LETTER Communicated by Stefano Panzeri

Dynamics and Computation of Continuous Attractors

Si Wu
siwu@sussex.ac.uk
Department of Informatics, University of Sussex, Brighton BN1 9QH, U.K.

Kosuke Hamaguchi
kosuke.hamaguchi@univ-paris5.fr
Shun-ichi Amari
amari@brain.riken.go.jp
Amari Research Unit, RIKEN Brain Science Institute, Saitama 351-0198, Japan

Continuous attractor is a promising model for describing the encoding of continuous stimuli in neural systems. In a continuous attractor, the stationary states of the neural system form a continuous parameter space, on which the system is neutrally stable. This property enables the neutral system to track time-varying stimuli smoothly, but it also degrades the accuracy of information retrieval, since these stationary states are easily disturbed by external noise. In this work, based on a simple model, we systematically investigate the dynamics and the computational properties of continuous attractors. In order to analyze the dynamics of a large-size network, which is otherwise extremely complicated, we develop a strategy to reduce its dimensionality by utilizing the fact that a continuous attractor can eliminate the noise components perpendicular to the attractor space very quickly. We therefore project the network dynamics onto the tangent of the attractor space and simplify it successfully as a one-dimensional Ornstein-Uhlenbeck process. Based on this simplified model, we investigate (1) the decoding error of a continuous attractor under the driving of external noisy inputs, (2) the tracking speed of a continuous attractor when external stimulus experiences abrupt changes, (3) the neural correlation structure associated with the specific dynamics of a continuous attractor, and (4) the consequence of asymmetric neural correlation on statistical population decoding. The potential implications of these results on our understanding of neural information processing are also discussed.

1 Introduction

External stimuli are encoded in neural activity patterns in the brain. The brain can reliably retrieve the stored information even when external inputs are incomplete or noisy, achieving the so-called associative memory or invariant object recognition. Mathematically this can be described as attractor computation, that is, the network dynamics enables the neural system to reach the same stationary state once external inputs fall into its basin of attraction. In the conventional models for attractor computation, such as the Hopfield model (Hopfield, 1984), it is often assumed that the stationary states of the neural system are discretely distributed in the state space, which are called discrete attractors.

Recently progress in both experimental and theoretical studies has suggested that there may exist another form of attractor, called continuous attractors, in biological systems (Amari, 1977; Georgopoulos, Kalaska, Caminiti, & Massey, 1982; Maunsell & Van Essen, 1983; Funahashi, Bruce, & Goldman-Rakic, 1989; Wilson & McNaughton, 1993; Rolls, Robertson, & Georges-François, 1995; Ben-Yishai, Lev Bar-Or, & Sompolinsky, 1995; Zhang, 1996; Seung, 1996; Ermentrout, 1998; Hansel & Sompolinsky, 1998; Taube, 1998; Deneve, Latham, & Pouget, 1999; Wang, 2001; Wu, Amari, & Nakahara, 2002; Stringer, Trappenberg, Rolls, & Aranjo, 2002; Brody, Romo, & Kepecs, 2003; Gutkin, Pinto, & Ermentrout, 2003; Trappenberg, 2003; Wu & Amari, 2005; Chow & Coombes, 2006). This type of attractor is appealing for encoding continuous stimuli, such as the orientation, the moving direction, and the spatial location of objects, or those continuous features that underlie the categorization of complicated objects. In a continuous attractor, the stationary states of the neural system are properly aligned in the state space according to the stimulus values they represent. They form a continuous parameter space, on which the neural system is neutrally stable. Figure 1 illustrates the typical structure difference between a continuous and a discrete attractor. We see that in a discrete point attractor, the system is stable only at the bottom of the bowl, whereas in a continuous line attractor, the system is neutrally stable on the one-dimensional valley.

Neutral stability is the key that distinguishes a continuous attractor from a discrete one. This property enables the neural system to change its stable state rather easily along the attractor space and hence provides the neural system the capacity of tracking time-varying stimuli in real time, an ability that is crucial for the brain to carry out many important computational tasks such as motion control and spatial navigation. On the other hand, neutral stability can have a negative effect on attractor computation. For example, because of neutral stability, the stationary states of the neural system are easily disturbed by input noise. Consequently, it degrades the accuracy of neural decoding (Seung, Lee, Reis, & Tank, 2000; Wang, 2001; Brody et al., 2003; Wu & Amari, 2005).

In this study, by using a simple model whose solution is analytically attractable, we systematically investigate the dynamics and the computational properties of continuous attractors. In particular, we elucidate how the neutral stability of the network dynamics leads to these properties. The investigated issues are (1) what the input noise components are that can or cannot be cleaned by the dynamics of a continuous attractor and their

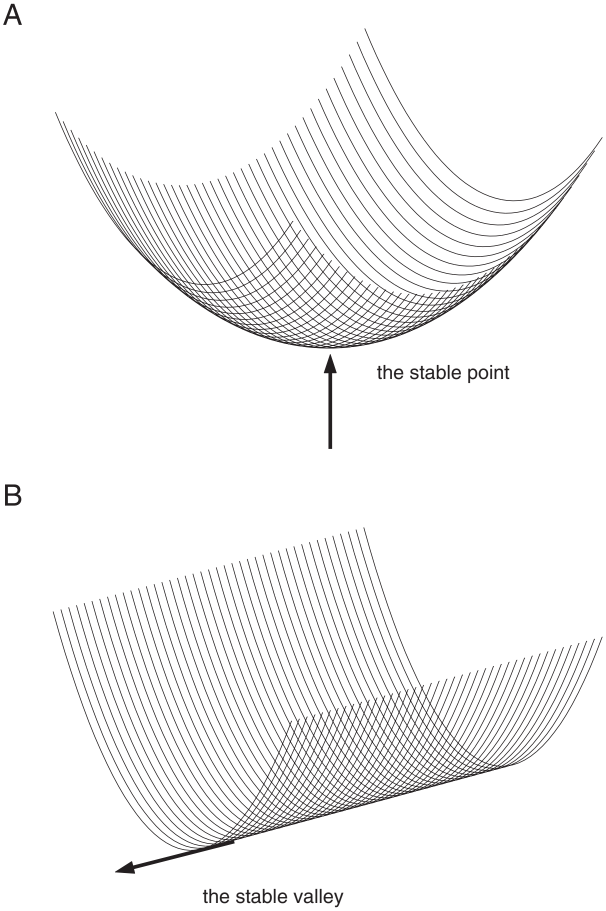

Figure 1: An illustration of the structural difference between discrete and continuous attractors. (A) An example of a discrete point attractor. The system is stable only at the bottom of the bowl. (B) An example of a line attractor, the one-dimensional version of a continuous attractor. The stationary states of the system form a one-dimensional valley. Along the valley, the system is neutrally stable.

effects on the performance of information retrieving; (2) what the tracking speed of a continuous attractor is when external stimuli experience abrupt changes; (3) how the neural correlation structure is shaped by the dynamics of a continuous attractor; and (4) what the consequence is of this correlation structure on statistical population decoding. We point out that some of these issues have been studied in the literature: for instance, Amari (1977) and Seung (1996) have pointed out that a continuous attractor retains the noise component only along the attractor space; Deneve et al. (1999) and Wu et al. (2002) have studied the link between a continuous attractor and neural population decoding; and Ben-Yishai et al. (1995), Spiridon and Gerstner (2001), and Miller (2006) have also identified the asymmetric neural correlation structure in continuous attractors. Compared with these studies, the contribution of this work is that we will elucidate these properties in more detail by giving formal analytical solutions.

A main challenge in analyzing the behavior of a large-size network is to handle the extremely high dimensionality of the network dynamics.[1] Here, by utilizing the specific nature of continuous attractors, we develop a technique to reduce its dimensionality. This takes into account the facts that neural population dynamics is extremely fast and that a continuous attractor can clean those noise components perpendicular to the attractor space very quickly. Therefore, we can project the network dynamics onto the tangent of the attractor space and simplify it to be a one-dimensional Ornstein-Uhlenbeck (OU) process. Simulation results show that this method works well.

Based on this simplified model, we then explore the decoding performance of a continuous attractor under the driving of external noisy inputs. In order to provide clues for checking experimently whether continuous attractors are really applied in neural systems, we also investigate two general properties associated uniquely with the dynamics of continuous attractors: the tracking speed and the correlation structure between neural response variabilities. It turns out that (1) the reaction time for a continuous attractor catching upto abrupt stimulus change increases logarithmically with the size of the stimulus change, and (2) the neural correlation is asymmetrical with respect to the stimulus position. The latter finding agrees with those in Ben-Yishai et al. (1995) and Miller (2006).

After obtaining the asymmetrical neural correlation, we further study how this correlation structure affects statistical population decoding. In experiments, we use a statistical inference method to reconstruct the stimulus based on the recorded neural data without referring to the underlying network dynamics. We particularly investigate an unfaithful decoding strategy that ignores the neural correlation. We find that this method works efficiently because the contributions of the positive and negative correlation on the decoding error cancel each other.

The organization of the letter is as follows. In section 2, we introduce a simple firing-rate-based model for a continuous attractor and analyze its stationary states. In section 3, the dynamics of a continuous attractor under the driving of noisy inputs is studied, and its tracking speed is quantified. In section 4, the decoding performance of a continuous attractor and the correlation structure of neural activities are investigated. In section 5,

[1] Consider that each neuron has an independent input component with respect to others; then the dimensionality of the network dynamics is equal to the number of neurons.

simulation experiments on both firing-rate- and spiking-based models are carried out to confirm the theoretical analysis. In section 6, we calculate the performance of statistical population decoding when neural activities are asymmetrically correlated. Finally, in section 7, we offer discussion and conclusions about this work.

2 The Model

Although diverse models exist for a continuous attractor in the literature, they all share two common features: (1) the network should have properly balanced excitatory and inhibitory interactions, so that it can hold persistent activities after external inputs are removed, and (2) the neuronal interactions should be translationally invariant, so that the network can have a continuous family of stationary states. We start by considering a simple firing-rate-based model for a continuous attractor. The advantage of this model is that it allows us to compute the network dynamics analytically, and its main conclusions can be extended to general cases, since they depend on only the common features of continuous attractor.

Consider a one-dimensional continuous stimulus x encoded by an ensemble of neurons. The neuronal preferred stimulus is denoted as c. We assume $c \in (-\infty, \infty)$ for convenience. The neurons are clustered according to their preferred stimuli, mimicking a column structure. The clusters are uniformly distributed in the parameter space c with density ρ. We denote γ_c to be the firing rate of the cluster c and U_c the population-averaged input (i.e., the synaptic drive as in Ermentrout, 1998, and Gutkin et al., 2003). The interaction between the two clusters c and c' is written as $J_{c,c'}$.

The dynamics of the network, in the unit of a cluster, is given by

$$\tau \frac{dU_c}{dt} = -U_c + \rho \int_{c'} J_{c,c'} \gamma_{c'} dc' + I_c^{ext}, \tag{2.1}$$

$$\gamma_c = \frac{U_c^2}{1 + k\rho \int_{c'} U_{c'}^2 dc'}, \tag{2.2}$$

where k is a small, positive constant and I_c^{ext} the external input. The parameter τ is the time constant for the population dynamics, which is on the order of 1 ms (see the justification in appendix C or the relevant references: e.g., Ermentrout, 1998, and Gutkin et al., 2003).

The recurrent interaction is set to be

$$J_{c,c'} = \frac{J}{\sqrt{2\pi}a} e^{-(c-c')^2/2a^2}, \tag{2.3}$$

where J is a constant that controls the magnitude of the recurrent interactions. $J_{c,c'}$ is the decay function of the difference between the preferred stimuli of the clusters, $(c - c')$. Here, $J_{c,c'}$ has only the excitatory components. The contribution of inhibition is achieved indirectly through the divisive normalization in equation 2.2.

When $I_c^{ext} = 0$, the stationary states of the network, referred to as $\bar{U}_c$ and $\bar{\gamma}_c$, satisfy the following conditions:

$$\bar{U}_c = \rho \int_{c'} J_{c,c'} \bar{\gamma}_{c'} dc', \tag{2.4}$$

$$\bar{\gamma}_c = \frac{\bar{U}_c^2}{1 + k\rho \int_{c'} \bar{U}_{c'}^2 dc'}. \tag{2.5}$$

It is straightforward to check that the network holds a continuous family of nontrivial stationary states (Amari, 1977; Wu et al., 2002),

$$\bar{U}_c(z) = \frac{A\rho J}{\sqrt{2}} e^{-(c-z)^2/4a^2}, \tag{2.6}$$

$$\bar{\gamma}_c(z) = Ae^{-(c-z)^2/2a^2}, \tag{2.7}$$

where $A = (1 + \sqrt{1 - 8\sqrt{2\pi}ak/(J^2\rho)})/(2\sqrt{2\pi}ak\rho)$, and $z \in (-\infty, \infty)$ is a free parameter.[2] These states are of a gaussian bell shape and can be retained after removing external inputs if $0 < k < J^2\rho/(8\sqrt{2\pi}a)$ (note that k controls the amount of inhibition). The parameter z is the peak position of the bump, which indicates the network representation and decoding of the external stimulus.

The stimulus information is conveyed to the neural system through the external input I_c^{ext}. In a conventional study, it is often assumed that I_c^{ext} is either a large transient or a small, constant input, which drives the network to be stable at the position determined by Deneve et al. (1999) and Wu et al. (2002):

$$\hat{x} = \max_z \int_c \bar{\gamma}_c(z) I_c^{ext} dc. \tag{2.8}$$

In this study, we consider a more general case when I_c^{ext} is time varying and contains gaussian white noise. Without loss of generality, we choose

[2] The network has two sets of stationary states: one at $A = (1 + \sqrt{1 - 8\sqrt{2\pi}ak/(J^2\rho)})/(2\sqrt{2\pi}ak\rho)$ and the other at $A = 0$ (the silent states). The network will be sustained at the active states if the initial value $A > (1 - \sqrt{1 - 8\sqrt{2\pi}ak/(J^2\rho)})/(2\sqrt{2\pi}ak\rho)$.

I_c^{ext} to be of the following form (for a more general choice of I_c^{ext}, see appendix A),

$$I_c^{ext} = \alpha \bar{U}_c(x) + \sigma \xi_c(t), \tag{2.9}$$

where both α and σ are small, positive constants and $\xi_c(t)$ is gaussian white noise with zero mean and unit variance. The first term, $\alpha \bar{U}_c(x)$, represents the stimulus signal, whose contribution is to drive the system to the location of the stimulus x. The second term, $\sigma \xi_c(t)$, represents the input noise with σ the noise strength. For simplicity, we assume ξ_c and $\xi_{c'}$, for $c \neq c'$, are independent to each other.

3 Dynamics of Continuous Attractor

In general it is difficult to solve the dynamics of a large, fully connected network. Here, by utilizing the specific features of a continuous attractor, we develop a strategy to assess its dynamics approximately.

To proceed, let us first check how neutral stability shapes the dynamics of a continuous attractor. Consider the network state to be initially at a position z. An input noise induces small fluctuations on the network state and the stationary inputs, which are denoted as $\delta\gamma_c(z)$ and $\delta U_c(z)$ for the cluster c, respectively. Then, according to the stability conditions in equations 2.4 and 2.5, we have

$$\begin{aligned}
\delta\gamma_c(z) &= \int_{c'} \frac{\partial \bar{\gamma}_c(z)}{\partial \bar{U}_{c'}(z)} \delta U_{c'}(z) dc', \\
&= \int_{c',c''} \frac{\partial \bar{\gamma}_c(z)}{\partial \bar{U}_{c'}(z)} \rho J_{c',c''} \delta\gamma_{c''}(z) dc' dc'', \\
&= \int_{c'} F_{c,c'}(z) \delta\gamma_{c'}(z) dc',
\end{aligned} \tag{3.1}$$

where the matrix $\mathbf{F}(z)$ is calculated to be

$$\begin{aligned}
F_{c,c'}(z) &= \rho \int_{c''} \frac{\partial \bar{\gamma}_c(z)}{\partial \bar{U}_{c''}(z)} J_{c',c''} dc'' \\
&= \frac{A\rho^2 J^2}{B\sqrt{\pi} a} e^{-(c-z)^2/4a^2} e^{-(c-c')^2/2a^2} \\
&\quad - \frac{k A^3 \rho^5 J^4}{\sqrt{3} B^2} e^{-(c-z)^2/2a^2} e^{-(c'-z)^2/6a^2},
\end{aligned} \tag{3.2}$$

with $B = 1 + A^2 J^2 \sqrt{2\pi} a k \rho^3 / 2$.

Neutral stability implies that if the change of the network state is along the attractor space (i.e., only the peak position is moved, whereas the bump shape is unchanged), then the network is stable at the new position; otherwise, the system will return to its original shape. Intuitively stated, a continuous attractor will clean only those noise components that are perpendicular to the attractor space. Mathematically, this means that the matrix $\mathbf{F}(z)$ has one eigenvector whose eigenvalue is one and all other eigenvalues are smaller than one. The eigenvector, belonging to unit eigenvalue, referred to as $\mathbf{e}^r(z)$, is along the tangent of the attractor space and is dependent on the position z, whose component is given by

$$\begin{aligned} e^r_c(z) &\sim \bar{\gamma}'_c(z), \\ &= D_r(c-z)e^{-(c-z)^2/2a^2}, \end{aligned} \tag{3.3}$$

where D_r is a constant (the exact value of D_r is not important here). It is straightforward to check that $\mathbf{e}^r$ is indeed the eigenvector of $\mathbf{F}$ with eigenvalue unit (i.e., $\int_c F_{c,c'}e^r_{c'}dc' = e^r_c$).

The vector $\mathbf{e}^r(z)$ specifies the direction in the state space along which the network state is neutrally stable. In the input space I_c, the corresponding direction, referred to as $\mathbf{e}^I(z)$, is given by

$$\begin{aligned} e^I_c(z) &\sim \bar{U}'_c(z), \\ &= D_I(c-z)e^{-(c-z)^2/4a^2}, \end{aligned} \tag{3.4}$$

where D_I is a constant. Similarly, it can be checked that $\mathbf{e}^I$ is the eigenvector of the matrix, $G_{c,c'} = \int_{c''} J_{c,c''}(\partial\bar{\gamma}_{c'}/\partial\bar{U}_{c''})dc''$, with eigenvalue unit (see appendix B).

We note that the neural population dynamics, in the unit of a cluster, is extremely fast, which is on the order of τ ($1 \sim 2$ ms), much smaller than the membrane time constant of single neurons ($10 \sim 20$ ms) (see the justification in appendix C). Combining this with the special stability of a continuous attractor, it means that the network dynamics can clean those noise components perpendicular to $\mathbf{e}^I(z)$ very quickly. Thus, if the time window for the neural system to read out the stimulus is sufficiently long (e.g., much larger than τ), we can reasonably assume the network dynamics is mainly driven by the projection of external inputs on the direction $\mathbf{e}^I(z)$ and ignore the contribution of other components. This implies that the network bump has only its position shifted, whereas its shape is unchanged. By this approximation, we reduce the dimensionality of the network dynamics from the original value of infinity (since $c \in (-\infty, \infty)$) to unity.

Without loss of generality, we consider in what follows that the true stimulus is at $x = 0$, and under the perturbation of external noise, the peak position of the bump deviates from the stimulus position only weakly.

Assuming the peak position is at z at time t, we project both sides of equation 2.1 on the direction $\mathbf{e}^I(z)$, and obtain

$$\begin{aligned}\text{Left-hand side} &= \tau \int_c \frac{dU_c}{dt} e_c^I(z) dc, \\ &= \left[\frac{\tau AJ\rho}{2\sqrt{2}a^2} \int_c (c-z) e^{-(c-z)^2/4a^2} e_c^I dc \right] \frac{dz}{dt}, \\ &= \left[\frac{\tau AJ\rho\sqrt{\pi} D_I a}{2} \right] \frac{dz}{dt}, \end{aligned} \tag{3.5}$$

and

$$\begin{aligned}\text{Right-hand side} &= -\int_c \left(U_c - \sum_{c'} J_{c,c'} \gamma_{c'} \right) e_c^I(z) dc + \alpha \int_c \bar{U}_c(0) e_c^I(z) dc \\ &\quad + \sigma \int_c \xi_c(t) e_c^I(z) dc, \\ &= -\frac{AJa\rho\sqrt{\pi} D_I}{2} \alpha z + \sigma (2\pi)^{1/4} a^{3/2} D_I \epsilon(t). \end{aligned} \tag{3.6}$$

To obtain the above results, we have used two approximations: (1) the distortion of the bump from its stationary state is small, that is, $U_c \approx \bar{U}_c(z)$, $\gamma_c \approx \bar{\gamma}_c(z)$, and (2) for $|z| \ll a$, by the first-order Taylor expansion, $\bar{U}_c(0) \approx \bar{U}_c(z) - zAJ\rho/(2\sqrt{2}a^2 D_I) e_c^I(z)$. The second term in equation 3.6 is the projection of the input noise on $\mathbf{e}^I(z)$, where $\epsilon(t)$ is the gaussian white noise of zero mean and unit variance.

Combining the above results, we get,

$$\tau \frac{dz}{dt} = -\alpha z + \beta \epsilon(t), \tag{3.7}$$

where β is a positive number and β^2 is given by

$$\begin{aligned}\beta^2 &= \frac{\sigma^2}{\int_c [\bar{U}_c'(z)]^2 dc}, \\ &= \frac{4\sqrt{2}a\sigma^2}{A^2 J^2 \rho^2 \sqrt{\pi}}. \end{aligned} \tag{3.8}$$

Equation 3.7 is the one-dimensional OU process for the peak position. The meaning of this equation is straightforward: whenever the bump position deviates from the true stimulus, the stimulus signal generates a force, $-\alpha z$, that pulls the bump back to the stimulus position ($z = 0$). The noise effect, $\beta\epsilon(t)$, tends to shift the bump position randomly.

To solve equation 3.7, we define a new stochastic variable $y(t) = e^{\alpha t/\tau} z(t)$, with $y(0) = z(0)$ and $z(0)$ being the initial bump position. It is easy to check that

$$\frac{dy}{dt} = \frac{\beta}{\tau} e^{\alpha t/\tau} \epsilon. \tag{3.9}$$

Integrating this equation gives

$$y(t) = y(0) + \frac{\beta}{\tau} \int_0^t e^{\alpha t'/\tau} \epsilon(t') dt'. \tag{3.10}$$

Thus, after averaging over many trials, we get

$$\langle y(t) \rangle = y(0), \tag{3.11}$$

$$\langle y(t)^2 \rangle = y(0)^2 + \frac{\beta^2}{2\alpha} (e^{2\alpha t/\tau} - 1). \tag{3.12}$$

Finally by using the relationship $z(t) = e^{-\alpha t/\tau} y(t)$, we obtain

$$\langle z(t) \rangle = z(0) e^{-\alpha t/\tau}, \tag{3.13}$$

$$\langle z(t)^2 \rangle = z(0)^2 e^{-2\alpha t/\tau} + \frac{\beta^2}{2\alpha} (1 - e^{-2\alpha t/\tau}). \tag{3.14}$$

3.1 The Tracking Speed. With the simplified dynamical model, equation 3.7, we can calculate the tracking speed of a continuous attractor. We consider a scenario that the stimulus value is abruptly changed from the initial value $z(0)$ to 0. We measure the reaction time for the system to catch upto this change.

Note that when z is approaching zero, the driving force $-\alpha z$ becomes smaller and smaller, which implies that when there is no noise, it takes $t \to \infty$ for the bump to reach the stimulus position. This, however, should not be a problem in practice, since the neural system does not have to wait for the bump to be exactly at $z = 0$ in order to make a judgment. Without loss of generality, we assume that after the distance $|z|$ is below a threshold θ, a predefined small, positive number, the tracking is finished.

Because of noise, the reaction time of the network is given by the first passage time for its bump position across the threshold θ. Following the standard procedure for solving the OU process (Tuckwell, 1988), we get the mean of the reaction time T to be

$$\langle T \rangle = \frac{\tau}{\alpha}\sqrt{\pi}\int_{d_1}^{d_2} e^{u^2}[1 + erf(u)]du, \tag{3.15}$$

where $d_1 = -z(0)\sqrt{\alpha\tau}/\beta$ and $d_2 = -\theta\sqrt{\alpha\tau}/\beta$.

To see this relationship more clearly, we consider that the noise is sufficiently small and can be ignored. Then equation 3.15 becomes

$$\langle T \rangle = \frac{\tau}{\alpha}\ln\frac{|z(0)|}{\theta}. \tag{3.16}$$

This equation reveals that the reaction time of a continuous attractor increases logarithmically with the stimulus change (here, the size of change is equal to $|z(0)|$). This logarithm relationship is intuitively understandable. It comes from the fact that the driving force for the bump movement is proportional to the discrepancy between the current bump position and the stimulus value (see the first term on the right-hand side of equation 3.7): the bump tends to move faster when it is far away from the stimulus and slower when it is close. We further confirm this result by simulation in section 5.1.

Since the property of smooth tracking is uniquely associated with the neutral stability of the dynamics of continuous attractors, we expect that this logarithm reaction time can provide a clue for checking experimently whether continuous attractors are actually applied in neural systems.

4 Computation of Continuous Attractor

We now assess the performance of a continuous attractor as a general model for information retrieval. We consider that a neural estimator reads out the stimulus based on the peak position of the bump.[3] Two ways of collecting data are distinguished: to infer the stimulus based on: the instant position of the bump or the accumulated cluster activities over a period.

In the first case, the decoding result is given by $z(t)$. From equations 3.13 and 3.14, we observe the following properties:

- The mean of $z(t)$ is determined by the initial position of the bump, which decays exponentially with time. When $t \to \infty$, $\langle z(t)\rangle = 0$, implying that at long times, the decoding accuracy of continuous attractor is unbiased.

[3] Note that due to neutral stability of the network dynamics, it is the discrepancy of the bump peak from the stimulus position that dominates the encoding error rather than the distortion of the bump. We may use some more complicated population decoding strategies, such as center of mass, template matching, or maximum likelihood inference, to read out the stimulus, but the result will not be much different.

- The mean square error,[4] $\langle z(t)^2 \rangle$, measures the discrepancy between the true stimulus and the estimation at time t and has two parts. The first is due to the initial position of the bump, which decays exponentially with time. The second is the noise contribution, which increases with time initially and saturates to a constant $\beta^2/(2\alpha)$ when $t \to \infty$. The saturating constant is determined by the ratio between the noise (β^2) and the signal (α) strengths.
- When $\alpha = 0$, that is, no stimulus signal, $\langle z(t)^2 \rangle = z(0)^2 + \beta^2 t/\tau$, indicating that the network state fluctuates on the attractor space like a random walk.
- When the height of the bump is fixed (i.e., A, and also ak, is a constant), the decoding error increases with the parameter a (note $\beta^2 \sim a$; see equation 2.7). This is understandable. The size of a determines the range of excitatory interaction (see equation 2.3) and the tuning width of neurons (see equation 2.7). The larger the value of a is, the quicker the bump can be moved under the driving of external noise, indicating "larger" neutral stability.[5]
- The decoding error decreases with signal strength α. This is intuitively correct.

In the second case, more effort is required to compute the decoding error. Let us denote the cluster activity observed by the estimator over a time window T to be $r(T)$, which is calculated to be

$$r_c(T) = \frac{1}{T} \int_0^T \gamma_c(t) dt. \tag{4.1}$$

By using the approximation that the change of the network activity is dominated by the bump position shift, we have

$$\begin{aligned} r_c(T) &\approx \frac{1}{T} \int_0^T \bar{\gamma}_c(z(t)) dt, \\ &\approx \bar{\gamma}_c(0) + \left[\frac{1}{T} \int_0^T z(t) dt \right] \bar{\gamma}_c'(0), \\ &= \bar{\gamma}_c(0) + h(T) \bar{\gamma}_c'(0), \\ &\approx \bar{\gamma}_c(h(T)), \end{aligned} \tag{4.2}$$

where $h(T) = (\int_0^T z(t) dt)/T$. In the above, we used the condition $|z(t)| \ll a$.

[4] Note that here we do not use variance, $\langle (z - \langle z \rangle)^2 \rangle$, to measure the decoding error. This is because the initial pump position is unknown to the estimator, and it can have a significant contribution on the decoding error.

[5] Intuitively, the bump movement is conducted through neuronal excitatory interactions. Larger a implies the bump can be moved more quickly (Wu & Amari, 2005).

The task of a neural estimator is to infer the stimulus value based on $\{r_c(T)\}$ for all c values. Since $\{r_c(T)\}$ can be approximated as a smooth bump centered at $h(T)$, the decoding result there is $h(T)$.

The averaged values of $h(T)$ and $h(T)^2$ are calculated to be

$$\langle h(T)\rangle = \frac{z(0)\tau}{\alpha T}(1 - e^{-\alpha T/\tau}), \tag{4.3}$$

$$\begin{aligned}\langle h(T)^2\rangle &= \left\langle \left[\frac{1}{T}\int_0^T z(t)dt\right]^2\right\rangle,\\ &= \frac{(z(0)\tau)^2}{(\alpha T)^2}(1 - e^{-\alpha T/\tau})^2 + \frac{\beta^2\tau}{2\alpha^2 T^2}\left(2T - \frac{3\tau}{\alpha} + \frac{4\tau}{\alpha}e^{-\alpha T/\tau} - \frac{\tau}{\alpha}e^{-2\alpha T/\tau}\right),\\ &= \langle h(T)\rangle^2 + \mathrm{Var}(h(T))\end{aligned} \tag{4.4}$$

where the average is over many trials and $\mathrm{Var}(h)$ denotes the variance of $h(T)$.

From equations 4.3 and 4.4, we observe the following properties:

- The mean of $h(T)$ is determined by the initial position of the bump. When $T \to \infty$, $\langle h(T)\rangle = 0$.
- When T is sufficiently large, $\langle h(T)^2\rangle \sim \beta^2\tau/(\alpha^2 T)$. This means that the decoding error, based on the accumulated cluster activities, decreases with time, which is different from the result based on the instant position of the bump. This is understandable, since although noise can shift the bump position randomly, by integrating over time, these fluctuations are averaged out due to the memoryless nature of gaussian white noise.
- When $\alpha = 0$, $\langle h(T)^2\rangle = z(0)^2 + \beta^2 T/(3\tau)$, reflecting the consequence of random walk.

It is worth noting that although in the second case the decoding accuracy of the continuous attractor is improved, it is at the expense of delaying the tracking speed of the network.

4.1 The Correlation Structure.

The correlation between neuronal activities is an important quantity to measure in experiment to infer the mechanism of brain functions. Here we investigate how the specific dynamics of continuous attractor shapes the neural correlation form.

Consider that neural activities are recorded over a time window T; the cross-correlation between cluster activities is defined as

$$R_{c,c'}(T) = \langle [r_c(T) - \langle r_c(T)\rangle][r_{c'}(T) - \langle r_{c'}(T)\rangle]\rangle. \tag{4.5}$$

By using the relationship in equation 4.2, we obtain

$$\begin{aligned} R_{c,c'}(T) &= \langle [h(T) - \langle h(T)\rangle]^2 \rangle \bar{\gamma}_c'(0)\bar{\gamma}_{c'}'(0) \\ &= \frac{\text{Var}(h) A^2 cc'}{a^4} e^{-c^2/2a^2} e^{-(c')^2/2a^2}. \end{aligned} \tag{4.6}$$

This correlation structure exhibits an interesting feature: $R_{c,c'}$ is asymmetric with respect to the stimulus x: when $x = 0$, $R_{c,c'} = -R_{c,-c'}$. It is straightforward to check that when $x \neq 0$, $R_{c,c'-x} = -R_{c,x-c'}$. The clusters' correlation is positive if their preferred stimuli are on the same side of the stimulus: $(c - x)(c' - x) > 0$; otherwise it is negative. Mathematically, this structure comes from that the bump $\bar{\gamma}_c(x)$ is symmetrical with respect to x, whereas $\bar{\gamma}_c'$ is asymmetrical (see equation 2.7). This is due to the fact that the bump fluctuations are dominated by the bump's position shift as a consequence of the neutral stability of the network dynamics. Thus, this property is associated with the specific dynamics of continuous attractors. It can serve as an important clue for experimentally checking the application of continuous attractors in neural systems.

5 Simulation Experiments

In this section, we give results of the simulation to further confirm our theoretical analysis. Two types of model are considered: one based on the firing rate model and the other on spiking neurons.

5.1 A Firing-Rate-Based Model. In this case, we do not consider the responses of individual neurons, but rather focus on the the population-averaged firing rates. The dynamics of the network is given by equations 2.1 and 2.2. We consider there are $N = 101$ clusters, whose preferred stimuli are uniformly distributed in the range $(-\pi, \pi]$. The clusters' interactions are periodic, that is, $J_{c,c'} = J_{c,c''}$ if $|c - c'| = 2\pi - |c - c''|$, where $|\cdot|$ denotes the absolute value. In this case, the stationary states of the network will not have the exact form as given by equations 2.4 and 2.5, but they are still bell shaped. Also, since now the number of clusters is finite, the integration in all the above calculations is replaced by the corresponding summation, $\rho \int_c = 2\pi/N \sum_i$. The parameters are set to be $x = 0$, $\tau = 1$ ms, $k = 10$, $J = 50$, $a = 0.5$, $\alpha = 0.1$, and $\sigma^2 = 0.1$. All simulation results are obtained by averaging over 100 trials.

Figures 2A and 2B illustrate examples of the stationary states and the synaptic drive of the network when there is no external input. They can be well fitted as gaussian functions: $\bar{\gamma}_c = 0.08e^{-(c-x)^2/0.5}$ and $\bar{U}_c = 2.75e^{-(c-x)^2/1.125}$. Figure 3A first shows, when no stimulus signal exists, how the mean squared discrepancy of the bump position with respect to its initial location varies with time under the driving of gaussian white noise.

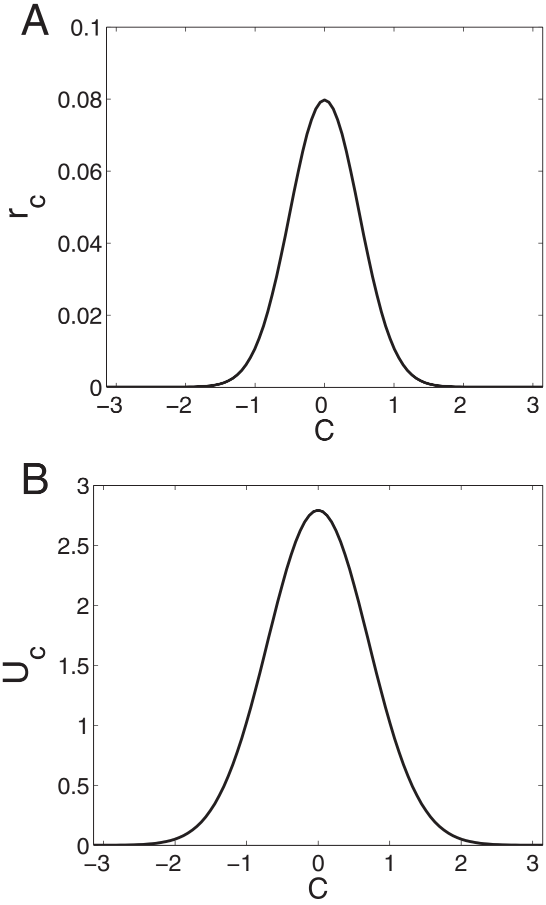

Figure 2: (A) An example of the stationary state of the network centered at $z = 0$. If approximated as a gaussian function, the width of the bump is 1. (B) An example of the stationary synaptic drive centered at $z = 0$. The width of the bump is 0.75.

We see that its value linearly increases with time, indicating the random walk behavior. We then add the stimulus signal $x = 0$. Figure 3B shows that the decoding error based on the instant position of the bump saturates to a constant value. Depending on the initial position of the bump, the decoding error may increase or decrease initially. Figure 3C shows that the decoding error based on the accumulated cluster activities decreases with time. If the initial bump position is very close to the true stimulus, the decoding error increases with time initially. All of these observations agree with our theoretical analysis in section 3.

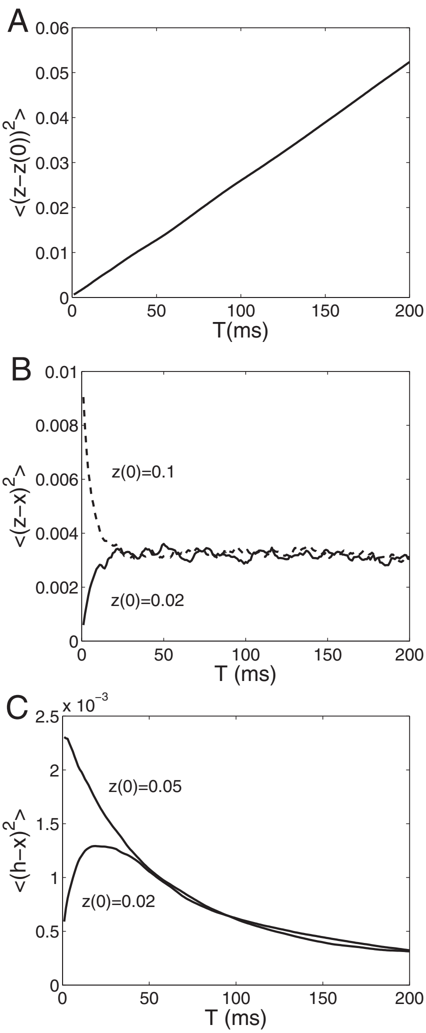

Figure 3: (A) When no stimulus signal exists, the bump position fluctuates randomly under the driving of external noise. Its mean squared distance to the initial location increases linearly with time, displaying the behavior of a random walk. (B) When the stimulus signal $x = 0$ is added, the decoding error based on the instant position of the bump saturates to a constant value. (C) When the stimulus signal is added, the decoding error based on the accumulated cluster activities decreases with time.

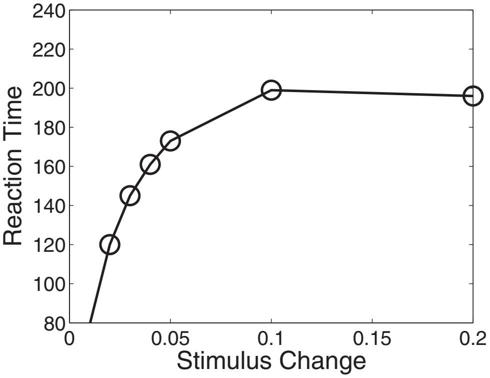

Figure 4: Reaction time versus the size of abrupt stimulus change.

Figure 4 shows the reaction time of the network for catching upto different sizes of abrupt stimulus change. We see that the reaction time increases logarithmically with the change size, supporting the theoretical analysis in section 3.1.

Figure 5 further illustrates the correlation structure between cluster activities. We see that they indeed have the asymmetric shape with respect to the stimulus position, agreeing well with our theoretical analysis.

5.2 A Spiking Neuron–Based Model. To further confirm that our theoretical analysis is applicable to general cases, we also carry out simulation on a spiking neuron network. We consider that there are $N = 64$ clusters, whose preferred stimuli are uniformly distributed in the range $(-\pi, \pi]$ with the periodic condition held. In each cluster, there are $M = 64$ neurons. The connections between neurons are random and sparse; the probability for two neurons being connected is $\rho_c = 0.1$. The connection strength between two clusters c and c' is $J_{c,c'}$. The dynamics of a single neuron is given by

$$\tau_m \frac{dv_c^i}{dt} = -\left(v_c^i - V_L\right) + I_c^i + I_c^{ext} + \sigma \xi_c^i(t), \tag{5.1}$$

$$\tau \frac{I_c^i}{dt} = -I_c^i + \sum_{c'} \sum_{j \to i} J_{c,c'} A_{c'}^j(t), \tag{5.2}$$

where v_c^i represents the membrane potentials of the ith neuron in the cluster c and τ_m the membrane time constant. The parameter τ is the synaptic current constant. We use $A_{c'}^j$ to denote the spike train generated by the jth neuron in the cluster c', which can be written as $A_{c'}^j = \sum_m \delta(t - t_m)$, with t_m the firing moment of the mth spike. The symbol $j \to i$ indicates the summation runs over all neurons connected to neuron i.

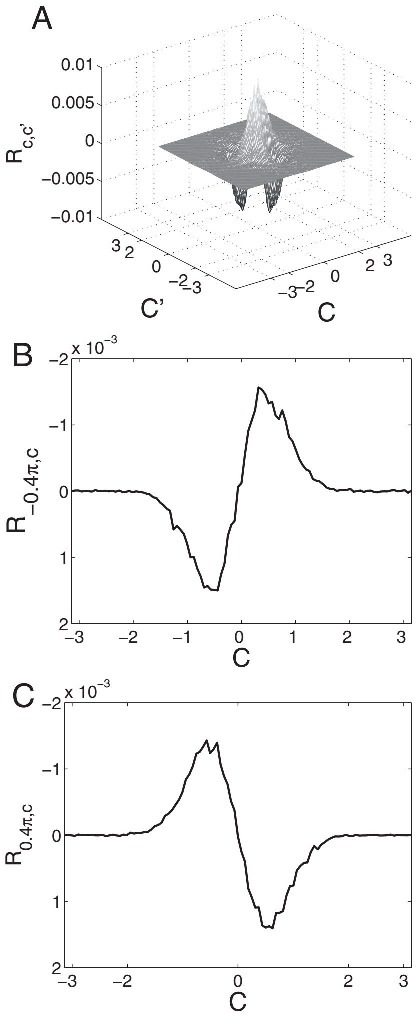

Figure 5: The correlation structures between cluster activities. The cluster activities are collected at $T = 20$ ms. (A) The correlation structure plotted in 3D. (B) The correlation strengths between the cluster $c = -0.4\pi$ and all others. (C) The correlation strengths between the cluster $c = 0.4\pi$ and all others.

We choose $J_{c,c'}$ to be of the Mexican hat type,

$$J_{c,c'} = \frac{J_0}{NM\rho} + \frac{J_1}{NM\rho}\cos(c - c'), \tag{5.3}$$

where J_0 and J_1 are constants.

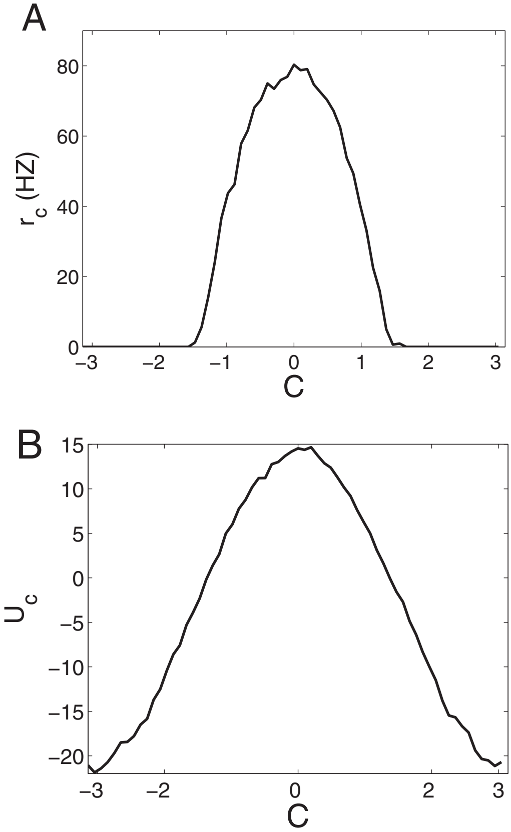

Figure 6: (A) An example of the stationary state of the network centered at $z = 0$. (B) An example of the stationary synaptic drive centered at $z = 0$. $U_c = 1/M \sum_i I_c^i$.

We further choose the external input that contains the information of the stimulus x to be

$$I_c^{ext} = I_0^{ext} + I_1^{ext} \cos(c - x). \tag{5.4}$$

A neuron fires when its potential exceeds the threshold $V_{th} = -55$ mV and subsequently resets to the resting potential $V_{reset} = -65$ mV. The refractory period is 5 ms. The other parameters used in the simulation are $\tau_m = 10$ ms, $\tau = 2$ ms, $J_0 = -149.5$, $J_1 = 897$, and $\sigma = 10$.

Figure 6 illustrates the examples of the stationary states and synaptic drive of the network when no external input exists. Both exhibit a bell shape. Figure 7 displays the computational properties of the network, which agree

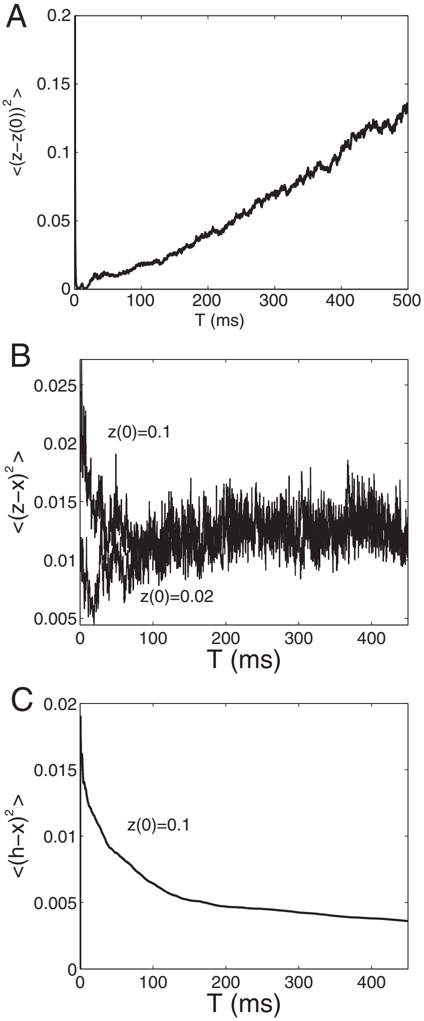

Figure 7: (A) When no stimulus signal exists, the bump position fluctuates randomly under the driving of external noise. Its mean squared distance to the initial location increases linearly with time, displaying the behavior of a random walk. (B) When the stimulus signal exists, the decoding error based on the instant position of the bump saturates to a constant value. Depending on the initial position of the bump, the decoding error may increase or decrease first. (C) When the stimulus signal exists, the decoding error based on the accumulated cluster activities decreases with time.

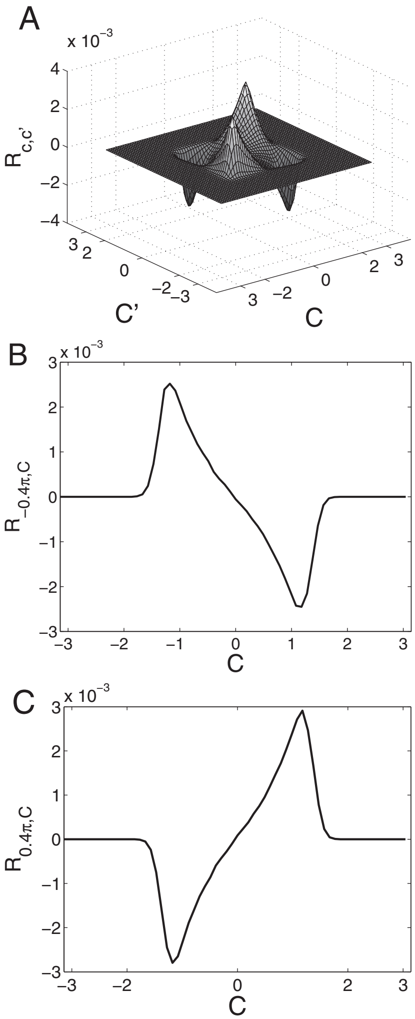

Figure 8: The correlation structures between cluster activities. (A) The correlation structure plotted in 3D. (B) The correlation strengths between the cluster $c = -0.4\pi$ and all others. (C) The correlation strengths between the cluster $c = 0.4\pi$ and all others.

well with our theoretical analysis and the simulation results based on the rate model shown in Figure 3. The correlation structure between clusters is shown in Figure 8, which is asymmetric with respect to the stimulus position as expected.

6 Effect of Asymmetrical Correlation on Population Decoding

We have observed that the neuronal correlations in a continuous attractor exhibit a specific feature, that is, they are asymmetrical with respect to the stimulus position. In this section, we investigate how this correlation structure affects the performance of statistical population decoding. In experiments, we use a statistical inference method to reconstruct the stimulus based on the recorded neural activities without referring to the underlying network dynamics.

The problem is formulated as follows. Consider a set of cluster activities, $\mathbf{r} = \{r_c\}$, for all c values. Their mean values averaged over many trials are given by $\bar{\mathbf{r}} = \{\bar{r}_c\}$. For simplicity, we ignore at first the discrepancy of the initial position of the bump from the stimulus value and assume $\bar{r}_c = \bar{r}_c(x)$, with x the stimulus value (if the initial discrepancy is considered, it should be $\bar{r}_c = \bar{r}_c(x + \langle h \rangle)$; see equation 4.3). According to the underlying dynamics of the continuous attractor, the correlation matrix can be written as $R_{c,c'} = \mathrm{Var}(h(T))\bar{\gamma}'_c(x)\bar{\gamma}'_{c'}(x) + \sigma_m^2\delta_{c,c'}$ (see equation 4.6, and consider that data are collected over a time window T). Here, for generality, we also include a term, $\sigma_m^2\delta_{c,c'}$, to represent the bump distortion and the measurement errors, which are assumed to be independent gaussian noises of zero mean and the variance σ_m^2. The probability of observing a particular set of cluster activities $\mathbf{r}$ is given by

$$Q(\mathbf{r} \mid x) = \frac{1}{Z} \exp\left[-\frac{\rho^2}{2} \int\int (r_c - \bar{\gamma}_c) R^{-1}_{c,c'} (r_{c'} - \bar{\gamma}_{c'}) dc dc' \right], \tag{6.1}$$

where $\mathbf{R}^{-1}$ is the inverse of $\mathbf{R}$.

We are particularly interested in using a simple unfaithful decoding strategy to estimate the stimulus.[6] This method ignores the neural correlation and calculates the stimulus x by

$$\begin{aligned} \hat{x} &= \mathrm{Max} \ln P(\mathbf{r} \mid x), \\ &= \mathrm{Min} \int (r_c - \bar{\gamma}_c)^2 dc, \\ &= \mathrm{Max} \int r_c \bar{r}_c dc, \end{aligned} \tag{6.2}$$

[6] There are two main reasons. First, unfaithful decoding is more feasible in practice. Faithful decoding needs full knowledge of the correlation structure, which itself depends on the stimulus value. This makes faithful decoding difficult to implement. Second, as shown later, a proper unfaithful decoding is already accurate enough.

where the unfaithful model

$$P(\mathbf{r} \mid x) = \frac{1}{Z} \exp\left[-\frac{\rho}{2} \int (r_c - \bar{\gamma}_c)^2 dc \right]. \tag{6.3}$$

From equation 6.2, we see that this decoding strategy is actually the template-matching method with $\bar{\mathbf{r}}$ the template.

Suppose the estimation $\hat{x}$ is close enough to x. We expand $\nabla \ln P(\mathbf{r} \mid \hat{x})$ at x,

$$\nabla \ln P(\mathbf{r} \mid \hat{x}) \approx \nabla \ln P(\mathbf{r} \mid x) + \nabla\nabla \ln P(\mathbf{r} \mid x)(\hat{x} - x). \tag{6.4}$$

Since $\nabla \ln P(\mathbf{r} \mid \hat{x}) = 0$, we have

$$\begin{aligned}(\hat{x} - x) &\approx -\frac{\nabla \ln P(\mathbf{r} \mid x)}{\nabla\nabla \ln P(\mathbf{r} \mid x)}, \\ &= -\frac{R}{S}.\end{aligned} \tag{6.5}$$

Here the variable R is calculated to be

$$\begin{aligned}R &= \nabla \ln P(\mathbf{r} \mid x), \\ &= \rho \int [r_c - \bar{\gamma}_c(x)]\bar{\gamma}_c'(x) dc,\end{aligned} \tag{6.6}$$

where $\bar{\gamma}_c'(x) = d\bar{\gamma}_c/dx$.

Its variance is given by

$$V[R] = \rho^2 \int \bar{\gamma}_c' R_{c,c'} \bar{\gamma}_{c'}' dc dc'. \tag{6.7}$$

The random variable S can be divided into two parts,

$$\begin{aligned}S &= \nabla\nabla \ln P(\mathbf{r} \mid x), \\ &= \rho \int [r_c - \bar{\gamma}_c(x)]\bar{\gamma}_c''(x) dc - \rho \int (\bar{\gamma}_c'(x))^2 dc, \\ &= S_1 + S_2,\end{aligned} \tag{6.8}$$

where S_2 is a constant and S_1 is a random number of zero mean, with the variance given by

$$V[S_1] = \rho \int \sigma_c^2 (\bar{\gamma}_c'')^2 dc + \rho^2 \int_{c \neq c'} \bar{\gamma}_c'' R_{c,c'} \bar{\gamma}_{c'}'' dc dc', \tag{6.9}$$

where $\sigma_c^2 = R_{c,c}$, the variance of the activity of the cluster c.

Since $\bar{\gamma}_c''(x)$ is symmetric with respect to x, whereas $R_{c,c'}$ is asymmetric, it is straightforward to check that the contribution of the cross-correlation in equation 6.9 vanishes. Therefore, $V[S_1]$ is on the order of ρ, and $S1/S2 \sim \rho^{1/2}$. This implies that S_1 can be neglected when the number of clusters is large, and the decoding error asymptotically satisfies a gaussian distribution with the variance given by

$$\begin{aligned}\langle(\hat{x}-x)^2\rangle &\approx \frac{V[R]}{(S_2)^2},\\ &= \frac{\rho^2\int \bar{\gamma}_c' R_{c,c'}\bar{\gamma}_{c'}' dc dc'}{\rho^2[\int(\bar{\gamma}_c'(x))^2 dc]^2},\\ &= \mathrm{Var}(h(T)) + \frac{\sigma_m^2}{\rho\int(\bar{\gamma}_c'(x))^2 dc}.\end{aligned} \tag{6.10}$$

It consists of two parts. The first part is equal to the decoding result based on the peak position of the bump (see equation 4.4), which reflects the error due to the neutral stability of the network dynamics and is independent of the neuronal density ρ. The second part represents the error due to the bump distortion and the measurement mistakes, which decreases with ρ. If we include the effect due to the discrepancy of the initial position, the real decoding error is given by $\langle(\hat{x}-x)^2\rangle = \mathrm{Var}(h) + (\langle h\rangle)^2 + \sigma_m^2/[\rho\int(\bar{\gamma}_c'(x))^2 dc]$, since in this case, $\bar{\gamma}_c = \bar{\gamma}_c(x+\langle h\rangle)$.

Remarks. Understanding the effect of correlation on population coding is important for us to elucidate theoretically the mechanism of neural information processing. Recently two aspects of this issue have been intensively studied: (1) whether correlation degrades the accuracy of population decoding (see, e.g., Abbott & Dayan, 1999; Sompolinsky, Yoon, Kang, & Shamir, 2001; Wu et al., 2002) and (2) whether the correlation information can be ignored in the decoding process (see, e.g., Wu, Nakahara, & Amari, 2001; Wu et al., 2002; Nirenberg & Latham, 2003; Averbeck, Latham, & Pouget, 2006; Amari & Nakahara, 2006). In one study (Wu et al., 2002; Wu, Amari, & Nakahara, 2004), Wu et al. found that a strong, positive correlation will make a simple decoding method, such as template matching or center of mass, inefficient.[7] In another work, Sompolinsky et al. (2001) have shown that negative correlation can increase the optimal population decoding accuracy (based on an analysis of Fisher information). Here we observe that

[7] A population decoding method is inefficient if its decoding error satisfies the Cathy, rather than the gaussian, distribution, and the variance of the decoding error averaged over many trials diverges. Mathematically, this is due to that the cross-correlation term in equation 6.9 is not small (Wu et al., 2002).

for the asymmetric correlation, the contributions of the negative and positive parts of the decoding error cancel each other (see equation 6.9), which makes a simple decoding method efficient. Because continuous attractors may be widely applied in neural systems, this finding can provide guidance on designing a proper population decoding method.

7 Discussion and Conclusion

This study investigates the dynamics of a continuous attractor when external inputs are noisy and evaluates its performance as a general model for information retrieval. In order to carry out the research, we develop a strategy to reduce the dimensionality of the network dynamics by utilizing the fact that a continuous attractor retains only the noise component along the attractor space. Therefore, we project the network dynamics onto the tangent of the attractor space and simplify it to be a one-dimensional OU process. Based on this simplification, the computational behaviors of a continuous attractor are clear. We observe that (1) if the network decoding is based on the instant position of the bump (i.e., fast decoding), then the decoding error saturates to a constant value determined by the ratio between the signal and the noise strengths; and (2) if the decoding is based on the accumulated neural activities (i.e., slow decoding), the error decreases with time. The latter, however, is achieved at the expense of delaying the track speed of a continuous attractor.

We also investigate two general properties associated with the unique dynamics of continuous attractors: the logarithm reaction time and the asymmetric neural correlation structure. We expect they can serve as important experimental clues for us to check whether continuous attractors are actually applied in neural systems.

At first, it may appear that the logarithm reaction time is in contradiction to the linear relationship as observed in many mental rotation experiments (see, e.g., Shepard & Metzler, 1971). However, as already pointed out in the literature (see, e.g., Shepard & Metzler, 1988; Koriat & Norman, 1989), the mechanisms underlying mental rotation can be rather complicated and may vary in different experimental settings. On the other hand, we find encouraging evidence in a special type of mental rotation: backward alignment (Koriat & Norman, 1989). In this experiment, the human subjects were instructed to judge whether the rotated letter or number is the one just presented. The data showed that the reaction time of the subjects tends to increases logarithmically with the rotation angle (since this experiment was not designed for checking this property, it is not clear yet how accurately the data fit the logarithm function). We will carry out psychophysical experiments to clarify this point.

The asymmetric neural correlation is another salient feature that indicates the specific dynamics of continuous attractors. A direct way to prove this property is to measure the neural correlation in a computational task where a continuous attractor is likely to be involved, such as orientation tuning and motion control (see, e.g., Montani, Kohn, Smith, & Schultz, 2007). Alternatively, we may use fMRI data to indirectly check this correlation form. The idea is that since the voxel activities measured in fMRI are associated with neural responses, the neural correlation structure may be embedded in the corresponding voxel activities. We will carry out research along these two lines to check our analysis.

Although our results are obtained by using an ideal mathematical model for continuous attractors, they are qualitively applicable to general cases. Consider, for instance, that in reality, the number of clusters and neurons is finite and the network structure is often heterogeneous; in such a situation, the attractor space of the network is no longer perfectly flat but will contain many local minimums (Zhang, 1996; Seung, 1996; Renart, Song, & Wang, 2003). Nevertheless, if the distortion of the attractor space is sufficiently small, then the direction along the attractor space still dominates the network dynamics; that is, we can still reasonably approximate the network dynamics as a one-dimensional OU process (Renart et al., 2003) and obtain the main results in this work. We will firmly prove this point in our future work.

We also analyze the effect of asymmetrical correlation on neural population coding and find that in this particular correlation structure, the contributions of the positive and negative correlation on the decoding error cancel each other, leading a simple decoding method such as template matching to be efficient.

Appendix A: On the Choice of I_c^{ext}

In this study we have modeled the external signal as an input form that drives the bump to be stable at the stimulus position. For convenience, we have chosen $I_c^{ext} = \bar{U}_c(x)$, when no noise exists. In principle, we can model the signal, referred to as $S_c(x)$, by any unimodal function centered at x and obtain the similar results. The only difference in calculation will be the second term in equation 3.6, where $\bar{U}_c(0)$ will be replaced by $S_c(0)$. Considering that z is sufficiently small, we have $S_c(0) \approx S_c(z) - zS_c'(z)$, with $S_c' = dS_c/dz$. By utilizing the fact that $S_c(z)$ is symmetric with respect to z, and hence $\int_c S_c(z)e_c^I(z)dc = 0$, we will reach the same dynamical equation as equation 3.7, except that the value of β is given by $\beta^2 = \sigma^2 \int_c [e_c^I(z)]^2 dc/[\int_c S_c'(z)e_c^I(z)dc]^2$.

Appendix B: The Neutral Direction in the Input Space

We can similarly calculate the direction in the input space, on which the projection of external inputs has a sustained effect on the stationary state of the network.

According to the stability conditions in equations 2.4 and 2.5, we have

$$\begin{aligned}\delta U_c(z) &= \int_{c'} \frac{\partial \bar{U}_c(z)}{\partial \bar{\gamma}_{c'}(z)} \delta\gamma_{c'}(z) dc', \\ &= \int_{c',c''} \rho J_{c,c'} \left[\frac{2\bar{U}_{c'}}{B} \delta(c'-c'') - \frac{2k\rho \bar{U}_{c'}^2 \bar{U}_{c''}}{B^2} \right] \delta U_{c''} dc' dc'', \\ &= \int_{c'} G_{c,c'}(z) \delta U_{c'}(z) dc', \end{aligned} \tag{B.1}$$

where the matrix $\mathbf{G}(z)$ is calculated to be

$$\begin{aligned}G_{c,c'}(z) &= \int_{c''} \rho J_{c,c''} \left[\frac{2\bar{U}_{c''}}{B} \delta(c'-c'') - \frac{2k\rho \bar{U}_{c''}^2 \bar{U}_{c'}}{B^2} \right] dc'' \\ &= \frac{AJ^2\rho^2}{B\sqrt{\pi}a} e^{-(c-c')^2/2a^2} e^{-(c'-z)^2/4a^2} \\ &\quad - \frac{kA^3\rho^5 J^4}{\sqrt{3}B^2} e^{-(c-z)^2/6a^2} e^{-(c'-z)^2/2a^2}. \end{aligned} \tag{B.2}$$

We check that $\mathbf{e}^I$ is indeed the eigenvector of $\mathbf{G}$ with the eigenvalue being one:

$$\begin{aligned}\int_{c'} G_{c,c'} e_{c'}^I dc' &= \int_{c'} \frac{AJ^2\rho^2 D_I}{B\sqrt{\pi}a} e^{-(c-c')^2/2a^2} e^{-(c'-z)^2/4a^2} (c'-z) e^{-(c'-z)^2/4a^2} dc', \\ &= \frac{AJ^2\rho^2 D_I}{B\sqrt{\pi}a} e^{-(c-z)^2/4a^2} \int_{c'} (c'-z) e^{-[(c'-z)-(c-z)/2]^2/a^2} dc', \\ &= \frac{AJ^2\rho^2 D_I}{B\sqrt{\pi}a} e^{-(c-z)^2/4a^2} \int_{l} [(c-z)/2 + l] e^{-l^2/a^2} dl, \\ &= \frac{AJ^2\rho^2 D_I}{2B} (c-z) e^{-(c-z)^2/4a^2}, \\ &= e_c^I. \end{aligned} \tag{B.3}$$

In the above we used the condition $AJ^2\rho^2/(2B) = 1$ due to the stability constraint of equation 2.5.

Appendix C: From Spiking- to Rate-Based Models

Here we show an example for linking a spiking-neuron-based attractor model to a firing-rate-based one. Much of this knowledge has been reported in the literature (see, e.g., Ermentrout, 1998; Gerstner, 2000; Gutkin et al., 2003). We include this part for completeness of the work and also for clarifying the meaning of several important parameters in our model.

C.1 The Dynamics of a Single Neuron. We consider that the dynamics of a single neuron is given by

$$\tau_m \frac{dv_c^i}{dt} = -v_c^i + \rho \int_{c'} I_{c,c'}^i dc' + I_c^{i,ext}, \tag{C.1}$$

where v_c^i represents the membrane potential of the ith neuron in the cluster c. The parameter τ_m is the membrane time constant, which is typically on the order of 10 to 20 ms. $I_{c,c'}^i$ is the recurrent input coming from the cluster c'. $I_c^{i,ext}$ is the external input.

We consider that the number of neurons in each cluster is N. Each neuron has, on average, M random connections to neurons in each of other clusters. The connection is sparse in the sense that the connection density $\rho_c = M/N \ll 1$. The connection strength between two connected neurons in the clusters c and c' is $J_{c,c'}$. Here, we do not distinguish excitatory and inhibitory neurons. Their effects are combined by using the Mexican hat form of $J_{c,c'}$.

We denote $A_{c'}^j(t)$ the discrete spike train generated by the jth neuron in the cluster c', which can be written as

$$A_{c'}^j(t) = \sum_m \delta\big(t - t_m^j\big), \tag{C.2}$$

where t_m^j is the firing moment of the mth spike of this neuron.

The postsynaptic current generated by a spike firing at time s is given by

$$\alpha(t-s) = \frac{1}{\tau} e^{-(t-s)/\tau} H(t-s), \tag{C.3}$$

where $H(t-s)$ is the Heaviside function, which equals one when $t > s$ and zero otherwise. The parameter τ is the synaptic current constant, whose value is typically on the order of 1 ms.

With the above notations, the recurrent input $I_{c',c}^i$ is written as

$$I_{c,c'}^i(t) = \sum_{j \to i} J_{c,c'} \int_{-\infty}^{t} \alpha(t-s) A_{c',j}(s) ds, \tag{C.4}$$

where the summation runs over those neurons in the cluster c' that have connections to the neuron i in the cluster c, illustrated by $j \to i$.

Thus, the total synaptic drive to neuron i in cluster c is given by

$$\begin{aligned} u_c^i(t) &= \rho \int_{c'} I_{c,c'}^i dc' + I_c^{i,ext}(t), \\ &= \rho \int_{c'} J_{c,c'} \sum_{j \to i} \int_{-\infty}^{t} \alpha(t-s) A_{c',j}(s) ds dc' + I_c^{i,ext}. \end{aligned} \tag{C.5}$$

C.2 The Dynamics of a Cluster. To maintain persistent activity in a spiking neuron network, it is important that neurons fire irregularly. Here we consider that this condition already holds and define the firing rate of a cluster to be

$$r_c(t) = \lim_{\Delta t \to 0} \frac{n_c(\Delta t)}{\Delta t}, \tag{C.6}$$

where $n_c(\Delta t)$ is the number of spikes generated in cluster c in a time window Δt.

Furthermore, we define the averaged synaptic drive to a cluster as

$$U_c(t) = \frac{1}{N} \sum_i u_c^i(t). \tag{C.7}$$

We take into account two properties of the network dynamics: (1) that since neurons are randomly connected, the number of spikes received by the neuron i from the cluster c' is approximately given by $\sum_{j \to i} A_{c',j}(t) \sim \rho_c r_{c'}(t)$; and (2) that because of sparse connectivity, the inputs to two neurons in the same cluster can be regarded as being largely independent of each other. According to the law of large numbers, $1/N \sum_i \sum_{j \to i} A_{c',j}(t) \approx \rho_c r_{c'}(t) + O(1/\sqrt{N})$. With the two properties, we obtain

$$U_c(t) = \rho \rho_c \int_{c'} J_{c,c'} \int_{-\infty}^{t} \alpha(t-s) \gamma_{c'}(s) ds dc' + I_c^{ext}, \tag{C.8}$$

where $I_c^{ext} = (\sum_i I_c^{i,ext})/N$, representing the common input and the common noise to the cluster c (independent noise components are averaged out due to large N).

Finally, by differentiating $U_c(t)$ with respect to t, we obtain

$$\tau \frac{dU_c}{dt} = -U_c + \rho \rho_c \int_{c'} J_{c,c'} \gamma_{c'} dc' + I_c^{ext}. \tag{C.9}$$

Without loss of generality, we can absorb the connection density ρ_c into $J_{c,c'}$. Then the above equation returns to equation 2.1 in our model. Note here that the time constant for the cluster dynamics is τ rather than τ_m, indicating that the population dynamics is much faster than that of single neurons.

In principle, we should be able to solve the cluster dynamics with the knowledge of a single neuron's dynamics and the profile of neuronal connections. But in practice, this is difficult. In order to illustrate network properties, we often start by assuming that the relationship between the synaptic drive and the cluster activity is known, for example,

$$r_c = g_c(U_c), \tag{C.10}$$

where g_c is a properly defined gain function, and confirm the obtained result by simulation. In our model, we assume the gain function is given by equation 2.2.

Acknowledgments

We acknowledge the Royal Society for supporting K. H. while visiting the U.K. when this work was conducted. We acknowledge David Waxman for correcting the writing of the paper. S.W. is supported by the BBSRC grant (BB/E017436/1). We are grateful for the valuable comments of the reviewers.

References

Abbott, L., & Dayan, P. (1999). The effect of correlated variability on the accuracy of a population code. *Neural Computation, 11,* 91–101.

Amari, S. (1977). Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological Cybernetics, 27,* 77–87.

Amari, S., & Nakahara, H. (2006). Correlation and independence in the neural code. *Neural Computation, 18,* 1259–1267.

Averbeck, B., Latham, P., & Pouget, A. (2006). Neural correlations, population coding and computation. *Nature Reviews Neurosci., 7,* 358–366.

Ben-Yishai, R., Lev Bar-Or, R., & Sompolinsky, H. (1995). Theory of orientation tuning in visual cortex. *Proc. Natl. Acad. Sci. USA, 92,* 3844–3848.

Brody, C. D., Romo, R., & Kepecs, A. (2003). Basic mechanisms for graded persistent activity: Discrete attractors, continuous attractors, and dynamic representations. *Current Opinion in Neurobiology, 13,* 204–211.

Chow, C., & Coombes, S. (2006). Existence and wandering of bumps in a spiking neural network model. *SIAM Journal of Applied Dynamical Systems, 5,* 552–574.

Deneve, S., Latham, P. E., & Pouget, A. (1999). Reading population codes: A neural implementation of ideal observers. *Nature Neuroscience, 2,* 740–745.

Ermentrout, B. (1998). Neural networks as spatial-temporal pattern-forming systems. *Reports on Progress in Physics, 61,* 353–430.

Funahashi, S., Bruce, C., & Goldman-Rakic, P. (1989). Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cotex. *J. Neurophysiology, 61,* 331–349.

Georgopoulos, A. P., Kalaska, J. F., Caminiti, R., & Massey, J. T. (1982). On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex. *J. Neurosci., 2,* 1527–1537.

Gerstner, W. (2000). Population dynamics of spiking neurons: Fast transients, asynchronous states, and locking. *Neural Computation, 12,* 43–89.

Gutkin, B. S., Pinto, D., & Ermentrout, B. (2003). Mathematical neuroscience: From neurons to circuits to system. *J. Physiol. Paris, 97,* 209–219.

Hansel, D., & Sompolinsky, H. (1998). Modelling feature selectivity in local cortical circuits. In C. Koch & I. Segev (Eds.), *Methods in neuronal modelling: From synapses to networks*. Cambridge, MA: MIT Press.

Hopfield, J. J. (1984). Neurons with graded responses have collective computational properties like those of two-state neurons. *Proc. Natl. Acad. Sci. USA, 81,* 3088–3092.

Koriat, A., & Norman, J. (1989). Establishing global and local correspondence between successive stimuli: The holistic nature of backward alignment. *Journal of Experimental Psychology: Human Perception, Memory and Cognition, 15,* 480–494.

Maunsell, J. H. R., & Van Essen, D. C. (1983). Functional properties of neurons in middle temporal visual area of the macaque monkey. I. Selectivity for stimulus direction, speed, and orientation. *J. Neurophysiology, 49,* 1127–1147.

Miller, P. (2006). Analysis of spike statistics in neuronal systems with continuous attractors or multiple, discrete attractor states. *Neural Computation, 18,* 1268–1317.

Montani, F., Kohn, A., Smith, M., & Schultz, S. (2007). The role of correlations in direction and contrast coding in the primary visual cortex. *J. Neuroscience, 27,* 2338–2348.

Nirenberg, S., & Latham, P. (2003). Decoding neuronal spike trains: How important are correlations? *Proc. Natl. Acad. Sci., 100,* 7348–7353.

Renart, A., Song, P., & Wang, X. (2003). Robust spatial working memory through homeostatic synaptic scaling in heterogeneous cortical networks. *Neuron, 38,* 473–485.

Rolls, E. T., Robertson, R. G., & Georges-François, P. (1995). The representation of space in the primate hippocampus. *Soc. Neurosci. Abstr., 21,* 1494.

Seung, H. S. (1996). How the brain keeps the eyes still. *Proc. Acad. Sci. USA, 93,* 13339–13344.

Seung, H. S., Lee, D. D., Reis, B. Y., & Tank, D. W. (2000). Stability of the memory of eye position in a recurrent network of conductance-based model neurons. *Neuron, 26,* 259–271.

Shepard, S., & Metzler, D. (1988). Mental rotation: Effects of dimensionality of objects and type of task. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 3–11.

Shepard, R., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 171,* 701–703.

Sompolinsky, H., Yoon, H., Kang, K., & Shamir, M. (2001). Population coding in neuronal systems with correlated noise. *Physical Review E, 64,* 051904.

Spiridon, M., & Gerstner, W. (2001). Effect of lateral connections on the accuracy of the population code for a network of spiking neurons. *Network, 12,* 209–421.

Stringer, S. M., Trappenberg, T. P., Rolls, E., & Aranjo, I. (2002). Self-organzing continuous attractor networks and path integration: One-dimensional models of head direction cells. *Network: Computation in Neural Systems, 13,* 217–242.

Taube, J. S. (1998). Head direction cells and the neurophysiological basis for a sense of direction. *Prog. Neurobiol., 55,* 225–256.

Trappenberg, T. (2003). Continuous attractor neural networks. In L. N. de Castro & F. J. V. Zuben (Eds.), *Recent developments in biologically inspired computing*. Hershey, PA: Idea Group.

Tuckwell, H. (1988). *Introduction to theoretical neurobiology*. Cambridge: Cambridge University Press.

Wang, X. J. (2001). Synaptic reverberation underlying mnemonic persistent activitity. *Trends in Neuroscience, 24,* 455–463.

Wilson, M. A., & McNaughton, B. L. (1993). Dynamics of hippocampal ensemble code for space. *Science, 261,* 1055–1058.

Wu, S., & Amari, S. (2005). Computing with continuous attractors: Stability and on-line aspects. *Neural Computation, 17,* 2215–2239.

Wu, S., Amari, S., & Nakahara, H. (2002). Population coding and decoding in a neural field: A computational study. *Neural Computation, 14,* 999–1026.

Wu, S., Amari, S., & Nakahara, H. (2004). Information processing in a neuron ensemble with the multiplicative correlation structure. *Neural Networks, 17,* 205–214.

Wu, S., Nakahara, H., & Amari, S. (2001). Population coding with correlation an unfaithful model. *Neural Computation, 13,* 775–797.

Zhang, K.-C. (1996). Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: A theory. *J. Neuroscience, 16,* 2112–2126.

Received October 24, 2006; accepted May 9, 2007.