Discrete-Attractor-like Tracking in Continuous Attractor Neural Networks

Chi Chung Alan Fung (馮志聰)* and Tomoki Fukai (深井朋樹)†
RIKEN Center for Brain Science, Hirosawa 2-1, Wako City, Saitama 351-0198, Japan

(Received 21 August 2018; revised manuscript received 31 October 2018; published 8 January 2019)

Continuous attractor neural networks generate a set of smoothly connected attractor states. In memory systems of the brain, these attractor states may represent continuous pieces of information such as spatial locations and head directions of animals. However, during the replay of previous experiences, hippocampal neurons show a discontinuous sequence in which discrete transitions of the neural state are phase locked with the slow-gamma (~30–50 Hz) oscillation. Here, we explore the underlying mechanisms of the discontinuous sequence generation. We find that a continuous attractor neural network has several phases depending on the interactions between external input and local inhibitory feedback. The discrete-attractor-like behavior naturally emerges in one of these phases without any discreteness assumption. We propose that the dynamics of continuous attractor neural networks is the key to generate discontinuous state changes phase locked to the brain rhythm.

DOI: 10.1103/PhysRevLett.122.018102

Ample evidence shows that the hippocampus performs sequence processing during the acquisition, consolidation, and retrieval of memory. The latter characteristic has been extensively studied in spatial navigation tasks, in which hippocampal neurons in awake [1] and sleep states [2,3] replay the firing sequences that were exhibited during the preceding spatial experiences. These replay events are internally generated by the hippocampal circuits and believed to be part of neural processes for memory consolidation [4–6]. On the other hand, the hippocampus has long been thought to operate as an attractor neural network in which episodes may be encoded into fixed-point attractors [7], and this hypothesis also receives some support from experiment. However, how the seemingly different dynamical characteristics, attractor dynamics, and sequence generation emerge and cooperate in memory processing remains a mystery.

Recently, this question was addressed in the activity of CA1 neurons during a spatial memory task [8]. Results of this experiment revealed that replay sequence during sleep, which represents sequence of the spatial locations visited previously by the animal, is discontinuous. Unlike smooth sequential firing during exploration in awake states, the sequence exhibited abrupt jumps between the decoded locations during replay events. Furthermore, these jumps were phase locked to the slow-gamma oscillation of the local field potentials (LFPs). In contrast, the hippocampal circuit showed attractorlike behavior during the replay of disconnected locations. These results suggested that slow decoding of accurate spatial information and fast movements between the remembered locations are alternated during spatial exploration.

However, the underlying mechanisms of this oscillatory coding remain unknown. Here, we propose a continuous attractor neural network (CANN) to account for the emergence of discretized local attractors. The CANN is a family of neural field models that can support a set of continuously connected attractor states, which generally form bump-shaped functions in the space of preferred stimuli [9,10]. The CANN models have been used to describe the tuning curves (i.e., place fields) of hippocampal place cells [11], orientation tuning in visual cortex [12], the direction of object movement in the middle temporal cortex [13], and head direction in the entorhinal cortex [14]. In addition, the pairwise correlation predicted by CANN has been observed in experiment [15]. By using a perturbative approach to the neural field dynamics of CANN [16,17], we demonstrate that the network model shows a phase transition from a continuous attractor state to discontinuous attractor states as the speed of an external stimulus is increased. This transition gives a discrete-attractor-like behavior similar to the experimental observation.

We use the following version of a CANN model to describe the activity state $u(x,t)$ of neurons with preferred location x at time t [10,16]:

$$\tau\frac{du(x,t)}{dt} = -u(x,t) + \rho\int dx' J(x,x')u(x',t)^2 B(t)^{-1} + I^{\mathrm{ext}}(x,t), \tag{1}$$

where ρ is their density in the preferred location space and τ is the corresponding time constant. Excitatory coupling function between neurons at x and x' is translational invariant and given as $J(x,x') = (J_0/\sqrt{2\pi}a)\exp\{-|x-x'|^2/(2a)^2\}$, where J_0 controls the average magnitude of excitatory couplings and a represents their average width. The function $B(t) = 1 + k\rho\int dx' u(x',t)^2$ expresses a divisive inhibition

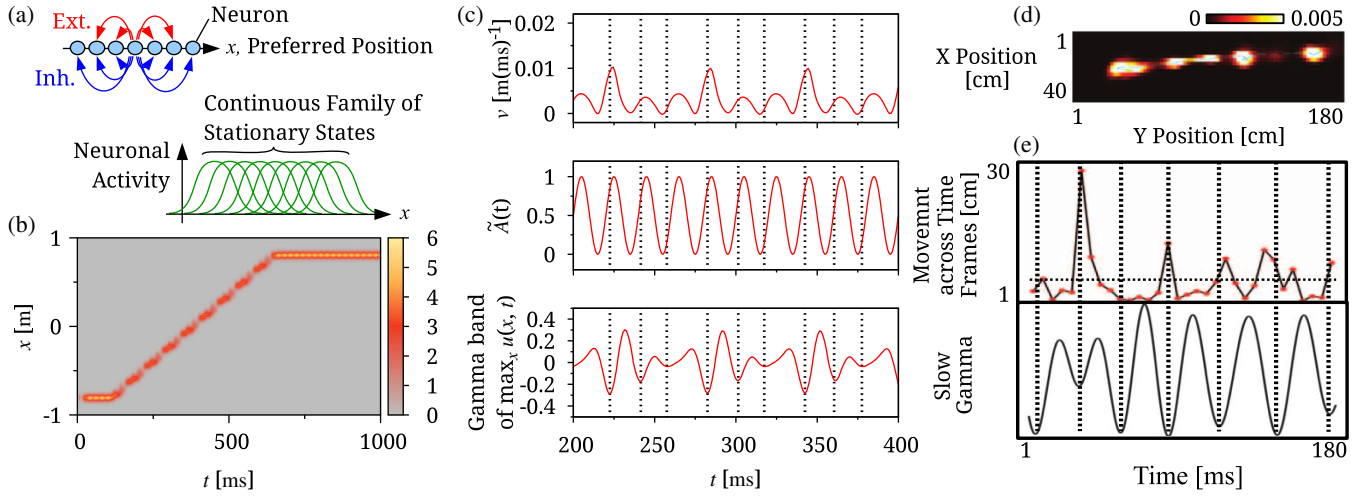

FIG. 1. Basic response properties of CANNs. (a) In CANNs, neurons with similar preferred locations are mutually connected. Excitatory (Ext.) connections have a narrower range than inhibitory (Inh.) connections. This architecture allows the network to support a continuous family of stationary states. (b) Neural responses to a moving input are shown. Parameters are $\tilde{k} = 1.0$, $a = 0.02$ m, $\tau = 2$ ms, $\tilde{v}_I = v_I/a = 0.15$ $(\mathrm{ms})^{-1}$, and $\tilde{A}_0 = 0.5$. (c) The instantaneous speed of the center of the localized activity profile (upper), magnitude of the external input function (middle), and the gamma-band oscillation of the neural field (lower) obtained by a bandpass filter (40–60 Hz) are shown (red curves) for the entire duration of moving input. Dotted lines indicate the troughs of gamma oscillations. Parameters are the same as those in (b). (d) The location of a rat, and (e) movements across 20-ms time frames with 5-ms increments and the slow-gamma power were decoded from hippocampal activity during replay. These panels were modified from Ref. [8] with reprint permission.

[10,18] with a positive parameter k. The larger the value of k, the stronger the inhibition. External input to the network is given by a Gaussian function as $I^{\mathrm{ext}}(x,t) = A(t)\exp\{-|x - z_I(t)|^2/(4a^2)\}$, when the current location of the animal is $z_I(t)$. Because external input influences network dynamics through the component projected onto the translational mode $(\partial u/\partial x)$ of attractor states [17], the specific functional form of external input does not change the essential results. We assume that Eq. (1) describes activity of hippocampal CA1 pyramidal cells. Local excitatory connections are less prominent in CA1 compared to CA3, but CA1 pyramidal cells are not devoid of recurrent synaptic connections [19]. External input may arise from the entorhinal cortex or CA3 [20].

To simplify the parameter dependence of the model, we performed the following rescaling [17]:

$$\begin{aligned} u(x,t) &\to \tilde{u}(x,t) \equiv \rho J_0 u(x,t), \\ k &\to \tilde{k} \equiv 8\sqrt{2\pi}ak/(J_0\rho)^{-1}, \\ A(t) &\to \tilde{A}(t) \equiv \rho J_0 A(t). \end{aligned} \tag{2}$$

By this rescaling, we can express the competition between excitation (J_0) and inhibition (k) with the rescaled inhibition ($\tilde{k}$).

We first study solutions to Eq. (1) for a vanishing external input [$\tilde{A}(t) = 0$]. For $\tilde{k} \geq 1$, we can show that $\tilde{u}(x,t) = 0$ is the only stable fixed-point solution. For $\tilde{k} < 1$, Eq. (1) has a stable fixed point $\tilde{u}_+(x)$ and an unstable fixed point $\tilde{u}_-(x)$, where $\tilde{u}_\pm(x) = \sqrt{8}[(1 \pm \sqrt{1-\tilde{k}})/\tilde{k}]\exp\{-|x-c|^2/(4a^2)\}$, with c being an arbitrary constant. The continuous family of stationary states is schematically illustrated in Fig. 1(a). Now we turn to solutions for nonvanishing input. Owing to local excitatory connections and widely spread inhibition, network activity tends to form a localized profile, even though $\tilde{k} \geq 1.0$. A solution to Eq. (1) is presented in Fig. 1(b). We assumed an oscillatory input within the slow-gamma band $\tilde{A}(t) = \tilde{A}_0[\sin(2\pi f t) + 1]$ with the amplitude and frequency of input set as $\tilde{A}_0 = 0.5$ and $f = 50$ [Hz], respectively. To mimic the observed replay sequence [8], we set $a = 0.02$ [m], which gives a range of network activity profile similar to the experimentally observed one [see Fig. 1(d)], and moved the input at the speed of 0.003 [m $(\mathrm{ms})^{-1}$] from −0.8 [m] to +0.8 [m] (after an initial transient period of 100 ms). In the simulation, $x \in [-1\ \mathrm{m}, 1\ \mathrm{m}]$ with a periodic boundary condition. The evoked activity faithfully tracked the movement of the external input, exhibiting fluctuations in the amplitude.

To explore the biological relevance of the network behavior, we calculated the instantaneous speed of the localized activity profile, $v(t) = dz(t)/dt$, with $z(t) = \int xu(x,t)dx / \int u(x,t)dx$, in the upper panel of Fig. 1(c), which shows discrete-attractor-like behavior during tracking. The time course of the magnitude of the external input is shown in the middle panel for reference. We further examined the phase-locking phenomenon between the gamma rhythm of the "local field potential" and the change of the decoded location, where the amplitude of the neural

field $\tilde{u}(x,t)$ was used to address the LFP. The time evolution of the gamma-band power of $\tilde{u}(x,t)$ is presented in the lower panel of Fig. 1(c). From the two figures [upper and lower panels in Fig. 1(c)], we can see that the peak speed of the activity packet is phase locked to the troughs of gamma-band LFP oscillation in the model. We may compare these results with experimental observations. Figure 1(d) shows neuronal activity recorded from the rat hippocampus during replay [8]. As in Fig. 1(b), the neuronal activity superimposed onto the decoded spatial locations was clearly discretized. Phase locking similar to that in Fig. 1(c) also occurred between the rat's movement and the slow-gamma LFP oscillation [Fig. 1(e)].

At first glance, the discretized response of the model looks trivial due to the presence of an oscillating external drive. However, as demonstrated below, this phenomenon emerges from internal network dynamics but not from the oscillatory drive *per se*. To understand how the dynamics of the continuous attractor model generate discretized responses, we analyzed the model using the perturbative method developed in Refs. [16,17]. In this approach, we construct a perturbative solution to the model by using a series of orthogonal functions.

For mathematical simplicity, we assume that the localized activity profile can be approximated by a Gaussian function: $\tilde{u}(x,t) = \tilde{u}_0(t)\exp\{-|x-z(t)|^2/(4a^2)\}$. Substituting this expression into Eq. (1), rescaling the variables according to Eq. (2), and considering orthogonality of functions, we have

$$\tau\frac{d\tilde{u}_0(t)}{dt} = -\tilde{u}_0(t) + \frac{1}{\sqrt{2}}\frac{\tilde{u}_0(t)^2}{1+\frac{1}{8}\tilde{k}\tilde{u}_0(t)^2} + \tilde{A}(t)e^{(-1/8)\tilde{s}^2}, \tag{3}$$

$$\tau\frac{d\tilde{s}(t)}{dt} = \tau\tilde{v} - \frac{\tilde{A}(t)}{\tilde{u}_0(t)}\tilde{s}e^{(-1/8)\tilde{s}^2}, \tag{4}$$

where $\tilde{s} = (z_I - z)/a$ and $\tilde{v} = (dz/dt)/a$ are the separation and velocity of the activity profile, respectively. We substitute $\tilde{v}_I = (dz_I/dt)/a$ for $\tilde{v}$ to examine whether the network can stably follow the input velocity. The derivation of the above two equations is found in the Supplemental Material [21].

Let $(\tilde{u}_0^*, \tilde{s}^*)$ be a fixed-point solution to Eqs. (3) and (4) for constant $\tilde{A}(t)$, namely, when the external input is a moving Gaussian packet without oscillatory modulations. Figure 2 shows the nullclines of Eqs. (3) and (4), in which a fixed-point solution corresponds to an intersection of the nullclines. The color code indicates the angle of $(d\tilde{u}_0(t)/dt/\tilde{u}_0(t), d\tilde{s}(t)/dt/\tilde{s}(t))$ relative to the horizontal axis in a counterclockwise direction. Four possible scenarios exist depending on parameter values. In Fig. 2(a), three fixed points exist, but only the solution with the smallest $\tilde{s}^*$ is stable. In Fig. 2(b), the value of $\tilde{v}_I$ is increased such that only a fixed point may survive, which corresponds to the fixed point with the largest $\tilde{s}^*$ in Fig. 2(a). This solution,

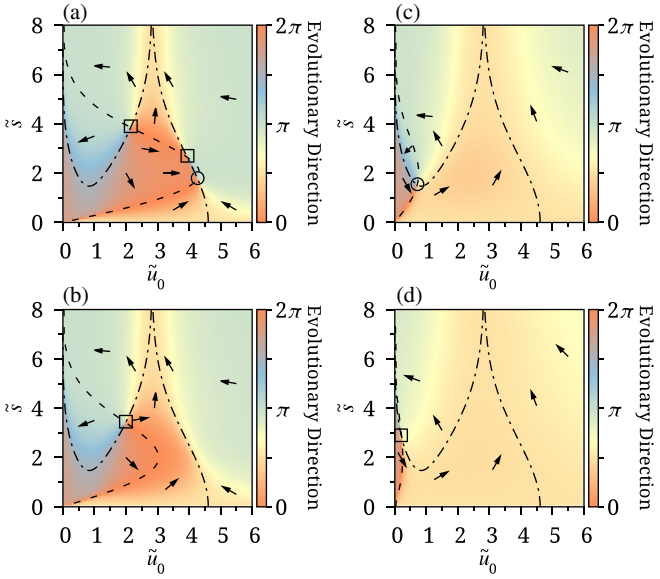

FIG. 2. Phase-plane diagrams for different parameter settings. Here, $\tilde{k} = 1.0$ and $\tilde{A} = 0.5$. (a) Phase-plane diagram shows three fixed-point solutions for $\tilde{v}_I = 0.07$. Dashed and dot-dashed curves indicate $\tilde{s}$ nullcline and $\tilde{u}$ nullcline, respectively. (b) Phase-plane diagram gives an unstable fixed point for $\tilde{v}_I = 0.1$. (c) Phase-plane diagram gives a stable fixed point for $\tilde{v}_I = 0.4$. (d) Phase-plane diagram yields an unstable fixed point for $\tilde{v}_I = 1.2$. Arrows are visualized tendencies of nearby regions. Stable (circles) and unstable (squares) fixed points are indicated.

however, is unstable. In Fig. 2(c), the value of $\tilde{v}_I$ is further increased and that of $\tilde{u}_0^*$ is decreased, as expected from Eq. (4), making $(\tilde{u}_0^*, \tilde{s}^*)$ a stable fixed point. In Fig. 2(d), the fixed point eventually turns unstable and $\tilde{s}(t)$ diverges.

We further studied the behavior of the network by solving Eqs. (3) and (4) in a broader range of parameter values. Figure 3(a) shows $\tilde{u}_0^*$ generally decreases with increases in $\tilde{v}_I$. If, however, $1 < \tilde{k} \lesssim 1.5$, as $\tilde{v}_I$ is increased from zero, the corresponding $\tilde{s}^*$ first increases but then decreases, taking a local maximum at some speed [Fig. 3(b)]. This shows the presence of a parameter region in which the average separation is larger compared to the neighboring regions. A linear stability analysis was preformed around solutions to Eqs. (3) and (4) and the real part of the lowest eigenvalue is shown in Fig. 3(c) for various parameter values. The fixed-point solution is unstable in the parameter regions in which the real part is positive.

In Figs. 3(d)–3(f), numerical simulations of Eq. (1) were performed to confirm the above predictions of the reduced system defined by Eqs. (3) and (4). In Fig. 3(d), the peak height of the localized activity profile $\langle \max_x \tilde{u}(x,t)\rangle_t$ averaged over $z_I(t) \in [-0.4\text{ m}, 0.4\text{ m}]$ is presented, while the temporal average of separation $\langle \tilde{s}(t)\rangle$ is shown in Fig. 3(e). They are comparable to the fixed-point solutions presented in Figs. 3(a) and 3(b), respectively. Instead of linear stability analysis, we evaluated the standard

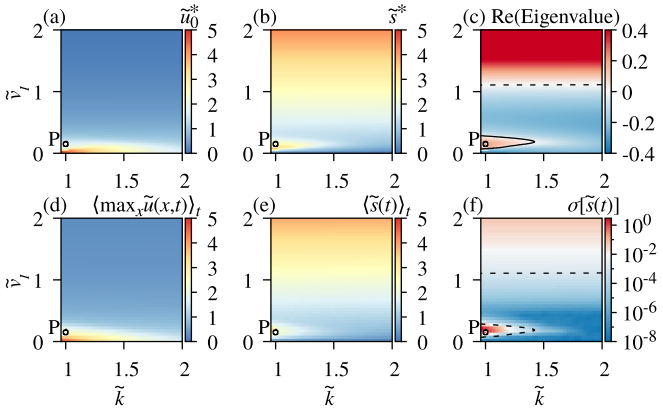

FIG. 3. Parameter dependence of localized activity profiles. (a) $\tilde{u}_0^*$ of the fixed-point solutions to Eqs. (3) and (4). (b) The value of $\tilde{s}^*$ in fixed-point solutions. In (a) and (b), only the fixed-point solutions with the smallest $\tilde{s}^*$ are presented if multiple fixed-point solutions exist. (c) Real part of the eigenvalues of the linearized system around the fixed-point solutions shown in (a) and (b). Dashed curves in (c) and (f) show contours on which the real part vanishes. (d) $\langle \max_x \tilde{u}(x,t) \rangle_t$ measured from simulations of Eq. (1). (e) $\langle \tilde{s}(t) \rangle_t$ measured from the simulations. (f) Standard deviation of the measured $\tilde{s}(t)$. The circles labeled by P show the parameter values used in Fig. 4.

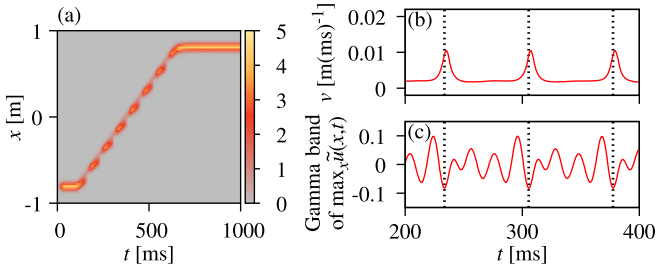

FIG. 4. Simulations with a nonoscillatory input. (a) Activity state $\tilde{u}(x,t)$ evolves with a nonoscillatory input, i.e., $f = 0$. Values of other parameters are the same as those in Fig. 1(b) and labeled by P in Fig. 3. (b) The instantaneous speed of the activity state profile presented in (a). (c) The gamma-band component of $\max_x \tilde{u}(x,t)$ shown in (a). Dotted lines show the aligned troughs of the gamma-band trace in (c).

deviation of $\tilde{s}(t)$ in Fig. 3(f). By comparing this and Fig. 3(c), we find that the fixed-point solution is unstable in the parameter regions where the standard deviation is larger than the minimum separation between neurons in the original model. The results indicate that the reduced system well replicates the dynamics of the original system.

The above results on the stability of activity profiles were derived for continuous attractor networks receiving constant inputs. Importantly, however, we found that the networks respond similarly to both constant and oscillatory inputs if the values of some parameters are adequately replaced. Actually, as shown in Fig. 4(a), a network receiving a constant input exhibits similar discrete-attractor-like dynamics to those shown in the same network driven by an oscillatory input [Fig. 1(b)]. In both models, the inputs had the same average magnitude, and the driving speed $\tilde{v}_I$ and inhibition $\tilde{k}$ were given by the same point P in Fig. 3. In addition, the instantaneous speed of the activity profile displays phase locking with the gamma-band oscillation of the network activity [Figs. 4(b) and 4(c)]. These results suggest that internal network dynamics, rather than external oscillatory drives, underlie the phase-locking behavior observed in the experiment.

We further investigated the robustness of the oscillatory responses shown in Figs. 1(b), 1(c), and 4. First, we examined how the maximum amplitude $\max_x \tilde{u}(x,t)$ varies with changes in the magnitude $\tilde{A}_0$ and speed $\tilde{v}_I$ of constant input. Figure 5(a) shows the differences between the peak and trough of $\max_x \tilde{u}(x,t)$ during the tracking behavior. The figure shows that oscillatory responses appear in considerably large ranges of $\tilde{v}_I$ and $\tilde{A}_0$. Second, we examined how the magnitude of $\max_x \tilde{u}(x,t)$ changes with the speed and frequency f of oscillatory input [Fig. 5(b)]. The result reveals that the oscillatory responses emerge in a wide range of the frequency. Note that two regimes of oscillatory responses can be seen, which are vaguely separated around $f \approx 30$ Hz. Below this frequency, oscillatory responses are primarily caused by slowly oscillating inputs, while for $f \gtrsim 30$ Hz oscillations occur due to intrinsic network dynamics without oscillatory drives. In the large f limit, the network behavior should agree with that for nonoscillatory inputs because it can only respond to the temporal average of input. Third, we checked the range of f in which the anti-phase-locking shown in Fig. 1(c) occurs. Correlations between the instantaneous speed $\tilde{v} \equiv v/a$ and gamma-band oscillation of $\max_x \tilde{u}(x,t)$ are negative in a broad range of f [Fig. 5(c)], implying that the anti-phase-locking behavior occurs for various frequencies. From these results, we may conclude that the qualitative

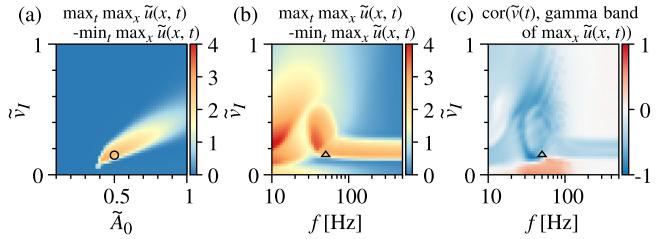

FIG. 5. Robustness of oscillatory responses against changes in $\tilde{v}_I$, $\tilde{A}_0$, and f. The parameter values used in Fig. 4 (circle) and Figs. 1(b) and 1(c) (triangles) are indicated. (a) $[\max_{t,x} \tilde{u}(x,t) - \min_t \max_x \tilde{u}(x,t)]$ versus $\tilde{A}_0$ is shown during the tracking of a constant input moving at speed $\tilde{v}_I$. (b) $[\max_{t,x} \tilde{u}(x,t) - \min_t \max_x \tilde{u}(x,t)]$ versus f is shown during the tracking of an oscillatory input moving at speed $\tilde{v}_I$ and oscillating at frequency f with magnitude $\tilde{A}_0 = 0.5$. (c) Correlations between $\tilde{v}(t)$ and the gamma-band component of $\max_x \tilde{u}(x,t)$ are presented during tracking of an oscillatory input moving at speed $\tilde{v}_I$ and oscillating at frequency f. Parameter values are $\tilde{k} = 1.0$ and $a = 0.02$ [m].

behavior shown in Figs. 1(b) and 1(c) is not due to a particular choice of parameter values.

The oscillatory response reported here is a result of the competition between attractor dynamics and a moving external input. We note that the external input needs not be oscillatory because the oscillatory responses are owing to the intrinsic network dynamics. It was previously shown that continuous attractors exist only for an adequate range of inhibition [16,17]. If the inhibitory feedback is too strong, only small-sized activity packets can exist in the network. However, these packets are unable to reflect any property of attractor dynamics, such as the presence of fixed points, because recurrent inputs from surrounding neurons are much inferior to the external input. Conversely, if the inhibition is too weak, excitatory interactions between neurons become too strong to stably maintain the activity packets. Our results indicate that the oscillatory coding of spatial information with discrete-attractor-like states [8] is possible only if the strength of inhibitory feedback falls into a marginal range. A testable prediction is that these states are impaired by the partial suppression of inhibition without crucially disturbing spatial memory.

In summary, this study presents the neural mechanisms that possibly underlie memory processing through oscillatory information coding. The slow-gamma oscillation is thought to play an active role in memory retrieval and consolidation during sleep or immobility [22–24]. How the intrinsic network behavior may contribute to memory consolidation is open for future studies.

This work was partly supported by Grants-in-Aid for Scientific Research (KAKENHI) from MEXT (17H06036, 18H05213) and CREST, JST (JPMJCR13W1) to T. F.

*chichung.fung@riken.jp
†tfukai@riken.jp

[1] H. S. Kudrimoti, C. A. Barnes, and B. L. McNaughton, J. Neurosci. **19**, 4090 (1999).
[2] C. Pavlides and J. Winson, J. Neurosci. **9**, 2907 (1989).
[3] W. E. Skaggs and B. L. McNaughton, Science **271**, 1870 (1996).
[4] O. Eschenko, W. Ramadan, M. Molle, J. Born, and S. J. Sara, Learn. Mem. **15**, 222 (2008).
[5] G. Girardeau, K. Benchenane, S. I. Wiener, G. Buzsáki, and M. B. Zugaro, Nat. Neurosci. **12**, 1222 (2009).
[6] M. F. Carr, S. P. Jadhav, and L. M. Frank, Nat. Neurosci. **14**, 147 (2011).
[7] J. J. Knierim and J. P. Neunuebel, Neurobiol. Learn. Mem. **129**, 38 (2016).
[8] B. E. Pfeiffer and D. J. Foster, Science **349**, 180 (2015).
[9] S.-I. Amari, Biol. Cybern. **27**, 77 (1977).
[10] S. Wu and S.-I. Amari, Neural Comput. **17**, 2215 (2005).
[11] A. Samsonovich and B. L. McNaughton, J. Neurosci. **17**, 5900 (1997).
[12] R. Ben-Yishai, R. L. Bar-Or, and H. Sompolinsky, Proc. Natl. Acad. Sci. U.S.A. **92**, 3844 (1995).
[13] T. D. Albright, J. Neurophysiol. **52**, 1106 (1984).
[14] K. Zhang, J. Neurosci. **16**, 2112 (1996).
[15] K. Wimmer, D. Q. Nykamp, C. Constantinidis, and A. Compte, Nat. Neurosci. **17**, 431 (2014).
[16] C. C. A. Fung, K. Y. M. Wong, and S. Wu, Europhys. Lett. **84**, 18002 (2008).
[17] C. C. A. Fung, K. Y. M. Wong, and S. Wu, Neural Comput. **22**, 752 (2010).
[18] S. Deneve, P. E. Latham, and A. Pouget, Nat. Neurosci. **2**, 740 (1999).
[19] M. J. Bezaire and I. Soltesz, Hippocampus **23**, 751 (2013).
[20] J. Yamamoto and S. Tonegawa, Neuron **96**, 217 (2017).
[21] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.122.018102 for the derivation of Eqs. (3) and (4) using the perturbative method.
[22] M. F. Carr, M. P. Karlsson, and L. M. Frank, Neuron **75**, 700 (2012).
[23] S. M. Montgomery and G. Buzsáki, Proc. Natl. Acad. Sci. U.S.A. **104**, 14495 (2007).
[24] A. Johnson and A. D. Redish, J. Neurosci. **27**, 12176 (2007).