

博士論文（要約）

Doctoral Thesis

Flexible Information Processing with Nonlinear Dynamics in Neural Networks

（ニューラルネットワークにおける非線形
ダイナミクスを用いた柔軟的な情報処理）

48-147207 徐 牧原 Xu Muyuan

Supervisor Professor 合原 一幸 Aihara Kazuyuki

Department of Mathematical Informatics

Graduate School of Information Science and Technology

the University of Tokyo

Abstract

The ability to process information in a flexible manner is crucial for an animal to adapt to the ever-changing environment. Although intensive efforts have been made to understand the neural mechanisms behind flexible information processing, little is known about how this is achieved in the brain. In this thesis, we explore several neural mechanisms that may contribute to this ability, through theoretical, as well as experimental studies.

Specifically, in the first study, we investigate the dynamics and stability of associative memory, by incorporating sparse coding and nonlinear, short-term synaptic dynamics. Using a mean field technique, we obtain the detailed bifurcation structure in stability and investigate how various dynamics depend on the sparseness of memory patterns, the time constants of synaptic dynamics and the strength of a negative feedback to the associative memory network. When the memory patterns become sparse, the appearance of spurious states is shown to hamper the performance of memory retrieval. However, we demonstrate that this can be compensated by applying an appropriate negative feedback. Furthermore, the oscillatory states induced by the short-term synaptic dynamics are found to depend on the sparseness of memory patterns. These results suggest how the encoding and synaptic properties affect the information storage in neural networks.

In the second study, we analyze the change of dynamics in random neural networks under the framework of reservoir computing. In particular, we focus on the influence of the statistical properties of external input. Using a mean field model and numerical methods, we obtain the critical points where the transition between the fixed-point state and the chaotic state occurs. In addition, we show that the statistical properties other than the first two moments of the external input have only inferior effect on the dynamics. The results can be used to design reservoirs in modeling and applications.

In the third study, we analyze data recorded from the prefrontal cortex of monkeys during a working memory task. The task requires the monkey to

perform logical computation which associates the outcome of each trial with a visual cue in a flexible way depending on the implicitly given context. We demonstrate that the information processing is underlain by a dynamical representational switching from the context to the behavioral relevant information. Based on our observations, we hypothesize that the prefrontal cortex actively maintains and integrates task-relevant information by separated neural populations, and propose a neural network model in which the neural populations are organized through feedforward, as well as recurrent structures. We show that this model can qualitatively reproduce the results from recorded data by using simulations with noisy integrate-and-fire neurons. Our results suggest that the feedforward and recurrent structures in the prefrontal cortex are crucial for the flexible information processing in the prefrontal cortex.

Collectively, our studies provide an insight into how various neural mechanisms contribute to the flexible information processing in neural networks.

Contents

Chapter 1	Introduction	1
Chapter 2	Stability of Associative Memory with Short-Term Synaptic Dynamics	7
Chapter 3	External Input-Forced Onset of Chaos in Random Neural Networks	9
3.1	Introduction	9
3.2	A Random Neural Network with Two Populations . . .	11
3.3	Discussion	18
3.4	Conclusion	20
Chapter 4	Logical Computation in Monkey's Prefrontal Cortex	21
Chapter 5	Conclusion	23
	Acknowledgments	25
	Bibliography	27

Chapter 1

Introduction

An animal is exposed to extensive information in every single moment of its life. How does the brain process such information in a flexible way is crucial for the animal to adapt to the ever-changing environment. Past studies have revealed a very important aspect of this processing, i.e. the large number of neurons involved in it. Although a single neuron is complex enough to exhibit sophisticated behaviors depending on its anatomy and physiology, neural networks provide a much richer repertoire of dynamics that enables cognitive tasks such as pattern recognition and classification. Therefore possible neural mechanisms underlying information processing in neural networks are of particular interest in neuroscience.

The processing of information depends critically on two aspects: transmission, and storage. On a cellular level, information is transmitted by action potentials. The membrane of a neuron selectively allows ions to pass what is known as ion channels through diffusion, trying to maintain a potential difference called the reversal potential. On the other hand, the net charge between the interior and exterior sides of the membrane, referred to as the membrane potential, depends on the concentrations of various types of ions. In the steady state, this potential balance the diffusion forces for each type of ions. This homeostasis process, which allows the neuron to behave like an electric battery, is first studied by Alan Hodgkin and Andrew Huxley (1952). In their model, the total current I through the membrane is described as

$$I = C_m \frac{dV_m}{dt} + g_K(V_m - V_K) + g_{Na}(V_m - V_{Na}) + g_L(V_m - V_L), \quad (1.1)$$

where C_m is the capacitance of the membrane, and V_m is the membrane potential. V_K and V_{Na} are the reversal potentials of potassium and sodium ions,

respectively, which are the two main types of ions in neural signaling. V_L is the reversal potential of the leakage current, whose conductance g_L is a constant. g_K and g_{Na} are the conductances of the currents induced by potassium and sodium ions. These quantities are associated to the opening and closing of corresponding ion channels, which in turn depend on the membrane potential. In particular, typical values for V_K and V_{Na} are -80 mV and 58 mV. In the steady (resting) state, the membrane potential is maintained near -70 mV, the sodium ions have a higher concentration on the exterior side and the potassium ions have a higher concentration on the interior side of the membrane (Rojas, 1996). An action potential is produced when the membrane potential is depolarized, resulting in the opening of sodium ion channels (increased value for g_{Na}) in a local area on the membrane. The influx of sodium ions causes an explosive increase in the membrane potential, and is followed by the opening of potassium ion channels (increased value for g_K) which leads to the repolarization and hyperpolarization of the membrane. This local process triggers the opening of ion channels in its adjacent area, and therefore the action potential (impulse) is transmitted along the axon. At the (chemical) synapse, the action potential is converted into the release of neurotransmitters, which in turn depolarize or polarize the membrane of the postsynaptic neuron by binding to receptors on it (**Figure 1.1**). In contrast to this, how information is transmitted on the network level is still not fully understood. In particular, what code is adopted by the neurons remains controversial. A prominent feature of neural activity is its irregularity. Because the duration of action potentials is short compared to other neuronal events, they are often treated as point events in time. The problem is, the spike train, which refers to the sequence of action potentials of a neuron, often appears random, reminiscent of a Poisson point process (Softky & Koch, 1993; Stein, 2005). So how does the neuron transmit information when its activity is not reliable? A well-established view is that the information is encoded by the number of action potentials emitted in a short period of time, i.e. the firing rate. Although recent evidence suggests that the timing of action potentials may be more reliable than has been thought (Mainen and Sejnowski, 1995), and the temporal features of the precise timing may also contribute to information transmission (O’Keefe and Recce, 1993; Petersen et al., 2002), rate-based model is still an often-used, very effective tool to understand the information processing in neural networks, since it describes the information transmission at single neurons with simple,

smooth dynamics such as

$$\frac{dx_i}{dt} = -\frac{x_i}{\tau} + \sum_j J_{ij}r_j + h_i, \quad (1.2)$$

$$r_i = f(x_i), \quad (1.3)$$

which is well-suited for analysis. In Eqs. (1.2) and (1.3), x_i and r_i denote the membrane potential (activation) and firing rate of the i th neuron, respectively. The second term and the third term on the right hand side of Eq. (1.2) correspond to the recurrent and external synaptic input to the neuron, whose function is to integrate the input into the firing rate through a nonlinear, monotonic function f . The temporal scale of this dynamics is given by the time constant τ .

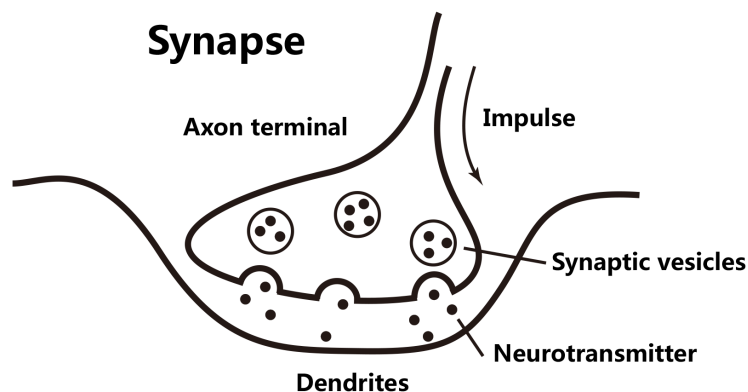


Figure 1.1: **Neural signaling at a chemical synapse.** When an action potential (impulse) arrives at the presynaptic axon terminal, neurotransmitters are released into the cleft between the pre- and postsynaptic neurons. These chemicals then bind to certain receptors on the postsynaptic membrane, and modulate the membrane potential of the postsynaptic neuron.

On the other hand, the study of the storage of information begins when Donald Hebb proposed his celebrated learning rule (Hebb, 1949). The theory explains how information processing in neural networks is affected by experience in an unsupervised way, and lies the foundation of associative learning. More importantly, in support of Ramon y Cajal's idea, it identifies synaptic efficacy, and plasticity that modifies the synaptic efficacy as the key for information storage. Later on, the theory is underpinned by the discovery of long-term potentiation, long-term depression (Lømo, 1966; Bliss & Lømo, 1973) and spike timing-dependent plasticity (Levy and Steward 1983; Debanne et al., 1994; Markram et al., 1997). From the computational point of view, this rule of

associative learning facilitated the development of attractor neural networks. In particular, the associative memory network proposed by Hopfield (1982) provides a general explanation of how information could be stored in recurrent neural networks. The storage of information is associated with stable states, referred to as attractors, in the intrinsic dynamics of the network. Following Hopfield's work, a number of neural network models have explored such stable states in various contexts. These models are of particular importance in the study of a central concept in cognition, known as the working memory. Working memory (WM) refers to the cognitive function that enables temporal holding and manipulation of information (Barak et al., 2013), and is crucial for more complex cognitive functions such as planning, reasoning, and comprehension (Baddeley and Hitch, 1974; Baddeley, 1986). In particular, neurophysiological studies showed that after the removal of stimulus, neural activity in certain brain areas still remain to be stimulus-specific, suggesting that this persistent stimulus-selective activity might be the internal neural correlate of the stimulus (Goldman-Rakic, 1987; Miyashita and Chang, 1988; Funahashi et al., 1989, 1990; Romo et al., 1999, 2002). In this context, the attractor neural networks provide a well-suited mechanism to explain how such activity is sustained in these areas. Specifically, when the stimulus information is of discrete nature, the network is assumed to possess a discrete set of attractors in the paradigm of Hopfield network (Hopfield, 1984; Amit & Brunel, 1997). When the stimulus information is of continuous nature, the network is assumed to possess a continuous set of attractors (Ben-Yishai et al., 1995; Seung, 1998; Hansel and Sompolinsky, 1998; Compte et al., 2000). These models have led to fruitful results on the way information is processed in the brain. In particular, a theory based on continuous attractor neural networks has been successful in explaining how the hippocampal-entorhinal system maintains a cognitive map for path integration (Samsonovich & McNaughton, 1997; McNaughton et al., 2006).

While the information processing based on attractor neural networks appears natural and robust (Wang, 2008), this framework often predicts a stationary dynamics due to the nature of the attractors (Barak et al., 2013). This contradicts the intuition that the brain processes information in a distributed way (Christophel et al., 2017), and the highly dynamic coding observed in experiments (Brody et al., 2003; Shafi et al., 2007; Jun et al., 2010; Stokes et al., 2013). In order to address these problems, a novel computational framework has proposed recently, known as the reservoir computing (Jaeger, 2001; Maass et al., 2002). This computational model suggests that the information could be processed in a highly flexible way without stable states, by combining

the transmission and storage of information, and exploiting the rich intrinsic dynamics in random neural networks to achieve real-time computation. In particular, the internal state of the random neural network which is referred to as the reservoir, changes continuously as a result of the self-interaction through recurrent connections, and the information processing under this framework can be considered as the information conveyed by the external input superposed onto this internal state. However, due to the complex nature of the dynamics in the reservoir, the interaction between the external input and the intrinsic dynamics is not fully understood.

A common feature of these different frameworks is the nonlinear dynamics involved in the information processing. In terms of the simplified, rate-based model, nonlinearity arises originally from the transfer function f . Under the framework of attractor neural networks, such simple, local dynamics allows for the appearance of nontrivial stable states, as well as the interaction between such states when coupled through structured synaptic connections. Under the framework of reservoir computing, where the coupling is random, the nonlinearity may lead to disordered dynamics which explores a broad region in the state space. On the other hand, the nonlinearity may also appear in the synaptic dynamics. In either case, nonlinear dynamics plays a crucial part in the flexible information processing by enriching the dynamics and allowing the processing to depend on previous states of the network.

In the following chapters, we introduce three studies. Our aim, in the broadest terms, is to explore neural mechanisms that contribute to flexible information processing. In the first two studies, we use rate-based models to investigate the dynamics of networks under the frameworks of associative memory network and reservoir computing. We focus on how the dynamics is affected by synaptic plasticity and external input. In the third study, we analyze the data recorded from monkey's prefrontal cortex during a working memory task, and construct a spiking neural network model which qualitatively reproduce the result of our analysis. Specifically, in **chapter 2**, we analyze the stability of associative memory in a network composed of stochastic neurons and dynamic synapses. Using a mean field model and bifurcation analysis, we describe the macroscopic dynamics in the stochastic model, and investigate how the structure in the dynamics changes according to parameters specifying the properties of associative memory patterns, short-term synaptic dynamics and an activity control mechanism. In **chapter 3**, we analyze the transition from a fixed-point state to a chaotic state in a random neural network consists of two populations. Using a mean field technique, we specify the fixed-point state, and numerically obtain the critical point of the transition. We focus on

how the critical point, near which the best performance of reservoir computing is known to be achieved, depends on the statistics of the external input to the network. In **chapter 4**, we first introduce the group reversal task which requires the monkey to predict the outcome of each trial based on a visual cue and the context that has to be inferred from previous trials. The data recorded while the monkey is performing this task is analyzed using various methods based on single neuron, as well as population responses. Then we propose a spiking neural network model with both recurrent and hierarchical structures, and demonstrate that the model can qualitatively reproduce the results from the recorded data by numerical simulations. Finally, in **chapter 5**, we conclude our studies by discussing their implications and limitations, as well as topics for future studies.

Chapter 2

Stability of Associative Memory with Short-Term Synaptic Dynamics

This part is omitted.

Chapter 3

External Input-Forced Onset of Chaos in Random Neural Networks

3.1 Introduction

In this chapter we study the dynamics of neural networks under another framework, i.e. the reservoir computing (Jaeger, 2001; Maass et al., 2002). Recent experimental studies revealed that during cognitive tasks, neural representations of task-relevant information can be quite flexible. In particular, the activity of a single neuron is tuned to multiple task-related aspects (Meyers et al., 2008; Barak et al., 2010; Rigotti et al., 2013), and exhibits dynamic, rather than static coding properties (Stokes et al., 2013). These characteristics are difficult to be accounted for by classical models, but are well-captured under the framework of reservoir computing. Specifically, reservoir computing often assumes a random structure in the recurrent neural network which is referred to as the reservoir. An external input to the reservoir interacts with the internal state of the network, and is transformed into a high dimensional response through the reverberation between nonlinear neurons. This response can be seen as a dynamic code of the external input. The computational benefit of this framework is that a desired output signal, no matter how complicated, can often be obtained by simple linear combinations trained to readout the response, by virtue of the intrinsic dynamics in the reservoir.

The performance of reservoir computing depends critically on the dynamical regime it works in. Specifically, according to different properties of the network, the dynamics expands a range from ordered to chaotic states. The border between these different types of dynamics, known as the edge of chaos, is found to enable complex computations with desirable features for sequential data processing (Bertschinger & Natschläger). Intuitively, this can be understood as follows. In order to effectively readout the information of the more recent external input, spurious information about the initial states should be washed out in the dynamics. In other words, trajectories of the network states driven by the same external input should asymptotically converge. On the other hand, for the computation to be useful, different external inputs must be transformed into different states of the network. These two conditions, referred to as the fading memory (Jaeger, 2001) and the separation property (Maass, et al., 2002), are often met at the edge of chaos because, roughly speaking, the former is guaranteed by the contractive dynamics in the ordered state and the latter is guaranteed by the sensitivity to the input in the chaotic state.

In a different context, the transition from fixed points to chaos has been studied by theoretical neuroscientists whose aim is to understand the origin of the irregularity in neural firing patterns. Although neurons communicate with each other in a noisy environment, the generation of a Poisson-like neural code is not a trivial problem because the neuron integrates over a large number of synaptic inputs, and such spatial average often results in regular dynamics (Denve & Machens, 2016). A widely accepted theory addressing this problem is based on the balance between excitatory and inhibitory inputs to individual neurons (Vreeswijk & Sompolinsky, 1996, 1998; Brunel, 2000). In these spiking models, when excitatory and inhibitory neurons are connected through random and sparse synapses, it is found that in certain parameter regimes the excitation and inhibition dynamically cancel each other, leaving the activity of the neuron to be driven by the stochastic fluctuations in the input, and therefore generates Poisson-like spike trains. Furthermore, using a rate-based model, Sompolinsky et al. (1988) proved that there is a sharp transition from fixed points to chaos in large random neural networks, when the synaptic gain exceeds a critical value. In a recent study, Kadmon and Sompolinsky (2015) investigated the properties of dynamical systems near this transition, and discussed under what condition the dynamics in spiking models can be predicted by that in rate-based models. Using dynamic mean field theory, the authors analyzed how the transition occurs in neural networks with various architectures. In particular, they focused on the effects of the synaptic gain and the

shape of the transfer function. These results have direct implications for the reservoir computing.

On the other hand, as a non-autonomous system, the dynamics in the reservoir also depends critically on the external input to it (Manjunath & Jaeger, 2013). For example, in the echo state network (Jaeger, 2001), which is an important implementation of reservoir computing, the echo state property necessary for the computation is defined with respect to the external input (Yildiz, et al., 2012). Nonetheless, how external input affects the dynamics in the reservoir is difficult to measure and is not fully understood. In this study, we focus on the influence of the statistical properties of external input on the critical point in a rate-based model. Using a mean field technique and numerical methods, we investigate how the critical point is affected by the statistics of simple, time-independent external inputs. Our results add to the study of Kadmon and Sompolinsky (2015) by clearly showing the dependence of the dynamics on external inputs, and may provide an insight to the designing of reservoir in modeling and applications.

3.2 A Random Neural Network with Two Populations

We consider a neural network composed of two populations: an excitatory population consists of N_E neurons and an inhibitory population consists of $N_I = \frac{1}{4}N_E$ neurons. The neurons are randomly connected, where each neuron receives C_E excitatory and C_I inhibitory synapses from local recurrent connections, and C_{ext} excitatory synapses from other (remote) brain areas. The synaptic efficacy of each synapse is chosen randomly from a Gaussian distribution, whose mean and standard deviation are both J for the excitatory synapses, and are both $g \cdot J$ for the inhibitory synapses. It should be noted that this assumption may actually violate the definition of excitatory and inhibitory synapses, since the synaptic efficacy could take a negative value. However, previous studies on large random systems revealed that when the number of synapses on a single neuron (i.e. C_E and C_I) is large, the dynamics of the network depends only on the first two moments of the connectivity matrix (Mzard et al., 1987; Fischer and Hertz, 1991; Tao, 2012). Therefore the network can be replaced by a randomly diluted network where the efficacies of synapses are identical. Based on these assumptions, the mean and variance of the synaptic efficacy are of the order $O(\frac{1}{N})$, where $N = N_E + N_I$ is the total number of neurons in the network.

Let E and I denote the excitatory and inhibitory population. The state of the i th neuron in the population k evolves according to

$$\frac{dx_i^k}{dt} = -\frac{x_i^k}{\tau} + \sum_{j=1}^{N_E} J_{ij}^{kE} r_j^E - \sum_{l=1}^{N_I} J_{il}^{kI} r_l^I + C_{ext} \cdot J \cdot h_i^k, \quad (3.1)$$

$$r_i^k = f(x_i^k), k \in \{E, I\}, \quad (3.2)$$

where x_i^k and r_i^k represent the activation and firing rate of the neuron, and τ is the time constant. J_{ij}^{kE} and J_{il}^{kI} are the synaptic efficacies from the j th excitatory neuron and the l th inhibitory neuron, respectively. When there is no synapse between the neurons, the corresponding synaptic efficacy is set to zero. The time-independent external input h_i^k is chosen from a given distribution, whose mean and standard deviation are denoted by μ_{ext} and σ_{ext} . In order to investigate the dependence of the dynamics on μ_{ext} and σ_{ext} separately, we assume that the efficacies of non-local synapses are identical. In this study, we adopt the often-used hyperbolic tangent transfer function $f(x) = \frac{1}{2}(1 + \tanh(x/T))$, where T specifies the shape of the transfer function. As shown by Kadmon and Sompolinsky (2015), when the transfer function arises sharply near the threshold, i.e, when T is small, the transition to chaos disappear in large networks.

In order to investigate the dynamics of the network, we simulate Eqs. (3.1)-(3.2) using Runge-Kutta method with time step $\Delta t = 0.1$. The parameter values used in the simulations are given as follows. For the network, $N_E = 10000$, $N_I = 2500$, $C_E = C_{ext} = 100$, and $C_I = 200$, and for the neurons and synapses, $J = 0.2$, $g = 5$, $\tau = 10$, and $T = 10$. Specifically, we are interested in the limit $N_E, N_I \rightarrow \infty$ with other parameters fixed. For the external input, we consider two different types of distribution: Gaussian distribution and uniform distribution. However, as we will see, these two types of distribution have essentially the same effects on the dynamics.

The two different types of dynamics are shown in **Figure 3.1**. When the external input is small, the dynamics is described by the convergence to a fixed-point state (**Figure 3.1A**), in which the firing rate of each neuron depends on the implementation of synapses. However, the macroscopic statistics of this state can be solved self-consistently (**Figure 3.1B**). As the external input increases, chaotic fluctuation is forced onset (**Figure 3.1C,3.1D**). Using a dynamic mean field theory, the temporal correlation in this state can be obtained (Sompolinsky et al., 1988; Kadmon & Sompolinsky, 2015). A further increase of the external input will result in saturation.

In order to characterize the fixed-point state, we first assume that the external input h_i^k subjects to Gaussian distribution. The Eq. (3.1) is rewritten as

$$\frac{dx_i^k}{dt} = -\frac{x_i^k}{\tau} + \eta_i^k. \quad (3.3)$$

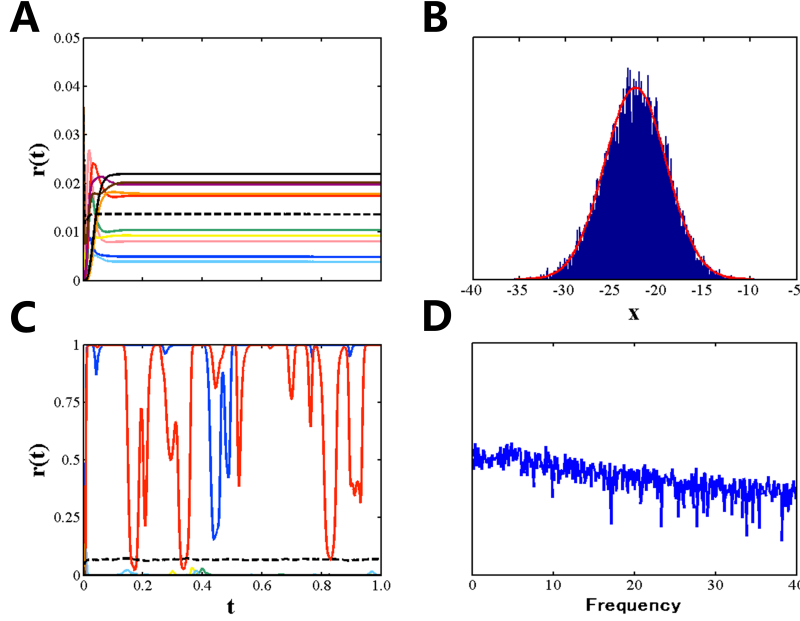


Figure 3.1: **Different types of dynamics of the network.** (A) The fixed-point state ($\mu_{ext} = 0.010, \sigma_{ext} = 0.010$). Colored curves show the temporal evolution of firing rates of 10 neurons. The temporal evolution of the mean firing rate in the network is indicated by the dashed curve. (B) The distribution of activation in the state shown in (A). The red curve indicates the Gaussian distribution fitted to the data. (C) The chaotic state ($\mu_{ext} = 0.200, \sigma_{ext} = 0.200$). The configuration is the same as in (A). Instead of converging to fixed values, the firing rate of each neuron fluctuates chaotically. (D) Power spectrum in the state shown in (C). The power spectrum is calculated for a sufficiently long time sequence of the mean firing rate, and plotted in logarithm. The continuous power spectrum is a signature of chaos.

In the limit of a large number of synaptic connections per neuron, $\eta_i^k = \sum_j J_{ij}^{kE} r_j^E - \sum_l J_{il}^{kI} r_l^I + C_{ext} \cdot J \cdot h_i^k$ can be described by a time-dependent Gaussian field, which represents the recurrent synaptic inputs to a single neuron (Sompolinsky et al., 1988). Therefore the activation x_i^k , driven by the Gaussian term, also obeys Gaussian statistics (**Figure 3.1B**). In particular, since the inputs to excitatory neurons and inhibitory neurons are assumed to be statistically equivalent in the current model, these two types of neurons subject to the same distribution. Therefore in the following, we simply denote the activation and firing rate by random variables x and r . In addition, the

mean and standard deviation of these variables are denoted by μ_x and σ_x , and μ_r and σ_r , respectively. The fixed-point state is described by these quantities. The probability density function of r is given by

$$P(r) = \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left[-\frac{(f^{-1}(r) - \mu_x)^2}{2\sigma_x^2}\right] \cdot [f^{-1}(r)]', \quad (3.4)$$

where the prime denotes the derivative of the function.

The correlation between synaptic efficacy and the firing rate of presynaptic neuron disappear when $N \rightarrow \infty$. Therefore the statistics of x can be derived as

$$\mu_x = \tau J[(C_E - gC_I)\mu_r + C_{ext}\mu_{ext}], \quad (3.5)$$

$$\sigma_x^2 = \tau^2 J^2[(C_E + g^2C_I)(2\sigma_r^2 + \mu_r^2) + C_{ext}^2\sigma_{ext}^2], \quad (3.6)$$

in the fixed-point state. Substituting Eqs. (3.5) and (3.6) into Eq. (3.4), and integrating for the first two moments of the distribution of r result in the following self-consistent equations for μ_r and σ_r :

$$\mu_r = \int_0^1 r P(r, \mu_r, \sigma_r) dr, \quad (3.7)$$

$$\mu_r^2 + \sigma_r^2 = \int_0^1 r^2 P(r, \mu_r, \sigma_r) dr. \quad (3.8)$$

Figure 3.2 shows the statistics of the firing rate as functions of the mean and standard deviation of the external input. The solution is found to exist for finite μ_{ext} and σ_{ext} , indicating the transition to chaos. In particular, the mean firing rate depends linearly on the mean external input, as in the balanced network with spiking neurons (Vreeswijk & Sompolinsky, 1996). The stability of the fixed-point state is determined by the Jacobian matrix of the original N dimensional system, which is given as follows:

$$\begin{aligned} \tilde{J} &= (\hat{J}_{ij} f'(x_j) - \frac{1}{\tau} \delta_{ij})_{ij} \\ &= \hat{J} \cdot \begin{bmatrix} f'(x_1) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & f'(x_N) \end{bmatrix} - \frac{1}{\tau} \cdot I, \end{aligned} \quad (3.9)$$

where \hat{J} is the N by N connectivity matrix whose first N_E columns correspond

to the synapses from excitatory neurons and last N_I columns correspond to the synapses from inhibitory neurons, and I denotes the identity matrix. The neurons are aligned similarly: $(x_1, x_2, \dots, x_N) = (x_1^E, \dots, x_{N_E}^E, x_1^I, \dots, x_{N_I}^I)$. In general, the activation of neurons are determined by the connectivity matrix. However, in the limit of large networks, the two random matrices in Eq. (3.9) can be evaluated separately. Therefore for each value of μ_{ext} and σ_{ext} , we first solve the statistics of the corresponding fixed-point state from Eqs. (3.7)-(3.8), and then calculate the eigenvalue spectrum of the Jacobian matrix by sampling a sufficiently large matrix based on these statistics.

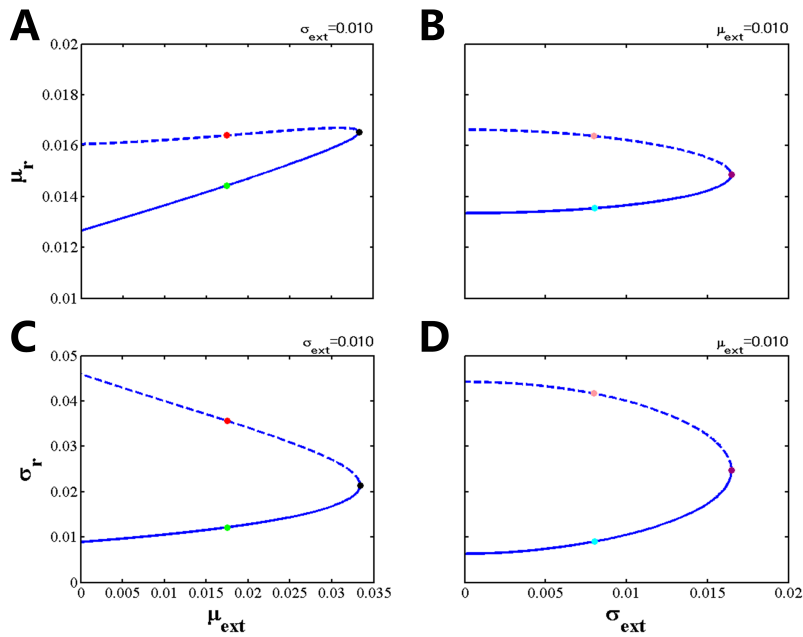


Figure 3.2: **Statistics in the fixed-point state.** Statistics of the firing rate in the fixed-point state are plotted as functions of statistics of the external input. Solid curves and dashed curves indicate the stable and unstable fixed-point state, respectively. The eigenvalue spectrums of the states indicated by the green, black, and red solid circles in (A) and (C), and the states indicated by the cyan, purple, and pink solid circles in (B) and (D) are illustrated in **Figure 3.3**. (A) The mean firing rate μ_r as a function of the mean external input μ_{ext} . (B) The mean firing rate μ_r as a function of the standard deviation of external input σ_{ext} . (C) The standard deviation of firing rate σ_r as a function of the mean external input μ_{ext} . (D) The standard deviation of firing rate σ_r as a function of the standard deviation of external input σ_{ext} .

Figure 3.3 illustrate how the eigenvalues of the Jacobian matrix change along the curves in **Figure 3.2**. The statistical properties of the eigenvalues of large random matrices are studied in the context of random matrix theory (Tao, 2012). In particular, eigenvalues of matrices whose entries are independent random variables distribute uniformly in a disk (Girko, 1985). When correlation is introduced among the entries, the disk becomes an ellipse (Sommer et

al., 1988). Furthermore, in the matrices with two populations, corresponding to the excitatory synapses and inhibitory synapses, the eigenvalue spectrum shows a layered structure (Rajan et al., 2006). The stability of the fixed-point state depends on the largest real part of the eigenvalues. In **Figure 3.3A,3.3B**, the eigenvalues stay away from the imaginary axis, and each has a negative real part. Therefore the fixed-point state is stable. As the fixed-point state approaches the critical point, the x-radius of the eigenvalue spectrum increases. The eigenvalue spectrum becomes tangent to the imaginary axis at the critical point, as illustrated in **Figure 3.3C,3.3D**. A further increase of the x-radius of the eigenvalue spectrum occurs when one or several eigenvalues cross the imaginary axis, resulting in the unstable fixed-point state shown in **Figure 3.3E,3.3F**. Therefore calculating the eigenvalue spectrum allows us to numerically identify the critical point. Specifically, critical points are found at the fixed-point states with maximal μ_{ext} and σ_{ext} values, respectively **Figure 3.2**.

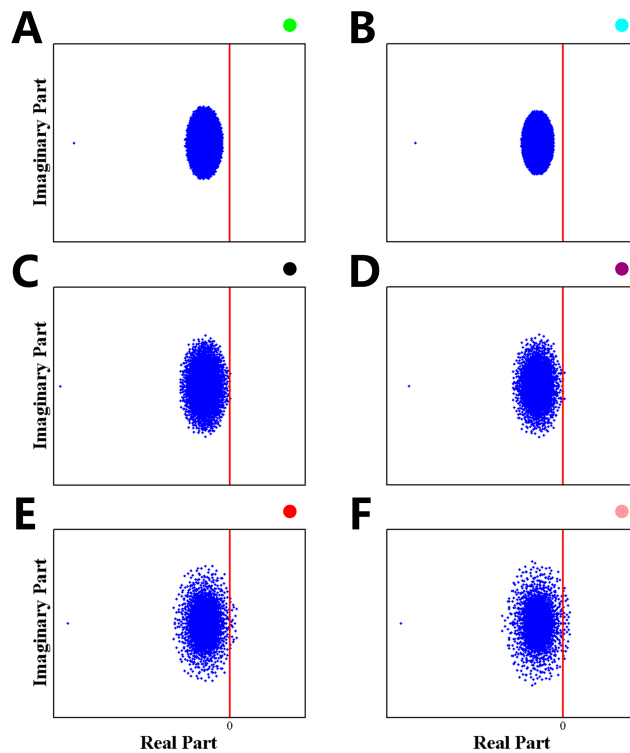


Figure 3.3: **Eigenvalue spectra of various states.** Each dot denotes an eigenvalue. The eigenvalue spectra are calculated using $N = 25000$ dimensional matrices. **(A)** The green solid circle ($\mu_{ext} = 0.0175, \sigma_{ext} = 0.010$) and **(B)** the cyan solid circle ($\mu_{ext} = 0.010, \sigma_{ext} = 0.008$) correspond to stable fixed-point states. **(C)** The black solid circle ($\mu_{ext} = 0.0334, \sigma_{ext} = 0.010$) and **(D)** the purple solid circle ($\mu_{ext} = 0.010, \sigma_{ext} = 0.0165$) correspond to critical fixed-point states. **(E)** The red solid circle ($\mu_{ext} = 0.0175, \sigma_{ext} = 0.010$) and **(F)** the pink solid circle ($\mu_{ext} = 0.010, \sigma_{ext} = 0.008$) correspond to unstable fixed-point states.

To investigate how the critical point depends on the statistics of the external input, we show in **Figure 3.4** the value of σ_{ext} at the critical points as a function of the value of μ_{ext} . This phase diagram indicates the position of the edge of chaos, where the best performance for information processing is expected to be achieved (Bertschinger & Natschläger, 2004). In particular, the mean and the standard deviation of the external input demonstrate a negative linear relationship on the edge of chaos. This result provides useful information regarding the design of reservoirs in modeling and applications.

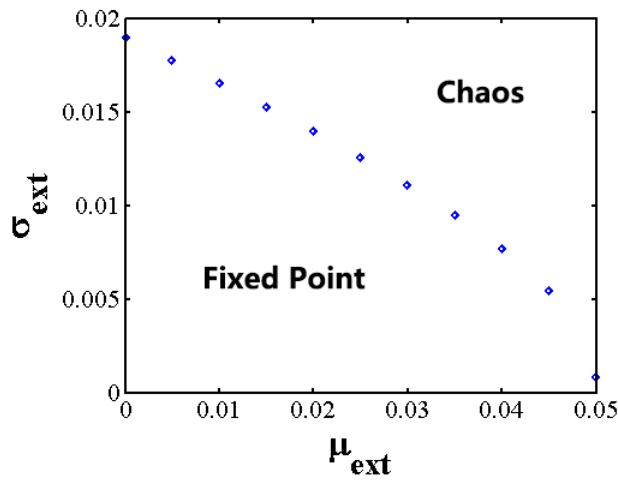


Figure 3.4: **The edge of chaos.** The critical value of the standard deviation of external input σ_{ext} as a function of the mean external input μ_{ext} .

Finally, we examine the consistency between the model and simulation results, as well as the effect of different types of external input. **Figure 3.5** and **Figure 3.6** show the results predicted by the model, and calculated in simulations using Gaussian and uniform external input with the same mean and standard deviation. In general, when the external input is not Gaussian, it is not obvious that η_i^k in Eq. (3.3) still obeys Gaussian statistics. Nonetheless, the results for these types of external inputs are essentially the same, indicating that statistics other than the first two moments of the external input only play an inferior role in determining the dynamics, consistent with previous knowledge of large random systems (Mzard et al., 1987; Fischer and Hertz, 1991; Tao, 2012). Furthermore, both results show good agreement with those predicted by the model, despite that the fixed-point state usually exists in a larger parameter region in finite networks, as indicated by the circles on the right hand side of each panel.

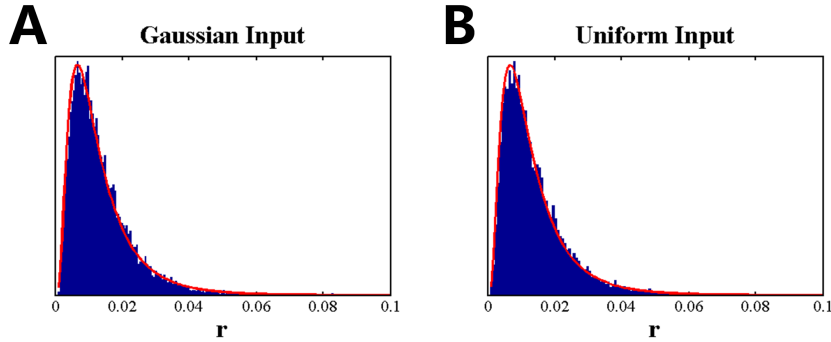


Figure 3.5: **The distribution of firing rate.** The red curves are plotted using Eq. (3.4). (A) Gaussian external input, and (B) uniform external input with the same first two moments are used in simulations ($\mu_{ext} = 0.010, \sigma_{ext} = 0.010$).

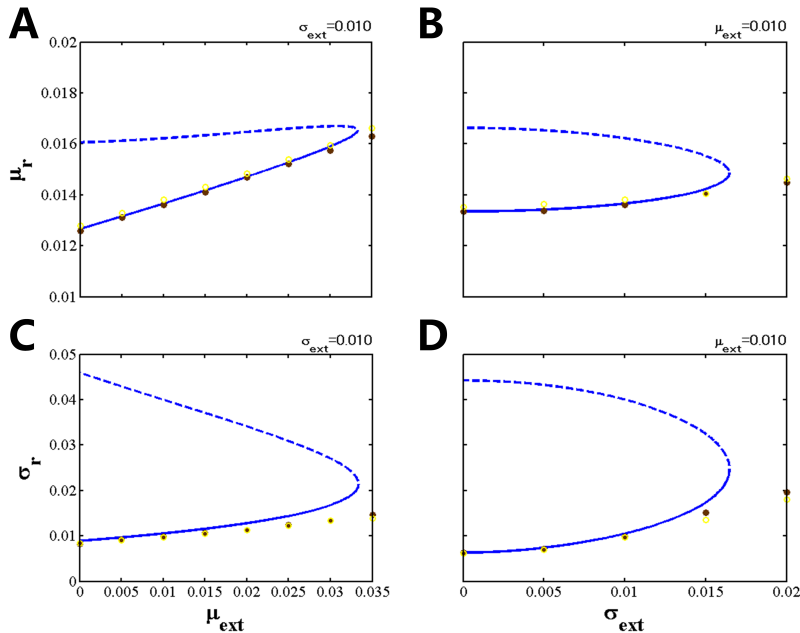


Figure 3.6: **Consistency between the model and simulation results.** The bifurcation diagrams are identical to those in Figure 3.2. Simulation results using Gaussian external input and uniform external input are indicated by brown solid circles and yellow open circles, respectively. In general, the results show good agreement to those predicted by the model. However, the fixed-point state exists in a larger parameter region in finite networks, as indicated by the simulation results with $\mu_{ext} = 0.035$ in (A) and (C), and those with $\sigma_{ext} = 0.020$ in (B) and (D).

3.3 Discussion

We investigated the dynamics in random neural networks by a rate-based model. In particular, we studied the transition from fixed-points to chaos. A similar model has been analyzed systematically in Kadmon and Sompolinsky

(2015), where a network composed of several populations, each receiving a homogeneous external input, is considered. In the current study, our focus is on the effect of external input. To investigate how external input affects the dynamics, we assumed that the external input can be described by a time-independent distribution. Using a mean field technique and numerical methods, we obtained the critical point and showed how it changes with respect to statistics of the external input. In the context of reservoir computing, computational performance depends on the combination of the state of the reservoir and the external input (Yildiz, et al., 2012). In particular, because the input weight matrix is often set randomly, the external input to the reservoir at each time point can be described by some distribution. Furthermore, in a benchmark problem of reservoir computing, known as NARMA10, the external input is given by a random number sequence. Our result therefore provides an insight into the dynamics in such reservoirs.

In this study, the critical point is identified based on numerical calculation of the eigenvalue spectrum. Compared with the dynamic mean field model (Sompolinsky et al., 1988; Kadmon & Sompolinsky, 2015), this approach cannot obtain the scaling properties at the critical point or the temporal structure in the chaotic state. However, it can still qualitatively provide an intuition on how the bifurcation occurs. Specifically, in the system with infinite size, the symmetry in the eigenvalue spectrum implies that the bifurcation occurs on the real axis, corresponding to the saddle-node bifurcation in **Figure 3.2**. In a finite system, the random fluctuation (quenched noise) in the connectivity matrix allows the crossing of imaginary axis to occur off the real axis. In particular, oscillatory states may emerge through the hopf bifurcation (**Figure 3.3D**). As more and more eigenvalues cross the imaginary axis, the nonlinear coupling between these oscillatory modes leads to the onset of chaos. However, in either case the frequency of the oscillation near the bifurcation, determined by the imaginary part of the eigenvalue crossing the imaginary axis, is expected to be small. This benefits the computation by allowing a relatively long integration time.

From the information processing point of view, the framework of reservoir computing contributes in two aspects. First, the intrinsic dynamics in the reservoir provides a set of basis very effective for pattern generation. In the simplest form of reservoir computing, the training of the readout weights only involves calculation of the inverse of a matrix, yet the intrinsic dynamics in the reservoir (near the edge of chaos) allows it to approximate almost arbitrarily complicated output signal. In this sense, the reservoir serves as a

universal function approximator. Second, reservoir computing, when incorporated with learning (Sussillo & Abbott, 2009; Sussillo & Barak, 2013; Barak et al., 2013; Mante et al., 2013; Sussillo et al., 2015; Rajan et al., 2016; Enel et al., 2016; Miconi, 2017), provides a novel way to understand the dynamics behind the highly dynamic, flexible neural responses observed in experiments. In contrast to the framework based on attractor neural networks, this novel approach captures the distributed nature of information processing in the brain (Rigotti et al., 2013; Christophel et al., 2017), and requires less tuning properties (Toyozumi & Abbott, 2011). It is therefore an important issue to reveal the dynamics in the reservoir with respect to various properties of the computation.

3.4 Conclusion

Reservoir computing has been proposed recently as an alternative framework for the information processing in the brain. Under this framework, we investigated the effect of external input to the transition from fixed-points to chaos in random neural networks. Using a mean field technique and numerical methods, we showed how the critical point changes according to the mean and the standard deviation of the external input. In addition, we showed by simulations that statistics other than the first two moments of the external input have only inferior effect on the dynamics. Our results specified the region in which high computational performance of information processing can be achieved, and may contribute to the design of reservoirs in modeling and applications.

Chapter 4

Logical Computation in Monkey's Prefrontal Cortex

This part is omitted.

Chapter 5

Conclusion

In this thesis we explored various neural mechanisms that may contribute to flexible information processing. In particular, the nonlinear dynamics in neural networks plays an essential part in the information storage and transmission. Under the framework based on attractor neural networks, the information storage is associated with stable states (**chapter 2**) which arise from mutual excitation between neurons. For these states to be nontrivial, the neuronal dynamics is required to be nonlinear. Additionally, when organized through synaptic connections, the stored information can be effectively transmitted through neurons serving as nonlinear logical gates to establish proper mapping between input and output, achieving flexible information processing (**chapter 4**). Furthermore, nonlinear synaptic processes may contribute to flexible information processing by largely enriching the dynamics on the network level (**chapter 2**). However, as we discussed in **chapter 4**, neuronal stimulus selectivity predicted by models based on attractor neural networks is often less variable compared with recorded data. This is because that attractor neural networks assume functionally equivalent neurons. Furthermore, such models often display stereotyped dynamics due to the static nature of the attractors (Barak et al., 2013). In order to address these problems, an alternative framework, i.e., reservoir computing has been proposed recently (Jaeger, 2001; Maass et al., 2002), where the information storage and transmission are combined in the rich, highly nonlinear dynamics of random neural networks near the onset of chaos (Bertschinger & Natschläger, 2004). Under this framework, we studied how the dynamics changes according to external input in **chapter 3**.

Meanwhile, flexible information processing depends on many other neural mechanisms as well. The flexibility essentially arises from the ability of neural

networks to maintain an internal state, in a way that is sensitive to changes in the environment. Therefore any neural mechanism that has an effect on the internal state (e.g., homeostatic plasticity which contributes to the maintenance of internal state) should also play a role in flexible information processing. In particular, in order to maintain and modulate the internal state, feedback loops on multiple scales have been established in the brain. Understanding the neural mechanisms involved in the individual pathways and their coordination may require simultaneous recording of a large number of neurons and multiple brain regions. These problems are beginning to be addressed by experimental studies (Harvey & Tank, 2012; Hamel et al., 2015; Rose et al., 2016). On the other hand, recently developed learning techniques (Sussillo & Abbott, 2009; Vanbiervliet et al., 2009; Martens, 2010) have provided powerful tools to reveal neural dynamics without imposing heavy assumptions on the network structures. Specifically, an initially random neural network may well adapt to perform certain cognitive tasks with structure adjusted by these techniques, and the dynamics can then be understood by analyzing the resulting network whose structure is already known (Sussillo & Barak, 2013). Theoretical studies have shown that this type of networks, lying between totally random and highly specified structures, is readily available for a broad range of tasks (Mante et al., 2013; Hennequin et al., 2014; Sussillo et al., 2015; Rajan et al., 2016), while preserving several prominent features of neural computation, such as mixed selectivity (Rigotti et al., 2013; Raposo et al., 2014) and dynamic coding (Stokes et al., 2013). We believe that the investigation of physiological counterparts of these learning techniques (i.e., plasticity mechanisms), as well as the generalization of a network to a variety of tasks is of particular importance for our understanding of flexible information processing in neural networks.

Acknowledgments

The author wishes to thank Kazuyuki Aihara at the University of Tokyo for providing an environment conducive to developing independent thinking; Yuichi Katori at Future University Hakodate, Takayuki Hosokawa and Ken-Ichiro Tsutsui at Tohoku University for helpful guidance and collaboration; Yoko Yamaguchi, Fumiyasu Komaki, Takayasu Matsuo, Gouhei Tanaka and Kazuyuki Aihara at the University of Tokyo for valuable advice and suggestions; Taro Toyozumi at RIKEN, Timothee Leleu at the University of Tokyo for enlightening discussions.

Bibliography

- [1] Abbott, L. F., Varela, J. A., Sen, K., and Nelson, S. B. (1997). Synaptic depression and cortical gain control. *Science*, 275(5297), 221-224.
- [2] Amari, S. I. (1972). Learning patterns and pattern sequences by self-organizing nets of threshold elements. *IEEE Transactions on Computers*, 100(11), 1197-1206.
- [3] Amari, S. I. (1989). Characteristics of sparsely encoded associative memory. *Neural networks*, 2(6), 451-457.
- [4] Amari, S. I., and Maginu, K. (1988). Statistical neurodynamics of associative memory. *Neural Networks*, 1(1), 63-73.
- [5] Amit, D. J., and Brunel, N. (1997). Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. *Cerebral cortex*, 7(3), 237-252.
- [6] Amit, D. J., Gutfreund, H., and Sompolinsky, H. (1985). Spin-glass models of neural networks. *Physical Review A*, 32(2), 1007.
- [7] Amit, D. J., Gutfreund, H., and Sompolinsky, H. (1987). Information storage in neural networks with low levels of activity. *Physical Review A*, 35(5), 2293.
- [8] Anderson, J. A. (1968). A memory storage model utilizing spatial correlation functions. *Biological Cybernetics*, 5(3), 113-119.
- [9] Anderson-Sprecher, R. (1994). Model comparisons and R 2. *The American Statistician*, 48(2), 113-117.
- [10] Baddeley, A. (1986). Oxford psychology series, No. 11. Working memory.
- [11] Baddeley, A. D., and Hitch, G. (1974). Working memory. *Psychology of learning and motivation*, 8, 47-89.

- [12] Barak, O., Sussillo, D., Romo, R., Tsodyks, M., and Abbott, L. F. (2013). From fixed points to chaos: three models of delayed discrimination. *Progress in neurobiology*, 103, 214-222.
- [13] Barak, O., Tsodyks, M., and Romo, R. (2010). Neuronal population coding of parametric working memory. *Journal of Neuroscience*, 30(28), 9424-9430.
- [14] Ben-Yishai, R., Bar-Or, R. L., and Sompolinsky, H. (1995). Theory of orientation tuning in visual cortex. *Proceedings of the National Academy of Sciences*, 92(9), 3844-3848.
- [15] Bertram, R. I. C. H. A. R. D., Sherman, A. R. T. H. U. R., and Stanley, E. F. (1996). Single-domain/bound calcium hypothesis of transmitter release and facilitation. *Journal of Neurophysiology*, 75(5), 1919-1931.
- [16] Bertschinger, N., and Natschläger, T. (2004). Real-time computation at the edge of chaos in recurrent neural networks. *Neural computation*, 16(7), 1413-1436.
- [17] Bibitchkov, D., Herrmann, J. M., and Geisel, T. (2002). Pattern storage and processing in attractor networks with short-time synaptic dynamics. *Network: Computation in neural systems*, 13(1), 115-129.
- [18] Bliss, T. V., and Lømo, T. (1973). Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *The Journal of physiology*, 232(2), 331-356.
- [19] Brody, C. D., Hernandez, A., Zainos, A., and Romo, R. (2003). Timing and neural encoding of somatosensory parametric working memory in macaque prefrontal cortex. *Cerebral cortex*, 13(11), 1196-1207.
- [20] Brunel, N. (2000). Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. *Journal of computational neuroscience*, 8(3), 183-208.
- [21] Candes, E. J., Romberg, J. K., and Tao, T. (2006). Stable signal recovery from incomplete and inaccurate measurements. *Communications on pure and applied mathematics*, 59(8), 1207-1223.
- [22] Candes, E. J., and Tao, T. (2006). Near-optimal signal recovery from random projections: Universal encoding strategies?. *IEEE transactions on information theory*, 52(12), 5406-5425.

- [23] Christophel, T. B., Klink, P. C., Spitzer, B., Roelfsema, P. R., and Haynes, J. D. (2017). The Distributed Nature of Working Memory. *Trends in Cognitive Sciences*.
- [24] Compte, A., Brunel, N., Goldman-Rakic, P. S., and Wang, X. J. (2000). Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cerebral Cortex*, 10(9), 910-923.
- [25] Coolen, A. C. C. (2001). Statistical mechanics of recurrent neural networks dynamics. *Handbook of biological physics*, 4, 619-684.
- [26] Debanne, D., G?hwiler, B. H., and Thompson, S. M. (1994). Asynchronous pre-and postsynaptic activity induces associative long-term depression in area CA1 of the rat hippocampus in vitro. *Proceedings of the National Academy of Sciences*, 91(3), 1148-1152.
- [27] Denve, S., and Machens, C. K. (2016). Efficient codes and balanced networks. *Nature neuroscience*, 19(3), 375-382.
- [28] Donoho, D. L. (2006). Compressed sensing. *IEEE Transactions on information theory*, 52(4), 1289-1306.
- [29] Enel, P., Procyk, E., Quilodran, R., and Dominey, P. F. (2016). Reservoir computing properties of neural dynamics in prefrontal cortex. *PLoS computational biology*, 12(6), e1004967.
- [30] Fischer, K. H., and Hertz, J. A. (1993). *Spin glasses (Vol. 1)*. Cambridge university press.
- [31] Funahashi, S., Bruce, C. J., and Goldman-Rakic, P. S. (1989). Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *Journal of neurophysiology*, 61(2), 331-349.
- [32] Funahashi, S., Bruce, C. J., and Goldman-Rakic, P. S. (1990). Visuospatial coding in primate prefrontal neurons revealed by oculomotor paradigms. *Journal of neurophysiology*, 63(4), 814-831.
- [33] Fuster, J. M. (1973). Unit activity in prefrontal cortex during delayed-response performance: neuronal correlates of transient memory. *Journal of Neurophysiology*.
- [34] Girko, V. L. (1984). Circular law. *Teoriya Veroyatnostei i ee Primeneniya*, 29(4), 669-679.

- [35] Goldman-Rakic, P. S. (1987). Circuitry of primate prefrontal cortex and regulation of behavior by representational memory. *Comprehensive Physiology*.
- [36] Goldman-Rakic, P. S. (1995). Cellular basis of working memory. *Neuron*, 14(3), 477-485.
- [37] Hamel, E. J., Grewe, B. F., Parker, J. G., and Schnitzer, M. J. (2015). Cellular level brain imaging in behaving mammals: an engineering approach. *Neuron*, 86(1), 140-159.
- [38] Hansel, D., and Sompolinsky, H. (1998). 13 Modeling Feature Selectivity in Local Cortical Circuits.
- [39] Harvey, C. D., Coen, P., and Tank, D. W. (2012). Choice-specific sequences in parietal cortex during a virtual-navigation decision task. *Nature*, 484(7392), 62.
- [40] Hebb, D. O. (1949). *The organization of behavior: A neuropsychological approach*. John Wiley & Sons.
- [41] Hennequin, G., Vogels, T. P., and Gerstner, W. (2014). Optimal control of transient dynamics in balanced networks supports generation of complex movements. *Neuron*, 82(6), 1394-1406.
- [42] Hodgkin, A. L., and Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of physiology*, 117(4), 500.
- [43] Holyoak, K. J., Lee, H. S., and Lu, H. (2010). Analogical and category-based inference: a theoretical integration with Bayesian causal models. *Journal of Experimental Psychology: General*, 139(4), 702.
- [44] Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, 79(8), 2554-2558.
- [45] Hopfield, J. J. (1984). Neurons with graded response have collective computational properties like those of two-state neurons. *Proceedings of the national academy of sciences*, 81(10), 3088-3092.
- [46] Hromdka, T., DeWeese, M. R., and Zador, A. M. (2008). Sparse representation of sounds in the unanesthetized auditory cortex. *PLoS Biol*, 6(1), e16.

- [47] Igarashi, Y., Oizumi, M., and Okada, M. (2010). Mean field analysis of stochastic neural network models with synaptic depression. *Journal of the Physical Society of Japan*, 79(8), 084001.
- [48] Ito, I., Ong, R. C. Y., Raman, B., and Stopfer, M. (2008). Sparse odor representation and olfactory learning. *Nature neuroscience*, 11(10), 1177-1184.
- [49] Itskov, V., Hansel, D., and Tsodyks, M. (2011). Short-term facilitation may stabilize parametric working memory trace. *Frontiers in computational neuroscience*, 5, 40.
- [50] Jaeger, H. (2001). The echo state approach to analysing and training recurrent neural networks-with an erratum note. Bonn, Germany: German National Research Center for Information Technology GMD Technical Report, 148(34), 13.
- [51] Jun, J. K., Miller, P., Hernandez, A., Zainos, A., Lemus, L., Brody, C. D., and Romo, R. (2010). Heterogenous population coding of a short-term memory and decision task. *Journal of Neuroscience*, 30(3), 916-929.
- [52] Kadmon, J., and Sompolinsky, H. (2015). Transition to chaos in random neuronal networks. *Physical Review X*, 5(4), 041030.
- [53] Katori, Y., Otsubo, Y., Okada, M., and Aihara, K. (2013). Stability analysis of associative memory network composed of stochastic neurons and dynamic synapses. *Front. Comput. Neurosci.*, 21.
- [54] Katori, Y., Sakamoto, K., Saito, N., Tanji, J., Mushiake, H., and Aihara, K. (2011). Representational switching by dynamical reorganization of attractor structure in a network model of the prefrontal cortex. *PLoS Comput Biol*, 7(11), e1002266.
- [55] Kiani, R., Esteky, H., Mirpour, K., and Tanaka, K. (2007). Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. *Journal of neurophysiology*, 97(6), 4296-4309.
- [56] Kuznetsov, Y. A. (2013). *Elements of applied bifurcation theory* (Vol. 112). Springer Science and Business Media.
- [57] Levy, W. B., and Steward, O. (1983). Temporal contiguity requirements for long-term associative potentiation/depression in the hippocampus. *Neuroscience*, 8(4), 791-797.

- [58] Lin, A. C., Bygrave, A. M., De Calignon, A., Lee, T., and Miesenbock, G. (2014). Sparse, decorrelated odor coding in the mushroom body enhances learned odor discrimination. *Nature neuroscience*, 17(4), 559-568.
- [59] Lisman, J. E., Fellous, J. M., and Wang, X. J. (1998). A role for NMDA-receptor channels in working memory. *Nature neuroscience*, 1(4), 273-275.
- [60] Lømo, T. (1966, January). Frequency potentiation of excitatory synaptic activity in dentate area of hippocampal formation. In *Acta Physiologica Scandinavica* (p. 128). PO BOX 88, OSNEY MEAD, OXFORD OX2 0NE, OXON, ENGLAND: BLACKWELL SCIENCE LTD.
- [61] Maass, W., Natschlager, T., and Markram, H. (2002). Real-time computing without stable states: A new framework for neural computation based on perturbations. *Neural computation*, 14(11), 2531-2560.
- [62] Machens, C. K. (2010). Demixing population activity in higher cortical areas. *Frontiers in computational neuroscience*, 4.
- [63] Machens, C. K., Romo, R., and Brody, C. D. (2010). Functional, but not anatomical, separation of what and when in prefrontal cortex. *Journal of Neuroscience*, 30(1), 350-360.
- [64] Mainen, Z. F., and Sejnowski, T. J. (1995). Reliability of spike timing in neocortical neurons. *Science*, 268(5216), 1503.
- [65] Manjunath, G., and Jaeger, H. (2013). Echo state property linked to an input: Exploring a fundamental characteristic of recurrent neural networks. *Neural computation*, 25(3), 671-696.
- [66] Mante, V., Sussillo, D., Shenoy, K. V., and Newsome, W. T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature*, 503(7474), 78-84.
- [67] Markram, H., Lbke, J., Frotscher, M., and Sakmann, B. (1997). Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. *Science*, 275(5297), 213-215.
- [68] Markram, H., and Tsodyks, M. (1996). Redistribution of synaptic efficacy between neocortical pyramidal neurons. *Nature*, 382(6594), 807.
- [69] Markram, H., Wang, Y., and Tsodyks, M. (1998). Differential signaling via the same axon of neocortical pyramidal neurons. *Proceedings of the National Academy of Sciences*, 95(9), 5323-5328.

- [70] Martens, J. (2010). Deep learning via Hessian-free optimization. In Proceedings of the 27th International Conference on Machine Learning (ICML-10) (pp. 735-742).
- [71] McNaughton, B. L., Battaglia, F. P., Jensen, O., Moser, E. I., and Moser, M. B. (2006). Path integration and the neural basis of the 'cognitive map'. *Nature Reviews Neuroscience*, 7(8), 663-678.
- [72] Mejias, J. F., and Torres, J. J. (2009). Maximum memory capacity on neural networks with short-term synaptic depression and facilitation. *Neural computation*, 21(3), 851-871.
- [73] Menzel, R., and Manz, G. (2005). Neural plasticity of mushroom body-extrinsic neurons in the honeybee brain. *Journal of Experimental Biology*, 208(22), 4317-4332.
- [74] Meyers, E. M., Freedman, D. J., Kreiman, G., Miller, E. K., and Poggio, T. (2008). Dynamic population coding of category information in inferior temporal and prefrontal cortex. *Journal of neurophysiology*, 100(3), 1407-1419.
- [75] Mzard, M., Parisi, G., and Virasoro, M. (1987). Spin glass theory and beyond: An Introduction to the Replica Method and Its Applications (Vol. 9). World Scientific Publishing Co Inc.
- [76] Miconi, T. Flexible decisionmaking in recurrent neural networks trained with a biologically plausible rule.
- [77] Miller, E. K., and Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual review of neuroscience*, 24(1), 167-202.
- [78] Miyashita, Y., and Chang, H. S. (1988). Neuronal correlate of pictorial short-term memory in the primate temporal cortex. *Nature*, 331(6151), 68-70.
- [79] Mongillo, G., Amit, D. J., and Brunel, N. (2003). Retrospective and prospective persistent activity induced by Hebbian learning in a recurrent cortical network. *European Journal of Neuroscience*, 18(7), 2011-2024.
- [80] Mongillo, G., Barak, O., and Tsodyks, M. (2008). Synaptic theory of working memory. *Science*, 319(5869), 1543-1546.
- [81] Mushiake, H., Saito, N., Sakamoto, K., Itoyama, Y., and Tanji, J. (2006). Activity in the lateral prefrontal cortex reflects multiple steps of future events in action plans. *Neuron*, 50(4), 631-641.

- [82] O'Keefe, J., and Recce, M. L. (1993). Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus*, 3(3), 317-330.
- [83] Otsubo, Y., Nagata, K., Oizumi, M., and Okada, M. (2010). Instabilities in associative memory model with synaptic depression and switching phenomena among attractors. *Journal of the Physical Society of Japan*, 79(8), 084002.
- [84] Palm, G. (2013). Neural associative memories and sparse coding. *Neural Networks*, 37, 165-171.
- [85] Pantic, L., Torres, J. J., Kappen, H. J., and Gielen, S. C. (2002). Associative memory with dynamic synapses. *Neural Computation*, 14(12), 2903-2923.
- [86] Petersen, R. S., Panzeri, S., and Diamond, M. E. (2002). Population coding in somatosensory cortex. *Current opinion in neurobiology*, 12(4), 441-447.
- [87] Rajan, K., and Abbott, L. F. (2006). Eigenvalue spectra of random matrices for neural networks. *Physical review letters*, 97(18), 188104.
- [88] Rajan, K., Harvey, C. D., and Tank, D. W. (2016). Recurrent network models of sequence generation and memory. *Neuron*, 90(1), 128-142.
- [89] Rao, S. C., Rainer, G., and Miller, E. K. (1997). Integration of what and where in the primate prefrontal cortex. *Science*, 276(5313), 821-824.
- [90] Raposo, D., Kaufman, M. T., and Churchland, A. K. (2014). A category-free neural population supports evolving demands during decision-making. *Nature neuroscience*, 17(12), 1784-1792.
- [91] Rigotti, M., Barak, O., Warden, M. R., Wang, X. J., Daw, N. D., Miller, E. K., and Fusi, S. (2013). The importance of mixed selectivity in complex cognitive tasks. *Nature*, 497(7451), 585-590.
- [92] Rojas, R. (1996). The backpropagation algorithm. In *Neural networks* (pp. 149-182). Springer Berlin Heidelberg.
- [93] Romo, R., Brody, C. D., Hernandez, A., and Lemus, L. (1999). Neuronal correlates of parametric working memory in the prefrontal cortex. *Nature*, 399(6735), 470-473.

- [94] Romo, R., Hernández, A., Zainos, A., Lemus, L., and Brody, C. D. (2002). Neuronal correlates of decision-making in secondary somatosensory cortex. *Nature neuroscience*, 5(11), 1217.
- [95] Rose, N. S., LaRocque, J. J., Riggall, A. C., Gossesies, O., Starrett, M. J., Meyering, E. E., and Postle, B. R. (2016). Reactivation of latent working memories with transcranial magnetic stimulation. *Science*, 354(6316), 1136-1139.
- [96] Saito, N., Mushiake, H., Sakamoto, K., Itoyama, Y., and Tanji, J. (2005). Representation of immediate and final behavioral goals in the monkey prefrontal cortex during an instructed delay period. *Cerebral Cortex*, 15(10), 1535-1546.
- [97] Samsonovich, A., and McNaughton, B. L. (1997). Path integration and cognitive mapping in a continuous attractor neural network model. *Journal of Neuroscience*, 17(15), 5900-5920.
- [98] Schultz, W., Dayan, P., and Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, 275(5306), 1593-1599.
- [99] Seung, H. S. (1997, December). Learning Continuous Attractors in Recurrent Networks. In *NIPS* (Vol. 97, pp. 654-660).
- [100] Shafi, M., Zhou, Y., Quintana, J., Chow, C., Fuster, J., and Bodner, M. (2007). Variability in neuronal activity in primate cortex during working memory tasks. *Neuroscience*, 146(3), 1082-1108.
- [101] Softky, W. R., and Koch, C. (1993). The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. *Journal of Neuroscience*, 13(1), 334-350.
- [102] Sommers, H. J., Crisanti, A., Sompolinsky, H., and Stein, Y. (1988). Spectrum of large random asymmetric matrices. *Physical review letters*, 60(19), 1895.
- [103] Sompolinsky, H., Crisanti, A., and Sommers, H. J. (1988). Chaos in random neural networks. *Physical Review Letters*, 61(3), 259.
- [104] Song, S., and Abbott, L. F. (2001). Cortical development and remapping through spike timing-dependent plasticity. *Neuron*, 32(2), 339-350.
- [105] Stein, R. B., Gossen, E. R., and Jones, K. E. (2005). Neuronal variability: noise or part of the signal?. *Nature Reviews Neuroscience*, 6(5), 389-397.

- [106] Stokes, M. G., Kusunoki, M., Sigala, N., Nili, H., Gaffan, D., and Duncan, J. (2013). Dynamic coding for cognitive control in prefrontal cortex. *Neuron*, 78(2), 364-375.
- [107] Sussillo, D., and Abbott, L. F. (2009). Generating coherent patterns of activity from chaotic neural networks. *Neuron*, 63(4), 544-557.
- [108] Sussillo, D., and Barak, O. (2013). Opening the black box: low-dimensional dynamics in high-dimensional recurrent neural networks. *Neural computation*, 25(3), 626-649.
- [109] Sussillo, D., Churchland, M. M., Kaufman, M. T., and Shenoy, K. V. (2015). A neural network that finds a naturalistic solution for the production of muscle activity. *Nature neuroscience*, 18(7), 1025-1033.
- [110] Sutton, R. S., and Barto, A. G. (1998). Reinforcement learning: An introduction (Vol. 1, No. 1). Cambridge: MIT press.
- [111] Tanji, J., and Hoshi, E. (2008). Role of the lateral prefrontal cortex in executive behavioral control. *Physiological reviews*, 88(1), 37-57.
- [112] Tao, T. (2012). Topics in random matrix theory (Vol. 132, p. 282). Providence, RI: American Mathematical Society.
- [113] Torres, J. J., Cortes, J. M., Marro, J., and Kappen, H. J. (2007). Competition between synaptic depression and facilitation in attractor neural networks. *Neural Computation*, 19(10), 2739-2755.
- [114] Toyozumi, T., and Abbott, L. F. (2011). Beyond the edge of chaos: Amplification and temporal integration by recurrent networks in the chaotic regime. *Physical Review E*, 84(5), 051908.
- [115] Tsodyks, M. V., and Feigel'Man, M. V. (1988). The enhanced storage capacity in neural networks with low activity level. *EPL (Europhysics Letters)*, 6(2), 101.
- [116] Tsodyks, M. V., and Markram, H. (1997). The neural code between neocortical pyramidal neurons depends on neurotransmitter release probability. *Proceedings of the National Academy of Sciences*, 94(2), 719-723.
- [117] Tsodyks, M., Pawelzik, K., and Markram, H. (1998). Neural networks with dynamic synapses. *Neural computation*, 10(4), 821-835.

- [118] Tsutsui, K. I., Hosokawa, T., Yamada, M., and Iijima, T. (2016). Representation of functional category in the monkey prefrontal cortex and its rule-dependent use for behavioral selection. *Journal of Neuroscience*, 36(10), 3038-3048.
- [119] Vanbiervliet, J., Vandereycken, B., Michiels, W., Vandewalle, S., and Diehl, M. (2009). The smoothed spectral abscissa for robust stability optimization. *SIAM Journal on Optimization*, 20(1), 156-171.
- [120] Vaughan, W. (1988). Formation of equivalence sets in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, 14(1), 36.
- [121] Vinje, W. E., and Gallant, J. L. (2000). Sparse coding and decorrelation in primary visual cortex during natural vision. *Science*, 287(5456), 1273-1276.
- [122] van Vreeswijk, C., and Sompolinsky, H. (1996). Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science*, 274(5293), 1724.
- [123] van Vreeswijk, C., and Sompolinsky, H. (1998). Chaotic balanced state in a model of cortical circuits. *Neural computation*, 10(6), 1321-1371.
- [124] Wang, X. J. (1999). Synaptic basis of cortical persistent activity: the importance of NMDA receptors to working memory. *Journal of Neuroscience*, 19(21), 9587-9603.
- [125] Wang, X. J. (2008). Decision making in recurrent neuronal circuits. *Neuron*, 60(2), 215-234.
- [126] Wang, Y., Markram, H., Goodman, P. H., Berger, T. K., Ma, J., and Goldman-Rakic, P. S. (2006). Heterogeneity in the pyramidal network of the medial prefrontal cortex. *Nature neuroscience*, 9(4), 534-542.
- [127] Yildiz, I. B., Jaeger, H., and Kiebel, S. J. (2012). Re-visiting the echo state property. *Neural networks*, 35, 1-9.