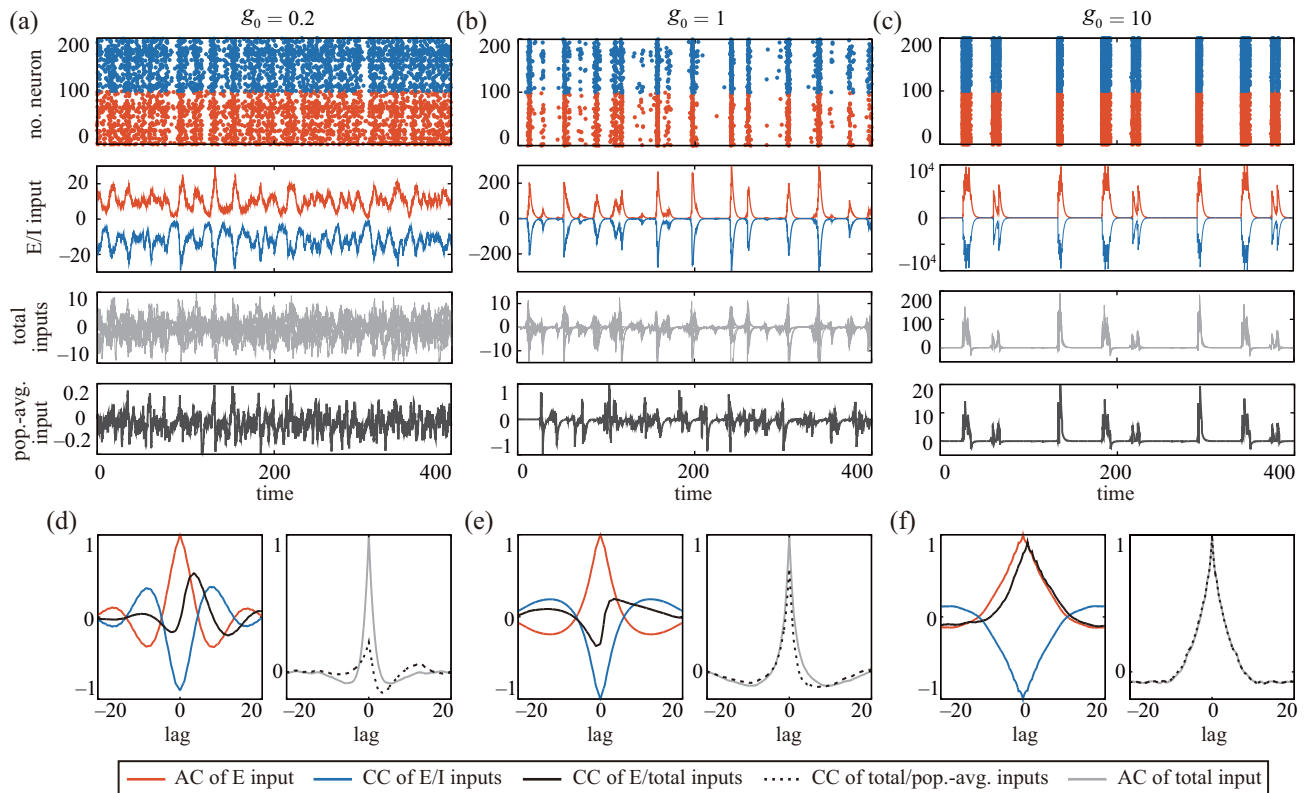


FIG. 12. [(a)–(c)] Multiscale dynamics obtained by numerically simulating a network of LIF neurons with  $g_0 = 0.2, 1,$  and  $10,$  respectively. The top panels are raster plots of firing activities of one hundred representative neurons from the excitatory (red) and inhibitory (blue) populations. The upper middle panels show the total excitatory (red) and inhibitory (blue) inputs to a single representative excitatory neuron. The lower middle panels show total inputs to five representative excitatory neurons (gray). The bottom panels show the population average of the inputs to all neurons. [(d)–(f)] The autocorrelation functions (ACs) of the excitatory input to a representative neuron and of the population average of the inputs to all neurons, and the cross-correlations functions (CCs) of those inputs and another fraction of the input to the representative neuron. The quantities shown are the AC of the excitatory input to the neuron (red); the CC of the excitatory and inhibitory inputs to the neuron (blue); the CC of the excitatory input to the neuron and the population-averaged input (black); the CC of the total input to the neuron and the population-averaged input (black dotted); and the AC of the total input to the neuron (gray). The following parameter values are used for the simulations:  $\tau_m = 20,$   $t_\Delta = 0.5,$   $t_{\text{ref}} = 0.55,$   $V_{\text{rest}} = 0,$   $V_{\text{reset}} = 10,$  and  $V_{\text{th}} = 20.$

as media for transmitting information in a state-dependent manner.

Furthermore, based on analogy, more biologically realistic networks of spiking neurons are expected to display similar multiscale dynamics. Although theoretical prediction of their dynamics is much more computationally demanding and beyond the scope of the current study, we numerically confirmed this hypothesis for networks of LIF neurons.

### A. Closely related results

The present study was largely inspired by a previous investigation of RNNs [32] and by a couple of published and unpublished studies on the same model as ours [62,64]. In the former study [32], the authors investigated a network with a balance between strong excitation and inhibition, providing a theoretical framework for dealing with a network with multiple neuronal populations and for analyzing its nontrivial fixed points and transitions to chaos. Our study used their theoretical framework as a starting point to analyze further the nontrivial population dynamics that they had not analyzed.

In the latter published study [64], the authors numerically analyzed the same network as ours and found similar oscillatory dynamics. They further analyzed the dynamics with approximate reduced equations and related them to the eigenvalues of the connectivity matrices. Although these results were quite inspiring, their approach—focusing on the eigenvalues with the largest real parts—was not always sufficient to characterize behavior of the model. It is known that the eigenvalues of the connectivity matrices of their networks with finely tuned weights in the  $N \rightarrow \infty$  limit are uniformly distributed over a disk [61] [Fig. 13(b)]. This implies that it is difficult to select a single eigenspace that effectively determines the dynamics in this limit, on which their theory relied. The unpublished study [62] also took a similar approach for the same model and came to a pessimistic conclusion about the usefulness of an MFT in this setting. Sometime after we initially publicized the present results, a very recent work [65] examined the case with  $g_0 \approx 0, \infty$  for our networks. Those authors developed a perturbative mean-field theory combined with the approach of Ref. [64]. Although they gained insights into the behavior of the model analytically,



their theory still largely relies on heuristic, approximate calculations, and its application is limited to nearly deterministic dynamics with small fluctuations. In contrast with these previous studies, we have here derived an MFT in a much more rigorous manner and have laid a foundation for further analysis. Our theory applies to the entire range of model parameters and gives accurate probabilistic descriptions of large macroscopic fluctuations.

### B. Stimulus-induced suppression of chaos and reservoir computing

Prior to the present study, several authors have theoretically studied the externally driven, nonchaotic dynamics of neuronal networks without balanced excitation and inhibition [29,66,67]. These studies have shown that the chaoticity of the dynamics of RNNs is suppressed by external random inputs and that a nontrivial nonchaotic regime appears after a transition at some amplitude of the input. In particular, a seminal study [29] showed that sinusoidal inputs with random phases induce a coherent state similar to ours. The transition to a coherent state in the present model is closely related to this suppression of chaos by stimulus, because the microscopic part of our model dynamics are statistically the same as the dynamics of a model without balanced excitation and inhibition that is suitably driven by a uniform external input, as shown by our MFT. In fact, the autocorrelation functions of neuronal activities of our networks (Fig. 9) behave in a similar manner to those observed in the previous study [29]. However, we note that the transition to this microscopic coherent state due to a uniform external input, not with random phases, has not been well studied to date. Besides the fact that the transition induced by uniform input is more difficult to analyze theoretically, the uniform application of an input often results in chaotic or trivial dynamics, and the transition to a coherent state is not found unless the waveform of the input is finely tuned. In the present model, the waveforms of the mean activity that induce a coherent state are determined by the network itself through the interactions between the microscopic and macroscopic dynamics, even in a case with external inputs. The main difference between the coherent states in the present study and those in previous studies lies in this spontaneity.

The spontaneously discovered coherent states discussed above may have implications for learning with RNNs. In previous studies, learning was first considered in the context of the “*edge of chaos*,” where the variety and stability of network dynamics at the transition point to chaos were exploited in learning [68–70]. More recent studies have focused on different nonchaotic dynamical phases induced by external or feedback inputs [40,67,71]. In particular, the authors in a seminal work [40] stably reconstructed desired patterns from the coherent dynamics induced by randomly weighted strong feedback from read-out values to all of the neurons in the network. This strong random global feedback is expected to induce coherent dynamics by a mechanism similar to that studied in Ref. [29] (see also Refs. [67,72] for a similar result with strong random global feed-forward input). This requirement for a strong global input, however, restricted the applicability of their framework to the supervised learning

of a small number of temporal patterns. Our results suggest a new regime of dynamics, in which nonchaotic coherent dynamics emerge spontaneously and stably reproduce output patterns [Fig. 11(b)], without being passively entrained by strong global inputs. Thus, investigating learning based on the dynamical phase we have found is a worthy challenge for a future study.

### C. Population dynamics and critical fluctuations

The relationship between population dynamics and individual neuronal activities has also been studied in previous models. In these studies, however, population dynamics and microscopic fluctuations in individual neurons were treated as statistically independent. Therefore, unlike our theory, none of the previously proposed theories for balanced networks could account for the experimentally observed strong impact of single neurons on the population activity [53,54].

In sparsely connected balanced networks of spiking neurons, population dynamics are unaffected by fluctuations in the irregular firing of individual neurons. In fact, a previous study [25] showed that even when individual neurons fire irregularly, the population-averaged activity exhibits regular slow oscillations except for tiny fluctuations due to finite-size effects. This indicates the fact that the irregular firing of neurons exerts only negligible effects on the population dynamics of the sparsely connected networks.

In densely connected balanced networks investigated previously, stimuli to a small number of excitatory neurons also induce a vanishingly small response in the entire population. Two recent studies investigated the responses of such networks with spatial structures to correlated external inputs applied to a large number [ $O(N)$ ] of neurons [51,52]. It was shown analytically that a non-negligible population response can be induced only when the spatial extent of the input correlation is narrower than that of recurrent connections from a single neuron [52]. This theoretical result should also hold when stimuli are given to a small number [ $O(\sqrt{N})$ ] of excitatory neurons because such stimuli generate correlated internal inputs to the surrounding neurons that are connected with the stimulated neurons. In this case, the previous theory indicates that the population response is negligibly small.

The difference in the impact of single neurons on the population dynamics between the previous models and our model can be understood from the strong ferromagnetic effects examined in Sec. IV A 3. Previous models focused on the dynamical regime in which the network activity is stabilized by strong inhibitory feedback that suppresses excessive excitations. This regime corresponds to the antiferromagnetic state we observed in Sec. IV A 3. The antiferromagnetic effects strongly suppress the responses of the neuronal population when a small number of neurons are stimulated. In contrast, we have focused on the dynamical regime emerging at the critical point between the ferromagnetic and antiferromagnetic states. Activated spontaneously or driven by stimuli to a small number of neurons, our model displays strong macroscopic fluctuations at the critical point. Remarkably, our MFT precisely describes the probabilistic behavior of these critical fluctuations, which was, to our knowledge, difficult for any of

the previous MFTs developed in the statistical mechanics of disordered systems. Although the parameter values yielding the critical point are not generic, experiments have shown evidence for self-organized critical dynamics in the brain [57], and thus we can reasonably expect some adaptive mechanisms to finely tune the system to the critical point.

The intrinsic origin of the population dynamics may also be supported by the experimentally observed large cross-correlations of EEG/LFPs and individual neuronal activities [15]—namely, strong coherence over the macroscopic and microscopic scales. EEG and LFPs are considered to reflect mainly a collective excitatory component of synaptic inputs to (apical dendrites of) neurons in the local circuit [73], and their waveforms are similar to those of the very large excitatory inputs to neurons [48]. When the strong excitatory and inhibitory inputs to neurons cancel out, the remaining fluctuations do not need to be strongly correlated with the original excitatory and inhibitory inputs, especially if the main driving-force for the population dynamics are extrinsic. In fact, from the multiscale dynamics of the previous model, small cross-correlations of population dynamics and recurrent excitatory input are expected (e.g., by comparing the activity of a population of neurons and the total excitatory inputs to those neurons from the inside and outside of the population in Fig. 1 of Ref. [52]). In contrast, our model shows large cross-correlations between the excitatory input to a neuron and population-averaged input to neurons [Figs. 12(d)–12(f)]. Since the large excitatory inputs are equal among neurons of our model in the leading order in  $N$ , our results imply that cross-correlations of excitatory inputs to the population and individual neuronal activities are large on average, explaining the experimentally observed large cross-correlations of EEG/LFPs and individual neuronal activities. Thus, modeling of coherent multiscale cortical dynamics with our theory is a promising challenge. From the fact that cortical activity displays switching behavior between states with small and large cross-correlations of EEG/LFPs and neuronal activities [15], our theory and previous theories are suggested to separately model two different operating regimes of the same cortical circuits.

In our theory, the dynamical nature of the critical fluctuations depends strongly on the detailed configuration of the synaptic connections. On the other hand, accumulating evidence suggests that the connectivity of local cortical circuits is rapidly remodeled [74]. Whether the fine connectivity structure has a strong impact on critical fluctuations in cortical network dynamics—and what functional implications such an impact has—need to be further clarified.

#### D. Limitations and future extensions of the theory

Despite the advantages of our theory mentioned above, it is fair to say that the validity of the theory is still restricted by the simplicity of the model settings. One of the most important steps for widening the applicability of our theory is to extend the theory to networks of spiking neurons. In the present study, we gave priority to the analytical tractability and simplicity in numerical simulations, and we restricted ourselves mainly to networks of firing-rate model neurons. However, the extension of our theory to networks of more

realistic model neurons should be straightforward. This can be done by regarding balanced inputs to individual neurons as Gaussian fluctuations and by determining the related statistics to ensure consistency with the nonlinear dynamics of single neurons. In this calculation, we can use our method in combination with a previous method [31] to describe the network dynamics. In the previous study, the mean-field equations for a simple RNN of nonlinear firing-rate units without balanced excitation and inhibition was solved by using the statistics calculated from extensive numerical simulations of single units driven by random forces. Applying the combined method to a critically balanced network of spiking neurons would be computationally demanding but in principle doable.

Although we leave this challenge for future study, we note that the computational costs associated with the approach of Ref. [31] cannot be reduced by commonly employed approximate treatments such as white Gaussian approximation of inputs to neurons [25,63]. This is because we must take into account the time-dependent nonlinear interactions between the microscopic neuronal fluctuations and macroscopic population activity underlying the critical multiscale dynamics. These interactions cannot be handled by the approximate treatments. This distinction between a full treatment and an approximate treatment may be related to the recent controversial argument about the transition in networks of spiking neurons from a state with irregular spiking at a constant rate to a state with irregular firing-rate fluctuations, as the mean strength of the synaptic connections is increased [63]. Although the interpretation of this observation based on an approximate description was controversial [75], a full treatment is expected to give an accurate description.

The other simplified aspects of the model include the neglect of different cellular properties of excitatory and inhibitory neurons [76] and of characteristic spatial patterns of cortical neuronal responses. Although our theory can be extended to include these elements, substantial works will be needed for that. For example, for models with different membrane constants for excitatory and inhibitory neurons, automatic cancellation of excitatory and inhibitory inputs with  $\bar{\phi}_E = \bar{\phi}_I$  is not ensured from the condition in Eq. (3). However, it is plausible that some feedback mechanism dynamically clamp the population activity to satisfy  $\bar{\phi}_E = \bar{\phi}_I$ . Then, similar critical multiscale dynamics to those we have observed are expected to emerge. In experimental studies, characteristic center-surround response patterns have been observed when a small number of neurons were stimulated [54,77]. Thus it is an important open challenge to understand how critical multiscale dynamics emerge in a spatially extended model and to examine whether those dynamics are consistent with the experimentally observed spatial patterns.

From the theoretical point of view, another question that remains unanswered concerns the way qualitatively different solutions bifurcate in our model when we increase the magnitudes of excitation and inhibition. Theoretical analysis of this bifurcation is hard due to the fact that the MFT is constructed based on an averaging over network configurations while the bifurcation point depends strongly on the individual configurations. Regarding this point, a recent study [33] developed a theory of linear dynamics for disordered

systems with individual configurations. The stochastic linear response theory shown in Appendices H and I also allows us to analyze the response dynamics around fixed points and regular coherent oscillations for individual configurations. However, to identify the type of a bifurcation, information is needed about the lowest-order nonlinear term relevant to the bifurcation. We expect lowest-order nonlinear corrections to our linear-response theory to provide nonlinear response dynamics valid for individual configurations and information about the bifurcations.

### ACKNOWLEDGMENTS

We thank Shun Ogawa, Naoki Hiratani, Yasuhiro Tanaka, Chigaku Itoi, Maximilian Schmidt, Takuma Tanaka, Milena M. Carvalho, Tomoki Kurikawa, and Toshitake Asabuki for their kind help with fruitful discussions and advice. We thank Keita Watanabe for his general advice on numerical simulations. We also thank two of our colleagues, Miho Itoi and Ryotaro Inoue, for their kind support to the work environment for the revision of the manuscript. This work was partly supported by the Brain Mapping by Integrated Neurotechnologies for Disease Studies (Brain/MINDS) from AMED. This work was also supported by JSPS KAKENHI Grant No. JP19K20359, and by MEXT KAKENHI Grants No. JP19H04494 and No. JP18H05213.

### APPENDIX A: SIMULATIONS OF NEURONAL NETWORKS THAT DO NOT VIOLATE DALE'S LAW

In the main text, we simulate the model equations, Eq. (1) or (51) together with synaptic weights described by Eq. (2) or (5), directly. In this section, we describe the details of the simulations. We first describe the random variables  $\mathcal{J}_{k\ell}^{ij}$ . As indicated in Ref. [62], we can choose random variables for the connectivity, so that the model does not violate Dale's law, a rule that prohibits neurons from feeding both excitatory and inhibitory connections. We use the following random variables with zero mean and unit variance:

$$\pm \mathcal{J}_{k\ell}^{ij} = \begin{cases} \sqrt{\frac{1-p}{p}} & \text{prob. } p \\ -\sqrt{\frac{p}{1-p}} & \text{otherwise} \end{cases}.$$

The sign before  $\mathcal{J}_{k\ell}^{ij}$  is positive if the population  $\ell$  is excitatory, and negative otherwise. With the same usage of sign  $\pm$  for  $J_{k\ell}^{ij}$ , these random variables give, for both Eqs. (2) and (5),

$$\pm \sqrt{N} J_{k\ell}^{ij} = \begin{cases} \sigma_0 \sqrt{\frac{1-p}{p}} + g_0 + O_p(1/\sqrt{N}) & \text{prob. } p \\ -\sigma_0 \sqrt{\frac{p}{1-p}} + g_0 + O_p(1/\sqrt{N}) & \text{otherwise} \end{cases}.$$

Note that the effect of the adjustment in the second of Eq. (5) is  $O_p(1/\sqrt{N})$ . For any finite value of  $g_0$ , we can thus choose  $p$  such that the value on the right-hand side of the above equation converges to positive values in distribution. In practice, for finite values of  $N$ , values of  $p$  that are too small reduce the reproducibility of the numerical results. Thus we use  $p = 0.2$  or  $0.4$  in all simulations, although fixing  $p$  violates Dale's law for small values of  $g_0$ .

With these random synaptic weights, we integrate the model equations using the fourth-order Runge-Kutta algorithm [78] with discrete timesteps of size  $\Delta t = 0.05$  for Eq. (1), and  $\Delta t = 0.02$  for Eq. (51), from *i.i.d.* random values of dynamical variables at  $t = 0$ . We use  $N = 10\,240$  for most of the results, except that we use  $N = 40\,960$  in Figs. 3(a) and 3(b),  $N = 20\,480$  in Figs. 3(c) and 3(d),  $N = 10\,000$  in Figs. 12 and 16, and different values of  $N$  in Figs. 11(c), 11(d), 17, and 18(c), and 18(d). For smaller values of  $N (< 8000)$ , we use  $p = 0.4$  to enhance the stability of the results. Each direct simulation is performed using 16 cores of recent versions of Intel Xeon processors in parallel and takes a few hours—a few days.

### APPENDIX B: EIGENVALUE SPECTRA OF SYNAPTIC WEIGHT MATRICES AND LOCAL STABILITY MATRICES

We show the entire eigenvalue spectra of the synaptic weight matrices of the untuned and finely tuned models in Figs. 13(a) and 13(b). As pointed out in previous studies [60–62], the synaptic weight matrices of the untuned model have configuration-dependent outliers while those of the finely tuned model do not.

Figure 14(b) shows the entire eigenvalue spectrum of the coefficient matrix  $B_{k\ell}^{ij}(t)$  of the following linear variational equation:

$$\frac{d}{dt} u_i^{(k)}(t) = \sum_{\ell, j} B_{k\ell}^{ij}(t) u_j^{(\ell)}(t),$$

$$B_{k\ell}^{ij}(t) \stackrel{\text{def}}{=} -\delta_{ij} \delta_{k\ell} + \sigma_0 J_{k\ell}^{ij} \phi'(h_j^{(\ell)}(t)). \quad (\text{B1})$$

The variational equation (B1) describes how an infinitesimal variation in  $h_i^{(k)}$ , denoted by  $u_i^{(k)}$ , evolves over time around the

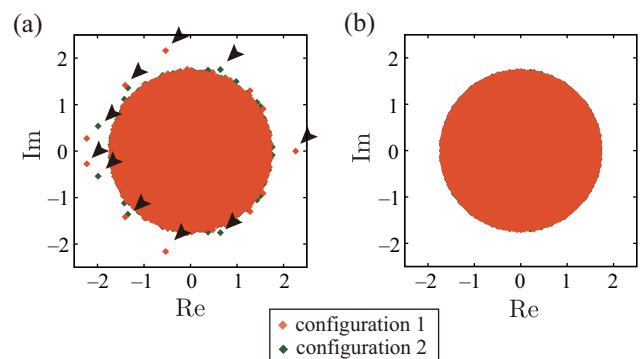


FIG. 13. The entire eigenvalue spectra of synaptic weight matrices for (a) the untuned model and (b) the finely tuned model. The spectra for two different configurations are calculated from networks with 10 240 neurons for each population and plotted in different colors. For both panels, we use  $(\sigma_0, g_0) = (1.2, 4)$ . In the spectra for the untuned model, outlier eigenvalues are observed [arrowheads in (a)], while most of the eigenvalues are distributed over a common disk. Such outlier eigenvalues are not observed for the finely tuned model. In both panels, most of the eigenvalues for configuration 2 in the common disk are hidden behind those for configuration 1, although the distribution in the common disk is quite similar between the two configurations.

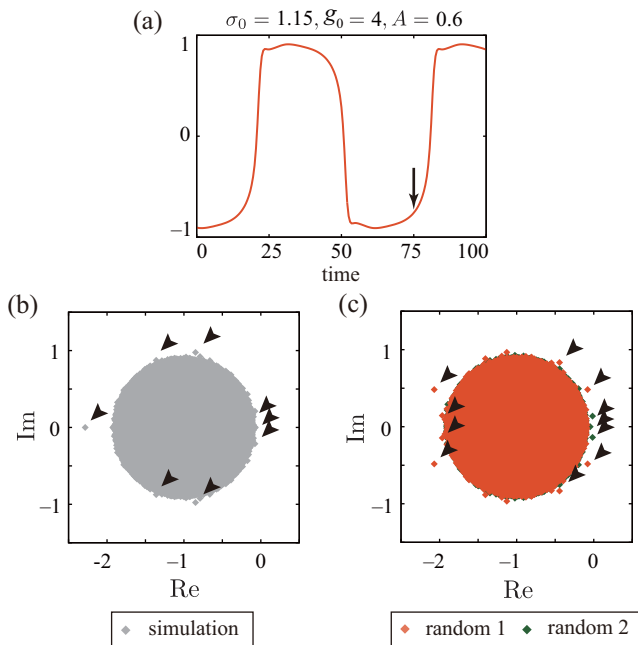


FIG. 14. (a) A simulation in the same setting as Fig. 10(c) is performed and the mean activity of the excitatory population is shown. The values of the model parameters are indicated in the panel. (b) The entire eigenvalue spectrum of the coefficient matrix  $B$  of the linear variational equation (B1), calculated for the indicated time point of the numerical simulation [(a) arrow]. (c) The eigenvalue spectra of  $B$  are also calculated by using two sets of independent random weight configurations and independent random neuronal activities of the same first- and second-order moments as those of the simulated activities used in (b) (indicated as “random 1” and “random 2,” respectively). In these spectra, outlier eigenvalues are observed [arrowheads in (b) and (c)], while most of the eigenvalues are distributed over a common disk. In (c), most of the eigenvalues for “random 2” in the common disk are hidden behind those for “random 1,” although the distributions in the common disk are quite similar in the two settings.

observed dynamics. The spectrum of  $B_{kl}^{ij}(t)$  is calculated with  $h_i^{(k)}(t)$  at an arbitrarily chosen time point  $t$  in the regular orbit of the finely tuned model with external inputs [the indicated point of Fig. 14(a)]. Since the sum of a row of  $B$  is not finely tuned to a fixed value even if  $J$  is the synaptic weight matrix of the finely tuned model, a previous result suggests that  $B$  has configuration-dependent outlier eigenvalues [60]. We observe this in Fig. 14(b). In Fig. 14(c), we also show the eigenvalue spectrum of the same random coefficient matrix,  $B_{kl}^{ij} = J_{kl}^{ij} \phi'(h_j^{(l)})$ , but for which  $J_{kl}^{ij}$  and  $\phi'(h_j^{(l)})$  are generated independently in such a manner that they have the same first- and second-order moments as those used for Fig. 14(b). These spectra have a common disk-form distribution of eigenvalues and outlier eigenvalues at different positions, as expected from the previous study [60]. A previous study [67] also calculated the spectra of the coefficient matrices of the linear variational equations around dynamics of RNNs and showed that they agreed with those estimated from random matrix theory. The results also agreed with the largest Lyapunov exponent calculated from the linear response theory based on

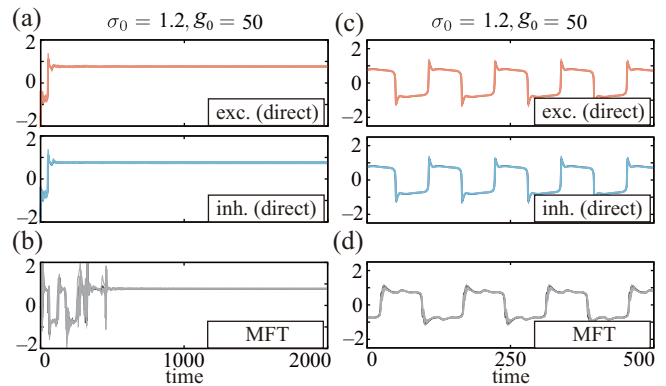


FIG. 15. Static states [(a) and (b)] and coherent oscillations [(c) and (d)] observed in direct simulations and simulations of the mean-field equations for the model with finely tuned synaptic weights. The simulations are conducted for the same condition as Fig. 2, except for the random numbers used for the simulations and the parameter values  $(\sigma_0, g_0) = (1.2, 50)$  used here. Activity patterns of the network are plotted in the same manner as in Fig. 2. These nonchaotic solutions are occasionally observed, depending on the configuration of the random connectivity or on the sequence of random numbers used for solving the mean-field equations.

an MFT. By analogy with this, our stochastic linear response theory derived in Sec. IV B 3, Sec. IV B 4, Appendix H, and Appendix I is expected to allow further quantitative evaluation of the agreement between random matrix theory with outliers and our stochastic MFT. We leave this as a challenge for the future.

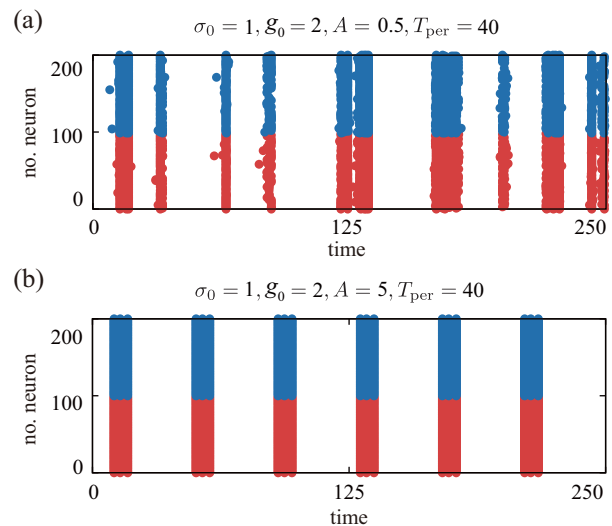


FIG. 16. The raster plots of firing activities of one hundred representative excitatory and inhibitory neurons of the LIF model with sinusoidal external inputs applied to  $\sqrt{N}$  excitatory neurons. The parameter values used for the simulations are indicated above the panels. As we increase the amplitude  $A$  of the sinusoidal inputs, the network undergoes a transition from (a) partially entrained dynamics with irregular firing to (b) coherent dynamics completely entrained to the input. The values of the other model parameters are the same as for Fig. 12, but the synaptic weights are finely tuned.



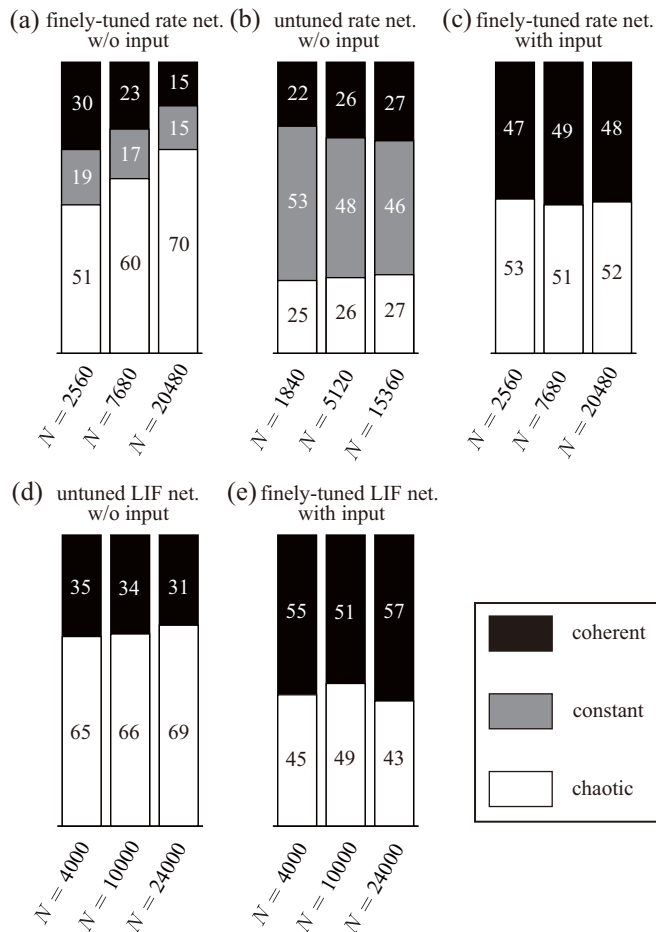


FIG. 17. [(a)–(e)] The numbers of the configurations for which each of the qualitatively different dynamics is observed are counted and summarized in a graph for one hundred direct simulations of (a) the finely tuned rate-neuron network without external inputs, (b) the untuned rate-neuron network without external inputs, (c) the finely tuned rate-neuron network with sinusoidal external inputs, (d) the untuned LIF-neuron network without external inputs, and (e) the finely tuned LIF-neuron network with sinusoidal external inputs, with the indicated number of neurons for each population. For (a)–(e), we use the same settings as for Figs. 15, 6(g), 10(c), 12(c), and 16(b), respectively, except for the random weight configurations.

### APPENDIX C: VARIETY OF NETWORK DYNAMICS FOR DIFFERENT SYSTEM SIZES

Here, we investigate how often we observe each of the qualitatively different solutions in simulations of the network models with different system sizes. We do this by conducting simulations of the models with random weight configurations. We use the same model equation and parameter values as for Figs. 6(g), 10(c), and 12(c), and for occasionally observed fixed-point and limit-cycle solutions for autonomous networks with finely tuned synaptic weights [Fig. 15]. We also examine stimulus-driven dynamics of networks of LIF neurons, by applying sinusoidal inputs to  $\sqrt{N}$  excitatory neurons as described by Eqs. (4) and (47). The input term in these equations are added to the right-hand side of Eq. (51). As mentioned in the main text, the entire network gets entrained to the

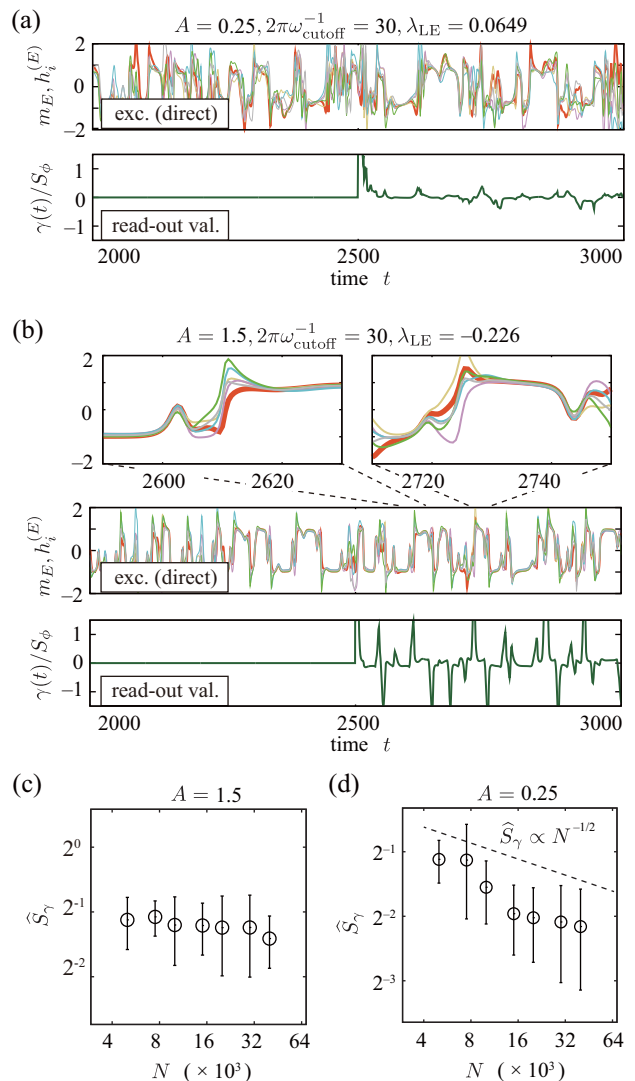


FIG. 18. [(a) and (b)] Typical mean activity of the excitatory population (thick red line) and typical activities of five representative excitatory neurons (thin lines in different colors) in networks with finely tuned synaptic weights for  $(\sigma_0, g_0) = (1.15, 4)$  with the irregular external input described by Eq. (D1) for the indicated parameter values are shown in the upper panels. For (b), two magnified images of the indicated parts are also shown. The largest Lyapunov exponents of the dynamics are also shown above these panels. The values read out from the networks in the same manner as in Figs. 11(a) and 11(b) ( $t_{\text{init}} = 2500$ ) are shown in the lower panels. [(c) and (d)] Normalized standard deviations of the read-out values from networks with the same model settings as for (a) and (b), respectively, except for different population sizes and random weight configurations. The averages of  $\hat{S}_\gamma$  over fifteen different configurations, together with their standard errors, are plotted on logarithmic scales. In (d), a straight line with slope  $-1/2$ , representing  $\hat{S}_\gamma \propto N^{-1/2}$ , is also shown.

sinusoidal input in a configuration-dependent manner, as we increase the amplitude of the input [Fig. 16].

In the model setting for Figs. 6(g), 10(c), 12(c), and 16(b), we find that the frequency with which each type of solution is observed does not depend much on the system size [Figs. 17(b)–17(e)], but it does in the model setting for Fig. 15

[Fig. 17(a)]. These findings are consistent with the results of the stability analysis we perform in Sec. IV B 3, Sec. IV B 4, Appendix H, and Appendix I.

#### APPENDIX D: APPLICATION OF IRREGULAR EXTERNAL INPUTS

A different type of external input of particular interest is an irregular input with no periodicity. As an example of such an input, we use the following filtered noise:

$$\begin{aligned}\tilde{I}(t) &= \sqrt{c_0} A \mathcal{F}^{-1}[\hat{I}_{\text{cutoff}}](t), \\ \hat{I}_{\text{cutoff}}(\omega) &\stackrel{\text{def}}{=} \Theta(\omega_{\text{cutoff}} - |\omega|) \mathcal{F}[I_{\text{WG}}](\omega), \\ c_0 &\stackrel{\text{def}}{=} \int_{\omega_{\text{min}}}^{\omega_{\text{max}}} |\mathcal{F}[I_{\text{WG}}](\omega)|^2 d\omega \bigg/ \int_{\omega_{\text{min}}}^{\omega_{\text{max}}} |\hat{I}_{\text{cutoff}}(\omega)|^2 d\omega,\end{aligned}\tag{D1}$$

where  $\mathcal{F}$  and  $\mathcal{F}^{-1}$  indicate Fourier and inverse-Fourier transforms, respectively. We define white Gaussian noise with unit variance,  $I_{\text{WG}}$ , with  $\langle I_{\text{WG}}(t) I_{\text{WG}}(t - \tau) \rangle = \delta(\tau)$ . The domain of integration is between  $\omega_{\text{min}}$  and  $\omega_{\text{max}}$ , which are given suitably in the discrete Fourier transforms we use for the numerical calculations below [78]. To filter out the high-frequency components, we use the step function  $\Theta(\omega) = 1$  for  $\omega \geq 0$  and  $\Theta(\omega) = 0$  for  $\omega < 0$ .

Examining the properties of the network dynamics resulting from these irregular inputs, we find that the largest Lyapunov exponent decreases and becomes negative as we increase the amplitude  $A$  of the input. By analogy with the results shown in Sec. IV C 3, we hypothesize that neuronal activities are coherent in the dynamics with the negative

largest Lyapunov exponents. In Figs. 18(a) and 18(b), we show typical activity patterns for the networks with the negative and positive largest Lyapunov exponents, given these irregular inputs. Coherence among the neurons is not seen in the activity patterns. To further test the hypothesis, we read out values linearly from these networks according to Eqs. (48) and (49) and examine the scaling of the read-out values in the same manner as for Figs. 11(c) and 11(d). Figures 18(c) and 18(d) show that the normalized standard deviations of the values read out from networks with negative exponents do not depend much on the system size, while those calculated from networks with positive exponents are roughly proportional to  $1/\sqrt{N}$ . This suggests that the networks undergo a transition to coherent dynamics as we increase the amplitude of the irregular external input, even though coherence among the neurons is not obvious from their activity patterns.

#### APPENDIX E: PATH-INTEGRAL REPRESENTATION OF THE DYNAMICS

In this section, we derive the MFT described in Sec. III based on a path-integral representation of dynamics. The approach based on path-integral representations has recently become increasingly popular in the analysis of RNNs and other disordered systems [34,67,79–83]. Our argument and notation follow Ref. [83]. We refer readers to the first two sections of Ref. [83] for precise definitions and notations for the path integral we use below.

We first analyze the autonomous dynamics of the model with finely tuned synaptic weights described in Sec. III A. The moment-generating functional for the dynamics of the present model from an initial condition  $a$  is given by

$$Z[\mathbf{j}] = \lim_{\Delta t \rightarrow +0} \prod_{\alpha=1}^{T/\Delta t} \left\{ \int_{-\infty}^{\infty} dh_{\alpha} \exp(j_{\alpha}^T h_{\alpha} \Delta t) \right\} p(h_1, h_2, \dots, h_{T/\Delta t} | h_0 = a).\tag{E1}$$

In the above, following Ref. [83], we collectively denote  $\{h_i^{(k)}\}_{k,i}$  by  $h$  as a single column vector, and add the subscript  $\alpha$  as the time index. The superscript  $T$  denotes transposition. The probability density over the sample paths is denoted by  $p$ . Hereafter, we sometimes use the same sort of collective notation without mentioning it. The main step in deriving the MFT for the present model is to transform the above functional into an integral with respect to the sample paths for the mean activity. Putting the right-hand side of the model equation equal to  $f(h)$ , and using the following Fourier representation of the Dirac delta functional:

$$\delta(h) = \frac{1}{(2\pi i)^{2N}} \int_{-\infty}^{i\infty} d\tilde{h} \exp(\tilde{h}^T h),\tag{E2}$$

the above generating functional is transformed as follows:

$$\begin{aligned}Z[\mathbf{j}, \tilde{\mathbf{j}}] &= \lim_{D \rightarrow +0} \lim_{\Delta t \rightarrow +0} \prod_{\alpha=1}^{T/\Delta t} \left\{ \int_{-\infty}^{\infty} dh_{\alpha} \right\} \prod_{\alpha'=0}^{T/\Delta t-1} \left\{ \int_{-\infty}^{\infty} \frac{d\tilde{h}_{\alpha'}}{(2\pi i)^{2N}} \right\} \\ &\times \exp \left( \sum_{\alpha''=0}^{T/\Delta t-1} \tilde{h}_{\alpha''}^T \left( h_{\alpha''+1} - h_{\alpha''} - f(h_{\alpha''}) \Delta t - a \delta_{\alpha'',0} + \frac{D}{2} \tilde{h}_{\alpha''} \right) + j_{\alpha''+1}^T h_{\alpha''+1} \Delta t + \tilde{j}_{\alpha''}^T \tilde{h}_{\alpha''} \Delta t \right) \\ &\stackrel{\text{def}}{=} \int \mathcal{D}h \mathcal{D}\tilde{h} \exp \left( \int_{-\infty}^{\infty} \tilde{h}(t)^T (\partial_t h(t) - f(h(t)) - a \delta(t)) + j(t)^T h(t) + \tilde{j}(t)^T \tilde{h}(t) dt \right).\end{aligned}\tag{E3}$$

In the above, we have introduced an auxiliary field  $\tilde{\mathbf{j}}$  for calculating the response function (see Ref. [83]). In the above definition, we adopt the Ito convention and take the noiseless limit in defining the path integrals for the dynamics. In this section, we assume

that the noiseless limit, thermodynamic limit, and stepsize limit all commute with one another. Note that we have

$$Z[\mathbf{j}, \tilde{\mathbf{j}}]_{\mathbf{j}=\mathbf{0}} = 1 \quad (\forall \tilde{\mathbf{j}}), \quad (\text{E4})$$

because this quantity is the limit of integrals of proper probability densities. In the following, we represent inner products with respect to time in the  $L^2$  sense, using a vectorial notation such as

$$\tilde{h}^T h \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} \tilde{h}(t)^T h(t) dt. \quad (\text{E5})$$

Rewriting the above with a concrete form for  $f$  and making the dependence on configurations explicit, we obtain

$$\begin{aligned} Z[\mathbf{j}, \tilde{\mathbf{j}}|J] &= \int \mathcal{D}h \mathcal{D}\tilde{h} \exp \left[ \sum_k S[h^{(k)}, \tilde{h}^{(k)}] - \sum_k \tilde{h}^{(k)T} \left\{ \sum_{\ell} \frac{\sigma_0}{\sqrt{N}} \tilde{\mathcal{J}}_{k\ell} \phi(h^{(\ell)}) + \sum_{\ell} \frac{g_{k\ell}}{\sqrt{N}} M_1 \phi(h^{(\ell)}) \right\} + \sum_k \mathbf{j}^{(k)T} h^{(k)} + \sum_k \tilde{\mathbf{j}}^{(k)T} \tilde{h}^{(k)} \right] \\ &= \int \mathcal{D}\theta \mathcal{D}\tilde{m} \mathcal{D}h \mathcal{D}\tilde{h} \exp \left[ \sum_k S[h^{(k)}, \tilde{h}^{(k)}] - \sum_k \tilde{h}^{(k)T} \sum_{\ell} \frac{\sigma_0}{\sqrt{N}} \tilde{\mathcal{J}}_{k\ell} \phi(h^{(\ell)}) - \sum_k \tilde{m}_k^T \sum_{\ell} \frac{g_{k\ell}}{\sqrt{N}} 1^T (\phi(h^{(\ell)}) - \bar{\phi}_{\ell} \mathbf{1}) \right. \\ &\quad \left. + \sum_k (\tilde{m}_k - 1^T \tilde{h}^{(k)})^T \theta_k + \sum_k \mathbf{j}^{(k)T} h^{(k)} + \sum_k \tilde{\mathbf{j}}^{(k)T} \tilde{h}^{(k)} \right]. \end{aligned} \quad (\text{E6})$$

Here, we define the action:

$$S[h^{(k)}, \tilde{h}^{(k)}] \stackrel{\text{def}}{=} \tilde{h}^{(k)T} (\partial_t + 1) h^{(k)}. \quad (\text{E7})$$

We define the vector and matrix for which all entries are unity at each timestep as  $\mathbf{1}$  and  $M_1$ , respectively. The left action of  $M_1$  is thus given by

$$M_1 : \{ \phi(h_j^{(\ell)}(t)) \}_{\ell, j, t} \rightarrow \left\{ \sum_{j'} \phi(h_{j'}^{(\ell)}(t)) \right\}_{\ell, j, t}. \quad (\text{E8})$$

The column vector which consists of  $\phi(h_i^{(k)})$  has been denoted by  $\phi(h^{(k)})$ . From the first line to the second line of Eq. (E6), we have inserted the Dirac delta functional equating  $\tilde{m}_k$  and  $1^T \tilde{h}^{(k)}$  for each  $t$ , where  $1^T$  denotes the following operation:

$$1^T : \{ \tilde{h}_i^{(k)}(t) \}_{i, t} \rightarrow \left\{ \sum_i \tilde{h}_i^{(k)}(t) \right\}. \quad (\text{E9})$$

Note that, under the condition in Eq. (3), the insertion of any value of  $\bar{\phi}_{\ell}$  does not affect the value of the integrand in the last line of Eq. (E6) as long as  $\bar{\phi}_E = \bar{\phi}_I$  and hence  $\sum_{\ell=E, I} g_{k\ell} \bar{\phi}_{\ell} = 0$  hold. We will determine the precise value of  $\bar{\phi}_k$  below.

Next, we take the configurational average of the integrand of the above equation. We consider unit Gaussian measures for  $\mathcal{J}_{k\ell}^{ij}$  denoted by  $\mathcal{N}(\mathcal{J}_{k\ell}^{ij})$ . Other distributions for these random variables can be analyzed in essentially the same manner, where distributions with zero mean and unit variance give the same result (to see this, expand the exponential in the integrand in terms of small values). Focusing on the term involving  $\tilde{\mathcal{J}}_{k\ell}^{ij}$ , we have

$$\int d\mathcal{N}(\mathcal{J}_{k\ell}^{ij}) \exp \left( -\tilde{h}_i^{(k)T} \frac{\sigma_0}{\sqrt{N}} \mathcal{J}_{k\ell}^{ij} \phi(h_j^{(\ell)}) + \sum_{j'} \tilde{h}_i^{(k)T} \frac{\sigma_0}{\sqrt{N}} \frac{\mathcal{J}_{k\ell}^{ij}}{N} \phi(h_{j'}^{(\ell)}) \right) = \exp \left[ \frac{\sigma_0^2}{2N} \left\{ \int \tilde{h}_i^{(k)} \phi(h_j^{(\ell)}) - \sum_{j'} \frac{1}{N} \tilde{h}_i^{(k)} \phi(h_{j'}^{(\ell)}) dt \right\}^2 \right]. \quad (\text{E10})$$

Taking the product of the right-hand side of the above equation over  $k$ ,  $\ell$ ,  $i$ , and  $j$ , we obtain the part of the integrand involving  $\{ \mathcal{J}_{k\ell}^{ij} \}_{k, \ell, i, j}$  as

$$\exp \left[ \sum_{k, \ell} \frac{\sigma_0^2}{2N} \left\{ \sum_{i, j} \int ds dt \tilde{h}_i^{(k)}(s) \tilde{h}_i^{(k)}(t) (\phi(h_j^{(\ell)}(s)) - \bar{\phi}_{\ell}(s)) (\phi(h_j^{(\ell)}(t)) - \bar{\phi}_{\ell}(t)) \right\} \right]. \quad (\text{E11})$$

Here,  $\check{\phi}_\ell$  denotes the average of  $\phi(h_j^{(\ell)})$  over the  $\ell$  population. We now introduce an auxiliary field by inserting the following Dirac delta functionals:

$$\begin{aligned} & \delta \left( -N\mathcal{Q}_{\ell,1}(s,t) + \sum_j (\phi(h_j^{(\ell)}(s)) - \check{\phi}_\ell(s))(\phi(h_j^{(\ell)}(t)) - \check{\phi}_\ell(t)) \right) \\ &= \int \mathcal{D}\mathcal{Q}_{\ell,2} \exp \left[ \int ds dt \mathcal{Q}_{\ell,2}(s,t) \left\{ -N\mathcal{Q}_{\ell,1}(s,t) + \sum_j (\phi(h_j^{(\ell)}(s)) - \check{\phi}_\ell(s))(\phi(h_j^{(\ell)}(t)) - \check{\phi}_\ell(t)) \right\} \right], \end{aligned} \quad (\text{E12})$$

$$\delta \left( -N\check{\phi}_\ell(t) + \sum_j \phi(h_j^{(\ell)}(t)) \right) = \int \mathcal{D}\tilde{\psi}_\ell \exp \left[ \tilde{\psi}_\ell^T \left\{ -N\check{\phi}_\ell + \sum_j \phi(h_j^{(\ell)}) \right\} \right]. \quad (\text{E13})$$

Following the convention in Ref. [83], we regard  $\mathcal{Q}_{\ell,i}$  ( $i = 1, 2$ ) as matrices and use the following notation:

$$\mathcal{Q}_{\ell,1}^T \mathcal{Q}_{\ell,2} = \int ds dt \mathcal{Q}_{\ell,1}(s,t) \mathcal{Q}_{\ell,2}(s,t), \quad \tilde{h}_i^{(k)T} \mathcal{Q}_{\ell,1} \tilde{h}_i^{(k)} = \int ds dt \tilde{h}_i^{(k)}(s) \mathcal{Q}_{\ell,1}(s,t) \tilde{h}_i^{(k)}(t). \quad (\text{E14})$$

Using Eqs. (E11)–(E13), we take the average of the moment-generating functional over the probability distribution  $P_J(J)$  of  $J_{k\ell}^{ij}$  as follows:

$$\begin{aligned} \bar{Z}[\tilde{j}, \tilde{j}] &= \int dP_J(J) Z[\tilde{j}, \tilde{j}|J] \\ &= \int \mathcal{D}\theta \mathcal{D}\tilde{m} \mathcal{D}\mathcal{Q}_1 \mathcal{D}\mathcal{Q}_2 \mathcal{D}\tilde{\psi} \mathcal{D}\check{\phi} \mathcal{D}h \mathcal{D}\tilde{h} \exp \left( -N \sum_\ell \mathcal{Q}_{\ell,1}^T \mathcal{Q}_{\ell,2} + \sum_{k,i} S[h_i^{(k)}, \tilde{h}_i^{(k)}] \right. \\ &\quad + \sum_{k,\ell,i} \frac{\sigma_0^2}{2} \tilde{h}_i^{(k)T} \mathcal{Q}_{\ell,1} \tilde{h}_i^{(k)} - \sum_{k,i} \tilde{h}_i^{(k)T} \theta_k + \sum_{\ell,j} \Delta\phi_j^{(\ell)T} \mathcal{Q}_{\ell,2} \Delta\phi_j^{(\ell)} + \sum_k \mathbf{j}^{(k)T} h^{(k)} + \sum_k \tilde{\mathbf{j}}^{(k)T} \tilde{h}^{(k)} \\ &\quad \left. + \sum_{\ell,j} \tilde{\psi}_\ell^T \phi(h_j^{(\ell)}) - N \sum_\ell \tilde{\psi}_\ell^T \check{\phi}_\ell \right) \exp \left( - \sum_k \tilde{m}_k^T \sum_\ell \frac{g_{k\ell}}{\sqrt{N}} \mathbf{1}^T (\phi(h^{(\ell)}) - \bar{\phi}_\ell \mathbf{1}) + \sum_k \tilde{m}_k^T \theta_k \right). \end{aligned} \quad (\text{E15})$$

Here,  $\phi(h_j^{(\ell)}) - \check{\phi}_\ell$  is denoted by  $\Delta\phi_j^{(\ell)}$ . Using this representation, we note that the argument of the first exponential takes the form of an independent interaction between each neuron and the auxiliary fields if the values of  $\theta_k$ ,  $\check{\phi}_\ell$ ,  $\tilde{\psi}_\ell$ ,  $\mathcal{Q}_{\ell,1}$ , and  $\mathcal{Q}_{\ell,2}$  are fixed.

In the above equation, we notice that, for fixed sample paths for  $\theta$  and  $\tilde{m}$ , the integral of the first exponential function with respect to  $\mathcal{Q}_1$ ,  $\mathcal{Q}_2$ ,  $\tilde{\psi}$ ,  $\check{\phi}$ ,  $h$ , and  $\tilde{h}$  gives the generating functional for the following dynamics of a fictitious RNN with a uniform external input  $\theta_k(t)$ :

$$\frac{d}{dt} h_i^{(k)}(t) = -h_i^{(k)}(t) + \sigma_0 \sum_{j,\ell} \frac{\tilde{J}_{k\ell}^{ij}}{\sqrt{N}} \phi(h_j^{(\ell)}(t)) + \theta_k(t). \quad (\text{E16})$$

The dynamics of this fictitious RNN are no longer under the effects of balanced excitation and inhibition, and therefore solved by a conventional MFT. We first rewrite the averaged generating functional for these dynamics as

$$\begin{aligned} \bar{Z}^*[\tilde{j}, \tilde{j}|\theta] &= \int \mathcal{D}\mathcal{Q}_1 \mathcal{D}\mathcal{Q}_2 \mathcal{D}\tilde{\psi} \mathcal{D}\check{\phi} \mathcal{D}h \mathcal{D}\tilde{h} \exp \left( -N \sum_\ell \mathcal{Q}_{\ell,1}^T \mathcal{Q}_{\ell,2} + \sum_{k,i} S[h_i^{(k)}, \tilde{h}_i^{(k)}] + \sum_{k,\ell,i} \frac{\sigma_0^2}{2} \tilde{h}_i^{(k)T} \mathcal{Q}_{\ell,1} \tilde{h}_i^{(k)} - \sum_{k,i} \tilde{h}_i^{(k)T} \theta_k \right. \\ &\quad \left. + \sum_{\ell,j} \Delta\phi_j^{(\ell)T} \mathcal{Q}_{\ell,2} \Delta\phi_j^{(\ell)} + \sum_k \mathbf{j}^{(k)T} h^{(k)} + \sum_k \tilde{\mathbf{j}}^{(k)T} \tilde{h}^{(k)} + \sum_{\ell,j} \tilde{\psi}_\ell^T \phi(h_j^{(\ell)}) - N \sum_\ell \tilde{\psi}_\ell^T \check{\phi}_\ell \right). \end{aligned} \quad (\text{E17})$$



The method for obtaining the dynamics described by the above moment-generating functional has been developed in previous studies [32,83]. First, the above functional is rewritten as

$$\begin{aligned} \bar{Z}^*[j, \tilde{j}|\theta] &= \int \mathcal{D}Q_1 \mathcal{D}Q_2 \mathcal{D}\tilde{\psi} \mathcal{D}\check{\phi} \exp \left( -N \sum_{\ell} Q_{\ell,1}^T Q_{\ell,2} + N \sum_k \ln \mathcal{Z}_k[Q_1, Q_2, \tilde{\psi}, \check{\phi}, j, \tilde{j}|\theta] - N \sum_{\ell} \tilde{\psi}_{\ell}^T \check{\phi}_{\ell} \right), \\ \mathcal{Z}_k[Q_1, Q_2, \tilde{\psi}, \check{\phi}, j, \tilde{j}|\theta] &\stackrel{\text{def}}{=} \int \mathcal{D}h \mathcal{D}\tilde{h} \mathcal{D}P_{j, \tilde{j}}(1, \tilde{1}) \exp \left( \tilde{h}^{(k)T} (\partial_t + 1) h^{(k)} + \sum_{\ell} \frac{\sigma_0^2}{2} \tilde{h}^{(k)T} Q_{\ell,1} \tilde{h}^{(k)} - \tilde{h}^{(k)T} \theta_k + 1^{(k)T} h^{(k)} + \tilde{1}^{(k)T} \tilde{h}^{(k)} \right. \\ &\quad \left. + (\phi(h^{(k)}) - \check{\phi}_k)^T Q_{k,2} (\phi(h^{(k)}) - \check{\phi}_k) + \tilde{\psi}_k^T \phi(h^{(k)}) \right). \end{aligned} \quad (\text{E18})$$

In the second equation,  $h^{(k)}$  and  $\tilde{h}^{(k)}$  are no longer collections of variables corresponding to individual neurons but instead are one-dimensional variables for a representative neuron feeling the mean fields. Similarly,  $1^{(k)}$  and  $\tilde{1}^{(k)}$  are one-dimensional variables that take values randomly drawn from the measure  $P_{j, \tilde{j}}$  corresponding to  $j$  and  $\tilde{j}$ :

$$P_{j, \tilde{j}} \stackrel{\text{def}}{=} \otimes_k \left( \frac{1}{N} \sum_i \delta_{j_i^{(k)}} \delta_{\tilde{j}_i^{(k)}} \right). \quad (\text{E19})$$

Applying the saddle-point method, we find that the entire probability mass of the path integral of the dynamics at  $j = \tilde{j} = 0$  concentrates at the values of  $\check{\phi}_{\ell}$ ,  $\tilde{\psi}_{\ell}$ ,  $Q_{\ell,1}$ , and  $Q_{\ell,2}$  that maximize the integrand of the first of Eq. (E18). Taking the functional derivatives and examining the stationarity conditions, we obtain these optimal values as

$$\begin{aligned} Q_{\ell,2}^* &= \sum_k \frac{\sigma_0^2}{2} \langle \tilde{h}^{(k)} \tilde{h}^{(k)T} \rangle = 0, \quad Q_{\ell,1}^* = \langle (\phi(h^{(\ell)}) - \check{\phi}_{\ell}) (\phi(h^{(\ell)}) - \check{\phi}_{\ell})^T \rangle, \\ \tilde{\psi}_{\ell}^* &= -2 \langle Q_{\ell,2}^* (\phi(h^{(\ell)}) - \check{\phi}_{\ell}) \rangle = 0, \quad \check{\phi}_{\ell}^* = \langle \phi(h^{(\ell)}) \rangle. \end{aligned} \quad (\text{E20})$$

Here, the bracket denotes the expected value of the argument for all possible sample paths weighted according to the path-integral representation of the fictitious dynamics. The zeros for  $Q_{\ell,2}^*$  and  $\tilde{\psi}_{\ell}^*$  are obtained by taking directional functional derivative with respect to  $\tilde{j}_i^{(k)}(t) = \alpha^{(k)}(t)$  at  $\alpha = 0$  and  $j = 0$  and using the normalization of the probability density (see also the arguments in Refs. [83,84]). We actually have

$$\left. \frac{\delta \bar{Z}^*[0, \{\tilde{j}_i^{(k)}(t) = \alpha^{(k)}(t)\}_{k,i,t}|\theta]}{\delta \alpha^{(k)} \delta \alpha^{(k)T}} \right|_{\alpha=0} = N \langle \tilde{h}^{(k)} \tilde{h}^{(k)T} \rangle = 0, \quad \left. \frac{\delta \bar{Z}^*[0, \{\tilde{j}_i^{(k)}(t) = \alpha^{(k)}(t)\}_{k,i,t}|\theta]}{\delta \alpha^{(k)}} \right|_{\alpha=0} = N \langle \tilde{h}^{(k)} \rangle = 0, \quad (\text{E21})$$

because we have  $\bar{Z}^*[0, \tilde{j}|\theta] = 1$  for any  $\tilde{j}$ , similarly to Eq. (E4). Then, following the argument in Ref. [83], we can regard the dynamics of individual neurons, as described by the second of Eq. (E18), as linear dynamics driven by a set of *i.i.d.* Gaussian noise with correlation  $Q_{\ell,1}^*$  and drift term  $\theta_k$ . Thus we obtain

$$\begin{aligned} (1 + \partial_t)(1 + \partial_s) \langle \delta h^{(k)}(t) \delta h^{(k)}(s) \rangle &= \sum_{\ell} \sigma_0^2 \langle (\phi(h^{(\ell)}(t)) - \langle \phi(h^{(\ell)}(t)) \rangle) (\phi(h^{(\ell)}(s)) - \langle \phi(h^{(\ell)}(s)) \rangle) \rangle, \\ (1 + \partial_t) \langle h^{(k)}(t) \rangle &= \theta_k(t). \end{aligned} \quad (\text{E22})$$

This is the solution for the fictitious RNN described by Eq. (E16).

We now return back to the generating functional for the original dynamics in Eq. (E15). Note that the integrand of Eq. (E17) gives a proper probability density over the sample paths for the fields. Then, because of the independence among neurons for fixed values of  $\theta_k$  and  $\tilde{m}_k$ , the integrand in Eq. (E15) can be rearranged into the following form with the aid of the central limit theorem:

$$\bar{Z}[0, 0] = \int \mathcal{D}\theta \mathcal{D}\tilde{m} \mathcal{D}\eta \exp \left( -\frac{1}{2} \sum_{\ell} \eta_{\ell}^T Q_{\ell,1}^{*-1} \eta_{\ell} - \sum_{\ell} F_{\ell} + \sum_k \tilde{m}_k^T (\theta_k - \sum_{\ell} g_{k\ell} \eta_{\ell}) - \text{const.} \right). \quad (\text{E23})$$

Here, we have set the values of  $\bar{\phi}_E = \bar{\phi}_I$  to  $\langle \phi(h_j^{(\ell)}) \rangle$ , the configuration and population average of  $\phi(h_j^{(\ell)})$  over the fictitious dynamics for each sample path for  $\theta_k$ . We have defined  $\eta_{\ell} \stackrel{\text{def}}{=} \frac{1}{\sqrt{N}} 1^T (\phi(h^{(\ell)}) - \bar{\phi}_{\ell} 1)$ . We determine the above normalization term,  $F_{\ell}$ , below. Noting that the auxiliary field  $\tilde{m}_k$  is imposing equality between  $\theta_k$  and  $\sum_{\ell} g_{k\ell} \eta_{\ell}$ , the above equation determines the probabilities with which the sample paths for  $\eta_{\ell}$  and  $\theta_k$  are realized, and the density for each value of  $\eta_{\ell}$  is given by

$$p_{(\eta)}(\{\eta_{\ell}\}) \propto \exp \left( -\frac{1}{2} \sum_{\ell} \eta_{\ell}^T Q_{\ell,1}^{*-1} \eta_{\ell} - \sum_{\ell} F_{\ell} \right). \quad (\text{E24})$$

Note that this density is at most  $O(1)$  in terms of the neuron count, and therefore, does not affect the values of  $Q_{\ell,1}$  and  $Q_{\ell,2}$ , which have most of the probability mass.

Because of the condition,  $g_{EE} = g_{IE} = g_0$  and  $g_{EI} = g_{II} = -g_0$ , we have  $\theta_E = \theta_I \stackrel{\text{def}}{=} \theta$  from  $\theta_k = \sum_{\ell} g_{k\ell} \eta_{\ell}$ . Since the right-hand sides of Eq. (E22) take the same values for  $k = E, I$ , we have

$$\langle \delta h^{(E)}(t) \delta h^{(E)}(s) \rangle = \langle \delta h^{(I)}(t) \delta h^{(I)}(s) \rangle \stackrel{\text{def}}{=} D(t, s), \quad (\text{E25})$$

$$\langle h^{(E)}(t) \rangle = \langle h^{(I)}(t) \rangle \stackrel{\text{def}}{=} m(t), \quad (\text{E26})$$

$$\langle (\phi(h^{(E)}(t)) - \langle \phi(h^{(E)}(t)) \rangle) (\phi(h^{(E)}(s)) - \langle \phi(h^{(E)}(s)) \rangle) \rangle = \langle (\phi(h^{(I)}(t)) - \langle \phi(h^{(I)}(t)) \rangle) (\phi(h^{(I)}(s)) - \langle \phi(h^{(I)}(s)) \rangle) \rangle \stackrel{\text{def}}{=} C(t, s). \quad (\text{E27})$$

We have  $Q_{E,1}^* = Q_{I,1}^* = C$  from Eq. (E20), and hence  $F_E = F_I \stackrel{\text{def}}{=} F$ . Defining  $\eta = \frac{\sqrt{2}}{2}(\eta_E - \eta_I)$  and  $\bar{\eta} = \frac{\sqrt{2}}{2}(\eta_E + \eta_I)$  and combining these with Eqs. (E22) and (E24) and the equality,  $\theta_E = \theta_I = g_0(\eta_E - \eta_I)$ , we obtain the realization probability of  $m(t)$  given by

$$(1 + \partial_t)m(t) = \sqrt{2}g_0\eta(t), \quad p_{\eta}(\eta) \propto \exp\left(-\frac{1}{2}\eta^T C \eta - \frac{1}{2}\bar{\eta}^T C \bar{\eta} - 2F - \text{const.}\right), \quad (\text{E28})$$

which is equivalent to Eqs. (17) and (19). We also obtain the dynamical equation for the correlation matrix for the microscopic fluctuations from Eqs. (E22), (E25), and (E27) as

$$(1 + \partial_t)(1 + \partial_s)D(t, s) = 2\sigma_0^2 C(t, s), \quad (\text{E29})$$

which is identical to Eq. (15).

Finally, we identify the normalization term  $F$ . Let us consider the following representation,

$$\eta = H\xi, \quad HH^T = C. \quad (\text{E30})$$

If  $H$  is independent of  $\eta$ , this transformation, together with the probability density in Eq. (E28), yields *i.i.d.* unit Gaussian variables  $\xi$ . In this case, the normalization term is given by  $\exp(-F) = \exp(-\frac{1}{2} \ln |C|) = |C|^{-1/2}$ . This is the Jacobian accompanying the transformation of probability densities over  $\eta$  and  $\xi$ . However, in the present setting,  $H$  depends on  $\eta$ . For this case, taking the nonlinear deformation of the coordinates spanned by  $\xi$  into account, we expect that the normalization term  $F$  is given by the Jacobian  $\ln |H + \frac{\partial H}{\partial \xi} \xi|$  and hence that the transformation  $\eta = H\xi$  still gives unit Gaussian variables for  $\xi$ . Here, the Jacobian matrix is rearranged as

$$H_{ij} + \sum_{\alpha} \frac{\partial H_{i\alpha}}{\partial \xi_j} \xi_{\alpha} = H_{ij} + \sum_{\alpha, \beta} \frac{\partial H_{i\alpha}}{\partial \eta_{\beta}} \frac{\partial \eta_{\beta}}{\partial \xi_j} (H^{-1}\eta)_{\alpha} = \frac{\partial \eta_i}{\partial \xi_j} \Rightarrow \frac{\partial \eta}{\partial \xi} = \left(1 - \frac{\partial H}{\partial \eta} H^{-1}\eta\right)^{-1} H, \quad (\text{E31})$$

which yields Eq. (18). Although rigorous treatment of this point encounters the mathematical difficulty in rigorously dealing with path-integrals, the validity of the above normalization term is supported by the fact that it yields the proper normalization of the path integral, which is consistent with Eq. (E4). This normalization needs to be kept in mind as we perform perturbative expansion below.

Below, we describe the outline of how the above theory is modified for the case with external input to  $\sqrt{N}$  excitatory neurons and for the case with untuned synaptic weights. For the case with  $\tilde{I}(t) \neq 0$  in Eq. (4), the following  $O(\sqrt{N})$  correction is introduced to Eq. (E18):

$$\begin{aligned} \bar{Z}^*[j, \tilde{j}|\theta] &= \int \mathcal{D}Q_1 \mathcal{D}Q_2 \mathcal{D}\tilde{\psi} \mathcal{D}\check{\phi} \exp\left(-N \sum_{\ell} Q_{\ell,1}^T Q_{\ell,2} + N \sum_k \ln \mathcal{Z}_k[Q_1, Q_2, \tilde{\psi}, \check{\phi}, j, \tilde{j}|\theta] \right. \\ &\quad \left. + \sqrt{N}(\ln \tilde{\mathcal{Z}}_E[Q_1, Q_2, \tilde{\psi}, \check{\phi}, j, \tilde{j}|\theta] - \ln \mathcal{Z}_E[Q_1, Q_2, \tilde{\psi}, \check{\phi}, j, \tilde{j}|\theta]) - N \sum_{\ell} \tilde{\psi}_{\ell}^T \check{\phi}_{\ell}\right), \\ \tilde{\mathcal{Z}}_E[Q_1, Q_2, \tilde{\psi}, \check{\phi}, j, \tilde{j}|\theta] &\stackrel{\text{def}}{=} \int \mathcal{D}h \mathcal{D}\tilde{h} \mathcal{D}P_{j, \tilde{j}}(\mathbb{1}, \tilde{\mathbb{1}}) \exp\left(\tilde{h}^{(E)T} (\partial_t + 1)h^{(E)} + \sum_{\ell} \frac{\sigma_0^2}{2} \tilde{h}^{(E)T} Q_{\ell,1} \tilde{h}^{(E)} - \tilde{h}^{(E)T} \theta_E - \tilde{I}^T \tilde{h}^{(E)} \right. \\ &\quad \left. + \mathbb{1}^{(E)T} h^{(E)} + \tilde{\mathbb{1}}^{(E)T} \tilde{h}^{(E)} + (\phi(h^{(E)}) - \check{\phi}_E)^T Q_{E,2} (\phi(h^{(E)}) - \check{\phi}_E) + \tilde{\psi}_E^T \phi(h^{(E)})\right). \quad (\text{E32}) \end{aligned}$$

In the subsequent saddle point method, however, we notice that the  $O(\sqrt{N})$  correction does not affect the values of  $Q_{\ell,j}^*$ ,  $\tilde{\psi}_\ell^*$ , and  $\tilde{\phi}_\ell^*$  in the leading order. Then, the neurons feel the same Gaussian fields as those described by Eqs. (E20). Then, the second exponential in Eq. (E15) is modified as

$$\exp \left[ - \sum_k \tilde{m}_k^T \left\{ \sum_\ell \frac{g_{k\ell}}{\sqrt{N}} 1_{\mathcal{S}}^T (\phi(h^{(\ell)}) - \bar{\phi}_\ell 1) + \frac{g_0}{\sqrt{N}} 1_{\mathcal{S}}^T (\phi(h^{(\ell)}) - \bar{\phi}_\ell) - \theta_k \right\} \right], \quad (\text{E33})$$

with  $\bar{\phi}_\ell = \langle \phi(h_j^{(\ell)}) \rangle_{j \notin \mathcal{S}}$ . In the above equation,  $1_{\mathcal{S}}$  ( $1_{\setminus \mathcal{S}}$ ) denotes a vector whose elements are one (zero) for indices belonging to  $\mathcal{S}$ , and zero (one) otherwise. We easily find that the second term in the above exponential gives the driving-force term,  $\tilde{\phi}$ , for the mean dynamics [Eq. (29)]. We also find that the exclusion of the sum of the  $\sqrt{N}$  random values in the first term followed by the division by  $\sqrt{N}$  does not affect the values of the driving-force term  $\eta$  in the leading order. Thus, with  $\eta$  given by Eq. (E28), we obtain the mean-field equations for the stimulus-driven dynamics as presented in Sec. III D.

For the untuned model, we easily see that the averaging with respect to the untuned random synaptic weights leads to the replacement of  $\Delta \phi_j^{(\ell)T} Q_{\ell,2} \Delta \phi_j^{(\ell)}$  in Eq. (E15) by  $\phi(h_j^{(\ell)})^T Q_{\ell,2} \phi(h_j^{(\ell)})$ . By a straightforward application of the same argument as above, we obtain the results presented in Sec. III C.

#### APPENDIX F: EFFICIENT METHOD FOR SOLVING THE MEAN-FIELD EQUATIONS

In this section, we show how the stochastic mean-field equations are numerically solved by recursively updating the statistics characterizing the microscopic and macroscopic dynamics of the system. In the main text and Appendix E, we found that self-consistent equations (14) and (15) determine the time evolution of the statistics of the microscopic fluctuations. These equations are rewritten as

$$(1 + \partial_{t_1})(1 + \partial_{t_2})D(t_1, t_2) = 2\sigma_0^2 C(t_1, t_2). \quad (\text{F1})$$

Dealing with the above equation requires us to retain large matrices. To reduce the computational cost, we recursively update matrices of smaller size by using the following auxiliary matrix that retains the effect of the boundary conditions:

$$R \stackrel{\text{def}}{=} 2\sigma_0^2(1 + \partial_{t_1})^{-1}C, \quad (\text{F2})$$

which means  $R$  is the solution of the ordinary differential equation,  $(1 + \partial_{t_1})R = 2\sigma_0^2 C$ , from suitable initial conditions, and the initial conditions do not affect the values of  $A$  at time indices distant from those of the initial conditions because of the decay with the unit decay constant. For the dynamics of the untuned model, matrix  $C$  should be suitably replaced by  $\tilde{C} = C + \bar{\phi}\bar{\phi}^T$ .

Suppose that we have values of  $m(t_1)$ ,  $\eta(t_1)$ ,  $\bar{\phi}(t_1)$ ,  $C(t_1, t_2)$ ,  $D(t_1, t_2)$ , and  $R(t_1, t_3)$  for timesteps  $t \geq t_1$ ,  $t_2 \geq t - T_1$ , and  $t - \Delta t \geq t_3 \geq t - T_1$ . For the model with external inputs, also suppose that we have values of  $\tilde{\phi}(t_1)$ . The length of the time interval  $T_1$  should not be too large. In what follows, we obtain the values of these vectors and matrices for a time index one step beyond the currently available entries.

We first update the value of  $m$ , which is obtained from the values of  $\eta(t)$ ,  $\bar{\phi}(t)$ , and  $\tilde{\phi}(t)$  by applying the Euler method to Eq. (19) or (29).

Next, we update  $R$ . The values of  $R(t + \Delta t, t_3)$  for  $t - \Delta t \geq t_3 \geq t - T_1$  are obtained by solving Eq. (F2) from the initial values,  $R(t, t_3)$ , by using the Euler method. For the values of  $R(t_1, t)$ , we need to solve Eq. (F2) from the unknown initial value,  $R(t - T_1, t)$ . We obtain the initial value by applying the discrete Fourier transform to Eq. (F2):

$$(1 + i\omega)\widehat{R}(\omega, t) = 2\sigma_0^2\widehat{C}(\omega, t). \quad (\text{F3})$$

This implicitly assumes that the regularity of the correlation function  $2\sigma_0^2 C$  is captured by the discrete Fourier transform. Actually, if the microscopic dynamics of the model are chaotic and have a relatively short correlation time, then,  $C(t_1, t) \approx 0$  for  $t_1 \leq t - T_1$  holds and we can use the initial condition,  $R(t - T_1, t) \approx 0$ , which is given also by Eq. (F3). If the entire dynamics have periodic components with a relatively short period, the Fourier transform detects the corresponding peak of the correlation function in the frequency domain and Eq. (F3) is expected to provide an accurate initial condition. Either one of these two situations is almost always the case if we use a sufficiently large value of  $T_1$ . The exception among the cases we have investigated in this study is one in which the external input is irregular and the network dynamics retain finite irregularly varying correlation over a very long time. In this case, the above two assumptions do not hold and we do not attempt to obtain a numerical solution from the MFT.

We use the initial value of the solution obtained by solving Eq. (F3) and then applying the inverse discrete Fourier transform. Note that, although the discrete Fourier transform captures decaying and regular patterns in the autocorrelation function, it does not necessarily provide a good approximation to the solution over the entire time domain. Thus we solve the ordinary differential equation (F2) again from the initial values thus obtained, not directly using the solution obtained from the inverse discrete Fourier transform.

Once we obtain the updated values of  $R$ , those of  $D$  are obtained straightforwardly with the Euler method by noting that Eqs. (F1) and (F2) yield

$$(1 + \partial_{t_2})D(s, t_2) = R(s, t_2). \quad (\text{F4})$$

We are not concerned about the initial conditions for this equation; for the values of  $D(s, t + \Delta t)$  with  $s \leq t$ , we can use the values of  $D(s, t)$  as initial conditions. The value of  $D(t + \Delta t, t + \Delta t)$  is almost independent of the initial value,  $D(t + \Delta t, t - T_1)$ , since the variation at time  $t_2 = t - T_1$  converges exponentially with unit time constant.

Next, we obtain the updated values of  $C(t + \Delta t, s)$ ,  $\bar{\phi}(t + \Delta t)$ , and  $\tilde{\phi}(t + \Delta t)$  from  $D$  and  $m$ . They are computed by using Eqs. (11), (14), and (29), which can be written more precisely as

$$\begin{aligned} C(t_1, t_2) &= \langle (\phi(z_1) - \bar{\phi}(t_1))(\phi(z_2) - \bar{\phi}(t_2)) \rangle \\ &= \int d\mathcal{N}(w) \int d\mathcal{N}(y_1) \left\{ \phi \left( D_{11}^{1/2} \left( \sqrt{1 - \frac{|D_{12}|}{(D_{11}D_{22})^{1/2}}} y_1 + \text{sgn} D_{12} \sqrt{\frac{|D_{12}|}{(D_{11}D_{22})^{1/2}}} w \right) + m_1 \right) - \bar{\phi}(t_1) \right\} \\ &\quad \times \int d\mathcal{N}(y_2) \left\{ \phi \left( D_{22}^{1/2} \left( \sqrt{1 - \frac{|D_{12}|}{(D_{11}D_{22})^{1/2}}} y_2 + \sqrt{\frac{|D_{12}|}{(D_{11}D_{22})^{1/2}}} w \right) + m_2 \right) - \bar{\phi}(t_2) \right\}, \end{aligned} \quad (\text{F5})$$

$$\bar{\phi}(t_1) = \langle \phi(z_1) \rangle = \int d\mathcal{N}(w) \int d\mathcal{N}(y_1) \phi \left( D_{11}^{1/2} \left( \sqrt{1 - \frac{|D_{12}|}{(D_{11}D_{22})^{1/2}}} y_1 + \text{sgn} D_{12} \sqrt{\frac{|D_{12}|}{(D_{11}D_{22})^{1/2}}} w \right) + m_1 \right), \quad (\text{F6})$$

$$\begin{aligned} \tilde{\phi}(t_1) + \bar{\phi}(t_1) &= \langle \phi(z_1 + \delta v(t_1)) \rangle \\ &= \int d\mathcal{N}(w) \int d\mathcal{N}(y_1) \phi \left( D_{11}^{1/2} \left( \sqrt{1 - \frac{|D_{12}|}{(D_{11}D_{22})^{1/2}}} y_1 + \text{sgn} D_{12} \sqrt{\frac{|D_{12}|}{(D_{11}D_{22})^{1/2}}} w \right) + m_1 + \delta v(t_1) \right). \end{aligned} \quad (\text{F7})$$

In the above,  $y_1, y_2$  and  $w$  are independent unit Gaussian variables integrated with  $d\mathcal{N}(w) = \exp(-w^2/2)dw/\sqrt{2\pi}$ . We have used the abbreviations  $D_{\alpha\beta} = D(t_\alpha, t_\beta)$  and  $m_\alpha = m(t_\alpha)$  for  $\alpha, \beta = 1, 2$ . In the calculation of  $\tilde{\phi}$ , we use  $\delta v(t) \stackrel{\text{def}}{=} (1 + \partial_t)^{-1} \tilde{I}(t)$ . To calculate the double integral in the above efficiently, we use a table that retains the values obtained by performing one of the two integrations with respect to  $y_1$  or  $y_2$  for a fixed value of  $w$ . More precisely, we prepare a table consisting of the values of the integral

$$\int d\mathcal{N}(y) \phi(\alpha y + \beta), \quad (\text{F8})$$

for different values of  $(\alpha, \beta)$ . In the calculation of Eq. (F5), we perform the double integration by interpolating the values in the table and integrating them with respect to  $w$ .

Finally, we obtain a realization of the random variable,  $\eta(t + \Delta t)$ . Recall that the realization probability of  $\eta$  is give by Eq. (20), which is rewritten as

$$P(\eta(t + \Delta t) | \{\eta(t - s)\}_{s \geq 0}) \propto \exp\left(-\frac{1}{2} \eta^T C^{-1} \eta\right). \quad (\text{F9})$$

The above conditional Gaussian distribution has mean  $\mu_{t+\Delta t}$  and variance  $\nu_{t+\Delta t}$ , with

$$\mu_{t+\Delta t} = c_{t+\Delta t}^T C_{t:t-T_2}^{-1} \eta, \quad (\text{F10})$$

$$\nu_{t+\Delta t} = C(t + \Delta t, t + \Delta t) - c_{t+\Delta t}^T C_{t:t-T_2}^{-1} c_{t+\Delta t}. \quad (\text{F11})$$

In these equations, we define the column vector  $c_{t+\Delta t}(s_1) \stackrel{\text{def}}{=} C(s_1, t + \Delta t)$  and matrix  $C_{t:t-T_2}(s_1, s_2) \stackrel{\text{def}}{=} C(s_1, s_2)$  for the restricted range of time indices  $t - T_2 \leq s_1, s_2 \leq t$ . Since approximation errors may be larger at the boundary of the time domain, we use a smaller value of  $T_2 (< T_1)$ . From the above, we obtain a realization of  $\eta(t + \Delta t)$  using independent unit Gaussian random variable  $\xi_{t+\Delta t}$  as

$$\eta(t + \Delta t) = \mu_{t+\Delta t} + \sqrt{\nu_{t+\Delta t}} \xi_{t+\Delta t}. \quad (\text{F12})$$

In this way, we obtain all necessary updated values.

For the computation in Eq. (F12), we need to calculate the inverse matrix,  $C_{t:t-T_2}^{-1}$ . Since our computation of matrix  $C$  is based on numerical integration, small amounts of errors are

inevitable. When the inverse matrix is computed, the effects of small errors in the small eigenvalues of the matrix can be large. Thus we introduce a small ridge, computing  $(C_{t:t-T_2} + \epsilon \text{diag}(C_{t:t-T_2}))^{-1}$  instead of  $C_{t:t-T_2}^{-1}$ , where the diagonal matrix  $\text{diag}(C_{t:t-T_2})$  consists of the diagonal elements of  $C_{t:t-T_2}$ . This amounts to ignoring the small eigenvalues of  $C$ . Also, since taking matrix inverse at every timestep is inefficient, we update the inverse matrix using the formula,

$$\begin{aligned} (A + pq^T + qp^T)^{-1} &= A^{-1} - \frac{1 + a_{pq}}{(1 + a_{pq})^2 - a_{pp}a_{qq}} \\ &\quad \times \left\{ b_q b_p^T + b_p b_q^T - \frac{a_{pp}}{1 + a_{pq}} b_q b_q^T - \frac{a_{qq}}{1 + a_{pq}} b_p b_p^T \right\}, \end{aligned} \quad (\text{F13})$$

for a square symmetric matrix  $A$  and column vectors  $p$  and  $q$  of the corresponding size. In the above, we define

$$\begin{aligned} a_{pp} &= p^T A^{-1} p, & a_{pq} &= p^T A^{-1} q, & a_{qq} &= q^T A^{-1} q, \\ b_p &= A^{-1} p, & b_q &= A^{-1} q. \end{aligned} \quad (\text{F14})$$

Applying this formula with  $A = C_{t:t-T_2}$ ,  $p(t - T_2) = -1$ ,  $p(t - s_1) = 0$  for  $s_1 \neq T_2$ ,  $q(t - T_2) = 0$  and  $q(s_2) = C(t - T_2, s_2)$  for  $t - T_2 < s_2 \leq t$  gives the inverse matrix of  $C_{t:t-T_2+\Delta t}$  in the upper left part of the output matrix. Applying this formula with  $A(t + \Delta t, t + \Delta t) = C(t + \Delta t, t + \Delta t)$ ,  $A(s_1, t + \Delta t) = A(t + \Delta t, s_1) = 0$ ,  $A(s_1, s_2) = C(s_1, s_2)$ ,  $p(t + \Delta t) = 1$ ,  $p(s_1) = 0$ ,  $q(t + \Delta t) = 0$ , and  $q(s_1) = C_\ell(t + \Delta t, s_1)$  for  $t - T_2 < s_1, s_2 \leq t$  then gives  $C_{\ell, t+\Delta t; t-T_2+\Delta t}^{-1}$ . To avoid the accumulation of numerical errors, we directly compute the inverse matrix every 500 time steps. In the main text, we use the following values of the parameters:  $T_1 = 480$ ,  $T_2 = T_1/2$  for Figs. 6 and 10, and  $T_1 = 960$ ,  $T_2 = T_1/2$  for Fig. 2. We use  $\epsilon = 1.0 \times 10^{-6}$  for  $g_0 = 0.25$  in Fig. 2,  $\epsilon = 1.0 \times 10^{-3}$  for Figs. 6 and 10, and  $\epsilon = 1.0 \times 10^{-4}$  for the rest. We use the discrete time steps with step size  $\Delta t = 15/128$ . Each simulation is performed using 16 cores of recent versions of Intel Xeon processors in parallel for a few hours—a few days.



**APPENDIX G: PERTURBATIVE EXPANSION FOR CHAOTIC SOLUTIONS WITH MEAN ACTIVITIES OF SMALL AMPLITUDES**

In Sec. IV A 1, we have mentioned a perturbative expansion around  $g_0 = 0$  for the calculation of the moments of the mean activity. This perturbative calculation scheme, however, turns out to not work well. We briefly explain how it is performed and why it does not work well. According to our MFT, the probability distribution over sample paths for the mean activity  $m(t)$  is given by  $(1 + \partial_t)m(t) = \sqrt{2}g_0\eta(t)$  with

$$p(\eta) = \exp\left(-\frac{1}{2}\eta^T C[\eta]\eta - F[\eta] - \text{const.}\right). \quad (\text{G1})$$

In this equation, we make explicit the dependence of  $C$  and  $F$  on  $\eta$ . Although the above equation is derived for sample paths from a fixed initial condition, we assume that it holds on the entire time axis. This is justified by the intuition that the mixing property of chaotic dynamics keep the calculated moments from being severely affected by the boundary values.

In what follows, for illustration, we focus on the calculation of autocorrelation

$$\langle m(t)m(t + \tau) \rangle = \frac{g_0^2}{\pi} \int_{\mathbf{R}^2} e^{ip_1 t - ip_2(t - \tau)} \frac{1}{(1 + ip_1)(1 - ip_2)} \langle \widehat{\eta}(p_1) \widehat{\eta}(p_2) \rangle dp_1 dp_2, \quad (\text{G2})$$

in the frequency domain, which is easier to carry out than that in the time domain. Here,  $\widehat{\cdot}$  denotes the Fourier transform. Thus our objective is to compute the following moment:

$$\langle \widehat{\eta}(p_1) \widehat{\eta}(p_2) \rangle = \int \mathcal{D}\widehat{\eta} \mathcal{D}\widehat{\eta} \widehat{\eta}(p_1) \widehat{\eta}(p_2) \frac{1}{Z} \exp\left(-\frac{1}{2}\widehat{\eta}^T \widehat{C}[\widehat{\eta}] \widehat{\eta} - \widehat{F}[\widehat{\eta}]\right), \quad (\text{G3})$$

where we define the normalization constant  $Z$  and the Fourier transform of the autocorrelation  $\widehat{C}[\widehat{\eta}](\omega_1, \omega_2) \stackrel{\text{def}}{=} \frac{1}{2\pi} \int_{\mathbf{R}^2} e^{-i\omega_1 t + i\omega_2 s} C[\eta](t, s) dt ds$ . We also define  $\widehat{F}$  with  $\widehat{F}[\widehat{\eta}] = F[\eta]$ .

For this calculation, we need to compute the perturbed correlation matrix, which can be carried out by differentiating Eq. (E29) with respect to  $\eta$ . The first-order response in  $D(t, s)$  then obeys the following equation:

$$(1 + \partial_t)(1 + \partial_s)\delta D^{(1)}(t, s) = a_1(t - s)\delta D^{(1)}(t, s) + a_2(t - s)\delta D^{(1)}(t, t) + a_2(s - t)\delta D^{(1)}(s, s) + a_3(t - s)\delta m(t) + a_3(s - t)\delta m(s), \quad (\text{G4})$$

where we define  $a_1(t - s) = 2\sigma_0^2 \langle \phi'(h(t))\phi'(h(s)) \rangle_0$ ,  $a_2(t, s) = \sigma_0^2 \{ \langle \phi''(h(t))\phi(h(s)) \rangle_0 - \overline{\phi_0''\phi_0} \}$ , and  $a_3(t - s) = 2\sigma_0^2 \{ \langle \phi'(h(t))\phi(h(s)) \rangle_0 - \overline{\phi_0'\phi_0} \}$ . Here, the average over the unperturbed dynamics with  $m(t) \equiv 0$  is denoted by the angle bracket with subscript 0, and we use abbreviations such as  $\overline{\phi_0} \stackrel{\text{def}}{=} \langle \phi(h(t)) \rangle_0$ . These coefficients originate from the differentiation of  $C(t, s)$  with respect to  $D(t, s)$  and  $m(t)$ , as summarized in Appendix J. Since  $\phi(h(t)) - \overline{\phi_0}$  and  $\phi'(h(t)) - \overline{\phi_0'}$  are odd and even functions of  $h(t)$ , and since the dynamical variable  $h(t)$  is distributed symmetrically around  $h(t) = 0$  in the unperturbed dynamics, we find  $a_3(t, s) = 0$ . Thus there is no driving force in the above equation, and we obtain  $\delta D^{(1)}(t, s) \equiv 0$ .

Next, we examine the second-order response in  $D(t, s)$ , which obeys the following nonhomogeneous linear equation:

$$(1 + \partial_t)(1 + \partial_s)\delta D^{(2)}(t, s) = a_1(t - s)\delta D^{(2)}(t, s) + a_2(t - s)\delta D^{(2)}(t, t) + a_2(s - t)\delta D^{(2)}(s, s) + b_1(t - s)\delta m(t)^2 + b_1(s - t)\delta m(s)^2 + b_2(t - s)\delta m(t)\delta m(s), \quad (\text{G5})$$

where we define  $b_1(t - s) = 2\sigma_0^2 \{ \langle \phi''(h(t))\phi(h(s)) \rangle_0 - \overline{\phi_0''\phi_0} \}$  and  $b_2(t - s) = 2\sigma_0^2 \{ \langle \phi'(h(t))\phi'(h(s)) \rangle_0 - \overline{\phi_0'\phi_0'} \}$ .

To obtain a solution to this equation, we first compute an approximate solution that satisfies

$$(1 + \partial_t)(1 + \partial_s)\delta D_0^{(2)}(t, s) = b_2(t - s)\delta m(t)\delta m(s) + \dots, \quad (\text{G6})$$

by ignoring the first three terms of the right-hand side of Eq. (G5). This equation is solved in the frequency domain as

$$\widehat{\delta D}_0^{(2)}(\omega_1, \omega_2) = \frac{2g_0^2}{\sqrt{2\pi}(1 + i\omega_1)(1 - i\omega_2)} \int_{\mathbf{R}} \frac{\widehat{b}_2(\omega_3)\widehat{\eta}(\omega_1 - \omega_3)\widehat{\eta}(\omega_2 - \omega_3)}{(1 + i(\omega_1 - \omega_3))(1 + i(\omega_2 - \omega_3))} d\omega_3 + \dots \quad (\text{G7})$$

The above solution leaves a residual error on the right-hand side of Eq. (G5). Then, we recursively make corrections to the solution by considering the residual error as a new nonhomogeneous term and solving

$$(1 + \partial_t)(1 + \partial_s)\delta D_{k+1}^{(2)}(t, s) = a_1(t - s)\delta D_k^{(2)}(t, s) + a_2(t - s)\delta D_k^{(2)}(t, t) + a_2(s - t)\delta D_k^{(2)}(s, s), \quad (\text{G8})$$

in the frequency domain.

Assuming that the series thus obtained is convergent, we obtain the second-order response in the form  $\delta D^{(2)}(t, s) = \sum_{k=0}^{\infty} \delta D_k^{(2)}(t, s)$ . Using this response, we obtain the response in  $C(t, s)$  as  $\widehat{\delta C}^{(2)}(\omega_1, \omega_2) = (1 + i\omega_1)(1 - i\omega_2)\widehat{\delta D}^{(2)}(\omega_1, \omega_2)/(2\sigma_0^2)$ .

Now, we obtain

$$\langle \widehat{\eta}(p_1) \widehat{\eta}(p_2) \rangle \approx \langle \widehat{\eta}(p_1) \widehat{\eta}(p_2) \rangle_0 + \langle \widehat{\eta}(p_1) \widehat{\eta}(p_2) \left( \frac{1}{2}\widehat{\eta}^T \widehat{C}[0]^{-1} \widehat{\delta C}^{(2)} \widehat{C}[0]^{-1} \widehat{\eta} - \frac{1}{2}\widehat{C}[0]^{-1} \odot \widehat{\delta C}^{(2)} + \dots \right) \rangle_c. \quad (\text{G9})$$

Here, the second angle bracket denotes average of just the connected contribution between the external legs and the interaction vertex, as is conventional in diagrammatic calculations [56]. The symbol  $\odot$  denotes the elementwise product of two matrices followed by integration with respect to the two argument variables. We omit to write down the terms originating from the complicated dependence of  $F$  on  $\eta$ . In principle, we can calculate the desired moment from the above expansion.

However, this calculation scheme turns out to be difficult to carry out. The difficulty originates from the slow convergence of the sum of the series. Because of this, we need to compute  $\delta D_k^{(2)}$  up to a large  $k$ . However, we find that the diagrammatic calculations of  $\delta D_k^{(2)}$  up to a large  $k$  requires us to carry out multiple integrals explicitly, and this is computationally intractable. This happens because of the lack of interchangeability between the diagrammatic averaging and the recursion in Eq. (G8). In the main text, because of this difficulty, we restrict ourselves just to the autocorrelation function calculated using the crudest approximation,  $\langle \tilde{\eta}(p_1) \tilde{\eta}(p_2) \rangle \approx \langle \tilde{\eta}(p_1) \tilde{\eta}(p_2) \rangle_0$  [Figs. 3(a) and 3(c)].

#### APPENDIX H: PERTURBATIVE STABILITY ANALYSIS OF FIXED POINTS

In this section, we analyze the stability of fixed-point solutions observed in Sec. IV B 3 of the main text. We consider

the case in which the mean activity initially takes a constant value,  $m(t) = m_f$  for  $t \leq 0$ . According to the MFT, the mean activity is determined by

$$(1 + \partial_t) \delta m(t) = \sqrt{2} g_0 \eta(t) - m_f, \quad (\text{H1})$$

where we define,  $\delta m(t) \stackrel{\text{def}}{=} m(t) - m_f$ . To analyze the dynamics of the mean activity, we first need to examine the associated microscopic Gaussian fluctuations determined by the past values of  $m(t)$ . For a certain range of constant values of  $m(t)$ , these Gaussian fluctuations have a stable fixed-point solution. For these fixed points, the correlation matrices  $D(t, s)$  and  $\tilde{C}(t, s)$  take constant values,  $D_0$  and  $\tilde{C}_0$ , which are obtained by solving the self-consistent equation, Eq. (38). As described in the main text, we assume that  $\eta(t) = m_f / \sqrt{2} g_0$  holds for  $t \leq 0$ . Then, the network state stays at the fixed point without requiring external inputs, and we have  $\delta m(t) = 0$  for  $t > 0$  if there is no external input in  $t > 0$ . This condition is expected to be satisfied for some value of  $m_f$  with a nonzero probability (see the discussion at the end of this section as well). We then examine the linear response for  $t > 0$  to temporary external perturbative inputs, collectively denoted by  $\mathbf{p}$ .

We first examine responses in the correlation matrices up to the second order in  $\mathbf{p}$ , which determines  $O(|\mathbf{p}|)$  response in  $\eta(t)$ . By simply differentiating the dynamical equation for the autocorrelation function, Eq. (E29), we find that the first-order response in  $D(t, s)$ , which is denoted by  $\delta D^{(1)}(t, s)$ , obeys

$$(1 + \partial_t)(1 + \partial_s) \delta D^{(1)}(t, s) = a_1 \delta D^{(1)}(t, s) + a_2 (\delta D^{(1)}(t, t) + \delta D^{(1)}(s, s)) + a_3 (\delta m(t) + \delta m(s)) + p_1(t) + p_1(s), \quad (\text{H2})$$

where the term,  $p_1(t)$ , represent the effects of the input component that is correlated with the unperturbed neuronal fluctuations, and we define  $a_1 = 2\sigma_0^2 \langle \phi'(h)^2 \rangle_0$ ,  $a_2 = \sigma_0^2 \langle \phi''(h) \phi(h) \rangle_0$  and  $a_3 = 2\sigma_0^2 \langle \phi'(h) \phi(h) \rangle_0$ . These coefficients originate from the differentiation of  $\tilde{C}(t, s)$  with respect to  $D(t, s)$  and  $m(t)$ , which is summarized in Appendix J. Note that these coefficients must be appropriately replaced by the values of centered statistics, such as  $\sigma_0^2 \langle (\phi''(h) - \overline{\phi''}) (\phi(h) - \overline{\phi_0}) \rangle_0$ , for the case with finely tuned synaptic weights. Here, we use abbreviations such as  $\overline{\phi''} = \langle \phi''(h) \rangle_0$  for the statistics averaged over the unperturbed dynamics. The solution for the above equation, with boundary condition,  $\delta D^{(1)}(t, s) = 0$  for  $t, s \leq 0$ , satisfies

$$\begin{aligned} \delta D^{(1)}(t, s) &= \int_0^t e^{-(1-a_1)(t-\tau)} (a_2 \delta D^{(1)}(\tau, \tau) + a_3 \delta m(\tau) + p_1(\tau)) d\tau \\ &+ \int_0^s e^{-(1-a_1)(s-\tau)} (a_2 \delta D^{(1)}(\tau, \tau) + a_3 \delta m(\tau) + p_1(\tau)) d\tau. \end{aligned} \quad (\text{H3})$$

Putting  $s = t$  and differentiating both sides of the above equation with respect to  $t$ , we have

$$((1 - a_1 - 2a_2) + \partial_t) \delta D^{(1)}(t, t) = 2(a_3 \delta m(t) + p_1(t)), \quad (\text{H4})$$

which then gives

$$\delta D^{(1)}(t, t) = 2 \int_0^t e^{-(1-a_1-2a_2)(t-\tau)} (a_3 \delta m(\tau) + p_1(\tau)) d\tau. \quad (\text{H5})$$

Comparing the above solution with Eq. (H3) gives the solution for  $\delta D^{(1)}(t, s)$ . This response is bounded if  $1 - a_1 - 2a_2 > 0$  and if  $\delta m$  and  $p_1$  are bounded.

From the above analysis, by putting  $\delta D^{(1)}(t, s) = r(t) + r(s)$ , we obtain

$$D(t, s) \approx D_0 + r(t) + r(s) + O(|\mathbf{p}|^2). \quad (\text{H6})$$

Noting that the first-order response in  $\overline{\phi}(t) \overline{\phi}(s)$  is given by

$$\delta(\overline{\phi}(t) \overline{\phi}(s)) = \overline{\phi_0} \left( \frac{1}{2} \overline{\phi_0''} \delta D(t, t) + \overline{\phi_0'} \delta m(t) + \frac{1}{2} \overline{\phi_0''} \delta D(s, s) + \overline{\phi_0'} \delta m(s) \right), \quad (\text{H7})$$

we have the correlation matrix for  $(1 + \partial_t)^{-1}\eta(t)$ ,

$$\begin{aligned} V(t, s) &\stackrel{\text{def}}{=} (1 + \partial_t)^{-1}(1 + \partial_s)^{-1}(\tilde{C}(t, s) - \bar{\phi}(t)\bar{\phi}(s)) \\ &= (\sqrt{V_0} + v(t))(\sqrt{V_0} + v(s)) + O(|\mathbf{p}|^2), \end{aligned} \quad (\text{H8})$$

where we define  $V_0 = \frac{1}{2\sigma_0^2}D_0 - \bar{\phi}_0^2$  and

$$v(t) = \frac{1}{2\sigma_0^2\sqrt{V_0}} \left\{ \left( 1 - \frac{2\sigma_0^2\bar{\phi}_0\phi_0''}{a_1 + 2a_2} \right) r(t) + 2\sigma_0^2 \left( \frac{\bar{\phi}_0\phi_0''a_3}{a_1 + 2a_2} - \bar{\phi}_0\phi_0' \right) (1 + \partial_t)^{-1}\delta m(t) + \frac{2\sigma_0^2\bar{\phi}_0\phi_0''}{a_1 + 2a_2} (1 + \partial_t)^{-1}p_1(t) \right\}. \quad (\text{H9})$$

Noting that  $V(t, s) = V_0$  and  $v(t) = v(s) = 0$  holds for  $t, s \leq 0$ , and that  $V(t, s)$  is the correlation matrix that determines the realization probability of  $(1 + \partial_t)^{-1}\eta(t)$ , we have

$$(1 + \partial_t)^{-1}\eta(t) = \frac{m_t}{\sqrt{2}g_0} \left( 1 + \frac{v(t)}{\sqrt{V_0}} \right) + \delta\eta^{(2)}(t) + O(|\mathbf{p}|^2), \quad (\text{H10})$$

where  $\delta\eta^{(2)}(t)$  represents  $O(|\mathbf{p}|)$  fluctuations due to the higher-order response in  $D(t, s)$ . Here, recall the argument about the representation of  $\eta$  in terms of unit Gaussian variables  $\xi$  in Appendix E. In the case with finely tuned synaptic weights, the correlation matrix for  $(1 + \partial_t)^{-1}\eta(t)$  is given by  $D(t, s)/2\sigma_0^2$  and we have the same representation with  $V_0 = D_0/2\sigma_0^2$  and  $v(t) = r(t)/2\sigma_0^2\sqrt{V_0}$ .

Next, we consider the second-order response in  $D(t, s)$ , which is denoted by  $\delta D^{(2)}(t, s)$ . Differentiating Eq. (H2) once again, we obtain

$$\begin{aligned} (1 + \partial_t)(1 + \partial_s)\delta D^{(2)}(t, s) &= a_1\delta D^{(2)}(t, s) + a_2(\delta D^{(2)}(t, t) + \delta D^{(2)}(s, s)) + b_1\delta D^{(1)}(t, s)^2 + b_2(\delta D^{(1)}(t, t)^2 + \delta D^{(1)}(s, s)^2) \\ &\quad + b_3\delta D^{(1)}(t, t)\delta D^{(1)}(s, s) + b_4(\delta D^{(1)}(t, t) + \delta D^{(1)}(s, s))\delta D^{(1)}(t, s) \\ &\quad + b_5(\delta m(t)^2 + \delta m(s)^2) + b_6\delta m(t)\delta m(s) + \sum_i p_{2,i}(t)p_{2,i}(s), \end{aligned} \quad (\text{H11})$$

where we define  $b_1 = 2\sigma_0^2\langle\phi''(h)\phi''(h)\rangle_0$ ,  $b_2 = \frac{\sigma_0^2}{2}\langle\phi''''(h)\phi(h)\rangle_0$ ,  $b_3 = \sigma_0^2\langle\phi''(h)^2\rangle_0$ ,  $b_4 = 2\sigma_0^2\langle\phi'''(h)\phi'(h)\rangle_0$ ,  $b_5 = 2\sigma_0^2\langle\phi''(h)\phi(h)\rangle_0$  and  $b_6 = 4\sigma_0^2\langle\phi'(h)\phi'(h)\rangle_0$ . We also define the random part of the input,  $p_{2,i}(t)$ .

The solution for the above nonhomogeneous linear equation is obtained as the superposition of special solutions for the equations with each of the nonhomogeneous terms on the right-hand side plus a solution for the homogeneous equation. Noting  $\delta D^{(1)}(t, s) = r(t) + r(s)$ , we rewrite the nonhomogeneous terms as

$$(b_1 + 4b_2 + 2b_4)(r(t)^2 + r(s)^2) + b_5(\delta m(t)^2 + \delta m(s)^2) + (2b_1 + 4b_3 + 4b_4)r(t)r(s) + b_6\delta m(t)\delta m(s) + \sum_i p_{2,i}(t)p_{2,i}(s). \quad (\text{H12})$$

We ignore the first two terms above, because they only yield responses of  $O(|\mathbf{p}|^2)$  magnitude in  $\eta(t)$ . This can be seen by checking that these two terms only make corrections to Eq. (H8) of the following form:

$$\begin{aligned} V(t, s) &\approx (\sqrt{V_0} + v(t) + O(|\mathbf{p}|^2))(\sqrt{V_0} + v(s)) \\ &\quad + O(|\mathbf{p}|^2) + o(|\mathbf{p}|^2). \end{aligned} \quad (\text{H13})$$

Also note that the response in  $\bar{\phi}(t)\bar{\phi}(s)$  due to  $\delta D^{(2)}(t, s)$  is negligible for the same reason. Thus we are interested in nonhomogeneous equations of the following form:

$$\begin{aligned} (1 + \partial_t)(1 + \partial_s)\delta D^{(2)}(t, s) &= a_1\delta D^{(2)}(t, s) + a_2(\delta D^{(2)}(t, t) \\ &\quad + \delta D^{(2)}(s, s)) + q_0(t)q_0(s). \end{aligned} \quad (\text{H14})$$

In solving this equation, we first ignore the unknown quantities on the right-hand side and obtain the following approximate solution:

$$\begin{aligned} \delta D_{q,1}^{(2)}(t, s) &= q_1(t)q_1(s) \\ &\stackrel{\text{def}}{=} \{(1 + \partial_t)^{-1}q_0(t)\}\{(1 + \partial_s)^{-1}q_0(s)\}. \end{aligned} \quad (\text{H15})$$

With this solution, residual error  $a_1q_1(t)q_1(s) + a_2(q_1(t)^2 + q_1(s)^2)$  remains on the right-hand side of Eq. (H14). We then solve Eq. (H14) by regarding the residual-error term as a new nonhomogeneous term and by ignoring the two unknown terms on the right-hand side again. Then, we obtain the following two corrections:

$$\begin{aligned} \delta D_{q,2}^{(2)}(t, s) &= q_2(t)q_2(s) \\ &\stackrel{\text{def}}{=} \{\sqrt{a_1}(1 + \partial_t)^{-1}q_1(t)\}\{\sqrt{a_1}(1 + \partial_s)^{-1}q_1(s)\}, \end{aligned} \quad (\text{H16})$$

$$\delta D_{q,2}^{(2)'}(t, s) = a_2\{(1 + \partial_t)^{-1}(q_1^2)(t) + (1 + \partial_s)^{-1}(q_1^2)(s)\}. \quad (\text{H17})$$

The second term in the above is again negligible because of its  $O(|\mathbf{p}|^2)$  contribution to  $\eta$ . We can then recursively obtain  $\delta D_{q,j}^{(2)}$ , ( $j \geq 2$ ) in the same manner as above and obtain the relevant part of the solution for Eq. (H11) as  $\sum_{1 \leq j < \infty} \delta D_{q,j}^{(2)}(t, s)$ . If we have  $a_1 < 1$ , this series is convergent because

we have

$$\|\widehat{q}_{j+1}\|_1 = \|\sqrt{a_1}(1 + \widehat{\partial}_t)^{-1}q_j\|_1 \leq \sqrt{a_1}\|\widehat{q}_j\|_1. \quad (\text{H18})$$

Here,  $\widehat{\cdot}$  denotes the Fourier transform, and  $\|\cdot\|_1$  denotes the  $L^1$ -norm. From the above analysis, we now have the first- and second-order responses in  $D(t, s)$ , which yields

$$D(t, s) \approx D_0 + r(t) + r(s) + \sum_{j,1 \leq \ell < \infty} d_{2,j\ell}(t)d_{2,j\ell}(s) + \sum_{j,1 \leq \ell < \infty} d_{3,j\ell}(t)d_{3,j\ell}(s). \quad (\text{H19})$$

In the above, we obtain  $d_{2,j\ell}(t)$  by applying the recursion relation in Eq. (H16) with  $q_\ell(t) = d_{2,j\ell}(t)$  and  $q_0(t) = d_{2,1,0}(t) = \sqrt{2b_1 + 4b_3 + 4b_4}r(t)$  or  $d_{2,2,0}(t) = \sqrt{b_6}\delta m(t)$ . We also define  $d_{3,j\ell}(t)$  by applying the recursion relation in Eq. (H16) with  $q_\ell(t) = d_{3,j\ell}(t)$  and  $q_0(t) = d_{3,j,0}(t) = p_{2,j}(t)$ . Note that  $v(t)v(s)$  needs to be suitably subtracted from these terms to compensate the corresponding term in Eq. (H8), which modifies the definitions of  $d_{2,j1}$  and  $d_{3,j1}$  described above. We omit the precise expressions of these terms, because they do not affect our conclusion.

Noting that the second-order response in  $D(t, s)$  does not evoke a response in  $\overline{\phi}(t)\overline{\phi}(s)$  which leads to  $O(|\mathbf{p}|)$  response in  $\eta$ , we have the relevant part of the response in  $V(t, s)$ ,

$$V(t, s) \approx (\sqrt{V_0} + v(t))(\sqrt{V_0} + v(s)) + \frac{1}{2\sigma_0^2} \left\{ \sum_{j,1 \leq \ell < \infty} d_{2,j\ell}(t)d_{2,j\ell}(s) + \sum_{j,1 \leq \ell < \infty} d_{3,j\ell}(t)d_{3,j\ell}(s) \right\}. \quad (\text{H20})$$

Putting  $d_1(t) = v(t)/\sqrt{V_0}$ , we now have  $\delta\eta^{(2)}$  in Eq. (H10) and obtain

$$(1 + \partial_t)^{-1}\eta(t) = \frac{m_f}{\sqrt{2}g_0}(1 + d_1(t)) + \frac{1}{\sqrt{2}\sigma_0} \left( \sum_{j,\ell} \xi_{j\ell}d_{2,j\ell}(t) + \sum_{j,\ell} \xi'_{j\ell}d_{3,j\ell}(t) \right) + O(|\mathbf{p}|^2), \quad (\text{H21})$$

where we define *i.i.d.* unit Gaussian random variables,  $\{\xi_{j\ell}\}_{j,\ell}$  and  $\{\xi'_{j\ell}\}_{j,\ell}$ . Once again, note that the outer-product representation in Eq. (H20) gives a transform of the form of  $(1 + \partial_t)^{-1}\eta = U\xi$ ,  $UU^T = V$ , which gives a representation of  $\eta$  with unit Gaussian variables,  $\xi$ , as we discussed in Appendix E. Together with the equation for  $m(t)$  for  $t > 0$ ,

$$(1 + \partial_t)\delta m(t) = \sqrt{2}g_0\delta\eta(t) + p_0(t), \quad \delta\eta(t) = \eta(t) - \frac{m_f}{\sqrt{2}g_0}, \quad (\text{H22})$$

Eq. (H21) gives a self-consistent equation that the first-order responses in  $m(t)$  must satisfy. Here, we define the uniform component of the input,  $p_0(t)$ .

From the above relation, we obtain the first-order response in  $m(t)$  for a given set of values of  $\{\xi_{j\ell}\}$  and  $\{\xi'_{j\ell}\}$  by further iteration. Initially, calculating  $d_1(t)$  and  $d_{3,j\ell}(t)$  for  $\delta m(t) = 0$  [denoted by  $d_1^{(0)}(t)$  and  $d_{3,j\ell}^{(0)}(t)$ , respectively], we have

$$\delta m^{(0)}(t) = (1 + \partial_t)^{-1}p_0(t) + m_f d_1^{(0)}(t) + \frac{g_0}{\sigma_0} \sum_{j,\ell} \xi'_{j\ell} d_{3,j\ell}^{(0)}(t). \quad (\text{H23})$$

The solution for the above equation needs the following correction to the right-hand side of the first of Eq. (H22):

$$\sqrt{2}g_0\delta\eta^{(1)}(t) = (1 + \partial_t) \left\{ m_f d_1^{(1)}(t) + \frac{g_0}{\sigma_0} \sum_{j,\ell} \xi_{j\ell} d_{2,j\ell}^{(1)}(t) \right\} + O(|\mathbf{p}|^2), \quad (\text{H24})$$

where  $d_1^{(1)}(t)$  and  $d_{2,j\ell}^{(1)}(t)$  are the corrections to  $d_1(t)$  and  $d_{2,j\ell}(t)$  due to the change  $\delta m^{(0)}(t)$ . We then recursively

obtain  $\delta m^{(j)}$  and  $\delta\eta^{(j)}$  by alternately correcting the errors in Eqs. (H21) and (H22). The sum of the series thus obtained gives the desired first-order solution,  $\delta m(t) = \sum_{j \geq 0} \delta m^{(j)}(t)$ , and the sum of this series converges with a nonzero probability if we have  $a_1 < 1$  and

$$\|\widehat{d}_1^{(k)}\|_1 \leq \theta_1 \|\widehat{\delta m}^{(k)}\|_1, \quad (\text{H25})$$

with  $m_f\theta_1 < 1$ . From Eqs. (H5) and (H9), we find

$$\theta_1 = \frac{1}{2\sigma_0^2 V_0} \sup_{\omega} \left| \frac{a_3}{1 - a_1 - 2a_2 + i\omega} \left( 1 - \frac{2\sigma_0^2 \overline{\phi_0} \phi_0''}{a_1 + 2a_2} \right) + \frac{2\sigma_0^2}{1 + i\omega} \left( \frac{a_3 \overline{\phi_0} \phi_0''}{a_1 + 2a_2} - \overline{\phi_0} \phi_0' \right) \right|. \quad (\text{H26})$$

Here, note that the norm of  $d_{2,j\ell}^{(k)}$  exponentially decreases as index  $\ell$  increases and the norm of  $d_{2,j0}^{(k)}$  is bounded by a certain multiple of the norm of  $\delta m^{(k)}$ . This implies that, for any positive  $\theta_2$ ,

$$\left\| \sqrt{2}g_0 \sum_{j,\ell} \xi_{j\ell} \widehat{d}_{2,j\ell}^{(k)} \right\|_1 < \theta_2 \|\widehat{\delta m}^{(k)}\|_1 \quad (\text{H27})$$

holds with a nonzero probability. The convergence in 1-norm in the frequency domain implies the uniform convergence in the time domain, which asserts that  $\mathbf{p}(t) \rightarrow \mathbf{0}$  implies  $\delta m(t) \rightarrow 0$  as  $t \rightarrow \infty$  with the aid of the dominated convergence theorem. Hence we have proved the linear stability with a nonzero probability. Also note that a single configuration of the random connectivity of the network corresponds to a single set of values for  $\{\xi_{j\ell}\}$ . Otherwise, the above solution is not consistent with the linearity of the response.



Besides the above stability result, we can also examine how the response dynamics diverge depending on the configuration. For example, suppose that the random coefficients  $\xi_{j\ell}$  take large values for small  $\ell$ , and that all the inputs initially take constant values for a sufficiently long time. Then, the constructed solution is likely to diverge at some  $t$ , because the divergence of  $d_{2,j\ell}^{(k)}$  for small  $\ell$  and for  $k \rightarrow \infty$  cannot be compensated by the other terms. However, even for such a divergent solution, if we restrict the domain of the solution to a short time interval,  $t \in [0, t_0]$ , the supremum of  $d_{2,j\ell}^{(k)}$  can be controlled to decay as index  $k$  increases. Here, recall that  $d_{2,j\ell}^{(k)}$  is obtained by recursively applying  $(1 + \partial_t)^{-1}$  to  $d_{2,j0}^{(k)}$  and suppose to choose a very short time interval compared with the unit decay constant of  $(1 + \partial_t)^{-1}$ . This consideration indicates that, for any values of  $\xi_{j\ell}$  and  $\xi'_{j\ell}$ , the constructed solution converges over a short time interval.

Also, suppose the case in which the equality,  $m_f = \sqrt{2}g_0\eta(0)$ , is slightly violated and that the mean activity is initially clamped to  $m_f$  by an external input until the clamping input is removed at  $t = 0$ . Then, the inequality introduces a small constant, uniform, input term to the above self-consistent equation for  $\delta m$ . Up to the first-order, this results in the convergence of  $\delta m$  and  $\eta - m_f/\sqrt{2}g_0$  to nonzero values for  $t \rightarrow \infty$ , with the same nonzero probability as that for the above stability to an external input. This indicates that a nonzero interval of values of  $\eta(0)$  around  $m_f/\sqrt{2}g_0$  results in convergence to nearby fixed points. Thus we reasonably expect the existence of fixed points with a nonzero probability, although a rigorous proof of this needs careful evaluation of all higher-order terms of the perturbative expansion.

### APPENDIX I: PERTURBATIVE STABILITY ANALYSIS OF REGULAR OSCILLATORY SOLUTIONS

In this section, we extend the perturbation analysis developed in Appendix H and analyze the stability of the regular oscillations observed in the main text. Although this analysis is along the same lines as that in Appendix H, much more complex calculations are required for the present case. The complexity of these calculations would make it harder to grasp the basic idea behind it. Thus we recommend readers to first check Appendix H and to familiarize themselves with the basic idea before reading this section. In what follows, we present a theoretical framework of the analysis first and then show the values of the derived bounding constants for concrete cases at the end of the section.

Suppose that the mean activity of a network is initially set to a periodic orbit  $m_0(t)$  for  $t \leq 0$  and that temporary external perturbative inputs, collectively denoted by  $\mathbf{p}$ , are applied for  $t > 0$ . Also suppose that the neuronal fluctuations are coherent and phase locked to the oscillation for  $t \leq 0$ . As we have seen in the main text, such coherent neuronal fluctuations are found by iteratively applying the self-consistent equation:

$$(1 + \partial_t)(1 + \partial_s)D(t, s) = 2\sigma_0^2 \tilde{C}(t, s), \quad (11)$$

with the fixed mean activity  $m_0(t)$ . Concretely, we first transform the autocorrelation function on the right-hand side to the

frequency domain:  $\tilde{C}(\omega_1, \omega_2) = \frac{1}{2\pi} \int e^{-i\omega_1 t} e^{i\omega_2 s} \tilde{C}(t, s) dt ds$ . Then, the left-hand side is obtained by a simple algebraic computation in the frequency domain. After transforming this back to the time domain, we update the values of  $\tilde{C}(t, s)$  using Eq. (F5). Although this type of iterative approach is thought to be a heuristic method for finding a solution, the convergence argument we make below provides criteria for judging whether the solutions thus obtained are stable. In the unperturbed dynamics, we have the following eigen-decomposition of the correlation matrix for  $(1 + \partial_t)^{-1}\eta(t)$ :  $V(t, s) \stackrel{\text{def}}{=} (1 + \partial_t)^{-1}(1 + \partial_s)^{-1}(\tilde{C}(t, s) - \tilde{\phi}(t)\tilde{\phi}(s)) = \sum_i \lambda_i v_i(t)\bar{v}_i(s)$  with real eigenvectors  $v_i(t)$ . If the microscopic part of the dynamics is coherent, the eigenvectors can be expanded with the Fourier basis,  $\mathbf{e}_\ell(t) = e^{i\ell\omega_0 t}$ , for the basic frequency  $\omega_0$ , as  $v_i(t) = \sum_\ell v_{i\ell} \mathbf{e}_\ell(t)$ . In the initial unperturbed dynamics, the driving-force term  $\eta(t)$  is given by

$$(1 + \partial_t)m_0(t) = \sqrt{2}g_0\eta(t) + g_0\tilde{\phi}(t), \\ \eta(t) = (1 + \partial_t) \sum_i \sqrt{\lambda_i} \xi_{0,i} v_i(t), \quad (t \leq 0), \quad (12)$$

for a suitable set of values for  $\{\xi_{0,i}\}_i$ . Otherwise, the sample path for  $\eta(t)$  is never realized [cf. Eq. (E28)]. Here, the term  $\tilde{\phi}$  is the external driving-force term defined in Eq. (29). For the model without external inputs, we set  $\tilde{\phi}(t)$  to zero. For  $t > 0$ , the mean activity obeys the following dynamical equation:

$$(1 + \partial_t)\delta m(t) = \sqrt{2}g_0\eta(t) - (1 + \partial_t)m_0(t) + g_0\delta\tilde{\phi}(t) + p_0(t), \quad (13)$$

where we define  $\delta m(t) = m(t) - m_0(t)$  and define  $\delta\tilde{\phi}(t)$  as the deviation of the function  $\tilde{\phi}(t)$  from the initial periodic orbit. The last term,  $p_0(t)$ , represents the effect of the uniform component of the input.

To analyze the above response dynamics, we first examine how a change in  $m(t)$  evokes a response in  $D(t, s)$  for  $t, s \geq 0$ . The first-order response in  $D(t, s)$  is denoted by  $\delta D^{(1)}(t, s)$  and obeys

$$(1 + \partial_t)(1 + \partial_s)\delta D^{(1)}(t, s) \\ = a_1(t, s)\delta D^{(1)}(t, s) + a_2(t, s)\delta D^{(1)}(t, t) \\ + a_2(s, t)\delta D^{(1)}(s, s) + a_3(t, s)\delta m(t) + a_3(s, t)\delta m(s) \\ + \sum_i \sqrt{\lambda_i}((1 + \partial_t)v_i(t)p_{1,i}(s) + p_{1,i}(t)(1 + \partial_s)v_i(s)), \quad (14)$$

where we define  $a_1(t, s) = 2\sigma_0^2 \langle \phi'(h(t))\phi'(h(s)) \rangle_0$ ,  $a_2(t, s) = \sigma_0^2 \langle \phi''(h(t))\phi(h(s)) \rangle_0$ , and  $a_3(t, s) = 2\sigma_0^2 \langle \phi'(h(t))\phi(h(s)) \rangle_0$ . For the model with finely tuned synaptic weights, the above coefficient functions must be suitably replaced. The coefficients originate from the differentiation of  $\tilde{C}(t, s)$  and  $C(t, s)$  with respect to  $D(t, s)$  and  $m(t)$ , which is summarized in Appendix J. The term  $p_{1,i}$  represents the effects of the component of the input that is correlated with the  $i$ th mode of the unperturbed neuronal fluctuations.

We also note that the first-order response in  $V(t, s)$  is given by

$$(1 + \partial_t)(1 + \partial_s)\delta V^{(1)}(t, s) = \frac{1}{2\sigma_0^2} \left( a_1(t, s)\delta D^{(1)}(t, s) + \tilde{a}_2(t, s)\delta D^{(1)}(t, t) + \tilde{a}_2(s, t)\delta D^{(1)}(s, s) + \tilde{a}_3(t, s)\delta m(t) + \tilde{a}_3(s, t)\delta m(s) + \sum_i \sqrt{\lambda_i}((1 + \partial_t)v_i(t)p_{1,i}(s) + p_{1,i}(t)(1 + \partial_s)v_i(s)) \right), \tag{15}$$

where we define  $\tilde{a}_2(t, s) = 2\sigma_0^2(\langle \phi''(h(t))\phi(h(s)) \rangle_0 - \overline{\phi''_0(t)\phi_0(s)})$ , and  $\tilde{a}_3(t, s) = 2\sigma_0^2(\langle \phi'(h(t))\phi(h(s)) \rangle_0 - \overline{\phi'_0(t)\phi_0(s)})$ . For the model with finely tuned synaptic weights, the response dynamics for  $D(t, s)$  and  $V(t, s)$  are the same up to a multiplication constant.

A solution of the form  $\delta V^{(1)}(t, s) = \sum_i \sqrt{\lambda_i}(v_i(t)\overline{\delta v_i(s)} + \delta v_i(t)\overline{v_i(s)})$  is obtained from the above equations. By initially ignoring the first three terms on the right-hand side of Eq. (14) and using the expansion  $a_k(t, s) = \sum_{ij} a_{k,ij}\mathbf{e}_i(t)\overline{\mathbf{e}_j(s)}$ , for  $k = 1, 2$ , and  $3$ , we obtain the following approximate solution:

$$\begin{aligned} \delta D_0^{(1)}(t, s) &= \sum_i \sqrt{\lambda_i}(v_i(t)\overline{\delta d_{0,i}(s)} + \delta d_{0,i}(t)\overline{v_i(s)}), \\ \delta d_{0,i}(t) &= \sum_j \overline{M}_{3,ji}(1 + \partial_t)^{-1}(\mathbf{e}_j(t)\delta m(t)) + (1 + \partial_t)^{-1}p_{1,i}(t), \end{aligned} \tag{16}$$

where we define  $\mathbf{e}_\ell(t) = \sum_i e_{\ell i}v_i(t)$  and  $\overline{M}_{3,ji} = \sum_\ell \tilde{a}_{3,j\ell}e_{\ell i}/\sqrt{\lambda_i}(1 + i\ell\omega_0)$ . We also have

$$\begin{aligned} \delta V_0^{(1)}(t, s) &= \sum_i \sqrt{\lambda_i}(v_i(t)\overline{\delta v_{0,i}(s)} + \delta v_{0,i}(t)\overline{v_i(s)}), \\ \delta v_{0,i}(t) &= \frac{1}{2\sigma_0^2} \sum_j \overline{M}_{3,ji}(1 + \partial_t)^{-1}(\mathbf{e}_j(t)\delta m(t)) + (1 + \partial_t)^{-1}p_{1,i}(t), \end{aligned} \tag{17}$$

where we define  $\tilde{M}_{3,ji} = \sum_\ell \tilde{a}_{3,j\ell}e_{\ell i}/\sqrt{\lambda_i}(1 + i\ell\omega_0)$  with  $\tilde{a}_k(t, s) = \sum_{ij} \tilde{a}_{k,ij}\mathbf{e}_i(t)\overline{\mathbf{e}_j(s)}$ , for  $k = 1, 2$ , and  $3$ .

Since this solution leaves a residual error on the right-hand side of Eq. (14), we make corrections recursively by solving

$$(1 + \partial_t)(1 + \partial_s)\delta D_{k+1}^{(1)}(t, s) = a_1(t, s)\delta D_k^{(1)}(t, s) + a_2(t, s)\delta D_k^{(1)}(t, t) + a_2(s, t)\delta D_k^{(1)}(s, s), \quad (k \geq 0). \tag{18}$$

It is easily seen that each additional response can be represented as  $\delta D_k^{(1)}(t, s) = \sum_i \sqrt{\lambda_i}(v_i(t)\overline{\delta d_{k,i}(s)} + \delta d_{k,i}(t)\overline{v_i(s)})$ , with

$$\begin{aligned} \delta d_{k+1,i}(t) &= \frac{1}{\sqrt{\lambda_i}} \sum_{j,\ell} M_{12,j\ell i}(1 + \partial_t)^{-1}(\mathbf{e}_j(t)\delta d_{k,\ell}(t)), \\ M_{12,j\ell i} &= \sqrt{\lambda_\ell} \sum_{m,n} \left\{ \frac{\overline{v_{\ell m}a_{1,jn}\overline{e}_{n+m,i}}}{1 - i(n+m)\omega_0} + 2 \sum_{m,n} \frac{v_{\ell m}a_{2,j-m,n}\overline{e}_{ni}}{1 - in\omega_0} \right\}. \end{aligned} \tag{19}$$

In the derivation of this equation, note that  $v_\ell(t)$  and  $\delta d_{k,\ell}(t)$  are real, and that expansion  $v_\ell(t) = \sum_m v_{\ell m}\mathbf{e}_m(t)$  is used. We also have the following similar equation for corrections to  $\delta V^{(1)}(t, s)$ ,  $\delta V_k^{(1)}(t, s) = \sum_i \sqrt{\lambda_i}(v_i(t)\overline{\delta v_{k,i}(s)} + \delta v_{k,i}(t)\overline{v_i(s)})$ :

$$\begin{aligned} \delta v_{k+1,i}(t) &= \frac{1}{2\sigma_0^2\sqrt{\lambda_i}} \sum_{j,\ell} \tilde{M}_{12,j\ell i}(1 + \partial_t)^{-1}(\mathbf{e}_j(t)\delta d_{k,\ell}(t)), \\ \tilde{M}_{12,j\ell i} &= \sqrt{\lambda_\ell} \sum_{m,n} \left\{ \frac{\overline{v_{\ell m}a_{1,jn}\overline{e}_{n+m,i}}}{1 - i(n+m)\omega_0} + 2 \sum_{m,n} \frac{v_{\ell m}\tilde{a}_{2,j-m,n}\overline{e}_{ni}}{1 - in\omega_0} \right\}. \end{aligned} \tag{110}$$

The first-order response in  $V(t, s)$  is then given by

$$\delta V^{(1)}(t, s) = \sum_{k=0}^{\infty} \delta V_k^{(1)}(t, s) = \sum_{k,i} \sqrt{\lambda_i}(v_i(t)\overline{\delta v_{k,i}(s)} + \delta v_{k,i}(t)\overline{v_i(s)}) = \sum_i \sqrt{\lambda_i}(v_i(t)\overline{\delta v_i(s)} + \delta v_i(t)\overline{v_i(s)}), \tag{111}$$

if the series of  $\delta V_k^{(1)}$  converges. Here, we define  $\delta v_i(t) = \sum_k \delta v_{k,i}(t)$ .

To check the convergence and magnitude of the above first-order response, we iterate the recursion relation from different initial values and check the 1-norm of the final result. More precisely, we numerically examine  $\{\delta v_i\}_i$  for initial input  $\delta m(t) = e^{i\omega t}$  with different values of  $\omega$ . From this calculation, we estimate the value of the bounding constant  $\theta_1$ , for  $\sqrt{2}g_0 \sum_i |\xi_{o,i}| \|\widehat{\delta v_i}\|_1 \leq \theta_1 \|\widehat{\delta m}\|_1 + \sum_i \theta'_{1,i} \|\widehat{p}_{1,i}\|_1$ . For use below, we also estimate the value of the bounding constant  $\theta_2$ , for  $\sum_i \|\widehat{\delta v_i}\|_1 \leq \theta_2 \|\widehat{\delta m}\|_1 + \sum_i \theta'_{2,i} \|\widehat{p}_{1,i}\|_1$ . Here, note that the 1-norm is evaluated as the sum of the discrete components over the multiples of the basic frequency  $\omega_0$  and the continuous component over the other frequencies.

From the convergent solution, we obtain

$$V(t, s) = \sum_i (\sqrt{\lambda_i} v_i(t) + \delta v_i(t)) (\sqrt{\lambda_i} \bar{v}_i(s) + \delta \bar{v}_i(s)) + O(|\mathbf{p}|^2). \tag{I12}$$

The first-order response in  $V(t, s)$  results in

$$\eta(t) = (1 + \partial_t) \left\{ \sum_i \xi_{0,i} (\sqrt{\lambda_i} v_i(t) + \delta v_i(t)) \right\} + \delta \eta^{(2)}(t), \quad (t \geq 0), \tag{I13}$$

where  $\delta \eta^{(2)}(t)$  is the response in  $\eta(t)$  due to the higher-order responses in  $D(t, s)$ .

Next, we examine the second-order response in  $D(t, s)$ , which is denoted by  $\delta D^{(2)}(t, s)$  and obeys the following equation:

$$\begin{aligned} (1 + \partial_t)(1 + \partial_s) \delta D^{(2)}(t, s) = & a_1(t, s) \delta D^{(2)}(t, s) + a_2(t, s) \delta D^{(2)}(t, t) + a_2(s, t) \delta D^{(2)}(s, s) + b_1(t, s) \delta D^{(1)}(t, s)^2 \\ & + b_2(t, s) \delta D^{(1)}(t, s) \delta D^{(1)}(t, t) + b_2(s, t) \delta D^{(1)}(t, s) \delta D^{(1)}(s, s) + b_3(t, s) \delta D^{(1)}(t, t)^2 \\ & + b_3(s, t) \delta D^{(1)}(s, s)^2 + b_4(t, s) \delta D^{(1)}(t, t) \delta D^{(1)}(s, s) + b_5(t, s) \delta D^{(1)}(t, s) \delta m(t) \\ & + b_5(s, t) \delta D^{(1)}(t, s) \delta m(s) + b_6(t, s) \delta D^{(1)}(t, t) \delta m(t) + b_6(s, t) \delta D^{(1)}(s, s) \delta m(s) \\ & + b_7(t, s) \delta D^{(1)}(t, t) \delta m(s) + b_7(s, t) \delta D^{(1)}(s, s) \delta m(t) + b_8(t, s) \delta m(t)^2 \\ & + b_8(s, t) \delta m(s)^2 + b_9(t, s) \delta m(t) \delta m(s) + \sum_i p_{2,i}(t) p_{2,i}(s), \end{aligned} \tag{I14}$$

where coefficient functions  $b_1(t, s) \dots b_9(t, s)$  are suitably defined and the random component of the input is denoted by  $\{p_{2,i}\}_i$ . Similarly to Appendix H, we see that the second-order response consist of two components: the first, denoted by  $\delta D^{(2)'}(t, s)$ , involves second-order products of either  $\delta m(t)$  or  $\delta m(s)$  and contributes to  $O(|\mathbf{p}|^2)$  response in  $\eta$ ; the second, denoted by  $\delta D^{(2)''}(t, s)$ , involves cross-terms such as  $\delta m(t) \delta m(s)$  and newly generates  $O(|\mathbf{p}|)$  fluctuations in  $\eta$ . Then, our aim is to show that the latter part of the second-order response in  $D(t, s)$  can be represented as the sum of a series of outer products of functions of time with exponentially decreasing magnitudes, as we did in Appendix H. Note that the relevant part of the second-order response in  $V(t, s)$  is the same as that in  $D(t, s)/2\sigma_0^2$ .

The approximate special solution  $\delta D_0^{(2)}(t, s)$  for Eq. (I14) is obtained by ignoring the first three terms on the right-hand side. Let us assume that the component of this solution that contributes to  $O(|\mathbf{p}|)$  response in  $\eta(t)$  can be represented as the sum of outer products with a certain convergence property. Concretely, we assume

$$\delta D_0^{(2)}(t, s) = \sum_i q_{0,i}(t) q_{0,i}(s), \tag{I15}$$

where  $q_{0,i}(t)$  is a linear transformation of  $\delta m(t)$  or input components. We assume that the magnitude of  $q_{0,i}$  is bounded as  $\sum_i \|\widehat{q_{0,i}}\|_1 \leq c \|\widehat{\delta m}\|_1, c \sum_j \|\widehat{p_{2,j}}\|_1$  for some finite constant  $c$ . This assumption is reasonable because the nonhomogeneous terms on the right-hand side of Eq. (I14) should represent a positive semidefinite correlation matrix for newly generated fluctuations, as we have observed in the case for fixed points.

We recursively make corrections to this solution by solving

$$\begin{aligned} (1 + \partial_t)(1 + \partial_s) \delta D_{\ell+1}^{(2)}(t, s) \\ = a_1(t, s) \delta D_{\ell}^{(2)}(t, s) + a_2(t, s) \delta D_{\ell}^{(2)}(t, t) \\ + a_2(s, t) \delta D_{\ell}^{(2)}(s, s). \end{aligned} \tag{I16}$$

Since the latter two terms on the right-hand side of this equation yield only  $O(|\mathbf{p}|^2)$  contributions in  $\eta(t)$ , we ignore

these terms. Then, for  $\delta D_{\ell}^{(2)}(t, s) = \sum_j q_{\ell,j}(t) q_{\ell,j}(s)$ , we have

$$\begin{aligned} \delta D_{\ell+1}^{(2)}(t, s) = \sum_{i,j} q'_{\ell+1,i,j}(t) q'_{\ell+1,i,j}(s), \\ q'_{\ell+1,i,j}(t) = (1 + \partial_t)^{-1} \sum_j \sqrt{\rho_{1,j}} u_{1,j}(\tau) q_{\ell,i}(t), \end{aligned} \tag{I17}$$

with eigendecomposition,  $a_1(t, s) = \sum_j \rho_{1,j} u_{1,j}(t) u_{1,j}(s)$ . Thus the iteration keeps the solution positive semidefinite.

If the solutions for this recursion equation converges, we obtain the relevant part of the second-order response as  $\delta D^{(2)}(t, s) = \sum_{\ell} \delta D_{\ell}^{(2)}(t, s)$ . To evaluate the magnitude of the solution, we iteratively solve the recursion equation and see that, after a sufficient number of iterations, the solution satisfies

$$\delta D_{\ell+1}^{(2)}(t, s) \approx \theta_3 \delta D_{\ell}^{(2)}(t, s). \tag{I18}$$

If we have  $\theta_3 < 1$ , the series converges.

Using the representation in terms of outer products,  $\delta D_{\ell}^{(2)}(t, s) = \sum_i q_{\ell,i}(t) q_{\ell,i}(s)$ , we obtain the response in  $\eta$ . Combined with Eq. (I13), the overall linear response in  $\eta(t)$  reads

$$\begin{aligned} \eta(t) = (1 + \partial_t) \left\{ \sum_i \xi_{0,i} (\sqrt{\lambda_i} v_i(t) + \delta v_i(t)) \right. \\ \left. + \frac{1}{\sqrt{2}\sigma_0} \sum_{\ell,j} \xi_{\ell,j} q_{\ell,j}(t) \right\}, \end{aligned} \tag{I19}$$

where  $\{\xi_{\ell,j}\}_{\ell,j}$  are *i.i.d.* unit Gaussian random variables. Substituting the above into Eq. (I3), we obtain a self-consistent equation,

$$\begin{aligned} \delta m(t) = \sqrt{2}g_0 \sum_i \xi_{0,i} \delta v_i(t) + \frac{g_0}{\sigma_0} \sum_{\ell,j} \xi_{\ell,j} q_{\ell,j}(t) \\ + g_0(1 + \partial_t)^{-1} \delta \tilde{\phi}(t) + (1 + \partial_t)^{-1} p_0(t). \end{aligned} \tag{I20}$$

TABLE I. The values of the bounding constants that determine stability of regular oscillatory solutions are numerically estimated and summarized in a table.

Model and orbit		$\theta_1$	$\theta_2$	$\theta_3$	$\theta_4$	$\theta_1 + g_0\theta_4$
untuned w/o input	Fig. 6(g)	0.159	–	0.487	0	0.159
finely tuned w/o input	Fig. 15(c)	$3.81 \times 10^{-3}$	–	1.02	0	$3.81 \times 10^{-3}$
finely tuned with input	Fig. 10(c)	$4.36 \times 10^{-3}$	$5.83 \times 10^{-4}$	0.896	$3.32 \times 10^{-5}$	$4.39 \times 10^{-3}$

The norms of  $\delta v_i(t)$  and  $q_{\ell,j}(t)$  were evaluated by comparison with the norm of  $\delta m(t)$ , as described above. For  $\delta\tilde{\phi}(t)$ , by putting  $\langle\phi''(h_j^{(E)}(t))\rangle_{j\notin S} - \langle\phi''(h_j^{(E)}(t))\rangle_{j\in S} = \sum_{\ell} \phi_{\ell}^{(2)} \mathbf{e}_{\ell}(t)$  and  $\langle\phi'(h_j^{(E)}(t))\rangle_{j\notin S} - \langle\phi'(h_j^{(E)}(t))\rangle_{j\in S} = \sum_{\ell} \phi_{\ell}^{(1)} \mathbf{e}_{\ell}(t)$ , we have

$$\begin{aligned} \delta\tilde{\phi}(t) &= \sum_{\ell} \phi_{\ell}^{(2)} \mathbf{e}_{\ell}(t) \delta D^{(1)}(t, t) \\ &+ \sum_{\ell} \phi_{\ell}^{(1)} \mathbf{e}_{\ell}(t) \delta m(t) + O(|\mathbf{p}|^2), \\ \|(1 + \partial_t)^{-1} \delta\tilde{\phi}\|_1 &\leq \theta_4 \|\widehat{\delta m}\|_1, \end{aligned} \quad (I21)$$

with

$$\begin{aligned} \theta_4 &= \theta_2 \sup_{i, \omega'} \left| \sum_{\ell, n} \frac{\sqrt{\lambda_i} \phi_{\ell}^{(2)} v_{in}}{1 + i((\ell + n)\omega + \omega')} \right| \\ &+ \sup_{\omega'} \left| \sum_{\ell} \frac{\phi_{\ell}^{(1)}}{1 + i(\omega' + \ell\omega)} \right|. \end{aligned} \quad (I22)$$

From this equation, we obtain the overall response in  $m(t)$  by recursive computation. We first calculate  $\delta m^{(0)}(t)$  from Eq. (I20), on the right-hand side of which  $\delta v_i$  and  $q_{\ell,j}$  are computed just from the input  $\mathbf{p}$  by setting  $\delta m = 0$ . Then, by calculating functions  $\delta v_i^{(k)}$ ,  $q_{\ell,j}^{(k)}$  and  $\delta\tilde{\phi}^{(k)}(t)$  from  $\delta m^{(k)}(t)$ , ( $k \geq 0$ ) according to Eqs. (I7), (I9), (I15), (I17), and (I21),

we make corrections recursively:

$$\begin{aligned} \delta m^{(k+1)}(t) &= \sqrt{2} g_0 \sum_i \xi_{\alpha,i} \delta v_i^{(k)}(t) + \frac{g_0}{\sigma_0} \sum_{\ell,j} \xi_{\ell,j} q_{\ell,j}^{(k)}(t) \\ &+ g_0(1 + \partial_t)^{-1} \delta\tilde{\phi}^{(k)}(t). \end{aligned} \quad (I23)$$

We finally obtain  $\delta m(t) = \sum_{k=0}^{\infty} \delta m^{(k)}(t)$ , which satisfies Eq. (I20), if  $\|\delta m^{(k+1)}\|_1 \leq \theta \|\delta m^{(k)}\|_1$  holds for some constant  $\theta < 1$ . This convergence condition is satisfied with a nonzero probability if

$$\theta_1 + g_0\theta_4 < 1, \quad \theta_3 < 1. \quad (I24)$$

Similarly to Appendix H, this convergence implies the linear stability of the regular oscillation under examination for a nonzero fraction of configurations of the random connectivity. A single configuration of the random connectivity corresponds to a single set of values for  $\{\xi_{\ell,j}\}_{\ell,j}$ .

We numerically calculate the bounding constants in the above condition for convergence, using the orbits of the mean activity observed in Figs. 15(c), 6(g), and 10(c) for  $m_o(t)$ . To reduce the computational cost, we ignore eigenmodes  $\{v_i(t)\}_i$  except for those for the 48 largest eigenvalues. The calculated values are summarized in Table I, which suggest that the regular oscillations observed in Figs. 6(g) and 10(c) are stable, while those observed in Fig. 15(c) are not.

## APPENDIX J: PRICE THEOREM AND DIFFERENTIATION OF CORRELATION MATRIX $\tilde{C}$

We have used the differentiation of correlation matrix  $\tilde{C}$  with respect to  $D$  and  $m$  in Appendix G–I. This is given by the following formula:

$$\begin{aligned} \frac{\partial \tilde{C}(t_1, t_2)}{\partial D_{11}} &= \frac{1}{2} \langle \phi''(z_1) \phi(z_2) \rangle, & \frac{\partial \tilde{C}(t_1, t_2)}{\partial D_{12}} &= \langle \phi'(z_1) \phi'(z_2) \rangle, & \frac{\partial \tilde{C}(t_1, t_2)}{\partial D_{22}} &= \frac{1}{2} \langle \phi(z_1) \phi''(z_2) \rangle, \\ \frac{\partial \tilde{C}(t_1, t_2)}{\partial m_1} &= \langle \phi'(z_1) \phi(z_2) \rangle, & \frac{\partial \tilde{C}(t_1, t_2)}{\partial m_2} &= \langle \phi(z_1) \phi'(z_2) \rangle, & \frac{\partial \bar{\phi}(t_1)}{\partial D_{11}} &= \frac{1}{2} \langle \phi''(z_1) \rangle, & \frac{\partial \bar{\phi}(t_1)}{\partial D_{12}} &= \frac{\partial \bar{\phi}(t_1)}{\partial D_{22}} = 0, \\ \frac{\partial \bar{\phi}(t_1)}{\partial m_1} &= \langle \phi'(z_1) \rangle, & \frac{\partial \bar{\phi}(t_1)}{\partial m_2} &= 0, & \frac{\partial \bar{\phi}(t_2)}{\partial m_2} &= \langle \phi'(z_2) \rangle, & \frac{\partial \bar{\phi}(t_2)}{\partial m_1} &= 0. \end{aligned} \quad (J1)$$

The variables  $\{z_{\alpha}\}_{\alpha=1,2}$  are Gaussian random variables that have the same first and second-order moments as  $\{h_j^{(\ell)}(t_{\alpha})\}_{\alpha=1,2}$  ( $\ell = E, I$ ), similarly to Eq. (F5). Also note that the above partial derivatives with respect to  $D_{\alpha\beta}$  are not total differentials with respect to  $D(t_{\alpha}, t_{\beta})$  but are derivatives with respect to the corresponding variables that appear in Eq. (F5). The twice differentiation is performed similarly and easily inferred from the above results. The differentiation of  $C$  is easily obtained from the results for the differentiation of  $\tilde{C}$  and  $\bar{\phi}$ . The above results can be derived by the differentiation of both sides of Eq. (F5) followed by integration by parts, but are obtained more easily by performing the differentiation in the frequency domain in a manner similar to the derivation of Price theorem [85] (also see [83] for a simpler example). We show the outline of how it is performed below.



Let us rewrite the representation of  $\tilde{C}$  in Eq. (F5) as

$$\begin{aligned}\tilde{C}(t_1, t_2) &= \int_{\mathbf{R}^3} \phi(f_1(\{D_{\alpha\beta}\})y_1 + f_2(\{D_{\alpha\beta}\})w + m_1)\phi(f_3(\{D_{\alpha\beta}\})y_2 + f_4(\{D_{\alpha\beta}\})w + m_2)d\mathcal{N}(y_1)d\mathcal{N}(y_2)d\mathcal{N}(w) \\ &= \frac{1}{2\pi} \int_{\mathbf{R}^5} \hat{\phi}(\omega_1)\hat{\phi}(\omega_2)e^{i\omega_1((f_1(\{D_{\alpha\beta}\})y_1 + f_2(\{D_{\alpha\beta}\})w + m_1) + i\omega_2(f_3(\{D_{\alpha\beta}\})y_2 + f_4(\{D_{\alpha\beta}\})w + m_2))}d\mathcal{N}(y_1)d\mathcal{N}(y_2)d\mathcal{N}(w)d\omega_1d\omega_2 \\ &= \frac{1}{2\pi} \int_{\mathbf{R}^2} \hat{\phi}(\omega_1)\hat{\phi}(\omega_2)e^{-\frac{1}{2}\omega_1^2 f_1(\{D_{\alpha\beta}\})^2 - \frac{1}{2}\omega_2^2 f_3(\{D_{\alpha\beta}\})^2 - \frac{1}{2}(\omega_1 f_2(\{D_{\alpha\beta}\}) + \omega_2 f_4(\{D_{\alpha\beta}\}))^2 + i\omega_1 m_1 + i\omega_2 m_2}d\omega_1d\omega_2,\end{aligned}\quad (J2)$$

where functions  $f_1$ – $f_4$  of  $\{D_{\alpha\beta}\}_{\alpha,\beta=1,2}$  are suitably defined. The Fourier transform is denoted by  $\hat{\cdot}$ . From the second line to the third line, we have used the characteristic function of a unit Gaussian measure,  $e^{-\frac{1}{2}\omega^2} = \int e^{i\omega X}d\mathcal{N}(X)$ . Differentiating the last line with respect to, e.g.  $D_{12}$ , and rearranging terms, we have

$$\begin{aligned}\frac{\partial \tilde{C}(t_1, t_2)}{\partial D_{12}} &= -\frac{1}{2\pi} \int_{\mathbf{R}^2} \omega_1 \omega_2 \hat{\phi}(\omega_1)\hat{\phi}(\omega_2)e^{-\frac{1}{2}\omega_1^2 f_1(\{D_{\alpha\beta}\})^2 - \frac{1}{2}\omega_2^2 f_3(\{D_{\alpha\beta}\})^2 - \frac{1}{2}(\omega_1 f_2(\{D_{\alpha\beta}\}) + \omega_2 f_4(\{D_{\alpha\beta}\}))^2 + i\omega_1 m_1 + i\omega_2 m_2}d\omega_1d\omega_2 \\ &= \frac{1}{2\pi} \int_{\mathbf{R}^2} \hat{\phi}'(\omega_1)\hat{\phi}'(\omega_2)e^{-\frac{1}{2}\omega_1^2 f_1(\{D_{\alpha\beta}\})^2 - \frac{1}{2}\omega_2^2 f_3(\{D_{\alpha\beta}\})^2 - \frac{1}{2}(\omega_1 f_2(\{D_{\alpha\beta}\}) + \omega_2 f_4(\{D_{\alpha\beta}\}))^2 + i\omega_1 m_1 + i\omega_2 m_2}d\omega_1d\omega_2 \\ &= \int_{\mathbf{R}^3} \phi'(f_1(\{D_{\alpha\beta}\})y_1 + f_2(\{D_{\alpha\beta}\})w + m_1)\phi'(f_3(\{D_{\alpha\beta}\})y_2 + f_4(\{D_{\alpha\beta}\})w + m_2)d\mathcal{N}(y_1)d\mathcal{N}(y_2)d\mathcal{N}(w) \\ &= \langle \phi'(z_1)\phi'(z_2) \rangle.\end{aligned}\quad (J3)$$

From the first line to the second line, we have used the formula for the Fourier transform of derivatives,  $\hat{\phi}'(\omega) = i\omega\hat{\phi}(\omega)$ . Other differentiations in Eq. (J1) are calculated similarly.

- 
- [1] J. M. Fuster and G. E. Alexander, Neuron activity related to short-term memory, *Science* **173**, 652 (1971).
- [2] S. Funahashi, M. V. Chafee, and P. S. Goldman-Rakic, Prefrontal neuronal activity in rhesus monkeys performing a delayed anti-saccade Task, *Nature (London)* **365**, 753 (1993).
- [3] R. Romo, C. D. Brody, A. Hernández, and L. Lemus, Neuronal correlates of parametric working memory in the prefrontal cortex, *Nature (London)* **399**, 470 (1999).
- [4] T. T. G. Hahn, J. M. McFarland, S. Berberich, B. Sakmann, and M. R. Mehta, Spontaneous persistent activity in entorhinal cortex modulates cortico-hippocampal interaction in vivo, *Nat. Neurosci.* **15**, 1531 (2012).
- [5] Z. V. Guo, H. K. Inagaki, K. Daie, S. Druckmann, C. R. Gerfen, and K. Svoboda, Maintenance of persistent activity in a frontal thalamocortical loop, *Nature (London)* **545**, 181 (2017).
- [6] J. A. Hobson and M. Steriade, The Neuronal basis of behavioral state control, in *Handbook of Physiology, The Nervous System, Intrinsic Regulatory Systems of the Brain* (American Physiological Society, Bethesda, MD, USA, 1986), pp. 701–823.
- [7] W. Klimesch, EEG alpha and theta oscillations reflect cognitive and memory performance: A review and analysis, *Brain Res. Rev.* **29**, 169 (1999).
- [8] G. Buzsáki, Theta oscillations in the hippocampus, *Neuron* **33**, 325 (2002).
- [9] O. Jensen, J. Kaiser, and J.-P. Lachaux, Human gamma-frequency oscillations associated with attention and memory, *Trends Neurosci.* **30**, 317 (2007).
- [10] C. M. Gray and W. Singer, Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex, *Proc. Natl. Acad. Sci. USA* **86**, 1698 (1989).
- [11] J. O’Keefe and M. L. Recce, Phase relationship between hippocampal place units and the EEG theta rhythm, *Hippocampus* **3**, 317 (1993).
- [12] A. K. Engel, P. Fries, and W. Singer, Dynamic predictions: Oscillations and synchrony in top-down processing, *Nat. Rev. Neurosci.* **2**, 704 (2001).
- [13] K. D. Harris, D. A. Henze, H. Hirase, X. Leinekugel, G. Dragoi, A. Czurkó, and G. Buzsáki, Spike train dynamics predicts theta-related phase precession in hippocampal pyramidal cells, *Nature (London)* **417**, 738 (2002).
- [14] P. Fries, D. Nikolić, and W. Singer, The gamma cycle, *Trends Neurosci.* **30**, 309 (2007).
- [15] J. F. A. Poulet and C. C. H. Petersen, Internal brain state regulates membrane potential synchrony in barrel cortex of behaving mice, *Nature (London)* **454**, 881 (2008).
- [16] J. Lisman, Working memory: The importance of theta and gamma oscillations, *Curr. Biol.* **20**, R490 (2010).
- [17] D. Strüber, S. Rach, S. A. Trautmann-Lengsfeld, A. K. Engel, and C. S. Herrmann, Antiphase 40hz oscillatory current stimulation affects bistable motion perception, *Brain Topography* **27**, 158 (2014).
- [18] T. Gulati, D. S. Ramanathan, C. C. Wong, and K. Ganguly, Reactivation of emergent task-related ensembles during slow-wave sleep after neuroprosthetic learning, *Nat. Neurosci.* **17**, 1107 (2014).
- [19] H. Sompolinsky, A. Crisanti, and H. J. Sommers, Chaos in Random Neural Networks, *Phys. Rev. Lett.* **61**, 259 (1988).
- [20] S.-I. Amari, Learning patterns and pattern sequences by self-organizing nets of threshold elements, *IEEE Trans. Comput. C-21*, 1197 (1972).
- [21] J. J. Hopfield, Neural networks and physical systems with emergent collective computational abilities, *Proc. Natl. Acad. Sci. USA* **79**, 2554 (1982).
- [22] D. J. Amit, *Modeling Brain Function: The World of Attractor Neural Networks* (Cambridge University Press, Cambridge, UK, 1989).

- [23] I. Ginzburg and H. Sompolinsky, Theory of correlations in stochastic neural networks, *Phys. Rev. E* **50**, 3171 (1994).
- [24] C. van Vreeswijk and H. Sompolinsky, Chaotic balanced state in a model of cortical circuits, *Neural Comput.* **10**, 1321 (1998).
- [25] N. Brunel, Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons, *J. Comput. Neurosci.* **8**, 183 (2000).
- [26] O. Faugeras, J. Touboul, and B. Cessac, A constructive mean-field analysis of multi-population neural networks with random synaptic weights and stochastic inputs, *Front. Comput. Neurosci.* **3**, 1 (2009).
- [27] A. Renart, J. De La Rocha, P. Bartho, L. Hollender, N. Parga, A. Reyes, and K. D. Harris, The asynchronous state in cortical circuits, *Science* **327**, 587 (2010).
- [28] L. Zhao, B. Beverlin 2nd, T. Netoff, and D. Q. Nykamp, Synchronization from second order network connectivity statistics, *Front. Comput. Neurosci.* **5**, 28 (2011).
- [29] K. Rajan, L. F. Abbott, and H. Sompolinsky, Stimulus-dependent suppression of chaos in recurrent neural networks, *Phys. Rev. E* **82**, 011903 (2010).
- [30] R. Rosenbaum and B. Doiron, Balanced Networks of Spiking Neurons with Spatially Dependent Recurrent Connections, *Phys. Rev. X* **4**, 021039 (2014).
- [31] M. Stern, H. Sompolinsky, and L. F. Abbott, Dynamics of random neural networks with bistable units, *Phys. Rev. E* **90**, 062710 (2014).
- [32] J. Kadmon and H. Sompolinsky, Transition to Chaos in Random Neuronal Networks, *Phys. Rev. X* **5**, 041030 (2015).
- [33] D. Dahmen, H. Bos, and M. Helias, Correlated Fluctuations in Strongly Coupled Binary Networks Beyond Equilibrium, *Phys. Rev. X* **6**, 031024 (2016).
- [34] D. Martí, N. Brunel, and S. Ostojic, Correlations between synapses in pairs of neurons slow down dynamics in randomly connected neural networks, *Phys. Rev. E* **97**, 062314 (2018).
- [35] C. van Vreeswijk and H. Sompolinsky, Chaos in neuronal networks with balanced excitatory and inhibitory activity, *Science* **274**, 1724 (1996).
- [36] W. Maass, T. Natschläger, and H. Markram, Real-time computing without stable states: A new framework for neural computation based on perturbations, *Neural Comput.* **14**, 2531 (2002).
- [37] H. Jaeger and H. Haas, Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication, *Science* **304**, 78 (2004).
- [38] D. Verstraeten, B. Schrauwen, D. Stroobandt, and J. Van Campenhout, Isolated word recognition with the liquid state machine: A case study, *Inf. Proc. Lett.* **95**, 521 (2005).
- [39] U. R. Karmarkar and D. V. Buonomano, Timing in the absence of clocks: Encoding time in neural network states, *Neuron* **53**, 427 (2007).
- [40] D. Sussillo and L. F. Abbott, Generating coherent patterns of activity from chaotic neural networks, *Neuron* **63**, 544 (2009).
- [41] L. Appeltant, M. C. Soriano, G. Van der Sande, J. Danckaert, S. Massar, J. Dambre, B. Schrauwen, C. R. Mirasso, and I. Fischer, Information processing using a single dynamical node as complex system, *Nat. Commun.* **2**, 468 (2011).
- [42] M. Lukoševičius, H. Jaeger, and B. Schrauwen, Reservoir computing trends, *KI-Künstliche Intelligenz* **26**, 365 (2012).
- [43] P. F. Dominey, Recurrent temporal networks and language acquisition—from corticostriatal neurophysiology to reservoir computing, *Front. Psychol.* **4**, 1 (2013).
- [44] P. Enel, E. Procyk, R. Quilodran, and P. F. Dominey, Reservoir computing properties of neural dynamics in prefrontal cortex, *PLoS Comput. Biol.* **12**, e1004967 (2016).
- [45] J. S. Anderson, M. Carandini, and D. Ferster, Orientation tuning of input conductance, excitation, and inhibition in cat primary visual cortex, *J. Neurophysiol.* **84**, 909 (2000).
- [46] Y. Shu, A. Hasenstaub, and D. A. McCormick, Turning on and off recurrent balanced cortical activity, *Nature (London)* **423**, 288 (2003).
- [47] B. Haider, A. Duque, A. R. Hasenstaub, and D. A. McCormick, Neocortical network activity in vivo is generated through a dynamic balance of excitation and inhibition, *J. Neurosci.* **26**, 4535 (2006).
- [48] B. V. Atallah and M. Scanziani, Instantaneous modulation of gamma oscillation frequency by balancing excitation with inhibition, *Neuron* **62**, 566 (2009).
- [49] M. Okun and I. Lampl, Instantaneous correlation of excitation and inhibition during ongoing and sensory-evoked activities, *Nat. Neurosci.* **11**, 535 (2008).
- [50] J. Barral and A. D. Reyes, Synaptic scaling rule preserves excitatory–inhibitory balance and salient neuronal network dynamics, *Nat. Neurosci.* **19**, 1690 (2016).
- [51] R. Darshan, W. E. Wood, S. Peters, A. Leblois, and D. Hansel, A canonical neural mechanism for behavioral variability, *Nat. Commun.* **8**, 15415 (2017).
- [52] R. Rosenbaum, M. A. Smith, A. Kohn, J. E. Rubin, and B. Doiron, The spatial structure of correlated neuronal variability, *Nat. Neurosci.* **20**, 107 (2017).
- [53] M. London, A. Roth, L. Beeren, M. Häusser, and P. E. Latham, Sensitivity to perturbations in vivo implies high noise and suggests rate coding in cortex, *Nature (London)* **466**, 123 (2010).
- [54] S. N. Chettih and C. D. Harvey, Single-neuron perturbations reveal feature-specific competition in V1, *Nature (London)* **567**, 334 (2019).
- [55] H. Nishimori and G. Ortiz, *Elements of Phase Transitions and Critical Phenomena* (Oxford University Press, Oxford, UK, 2010).
- [56] A. Altland and B. D. Simons, *Condensed Matter Field Theory* (Cambridge University Press, Cambridge, UK, 2010).
- [57] L. Cocchi, L. L. Gollo, A. Zalesky, and M. Breakspear, Criticality in the brain: A synthesis of neurobiology, models and cognition, *Prog. Neurobiol.* **158**, 132 (2017).
- [58] B. A. Olshausen and D. J. Field, Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision Res.* **37**, 3311 (1997).
- [59] B. A. Olshausen and D. J. Field, Sparse coding of sensory inputs, *Curr. Opin. Neurobiol.* **14**, 481 (2004).
- [60] K. Rajan and L. F. Abbott, Eigenvalue Spectra of Random Matrices for Neural Networks, *Phys. Rev. Lett.* **97**, 188104 (2006).
- [61] T. Tao, Outliers in the spectrum of iid matrices with bounded rank perturbations, *Probab. Theory Relat. Fields* **155**, 231 (2013).
- [62] M. Stern and L. Abbott, Dynamics of rate-model networks with separate excitatory and inhibitory populations, *The Annual*

- Meeting of the Society for Neuroscience* (Society for Neuroscience, SanDiego, CA, USA, 2016).
- [63] S. Ostoic, Two types of asynchronous activity in networks of excitatory and inhibitory spiking neurons, *Nat. Neurosci.* **17**, 594 (2014).
- [64] L. C. García del Molino, K. Pakdaman, J. Touboul, and G. Wainrib, Synchronization in random balanced networks, *Phys. Rev. E* **88**, 042824 (2013).
- [65] I. D. Landau and H. Sompolinsky, Coherent chaos in a recurrent neural network with structured connectivity, *PLoS Comput. Biol.* **14**, e1006309 (2018).
- [66] L. Molgedey, J. Schuchhardt, and H. G. Schuster, Suppressing Chaos in Neural Networks by Noise, *Phys. Rev. Lett.* **69**, 3717 (1992).
- [67] J. Schuecker, S. Goedeke, and M. Helias, Optimal Sequence Memory in Driven Random Networks, *Phys. Rev. X* **8**, 041029 (2018).
- [68] N. Bertschinger and T. Natschläger, Real-time computation at the edge of chaos in recurrent neural networks, *Neural Comput.* **16**, 1413 (2004).
- [69] R. Legenstein and W. Maass, What makes a dynamical system computationally powerful, in *New Directions in Statistical Signal Processing: From Systems to Brains* (MIT Press, Cambridge, MA, USA, 2007), pp. 127–154.
- [70] J. Boedecker, O. Obst, J. T. Lizier, N. M. Mayer, and M. Asada, Information processing in echo state networks at the edge of chaos, *Theo. Biosci.* **131**, 205 (2012).
- [71] R. Laje and D. V. Buonomano, Robust timing and motor patterns by taming chaos in recurrent neural networks, *Nat. Neurosci.* **16**, 925 (2013).
- [72] R. Pyle and R. Rosenbaum, Spatiotemporal Dynamics and Reliable Computations in Recurrent Spiking Neural Networks, *Phys. Rev. Lett.* **118**, 018103 (2017).
- [73] G. Buzsáki, R. D. Traub, and T. A. Pedley, The Cellular Basis of EEG Activity, in *Current Practice of Clinical Electroencephalography*, 3rd edition (Lippincott Williams & Wilkins, Philadelphia, PA, USA, 2003), pp. 1–11.
- [74] J. T. Trachtenberg, B. E. Chen, G. W. Knott, G. Feng, J. R. Sanes, E. Welker, and K. Svoboda, Long-term in vivo imaging of experience-dependent synaptic plasticity in adult cortex, *Nature (London)* **420**, 788 (2002).
- [75] R. Engelken, F. Farkhooi, D. Hansel, C. van Vreeswijk, and F. Wolf, A reanalysis of “Two types of asynchronous activity in networks of excitatory and inhibitory spiking neurons” [version 1; peer review: 2 approved], *F1000Research* **5**, 2043 (2016).
- [76] G. M. Shepherd, *The Synaptic Organization of the Brain*, 5th edition (Oxford University Press, Oxford, UK, 2004).
- [77] L. E. Russell, Z. Yang, L. Pei Tan, M. Fişek, A. M. Packer, H. W. P. Dalgleish, S. Chettih, C. D. Harvey, and M. Häusser, The influence of visual cortex on perception is modulated by behavioural state, bioRxiv, doi:10.1101/706010.
- [78] M. Galassi, J. Davies, J. Theiler, B. Gough, G. Jungman, P. Alken, M. Booth, F. Rossi, and R. Ulerich, *GNU Scientific Library Release 2.6* (Network Theory Ltd., 2019), <http://www.gnu.org/software/gsl/>.
- [79] M. A. Buice and J. D. Cowan, Field-theoretic approach to fluctuation effects in neural networks, *Phys. Rev. E* **75**, 051919 (2007).
- [80] M. A. Buice and C. C. Chow, Dynamic finite size effects in spiking neural networks, *PLoS Comput. Biol.* **9**, e1002872 (2013).
- [81] C. C. Chow and M. A. Buice, Path integral methods for stochastic differential equations, *J. Math. Neurosci.* **5**, 8 (2015).
- [82] J. A. Hertz, Y. Roudi, and P. Sollich, Path integral methods for the dynamics of stochastic and disordered systems, *J. Phys. A: Math. Theor.* **50**, 033001 (2017).
- [83] J. Schuecker, S. Goedeke, D. Dahmen, and M. Helias, Functional methods for disordered neural networks, [arXiv:1605.06758v2](https://arxiv.org/abs/1605.06758v2).
- [84] H. Sompolinsky and A. Zippelius, Relaxational dynamics of the Edwards-Anderson model and the mean-field theory of spin-glasses, *Phys. Rev. B* **25**, 6860 (1982).
- [85] A. Papoulis, *Probability, Random Variables, Stochastic Processes*, 3rd edition (McCraw-Hill, New York, 1991).