Solvable neural network model for input-output associations: Optimal recall at the onset of chaos

Tomoki Kurikawa *
*Department of Physics, Kansai Medical University, Shinmachi 2-5-1, Hirakata, Osaka 573-1010, Japan
and Department of Complex and Intelligent systems, Future University Hakodate, 116-2 Kamedanakano-cho, Hakodate, Hokkaido 041-8655, Japan*

Kunihiko Kaneko
*The Niels Bohr Institute, University of Copenhagen, Blegdamsvej 17, Copenhagen 2100-DK, Denmark
and Center for Complex Systems Biology, Universal Biology Institute, University of Tokyo, Komaba, Tokyo 153-8902, Japan*

(Received 14 July 2023; accepted 22 November 2023; published 8 December 2023)

In neural information processing, inputs modulate neural dynamics to generate desired outputs. To unravel the dynamics and underlying neural connectivity enabling such input-output association, we propose an exactly solvable neural-network model with a connectivity matrix explicitly consisting of inputs and required outputs. An analytic form of the response under the input is derived, while three distinctive types of responses including chaotic dynamics are obtained as distinctive bifurcations against input strength, depending on the neural sensitivity and number of inputs. Optimal performance is achieved at the onset of chaos. The relevance of the results to cognitive dynamics is discussed.

DOI: 10.1103/PhysRevResearch.5.043221

I. INTRODUCTION

Neural systems exhibit rich dynamics generated by strong recurrent connections [1]. For performing cognitive tasks in neural systems, sensory inputs modulate the neural dynamics to generate specific output patterns resulting in suitable behaviors. In the association task between the input signals and output choices, for instance, the different input deferentially modifies ongoing (spontaneous) neural dynamics, leading to the emergence of an appropriate attractor that guides the correct choice [2,3], as strongly contrasted with input-output transformation in feed-forward networks [4,5]. Unveiling the mechanisms behind such modulation and the type of connectivity relevant to it is essential for understanding information processing in neural systems.

One widespread and powerful approach to understanding information processing involves recurrent neural networks trained using machine learning techniques [2,3,6–8]. However, these trained networks are finely tuned for specific tasks, which obscures the connectivity relevant to cognitive functions.

Another approach, the autoassociative memory model, offers network connectivity explicitly represented by memorized patterns [9–11]. In this approach, a fixed flow structure in neural dynamics exists, where different fixed-point attractors correspond to distinct memorized patterns. A neural state converges to one of these attractors depending on its initial state, corresponding to the input. Thus, neural dynamics themselves are not modulated by the input.

In the present Letter, we explore how the modulation of neural dynamics by input results in the associated outputs, based on an alternative view of memory, termed "memories as bifurcations," [12,13]. In this view, an input changes the flow structure in the neural dynamical system itself and multiple inputs generate distinctive flow structures, which is consistent with experimental studies [2]. In the case of a successful memory, for a given flow structure under an input, a unique attractor as a memory exists. When multiple input-output associations are successfully embedded, distinct bifurcations to different memory attractors occur for each input. How such a system is designed and the analysis of its memory recall process is focused on in the present study.

We propose a neural network model with a connectivity matrix that enables such bifurcations in an association task between inputs and outputs. The connectivity is explicitly represented on a set of input and output patterns. Bifurcations against the change in input strength are analytically obtained regarding fixed points. Besides these fixed points, a chaotic attractor in the absence of input emerges as the number of memories and/or the gain parameter increases. Here, fixed points corresponding to different memorized patterns emerge from the spontaneous chaos upon the associated inputs. In contrast, with the further increase in the above parameters, the chaotic attractor remains even in the presence of the input. We found that the optimal recall is achieved at the onset of the remained chaos. Finally, computational roles of chaotic internal dynamics are discussed in possible relation to experimental observations.

*kurikawa@fun.ac.jp

II. MODEL

We consider a neural network model composed of N neurons. The network is required to generate target patterns $\boldsymbol{\xi}^{\boldsymbol{\mu}}$ ($\mu = 1, 2, \ldots, M$) in response to input patterns $\boldsymbol{\eta}^{\boldsymbol{\mu}}$, where $M = \alpha N$, and $\boldsymbol{\eta}^{\boldsymbol{\mu}}$ and $\boldsymbol{\xi}^{\boldsymbol{\mu}}$ are N-element vertical vectors. Each element of these vectors takes a binary value (± 1) that is randomly generated according to the probability distribution $P(\xi_i^\mu = \pm 1) = P(\eta_i^\mu = \pm 1) = 1/2$. The neural activity x_i evolves according to the following equation:

$$\dot{x}_i = f\left(\Sigma_j J_{ij} x_j + \gamma \eta_i^\mu\right) - x_i, \tag{1}$$

where β and γ are the gain of the activation function $f(x)$ and the input strength, respectively. We use $f(x) = \tanh(\beta x)$.

For memorizing input/output associations between $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$, we have designed the connectivity matrix J that is composed of $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$, in contrast to the Hopfield network that incorporates only $\boldsymbol{\xi}$ as follows:

$$J = X \begin{pmatrix} I & I \\ -I & -I \end{pmatrix} X^+, \tag{2}$$

$$X = [\boldsymbol{\xi}^{\mathbf{1}}, \boldsymbol{\xi}^{\mathbf{2}}, \ldots, \boldsymbol{\xi}^{\boldsymbol{M}}, \boldsymbol{\eta}^{\mathbf{1}}, \boldsymbol{\eta}^{\mathbf{2}}, \ldots, \boldsymbol{\eta}^{\boldsymbol{M}}], \tag{3}$$

where I is an M-dimensional identity matrix, X is an $(N, 2M)$-matrix, and $X^+ \triangleq (X^T X)^{-1} X^T$ is a pseudoinverse matrix of X, where X^T is a transpose matrix of X. The pseudoinverse matrix X^+ is introduced in the designed connectivity to mitigate the effects of potential interference across memories that could impair recall performance of the memory patterns [14–16]. Due to the pseudoinverse matrix, $J\boldsymbol{\xi}^\mu + \gamma \boldsymbol{\eta}^{\boldsymbol{\mu}} = \boldsymbol{\xi}^{\boldsymbol{\mu}} + (\gamma - 1)\boldsymbol{\eta}^{\boldsymbol{\mu}}$ and, consequently, the target $\boldsymbol{\xi}^{\boldsymbol{\mu}}$ is a fixed point under $\boldsymbol{\eta}^{\boldsymbol{\mu}}$ with $\gamma = 1$ for $\beta \to \infty$ in Eq. (1). This property applies to all μ, indicating that all $\boldsymbol{\xi}^\mu$ are the fixed points under the corresponding inputs with $\gamma = 1$. In other words, all associations are successfully memorized in this model. To satisfy the pseudoinverse matrix, however, the number of vectors, $2M$, that are linearly independent of each other should be less than N. As a consequence, at best, $M = N/2$ associations are allowed and the memory capacity is bounded by $\alpha = 0.5$ at a maximum.

III. RESULTS

A. Analytical solution of the response

How does the network respond to the input except for $\gamma = 1$ and $\beta \to \infty$? We, now, derive an analytical form of a fixed point of the neural dynamics upon input for any value of γ with finite β. For it, we consider $x^{\mathrm{fp}}(\gamma) = (a(\gamma)\boldsymbol{\xi} + b(\gamma)\boldsymbol{\eta})$ and derive $a(\gamma)$ and $b(\gamma)$ such that satisfy the fixed point condition for any γ as follows. Below, the superscript μ is omitted for clarity unless otherwise noted since the result is not dependent on μ. By using $J\boldsymbol{\xi} = J\boldsymbol{\eta} = \boldsymbol{\xi} - \boldsymbol{\eta}$, we have

$$Jx^{\mathrm{fp}} = (a + b)(\boldsymbol{\xi} - \boldsymbol{\eta}), \tag{4}$$

and, subsequently, by substituting x^{fp} to $\dot{x} = 0$ in Eq. (1),

$$a\boldsymbol{\xi} + b\boldsymbol{\eta} = f((a + b)(\boldsymbol{\xi} - \boldsymbol{\eta}) + \gamma\boldsymbol{\eta}). \tag{5}$$

Considering ith elements such that ξ_i equals η_i, $a + b = f(\gamma)$ should be satisfied and, similarly, by considering ith elements such that ξ_i equals $-\eta_i$, $a - b = f(2(a + b) - \gamma)$ should be satisfied. Thus, we derive a and b as

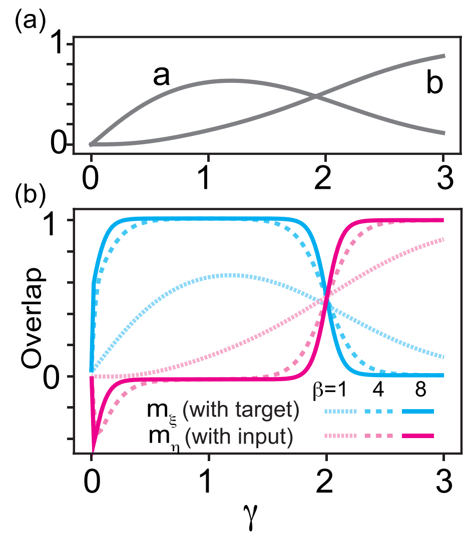

FIG. 1. Analytically obtained response of the network to input with increasing the input strength γ. (a) $a(\gamma)$ and $b(\gamma)$ in Eqs. (6) and (7), for $\beta = 1$. (b) The overlaps of x^{fp} with a target and an input for increasing γ, plotted for different β in blue and red, respectively.

$$a = (f(\gamma) + f(2f(\gamma) - \gamma))/2, \tag{6}$$

$$b = (f(\gamma) - f(2f(\gamma) - \gamma))/2, \tag{7}$$

where a and b depend solely on γ for a given $f(x)$ [17] while they are independent of N, α. It is straightforward to check that Eq. (5) is satisfied for any binary $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$. Although not proven analytically, we have confirmed numerically that x^{fp} is a unique fixed point for given μ and γ. As γ increases from zero, $a(\gamma)$ increases and takes a peak for $\gamma = 1$, while $b(\gamma)$ increases more slowly as plotted in Fig. 1(a). For γ less than two, $a(\gamma)$ is larger than $b(\gamma)$ and, oppositely, beyond $\gamma = 2$, $b(\gamma)$ is larger than $a(\gamma)$ [18]. The overlap of x^{fp} with $\boldsymbol{\xi}$, termed $m_\xi \triangleq \Sigma_i x_i^{\mathrm{fp}} \xi_i / N$, is also plotted in Fig. 1(b). For a given β, m_ξ increases up to $\gamma = 1$ and, subsequently, decreases. As β increases, the curve of m_ξ is steeper so that x^{fp} nearly equals one even for the weak input. The overlap of x^{fp} with $\boldsymbol{\eta}$, termed $m_\eta \triangleq \Sigma_i x_i^{\mathrm{fp}} \eta_i / N$, slowly increases with γ, followed by a sharp rise at $\gamma \approx 2$ beyond which it approaches unity, i.e., the network just outputs the input as it is [Fig. 1(b)]. Thus, in the following part, we consider the range of $0 \leqslant \gamma \leqslant 2$.

B. β dependence of the stability of the analytical solution and additional chaotic attractor

Although x^{fp} is a fixed point for any value of parameters, it is necessary to ascertain its stability and the existence of other attractors, to assess whether the recall of x^{fp} really works from any initial states. We numerically solved Eq. (1) ($N = 2048$, unless otherwise noted), and found another chaotic attractor in addition to x^{fp}. By varying α and β, three types of response behaviors are observed depending on the stability of the chaotic attractor, which are characterized by the distinct bifurcation

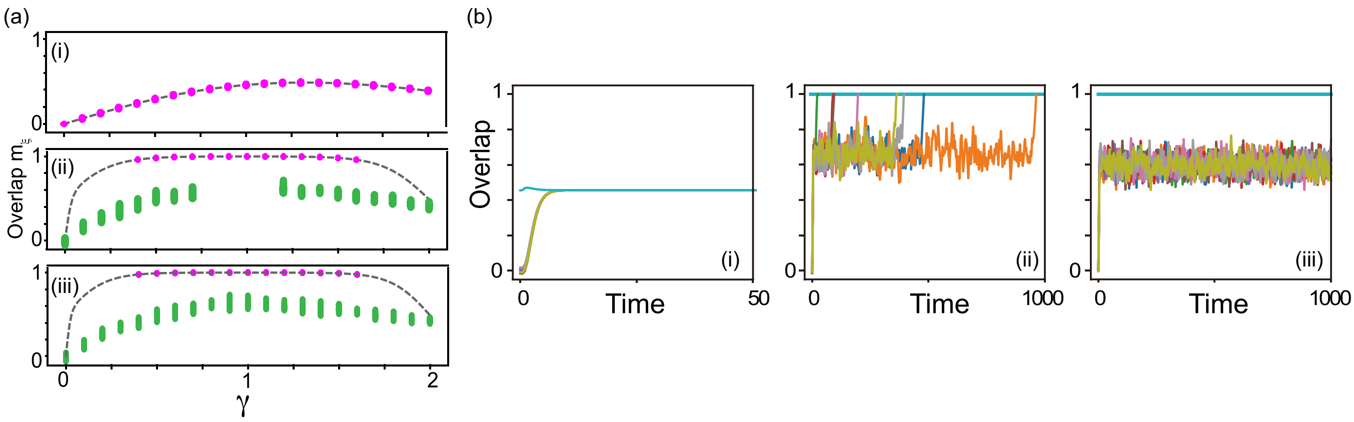

FIG. 2. Three response behaviors to $\boldsymbol{\eta}$ depending on β. (a) The overlaps m_{ξ} against the increase in γ are shown for $\beta = 0.8, 4, 4.7$ in panels (i)–(iii), respectively. In each panel, each of the colored dots for a given γ represents the overlap of x with $\boldsymbol{\xi}^1$ averaged over 100 unit time after the transient period. Magenta and green dots represent convergence into the x^{fp} and the chaotic attractor, respectively. To confirm the stability of x^{fp} and explore another attractor, we sampled the dynamics from 20 random initial states in addition to an initial state equal to x^{fp}. The dotted lines represent the overlap of x^{fp} with $\boldsymbol{\xi}^1$. (b) Dynamics of neural activities are shown with their overlap with the target for $\beta = 0.8$, 3.8, and 8 in (i)–(iii) panels, respectively, for $\gamma = 1$. Panels (i)–(iii) exhibit the response behavior type (i) and (ii) that is close to (iii) and (iii), respectively. Different colored lines represent trials starting from different initial states; one from x^{fp} (in cyan) and the others from states that are uniform randomly chosen from $(-1, 1)^N$. All results are obtained for $\alpha = 0.38$.

diagrams of m_{ξ} against γ. The stability of x^{fp} is not directly related to these types and is analyzed in detail in Appendix A.

(i) Stable recall of x^{fp} for any strength of the input [Figs. 2(a), panel (i) and (b), panel (i)]: x^{fp} is a unique attractor for any γ.

(ii) Stable recall of x^{fp} only for a certain range of γ [Figs. 2(a), panel (ii) and (b), panel (ii)]: x^{fp} is a unique attractor for $\gamma \sim 1$, whereas for smaller γ a chaotic attractor coexists, which exhibits a smaller overlap with ξ compared with the overlap of x^{fp} [19]. For smaller γ values, the neural state fails to converge into x^{fp}, and instead, it converges into the chaotic attractor from most initial states, meaning that the network fails to recall x^{fp}. Still, for $\gamma \sim 1$, the neural state from any initial state converges to x^{fp} whose overlap with the target is close to unity, resulting in the recall of the target.

(iii) No stable recall of x^{fp} for any γ [Figs. 2(a), panel (iii) and (b), panel (iii)]: the chaotic attractor exists across all ranges of γ, even though x^{fp} coexists around $\gamma = 1$. The chaotic attractor has a much larger basin of attraction than x^{fp} even for $\gamma \sim 1$ [Fig. 2(b), panel (iii)]. Consequently, the recall of x^{fp} is impaired.

To analyze these three behaviors, we explored the stability of x^{fp} and of the chaotic attractor across a range of β with a constant $\alpha = 0.38$. We found that for a small value of β ($\beta = 0.8$), the stable recall (i) is achieved. The neural states from any initial states for any γ converge rapidly into x^{fp} [as shown in Fig. 2(b), panel (i)], indicating high robustness in the successful recall of x^{fp}. However, the degree of overlap with the target is notably below the unity [20].

As β increases, x^{fp} approaches the target for all ranges of γ. Beyond the critical β, denoted by β_F, x^{fp} turns to be unstable for a certain range of γ, while the chaotic attractor emerges (see Appendix B), corresponding to the recall type (ii) as shown in Fig. 2(a), panel (ii). The overlap of the chaotic attractor with the target is much lower than that of x^{fp}. Although, for $\gamma = 1$, x^{fp} is the unique attractor; there exists long-term transient chaos before the neural state converges into x^{fp} [see Fig. 2(b), panel (ii) and "Transient chaos" in the following].

With the further increase in β, the range of γ within which the chaotic attractor exists expands, eventually, covering all the range of $0 \leqslant \gamma \leqslant 2$ at another critical value of β (termed β_I). Beyond β_I, the system exhibits the recall type (iii). Even for $\gamma = 1$, the basin of the chaotic attractor covers the full state space, and most orbits from random initial conditions converge into it [Fig. 2(b), panel (iii)]. Thus, the converged states are far from the target.

To comprehensively understand the recall behavior across β and γ, we draw the regions where the chaotic attractor is present in Fig. 3(a). We also investigated the stability of x^{fp}, which, however, is not directly related to the type of recall and is shown in Appendix A. In the area above the curve, the chaotic attractor is present. β_F is the minimum value of the

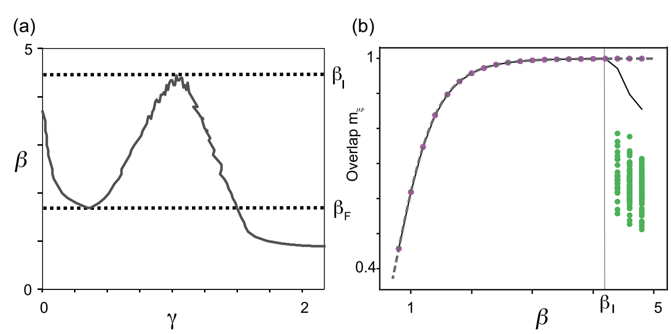

FIG. 3. (a) The boundary separating the parameter regions in the presence of and the absence of the chaotic attractor is plotted. The chaotic attractor is present above the solid line. (b) The overlap at $\gamma = 1$ with the increase in β. Dots represent the overlaps obtained from 100 randomly chosen initial states, while the solid line exhibits the overlap averaged over them. The dotted lines represent the overlap of x^{fp} with $\boldsymbol{\xi}^1$ as also shown in Fig. 2(a).

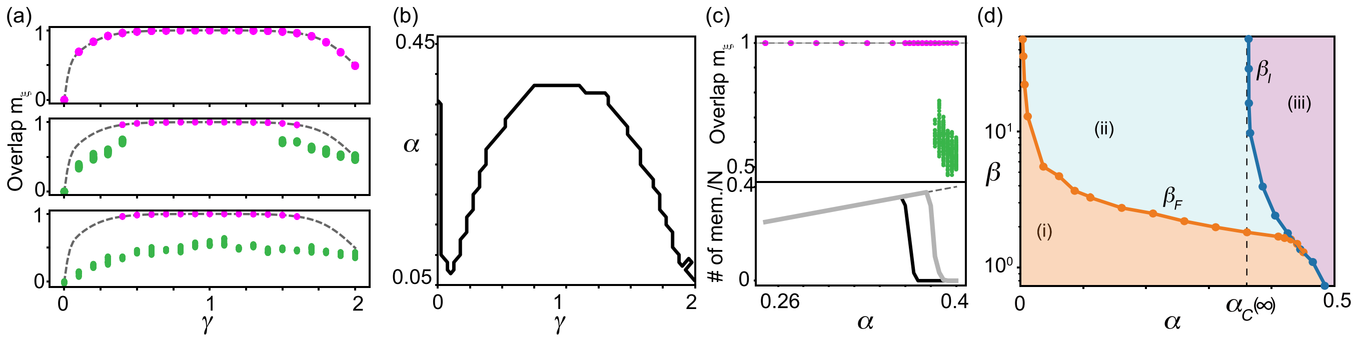

FIG. 4. Three recall behaviors in response to $\boldsymbol{\eta}$ depending on α. (a) The overlaps against the increase in γ are shown for $\alpha = 0.05, 0.3, 0.4$ in panels (i)–(iii), respectively. Colored dots and dotted lines exhibit the neural states and x^{fp} in the same way as in Fig. 2(a). (b) The boundary separating the parameter regions with or without the chaotic attractor is plotted, similar to Fig. 3(a). The chaotic attractor is present above the solid line. All results in (a) and (b) are obtained for $\beta = 4$. (c) (Upper panel) Bifurcation diagram of the overlap at $\gamma = 1$ for $\beta = 4$ with the increase in α is shown in the same way as in (a). (Lower panel) The number of successfully recalled $x^{\text{fp},\mu}$ ($\mu = 1, \ldots, \alpha N$) normalized by N is plotted. Filled lines in gray and black show the behavior for $\beta = 4$ and 32, respectively. The dotted line represents the maximum number of possible memories normalized by N (that is α). (d) Three recall behaviors in response to $\boldsymbol{\eta}$. A phase diagram of the recall regimes (i, ii, and iii) against α and β. β_F (orange) gives the boundary of the stable $\boldsymbol{x}^{\text{fp}}$, while β_I (magenta) shows the border of the impaired recall regime.

curve in $\beta \leqslant 1$, whereas β_I is the maximum value of β on the curve at $\gamma = 1$. These two critical values of β determine the phase boundary of three recall behaviors (i)–(iii). With an increase in β, x^{fp} approaches the target, and accordingly, the final states in all the recall trials overlap almost perfectly with the target below $\beta = \beta_I$ at which the chaotic attractor emerges [Fig. 3(b)]. As β increases beyond β_I, the basin of the x^{fp} attractor shrinks, while that of the chaotic attractor expands. Consequently, the overlap between the final state and the target averaged over randomly chosen initial states significantly decreases, as depicted in Fig. 3(b). Thus, the recall performance reaches its peak (i.e., at the onset of chaos) across all ranges of γ.

C. α dependence of the stability of the analytical solution and additional chaotic attractor

So far, we presented the results with the fixed number of memories αN ($\alpha = 0.38$). Note that standard associative memory models, such as the Hopfield network, recall fails beyond a critical number of memories. We next analyze the change in the response process with increasing α and demonstrate that it exhibits similar behavior with the change with the increases in β: Three types of response behavior emerge as α varies, as shown in Fig. 4(a). For small α, x^{fp} is stable and a unique attractor for any γ [response type (i)]. With the increase in α, the chaotic attractor emerges [21] within a certain range of γ [response type (ii)], and this range expands [Fig. 4(b)]. Finally, the range within which the chaotic attractor is present covers all the ranges of γ [response type (iii)]. In contrast to the change with increasing β, the value of x^{fp} remains unchanged during the increase in α [Fig. 4(c)].

We now explore the number of successfully recalled memories by focusing on the stability of attractors for $\gamma = 1$ in Fig. 4(c). We found that at $\alpha = \alpha_C(\beta)$, the chaotic attractors emerge when any input μ is applied [$\alpha_C(\beta)$ is the inverse function of $\beta_I(\alpha)$]. For $\alpha < \alpha_C$, the fixed points $x^{\text{fp},\mu} = (a\boldsymbol{\xi}^\mu + b\boldsymbol{\eta}^\mu)$ for all μ are stable and successfully recalled, whereas for $\alpha > \alpha_C$, almost all the recall trials fail for all μ due to the emergence of the chaotic attractors whose basins of attraction are much larger than those of $x^{\text{fp},\mu}$ [see Fig. 2(b)]. The number of successfully recalled $x^{\text{fp},\mu}$ equals that of embedded patterns αN below $\alpha = \alpha_C(\beta)$ and then drops to zero drastically [see Fig. 4(c)] signifying that $\alpha_C(\beta)N$ is the memory capacity in this model [e.g., $\alpha_C(4) = 0.38$]. $\alpha_C(\beta)$ decreases toward a certain finite value $\alpha_C(\infty)$ with the increase in β as analyzed in detail in the following.

By taking these results together, we show the phase diagram of the response process against α and β by identifying $\beta_F(\alpha)$ and $\beta_I(\alpha)$, as shown in Fig 4(d). As α approaches zero, β_F diverges, meaning that if α is set to a sufficiently small value, x^{fp} is stable throughout all γ, even for quite large β. For such a limit, x^{fp} matches $\boldsymbol{\xi}$ for $0 < \gamma < 2$. Consequently, the network perfectly recalls the target for $0 < \gamma < 2$. Here, β_I increases drastically as α decreases from 0.5 and diverges at $\alpha_C(\infty)$. For α below $\alpha_C(\infty)$, the neural state converges to x^{fp} for $\gamma = 1$ even for $\beta \to \infty$. The asymptotic analysis demonstrates that $\alpha_C(\infty) \sim 0.340$ for $N \to \infty$ (See Appendix C), indicating that the memory capacity is $\alpha = 0.340$ when β is sufficiently large.

D. Transient chaos

We finally investigated the transient behavior to reach x^{fp}, just below the transition point of α_C to the chaotic attractor. Generally, with approaching the transition point to the chaos, the long-transient chaotic behavior before converging to the fixed point is observed as is studied as transient chaos [22]. To analyze the behavior in detail, we computed the transient time before the neural state converges into x^{fp} for α that approaches α_C.

Figure 5 exhibits the distributions of the transient time for different α. We found that the distributions show long-tailed behaviors and these tails are extended as α is increased toward α_C (see Fig. 8 in Appendix C for detail). To quantify the change in these distributions, we fit these distributions by Weibull distribution $W(x; m, \lambda) = \frac{m}{\lambda}(\frac{x}{\lambda})^{m-1}\exp(-(\frac{x}{\lambda})^m)$ as shown in Fig. 5. The parameters for the fitting are

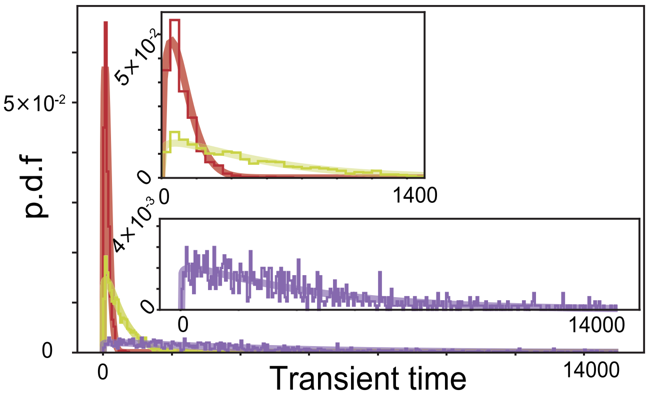

FIG. 5. The distributions of the transient time for $N = 2048$ are plotted for $\alpha = 0.355, 0.36$, and 0.365 in red, yellow, and purple, respectively. Here, $\alpha_C = 0.37$ (see Appendix C). (inset) Expansion of the distributions for $\alpha = 0.355, 0.36$.

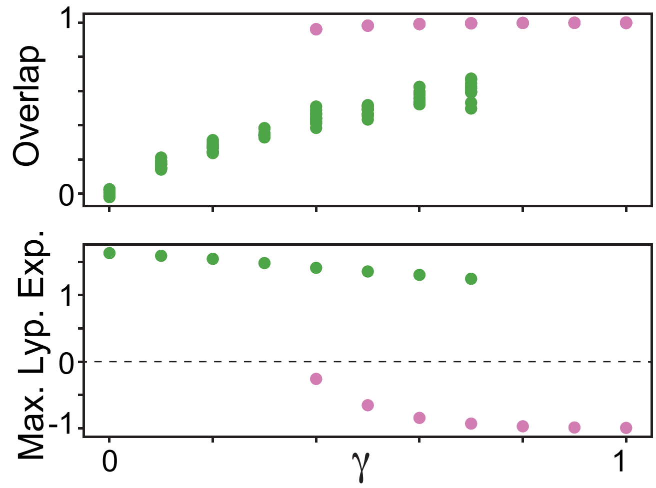

FIG. 7. The maximum Lyapunov exponent for $(\alpha, \beta) = (0.38, 4)$ is plotted in the lower panel. For reference, the overlap with x^{fp} is also plotted in the same way as in Fig. 2(a) in the upper panel. The points in magenta in the upper and lower panels show the overlaps and the Lyapunov exponent for x^{fp}, respectively, while those in green show them for the chaotic attractors.

$(m, \lambda) = (1.44, 258)$, (1.13, 1080), and (1.06, 8540) for $\alpha = 0.355, 0.36$, and 0.365, respectively. Note that $m = 1$ indicates that the distribution is exponential one, which is typically observed in the transient chaos; it is suggested that the distribution of the transient time approaches a standard behavior of the transient chaos as α approaches α_C (namely the onset of chaos), whereas there is a deviation from it given by the Weibull distribution for smaller α.

IV. SUMMARY AND DISCUSSION

In summary, we present an analytically solvable neural network model for I/O associations in which each input stabilizes the target pattern as a unique fixed point under the memory capacity for $\beta \to \infty$. The network's connectivity comprises both target and input patterns with the pseudoinverse matrix, which allows for recalls of any (correlated) targets without interference. This is in contrast to our previous model [12] in which the interference among targets and inputs hinders the recall of the target unless the patterns are mutually orthogonal to each other. The interference also masks the transition from chaos to the fixed point, which makes the classification of the response behaviors harder. In contrast, in the present model, by virtue of the elimination of interference, the analytical expression of the fixed point in the recall of any memory for any input strength is derived, whereas the response dynamics were explored in random networks (without embedded patterns) [23–25] and low-rank networks [26,27]. We also numerically demonstrate the emergence of the additional chaotic attractor. By exploring the stability of these attractors, we identified three distinct response processes.

Introducing the pseudoinverse matrix [X^+ in Eq. (3)] into the connectivity generally requires the global information of the network, which may be difficult to implement biologically (see [16] for Hopfield network). In our previous study [13,28], however, a Hebbian and anti-Hebbian learning that only requires local information can shape the connectivity that is similar to our current connectivity. Still, further studies need

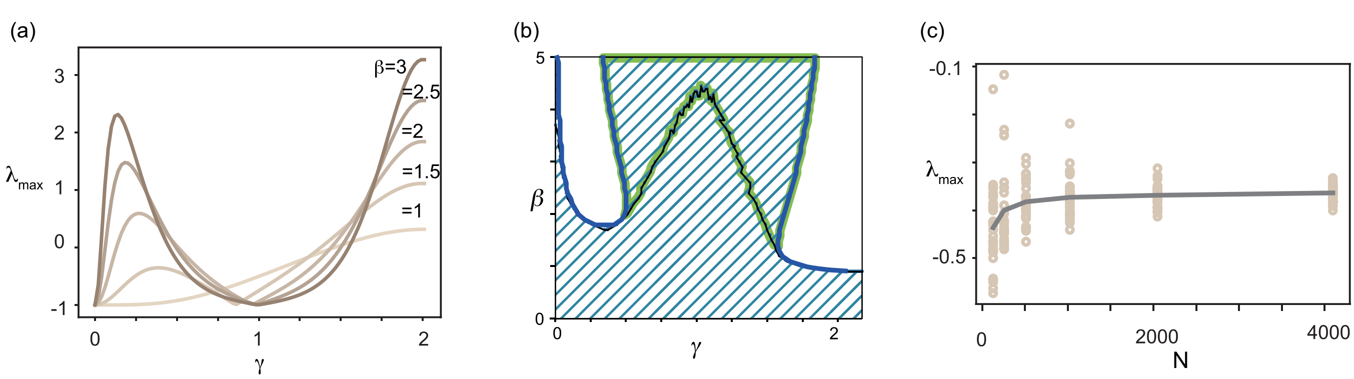

FIG. 6. The stability of x^{fp} dependence on β, γ and N. (a) Maximum eigenvalue $\lambda_{\max}$ of the Jacobian matrix of $F(x)$ in Eq. (1) at x^{fp} is plotted for varying γ and β ($N = 2048$). (b) Phase diagram of the stability of x^{fp} against γ and β is shown. The blue curve represents the line at which $\lambda_{\max} = 0$, while the black line shows the boundary at which the chaotic attractor loses its stability given in Fig. 3(a) for reference. (c) Dependence of $\lambda_{\max}$ on N is plotted. $\lambda_{\max}$ of ten realizations of networks are computed for $\beta = 1.5$ and $\gamma = 0.4$ and are plotted as ten circles. The black solid line is the value averaged over ten networks.

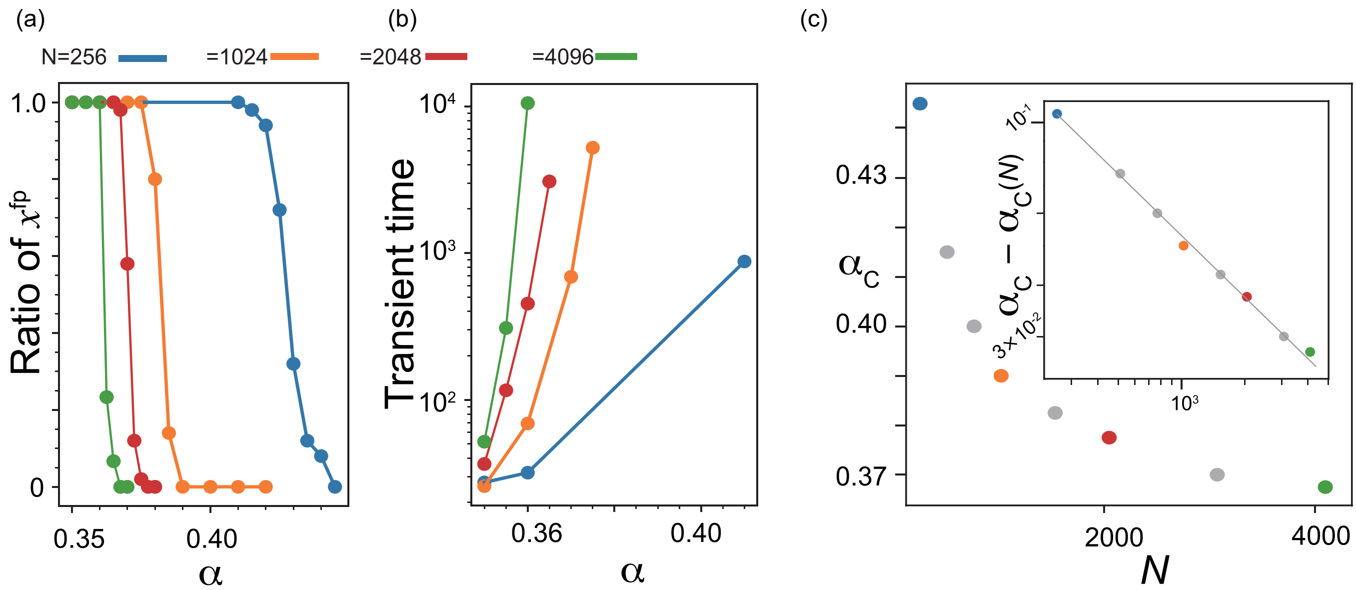

FIG. 8. (a) The ratio of trajectories converging into $x^{\rm fp}$ is plotted with the increase in α for different N. We calculated the ratio over ten initial states in each of the ten realizations of networks for each α. (b) The transient time before convergence into $x^{\rm fp}$. We calculated the transient time with the increase in α. The transient time diverges quickly when α approaches the transition point α_C. (c) Dependence of α_C on N. (inset) $\alpha_C(N) - \alpha_C(\infty)$ on the double logarithmic plot.

to fill the gap between the learning-shaped connectivity and the current connectivity.

Here, we identified three phases of responses, concerning the dominance of the chaotic attractor. Interestingly, the recall performance of the target is maximized at the onset of the chaos for $\gamma = 1$. For this parameter range, the spontaneous chaotic activity is bifurcated to the fixed point that corresponds to the target output with the increase in γ. In fact, similar transitions in activities due to stimulus changes are observed in various cortical areas [29,30]. These findings are consistent with our findings of the optimal performance under spontaneous chaotic dynamics, whereas the roles of the chaotic dynamics in the response and learning require further elucidation. Indeed, the relevance of spontaneous chaotic (and high-dimensional) dynamics to computational neuroscience has been discussed in areas such as reservoir computing [31–34], memories [35], mixed selectivity for efficient separation [36], sampling [37], neural avalanche [38,39], and learning [40,41]. Our study has demonstrated a new role of chaotic dynamics in recall performance.

Further, on the regime for the optimal recall, the long transient from the chaotic state to $x^{\rm fp}$ matching with the target is observed (Fig. 5). The distribution of the transient time before reaching the target follows the Weibull distribution below the transition point α_C and the exponential distribution at the vicinity of the transition point. The Weibull distribution is also observed in reaction time in human decision-making tasks [42]. The decision process is usually modeled by a mutually inhibited cluster network that generates bistable attractors corresponding to the decision state [43], which is not related to the chaotic behavior. On the other hand, the exponential distribution is typical in the transient chaos [22]. Our model, thus, could shed light on the role of chaotic behavior in the decision-making process in neural systems.

Although Hopfield networks [9,10] and their variants [14–16] made significant contributions to associative memory, the modulation of the neural dynamics by external input that is essential for performing cognitive functions has not been included. Our model presents a prototype connectivity underlying such modulation, which advances our understanding of neural processing.

ACKNOWLEDGMENTS

We thank Tatsuya Haga and Shun Ogawa for valuable discussion. The present work is supported by JSPS KAKENHI (No. 20H00123, T.K. and K.K.), NNF21OC0065542 (K.K.), and Special Research Expenses in Future University Hakodate (T.K.). K.K. is also supported by MIC under a grant entitled "R&D of ICT Priority Technology 421 (JPMI00316)".

APPENDIX A: DETAILED ANALYSIS FOR THE STABILITY OF $x^{\rm fp}$

We analyze the stability of $x^{\rm fp}$ for varying β and γ, while α is fixed at 0.38. To evaluate the stability quantitatively, we calculated the maximum eigenvalue $\lambda_{\rm max}$ of the Jacobian matrix of $F(x)$ in Eq. (1) at $x^{\rm fp}$ as a function of γ, plotted in Fig. 6(a). For $\beta = 1$, $\lambda_{\rm max}$ equals -1 at $\gamma = 0$ and increases monotonically as γ increases. Finally, at $\gamma = 1.4$, it turns positive. For the larger value of β, $\lambda_{\rm max}$ around $\gamma = 0.3$ raises rapidly, which is finally positive, while the interval with the positive eigenvalues above $\gamma = 1.4$ still remains. Thus, for the higher value of β, $x^{\rm fp}$ is unstable around $\gamma = 0.3$ and above $\gamma = 1.4$.

To understand the change of the stability of $x^{\rm fp}$ comprehensively, we exhibit the boundary at which $\lambda_{\rm max}$ turns positive in

Fig. 6(b). In the blue-shaded area, x^{fp} is stable, whereas x^{fp} is unstable in the white area. Either the x^{fp} attractor or the chaotic attractor exists for any parameter space. Especially, x^{fp} and the chaotic attractors coexist within the area enclosed by the green line.

Finally, we investigate the variability of $\lambda_{\max}$ with N increased, plotted in Fig. 6(c). The variability over different networks reduces as N increases, implying that $\lambda_{\max}$ is determined independent of network realizations and, consequently, the boundary in Fig. 6(b) is also determined independent of the realization of networks, if N is sufficiently large.

APPENDIX B: LYAPUNOV EXPONENT ANALYSIS

To quantitatively analyze the chaotic behavior in response dynamics in Fig. 2, we calculated the maximum Lyapunov exponent for the neural dynamics on different γ. Figure 7 exhibits the maximum Lyapunov exponent for $(\alpha, \beta) = (0.38, 4)$ as a function of γ. For $0.4 \leqslant \gamma \leqslant 0.7$, x^{fp} and the chaotic attractor coexist as in Fig. 6(b).

APPENDIX C: ESTIMATION OF THE CAPACITY $\alpha_C(\infty)$

The capacity $\alpha_C(\beta)N$ of the number of successfully recalled patterns is defined as α above which the chaotic attractor emerges, namely the transition value of α from the response type (ii) to (iii). To estimate the capacity, we analyze the transition from x^{fp} to the chaotic attractor for quite high β, here $\beta = 32$, depending on α with $N \to \infty$. γ is set at one.

First, we numerically computed the ratio of trajectories converging into x^{fp} by averaging over initial states, as shown in Fig. 8(a). The ratio that is equal to unity indicates the response (ii), whereas that equal to zero means that all trajectories converge into the chaotic attractor and consequently the response (iii). We explored the ratio by increasing α and found a clear transition from unity to zero at a certain value of α, α_C. Simultaneously, the transient time before convergence to x^{fp} is increased rapidly as α increases, as shown in Fig. 8(b). As N increases, α_C decreases.

To estimate α_C at $N \to \infty$, we calculated α_C with increasing N. α_C monotonically decreases as the increase in N as shown in Fig. 8(b). By fitting α_C as a function of N by $aN^{-1/2} + \alpha_C(\infty)$, we derived $\alpha_C(\infty) = 0.340$ ($a = 1.67$).

[1] A. Luczak, P. Bartho, and K. D. Harris, Spontaneous events outline the realm of possible sensory responses in neocortical populations, Neuron **62**, 413 (2009).

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

[3] J. Wang, D. Narain, E. A. Hosseini, and M. Jazayeri, Flexible timing by temporal scaling of cortical responses, Nat. Neurosci. **21**, 102 (2018).

[4] M. Minsky and P. Seymour, *Perceptrons* (MIT Press, Cambridge, 1969).

[5] N. Aloysius and M. Geetha, A review on deep convolutional neural networks, in *2017 International Conference on Communication and Signal Processing (ICCSP)* (IEEE, 2017) pp. 0588–0592.

[6] O. Barak, Recurrent neural networks as versatile tools of neuroscience research, Curr. Opin. Neurobiol. **46**, 1 (2017).

[7] W. Chaisangmongkon, S. K. Swaminathan, D. J. Freedman, and X.-J. Wang, Computing by robust transience: How the fronto-parietal network performs sequential, category-based decisions, Neuron **93**, 1504 (2017).

[8] T. Kurikawa, T. Haga, T. Handa, R. Harukuni, and T. Fukai, Neuronal stability in medial frontal cortex sets individual variability in decision-making, Nat. Neurosci. **21**, 1764 (2018).

[9] S.-I. Amari, Neural theory of association and concept-formation, Biol. Cybern. **26**, 175 (1977).

[10] J. J. Hopfield, Neurons with graded response have collective computational properties like those of two-state neurons, Proc. Nat. Acad. Sci. USA **81**, 3088 (1984).

[11] D. Amit, H. Gutfreund, and H. Sompolinsky, Statistical mechanics of neural networks near saturation, Ann. Phys. **173**, 30 (1987).

[12] T. Kurikawa and K. Kaneko, Associative memory model with spontaneous neural activity, EPL (Europhys. Lett.) **98**, 48002 (2012).

[13] T. Kurikawa and K. Kaneko, Embedding responses in spontaneous neural activity shaped through sequential learning, PLoS Comput. Biol. **9**, e1002943 (2013).

[14] L. Personnaz, I. Guyon, and G. Dreyfus, Collective computational properties of neural networks: New learning mechanisms, Phys. Rev. A **34**, 4217 (1986).

[15] I. Kanter and H. Sompolinsky, Associative recall of memory without errors, Phys. Rev. A **35**, 380 (1987).

[16] S. Diederich and M. Opper, Learning of correlated patterns in spin-glass networks by local learning rules, Phys. Rev. Lett. **58**, 949 (1987).

[17] Eqs. (6) and (7) are applicable not only to $\tanh(\cdot)$ but also to any arbitrary antisymmetric function. However, the behaviors as shown in Fig. 1 are held for $f(x) = \tanh(\beta x)$.

[18] When $\beta \to \infty$, $a = 1$, and $b = 0$ for $0 < \gamma < 2$, while $a = 0$ and $b = 1$ for $2 < \gamma$, which are obtained from Eqs. (6) and (7).

[19] x^{fp} and the chaotic attractor coexist for some range of parameters and do not otherwise. Even when they coexist, the neural states converge into the chaotic attractor for almost all initial states, as shown in Fig. 2(b), panel (iii). Thus, the type of response is independent of their coexistence.

[20] The mismatch between x^{fp} and the target is just because the activation function cannot reach ± 1 for finite beta. If the output is transformed to a binarized pattern, it matches rigorously the target. In this mean, our model generates optimal output even for any beta.

[21] The chaotic attractors are analyzed by computing the Lyapunov exponent in Appendix B. In the spontaneous chaotic attractor, the neural state approaches and departs from the target patterns intermittently, which is distinctive from the chaotic behaviors in random network [44].

[22] T. Tél, The joy of transient chaos, Chaos **25**, 097619 (2015).

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

[24] T. Toyoizumi and L. F. Abbott, Beyond the edge of chaos: Amplification and temporal integration by recurrent networks in the chaotic regime, Phys. Rev. E **84**, 051908 (2011).

[25] S. Takasu and T. Aoyagi, Suppression of chaos in a partially driven recurrent neural network, arXiv:2306.00900 (2023).

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

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

[28] T. Kurikawa, O. Barak, and K. Kaneko, Repeated sequential learning increases memory capacity via effective decorrelation in a recurrent neural network, Phys. Rev. Res. **2**, 023307 (2020).

[29] M. M. Churchland, B. M. Yu, J. P. Cunningham, L. P. Sugrue, M. R. Cohen, G. S. Corrado, W. T. Newsome, A. M. Clark, P. Hosseini, B. B. Scott, D. C. Bradley, M. a. Smith, A. Kohn, J. A. Movshon, K. M. Armstrong, T. Moore, S. W. Chang, L. H. Snyder, S. G. Lisberger, N. J. Priebe, I. M. Finn, D. Ferster, S. I. Ryu, G. Santhanam, M. Sahani, and K. V. Shenoy, Stimulus onset quenches neural variability: A widespread cortical phenomenon, Nat. Neurosci. **13**, 369 (2010).

[30] J. J. Niessing and R. W. Friedrich, Olfactory pattern classification by discrete neuronal network states, Nature (London) **465**, 47 (2010).

[31] 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).

[32] H. Jaeger and H. Haas, Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication, Science **304**, 78 (2004).

[33] N. Bertschinger and T. Natschläger, Real-time computation at the edge of chaos in recurrent neural networks, Neural Comput. **16**, 1413 (2004).

[34] R. Legenstein and W. Maass, Edge of chaos and prediction of computational performance for neural circuit models, Neural Netw. **20**, 323 (2007).

[35] I. Tsuda, Toward an interpretation of dynamic neural activity in terms of chaotic dynamical systems, Behavioral and Brain Sciences **24**, 793 (2001).

[36] S. Fusi, E. K. Miller, and M. Rigotti, Why neurons mix: High dimensionality for higher cognition, Curr. Opin. Neurobiol. **37**, 66 (2016).

[37] Y. Terada and T. Toyoizumi, Chaotic neural dynamics facilitate probabilistic computations through sampling, bioRxiv 2023.05.04.539470 (2023), doi:10.1101/2023.05.04.539470.

[38] T. Petermann, T. C. Thiagarajan, M. A. Lebedev, M. A. L. Nicolelis, D. R. Chialvo, and D. Plenz, Spontaneous cortical activity in awake monkeys composed of neuronal avalanches, Proc. Nat. Acad. Sci. USA **106**, 15921 (2009).

[39] Ł. Kuśmierz, S. Ogawa, and T. Toyoizumi, Edge of chaos and avalanches in neural networks with heavy-tailed synaptic weight distribution, Phys. Rev. Lett. **125**, 028101 (2020).

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

[41] A. E. Orhan and W. J. Ma, A diverse range of factors affect the nature of neural representations underlying short-term memory, Nat. Neurosci. **22**, 275 (2019).

[42] E. M. Palmer, T. S. Horowitz, A. Torralba, and J. M. Wolfe, What are the shapes of response time distributions in visual search? J. Exp. Psychol. Hum. Percept. Perform. **37**, 58 (2011).

[43] K.-F. Wong and X.-J. Wang, A recurrent network mechanism of time integration in perceptual decisions, J. Neurosci. **26**, 1314 (2006).

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