International Journal of Bifurcation and Chaos, Vol. 25, No. 4 (2015) 1550054 (25 pages)

DOI: 10.1142/S0218127415500546

Synthetic Modeling of Autonomous Learning with a Chaotic Neural Network

Masatoshi Funabashi
Sony Computer Science Laboratories, Inc.,
3-14-13 Higashi-Gotanda, Shinagawa-ku,
Tokyo 141-0022, Japan
masa_funabashi@csl.sony.co.jp

Received July 9, 2014

We investigate the possible role of intermittent chaotic dynamics called chaotic itinerancy, in interaction with nonsupervised learnings that reinforce and weaken the neural connection depending on the dynamics itself. We first performed hierarchical stability analysis of the Chaotic Neural Network model (CNN) according to the structure of invariant subspaces. Irregular transition between two attractor ruins with positive maximum Lyapunov exponent was triggered by the blowout bifurcation of the attractor spaces, and was associated with riddled basins structure. We secondly modeled two autonomous learnings, Hebbian learning and spike-timing-dependent plasticity (STDP) rule, and simulated the effect on the chaotic itinerancy state of CNN. Hebbian learning increased the residence time on attractor ruins, and produced novel attractors in the minimum higher-dimensional subspace. It also augmented the neuronal synchrony and established the uniform modularity in chaotic itinerancy. STDP rule reduced the residence time on attractor ruins, and brought a wide range of periodicity in emerged attractors, possibly including strange attractors. Both learning rules selectively destroyed and preserved the specific invariant subspaces, depending on the neuron synchrony of the subspace where the orbits are situated. Computational rationale of the autonomous learning is discussed in connectionist perspective.

Keywords: Chaotic itinerancy; invariant subspace; blowout bifurcation; riddled basins; Hebbian and STDP learning.

1. Introduction

Recent development of neuroscience has been largely promoted by the interaction between biological experiment and mathematical modeling. Although experimental setting is limited to a part of entire phenomenon situated in the ultimate interest to understand brain functions, different measurement methods varying in both temporal and spacial scales enable us to acquire experimental evidences ranging from single neuron to cortical and cerebral activities, and integrate them into a simulated model with certain assumption. This interaction between experimental measurement and mathematical modeling helps us not only to reproduce the observed brain dynamics itself, but to discover appropriate levels of description that are accessible to mathematical analysis and eventually help hypothesis testing and further hypothesis forming in experiment. Understanding complex brain dynamics such as chaotic behavior of neurons especially postulates both theoretical and experimental approaches.

Aihara *et al.* first discovered chaotic dynamics of single neuron activity in the squid giant axon with such experiment-modeling interaction [Aihara *et al.*, 1984, 1985]. Skarda and Freeman investigated the chaotic dynamics in olfactory bulb of rabbit, and discovered the function of chaos as catalyst of learning in macroscopic level [Skarda & Freeman, 1987]. In their study, the chaotic wandering state

among memorized patterns is shown to additionally create another attractor corresponding to the novel odor.

Later Tsuda showed in a modeling study of cortical chaos that such intermittent chaos associated with temporal laminar phase realizes the acquisition of externally exposed patterns without losing the former memory structure [Tsuda, 1992]. This dynamical linking of memory with chaotic transition is further investigated in more physiologically detailed settings, and are conceptualized in the name of *chaotic itinerancy* [Fujii *et al.*, 2004; Fujii & Tsuda, 2004].

Chaotic itinerancy is based on the concept of attractor ruins, where plural attractors localized in low-dimensional states lose their transverse stability. Possible candidate of attractor ruin is considered as destabilized Milnor attractor that went through the crisis [Tsuda & Fujii, 2004; Kaneko *et al.*, 2004; Kaneko & Tsuda, 2003]. Chaotic itinerancy is also observed in a Chaotic Neural Network model (CNN), with the chaotic neuron based on the physiological property [Aihara *et al.*, 1990; Adachi & Aihara, 1997].

Besides the learning of the exposed patterns, there has recently been studies on the self-organizing change of the network's connectivity without any supervisor signals in neural network model [Kang *et al.*, 2008] and in relevance to neural network [Ito & Kaneko, 2002]. These spontaneous changes of connectivity are based on the local learning rule such as Hebbian learning and the spike-time-dependent plasticity (STDP) learning. Indeed, such autonomous learning that depends on the dynamics of the network itself can be considered to happen widely and permanently as long as the brain manifests spontaneous activity and possesses plasticity. It is of further interest whether and how such internal dynamics can lead to structure formation of neural network and realize adaptive function.

In this article, we investigate the possible role of chaotic itinerancy associated with autonomous learning without supervisor signal. For this purpose, we combine the chaotic neural network model with Hebbian and STDP learning rules and investigate the dynamical reconfiguration of attractor patterns.

2. Dynamics of Chaotic Neural Network

We first define the chaotic neural network model (CNN) with two periodic attractors in different invariant subspaces, and investigate the dynamics without any learning rule.

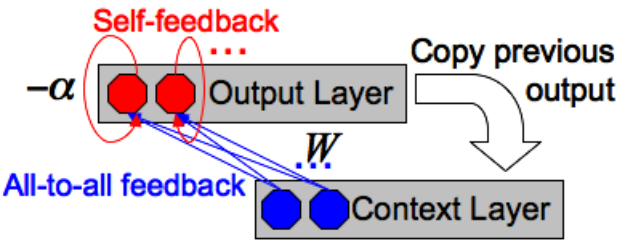


Fig. 1. Architecture of CNN.

2.1. *Definition of chaotic neural network*

The architecture of CNN is defined as in Fig. 1. CNN is a discrete time system that consists of two layers of neuron model: Context and Output. Each neuron of the Context layer is connected to all neurons in the Output layer. The output of the Context layer at time t is identical to that of the Output layer at time $t-1$. The chaotic neurons used in the Output layer of CNN are defined as follows [Aihara *et al.*, 1990]:

$$x_i(t+1) = f\left(\sum_{j=1}^{n} w_{ij} \sum_{d=0}^{t} k_f^d x_j(t-d) - \alpha \sum_{d=0}^{t} k_r^d x_i(t-d) + \theta_{\text{out}}\right), \quad (1)$$

where at time t, $x_i(t)$ is the output of the ith chaotic neuron, and θ_i^{out} is its threshold. The parameters $\alpha \geq 0$ and k_r $(0 \leq k_r \leq 1)$ control the refractoriness of neuron. By augmenting these parameters, the orbital stability changes, and can induce chaotic dynamics. Hence, the situation $\alpha = k_r = 0$ corresponds to simple analog neuron model. The exponentially decreasing influences of past outputs (outputs of the Context layer) are controlled by k_f $(0 \leq k_f \leq 1)$. n is the number of chaotic neurons. In this study, eight chaotic neurons with fixed parameter $k_f = 0.1$ were used in the Output layer of CNN. θ_{out} will be defined as a function of k_r in the following section.

The sigmoid function $f(\cdot)$ is defined as follows, with the increment parameter β.

$$f(x) = \frac{1}{1+\exp(-\beta x)}. \quad (2)$$

We used $\beta = 5.0$ in this study.

Table 1. Definition of the four patterns A, B, C, D to be memorized in the Output layer of CNN.

Pattern Symbol	Pattern Vector
A	$(1,1,1,1,0,0,0,0)^T$
B	$(0,0,0,0,1,1,1,1)^T$
C	$(1,1,0,0,0,0,1,1)^T$
D	$(0,0,1,1,1,1,0,0)^T$

The connection matrix $W = (w_{ij})$ was set with Hebbian rule to memorize two two-periodic patterns A $\rightleftharpoons$ B and C $\rightleftharpoons$ D defined in Table 1. Let $A = (a_1, \ldots, a_8)^T$, $B = (b_1, \ldots, b_8)^T$, $C = (c_1, \ldots, c_8)^T$, $D = (d_1, \ldots, d_8)^T$, where T is the transpose of vector. Then $W = (w_{ij})$ $(1 \le i, j \le 8)$ was defined as follows:

$$w_{ij} = \frac{1}{4}\{(2a_i - 1)(2b_j - 1) + (2b_i - 1)(2a_j - 1) + (2c_i - 1)(2d_j - 1) + (2d_i - 1)(2c_j - 1)\}. \tag{3}$$

Consequently,

$$W = \begin{pmatrix} -1 & -1 & 0 & 0 & 1 & 1 & 0 & 0 \\ -1 & -1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & -1 & -1 & 0 & 0 & 1 & 1 \\ 0 & 0 & -1 & -1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & -1 & -1 & 0 & 0 \\ 1 & 1 & 0 & 0 & -1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & -1 & -1 \\ 0 & 0 & 1 & 1 & 0 & 0 & -1 & -1 \end{pmatrix}, \tag{4}$$

was obtained. Note that the number of neurons and the patterns of the attractors was chosen to realize chaotic itinerancy between attractor ruins situated in different invariant subspaces with the simplest network, which is described in the following section.

We judge the retrieval of a pattern at time t by interpreting the output $\mathbf{x}(t) = (x_1(t), \ldots, x_8(t))^T$ into binary value $\mathbf{h}(t) = (h_1(t), \ldots, h_8(t))^T$ as follows:

$$h_i(t) = \begin{cases} 0 & \text{if } x_i(t) < 0.5 \\ 1 & \text{else.} \end{cases} \tag{5}$$

We state that the network retrieved a pattern when the value of $\mathbf{h}(t)$ coincides with any of the patterns A, B, C, D. By setting the refractoriness parameters $k_r = \alpha = 0$, the network retrieved either of the two-periodic patterns A $\rightleftharpoons$ B or C $\rightleftharpoons$ D depending on the initial condition. Note that almost all initial conditions converge to either of these patterns, unless specifically taken on the invariant subspaces where (11335577) is not included (see next section).

2.2. *Invariant subspaces*

In discrete time neural networks, considering certain symmetry of the model, it is possible to decompose its dynamics into subspaces of the whole state space. The decomposition of the dynamics into lower dimensions has a possibility to characterize the network dynamics more precisely. The decomposed subspaces in which the dynamics can fall and remain permanently are called invariant subspaces. In this section, we derive the invariant subspaces of CNN based on [Komuro & Aihara, 2001].

First of all, let the CNN be a differentiable map

$$\Phi = (\Phi_f, \Phi_r) : R^n \times R^n \to R^n \times R^n, \tag{6}$$

defined by

$$\begin{cases} \Phi_f(\eta, \zeta) = k_f \eta + W\mathbf{x} \\ \Phi_r(\eta, \zeta) = k_r \zeta + \theta_r - \alpha \mathbf{x}, \end{cases} \tag{7}$$

where

$$\mathbf{x} = \mathbf{x}(\eta, \zeta) = (f(\eta_i + \zeta_i))_{i=1}^n \tag{8}$$

and

$$\theta_r = \theta_{\text{out}}(1 - k_r). \tag{9}$$

$\theta_r = 0.0$ was used for the following simulation.

Now, let S_n be a symmetric group with degrees of n, and $C(W)$ be a subgroup of S_n. For $\sigma \in S_n$, we define a linear transformation $P_\sigma : R^n \to R^n$ by

$$P_\sigma : (u_1, \ldots, u_n)^T \mapsto (u_{\sigma(1)}, \ldots, u_{\sigma(n)})^T. \tag{10}$$

For $\sigma \in S_n$, we define that σ belongs to $C(W)$ if and only if

$$P_\sigma W = W P_\sigma. \tag{11}$$

This condition corresponds to the condition $w_{ij} = w_{\sigma(i)\sigma(j)}$ of the connection matrix between the Output and Context layers. Using this symmetry of the connection matrix, we define the invariant subspaces of CNN as follows. Let $\sigma \in C(W)$. Then $\Phi = (\Phi_f, \Phi_r)$ of a CNN is $P_\sigma \times P_\sigma$-invariant;

$$(P_\sigma \times P_\sigma)\Phi = \Phi(P_\sigma \times P_\sigma). \tag{12}$$

We define a linear subspace $H_1(\sigma)$ of R^n by

$$H_1(\sigma) = \{(u_1, \ldots, u_n)^T \in R^n \,|\, u_i = u_{\sigma(i)}, \quad 1 \le i \le n\}. \tag{13}$$

We also define the invariant subspace $H(\sigma)$ of R^{2n} by

$$\begin{aligned} H(\sigma) &= H_1(\sigma) \times H_1(\sigma) \\ &= \{(\eta_1, \ldots, \eta_N; \zeta_1, \ldots, \zeta_N)^T \in (\eta_i, \zeta_i) \\ &= (\eta_{\sigma(i)}, \zeta_{\sigma(i)}), 1 \le i \le n\}. \end{aligned} \tag{14}$$

For each $\sigma \in S_n$, $H(\sigma) \subset R^{2n}$ is Φ-invariant;

$$\Phi(H(\sigma)) \subset H(\sigma). \tag{15}$$

Hence, this holds for all time steps of CNN. Consequently, it means that if the initial states of CNN were taken inside of $H(\sigma)$, the ith neuron and the $\sigma(i)$th neuron of the Output layer are always synchronized regardless of transversal stability, and output the same value at each time step.

Here, we give an example of the invariant subspace. If

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 2 & 1 & 4 & 3 & 6 & 5 & 8 & 7 \end{pmatrix} \in C(W), \tag{16}$$

then

$$\begin{aligned} H(\sigma) = \{&(\eta_1, \eta_1, \eta_3, \eta_3, \eta_5, \eta_5, \eta_7, \eta_7; \\ &\zeta_1, \zeta_1, \zeta_3, \zeta_3, \zeta_5, \zeta_5, \zeta_7, \zeta_7)^T\}. \end{aligned} \tag{17}$$

For simplicity, let us denote

$$H(\sigma) = (11335577), \tag{18}$$

representing $H(\sigma)$ with the minimum subscripts of vector η and ζ that are identical by P_σ. We take this notation rule for the following sections.

The defined patterns A and B are situated inside of (11115555), while C and D are inside of (11333311). These two invariant subspaces are isomorphic from the symmetry of the model, and therefore it is sufficient to investigate only one subspace to know the dynamics in both.

2.3. *Periodicity analysis*

In chaotic itinerancy state of CNN, the orbit is reported to visit irregularly among memorized patterns [Adachi & Aihara, 1997]. We investigated the periodicity of the temporal dynamics to search for irregular orbits where chaotic itinerancy may occur, by changing empirically the refractoriness parameters k_r and α. The result is shown in Fig. 3.

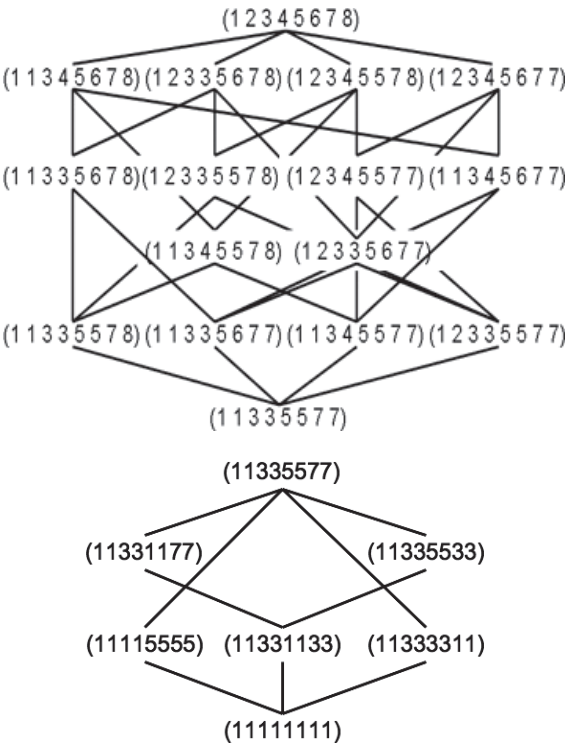


Fig. 2. Invariant subspaces of CNN in the defined setting. (top) (11335577) and (bottom) (11335577). The lines denote inclusion relation. Although there exist actually 76 invariant subspaces, for simplicity only those in the above/below hierarchy of (11335577) are shown.

2.4. *Deviation rate from attractor ruins*

Chaotic itinerancy is characterized not only by the lack of periodicity, but the intermittent visit to the

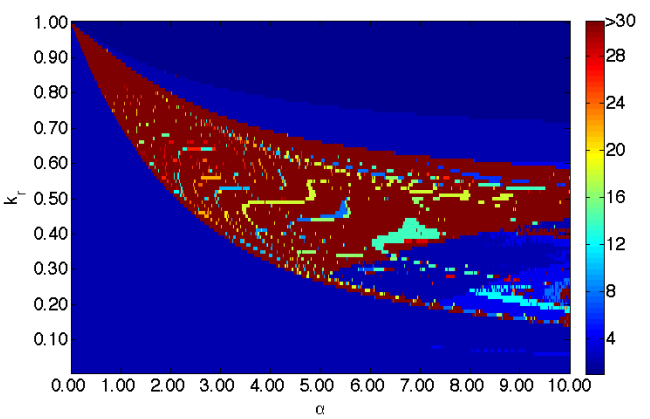


Fig. 3. Periodicity of CNN. Horizontal axis: α. Vertical axis: k_r. After cutting 5000 transient, periodicity was judged with the tolerance of 1.0e−6 for each neuron output. More than 30 periodic orbits are depicted in brown. After cutting 5000 transient, 10 000 steps were used to calculate the periodicity at each 0.01 step of k_r and α. Initial conditions were taken randomly.

attractor ruins. To find such parameter region in k_r–α plane, we calculated the deviation rate $DR(k_r, \alpha)$ from the patterns A, B, C, D, which is defined as follows:

$$DR(k_r,\alpha) = \frac{1}{N}\lim_{N\to\infty}\sum_{t=1}^{N}\Bigg[1 - \prod_{i=1}^{n}\{\delta(h_i(t), a_i)\} - \prod_{i=1}^{n}\{\delta(h_i(t), b_i)\} - \prod_{i=1}^{n}\{\delta(h_i(t), c_i)\} - \prod_{i=1}^{n}\{\delta(h_i(t), d_i)\}\Bigg], \quad (19)$$

where $\delta(\cdot,\cdot)$ is the delta function. $N = 10\,000$ was used for the calculation. The result is shown in Fig. 4. In case of $DR(k_r, \alpha) = 0$, the orbit stays in either of the periodic cycle A $\rightleftharpoons$ B or C $\rightleftharpoons$ D. While in case $DR(k_r, \alpha) = 1$, the orbit never visits any of the memorized patterns. Chaotic itinerancy can occur in $0 < DR(k_r, \alpha) < 1$.

We also investigated the wandering range among the patterns A, B, C, D in k_r–α plane to find the parameter region where the orbits visit both pairs of patterns A $\rightleftharpoons$ B and C $\rightleftharpoons$ D. In such a case, both periodic cycles are partially destabilized and become the attractor ruins that allow the orbits to escape and be attracted intermittently. The result is shown in Fig. 5.

The irregular transition between two attractor ruins therefore occurs in the parameter regions where there is no obvious periodicity (inside brown regions in Fig. 3), more than 0 but less than 1 deviation rate in Fig. 4, and wandering orbit ranging among all patterns (red regions in Fig. 5).

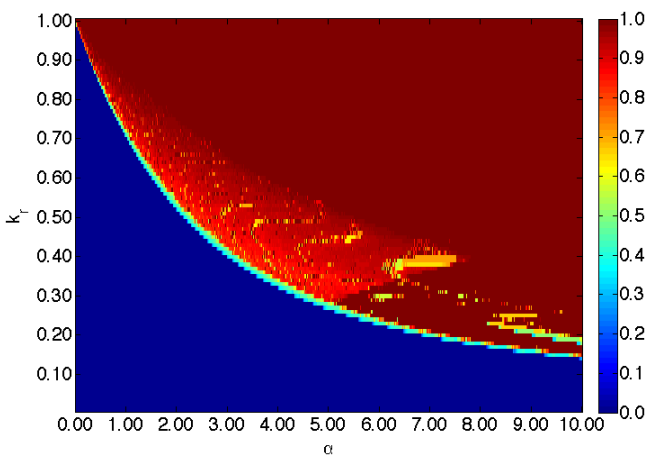


Fig. 4. Deviation rate from memorized patterns of CNN. Horizontal axis: α. Vertical axis: k_r. After cutting 5000 transient, 10 000 steps were used to calculate the deviation rate at each 0.01 step of k_r and α. Initial conditions were taken randomly.

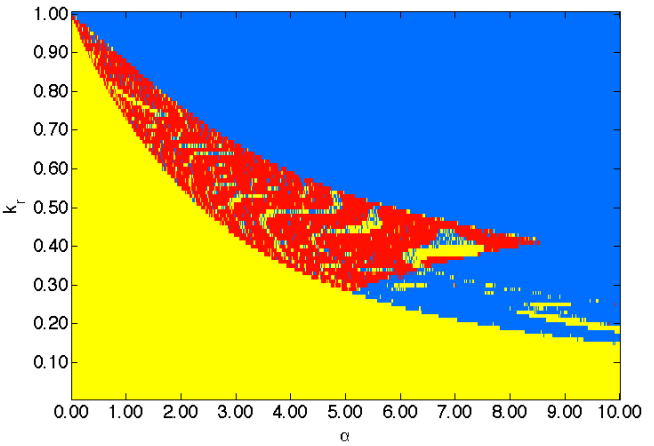


Fig. 5. Wandering range among memorized patterns of CNN. Horizontal axis: α. Vertical axis: k_r. The blue region is where the dynamics visit none of the patterns A, B, C, D. The yellow region is where the dynamics visit only either pair of patterns A $\rightleftharpoons$ B or C $\rightleftharpoons$ D. The red region is where the dynamics visit both pairs of patterns A $\rightleftharpoons$ B and C $\rightleftharpoons$ D. After cutting 5000 transient, 10 000 steps were used to calculate the wandering range at each 0.01 step of k_r and α. Initial conditions were taken randomly.

2.5. *Linear stability analysis*

Chaotic itinerancy state is assumed to follow chaotic orbit in the transition process between attractor ruins. Therefore, the orbital stability should be investigated to prove the existence of chaos. Since possible attractor ruins A $\rightleftharpoons$ B and C $\rightleftharpoons$ D are situated inside of the invariant subspaces (11115555) and (11333311) respectively, chaotic wandering orbits between these patterns are situated in the upper hierarchy of invariant subspaces. Indeed, in most of the parameter region, the dynamics is situated inside of (11335577). To properly examine the stability of each invariant subspace, it is important to consider the effect of such hierarchical structure on the dynamics. Classical method to calculate Lyapunov spectrum such as Gram–Schmidt orthonormalization simply calculates the Lyapunov exponents in the descending order, and does not necessarily consider such symmetry of the model. We derive here the method to decompose the Lyapunov spectrum according to a series of chosen invariant subspaces.

First, we formulate the Jacobian matrix $D\Phi(\eta, \zeta)$ of CNN as follows:

$$D\Phi(\eta,\zeta)=\begin{pmatrix} D_\eta\Phi_f(\eta,\zeta) & D_\zeta\Phi_f(\eta,\zeta)\\ D_\eta\Phi_r(\eta,\zeta) & D_\zeta\Phi_r(\eta,\zeta)\end{pmatrix}, \tag{20}$$

where

$$D_\eta\Phi_f(\eta,\zeta) = \begin{pmatrix} k_f & & 0\\ & \ddots & \\ 0 & & k_f\end{pmatrix} + W\begin{pmatrix} Df(\eta_1+\zeta_1) & & 0\\ & \ddots & \\ 0 & & Df(\eta_n+\zeta_n)\end{pmatrix}, \tag{21}$$

$$D_\zeta\Phi_f(\eta,\zeta) = W\begin{pmatrix} Df(\eta_1+\zeta_1) & & 0\\ & \ddots & \\ 0 & & Df(\eta_n+\zeta_n)\end{pmatrix}, \tag{22}$$

$$D_\eta\Phi_r(\eta,\zeta) = -\alpha\begin{pmatrix} Df(\eta_1+\zeta_1) & & 0\\ & \ddots & \\ 0 & & Df(\eta_n+\zeta_n)\end{pmatrix}, \tag{23}$$

$$D_\zeta\Phi_r(\eta,\zeta) = \begin{pmatrix} k_r & & 0\\ & \ddots & \\ 0 & & k_r\end{pmatrix} - \alpha\begin{pmatrix} Df(\eta_1+\zeta_1) & & 0\\ & \ddots & \\ 0 & & Df(\eta_n+\zeta_n)\end{pmatrix}. \tag{24}$$

D_η and D_ζ denote the differential operator with respect to η and ζ, respectively.

We consider the perturbation in the state space of the chaotic neuron $(\eta,\zeta)\in R^{2n}$ to the directions inside and outside of the invariant subspaces (11115555) and (11335577), where chaotic itinerancy takes place. For simplicity, we define the transpose σ_2 and σ_4 as follows, under which the orbits inside of two-dimensional (11115555) and four-dimensional (11335577) are invariant, respectively.

$$\sigma_2=\begin{pmatrix}1&2&3&4&5&6&7&8\\4&1&2&3&8&5&6&7\end{pmatrix}\in C(W), \tag{25}$$

$$\sigma_4=\begin{pmatrix}1&2&3&4&5&6&7&8\\2&1&4&3&6&5&8&7\end{pmatrix}\in C(W). \tag{26}$$

We then define the perturbation vectors $p_1^{2\text{in}},\ldots,p_4^{2\text{in}}$ to the directions inside of (11115555), and $p_1^{2\text{out}},\ldots,p_4^{2\text{out}}$ to the directions outside of (11115555) but inside of (11335577) as follows:

$$(p_1^{2\text{in}},\ldots,p_4^{2\text{in}})=\delta\begin{pmatrix}1&0&0&0\\1&0&0&0\\1&0&0&0\\1&0&0&0\\0&1&0&0\\0&1&0&0\\0&1&0&0\\0&1&0&0\\0&0&1&0\\0&0&1&0\\0&0&1&0\\0&0&1&0\\0&0&0&1\\0&0&0&1\\0&0&0&1\\0&0&0&1\end{pmatrix}, \tag{27}$$

$$(p_1^{2\text{out}},\ldots,p_4^{2\text{out}})=\delta\begin{pmatrix}1&0&0&0\\1&0&0&0\\-1&0&0&0\\-1&0&0&0\\0&1&0&0\\0&1&0&0\\0&-1&0&0\\0&-1&0&0\\0&0&1&0\\0&0&1&0\\0&0&-1&0\\0&0&-1&0\\0&0&0&-1\\0&0&0&-1\\0&0&0&-1\\0&0&0&-1\end{pmatrix},$$

where δ represents infinitesimal quantity. Hence the eight vectors $p_1^{2\text{in}}, \ldots, p_4^{2\text{in}}$, $p_1^{2\text{out}}, \ldots, p_4^{2\text{out}}$ form the linearly independent basis of the state space $(\eta, \zeta) \in (11335577)$.

In the same way, we also define the perturbations $p_1^{4\text{in}}, \ldots, p_8^{4\text{in}}$ to the direction inside of (11335577), and $p_1^{4\text{out}}, \ldots, p_8^{4\text{out}}$ to the direction outside of (11335577) as follows:

$$(p_1^{4\text{in}}, \ldots, p_8^{4\text{in}}) = \delta \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \tag{28}$$

$$(p_1^{4\text{out}}, \ldots, p_8^{4\text{out}}) = \delta \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \end{pmatrix}. \tag{29}$$

The 16 vectors $p_1^{4\text{in}}, \ldots, p_8^{4\text{in}}$, $p_1^{4\text{out}}, \ldots, p_8^{4\text{out}}$ form the linearly independent basis of the state space $(\eta, \zeta) \in (12345678)$.

We then consider the development of these perturbations with time development. Using the numerical symmetry of Jacobian matrix $D\Phi(\eta, \zeta)$, the linear approximation of the perturbation development for one time step $D\Phi(\eta, \zeta)p$ has the following relation.

$$D\Phi(\eta, \zeta)p \in (11115555) \quad \text{if } p \in \{p_1^{2\text{in}}, \ldots, p_4^{2\text{in}}\}, \quad (\eta, \zeta) \in (11115555), \tag{30}$$

$$D\Phi(\eta,\zeta)p \in (11115555)^{\perp} \quad \text{if } p \in \{p_1^{2\text{out}},\ldots,p_4^{2\text{out}}\}, \quad (\eta,\zeta) \in (11115555), \tag{31}$$

$$D\Phi(\eta,\zeta)p \in (11335577) \quad \text{if } p \in \{p_1^{4\text{in}},\ldots,p_8^{4\text{in}}\}, \quad (\eta,\zeta) \in (11335577), \tag{32}$$

$$D\Phi(\eta,\zeta)p \in (11335577)^{\perp} \quad \text{if } p \in \{p_1^{4\text{out}},\ldots,p_8^{4\text{out}}\}, \quad (\eta,\zeta) \in (11335577). \tag{33}$$

Here, $(11115555)^{\perp}$ is the state space outside of (11115555) but inside of (11335577), while $(11335577)^{\perp}$ is outside of (11335577). For simplicity, we call $(11115555)^{\perp}$ simply as "outside of (11115555)". These relations mean that the perturbation developments are strictly separated according to the hierarchical structure of the invariant subspaces, and it is possible to decompose them into the directions inside and outside of where the orbit is situated.

Since the invariant subspaces are defined on the synchrony of the state space, their geometrical compositions are situated always in the diagonal lines of the coordinates (η,ζ). We are now motivated to choose another coordinate system to separate the independent development of the perturbations. We first define the coordinates $(\mathbf{H},\mathbf{Z}) = (H_1,\ldots,H_8,Z_1,\ldots,Z_8)^T$ which separate (11335577) and $(11335577)^{\perp}$ as follows.

$$(\mathbf{H},\mathbf{Z}) = A(\eta,\zeta), \tag{34}$$

where

$$A = (a_{ij}) \tag{35}$$

$$= \begin{pmatrix} \cos\left(\frac{\pi}{4}\right) & \sin\left(\frac{\pi}{4}\right) & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \cos\left(\frac{\pi}{4}\right) & \sin\left(\frac{\pi}{4}\right) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \cos\left(\frac{\pi}{4}\right) & \sin\left(\frac{\pi}{4}\right) & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \cos\left(\frac{\pi}{4}\right) & \sin\left(\frac{\pi}{4}\right) \\ -\cos\left(\frac{\pi}{4}\right) & \sin\left(\frac{\pi}{4}\right) & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -\cos\left(\frac{\pi}{4}\right) & \sin\left(\frac{\pi}{4}\right) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -\cos\left(\frac{\pi}{4}\right) & \sin\left(\frac{\pi}{4}\right) & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -\cos\left(\frac{\pi}{4}\right) & \sin\left(\frac{\pi}{4}\right) \end{pmatrix}. \tag{36}$$

The transformation matrix A is chosen so that $H_1,\ldots,H_4$ and $Z_1,\ldots,Z_4$ form the coordinates inside of (11335577), while $H_5,\ldots,H_8$ and $Z_5,\ldots,Z_8$ outside of (11335577).

Then the CNN $\Phi : R^{2n} \to R^{2n}$ can be reformulated on the coordinates $(\mathbf{H},\mathbf{Z})$ as $\Psi : R^{2n} \to R^{2n}$:

$$\Psi = (\Psi_f,\Psi_r) : R^n \times R^n \to R^n \times R^n, \tag{37}$$

$$\begin{cases} \Psi_f(\mathbf{H}(t),\mathbf{Z}(t)) = \mathbf{H}(t+1), \\ \Psi_r(\mathbf{H}(t),\mathbf{Z}(t)) = \mathbf{Z}(t+1), \end{cases} \tag{38}$$

$$\begin{cases} H_i(t+1) = k_f H_i(t) + \sum_k a_{ik} \sum_j w_{kj} f\left(\sum_l a_{lj}(H_l(t) + Z_l(t))\right) \\ Z_i(t+1) = k_r Z_i(t) + \sum_k a_{ik}\theta_r - \alpha \sum_k a_{ik} f\left(\sum_l a_{lk}(H_l(t) + Z_l(t))\right) \end{cases} \quad (1 \le i \le 8). \tag{39}$$

The corresponding Jacobian matrix $D\Psi(\mathbf{H}, \mathbf{Z})$ becomes as follows:

$$D\Psi(\mathbf{H}, \mathbf{Z}) = \begin{pmatrix} D_H\Psi_f(\mathbf{H}, \mathbf{Z}) & D_Z\Psi_f(\mathbf{H}, \mathbf{Z}) \\ D_H\Psi_r(\mathbf{H}, \mathbf{Z}) & D_Z\Psi_r(\mathbf{H}, \mathbf{Z}) \end{pmatrix}, \tag{40}$$

where

$$\begin{aligned} &D_H\Psi_f(\mathbf{H}, \mathbf{Z}) \\ &= \begin{pmatrix} k_f & & 0 \\ & \ddots & \\ 0 & & k_f \end{pmatrix} \\ &\quad + \begin{pmatrix} \sum_k a_{1k} \sum_j w_{kj} D_{H_1} f\left(\sum_l a_{lj}(H_l + Z_l)\right) & \cdots & \sum_k a_{1k} \sum_j w_{kj} D_{H_8} f\left(\sum_l a_{lj}(H_l + Z_l)\right) \\ \vdots & \ddots & \vdots \\ \sum_k a_{8k} \sum_j w_{kj} D_{H_1} f\left(\sum_l a_{lj}(H_l + Z_l)\right) & \cdots & \sum_k a_{8k} \sum_j w_{kj} D_{H_8} f\left(\sum_l a_{lj}(H_l + Z_l)\right) \end{pmatrix}, \end{aligned} \tag{41}$$

$$\begin{aligned} &D_Z\Psi_f(\mathbf{H}, \mathbf{Z}) \\ &= \begin{pmatrix} \sum_k a_{1k} \sum_j w_{kj} D_{Z_1} f\left(\sum_l a_{lj}(H_l + Z_l)\right) & \cdots & \sum_k a_{1k} \sum_j w_{kj} D_{Z_8} f\left(\sum_l a_{lj}(H_l + Z_l)\right) \\ \vdots & \ddots & \vdots \\ \sum_k a_{8k} \sum_j w_{kj} D_{Z_1} f\left(\sum_l a_{lj}(H_l + Z_l)\right) & \cdots & \sum_k a_{8k} \sum_j w_{kj} D_{Z_8} f\left(\sum_l a_{lj}(H_l + Z_l)\right) \end{pmatrix}, \end{aligned} \tag{42}$$

$$\begin{aligned} &D_H\Psi_r(\mathbf{H}, \mathbf{Z}) \\ &= -\alpha \begin{pmatrix} \sum_k a_{1k} D_{H_1} f\left(\sum_l a_{lk}(H_l + Z_l)\right) & \cdots & \sum_k a_{1k} D_{H_8} f\left(\sum_l a_{lk}(H_l + Z_l)\right) \\ \vdots & \ddots & \vdots \\ \sum_k a_{8k} D_{H_1} f\left(\sum_l a_{lk}(H_l + Z_l)\right) & \cdots & \sum_k a_{8k} D_{H_8} f\left(\sum_l a_{lk}(H_l + Z_l)\right) \end{pmatrix}, \end{aligned} \tag{43}$$

$$D_Z\Psi_r(\mathbf{H},\mathbf{Z}) = \begin{pmatrix} k_r & & 0 \\ & \ddots & \\ 0 & & k_r \end{pmatrix} - \alpha \begin{pmatrix} \sum_k a_{1k} D_{Z_1} f\left(\sum_l a_{lk}(H_l+Z_l)\right) & \cdots & \sum_k a_{1k} D_{Z_8} f\left(\sum_l a_{lk}(H_l+Z_l)\right) \\ \vdots & \ddots & \vdots \\ \sum_k a_{8k} D_{Z_1} f\left(\sum_l a_{lk}(H_l+Z_l)\right) & \cdots & \sum_k a_{8k} D_{Z_8} f\left(\sum_l a_{lk}(H_l+Z_l)\right) \end{pmatrix}. \quad (44)$$

For simplicity, we denote the (i,j)th element of the matrix $D_H\Psi_f(\mathbf{H},\mathbf{Z})$, $D_Z\Psi_f(\mathbf{H},\mathbf{Z})$, $D_H\Psi_r(\mathbf{H},\mathbf{Z})$, $D_Z\Psi_r(\mathbf{H},\mathbf{Z})$ as $D_{H_i}\Psi_{f_j}$, $D_{Z_i}\Psi_{f_j}$, $D_{H_i}\Psi_{r_j}$, $D_{Z_i}\Psi_{r_j}$, respectively.

From the fact (32) and (33), the Jacobian $D\Psi(\mathbf{H},\mathbf{Z})$ can be decomposed into the following, since the perturbation inside of (11335577) never turns outside and vice versa.

$$D\Psi(\mathbf{H},\mathbf{Z}) = \begin{pmatrix} D_{\mathbf{H}_{\text{in}}}\Psi_{f_{\text{in}}} & 0 & D_{\mathbf{Z}_{\text{in}}}\Psi_{f_{\text{in}}} & 0 \\ 0 & D_{\mathbf{H}_{\text{out}}}\Psi_{f_{\text{out}}} & 0 & D_{\mathbf{Z}_{\text{out}}}\Psi_{f_{\text{out}}} \\ D_{\mathbf{H}_{\text{in}}}\Psi_{r_{\text{in}}} & 0 & D_{\mathbf{Z}_{\text{in}}}\Psi_{r_{\text{in}}} & 0 \\ 0 & D_{\mathbf{H}_{\text{out}}}\Psi_{r_{\text{out}}} & 0 & D_{\mathbf{Z}_{\text{out}}}\Psi_{r_{\text{out}}} \end{pmatrix}, \quad (45)$$

where

$$D_{\mathbf{H}_{\text{in}}}\Psi_{f_{\text{in}}} = \begin{pmatrix} D_{H_1}\Psi_{f_1} & \cdots & D_{H_4}\Psi_{f_1} \\ \vdots & \ddots & \vdots \\ D_{H_1}\Psi_{f_4} & \cdots & D_{H_4}\Psi_{f_4} \end{pmatrix}, \quad (46)$$

$$D_{\mathbf{Z}_{\text{in}}}\Psi_{f_{\text{in}}} = \begin{pmatrix} D_{Z_1}\Psi_{f_1} & \cdots & D_{Z_4}\Psi_{f_1} \\ \vdots & \ddots & \vdots \\ D_{Z_1}\Psi_{f_4} & \cdots & D_{Z_4}\Psi_{f_4} \end{pmatrix}, \quad (47)$$

$$D_{\mathbf{H}_{\text{out}}}\Psi_{f_{\text{out}}} = \begin{pmatrix} D_{H_5}\Psi_{f_5} & \cdots & D_{H_8}\Psi_{f_5} \\ \vdots & \ddots & \vdots \\ D_{H_5}\Psi_{f_8} & \cdots & D_{H_8}\Psi_{f_8} \end{pmatrix}, \quad (48)$$

$$D_{\mathbf{Z}_{\text{out}}}\Psi_{f_{\text{out}}} = \begin{pmatrix} D_{Z_5}\Psi_{f_5} & \cdots & D_{Z_8}\Psi_{f_5} \\ \vdots & \ddots & \vdots \\ D_{Z_5}\Psi_{f_8} & \cdots & D_{Z_8}\Psi_{f_8} \end{pmatrix}, \quad (49)$$

$$D_{\mathbf{H}_{\text{in}}}\Psi_{r_{\text{in}}} = \begin{pmatrix} D_{H_1}\Psi_{r_1} & \cdots & D_{H_4}\Psi_{r_1} \\ \vdots & \ddots & \vdots \\ D_{H_1}\Psi_{r_4} & \cdots & D_{H_4}\Psi_{r_4} \end{pmatrix}, \quad (50)$$

$$D_{\mathbf{Z}_{\text{in}}}\Psi_{r_{\text{in}}} = \begin{pmatrix} D_{Z_1}\Psi_{r_1} & \cdots & D_{Z_4}\Psi_{r_1} \\ \vdots & \ddots & \vdots \\ D_{Z_1}\Psi_{r_4} & \cdots & D_{Z_4}\Psi_{r_4} \end{pmatrix}, \quad (51)$$

$$D_{\mathbf{H}_{\text{out}}}\Psi_{r_{\text{out}}} = \begin{pmatrix} D_{H_5}\Psi_{r_5} & \cdots & D_{H_8}\Psi_{r_5} \\ \vdots & \ddots & \vdots \\ D_{H_5}\Psi_{r_8} & \cdots & D_{H_8}\Psi_{r_8} \end{pmatrix}, \quad (52)$$

$$D_{\mathbf{Z}_{\text{out}}}\Psi_{r_{\text{out}}} = \begin{pmatrix} D_{Z_5}\Psi_{r_5} & \cdots & D_{Z_8}\Psi_{r_5} \\ \vdots & \ddots & \vdots \\ D_{Z_5}\Psi_{r_8} & \cdots & D_{Z_8}\Psi_{r_8} \end{pmatrix}. \quad (53)$$

Therefore, the Lyapunov spectrum to the direction inside and outside of (11335577) can simply be calculated from the corresponding nonzero Jacobian elements. Let

$$D_{\text{in}}\Psi_{\text{in}} = \begin{pmatrix} D_{\mathbf{H}_{\text{in}}}\Psi_{f_{\text{in}}} & D_{\mathbf{Z}_{\text{in}}}\Psi_{f_{\text{in}}} \\ D_{\mathbf{H}_{\text{in}}}\Psi_{r_{\text{in}}} & D_{\mathbf{Z}_{\text{in}}}\Psi_{r_{\text{in}}} \end{pmatrix} \quad (54)$$

and

$$D_{\text{out}}\Psi_{\text{out}} = \begin{pmatrix} D_{\mathbf{H}_{\text{out}}}\Psi_{f_{\text{out}}} & D_{\mathbf{Z}_{\text{out}}}\Psi_{f_{\text{out}}} \\ D_{\mathbf{H}_{\text{out}}}\Psi_{r_{\text{out}}} & D_{\mathbf{Z}_{\text{out}}}\Psi_{r_{\text{out}}} \end{pmatrix}. \quad (55)$$

Then the Lyapunov spectrum to the directions inside of (11335577) and outside of (11335577) can be obtained separately from the time series of $D_{\text{in}}\Psi_{\text{in}}$ and $D_{\text{out}}\Psi_{\text{out}}$, with a classical method such as QR decomposition.

In the same way, we can also decompose the Lyapunov spectrum into the directions inside of (11115555), outside of (11115555) in (11335577), and outside of (11335577). To avoid redundancy, we only give here the definition of the coordinates $(\mathbf{H}', \mathbf{Z}') = (H'_1, \ldots, H'_8, Z'_1, \ldots, Z'_8)^T$ which separate the elements of Jacobian $D\Psi(\mathbf{H}', \mathbf{Z}')$ according to the hierarchy.

$$(\mathbf{H}', \mathbf{Z}') = A' A(\eta, \zeta), \tag{56}$$

where

$$A' = (a'_{ij}) \tag{57}$$

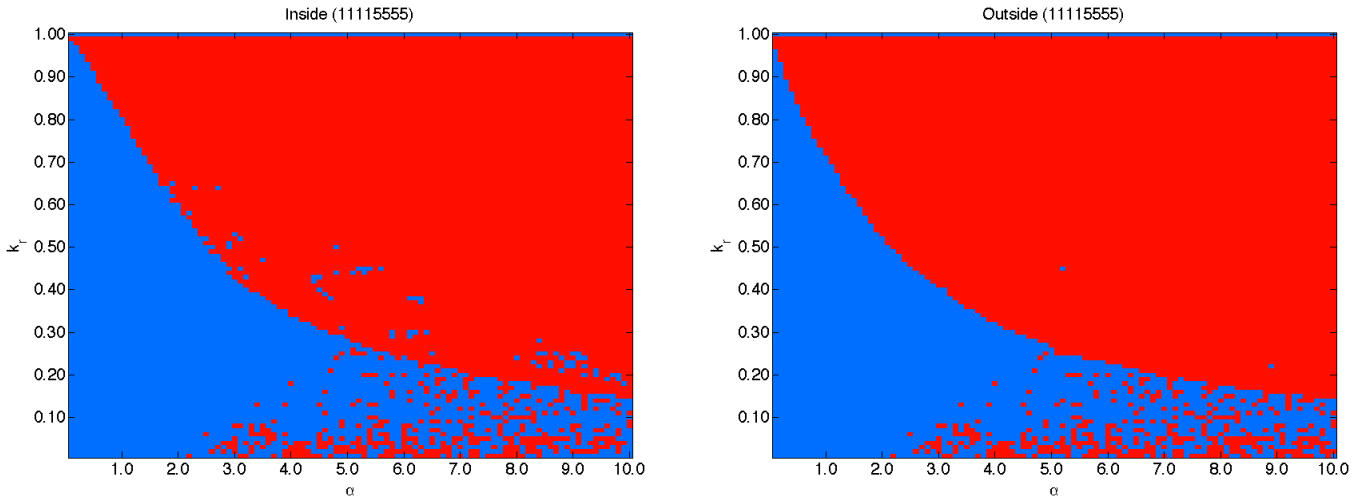
$$= \begin{pmatrix} \cos\left(\frac{\pi}{4}\right) & \sin\left(\frac{\pi}{4}\right) & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \cos\left(\frac{\pi}{4}\right) & \sin\left(\frac{\pi}{4}\right) & 0 & 0 & 0 & 0 \\ -\cos\left(\frac{\pi}{4}\right) & \sin\left(\frac{\pi}{4}\right) & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -\cos\left(\frac{\pi}{4}\right) & \sin\left(\frac{\pi}{4}\right) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{58}$$

In this case, $(H'_1, H'_2, Z'_1, Z'_2)^T$ are the coordinates inside of (11115555), $(H'_3, H'_4, Z'_3, Z'_4)^T$ are outside of (11115555), and $(H'_5, \ldots, H'_8, Z'_5, \ldots, Z'_8)^T$ are outside of (11335577). The last coordinates $(H'_5, \ldots, H'_8, Z'_5, \ldots, Z'_8)^T$ coincide with $(H_5, \ldots, H_8, Z_5, \ldots, Z_8)^T$.

The result of calculation of the maximum Lyapunov exponents in the decomposed directions are shown in Figs. 6 and 7. This enables us to verify the existence of chaos in each hierarchy of the invariant subspace. To investigate the possible role of chaos in this model, chaotic itinerancy between patterns A $\rightleftharpoons$ B and C $\rightleftharpoons$ D should have at least positive maximum Lyapunov exponent inside of (11335577).

2.6. *Temporal dynamics of chaotic itinerancy state*

Among the possible regions of chaotic itinerancy state, we investigate the temporal dynamics on specific refractory parameters, $k_r = 0.4$ and $\alpha = 5.0$. In these parameters, there exists a chaotic attractive set inside of (11115555) and (11333311) including the patterns A, B and C, D, respectively (Fig. 8), and the maximum Lyapunov exponents to the direction inside and outside of the (11115555) and (11333311) are positive (Fig. 6). The transversal instability of (11115555) and (11333311) to the direction inside of (11335577) implies these subspaces went through the blowout bifurcation. The temporal dynamics of the maximum moment Lyapunov exponents inside and outside of (11115555) and (11333311) are also dominantly positive (Fig. 10), and the orbits inside of (11335577) fall neither inside of (11115555) nor (11333311) at least for a quite long period (we tested the case until $t = 1000\,000\,000$). The orbits starting from almost all points of (12345678) are attracted inside of (11335577), which reflect the transversal stability of (11335577) (Fig. 7). The orbits inside of (11335577) irregularly visit the patterns A, B, C, D, including partially preserved temporal sequences of two-periodic cycles A $\rightleftharpoons$ B and C $\rightleftharpoons$ D (Fig. 8), and the maximum Lyapunov exponent and moment Lyapunov exponent are positive (Figs. 7 and 9).

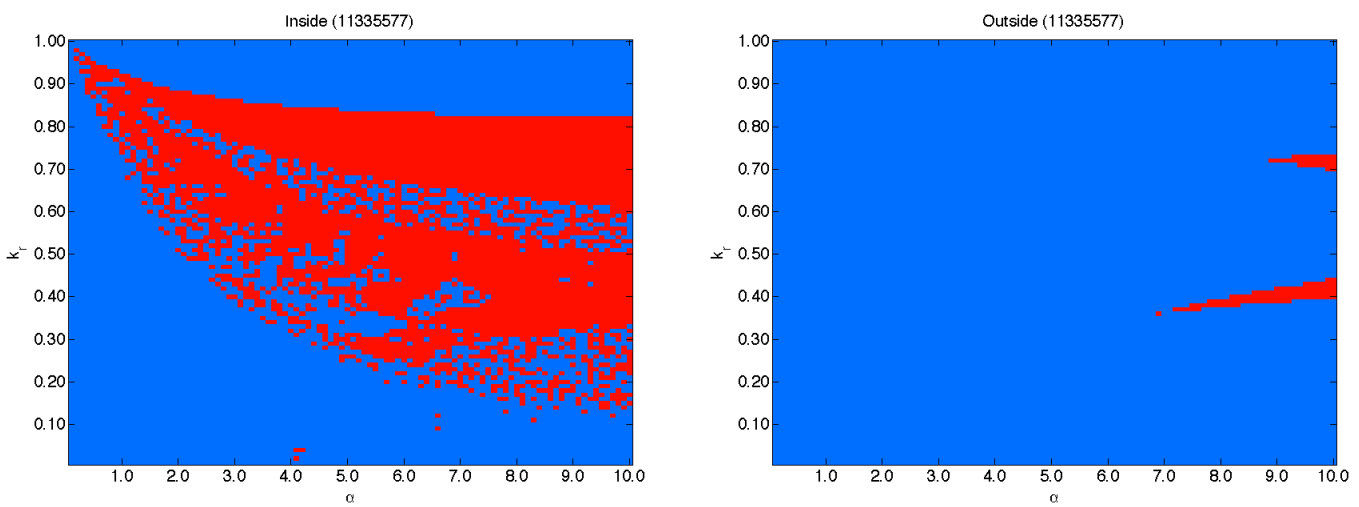


(a) (b)

Fig. 6. Positive/negative sign of the maximum Lyapunov exponents in (11115555). (a) Directions inside of (11115555) and (b) outside of (11115555). Vertical axis: k_r. Horizontal axis: α. Initial conditions were taken randomly inside of (11115555). The blue region is where the maximum Lyapunov exponent is negative, while the red region is positive. After cutting 5000 transient, 1000 steps were used for calculation at each k_r and α. The values of k_r and α were taken for each 0.01 and 0.1 step, respectively.

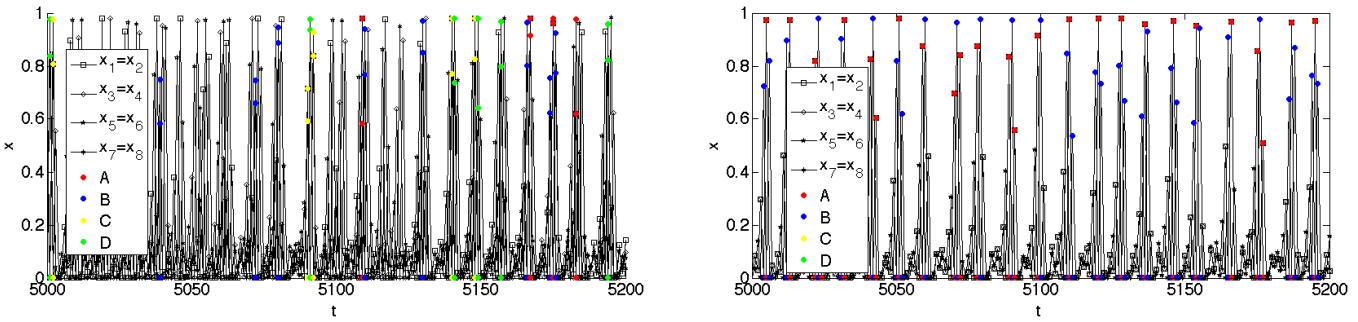
This supports the existence of chaos by the positive maximum Lyapunov exponent in this chaotic itinerancy state.

The loss of stability to the transversal directions of (11115555) and (11333311) can be considered as the blowout bifurcation. It has been widely reported in continuous time systems that the blowout bifurcations of plural attractors are associated with the riddled structure of the attractors' basins [Ott & Sommere, 1994; Ott *et al.*, 1995]. In this CNN, there exist two attractors $\mathrm{A} \rightleftharpoons \mathrm{B}$ and $\mathrm{C} \rightleftharpoons \mathrm{D}$ that are destabilized to their transversal directions. To reveal whether there exists similar phenomenon in the chaotic itinerancy state, we investigated the basin structure of CNN inside of (11335577). Since the dynamics seems to settle on neither of the patterns $\mathrm{A} \rightleftharpoons \mathrm{B}$ nor $\mathrm{C} \rightleftharpoons \mathrm{D}$, these attractors are considered to have gone through



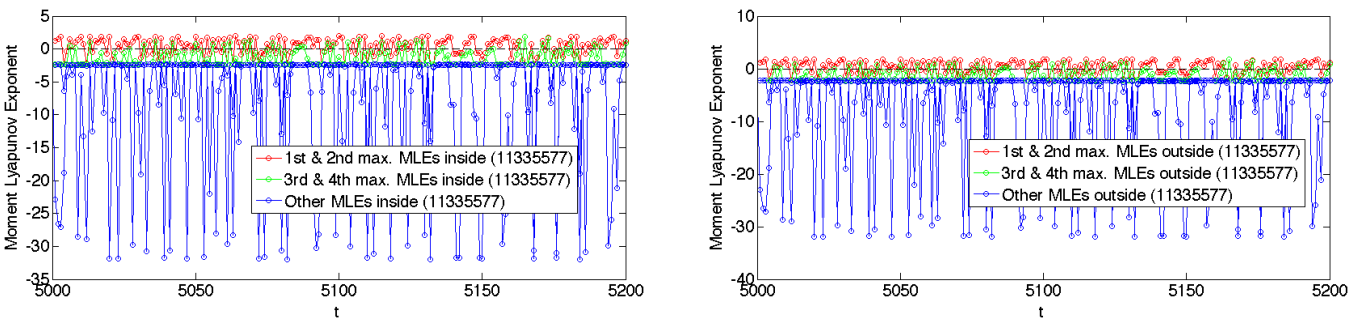
(a) (b)

Fig. 7. Positive/negative sign of the maximum Lyapunov exponents in (11335577). (a) Inside of (11335577) and (b) outside of (11335577). Vertical axis: k_r. Horizontal axis: α. Initial conditions were taken randomly inside of (11335577). The blue region is where the maximum Lyapunov exponent is negative, while the red region is positive. After cutting 5000 transient, 1000 steps were used for calculation at each k_r and α. The values of k_r and α were taken for each 0.01 and 0.1 step, respectively.



(a) (b)

Fig. 8. Temporal dynamics of CNN. (a) Initial condition inside of (11335577) and (b) initial condition inside of (11115555). Horizontal axis: Time step t. Vertical axis: CNN output $\mathbf{x}(t)$. The color dots representing the patterns A, B, C, D are superimposed on the value of $\mathbf{x}(t)$ when the network retrieved them. $k_r = 0.4$, $\alpha = 5.0$.

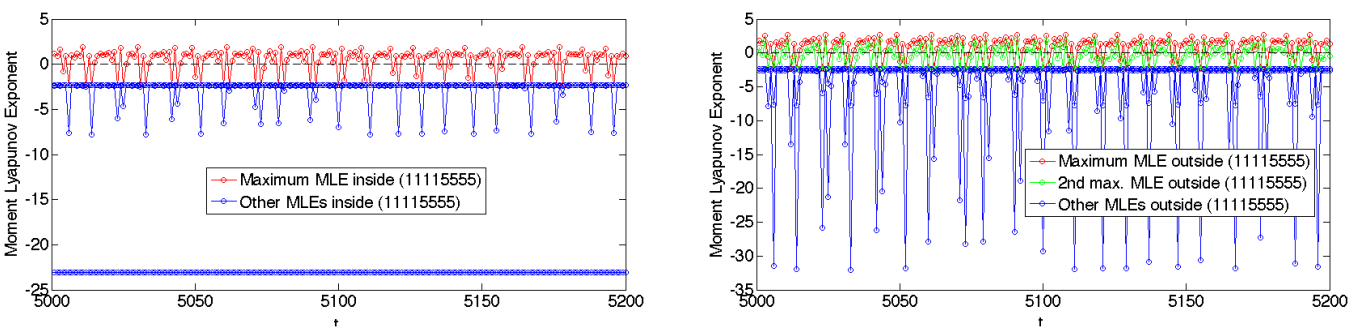


(a) (b)

Fig. 9. Temporal dynamics of moment Lyapunov exponents with initial condition inside of (11335577). (a) Moment Lyapunov exponents to the direction inside of (11335577) and (b) moment Lyapunov exponents to the direction outside of (11335577). Horizontal axis: Time step t. Vertical axis: Moment Lyapunov exponent. $k_r = 0.4$, $\alpha = 5.0$.

the transversal crisis, and became the attractor ruins with attractive basins of positive measure. We judged the basins according to the first retrieval of the patterns A, B or C, D, with respect to the values of $h_i(t)$ $(1 \leq i \leq 8)$.

The result shows a fractal boundary structure of two basins (Fig. 11). The spatially complex mixture strongly supports the occurring mechanism of the intermittent transition between two attractor ruins. Due to the fractal boundary structure and



(a) (b)

Fig. 10. Temporal dynamics of moment Lyapunov exponents with initial condition inside of (11115555). (a) Moment Lyapunov exponents to the direction inside of (11115555) and (b) moment Lyapunov exponents to the direction outside of (11115555). Horizontal axis: Time step t. Vertical axis: Moment Lyapunov exponent. $k_r = 0.4$, $\alpha = 5.0$.

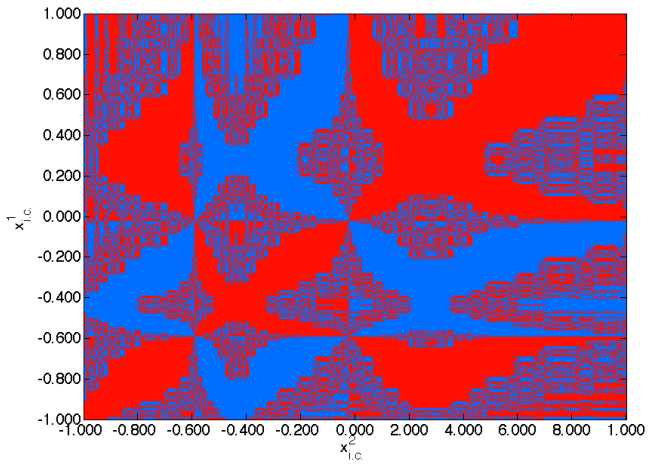


Fig. 11. Basins structure of CNN. Horizontal axis: Value of $x^2_{i.c.}$. Vertical axis: Value of $x^1_{i.c.}$. Initial condition were taken inside (11335577) and close to (11331133) as $(\eta_1, \ldots, \eta_8)^T = (\zeta_1, \ldots, \zeta_8)^T = (x^1_{i.c.} + \delta, x^1_{i.c.} + \delta, x^2_{i.c.} + \delta, x^2_{i.c.} + \delta, x^1_{i.c.} - \delta, x^1_{i.c.} - \delta, x^2_{i.c.} - \delta, x^2_{i.c.} - \delta)^T$ with $\delta = 1.0\text{e}{-6}$. The red region is where the orbits reach the neighbor of the patterns A $\rightleftharpoons$ B first, while the blue region reaches the neighbor of the patterns C $\rightleftharpoons$ D first. $x^1_{i.c.}$ and $x^2_{i.c.}$ were taken for each 0.002 step with parameters $k_r = 0.4$ and $\alpha = 5.0$.

orbital instability, each orbit is expected to show a unique visiting sequence order and residence time to the attractor ruins, which becomes distinguishable from others in a short time period. Note that a further question is whether the blowout bifurcation associated with riddled basins is the general occurring mechanism of chaotic itinerancy.

Similar situation is also investigated in other stochastic and dynamical system, with the interest of defining "*emergence*" with respect to the temporal development of probability distribution of the system variables, namely space-time phase [MacKay, 2008, 2005]. MacKay defined the emergence as the existence of plural states of convergence for space-time phase, with weak/strong distinction. The converged space-time phase corresponds to the SRB measure on an attractor with measure positive attractive basin in dynamical system. "*Weak emergence*" is defined as due to the trivial topological reason of the system, such as the existence of simple separatrix between plural attractors in dynamical system. The coexistence of stable attractors in CNN at low refractoriness parameters region corresponds to such a case. While "*strong emergence*" is referred to as the sensitivity of converged space-time phase to the initial condition, where tiny perturbation at the beginning can lead to a different phase. This is similar to the riddled basin situation at the bifurcation point in CNN. Furthermore, MacKay mentioned a more complex class of dynamics where space-time phase never settles down (e.g. biological evolution, gravitational systems, aggregation and coarsening models). As counterpart to the dynamical system, chaotic itinerancy in CNN is an example of such a class of complex system which goes beyond the definition of strong emergence.

3. Synthetic Modeling of Autonomous Learning

Based on the chaotic itinerancy state we have investigated in the above section with refractoriness parameters $k_r = 0.4$ and $\alpha = 5.0$, we add and simulate the effect of autonomous learning. The combination of chaotic itinerancy in CNN with intrinsic learning rule is proposed as a dynamical model of developmental brain activity such as the complexification process of birdsong syntax [Funabashi & Aihara, 2007]. We chose two basic ways of autonomous learning that are expected to exist widely and generally in neuronal activity: Hebbian learning and the spike-timing-dependent plasticity (STDP) rule [Hebb, 1949; Markram *et al.*, 1997]. Naturally, actual learning dynamics of *in situ* neurons have their own specificity and variation, and cannot be reduced into simple rules with a small number of parameters. What we investigate here is the possible role of mathematically simplified form of autonomous learning during chaotic itinerancy. This is not necessarily the modeling of experimentally observed neural dynamics, but the combination of prototypical learning rules with a biological neuron model expressing candidate dynamics of cortical transitory activity, which is accessible to mathematical analysis. In this sense, we call synthetic modeling as a part of the dynamical basis of hermeneutics in the brain [Tsuda, 1991]. In a wide sense, synthetic modeling is a part of constructive modeling in complex system sciences, but rather emphasizes on the generic and qualitative property which appears in the combination of candidate models, than the reconstruction of some observed phenomenon with quantitative reproducibility.

3.1. *Autonomous learning with Hebb's learning rule*

Hebbian learning is the most classical rule to explain the self-organization of neural network,

and is based on the "cells that fire together, wire together" principle. We implemented this rule to CNN to reinforce the connection matrix $W = (w_{ij})$ $(1 \leq i, j \leq 8)$ depending on the dynamics itself of the network. The additional Hebbian learning in CNN was defined as follows:

$$w_{ij}(t+1) = w_{ij}(t) + \epsilon(2x_i(t) - 1) \times (2x_j(t-1) - 1), \quad (59)$$

where t is the time step of CNN and ϵ is the learning coefficient. We chose $\epsilon = 0.001$ for the following simulation.

3.2. *Autonomous learning with STDP rule*

The STDP rule is known to modify the synaptic conductance depending asymmetrically on the difference of pre-synaptic and post-synaptic firing time. We defined the discrete time version of STDP rule as follows:

$$w_{ij}(t+1) = w_{ij}(t) + \Delta w_{ij}(t), \quad (60)$$

$$\Delta w_{ij}(t) = g_{ij}(t) - g_{ji}(t), \quad (61)$$

$$g_{ij}(t) = A \sum_{d=1}^{d_f} k_{\text{STDP}}^{d-1} \text{ AND } (x_i(t), x_j(t-d)), \quad (62)$$

where A is the norm of learning coefficient, and k_{STDP} controls the exponential decrease of learning effect with respect to the spike timing difference d. We use $A = 1$ and $k_{\text{STDP}} = 0.1$ for the following simulation. To implement the "nearest neighbor model" of STDP learning between a single pair of neuron spikes, d_f is defined at each time step and by the neuron in relation to the past neuron activity, so that

$$x_j(t-d) < 0.5 \quad (1 \leq d < d_f), \quad (63)$$

$$x_j(t-d_f) \geq 0.5, \quad (64)$$

hold. In actual learning dynamics of CNN with STDP rule, neuron spikes are generally sparse and past influence quickly decays in exponential. For the given parameters, it is approximately equal to set $d_f = \infty$. The AND$(\cdot, \cdot)$ is the extended Boolean expression defined as follows:

$$\text{AND}(x_i(t), x_j(t-d)) = h_i(t)h_j(t-d)x_i(t)x_j(t-d). \quad (65)$$

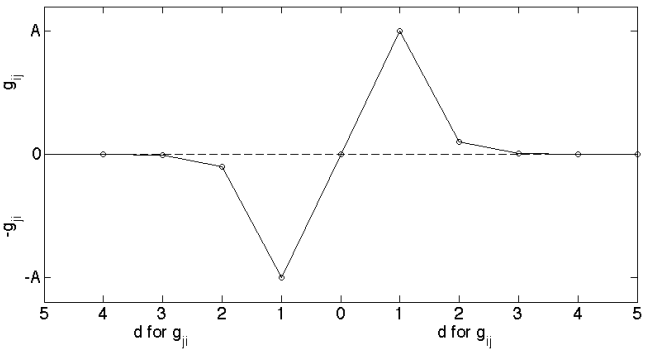


Fig. 12. STDP rule for discrete time CNN. Horizontal axis: Spike time interval d. Vertical axis: Modification to w_{ij} defined in Eq. (62). $k_{\text{STDP}} = 0.1$.

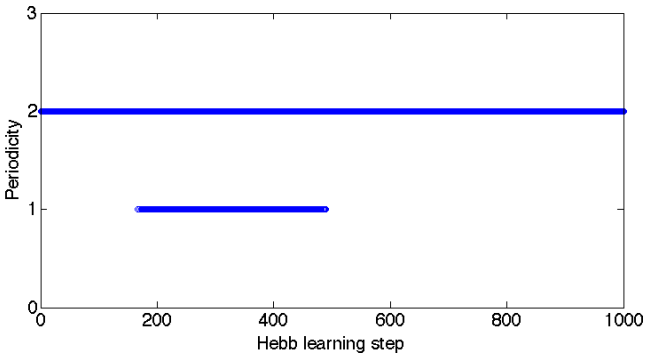
This means the function returns the value $x_i(t) \times x_j(t-d)$ only when there exist spikes $x_i(t) \geq 0.5$ and $x_j(t-d) \geq 0.5$ of interval d. The defined STDP rule is depicted in Fig. 12.

4. Simulation Result and Discussion

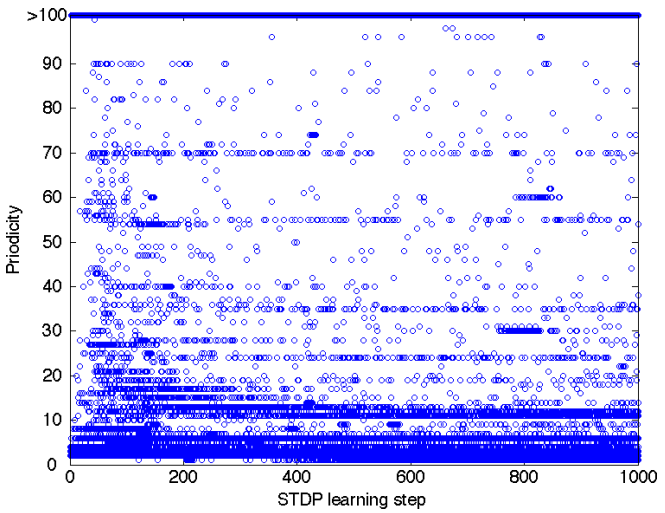
The simulation is performed by adding Hebbian or STDP rule on the chaotic itinerancy state with the parameters $k_r = 0.4$ and $\alpha = 5.0$, over an empirical range of the learning time steps. After the learning, the refractoriness parameters were set to $k_r = \alpha = 0$ to investigate the memory structure. Note that such transition between chaotic wandering state and ordered state is also reported in EEG data [Skarda & Freeman, 1987]. We summarize here the observed modifications of the dynamics.

4.1. *Periodicity of emergent attractors*

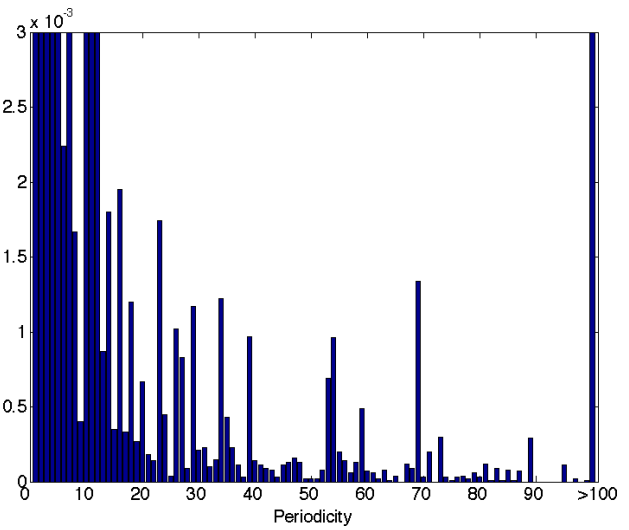
Both Hebbian learning and STDP rules brought the network novel attractors, based on the patterns A, B, C, D and transitional chaotic dynamics. We investigated the periodicity of the emergent attractors, which is shown in Fig. 13(a). Hebbian learning did not change the periodicity and the modified attractors are always two-periodic, except in the divergent case where the only attractor is the fixed point $\mathbf{x} = (1, 1, 1, 1, 1, 1, 1, 1)^T$. As for STDP learning, the periodicity of newly emerged attractors varies in quite a wide range [Fig. 13(b)]. There exist orbits with even more than 10 000 period, which strongly implies the synthesis of strange attractors. Almost all periodicity between 1 to 100 are observed. The histogram of the obtained periodicity is shown in Fig. 13(c).



(a)



(b)



(c)

Fig. 13. Periodicity of attractor after autonomous learning. (a) Hebbian learning. Horizontal axis: Time step of Hebb learning. Vertical axis: Periodicity of attractor after the learning, (b) STDP learning. Horizontal axis: Time step of STDP learning. Vertical axis: Periodicity of attractor after the learning and (c) histogram of periodicity after STDP learning. Horizontal axis: Periodicity of attractor after the learning. Vertical axis: Percentage (values more than 3.0e−3 are cut off). The periodicity after the learning was judged by resetting $k_r = \alpha = 0$ and cutting 5000 transient. 100 trials are superimposed for each learning step. Initial conditions were taken randomly.

Such difference of results between Hebbian and STDP rules can be considered to relate the temporal symmetry of the learning rules. STDP rule has the temporal asymmetry which is not expressed in Hebbian rule. The effect of the past inputs with more than one time step interval may also bring variation to periodicity.

4.2. *Spatial configuration of emergent attractors*

The two-periodic patterns that emerged from Hebbian learning consists of two kinds; the former attractor $A \rightleftharpoons B$, $C \rightleftharpoons D$ or the novel patterns which are different from any of A, B, C, D. The cases of $A \rightleftharpoons B$ and $C \rightleftharpoons D$ are investigated in the following section. We examine here the spatial configuration of the novel patterns, which are depicted in Fig. 14.

The newly emerged patterns are also limited to two kinds, and they are both situated at equidistance points from $A \rightleftharpoons B$ and $C \rightleftharpoons D$. Furthermore, these patterns are situated in the invariant subspace (11335577), which is the minimum union of the (11115555) and (11333311) where former attractors were situated. This geometry implies that the novel patterns are in a sense the integration of the attractors $A \rightleftharpoons B$ and $C \rightleftharpoons D$ in the smallest subspace preserving their periodicity. This is interesting as an analogical process of dialectic, where the

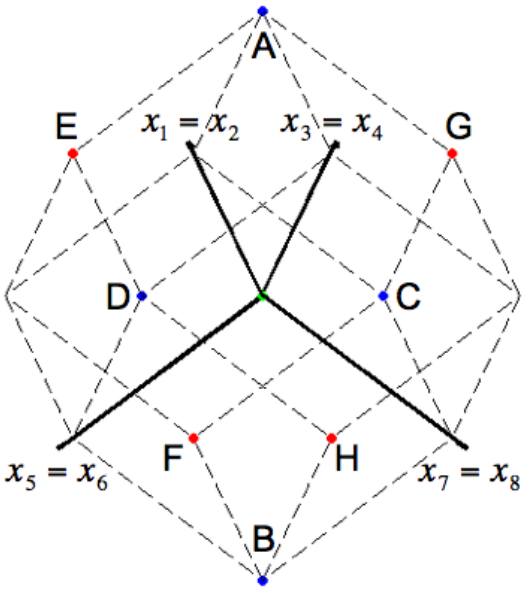


Fig. 14. Possible emergent attractors after Hebb learning. The arrows are four-dimensional coordinates $(x_1 = x_2, x_3 = x_4, x_5 = x_6, x_7 = x_8)^T$ of the CNN output inside of (11335577). The blue points are memorized patterns A, B, C, D. The red points E, F, G, H are possible emergent two-periodic attractors, and are situated at the equidistance points from A $\rightleftharpoons$ B and C $\rightleftharpoons$ D. Only two-periodic patterns E $\rightleftharpoons$ F and G $\rightleftharpoons$ H are possible to emerge. The green point is the divergent case that corresponds to the fixed point $(1, 1, 1, 1)^T$.

opposing two theses find the way to be sublated not by logically solving the contradiction but by changing the way of asking itself. Here, we used the analogy of chaotic itinerancy state as the contemplation process comparing two symmetric patterns [Tsuda, 1991, 2004], and the result of autonomous learning as the formation of an sublated idea. Although it is a quite simplified realization, this may lead to another adaptive computing principle with neural network, where neuronal plasticity plays an essential role to produce internal complexity of the system dynamics to cope with the diversification of external environment. Dynamical acquisition of olfactory memories is a classical example observed in experiment [Skarda & Freeman, 1987]. This perspective is further discussed in a later section, in relation to the formal logic in topological psychology. On the other hand, computing with dynamical systems such as associative memory generally utilizes the retrieval process to classify the input patterns according to the memory structure, and does not consider the modification of the memory itself in interaction with the retrieval dynamics.

Although the possible emergent patterns are limited with Hebbian learning, its reproducibility requires a quite high numerical precision of initial condition down to the order of 1.0e−17 (in case of 350 learning steps). This may be a reflection of the observed riddled basins structure to the learning dynamics and orbital instability, since the global behaviors of the orbits differ being highly sensitive on the initial conditions.

The emerged attractors from STDP rule are more suppressive, and in most cases express the output patterns $\mathbf{h}(t) = (0, 0, 0, 0, 0, 0, 0, 0)^T$ and a few patterns with only a single pair of neurons firing inside (11335577) according to the varying periodicity, such as $\mathbf{h}(t) = (1, 1, 0, 0, 0, 0, 0, 0)^T$. The sensitive dependence on initial conditions was also observed numerically in the same order as Hebbian learning.

4.3. *Modification of invariant subspaces*

Since the connection matrix W is modified by autonomous learning, the structure of the invariant subspaces reflecting its symmetry also changes temporally. In both Hebbian and STDP learning, the remaining invariant subspaces in rigorous definition are only those in the above hierarchy of (11335577) [Fig. 2(top)]. This is due to the loss of the other synchronizations in the orbits. Since chaotic itinerancy takes place in (11335577), four pairs of neurons are always synchronized regardless of its irregularity. Therefore, although all values of connection matrix are being modified, the symmetry of the connection matrix supporting the above hierarchy of (11335577) is completely maintained, while other invariant subspaces are destroyed by the chaotic irregularity.

Since (11115555) and (11333311) are situated below (11335577), the attractors A $\rightleftharpoons$ B and C $\rightleftharpoons$ D are destabilized along with the destruction of the invariant subspaces [Fig. 2(bottom)]. The invariant subspaces which are neither above nor below (11335577) also possess an attractor inside (11111111). Although these are measure zero subspaces with respect to the whole state space, their dimensions vary from one to seven. These can also be considered as a part of memory structure. As the invariant subspaces are trimmed leaving only the structure used in retrieval process, the observed modification can be interpreted as the simplification

or refinement of the memory. Hence, the selective destruction and conservation of invariant subspaces are impossible to realize with stochastic noise. Such utility of chaos lies in its completely deterministic property.

Although the rigorous invariant subspaces other than above (11335577) are instantaneously destroyed by the autonomous learning, approximate invariant subspaces shows certain dynamics which reflects the nature of dynamics. We consider the case of Hebbian learning in which the competition between two attractors $\mathrm{A} \rightleftharpoons \mathrm{B}$ and $\mathrm{C} \rightleftharpoons \mathrm{D}$ takes place. We define the approximate invariant subspaces by admitting certain tolerance to the condition (11), and investigate the asymmetric convergence to one attractor.

Table 2 shows an example with $k_r = 0.4$ and $\alpha = 4.0$. In this case, the number of the approximate invariant subspaces is oscillating with respect to the time steps of Hebbian learning. The dynamics finally settled to the pattern $\mathrm{C} \rightleftharpoons \mathrm{D}$ after more than 10 000 learning steps. This dynamics implies that certain invariant subspaces are repeatedly destroyed and recovering during the learning. The prevalence of certain invariant subspaces is assumed to relate to the localization of orbit between two attractors. The oscillation of the number of invariant subspaces continues until one basin structure becomes completely dominant and fixed. This can be analogically interpreted as the decision-making process, where one is supposed to choose only one side from the two possibilities. The chaotic itinerancy state is again the analogy of contemplation process comparing two patterns. The final settlement to the pattern $\mathrm{C} \rightleftharpoons \mathrm{D}$ that drove out the pattern $\mathrm{A} \rightleftharpoons \mathrm{B}$ corresponds to the decision-making.

Table 2. Oscillation of the number of invariant subspaces in case of the asymmetric convergence to one attractor $\mathrm{C} \rightleftharpoons \mathrm{D}$. Invariant subspaces were calculated for each 100 time steps with the allowance of 0.01. $k_r = 0.4$, $\alpha = 4.0$. Only the relative differences with respect to the past time step are listed.

Number of Invariant Subspaces	Time Steps of Hebbian Learning
76	0
49	100
34	200
43	1100
25	10000+

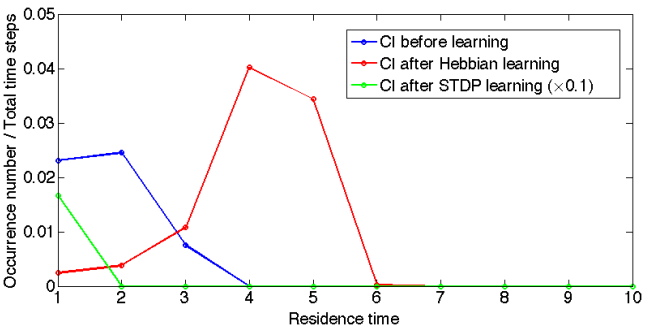


Fig. 15. Residence time distribution on memorized patterns before and after autonomous learning during chaotic itinerancy (CI). Horizontal axis: Residence time step on patterns A, B, C, D. Vertical axis: Occurrence number of each residence time divided by the total time step 10 000. The blue line is calculated from the dynamics without any modification to connection matrix W, while the red line is after adding 350 steps of Hebbian learning. The green line is after adding 350 steps of STDP learning. The distribution of STDP learning is multiplied by 0.1. Initial condition was chosen to generate the patterns $\mathrm{E} \rightleftharpoons \mathrm{F}$ (see Fig. 14) for Hebbian learning, and three-periodic patterns $\mathbf{h}(t) = (1,0,0,0)^T \to (0,0,0,0)^T \to (0,0,0,0)^T \cdots$ for STDP learning in (11335577). $k_r = 0.4$, $\alpha = 5.0$.

4.4. *Change of residence time distribution*

The defined autonomous learnings always follow the principle that the more frequently causal temporal structures appear, the more they are reinforced. The way of visiting each attractor ruin during the learning gives us the information how the self-organization of novel attractors proceeds. We investigated the residence time distributions for the typical emergent cases of the novel two-periodic attractor $\mathrm{E} \rightleftharpoons \mathrm{F}$ (see Fig. 14) in Hebbian learning and three-periodic attractor in STDP learning in Fig. 15.

Residence time distribution of the chaotic itinerancy state with $k_r = 0.4$ and $\alpha = 5.0$ shows unimodal pattern that may be approximated with gamma-distribution. Additional Hebbian learning triggers the right shift of the distribution and the overall increase of the visiting rate. This means that the more Hebbian learning proceeds, the more frequently and the longer the learning orbit visits for patterns $\mathrm{A} \rightleftharpoons \mathrm{B}$ and $\mathrm{C} \rightleftharpoons \mathrm{D}$. This can be considered as the self-organized reinforcement of attractiveness to each attractor ruin, and the learning was revealed to augment the neural correlations of memorized patterns during chaotic itinerancy. The increase of correlations may also be supported by the learning of transitional chaotic state with low

neural activity. The continuous sequences of low outputs also increase the connection weight in the defined setting, even if the simple increase of the Hebbian learning step does not lead the network to the divergent pattern.

As for STDP learning, generally the learning orbits become periodic regardless of the refractoriness, and the residence time distribution shrinks to the residence time 1. The dynamics do not stay more than one step on each pattern during learning, including no residence time on the four patterns. STDP learning on attractor ruins seems to be cancelled out with the periodicity two of the patterns $A \rightleftharpoons B$ and $C \rightleftharpoons D$, since the reinforced connections at $A \to B$ and $C \to D$ are weakened by the successive $B \to A$ and $D \to C$. On the other hand, the exponential duration of STDP rule seems also to reinforce the firing patterns other than A, B, C, D in transitional chaotic state which are generally sparse. Such isolated patterns are weakened in Hebbian learning. In chaotic dynamics, it is known that there exist countably infinite sets of unstable periodic orbits with arbitrary periodicity. It is of interest whether the STDP rule brings rich variation of periodicity by stabilizing long periodicities from chaotic orbit.

4.5. *Change of system decompositionability by Hebbian learning*

The effect of Hebbian learning during chaotic itinerancy was revealed to reinforce simultaneous neural correlation in this model. It is of interest which combinations of the neuronal synchrony arise. The augmentation of coincident firing patterns rate can be interpreted as the formation of modularity, which can lead to the functional differentiation. To investigate the emerged system modularity during Hebbian learning, we consider the time-averaged four-dimensional discrete distribution $P(y_1, y_2, y_3, y_4)$ of binary outputs $\mathbf{h}(t)$ in (11335577) as

$$\begin{aligned} &P(y_1, y_2, y_3, y_4) \\ &= \frac{1}{T}\sum_{t=1}^{T} \delta(h_1(t) = h_2(t), y_1)\delta(h_3(t) = h_4(t), y_2) \\ &\quad \cdot \delta(h_5(t) = h_6(t), y_3)\delta(h_7(t) = h_8(t), y_4), \end{aligned} \tag{66}$$

where $\delta(\cdot, \cdot)$ is the delta function, T is the temporal average window and $y_i(t) \in \{0, 1\}$ $(1 \le i \le 4)$.

To measure the statistical modularity expressed as correlation among (y_1, y_2, y_3, y_4), we define the decomposed distribution $P^{\text{dec}}(y_1, y_2, y_3, y_4)$ as the product of two joint distributions $P(\mathbf{y}^1)$ and $P(\mathbf{y}^2)$ of the decomposed subsystems $\mathbf{y}^1$ and $\mathbf{y}^2$.

$$P^{\text{dec}}(y_1, y_2, y_3, y_4) = P(\mathbf{y}^1)P(\mathbf{y}^2), \tag{67}$$

where

$$\begin{aligned} &\mathbf{y}^1 = (y_{i_1}, \ldots, y_{i_k}), \\ &\qquad \{i_1, \ldots, i_k\} \subset \{1, \ldots, 4\}, \quad k \in \{1, 2, 3\}, \end{aligned} \tag{68}$$

$$\begin{aligned} &\mathbf{y}^2 = (y_{j_1}, \ldots, y_{j_{4-k}}), \\ &\qquad \{i_1, \ldots, i_k\} \not\supset \{j_1, \ldots, j_{4-k}\} \subset \{1, \ldots, 4\}. \end{aligned} \tag{69}$$

Note that the system decomposition is not limited to two parts, and there exist up to the four subsystems in (11335577) which can be easily generalized such as

$$P^{\text{dec}}(y_1, y_2, y_3, y_4) = P(\mathbf{y}^1)P(\mathbf{y}^2)P(\mathbf{y}^3), \tag{70}$$

$$\mathbf{y}^1 = (y_{i_1}, y_{i_2}), \quad \{i_1, i_2\} \subset \{1, \ldots, 4\}, \tag{71}$$

$$\mathbf{y}^2 = (y_j), \quad \{i_1, i_2\} \not\ni \{j\} \in \{1, \ldots, 4\}, \tag{72}$$

$$\mathbf{y}^3 = (y_k), \quad \{i_1, i_2, j\} \not\ni \{k\} \in \{1, \ldots, 4\}, \tag{73}$$

for three subsystems, and

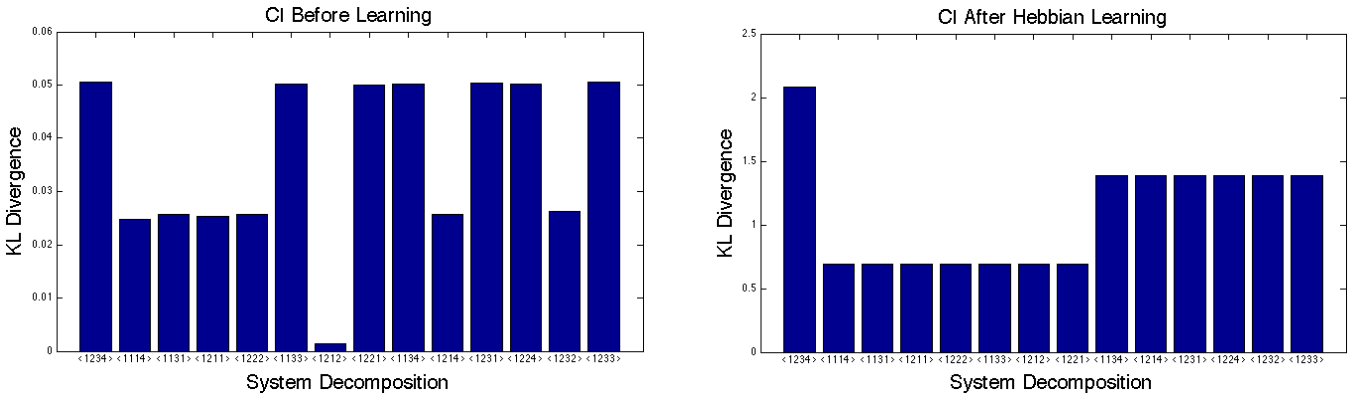
$$\begin{aligned} P^{\text{dec}}(y_1, y_2, y_3, y_4) &= P(\mathbf{y}^1)P(\mathbf{y}^2)P(\mathbf{y}^3)P(\mathbf{y}^4) \\ &= P(y_1)P(y_2)P(y_3)P(y_4), \end{aligned} \tag{74}$$

for four subsystems. In each decomposed distribution $P^{\text{dec}}(y_1, y_2, y_3, y_4)$, only correlations between the defined subsystems are set to be independent, and the correlations inside of each subsystem are preserved.

We measured the decompositionability of the system $P(y_1, y_2, y_3, y_4)$ into $P^{\text{dec}}(y_1, y_2, y_3, y_4)$ with Kullback–Leibler (KL) divergence $D[\cdot : \cdot]$ as follows.

$$\begin{aligned} &D[P(y_1, y_2, y_3, y_4) : P^{\text{dec}}(y_1, y_2, y_3, y_4)] \\ &\quad = \sum_{y_1, y_2, y_3, y_4 \in \{0,1\}} P(y_1, y_2, y_3, y_4) \\ &\qquad \times \log \frac{P(y_1, y_2, y_3, y_4)}{P^{\text{dec}}(y_1, y_2, y_3, y_4)}. \end{aligned} \tag{75}$$

This means that if the value of $D[P(y_1, y_2, y_3, y_4) : P^{\text{dec}}(y_1, y_2, y_3, y_4)]$ is large, the system loses much



(a) (b)

Fig. 16. KL divergence of the system decompositions. (a) Chaotic itinerancy (CI) state before learning and (b) chaotic itinerancy (CI) state after 350 steps of Hebbian learning with learning coefficient $\epsilon = 0.01$. Horizontal axis: $\langle \cdots \rangle$ represent system decompositions in which correlations between different numbers of elements are set to be independent. e.g. $\langle 1212 \rangle$ corresponds to $P^{\mathrm{dec}}(y_1, y_2, y_3, y_4) = P(\mathbf{y}^1)P(\mathbf{y}^2)$ where $\mathbf{y}^1 = (y_1, y_3)$ and $\mathbf{y}^2 = (y_2, y_4)$. Vertical axis: KL divergence $D[P : P^{\mathrm{dec}}]$. $k_r = 0.4$, $\alpha = 5.0$, $T = 10\,000$ were used for the calculation.

information by the decomposition, therefore, there exist high modularity on the nullified interactions. Note that the total system decompositionability $D[P : P^{\mathrm{dec}} = P(y_1)P(y_2)P(y_3)P(y_4)]$ is referred to as a complexity measure in several studies [MacKay, 2008; Tononi *et al.*, 1994; Ay *et al.*, 2006].

The calculation result of KL divergence for each system decomposition during chaotic itinerancy after 350 steps of Hebbian learning is shown in Fig. 16. The parameters were chosen to realize the emergence of the novel two-periodic attractor $\mathrm{E} \rightleftharpoons \mathrm{F}$ (see Fig. 14) after the learning. Before adding the learning, the chaotic itinerancy state showed modularity in $\mathbf{y} = (y_1, y_3)$ and $\mathbf{y} = (y_2, y_4)$ because cutting between them does not lose much information measured with KL divergence. As the Hebbian learning proceeded, the KL divergences strongly augmented, meaning the system increased the corresponding statistical coherence. After 350 steps of learning, chaotic itinerancy state showed three kinds of uniform KL divergence values proportional to the number of decomposed subsystems. Interestingly, the relation

$$\begin{aligned} &\frac{1}{4-1} D[P : P^{\mathrm{dec}} = P(y_1)P(y_2)P(y_3)P(y_4)] \\ &\quad = \frac{1}{3-1} D[P : P^{\mathrm{dec}} = P(\mathbf{y}^1)P(\mathbf{y}^2)P(\mathbf{y}^3)] \\ &\quad = \frac{1}{2-1} D[P : P^{\mathrm{dec}} = P(\mathbf{y}^1)P(\mathbf{y}^2)], \end{aligned} \tag{76}$$

holds. This relation stands for the fact that the system loses the same amount of information when isolating any arbitrary subsystem. This means that the neuronal synchrony rate augmented in the way that the degrees of interactions became identical among all elements. In such distribution, there exists no easiest decomposition [Funabashi, 2014]. In other words, the existing modularity of the dynamics is integrated in the total coherence of the system after the learning.

In the integrated information theory of consciousness (IIT), the system decomposition is axiomatically defined as the basic element of the integrated information [Tononi, 2004; Oizumi *et al.*, 2014]. In this interpretation, the Hebbian learning increased the "level of consciousness" of CNN dynamics as postulated in IIT.

4.6. *Computational rationale of autonomous learning in relation to topological psychology*

We finally revisit the analogy with the dialectic of the emerged attractors during the autonomous learning, in relation to a formal logical description of dialectic in topological psychology. We aim here to investigate the computational rationale of autonomous learning in view of realization of a system which incorporates dialectical information processing.

Usually dialectic describes one's mental process which wanders around contradictory experience or thoughts, meanwhile creating a solution which trivializes the cognitive dissonance to some extent. Such

dynamical view on mental activity was also one of the connectionist's motivations concerning the versatility and generativity of a cognitive system. Neural network model based on parallel distributed information processing was proposed analogous to the physiological structure of neural systems, with the expectation to realize certain aspect of our mental activity missing in the classicist's way of serial-processing of symbolic computation [Morra *et al.*, 2007].

Indeed, several authors have proposed connectionist models of dialectical psychology with dynamical system, in both on-going mental process [Fischer & Bidell, 2006; Fischer & Kennedy, 1997; Lewis, 2000, 2001; Pascual-Leone, 1970] and long-term developmental process [Demetriou & Raftopoulos, 2004].

Our model with CNN also finds common interest with this stream, but gives distinguished emphasis on chaotic dynamics with neurophysiological basis. The representation of memory structure is also particularly grounded on intermittent chaotic dynamics in CNN called chaotic itinerancy, in which the interior crisis of attractors leads to the dynamical transition between plural attractor ruins. The analogy of chaotic itinerancy with the volatile aspect of mental process was assigned by Tsuda and insisted to be a core concept of hermeneutics and mathematical model of dynamical brain activity [Tsuda & Fujii, 2004; Tsuda, 1991, 2004].

Acknowledging the background of connectionist's interest on dialectic, we go back to the relation between our model and topological psychology. One of the classical philosophical descriptions of dialectic with particular emphasis on the contradiction process was developed by celebrated philosopher Hegel [1970]. It was further adapted in the context of Piagetian theory of cognitive development in dialectical psychology by Riegel [1973]. Riegel's psychological interpretation of Hegel's dialectic are summarized by Hoffman as follows, namely Hegel–Riegel laws [Hoffman, 2003a]:

(1) The unity and struggle of opposites,
(2) The transformation of quantitative into qualitative change,
(3) The negation of the negation.

Hoffman expressed these principles with formal logic by means of set-theoretic operation with symmetric difference structure, and further relation to various mental activities is investigated in the name of topological psychology. We briefly introduce his framework for the following argument [Hoffman, 2003a, 2003b]. The symmetric difference \$ between two cognitive sets S and S' is defined using set-theoretic intersection ($\cap$) and union ($\cup$) as follows:

$$S\$S' = (S \cup S') \backslash (S \cap S'). \tag{77}$$

The symmetric difference and the complements $\neg(S \cup S')$ are equivalent to Sheffer stroke in logic, which is the NAND operation in Boolean algebra and the generator of other operators in ordinary first-order logic such as AND, OR, NOT, and IMP [Sheffer, 1913]. Formal operation is thus realizable by the combination of the symmetric difference and NOT ($\neg$), AND ($\cap$) operation. For example, the OR ($\cup$) operation is derived as follows:

$$S \cup S' = (S\$S')\$(S \cap S'). \tag{78}$$

Hoffman insists that the symmetric difference represents the first two of the Hegel–Riegel laws: In the definition (77), the term $(S \cup S')$ is naturally the unity, from which the commonality $(S \cap S')$ is substituted as the struggle of opposites. The quantitative change between two sets $|S| - |S'|$ is transformed into qualitative one $S\$S'$ by set-theoretic definition of commonality $(S \cap S')$. Here, $|S|$ corresponds to a simple quantitative measure of the set S, such as the element number. The third law, "the negation of the negation" corresponds to a part of the complementary set of symmetric difference $-(S\$S')$. From

$$\neg(S\$S') = (S \cap S') \cup (\neg S \cap \neg S'), \tag{79}$$

the complementary set $\neg(S\$S')$ is the union between the Hegel's union of opposite $S \cap S'$, and the intersection of not S and not S', as the negation of the negation. The second term can be considered as the context within the universe of discourse. The commonality $S \cap S'$ is nothing but the synthesis through the struggle between the thesis $S\backslash(S \cap S')$ and anti-thesis $S'\backslash(S \cap S')$. The cognitive sets S and S' are not appropriate in themselves to be called thesis and anti-thesis, since the synthesis $S \cap S'$ is supposed to contain a novel dimension which does not exist beforehand. The substitution of $S \cap S'$ in both theses represent the novelty of synthesis, which is at the same time a cognitive commonality between them. Figure 17 shows the Venn diagram of the symmetric difference \$ in relation to ordinary set-theoretic operations. Dialectical pair with symmetric difference is an elemental unit of our symbolic conceptualization, which establishes semantic

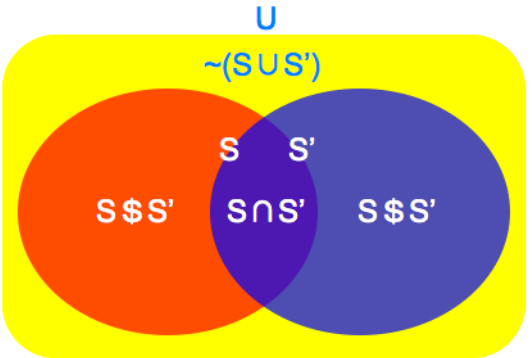


Fig. 17. Venn diagram of symmetric difference \$. The exclusive disjunction or XOR of two cognitive sets S and S' coincides with $S\$S'$, within the universe of possible cognitive sets U. Based on [Hoffman, 2003a].

relation where the meaning is defined by contrastive reference to symmetric counterpart [Riegel, 1973].

We utilize the introduced dialectical pair S and S' to formalize symbolically the results of autonomous learnings in CNN in a way that is compatible to set-theoretic operation. Chaotic itinerancy actually incorporates the dynamics on dialectical pair: The orbit intermittently transits between symmetric attractor ruins, which can be regarded as the symmetric sets without commonality. The presence of orbit in each attractor ruin is mutually exclusive. Therefore, the memory attractors $\mathrm{A} \rightleftharpoons \mathrm{B}$ and $\mathrm{C} \rightleftharpoons \mathrm{D}$ correspond to the thesis T and anti-thesis T' defined as a part of the cognitive sets S and S' as follows:

$$T = S \backslash (S \cap S'), \tag{80}$$

$$T' = S' \backslash (S \cap S'), \tag{81}$$

where T is $\mathrm{A} \rightleftharpoons \mathrm{B}$ and T' is $\mathrm{C} \rightleftharpoons \mathrm{D}$ in binary output space of CNN.

The synthesis $S \cap S'$ corresponds to the emerged attractors such as $\mathrm{E} \rightleftharpoons \mathrm{F}$, $\mathrm{G} \rightleftharpoons \mathrm{H}$ in Hebb learning, and various periodic attractors in STDP learning. Figure 18 shows the relation between thesis, anti-thesis, synthesis, and the underlying cognitive sets in case of Hebb learning. The commonality of symmetric cognitive sets $S \cap S'$ can be expressed as the OR operation of CNN binary output patterns. According to the simulation results, the cognitive sets can be directly defined from the geometrical composition of thesis and anti-thesis, but usually there is no analytical way to derive it: The synthesis is only possible in the OR product space of the memory attractors. In case of STDP learning, the situation becomes more complex since it includes further information of the change in periodicity. Empirical simulation can still incorporate statistically major results of emerged attractors into Boolean logic expression.

One would question what is the use of the autonomous learning, if the emerged results can be simply classified and expressed as Boolean logic. Actually our model does not allow simple symbolic classification of the results with respect to the parameter space: There exist high sensitivity on initial conditions as the property of chaotic dynamics. The group of the initial conditions giving the same result of autonomous learning is supposed to form a Cantor set. As shown in Fig. 18, we can only represent the results of emerged attractors with parameters specified at the digit limit of computation. Theoretically, it may reach real value precision. If one tries to simulate the same system as CNN with autonomous learning only by a system composed of Boolean logic, the infinite length of the program is needed to perfectly follow the sensitivity to initial conditions. The characterization of autonomous learning with Boolean logic only shows us the computational rationale in classified results with specified parameters, and does not reduce the complexity of chaos.

This raises a deeper consideration on what has been severely argued between classicists and connectionists. With the presence of chaotic dynamics, both approaches encounter the same difficulty in different aspects. The connectionists emphasize biological realism. They try to imitate the structure of the system and expect to reconstruct the mechanism of the input–output relations. The difficulty lies in the observational resolution of a cognitive

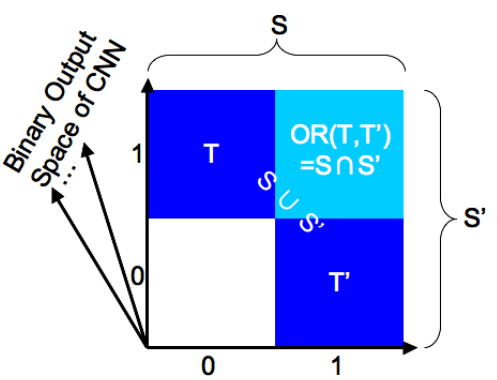


Fig. 18. Relation between cognitive sets S and S', including thesis T, anti-thesis T' respectively, and their OR product $\mathrm{OR}(T, T')$ in binary phase space of CNN outputs. Results of Hebb learning.

system, which is extremely complex in different scales. One cannot incorporate hierarchical structures including chaos of molecular scale in a macroscopic model. The artificial intelligence, on the other hand, is not necessarily attached to biological mechanisms but rather seeks to realize "intelligent" behaviors in a more metaphysical level. They start from abstracting human mental activity, then try to construct an algorithmic system and implement it in artificial systems such as computer simulation and robots to test the degree of achievement [Russell & Norvig, 2009].

The connectionist model including chaos such as CNN cannot specify the control resolution of initial conditions with respect to the long-term prediction. No matter how small we divide the phase space, miniscule changes quickly propagate to the system size. Therefore, one cannot specify a finite set of symbols which would accurately reproduce input–output patterns in a long run. In a more complex case such as the emergence of strange attractors in STDP learning, the condition requires much more resolution in shorter term.

Even without the autonomous learning, the approximation of chaotic itinerancy in CNN with time-series analysis model such as high-order Markov process encounters a similar difficulty from chaos. Chaotic dynamics usually contain unstable periodic orbits of arbitrary periodicity, which cannot be expressed as a deterministic finite state automaton.

Although the CNN is defined with a finite length of algorithm, the reconstruction of the generated sequences without the mechanism becomes impossible with a finite code due to the Cantor set structure of the phase space. The effort from the outside to capture the chaotic dynamics in an algorithmic way with a finite symbolic set is nullified by the infinite riddled structure of Cantor sets.

In other words, the perfect emulation of the autonomous learning in CNN is generally impossible, because the initial condition giving the same results is undecidable in computability theory. A set of natural numbers is called recursive, computable or decidable if there is an algorithm which terminates after a finite amount of time and correctly decides whether or not a given number belongs to the set. Since the Cantor set is defined on the infinite self-recurrence, such algorithm does not exist. Therefore, the emulation of CNN is undecidable in the same way as the halting problem of Turing machines.

The approach of artificial intelligence generally considers the input–output relations from the outside of the system, and propose the algorithm which would replace the black box, the inside of the system. Connectionist model including CNN try to reproduce the input–output patterns by reconstructing the internal mechanism of the system. With this, it is possible to abstract the top-down approach of artificial intelligence as the "*out–in emulation*", while the bottom-up connectionist model as the "*in–out reconstruction*", considering the difficulty of the same origin that both encounter in chaotic systems. Figure 19 schematically shows

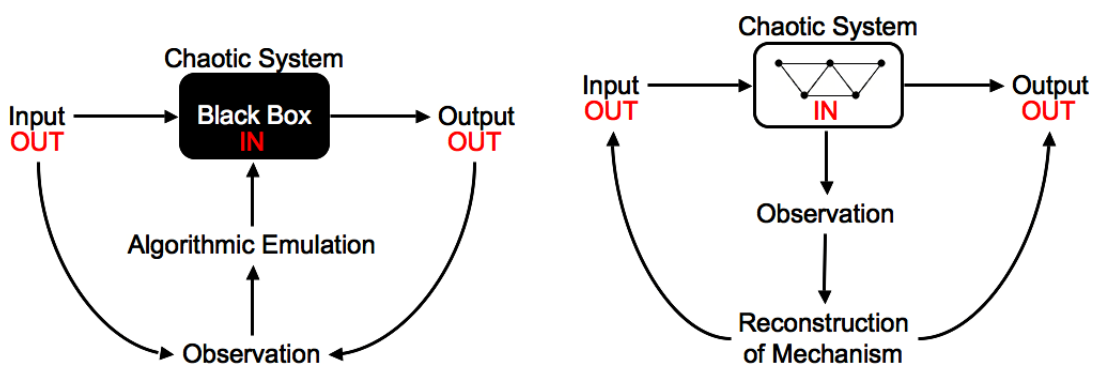


Fig. 19. Difference of approaches to chaotic systems in artificial intelligence and connectionism. (a) Out–in emulation of artificial intelligence and (b) in–out reconstruction of connectionism. Artificial intelligence generally tries to imitate the complex input–output relation of the system from the outside, by proposing the algorithm which realizes observed patterns. The model is not necessary following the neural basis, and in most cases conceptual. The connectionism starts from knowing the neural structure of cognitive system from the inside, and imitate its mechanism, expecting the reconstructed system would produce similar input–output relation.

the difference of these conceptualizations. Besides the conventional top-down and bottom-up classification, we propose to consider the complexity of chaotic input–output relations for the modeling of a developmental dynamics of the brain, which originates from inside of the system.

Autonomous learning is difficult to control and takes much more time to calculate than simple symbolic operation. Though, it is profoundly grounded to the intrinsic mechanism of cognitive systems and seek for the autonomous dynamics of information processing which actually produces novelty in living system. Interest of connectionism in terms of parallel information processing also shares the same perspective.

It is of further interest whether the other basic Boolean operators such as AND and NOT are realizable with autonomous learning. Actually, the various emergent periodic attractors from STDP learning contain the AND patterns of the memorized attractors. The NOT sets of the synthesis $E \rightleftharpoons F$ and $G \rightleftharpoons H$ can also be interpreted as the NAND of the two cognitive sets in Fig. 18. Empirical listing of the learning rules and parameters with respect to the modification of memory structure is needed to realize the application in analog computing.

5. Conclusion

We first analyzed the dynamics of CNN mainly in terms of the periodicity, the deviation rate from attractors, the wandering range, and the linear stability according to the hierarchical structure of invariant subspaces. The chaotic itinerancy state of CNN was revealed to occur with the blowout bifurcation of the attractors, and was associated with the riddled basins structure.

Next, we synthetically investigated the simulation of the possible constructive role of chaotic itinerancy state in interaction with the autonomous learning rules. Hebbian learning was shown to be able to converge the memorized attractors to form novel ones at their equidistance points, conserving the periodicity and augmenting neural correlation. STDP learning rather suppressed average neural activity but derived a rich variation of periodicity in the emerged attractors. The deterministic property of the system allowed both Hebbian and STDP learning to conserve the invariant subspaces better than where the orbits are situated, which can be interpreted as the selective destruction and preservation of memory structure.

The results provide basic facts for the construction of an analog computing system which would implement the analyzed aspect of dialectical information processing with autonomous learning. Using this model as a dynamical basis of hermeneutics in the brain [Tsuda, 1991], further experimental verification can also be designed for the generalization of the results in more physiology-grounded situations.

Acknowledgments

This study was partially supported by CNRS, the long term study abroad support program of the University of Tokyo, and the French government (Promotion Simone de Beauvoir).

References

Adachi, M. & Aihara, K. [1997] "Associative dynamics in a chaotic neural network," *Neural Netw.* **10**, 83–98.

Aihara, K., Matsumoto, G. & Ikegaya, Y. [1984] "Periodic and non-periodic responses of a periodically forced Hodgkin–Huxley oscillator," *J. Theoret. Biol.* **109**, 249–269.

Aihara, K., Matsumoto, G. & Ichikawa, M. [1985] "An alternating periodic-chaotic sequence observed in neural oscillators," *Phys. Lett. A* **111**, 251–255.

Aihara, K., Takabe, T. & Toyoda, M. [1990] "Chaotic neural networks," *Phys. Lett. A* **144**, 333–340.

Ay, N., Olbrich, E., Bertschinger, N. & Jost, J. [2006] "A unifying framework for complexity measures of finite systems," number 06-08-028 (Santa Fe Institute).

Demetriou, A. & Raftopoulos, A. [2004] "The shape and direction of development: Teleologically but erratically lifted up or timely harmonious?" *J. Cogn. Devel.* **5**, 89–95.

Fischer, K. W. & Kennedy, B. P. [1997] *Change and Development*: *Issues of Theory*, *Method*, *and Application* (Lawrence Erlbaum Associates, Mahwah, NJ), pp. 117–152.

Fischer, K. W. & Bidell, T. R. [2006] *Theoretical Models of Human Development*, *Handbook of Child Psychology*, 6th edition, Vol. 1 (Wiley, NY), pp. 313–399.

Fujii, H. & Tsuda, I. [2004] "Neocortical gap junction-coupled interneuron systems may induce chaotic behavior itinerant among quasi-attractors exhibiting transient synchrony," *Neurocomputing* **58–60**, 151–157.

Fujii, H., Aihara, K. & Tsuda, I. [2004] "Functional relevance of 'excitatory' GABA actions in cortical interneurons: A dynamical systems approach," *J. Integr. Neurosci.* **3**, 183–205.

Funabashi, M. & Aihara, K. [2007] "Modeling birdsong learning with a chaotic Elman network," *Artif. Life Robot.* **11**, 162–166.

Funabashi, M. [2014] "Network decomposition and complexity measures: An information geometrical approach," *Entropy* **16**, 4132–4167.

Hebb, D. O. [1949] *The Organization of Behavior*: *A Neuropsychological Theory* (Wiley).

Hegel, G. W. F. [1970] *Phänomenologie des Geistes* (Suhrkamp).

Hoffman, W. C. [2003a] *Mind in Time*: *The Dynamics of Thought*, *Reality*, *and Consciousness* (Hampton Press).

Hoffman, W. C. [2003b] "Symmetry in psychology," *Vis. Math.* **5**.

Ito, J. & Kaneko, K. [2002] "Spontaneous structure formation in a network of chaotic units with variable connection strengths," *Phys. Rev. Lett.* **88**, 028701.

Kaneko, K. & Tsuda, I. [2003] "Chaotic itinerancy," *Chaos* **13**, 937–946.

Kaneko, K., Tsuda, I., Tadokoro, S., Yasuoka, T. & Yamaguti, Y. [2004] "Chaotic itinerancy as a mechanism of irregular changes between synchronization and desynchronization in a neural network," *J. Integr. Neurosci.* **3**, 159–182.

Kang, S., Kitano, K. & Fukai, T. [2008] "Structure of spontaneous UP and DOWN transitions self-organizing in a cortical network model," *PLoS Comput. Biol.* **4**, e1000022.

Komuro, K. & Aihara, K. [2001] "Hierarchical structure among invariant subspaces of chaotic neural networks," *Japan J. Industr. Appl. Math.* **18**, 335–357.

Lewis, M. D. [2000] *Emotion*, *Development*, *and Self-Organization* (Cambridge University Press), pp. 37–69.

Lewis, M. D. [2001] *Appraisal Processes in Emotion*: *Theory*, *Methods*, *Research* (Oxford University Press, NY, US), pp. 205–220.

MacKay, R. S. [2005] "Indecomposable coupled map lattices with non-unique phase," *Dynamics of Coupled Map Lattices and of Related Spatially Extended Systems* (Springer, Berlin/Heidelberg), pp. 65–94.

MacKay, R. S. [2008] "Nonlinearity in complexity science," *Nonlinearity* **21**, T273–T281.

Markram, H., Lubke, J., Frotscher, M. & Sakmann, B. [1997] "Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs," *Science* **10**, 213–215.

Morra, S., Gobbo, C., Marini, Z. & Sheese, R. [2007] *Cognitive Development*: *Neo-Piagetian Perspectives* (Lawrence Erlbaum), pp. 367–368.

Oizumi, M., Albantakis, L. & Tononi, G. [2014] "From the phenomenology to the mechanisms of consciousness: Integrated information theory 3.0," *PLoS Comput. Biol.* **10**, e1003588.

Ott, E. & Sommere, J. C. [1994] "Blowout bifurcations: The occurrence of riddled basins and on-off intermittency," *Phys. Lett. A* **188**, 39–47.

Ott, E., Sommere, J. C., Antosen, T. M. Jr. & Venkataramani, S. [1995] "Blowout bifurcations: Symmetry breaking of spatially symmetric chaotic states," *Lecture Notes in Physics*, Vol. 450 (Springer, Berlin/Heidelberg), pp. 182–195.

Pascual-Leone, J. [1970] "A mathematical model for the transition rule in Piaget's development stages," *Acta Psychologica* **32**, 302–345.

Riegel, K. F. [1973] "Dialectical operations: The final period of cognitive development," *Human Devel.* **16**, 346–370.

Russel, S. J. & Norvig, P. [2009] *Artificial Intelligence*: *Modern Approach*, 3rd edition (Prentice Hall).

Sheffer, H. M. [1913] "A set of five independent postulates for Boolean algebras, with application to logical constants," *Trans. Amer. Math. Soc.* **14**, 481–488.

Skarda, C. A. & Freeman, W. J. [1987] "How brain make chaos in order to make sense of the world," *Behav. Brain Sci.* **10**, 161–195.

Tononi, G., Sporns, O. & Edelman, M. [1994] "A measure for brain complexity: Relating functional segregation and integration in the nervous system," *Proc. Natl. Acad. Sci. USA* **91**, 5033.

Tononi, G. [2004] "An information integration theory of consciousness," *BMC Neurosci.* **5**, 42.

Tsuda, I. [1991] "Chaotic itinerancy as a dynamical basis of hermeneutics in brain and mind," *World Futures* **32**, 167–184.

Tsuda, I. [1992] "Dynamic link of memory — Chaotic memory map in nonequilibrium neural network," *Neural Netw.* **5**, 313–326.

Tsuda, I. [2004] "Chaotic itinerancy is a key to mental diversity," *Behav. Brain Sci.* **27**, 585–602.

Tsuda, I. & Fujii, H. [2004] "A complex systems approach to an interpretation of dynamic brain activity I: Chaotic itinerancy can provide a mathematical basis for information processing in cortical transitory and nonstationary dynamics," *Computational Neuroscience*: *Cortical Dynamics*, Lecture Notes in Computer Science, Vol. 3146 (Springer, Berlin/Heidelberg), pp. 109–128.