Neural Networks 136 (2021) 72–86

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

Chaos may enhance expressivity in cerebellar granular layer

Keita Tokuda [a,*], Naoya Fujiwara [b,c,d], Akihito Sudo [e], Yuichi Katori [f,c]

[a] Department of Computer Science, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8577, Japan
[b] Graduate School of Information Sciences, Tohoku University, 6-3-09 Aoba, Aramaki-aza Aoba-ku, Sendai, Miyagi, 980-8579, Japan
[c] Institute of Industrial Science, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan
[d] Center for Spatial Information Science, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa-shi, Chiba 277-8568, Japan
[e] Faculty of Informatics, Shizuoka University, 3-5-1 Johoku, Naka-ku, Hamamatsu-shi, Shizuoka, Japan
[f] The School of Systems Information Science, Future University Hakodate, 116-2 Kamadanakano-cho, Hakodate, Hokkaido 041-8655, Japan

ARTICLE INFO

Article history:
Received 17 June 2020
Received in revised form 23 November 2020
Accepted 20 December 2020
Available online 28 December 2020

Keywords:
Cerebellar granular layer
Reservoir computing
Gap junction
Chaotic dynamics
Sierpinski gasket
Reaction–diffusion system

ABSTRACT

Recent evidence suggests that Golgi cells in the cerebellar granular layer are densely connected to each other with massive gap junctions. Here, we propose that the massive gap junctions between the Golgi cells contribute to the representational complexity of the granular layer of the cerebellum by inducing chaotic dynamics. We construct a model of cerebellar granular layer with diffusion coupling through gap junctions between the Golgi cells, and evaluate the representational capability of the network with the reservoir computing framework. First, we show that the chaotic dynamics induced by diffusion coupling results in complex output patterns containing a wide range of frequency components. Second, the long non-recursive time series of the reservoir represents the passage of time from an external input. These properties of the reservoir enable mapping different spatial inputs into different temporal patterns.

1. Introduction

Recent experimental studies have revealed that neighboring Golgi cells in the granular layer of the cerebellar cortex are densely interconnected with gap junctions that allow direct diffusion of ions between neuronal intracellular spaces[1] (Dugué et al., 2009; Vervaeke et al., 2010). Vervaeke et al. (2010) reported that more than 80% of neighboring neuron pairs are interconnected with gap junctions, and that each Golgi cell is connected to approximately 10 other Golgi cells via gap junctions. They also showed that the diffusion current between neighboring Golgi cells has the effect of transiently desynchronizing the spike activities after external excitation (Vervaeke et al., 2010). This is contradictory to the classical view of the role of gap junctions, which is to synchronize nearby neurons (Watanabe, 1958). In spite of the complex effect of diffusion coupling between Golgi cells on the ongoing dynamics, the causal relationship between this dynamics and cerebellar computation has yet to be elucidated.

Several theoretical studies have pointed out that diffusion coupling between nonlinear oscillators not necessarily realizes synchronization, but also induces instability (Turing, 1952) or even chaotic activity (Fujii & Tsuda, 2004; Katori et al., 2010; Schweighofer et al., 2004; Tadokoro et al., 2011; Tokuda et al., 2010, 2019; Tsuda et al., 2004; Yamada & Kuramoto, 1976). Fujii and Tsuda (2004) and Tsuda et al. (2004) reported that introducing diffusion coupling through gap junctions between class 1 neurons induces chaotic dynamics. Schweighofer et al. (2004) proposed a theory that the abundant gap junctions in the inferior olive produce chaotic neural activity that enables efficient transmission of information in the high-frequency components of inputs. It has also been proposed that the adaptive strength of the gap junction in the inferior olive regulates the degrees of freedom of the system, and the brain modifies the gap junction strength during learning to ensure that the system operates at an optimal level of degrees of freedom (Hoang et al., 2020; Kawato et al., 2011; Tokuda et al., 2017, 2013). The possible computational role of diffusion coupling through gap junction in the granular layer should be elucidated as well.

The majority of cerebellar computational theories assume that the cerebellum is a supervised machine that learns a desirable input–output relationship (Albus, 1971; Buonomano &

* Corresponding author.
E-mail address: tokuda@cs.tsukuba.ac.jp (K. Tokuda).

[1] Gap junctions are the structures that directly connect the intracellular spaces of neighboring cells, thereby forming electrical synapses that allow direct diffusion of ions between the neuronal intracellular spaces. Several properties of an electrical synapse differ from those of a chemical synapse; an electrical synapse has almost no conduction delay, does not require spiking, and the connection is usually bidirectional and symmetric. In contrast, an interaction between neurons with chemical synapses consists of multiple active steps involving several structures (axon conduction, transmitter release, sensing and response by the receptors on the postsynaptic cells) that result in a variety of dynamic responses with finite delays (∼ a few milliseconds) (Kandel et al., 2013).

https://doi.org/10.1016/j.neunet.2020.12.020

Mauk, 1994; Ito, 1970; Kawato et al., 1987; Marr, 1969; Raymond & Medina, 2018; Schweighofer et al., 2004; Wolpert et al., 1998; Yamazaki & Tanaka, 2007). It is well known that two major input pathways converge on the Purkinje cells; the mossy fiber–granular layer–Purkinje cell pathway originating from a precerebellar nucleus such as the pontine nucleus, and the climbing fiber–Purkinje cell pathway originating from the inferior olive (Ruigrok et al., 2015). These theories assume that the former pathway is the input layer of the supervised machine and the latter pathway conveys the supervising signals. The computational role of the cerebellar granular layer is assumed to be the preprocessing – feature engineering – of incoming signals from the mossy fibers. It transforms an input to a dynamical representation in a high-dimensional space realized by the enormous number ($\sim 10^{11}$ in human) of the granular neurons (Albus, 1971; Badura & De Zeeuw, 2017; Marr, 1969; Raymond & Medina, 2018). The granule cells and Golgi cells are the two major components of the cerebellar granular layer. Even though the major outputs of the granular layer are conveyed by the parallel fibers of the granule cells, the Golgi cells are also thought to play the central role of forming the representation because of the lack of direct recurrent connection within granule cells (Albus, 1971; Marr, 1969; Raymond & Medina, 2018).

It has long been known that the cerebellum plays a crucial role in motor learning, which requires execution of sequential movements with temporally precise timing (Ito, 1984). For example, a vast amount of experimental studies have characterized the essential information flow in the cerebellum that supports motor learning called classical eyeblink conditioning (Thompson, 2005). In a typical eyeblink conditioning, the animal is exposed to paired presentation of tone stimulus and a periorbital air puff stimulus intervened with a fixed interval (typically 250 ms) repetitively. After learning occurs, the animal acquires a temporally precise motor response (eyeblink) to the tone. The eyeblink response is precisely timed at the air puff onset with millisecond precision. It is also known that the interval discrimination task can be learned in eyeblink conditioning: animals can learn to elicit eyeblink responses at different latencies to different tone stimuli (Green & Steinmetz, 2005; Kehoe et al., 1993). Vast amount of evidence supports the fact that the cerebellum acquires the desired map to return a specific spatiotemporal pattern to a specific input. To explain this computation of the cerebellum, Buonomano and Mauk (1994) proposed a model of the granular layer consisting of sparse reciprocally connected granule cells and Golgi cells that are capable of representing the passage of time from the onset of an external sensory stimulus. In their model, mossy fiber excitation conveying the information of external tone stimulus elicits activity of the granule cells and the Golgi cells, with different sub-populations activated at different times. As a result, a specific sub-population of the granule cells is activated at a specific time from the onset of the stimulus, thereby representing the passage of time. This model successfully explains an important aspect of the behavioral and physiological traits in eyeblink conditioning.

Yamazaki and Tanaka (2007) extended Buonomano's model, and proposed the view that the cerebellum is a liquid state machine — a type of reservoir machine (Jaeger, 2001; Maass et al., 2002). In the reservoir computing framework, the input signals to the system project to a recurrent network called reservoir that has a highly nonlinear dynamics in a high dimensional space, and only the readout connections from this reservoir are modified to give the desired output signals. In Yamazaki's model, the general computational role of the cerebellum is to acquire a map between spatiotemporal input patterns and desired spatiotemporal output patterns. This is a natural elaboration of the classical Marr-Albus-Ito model regarding the cerebellum as a supervised machine, in that the reservoir machine can process spatiotemporal patterns. The granular layer works as the reservoir, and long-term depression (LTD) of the parallel fiber–Purkinje cell connection works as the learning rule. The Purkinje cell corresponds to the output neuron of the reservoir. They showed that the network of random recurrent connections between granule and Golgi cells realize temporally specific activation of different sub-populations of granule cells in response to external inputs. Their model successfully acquires a function that maps specific different inputs to specific different temporal patterns. To date, several studies of cerebellar function with the reservoir computing framework have been conducted (Rössert et al., 2015; Yamazaki & Nagao, 2012). However, even though these theories assume the inevitable functional role of the Golgi cells in realizing the reservoir, to our knowledge, no study has focused on the functional role of massive gap junctions between the Golgi cells. Considering the facts that chaotic activity is related to the performance of the reservoir (Bertschinger & Natschläger, 2004; Jaeger, 2001; Laje & Buonomano, 2013; Natschläger et al., 2005; Sussillo & Abbott, 2009; Yildiz et al., 2012) and that gap junction often induces chaotic activity (Fujii & Tsuda, 2004; Katori et al., 2010; Schweighofer et al., 2004; Tadokoro et al., 2011; Tokuda et al., 2010, 2019; Tsuda et al., 2004; Yamada & Kuramoto, 1976), it is necessary to elucidate how the gap junctions affect the computational performance of the granular layer as the reservoir, especially in terms of the effect of chaotic dynamics it may produce.

2. Methods

2.1. The network architecture

Fig. 1 shows the schematic diagrams of the two network architectures of the reservoir machines studied in the current study. We take the view that the cerebellum is a reservoir machine (Yamazaki & Tanaka, 2007). In this view, the granular layer of the cerebellum works as the reservoir, the mossy fiber projecting to the granular layer is the input, and the Purkinje cell is the output neuron. Learning via synaptic modification is realized by changing the connection strength of the parallel fibers, which are the readout connections. The reservoir maps an incoming input into a high-dimensional time series by nonlinear dynamics. Fig. 1(a) shows a simple reservoir machine composed of only Golgi cells that are mutually connected with diffusion coupling through gap junctions. In this study, we mainly focused on this model to evaluate the computational performance of diffusion-induced chaos as a reservoir machine. To confirm whether the results obtained with this model can also be observed in a more realistic situation, we perform a simulation incorporating the granule cells as well (Fig. 1(b)). This model incorporates other major components of the granular layer: the granule cells, the excitatory projections from the granule cells to the Golgi cells, and the inhibitory projections from the Golgi cells to the granule cells. The readout projection to the Purkinje cell originates from the granule cells as the actual cerebellar anatomical structure. Note that we do not incorporate a feedback loop from the reservoir output to the input in this study.

2.2. The model dynamics

A cerebellar Golgi cell is known to have the following properties: (1) it shows a periodic activity *in vitro* (Forti et al., 2006; Solinas et al., 2007; Vervaeke et al., 2010), (2) its spike frequency increases as external input increases (Forti et al., 2006; Solinas et al., 2007), and (3) its diffusion coupling induces desynchronization of neighboring cells (Vervaeke et al., 2010). We model the Golgi cells with the μ-model, which is a simple model described by a two-dimensional ordinary differential equation (Tsuda et al.,

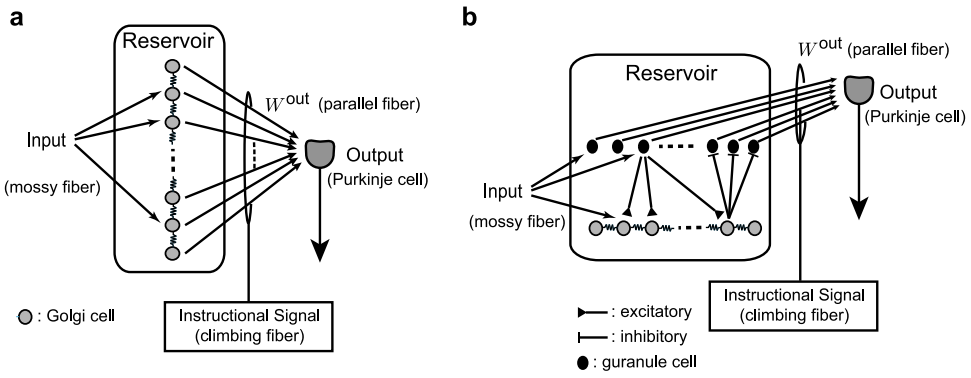

Fig. 1. Two network architectures of the reservoir machine considered in the current study. The assumed corresponding anatomical structures of the cerebellum are also indicated. (a) The Golgi cells coupled with gap junctions are the reservoir. Other structures in the granular layer, such as the granule cells and chemical synapses, are omitted in order to focus on studying the computational performance of neurons with gap junctions acting as the reservoir. (b) The model incorporating the granule cells, the excitatory projections from the granule cells to the Golgi cells, and the inhibitory projections from the Golgi cells to the granule cells.

2004). The μ-model is a class 1 neuron model that shows spiking activity after a saddle–node bifurcation occurs as tonic external input increases. This model shows periodic activity under isolated conditions with a tonic input, increases spike frequency to increasing tonic input, and shows aperiodic activity when coupled with diffusion (Katori et al., 2010; Tadokoro et al., 2011; Tokuda et al., 2013, 2019; Tsuda et al., 2004). The simple model shown in Fig. 1(a) is described as follows:

$$\frac{dV_i^{go}}{dt} = -R_i^{go} - \mu \cdot \left(V_i^{go}\right)^2 \left(V_i^{go} - \frac{3}{2}\right) + J_i^{go} + I^{go,tonic} + I_i^{go,input}, \tag{1}$$

$$\frac{dR_i^{go}}{dt} = -R_i^{go} - \mu \cdot \left(V_i^{go}\right)^2, \tag{2}$$

$$J_i^{go} = \begin{cases} g_{GJ}(V_2^{go} - V_1^{go}) & (\text{for} \quad i = 1), \\ g_{GJ}(V_{i+1}^{go} + V_{i-1}^{go} - 2V_i^{go}) & (\text{for} \quad i = 2, \ldots, N_{go} - 1), \\ g_{GJ}(V_{N_{go}-1}^{go} - V_{N_{go}}^{go}) & (\text{for} \quad i = N_{go}), \end{cases} \tag{3}$$

where V_i^{go} is the membrane potential of the ith Golgi cell, R_i^{go} is the recovery variable representing the ion channel activity of the ith Golgi cell, μ is the parameter of the μ-model, $I^{go,tonic}$ is the common tonic input to all cells, J_i^{go} is the diffusion current into the ith Golgi cell from the neighboring cells conducted through the gap junctions, g_{GJ} is the conductance of a gap junction, $I_i^{go,input}$ is the input signal to the reservoir described below, and N_{go} is the number of Golgi cells. The model only differs from that of the former studies (Fujii & Tsuda, 2004; Tadokoro et al., 2011; Tsuda et al., 2004) in that the input signal $I_i^{go,input}$ is incorporated in Eq. (1). In the μ-model, both the units of time and the variables are arbitrary. We use milliseconds as the unit of time for convenience. We use parameters $\mu = 1.7$, $g_{GJ} = 0.08$ (typical parameter set showing chaotic dynamics) except in Figs. 4, 5 where the dependency of the dynamics on these parameters is studied, and in Figs. 3, 10 where $g_{GJ} = 0$ is used. For the parameter $I^{go,tonic}$, we use $I^{go,tonic} = 0.004$ other than in a simulation shown in Fig. 2(b) and (c). As in the former studies (Katori et al., 2010; Tadokoro et al., 2011; Tokuda et al., 2013, 2019; Tsuda et al., 2004), we choose a one-dimensional chain with nearest-neighbor coupling (Eq. (3)). This setting makes it easy to compare our results to those of previous studies such as the one that conducted a detailed bifurcation analysis (Tadokoro et al., 2011) and it reduces the computational cost of the numerical simulations. It has been shown that this network topology is not a necessary condition for chaotic dynamics, but it appears in the two-dimensional lattice of the μ-model as well (Tokuda et al., 2019). Quantitatively, the same results as those of the current study are obtained in two- and three-dimensional topology (see the Supplementary material (MMC S1) and Video (Video S3)).

We restrict the form of the input signal $\boldsymbol{I}^{go,input} \in \mathbb{R}^{N_{go}}$ to an instantaneous pulse as follows:

$$\boldsymbol{I}^{go,input} = \boldsymbol{x}^{in}\delta(t - t_1), \tag{4}$$

where $\delta(t)$ is the Dirac delta function, $\boldsymbol{x}^{in,i} \in \mathbb{R}^{N_{go}}$ is the N_{go}-dimensional vector representing the amplitude of the input, t_1 is the time when the input is given to the reservoir. Practically, giving an input pulse is done by setting $\boldsymbol{V}^{go} \to \boldsymbol{V}^{go} + \boldsymbol{x}^{in}$ at time t_1, where $\boldsymbol{V}^{go} = (V_i^{go}) \in \mathbb{R}^{N_{go}}$ is the vector representation of the membrane potentials. In Section 3.4, a series of two different pulses are given to the system. In this case, the input signal $\boldsymbol{I}^{go,input}$ is described as follows:

$$\boldsymbol{I}^{go,input} = \boldsymbol{x}^{in,1}\delta(t - t_1) + \boldsymbol{x}^{in,2}\delta(t - t_2), \tag{5}$$

where t_1 and t_2 are the times when the input pulses are given to the reservoir.

The model incorporating the granule cells shown in Fig. 1(b) is described as follows:

$$\frac{dV_i^{gr}}{dt} = -R_i^{gr} - \mu \cdot \left(V_i^{gr}\right)^2 \left(V_i^{gr} - \frac{3}{2}\right) + I^{gr,tonic} + I_i^{ei} + I_i^{gr,input}, \tag{6}$$

$$\frac{dR_i^{gr}}{dt} = -R_i^{gr} - \mu \cdot \left(V_i^{gr}\right)^2, \tag{7}$$

$$\frac{dV_i^{go}}{dt} = -R_i^{go} - \mu \left(V_i^{go}\right)^2 \left(V_i^{go} - \frac{3}{2}\right) + J_i^{go} + I^{go,tonic} + I_i^{ie} + I_i^{go,input}, \tag{8}$$

$$\frac{dR_i^{go}}{dt} = -R_i^{go} - \mu \left(V_i^{go}\right)^2, \tag{9}$$

where V_i^{gr} is the membrane potential of the ith granule cell, R_i^{gr} is the recovery variable representing the ion channel activity of the ith granule cell, $I^{gr,tonic}$ is the common tonic input to the granule cells, and $I_i^{gr,input}$ is the input signal to the reservoir projecting to the ith granule cell. The diffusion currents are described as follows:

$$J_i^{go} = \begin{cases} g_{GJ}(V_2^{go} - V_1^{go}) & (\text{for} \quad i = 1), \\ g_{GJ}(V_{i+1}^{go} + V_{i-1}^{go} - 2V_i^{go}) & (\text{for} \quad i = 2, \ldots, N_{go} - 1), \\ g_{GJ}(V_{N_{go}-1}^{go} - V_{N_{go}}^{go}) & (\text{for} \quad i = N_{go}). \end{cases} \tag{10}$$

The currents I_i^{ei}, I_i^{ie} are the currents caused by chemical synapses, each representing the inhibition of the ith granule cell by the Golgi cells and the excitation of the ith Golgi cell by the granule cells, respectively, described with the following equations:

$$I_i^{\mathrm{ei}} = \sum_{j=1}^{N_{\mathrm{go}}} w_{ij}^{\mathrm{ei}} f(V_j^{\mathrm{go}} - \theta), \tag{11}$$

$$I_i^{\mathrm{ie}} = \sum_{j=1}^{N_{\mathrm{gr}}} w_{ij}^{\mathrm{ie}} f(V_j^{\mathrm{gr}} - \theta), \tag{12}$$

where w_{ij}^{ei} is the strength of the synaptic connection from the jth Golgi cell to the ith granule cell, w_{ij}^{ie} is the strength of the synaptic connection from the jth granule cell to the ith Golgi cell, θ is a parameter defining the threshold above which each neuron can be regarded as emitting a spike, and $f(V)$ is an activation function. In this study, we use $\theta = 0.7$, $f(V) = \frac{1}{1+\exp(-50V)}$. The synaptic strengths are determined using the following procedure. For each granule cell i, n_{ei} presynaptic Golgi cells are randomly chosen. Then, the strength is set at $w_{ij}^{\mathrm{ei}} = c^{\mathrm{ei}}/n_{\mathrm{ei}}$, if the jth Golgi cell is in the chosen group. The strength is set at $w_{ij}^{\mathrm{ei}} = 0$, if the jth Golgi cell is not in the chosen group. Similarly, for each Golgi cell i, n_{ie} presynaptic granule cells are randomly chosen. Then, the strength is set at $w_{ij}^{\mathrm{ie}} = c^{\mathrm{ie}}/n_{\mathrm{ie}}$, if the jth granule cell is in the chosen group. The strength is set at $w_{ij}^{\mathrm{ie}} = 0$, if the jth granule cell is not in the chosen group. The parameter values used are $\mu = 1.7$, $g_{\mathrm{GJ}} \in \{0,\ 0.08\}$, $N_{\mathrm{gr}} = 10^4$, $N_{\mathrm{go}} = 10^2$, $n_{\mathrm{ei}} = 4$, $n_{\mathrm{ie}} = 10^2$, $c^{\mathrm{ie}} = -c^{\mathrm{ei}} = 0.2$. These parameters are determined such that the ratio of the numbers of granule cells and Golgi cells, $N_{\mathrm{gr}}/N_{\mathrm{go}}$, and the number of synapses each neuron receive, n_{ei}, n_{ie}, are compatible with those described in the former study (Sudhakar et al., 2017).

Neurons in a specific subset of the reservoir neurons are connected to the outputs, which we refer to as the *projecting neurons* hereafter. Let V_i^{out} be the membrane potential of the ith projecting neuron. In the simple model shown in Fig. 1(a), all the Golgi cells are the projecting neurons. Thus, $V_i^{\mathrm{out}} = V_i^{\mathrm{go}}$. In the model shown in Fig. 1(b), all (and only) the granule cells are the projecting neurons. Thus, $V_i^{\mathrm{out}} = V_i^{\mathrm{gr}}$. The ith output of the reservoir, y_i is defined as follows:

$$y_i = \sum_{j=1}^{N_{\mathrm{out}}} w_{ij}^{\mathrm{out}} V_j^{\mathrm{out}}, \tag{13}$$

where N_{out} is the number of projecting neurons in the reservoir, and w_{ij}^{out} is the synaptic weight of the connection from the jth projecting neuron in the reservoir to the ith output y_i. The vector $\boldsymbol{y} = (y_i) \in \mathbb{R}^{N_y}$ defines the instantaneous output of the reservoir machine, where N_y is the number of outputs (the number of Purkinje cells considered). The output synaptic weight w_{ij}^{out} is time independent, and its value is modified only in the batch learning procedure described below.

2.3. Learning of the readout connection

Let $\boldsymbol{y}^{\mathrm{target}} \in \mathbb{R}^{N_y}$ be the target pattern consisting of N_y-dimensional time series defined over a time interval $[t_0,\ t_0 + T_{\mathrm{train}}]$, where T_{train} is the length of the time series. We determine the readout weight matrix $W^{\mathrm{out}} = \left[w_{ij}^{\mathrm{out}}\right] \in \mathbb{R}^{N_y \times N_{\mathrm{out}}}$ to minimize the following residual value:

$$\int_{t_0}^{t_0+T_{\mathrm{train}}} |\boldsymbol{y}^{\mathrm{target}} - \boldsymbol{y}|^2 \mathrm{d}t = \int_{t_0}^{t_0+T_{\mathrm{train}}} |\boldsymbol{y}^{\mathrm{target}} - W^{\mathrm{out}}\boldsymbol{V}^{\mathrm{out}}|^2 \mathrm{d}t, \tag{14}$$

where $|\boldsymbol{x}|$ is the Euclidean norm of a vector $\boldsymbol{x}$. In practice, this is conducted by sampling both the output vectors and the target vectors with a small sampling interval Δ. Let $\boldsymbol{\Omega} \in \mathbb{R}^{N_y \times K}$ be the discretized time series of the numerically integrated membrane potentials of the reservoir dynamics over the time interval $[t_0, t_0 + T_{\mathrm{train}}]$ as follows:

$$\boldsymbol{\Omega} = \left(\boldsymbol{V}^{\mathrm{out}}(t_0),\ \boldsymbol{V}^{\mathrm{out}}(t_0+\Delta),\ \boldsymbol{V}^{\mathrm{out}}(t_0+2\Delta),\ \ldots,\ \boldsymbol{V}^{\mathrm{out}}(t_0+K\Delta)\right), \tag{15}$$

where K is the natural number satisfying $K\Delta \le T_{\mathrm{train}} < (K+1)\Delta$, and $\boldsymbol{V}(t) = (V_i^{\mathrm{out}}(t)) \in \mathbb{R}^{N_{\mathrm{out}}}$ is the instantaneous membrane potentials of the projecting neurons. The target pattern matrix $\boldsymbol{Y}^{\mathrm{target}}$ is also defined by the same discretization as follows:

$$\boldsymbol{Y}^{\mathrm{target}} = \left(\boldsymbol{y}^{\mathrm{target}}(t_0),\ \boldsymbol{y}^{\mathrm{target}}(t_0+\Delta),\ \boldsymbol{y}^{\mathrm{target}}(t_0+2\Delta),\ \ldots,\ \boldsymbol{y}^{\mathrm{target}}(t_0+K\Delta)\right). \tag{16}$$

Then, the optimal readout weight matrix $\widehat{W}^{\mathrm{out}}$ is obtained by solving the following linear least square regression:

$$\widehat{W}^{\mathrm{out}} = \arg\min\ |\boldsymbol{Y}^{\mathrm{target}} - W^{\mathrm{out}}\boldsymbol{\Omega}|_{\mathrm{fro}}^2. \tag{17}$$

where $|W|_{\mathrm{fro}}$ is the Frobenius norm of a matrix W.

2.4. Evaluation of model performance

In order to evaluate the performance of the model, we use root mean square error normalized by that of the target pattern (nRMSE), as follows:

$$\mathrm{nRMSE} = \left(\frac{\int_{t_0}^{t_0+T_{\mathrm{train}}} (|\boldsymbol{y}^{\mathrm{target}} - \widehat{W}^{\mathrm{out}}\boldsymbol{V}^{\mathrm{out}}|^2)\mathrm{d}t}{\int_{t_0}^{t_0+T_{\mathrm{train}}} (|\boldsymbol{y}^{\mathrm{target}}|^2)\mathrm{d}t.}\right)^{\frac{1}{2}} \tag{18}$$

With the discretized time series, nRMSE is calculated as follows:

$$\mathrm{nRMSE} = \frac{|\boldsymbol{Y}^{\mathrm{target}} - \widehat{W}^{\mathrm{out}}\boldsymbol{\Omega}|_{\mathrm{fro}}}{|\boldsymbol{Y}^{\mathrm{target}}|_{\mathrm{fro}}}. \tag{19}$$

2.5. Lyapunov dimension

For the simple model (Fig. 1(a)) under no dynamical input, we characterized the strength of chaotic activity of the reservoir with the Lyapunov dimension (Kaplan & Yorke, 1979). First, we calculate the Lyapunov spectrum by the standard method with continuous Gram–Schmidt orthonormalization of the fundamental solutions to the linearized differential equation along the trajectory (Shimada & Nagashima, 1979). Then, let $\lambda_1 \ge \lambda_2 \ge \lambda_3 \cdots \ge \lambda_{2N_{\mathrm{go}}}$ be the Lyapunov exponents of the reservoir dynamics, and k be the maximal value of j such that $\sum_{i=1}^{j} \lambda_i \ge 0$, the Lyapunov dimension of the system is defined as follows:

$$D_{\mathrm{L}} = k + \frac{\sum_{i=1}^{k} \lambda_i}{|\lambda_{k+1}|}. \tag{20}$$

In this study, the Lyapunov dimension of the system is calculated under no external input to the system, where the system can be regarded as an autonomous dynamical system. Because we restrict the form of the input to the system as an instantaneous pulse (Eq. (4)), the property of the reservoir without any external input characterizes the system's response to the external input.

2.6. Similarity index

It is commonly assumed that the cerebellar granular layer is able to exhibit activity specific to the passage of time in response to an input (Buonomano & Mauk, 1994; Yamazaki & Tanaka, 2007). In order to evaluate the specificity of the instantaneous

$$C^{(1),(2)}(t_1, t_2) = \left| \frac{\sum_{i=1}^{N_{\text{out}}} (V_i^{\text{out},(1)}(t_1) - \langle V^{\text{out},(1)}(t_1)\rangle)(V_i^{\text{out},(2)}(t_2) - \langle V^{\text{out},(2)}(t_2)\rangle)}{\sqrt{\sum_{i=1}^{N_{\text{out}}} (V_i^{\text{out},(1)}(t_1) - \langle V^{\text{out},(1)}(t_1)\rangle)^2}\sqrt{\sum_{i=1}^{N_{\text{out}}} (V_i^{\text{out},(2)}(t_2) - \langle V^{\text{out},(2)}(t_2)\rangle)^2}} \right|, \tag{21}$$

Box I.

activity of the reservoir with respect to the passage of time and external input, we use the similarity index between different states of the reservoir. Let $\boldsymbol{V}^{\text{out},(1)}(t)$ and $\boldsymbol{V}^{\text{out},(2)}(t)$ be two different time series of the projecting neurons of the reservoir generated with two different external inputs. We use the absolute value of the Pearson correlation coefficient between the instantaneous values of $\boldsymbol{V}^{\text{out},(1)}(t)$ and $\boldsymbol{V}^{\text{out},(2)}(t)$ at two time points t_1 and t_2, which we call the similarity index Eq. (21) as given in Box I: where $|x|$ denotes the absolute value of x, and $\langle V^{\text{out},(1)}(t)\rangle$ is the mean membrane potential at time t as follows:

$$\langle V^{\text{out},(1)}(t)\rangle = \frac{1}{N_{\text{out}}} \sum_{i=1}^{N_{\text{out}}} V_i^{\text{out},(1)}(t). \tag{22}$$

2.7. Continuous-time recurrent neural network with random connections

To characterize the current model of the cerebellar Golgi cells with diffusion, the temporal specificity of the model neuron activity is compared to that of the commonly used model that shows chaotic activity – continuous-time recurrent neural network (CTRNN) – with random connections (Sompolinsky et al., 1988; Sussillo & Abbott, 2009). The CTRNN model is described as follows:

$$\tau \frac{dV_i^{\text{ctrnn}}}{dt} = -V_i^{\text{ctrnn}} + \sum_{j=1}^{N_{\text{ctrnn}}} W_{ij} \tanh(V_j^{\text{ctrnn}}), \tag{23}$$

where V_i^{ctrnn} is the membrane potential of the ith neuron of the CTRNN, W_{ij} is the coupling strength of the synaptic connection from the jth neuron to the ith neuron, $\tanh(\cdot)$ denotes the hyperbolic tangent, and τ is the time constant. The coupling matrix W_{ij} is generated as follows: first, each entry of W_{ij} is sampled independently from the standard normal distribution, $\mathcal{N}(0, 1)$. Then, the largest absolute value of the eigenvalues of the connection matrix, $\rho_{\max} = \max(\text{abs}(\text{eig}(W_{ij})))$, is calculated. Lastly, W_{ij} is multiplied by $\rho/\rho_{\max}$, where ρ is a pre-determined scalar that specifies the spectral radius of W_{ij}. In this study, we use the parameter values $\tau = 1$ ms, $\rho = 10$ (The results of this study have little dependency on these parameters, unless τ is too large or ρ is too small).

2.8. Numerical calculation

The numerical simulations were conducted using the *ode45* function in Matlab R2019a (MathWorks Inc., Natick, MA, USA). The obtained time series of the reservoir states were further discretized with time step $\Delta = 0.1$ ms using the *interp1* function (Eq. (15)). The learning procedure to obtain the optimal readout weight matrix $\widehat{W}^{\text{out}}$ was conducted by solving a linear least squares regression in Eq. (17) using the Matlab function *mldivide*.

3. Results

3.1. Broad distribution of the interspike interval caused by gap junctions

First, we show that introducing diffusion coupling causes chaotic dynamics, which in turn results in a broad distribution of interspike interval (ISI) of a neuron in the model. Fig. 2 shows examples of the evolution of the simple reservoir (Fig. 1(a)) described by Eqs. (1)–(3). Namely, the model network is a one-dimensional chain of neurons coupled to nearest neighbors with gap junctions and does not consist of any connections via chemical synapses. Fig. 2(a) illustrates an evolution of the membrane potentials ($N_{\text{go}} = 100$) under a positive tonic input, $I^{\text{go,tonic}} = 0.004$. An external perturbation at $t = 0$ is given with $\boldsymbol{x}^{\text{in}}$ in Eq. (4) as $\boldsymbol{x}^{\text{in}} = 0.2\mathbf{e}_{50}$, to the all-synchronized state, where $\mathbf{e}_{50}$ is the 50th standard basis. Namely, the membrane potential of the neuron in the center (50th neuron) is set as $V_{50}^{\text{go}} \to V_{50}^{\text{go}} + 0.2$. The network shows chaotic activity induced by diffusion coupling, as reported previously (Tadokoro et al., 2011; Tsuda et al., 2004). Periodic synchronous activity is observed before the propagation of chaotic dynamics is elicited by an external perturbation. The behavior of the cell in the center of the network is shown in Fig. 2(d). The activity is quite different from an isolated μ-model neuron with the same parameter, as shown in Fig. 2(e) ($I^{\text{go,tonic}} = 0.004$, $\mu = 1.7$). An isolated μ-model shows a saddle–node bifurcation at $I^{\text{go,tonic}} = 0$, and it has a periodic spiking activity with $I^{\text{go,tonic}} > 0$ and stable fixed point at resting potential with $I^{\text{go,tonic}} < 0$ (Tsuda et al., 2004). The ISI of the neuron in the center of the network (cell #50) in Fig. 2(a) is calculated for the subsequent 1×10^3 seconds (Fig. 2(f)) by regarding the neuron emitting a spike when crossing the threshold $\theta = 0.7$ from negative to positive. As visually evident in Fig. 2(a), (d), the ISI of this neuron shows a broad distribution over a wide range of periods, which is quite different from the periodic spiking activity without diffusion coupling shown in Fig. 2(e). The distribution has a small peak at around 50 ms, which is close to the period of the isolated single neuron (without the effect of the gap junction) shown in Fig. 2(e). Interestingly, the spatiotemporal pattern of the membrane potentials shows the fractal known as the Sierpinski gasket (Mandelbrot, 1983) under a small negative tonic input, $I^{\text{go,tonic}} = -0.00095$ (Fig. 2(b)). The spatiotemporal pattern at larger scale ($N_{\text{go}} = 1000$) shown in Fig. 2(c) clearly depicts self-similarity of the spatiotemporal pattern. Namely, the spatiotemporal pattern has no characteristic scale. The spatiotemporal pattern with positive tonic input (Fig. 2(a)) could be interpreted as a ruined pattern of the Sierpinski gasket (Fig. 2(b)), thus inheriting the fractal's property of scale invariance over multiple time scales (i.e., broad distribution of ISI).

3.2. Gap junctions in the reservoir realize producing a target pattern with a broad range of frequencies.

Next, we evaluate the effect of introducing gap junctions on the expressivity of the reservoir. More specifically, we examine how closely the model can output a sinusoidal temporal pattern with various temporal frequencies. Namely, we use the following sinusoidal wave (a scalar function of time t) as the target pattern $\boldsymbol{y}^{\text{target}}$ described in Eq. (14):

$$f(t; T^{\text{wave}}) = \sin\left(\frac{2\pi t}{T^{\text{wave}}}\right), \tag{24}$$

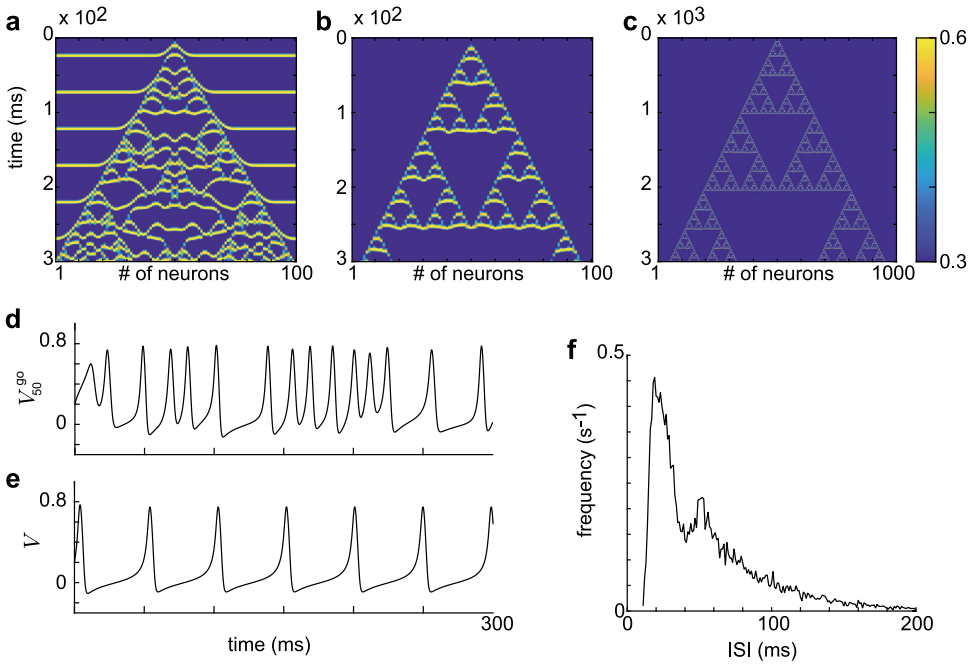

Fig. 2. The chaotic dynamics induced by gap junctions in the simple reservoir consist of a one-dimensional chain (Eqs. (1)–(3), Fig. 1(a)). (a) The dynamics of the model with a small positive tonic input, $I^{\mathrm{go,tonic}} = 0.004$. The colormap shows the dynamics of the membrane potentials, $\boldsymbol{V}^{\mathrm{go}}$. The parameter values are $N_{\mathrm{go}} = 100$, $\mu = 1.7$, $g_{\mathrm{GJ}} = 0.08$, $I^{\mathrm{go,tonic}} = 0.004$. (b–c) A Sierpinski gasket appears under small tonic inhibition, $I^{\mathrm{go,tonic}} = -0.00095$. The number of neurons is $N = 100$ in (b) and $N_{\mathrm{go}} = 1000$ in (c). (d) The time evolution of the membrane potentials of the 50th neuron in (a). (e) Periodic activity of an isolated μ-model neuron under a small positive tonic input, $I^{\mathrm{go,tonic}} = 0.004$. (f) Broad distribution of the ISI of a neuron in chaotic dynamics. Data of the 50th neuron in (a) for 1×10^3 seconds is shown. The bin width is 1 ms.

where T^{wave} is the period of the sinusoidal wave. We quantify the dependency of the following nRMSE value on the period of the sinusoidal target wave, T^{wave}, and on the model parameters:

$$\mathrm{nRMSE}(T^{\mathrm{wave}}) = \left(\frac{\mathrm{RSS}(T^{\mathrm{wave}})}{\int_{t_0}^{t_0+T_{\mathrm{train}}} |f(t; T^{\mathrm{wave}})|^2 \, \mathrm{d}t} \right)^{\frac{1}{2}}, \tag{25}$$

where $\mathrm{RSS}(T^{\mathrm{wave}})$ is the value of the objective function described in Eq. (14), which depends on T^{wave}. The spectrum of the value $\mathrm{nRMSE}(T^{\mathrm{wave}})$ over various T^{wave} gives a way to evaluate the model's ability to output a more general complex temporal pattern that contains various frequency components. Suppose there is a target pattern, y^{target} that is composed of a linear superposition of various sinusoidal waves as follows:

$$y^{\mathrm{target}} = \sum_{k=1}^{N^{\mathrm{wave}}} c_k f(t; T_k^{\mathrm{wave}}), \tag{26}$$

where N^{wave} is the number of different sinusoidal waves that compose the target pattern, T_k^{wave} is the period of the kth sinusoidal pattern. Let $\mathrm{nRMSE}^{\mathrm{target}}$ be the optimal nRMSE values for y^{target} that minimizes the objective function described in Eq. (14). Then, from the triangle inequality, the following inequality holds:

$$\left(\mathrm{nRMSE}^{\mathrm{target}}\right)^2 \leq \sum_{k=1}^{N^{\mathrm{wave}}} c_k^2 \left(\mathrm{nRMSE}(T_k^{\mathrm{wave}})\right)^2. \tag{27}$$

Thus, the right hand side of Eq. (27) gives the upper limit of the nRMSE value for the target pattern.

To evaluate $\mathrm{nRMSE}(T^{\mathrm{wave}})$ (Eq. (25)), the following analysis is conducted. Firstly, a time evolution over $[t_0,\ t_0 + T_{\mathrm{train}}]$ of the reservoir with a random initial input at $t = 0$ is numerically generated. Then, we use various sinusoidal waves (scalar function of time) over the same time interval $[t_0,\ t_0 + T_{\mathrm{train}}]$ as the target patterns. The readout connection is fitted to match each sinusoidal wave separately by the linear least squares. The dashed lines in Fig. 3(a) indicate the sinusoidal target patterns, and the solid lines indicate the fitted outputs of the model when the system is chaotic ($g_{\mathrm{GJ}} = 0.08$). The difference between the target patterns and the model outputs are visually not detectable when $T^{\mathrm{wave}} \geq 3$ ms. Note that, as shown in Fig. 2(d) and (e), the width of a spike of the model neuron is ~ 5–8 ms. The results of the same analysis without gap junctions are shown in Fig. 3(b). All settings of the analysis are the same as in Fig. 3(a) except for two conditions. First, the conductance of the gap junction is changed from $g_{\mathrm{GJ}} = 0.08$ to $g_{\mathrm{GJ}} = 0$. Second, the initial state of the reservoir is generated by randomly shuffling the phases of neurons. This is because, with $g_{\mathrm{GJ}} = 0$, all neurons behave as parallel isolated neurons with a shared identical period. Thus, the distribution of the phase of the neurons is time invariant, and biased distribution of the phase of the neurons should be disadvantageous in producing the sinusoidal target patterns. The phase of the initial state of each neuron is shuffled by separately simulating the evolution of each neuron for a sufficiently long period (3000 ms), discarding the initial transient, choosing a time point in the remaining time series from a uniform distribution, and employing the state variable, $(V^{\mathrm{go}}, R^{\mathrm{go}})$, as the initial condition of the neuron. The precision of the model output drastically decreases compared to the condition with diffusion coupling through gap junctions (Fig. 3(b)). The nRMSE values for both conditions are shown in Fig. 3(c). The nRMSE takes very small values over a wide range of the period of the target pattern when the gap junctions are incorporated in the model, whereas the nRMSE takes small values only at some specific range of the period if the model lacks the gap junctions. This result suggests that incorporating diffusion coupling in a reservoir enhances the output's expressivity that the network can generate by inducing chaotic dynamics.

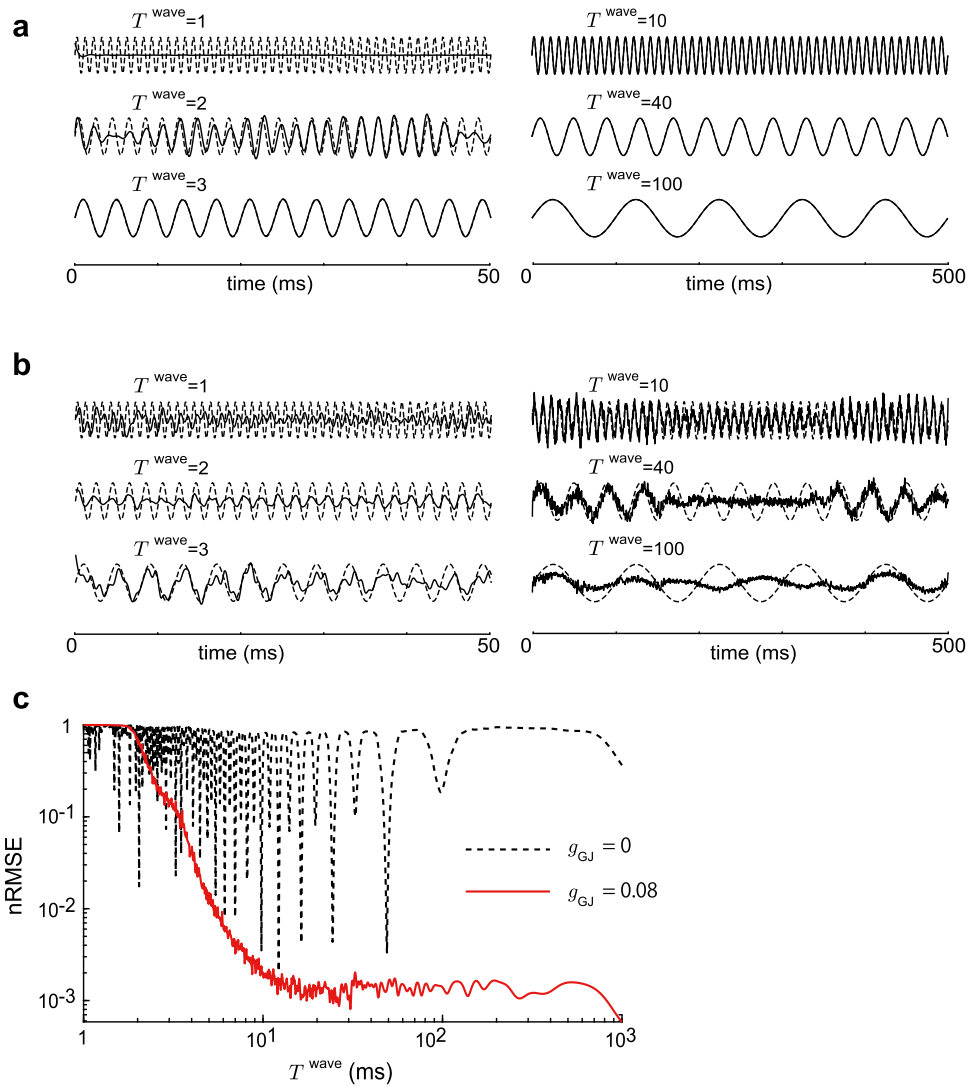

Fig. 3. Evaluation of the expressivity in the frequency domain. (a) The sinusoidal target patterns (dashed lines) and the fitted outputs of the model (solid lines) when the reservoir shows chaotic dynamics with diffusion coupling. The readout weights are determined separately for each sinusoidal target pattern with a different period, T^{wave}. The parameters are $T_{train} = 500$ ms, $\mu = 1.7$, $g_{GJ} = 0.08$, $N_{go} = 500$, $I^{go,tonic} = 0.004$, $\Delta = 0.1$ ms. Note that the spike width of a neuron is $\sim$ 5–8 ms, as shown in Fig. 2(b) and (c). Data only at $t = [0\ 50]$ are shown for $T^{wave} \leq 3$, though fitting is done for $t = [0\ 500]$. (b) Without diffusion through the gap junctions ($g_{GJ} = 0$). Each solid line indicates the fitted output of the model. (c) The nRMSE dependency on the period of the target pattern, with gap junctions ($g_{GJ} = 0.08$) (solid line), and without gap junctions ($g_{GJ} = 0$) (dashed line).

3.3. Inverse correlation of the Lyapunov dimension and nRMSE

Figs. 4 and 5 illustrate the parameter dependency of nRMSE and the Lyapunov dimension. The color map in Fig. 4(a) depicts the dependency of the nRMSE on the strength of gap junction coupling, g_{GJ}/N_{go}^2, and the target wave period T^{wave}. Fig. 4(b) illustrates the nRMSE dependence on g_{GJ}/N_{go}^2 when the target pattern period is $T^{wave} = 100$ ms. We employ g_{GJ}/N_{go}^2 rather than g_{GJ} because different network models with different sizes, but a same value of g_{GJ}/N_{go}^2, behave qualitatively the same (Tadokoro et al., 2011). This is because the model described with Eqs. (1)–(3) can be regarded as a discretization of a partial differential equation of a continuous one-dimensional excitable media. Models with different network sizes, N_{go}, and the same value of g_{GJ}/N_{go}^2 correspond to discretizations of the same partial differential equation with different spatial resolutions. Equivalently, different models with a same network size N_{go}^2 and different g_{GJ} values show spatiotemporal patterns with different spatial scales proportional to $1/\sqrt{g_{GJ}}$. As illustrated in Fig. 4(a) and (b), the nRMSE takes small value at a wide but specific range of g_{GJ} ($10^{-7.5} \leq g_{GJ} \leq 10^{-5.5}$). At the same time, the Lyapunov dimension, D_L, takes a large value at the same range of g_{GJ} (Fig. 4(c) and (d)). The Lyapunov dimension characterizes the strength of chaos, and represents the degrees of freedom of the dynamics (Kaplan & Yorke, 1979). Fig. 5 illustrates the nRMSE dependency on the parameter μ. Chaotic activity induced by diffusion coupling appears at a specific range of μ, as shown in the positive Lyapunov dimension in Fig. 5(c) and (d)). As in the case of the dependency of the nRMSE value on g_{GJ} (Fig. 4), the nRMSE takes very small values when the dynamics is chaotic and the Lyapunov dimension is large. These results showing the inverse correlation between the Lyapunov dimension and nRMSE suggest that diffusion-induced chaotic dynamics enhances the complexity of the reservoir's representation. Interestingly, the Lyapunov dimension takes a larger value than the system size (the number of the neurons in the model), $N_{go} = 500$, with some parameter values (Fig. 4(d)). This large Lyapunov dimension is distinct from that of the chaotic dynamics in the commonly used CTRNN model, which can only take less than 10% of the system size (Engelken et al., 2020).

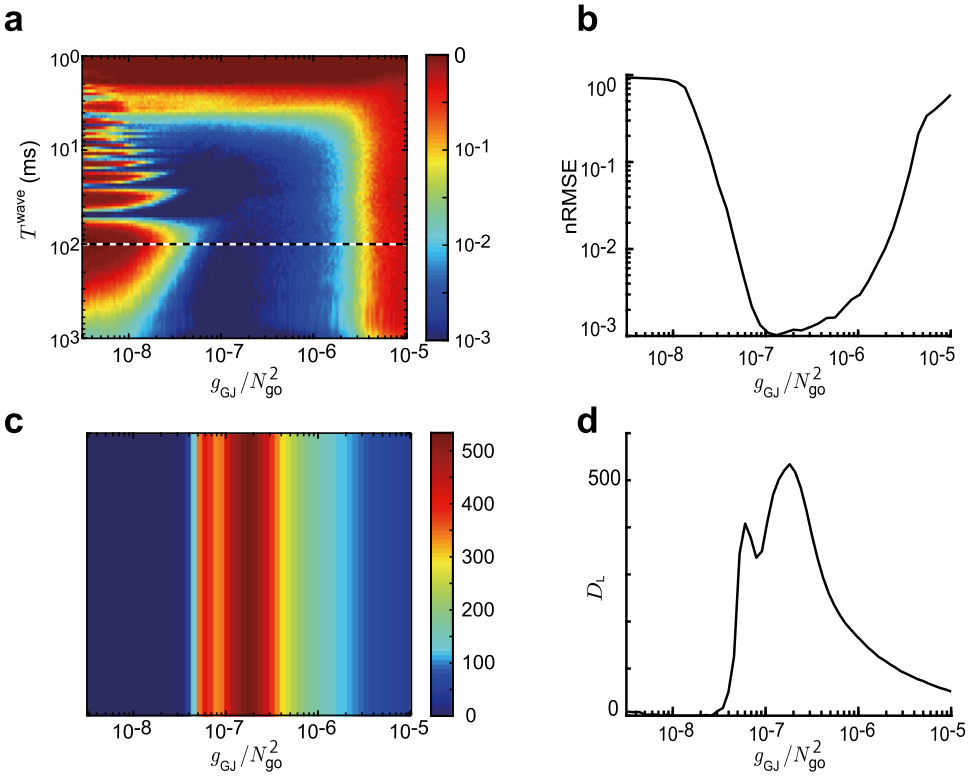

Fig. 4. The dependency of the nRMSE on the period of the target pattern, T^{wave}, and the strength of the gap junction, $g_{\text{GJ}}/N_{\text{go}}^2$. (a) The nRMSE is illustrated over various values of T^{wave} and $g_{\text{GJ}}/N_{\text{go}}^2$. The colormap shows the nRMSE value on a logarithmic scale. The parameter values are $N_{\text{go}} = 500$, $T_{\text{train}} = 500$ ms, $\mu = 1.7$, $I^{\text{go,tonic}} = 0.004$, $\Delta = 0.1$ ms. (b) The nRMSE value for a target pattern with $T^{\text{wave}} = 100$ ms plotted against $g_{\text{GJ}}/N_{\text{go}}^2$, corresponding to the section indicated by the dashed line in panel (a). (c) Colormap showing the Lyapunov dimension against $g_{\text{GJ}}/N_{\text{go}}^2$ (no variable is assigned to the vertical axis of the colormap since D_{L} does not depend on T_{train}). (d) The Lyapunov dimension plotted against $g_{\text{GJ}}/N_{\text{go}}^2$ (same value as panel (c)).

3.4. The reservoir state represents the passage of time from a specific input

We evaluate the reservoir's ability to represent the passage of time with the similarity index (Eq. (21)). The color map in Fig. 6(a) shows similarity indices $C^{(1),(1)}(t_1, t_2)$ defined by Eq. (21), calculated within a time series, $\boldsymbol{V}^{\text{go},1}(t)$, shown in the upper and left panels in Fig. 6(a). An external perturbation is given at $t = 0$ to an all-synchronized state as $V_{50}^{\text{go}} \rightarrow V_{50}^{\text{go}} + 0.5$. The length of the time series $\boldsymbol{V}^{\text{go},1}(t)$ is 1000 ms, and it is discretized with a time step of 1 ms to calculate the similarity indices, yielding a 1000×1000 matrix. Each element of the matrix corresponds to the correlation coefficient between the membrane potentials of the reservoir neurons at two different time points. Because $C^{(1),(1)}(t_1, t_2)$ is calculated within one time series, the diagonal elements are all 1. Fig. 6(b) shows the histogram of the values of all the elements above the main diagonal of $C^{(1),(1)}(t_1, t_2)$ shown in panel (a). Most of the elements have smaller values than ~ 0.4. This suggests that the state of the system does not come back close to the same point in the phase space – close enough that the similarity index takes high value close to 1 – within 1000 ms. In other words, the state of the reservoir activity has specificity to the passage of time.

The discriminative ability of the reservoir to different inputs is also important. Fig. 6(c) shows the similarity indices calculated between two time series, $\boldsymbol{V}^{\text{go},(1)}(t)$ and $\boldsymbol{V}^{\text{go},(2)}(t)$, that have slightly different inputs. The time series shown in the left panel, $\boldsymbol{V}^{\text{go},(2)}(t)$, evolves with the same initial condition as $\boldsymbol{V}^{\text{go},(1)}(t)$ shown in Fig. 6(a), but an additional input pulse is given to the cell at the edge of the network at $t = 200$ as $V_1^{\text{go}} \rightarrow V_1^{\text{go}} + 0.5$ (shown with an open magenta circle). Because of the chaotic property of the dynamics, the orbit diverges from the original unperturbed orbit of $\boldsymbol{V}^{\text{go},(1)}(t)$, and the correlation between the two time series vanishes at around $t = 350$ ms (Fig. 6(c), lower panel). This can be explained by the sensitivity to the initial condition of chaotic dynamics. Thus, the reservoir's response to input has high specificity to the input.

By conducting the same analysis, we compare the current model to the commonly used model that exhibits chaotic dynamics, i.e., CTRNN with a random connection matrix (Sompolinsky et al., 1988; Sussillo & Abbott, 2009), in terms of the temporal specificity of the model activity. This model is often used in reservoir computing frameworks. Fig. 7 shows the same analysis as shown by Fig. 6, conducted for both the current model (μ-model) and the CTRNN. Here, the initial state of the model is generated from a multivariate Gaussian distribution, simulation is conducted for 2000 ms, and the data from the first 1000 ms is discarded to exclude the transient state. Within the remaining time series of 1000 ms, the similarity index is calculated. The random parameter W_{ij} is generated every time the simulation is conducted. We use the parameter values $\tau = 1$ ms, $\rho = 10$ for the CTRNN model, but the results shown in Fig. 7(c) and (f) depend little on these parameters as long as τ is sufficiently small compared to the length of the time series (1000 ms), and ρ is sufficiently large that the probability of the system showing chaotic dynamics is high. As the parameter τ purely determines the time scale of the system, if τ is too large, the state of the system does not evolve much within the simulation time, resulting in a large similarity index within the different time points. When ρ is not sufficiently larger than 1, the system often does not show chaotic behavior, which results in a large similarity index as well. Fig. 7 (a–c) show the results when the number of

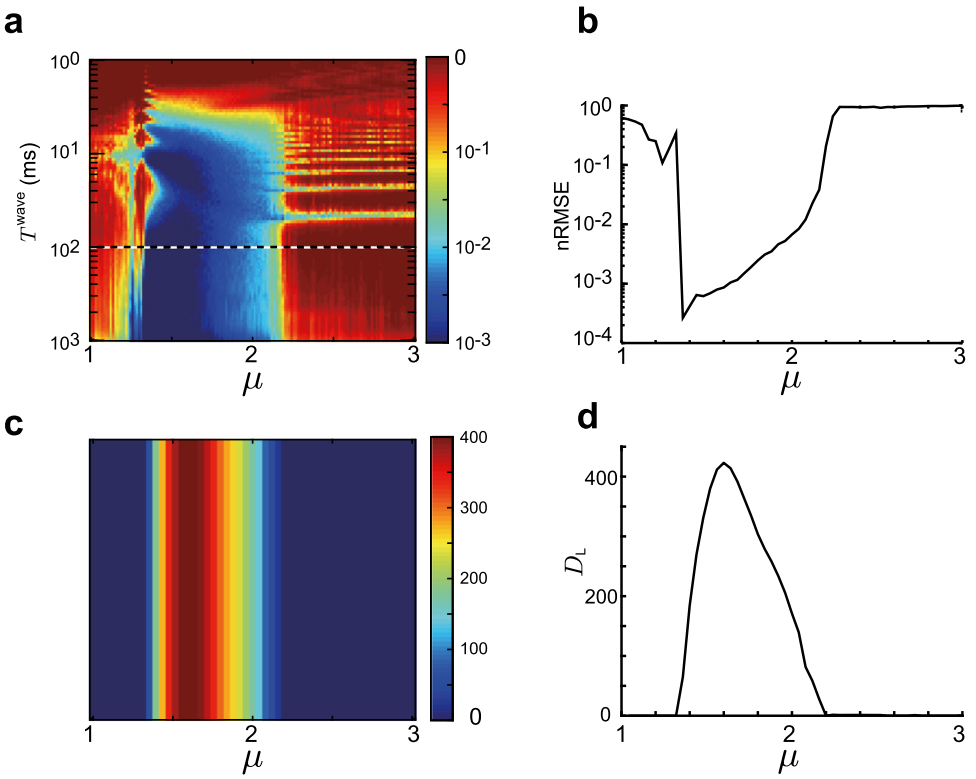

Fig. 5. The dependency of the nRMSE on the period of the target pattern, T^{wave}, and the parameter, μ. (a) The nRMSE is illustrated over various values of T^{wave} and μ. The colormap shows the nRMSE value on a logarithmic scale. The parameter values are $N_{\text{go}} = 500$, $T_{\text{train}} = 500$ ms, $g_{\text{GJ}} = 0.08$, $I^{\text{go,tonic}} = 0.004$, $\Delta = 0.1$ ms. (b) The nRMSE value for a target pattern with $T^{\text{wave}} = 100$ ms plotted against μ, corresponding to the section indicated by the dashed line in panel (a). (c) Colormap showing the Lyapunov dimension against μ (no variable is assigned to the vertical axis of the colormap since D_{L} does not depend on T_{train}). (d) The Lyapunov dimension plotted against μ (same value as panel (c)).

neurons is 100. Fig. 7(a) shows the result from the current model, showing a small similarity index between different time points, as in Fig. 6. In contrast, the similarity index has large values between different time points in the CTRNN (Fig. 7(b)). This difference is clearly shown in the histograms of the similarity index in Fig. 7(c) with flat distributions between 0 and 1 for the CTRNN. In the CTRNN, there are some trials with large frequencies at very large similarity indices near 1. This suggests that the system sometimes does not show strong chaotic activity because of the randomness of the value of parameter W_{ij} and the small number of the neurons, $N_{\text{ctrnn}} = 100$. Recently, Engelken et al. (2020) reported the size-invariant Lyapunov spectrum and the linear relationship between the attractor dimension and the systems size in the CTRNN model. According to their results, the CTRNN should show chaotic dynamics with the current parameter value with a larger system size. Fig. 7(d–f) show the results for the analyses with a larger system size $N_{\text{go}} = N_{\text{ctrnn}} = 500$. As expected, the histograms' variance across different simulations shown in Fig. 7(f) is smaller than that in Fig. 7(c). Additionally, both the histograms for the μ-model and the CTRNN shift toward smaller values than Fig. 7(c). This difference may be reflecting the increase of the attractor dimension with a larger system size in both models. The μ-model shows smaller distributions of the similarity indices than the CTRNN model with different system sizes. The histogram for the μ-model with $N_{\text{go}} = 100$ (Fig. 7(c)) shows even smaller values compared to the histogram for the CTRNN model with $N_{\text{ctrnn}} = 500$ (Fig. 7(f)). The dimension of the phase space is $2N_{\text{go}} = 200$ and $N_{\text{ctrnn}} = 500$, respectively, with the former smaller than the latter. Thus, the smaller value of the similarity index between different time points in the μ-model cannot be simply explained by the fact that the degrees of freedom of its phase space is two times larger than the number of its neurons. These results suggest that chaotic dynamics induced by diffusion results in activity with a higher temporal specificity with a smaller number of neurons, compared to the CTRNN model.

3.5. Generation of different activities for different inputs

Next, we examine whether a model with a fixed readout weight matrix is actually able to generate different temporal outputs for different inputs. Firstly, we show a simulation of eyeblink conditioning: an extensively studied model of cerebellar dependent learning, which has been reproduced in several computational studies as well (Bullock et al., 1994; Buonomano & Mauk, 1994; Li et al., 2013; Medina et al., 2000; Yamazaki & Igarashi, 2013; Yamazaki & Tanaka, 2007). We simulated a situation where the model outputs a specific time series with respect to a specific external input. In eyeblink conditioning, animals are able to acquire motor response with different timings to different types of tone stimuli (Green & Steinmetz, 2005; Kehoe et al., 1993). For example, an animal can learn to elicit a motor reflex to a tone stimulus 200 ms after the stimulus onset if the pitch of the tone is 600-Hz, and 600 ms after the onset if its pitch is 1-kHz (Kehoe et al., 1993). To reproduce this phenomenon, we consider the case where the model output, y, is a scalar function of time representing the eyeblink response, and train the model with multiple target patterns with its peaks at different latencies from the external input. Fig. 8(a) shows similarity indices calculated between four time series of the reservoir state with a same initial condition and four different inputs $\boldsymbol{x}^{\text{in},0}$, $\boldsymbol{x}^{\text{in},1}$, $\boldsymbol{x}^{\text{in},2}$, $\boldsymbol{x}^{\text{in},3} \in \mathbb{R}^{N_{\text{go}}}$. The four inputs, $\boldsymbol{x}^{\text{in},0}$, $\boldsymbol{x}^{\text{in},1}$, $\boldsymbol{x}^{\text{in},2}$, $\boldsymbol{x}^{\text{in},3} \in \mathbb{R}^{N_{\text{go}}}$ are vectors representing the spatial patterns that are generated from a multivariate Gaussian distribution $\mathcal{N}(\mathbf{0}, (0.2)^2\mathbf{E})$, where $\mathbf{E}$ is the identity matrix. The input pattern is given at time $t = 50$ as

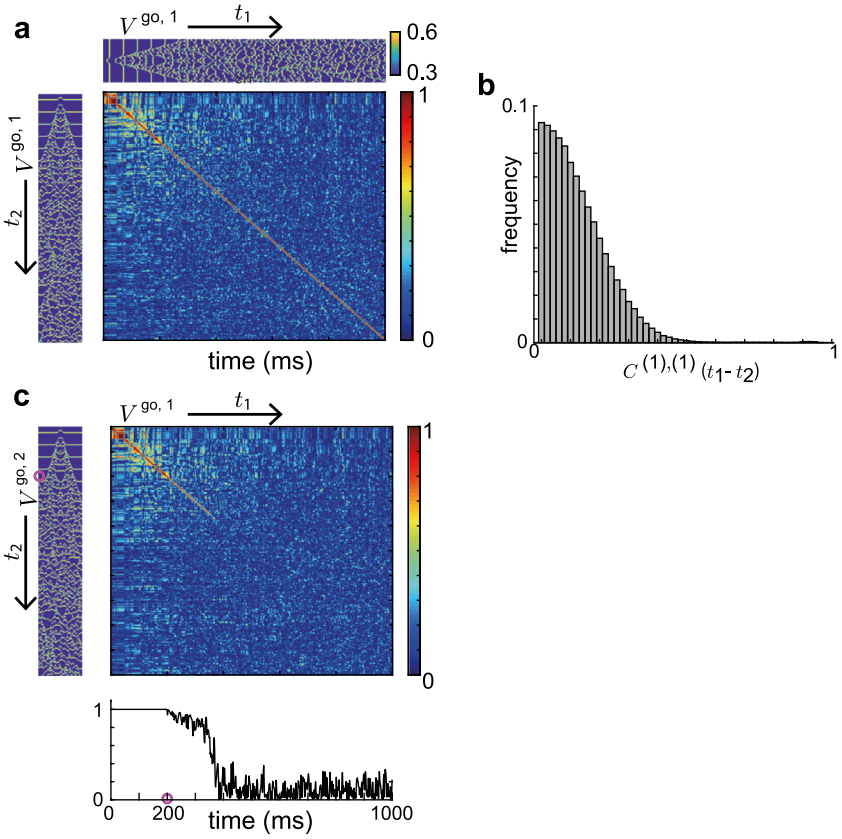

Fig. 6. The reservoir state is specific to the passage of time from a specific input. (a) Similarity indices $C^{(1),(1)}(t_1, t_2)$ calculated within a time series. The upper and left panels show the time series, $\boldsymbol{V}^{\mathrm{go},1}(t)$: the membrane potentials of reservoir neurons over 1000 ms. The neuron in the center (cell #50) is stimulated at $t = 0$ as $V_{50}^{\mathrm{go}} \to V_{50}^{\mathrm{go}} + 0.5$. The parameter values are $\mu = 1.7$, $g_{\mathrm{GJ}} = 0.08$, $N_{\mathrm{go}} = 100$, $I^{\mathrm{go,tonic}} = 0.004$. (b) Histogram of all the elements above the main diagonal of the similarity indices shown in (a). The bin width is 0.02. Normalized with total counts. (c) The similarity indices calculated between two time series, $\boldsymbol{V}^{\mathrm{go},1}(t)$ and $\boldsymbol{V}^{\mathrm{go},2}(t)$. The time series in the left, $\boldsymbol{V}^{\mathrm{go},2}(t)$, evolves with the same initial condition as $\boldsymbol{V}^{\mathrm{go},1}(t)$, but an additional input pulse is given at the 100 th cell at $t = 200$ ms (shown with an open magenta circle). The lower panel shows the diagonal elements versus time.

an impulse. (See Eq. (4) for the detailed description as to how the input stimulus is given to the model). As shown in Fig. 8(a), the four spatiotemporal patterns with the same initial condition but different external inputs $\boldsymbol{x}^{\mathrm{in},0}$, $\boldsymbol{x}^{\mathrm{in},1}$, $\boldsymbol{x}^{\mathrm{in},2}$, $\boldsymbol{x}^{\mathrm{in},3}$ at $t = 50$ rapidly lose the similarity among them after the inputs are given. When the time series with the inputs $\boldsymbol{x}^{\mathrm{in},0}$, $\boldsymbol{x}^{\mathrm{in},1}$, $\boldsymbol{x}^{\mathrm{in},2}$, $\boldsymbol{x}^{\mathrm{in},3}$ are fitted to different target patterns (dashed lines in Fig. 8(b)) simultaneously, the model acquires different outputs assigned for each time series of the reservoir (solid lines). This procedure of fitting a shared output connection for different output patterns is conducted as follows: first, four pairs of each desired pattern and corresponding time evolution of the reservoir dynamics are obtained by the numerical simulation of the reservoir. The four desired patterns are then concatenated to yield a novel vector whose length, K, is four times longer than the original patterns. This vector is used as the final target vector for training, $\boldsymbol{Y}^{\mathrm{target}} \in \mathbb{R}^{1\times K}$ in Eq. (16). Similarly, the four time series of the reservoir's membrane potentials are concatenated to yield a novel matrix whose size is $\mathbb{R}^{N_{\mathrm{go}}\times K}$, which is used as the matrix Ω in Eq. (15).

We also demonstrate the model's ability to generate two different human motions responding to two different inputs. Fig. 9 shows the two outputs of a learned model with a fixed output matrix: walking (Fig. 9(a)) and boxing (Fig. 9(b)). The simulation settings are the same as in Fig. 8. The model is trained on two time series simultaneously, each for a specific input at time $t = 0$. The two external inputs are drawn from a multivariate Gaussian distribution in the same way as the inputs $\boldsymbol{x}^{\mathrm{in},0}$, $\boldsymbol{x}^{\mathrm{in},1}$, $\boldsymbol{x}^{\mathrm{in},2}$, $\boldsymbol{x}^{\mathrm{in},3}$ in Fig. 8. The motion capture data provided by the Carnegie Mellon University Motion Capture Library (MOCAP) (http://mocap.cs.cmu.edu/) were used. Datasets 08_01.amc (walking) and 13_17.amc (boxing) were used as the target output patterns. Three variables representing the spatial offset of the person are discarded from the 62 dimensional signal. Thus, the target patterns are the remaining 59-dimensional temporal patterns. The result demonstrates that the model is able to generate completely different temporal patterns with a fixed readout connection, when the initial condition is different. See also the Supplementary Videos (Videos S1–2).

3.6. Incorporation of excitatory granular neurons

Lastly, we briefly confirm that the observed property of the simple model (Fig. 1(a)) can also be reproduced in the model incorporating the excitatory granule cells and the chemical synapses, shown in Fig. 1(b) (Eqs. (6)–(10)). A model incorporating 10^4 granule cells, 100 Golgi cells, and the reciprocal connections between the granule cells and the Golgi cells with the chemical synapses is composed. We evaluate the similarity index using the membrane potentials of the granule cells because the projecting neurons of the real cerebellar granular layer to the Purkinje cells are the granule cells. Fig. 10 shows examples of the dynamics of the model after a random initial input, with the gap junctions ($g_{\mathrm{GJ}} = 0.08$, Fig. 10(a)–(e)) and without gap junctions ($g_{\mathrm{GJ}} = 0$, Fig. 10(f)–(j)). Fig. 10(a) and (b) show the raster plots of the spikes of the granule cells and the Golgi cells, respectively. Both the granule cells and the Golgi cells show irregular activity with no apparent repetitive pattern. Fig. 10(c) is a scattergram

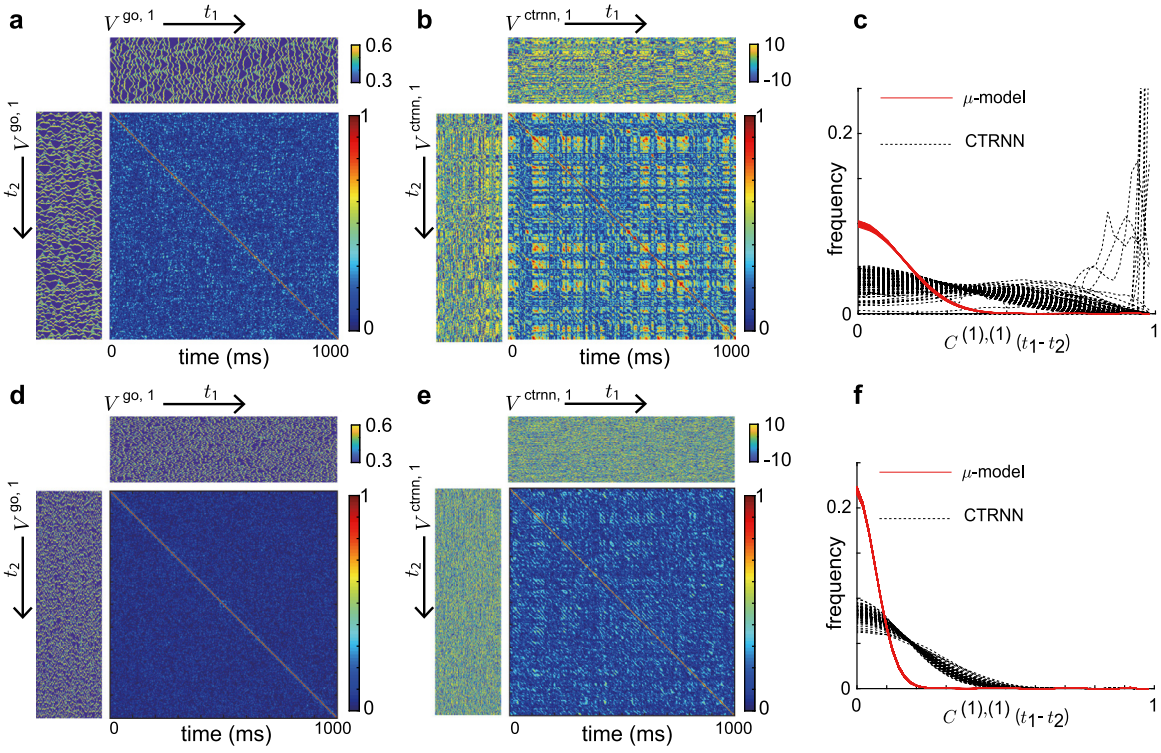

Fig. 7. Comparison of the similarity index with the CTRNN model. The panels (a–c) show the results for $N_{go} = N_{ctrnn} = 100$, and the panels (d–f) show the results for $N_{go} = N_{ctrnn} = 500$. The panels (a) and (d) show the results for the current model (μ-model), and the panels (b) and (e) show the results for the CTRNN. (a) Similarity indices $C^{(1),(1)}(t_1, t_2)$ calculated within a representative time series of the chaotic activity in the current model (same analysis as in Fig. 6 other than the initial condition). The parameter values are $\mu = 1.7$, $g_{GJ} = 0.08$, $N_{go} = 100$, $I^{go,tonic} = 0.004$. (b) The similarity index calculated within a representative time series of the evolution of the CTRNN model. The parameter values are $\tau = 1$ ms, $\rho = 10$, $N_{ctrnn} = 100$. For the randomly chosen parameter, W_{ij}, see the main text. (c) Histogram of all the elements above main diagonal of the similarity indices for the calculations shown in (a) and (b). Because the model parameter has a random parameter W_{ij}, the results of a 100 simulations are shown. The bin width is 0.01. Each trace normalized with total counts. (d–e) The same analysis as in (a) and (b) with the number of the neurons being $N_{go} = N_{ctrnn} = 500$. (f) Histogram of all the elements above the main diagonal of the matrix of similarity indices for the calculations shown in (d) and (e).

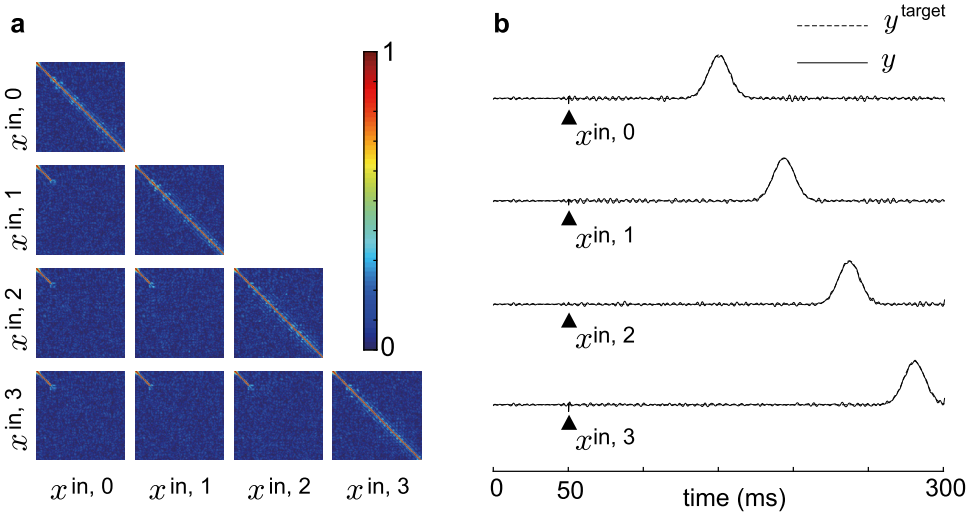

Fig. 8. Different output patterns for different inputs acquired by the model with a same readout connection matrix. (a) Each heat map shows the similarity index calculated between two different time series corresponding to two different inputs (as in Fig. 6(a)). Each side of each square corresponds to the time interval from $t = 0$ to $t = 300$. For example, the square that is lowermost and leftmost shows the similarity index calculated between two time series, one with the external input $\boldsymbol{x}^{in,0}$ given at $t = 50$, and one with the external input $\boldsymbol{x}^{in,3}$ given at $t = 50$. All the time series start from an identical initial condition. Since two time series receive different spatial inputs at $t = 50$, their orbits diverge after that, resulting in a decrease in the value of the similarity index, as shown by the diagonal line. The parameter values are $\mu = 1.7$, $g_{GJ} = 0.08$, $N_{go} = 500$, $I^{go,tonic} = 0.004$. (b) The time series with the initial inputs $\boldsymbol{x}^{in,0}$, $\boldsymbol{x}^{in,1}$, $\boldsymbol{x}^{in,2}$, $\boldsymbol{x}^{in,3}$ are fitted to different target patterns (dashed lines) with a shared readout connection. The output patterns are plotted with solid lines. The differences between the target patterns and the fitted patterns are so small that they are almost indistinguishable.

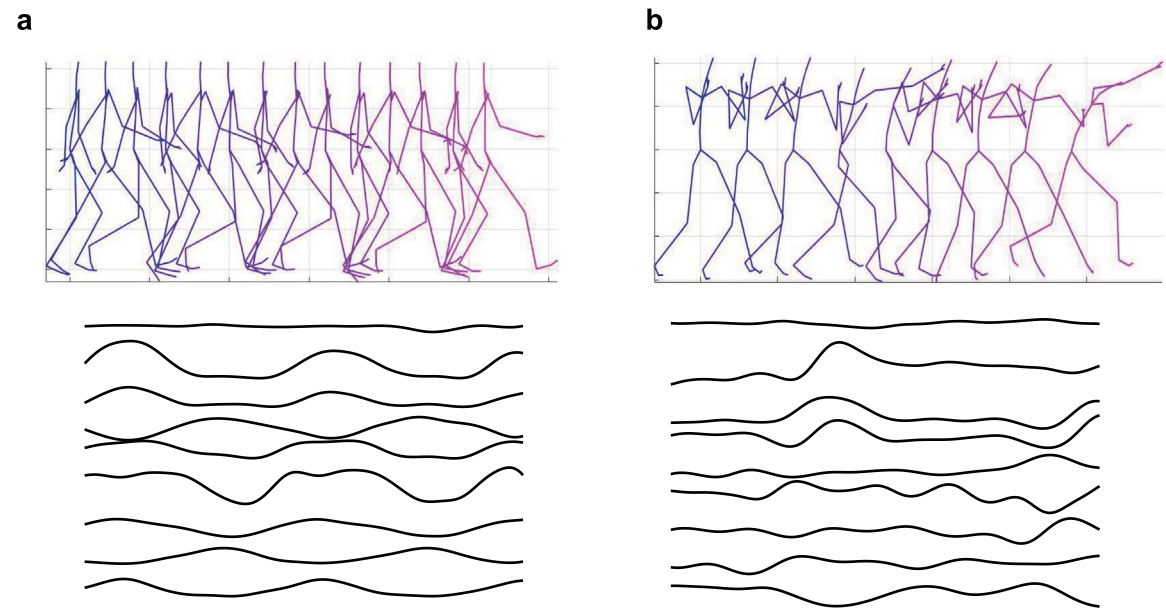

Fig. 9. The network showing walking (a) and boxing (b) patterns for two different initial inputs. The upper panels show the human motion generated by the model output visualized using a skeleton, and the lower panels show the corresponding raw output signals of the model representing the angles of the bones. Only the representative nine out of 59 traces are shown for simplicity. The same readout connection matrix is used for both movements, walking and boxing. The motion capture data from a human are learned. The parameters are $\mu = 1.7$, $g_{GJ} = 0.08$, $N_{go} = 500$, $I^{go,tonic} = 0.004$. Visualization was done using a toolbox, MATLAB Motion Capture Toolbox, at https://github.com/lawrennd/mocap. See also the Supplementary Videos (Videos S1–2).

showing the instantaneous membrane potentials of the granule cells, V_i^{gr}, at two different time points, $t = 954$ and $t = 990$. The similarity index between two different time points within this time series verifies that the state, $\boldsymbol{V}^{gr}$, is specific to time (Fig. 10(d)). The histogram of all the elements above the main diagonal of this matrix reveals the small similarity indices between two different time points (Fig. 10(e)). The similarity index is distributed at far smaller values than that of the Golgi membrane potentials in the simple model (Fig. 6(c) and (d)), presumably because of the large number of granule cells. Fig. 10(f)–(j) show the result of the same analysis with the parameter $g_{GJ} = 0$. All the other settings are identical. It is observed that, without diffusion coupling, the model dynamics converges to an all-synchronized periodic orbit by the interaction through the chemical synapses. The scattergram of $\boldsymbol{V}^{gr}$ in Fig. 10 (h) shows a high correlation between the state of the membrane potential at different time points corresponding to the two last peaks, even though the absolute value of its variation across different cells is very small. This is because the small deviations of the neurons' phases have not completely vanished within the first 1000 ms. The resulting small deviations from the instantaneous mean value of the membrane potentials across all the neurons are highly correlated at different phases of the periodic oscillation. As a result, the similarity index takes very high values after it converges to periodic activity at around $t = 250$ ms, independent of the phase of the global oscillation over multiple periods (Fig. 10 (i)). Thus, it is not the case that the small deviations of the phases of the neurons differ at different peaks of oscillation, and the system is able to represent the passage of time using different orders of the precise spike sequence at the peaks. Rather, the spike sequence of the neurons around the peaks of the global oscillation is highly preserved across different periods (although its absolute difference is less than 1 ms), and the system cannot represent the passage of time using different activity patterns. The distribution of non-diagonal elements of the similarity index clearly shows that changing the parameter g_{GJ} to 0 abolished the model's ability to represent the passage of time (Fig. 10(j)). These results suggest that the stable all-synchronized orbit exists in the system without gap junction, but it is destroyed by the chaotic dynamics if diffusion current through gap junctions exists.

4. Discussion

In the current study, we investigated the computational role of the gap junction between Golgi cells in the cerebellar granular layer. Specifically, we evaluated the computational performance of the model of the cerebellar cortex using a reservoir computing framework. First, we showed that introducing gap junctions in the model induces chaotic dynamics that enables the reservoir to output complex patterns containing a wide range of frequency components (Figs. 2–5). Second, we showed that the chaotic dynamics has a long non-recursive time series that is capable of representing the passage of time (Figs. 6–7). These properties of the chaotic dynamics realize the reservoir's ability to output the desired temporal patterns (Figs. 8, 9). Yamazaki and their colleagues have proposed a model with these abilities based on a different mechanism, i.e., the random connection by chemical synapses between granule cells and Golgi cells (Yamazaki & Tanaka, 2007). In the current study, we pointed out another possible mechanism (diffusion through gap junctions) that would be capable of reproducing the aforementioned abilities of the model. Because the gap junctions connect neighboring neurons, the connections realized by the gap junctions must inevitably be local rather than distant. Thus, the average number of neighboring cells a neuron can contact cannot be as large as the case of chemical synapses. The average degree of the network realized by gap junctions may therefore be small. In the literature, it has been argued that the desirable feature of a reservoir is that the connection should be sparse (Jaeger, 2001). This is in line with our hypothesis that diffusion coupling by the gap junction constitutes the reservoir in the cerebellum. On the other hand, one report showed that the small-worldness of the reservoir contributes to better performance (Kawai et al., 2019). In the granular layer of the real brain, the chemical synapses and the gap junctions may work in concert to realize preferable properties as a reservoir.

The ISI of a neuron in the chaotic dynamics induced by diffusion coupling through gap junctions shows a broad distribution over a wide range of periods, unlike that of an isolated neuron model (Fig. 2). Our result may explain the fact that a Golgi cell

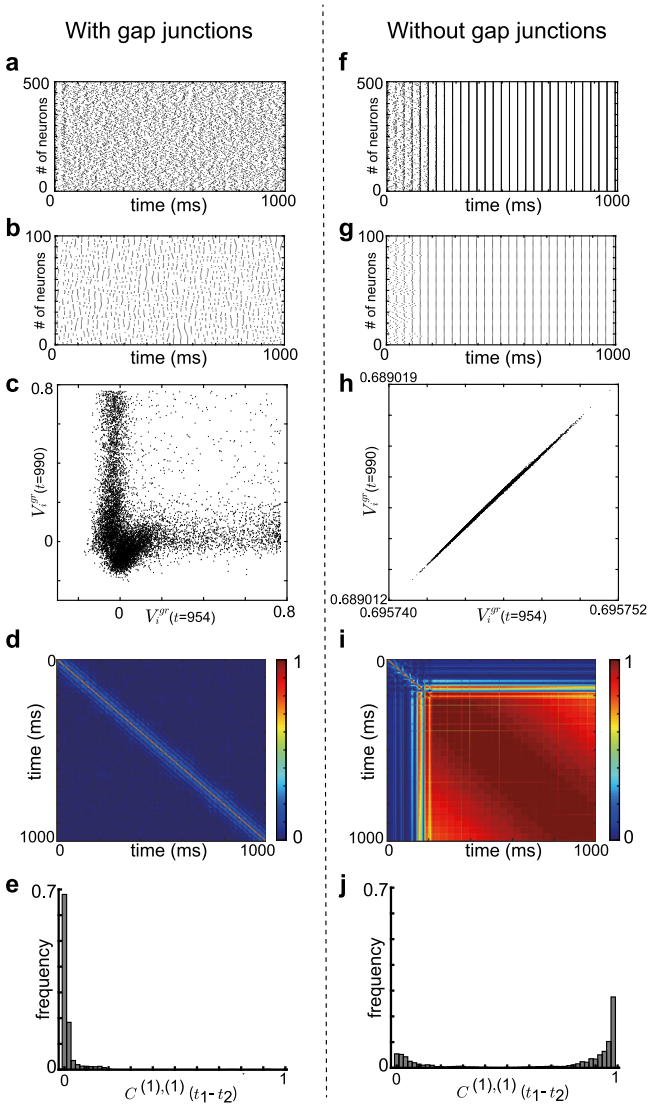

Fig. 10. The dynamics of the model incorporating the granule cells in the reservoir. (a–e) The model behavior with gap junctions under random initial input. The parameters are $\mu = 1.7$, $g_{\mathrm{GJ}} = 0.08$, $N_{\mathrm{gr}} = 10^4$, $N_{\mathrm{go}} = 100$, $I^{\mathrm{gr,tonic}} = 0.01$, $I^{\mathrm{go,tonic}} = 0.004$, $\theta = 0.7$. (a) The spike raster plot of the granule cells. Only the 1–500th cells out of $N_{\mathrm{gr}} = 10^4$ cells are shown. (b) The spike raster plot of all the Golgi cells. (c) Scattergram showing V_i^{gr} at two different time points, $t = 954$ and $t = 990$. (d) The similarity index calculated between the granule cells' instantaneous membrane potentials at different time points in the time series shown in (a). (e) Histogram of all the elements above the main diagonal of the matrix of similarity indices shown in (d). The histogram is normalized with total counts. (f–j) The model dynamics without gap junctions. All of the conditions of simulation and properties of the model are the same as those in (a–e), other than that parameter value $g_{\mathrm{GJ}} = 0$ is used. (f) Spike raster plot showing the activities of the granule cells of a model without gap junctions. (g) The spike raster plot of all the Golgi cells. (h) Scattergram showing a high correlation between V_i^{gr} at two different time points, $t = 954$ and $t = 990$, which correspond to the time of the last two peaks of the oscillation. (i) The similarity indices showing high values over different time points because of the rapid relaxation to a synchronized activity. (j) Histogram of all the elements above the main diagonal of the matrix of similarity indices shown in (i).

exhibits periodic activity in some experimental settings such as *in vitro* recordings (Forti et al., 2006; Solinas et al., 2007; Vervaeke et al., 2010), but also shows an irregular activity with broad ISI distribution *in vivo* (Holtzman et al., 2006).

Reservoir computing has drawn considerable attention in recent years, as it is expected to be a suitable and powerful framework for processing temporal sequences. However, the desirable dynamical properties that the reservoir must have, and its proper implementation, have not been well characterized, and still remain an open question (Bertschinger & Natschläger, 2004; Boyd & Chua, 1985; Jaeger, 2001; Maass et al., 2002; Natschläger et al., 2005; Yildiz et al., 2012). A variety of systems, including both physical systems and mathematical models, have been used as implementations of the reservoir (Tanaka et al., 2019). The current study investigated a reservoir consisting of a reaction–diffusion system (i.e., neurons coupled with gap junctions) and obtained results suggesting chaos in a reaction–diffusion system may contribute to the performance of the model. Interestingly, the Lyapunov dimension of the chaotic attractor induced by diffusion in the current model can be larger than the number of the neurons in the model, which is 50% of the freedom of the whole phase space (Fig. 4). In contrast, a recent study by Engelken et al. (2020) reported that in the CTRNN model, which is often used as the reservoir that shows chaotic dynamics, the ratio of the Lyapunov dimension of the chaotic attractor to the system size is invariant to the system size and less than 10%. This large difference in the ratio of the Lyapunov dimension to the system size is consistent with the smaller similarity index between two different time points along a temporal evolution of the system in the current model than the CTRNN model (Fig. 7). The specificity of the activity pattern was higher in the current model than in the CTRNN model having the same system size. Thus, incorporating diffusion coupling in a reservoir could be an advantageous implementation to raise the expressivity of the system. Further investigation of a reservoir machine using chaotic dynamics in a reaction–diffusion system may be an interesting direction for future study.

We found that the spatiotemporal pattern of the chaotic dynamics of μ-models coupled with diffusion in a one-dimensional chain shows the Sierpinski gasket (Fig. 2). It is reported that some other nonlinear reaction–diffusion systems also show the Sierpinski gasket (Hayase & Ohta, 1998, 2000). The simple model in our current study (Fig. 1(a)) also belongs to the class of reaction–diffusion system, as it consists of a one-dimensional chain of the neurons with nearest–neighbor connections. There maybe a common mathematical structure behind our model and these former studies. It is well known that the Sierpinski gasket appears in the spatiotemporal patterns of a cellular automaton (CA) such as Wolfram's Rule 90 (Wolfram, 1994). Some previous studies attempted to implement a reservoir with a CA (Morán et al., 2020; Yilmaz, 2015). Morán et al. used CA as the reservoir and constructed a classifier for handwritten characters of the MNIST dataset (Lecun et al., 1998), and compared the performance across the rules in CA. They reported that Wolfram's Rule 90 gives the best performance in the test data of the cross-validation (Morán et al., 2020). In the current study, we did not construct a classifier with our model. Thus, it is difficult to directly compare the current results with Morán and their colleagues' work. However, it should be an interesting direction to evaluate the performance of a classifier model with a reaction–diffusion system as the reservoir. Additionally, studies with CA may contribute to the elucidation of the cerebellar computation.

An important issue we did not consider in depth in the current study is the generalization ability of the model. Namely, the ability of the model to generate the same output from similar input. We showed that the chaotic dynamics realizes the specificity of the response to the input (Fig. 6). This can be interpreted by the chaotic dynamics' high sensitivity to the initial condition. However, the high sensitivity to the initial condition may cause poor generalization ability, because a small noise or a deviation in the input signal grows rapidly over time. Thus, systematic evaluation of the generalization ability of the current model should be an important issue to be elucidated in the

future study. The aforementioned study by Morán et al. showed that Wolfram's Rule 90, which shows the same Sierpinski gasket pattern as the current model we use, shows the best performance in the test data of cross-validation (i.e., it shows the highest generalization ability). In their study, the MNIST data is used as the initial state of the reservoir, and the spatiotemporal pattern of the evolution of CA over specific steps is used as the feature vector used for the subsequent classification task. Similarly, one could modify our current model so that the activity caused by the input decays within a specific time scale, before the small deviation in the input grows to the system size. This situation is actually similar in the real cerebellum because the cerebellum is believed to be able to maintain the input information for a fixed time, approximately $\sim$ 500 ms (Kotani et al., 2003; Thompson, 2005). It should be an interesting issue to elucidate the relationship to previously proposed properties of the reservoir such as the echo state property (Jaeger, 2001; Yildiz et al., 2012) or the edge of chaos (Bertschinger & Natschläger, 2004; Natschläger et al., 2005). Another important issue to be elucidated is the relationship between the strength of gap junctions and the generalization ability of the model. Fig. 4 shows the Lyapunov dimension of the system takes a large value at a specific range of the strength of gap junctions. Some previous studies pointed out the possibility that the strength of gap junctions changes the degrees of freedom of the network dynamics, which is crucial for the generalization ability (Hoang et al., 2020; Kawato et al., 2011; Schweighofer et al., 2013; Tokuda et al., 2017). Additionally, it should also be noted that cerebellar dependent motor learning requires repetitive training (Thompson, 2005), which would help generalization. This is very different from hippocampal dependent learning, where an episode is learned one-shot.

In the last part of the Results section, we confirmed that the non-recursiveness of the system is inherited in the model incorporating the excitatory granule cells (Fig. 10). Actually, the similarity index between different time points within a time series (Fig. 10(d)) shows higher specificity to the passage of time than in the simple model consisting of only the Golgi cells (Fig. 6(c)). This suggests that the large number of granule cells contributes to the specificity. In this study, we incorporated granular neurons with a population size only 10^2 times larger than the Golgi cells ($N_{gr} = 10^4$, $N_{go} = 10^2$) because of the computational cost. However, the ratio of the number of the granule cells to the Golgi cells in the real brain is reported to be even larger, as much as 430 times (Korbo et al., 1993). The numerous granule cells may serve a role in multiplying the representational ability of the Golgi neurons.

In conclusion, we proposed the hypothesis that the massive gap junctions between the Golgi cells in the cerebellar granular layer contributes to expressivity by inducing chaotic dynamics.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by JSPS KAKENHI, Japan (Nos. 17K16365, 19K12235, 20H04258 and 20K19882). This study was also partially supported by the JST Strategic Basic Research Programs (Symbiotic Interaction: Creation and Development of Core Technologies Interfacing Human and Information Environments, Japan, CREST Grant Number JPMJCR17A4). This paper is based on results obtained from a project, JPNP16007, commissioned by the New Energy and Industrial Technology Development Organization (NEDO).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.neunet.2020.12.020.

References

Albus, J. S. (1971). A theory of cerebellar function. *Mathematical Biosciences*, *10*(1), 25–61.
Badura, A., & De Zeeuw, C. I. (2017). Cerebellar granule cells: Dense, rich and evolving representations. *Current Biology*, *27*(11), R415–R418.
Bertschinger, N., & Natschläger, T. (2004). Real-time computation at the edge of chaos in recurrent neural networks. *Neural Computation*, *16*(7), 1413–1436.
Boyd, S., & Chua, L. (1985). Fading memory and the problem of approximating nonlinear operators with volterra series. *IEEE Transactions on Circuits and Systems*, *32*(11), 1150–1161.
Bullock, D., Fiala, J. C., & Grossberg, S. (1994). A neural model of timed response learning in the cerebellum. *Neural Networks*, *7*(6), 1101–1114.
Buonomano, D. V., & Mauk, M. (1994). Neural network model of the cerebellum: Temporal discrimination and the timing of motor responses.. *Neural Computation*, *6*(1), 38–55.
Dugué, G. P., Brunel, N., Hakim, V., Schwartz, E., Chat, M., Lévesque, M., Courtemanche, R., Léna, C., & Dieudonné, S. (2009). Electrical coupling mediates tunable low-frequency oscillations and resonance in the cerebellar golgi cell network. *Neuron*, *61*(1), 126–139.
Engelken, R., Wolf, F., & Abbott, L. F. (2020). Lyapunov spectra of chaotic recurrent neural networks. arXiv, 2006.02427.
Forti, L., Cesana, E., Mapelli, J., & D'Angelo, E. (2006). Ionic mechanisms of autorhythmic firing in rat cerebellar golgi cells. *The Journal of Physiology*, *574*(3), 711–729.
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.
Green, J. T., & Steinmetz, J. E. (2005). Purkinje cell activity in the cerebellar anterior lobe after rabbit eyeblink conditioning. *Learning & Memory*, *12*(3), 260–269.
Hayase, Y., & Ohta, T. (1998). Sierpinski gasket in a reaction-diffusion system. *Physical Review Letters*, *81*, 1726–1729.
Hayase, Y., & Ohta, T. (2000). Self-replicating pulses and sierpinski gaskets in excitable media. *Physical Review E*, *62*, 5998–6003.
Hoang, H., Lang, E. J., Hirata, Y., Tokuda, I. T., Aihara, K., Toyama, K., Kawato, M., & Schweighofer, N. (2020). Electrical coupling controls dimensionality and chaotic firing of inferior olive neurons. *PLoS Computational Biology*, *16*(7), 1–26.
Holtzman, T., Rajapaksa, T., Mostofi, A., & Edgley, S. A. (2006). Different responses of rat cerebellar purkinje cells and golgi cells evoked by widespread convergent sensory inputs. *The Journal of Physiology*, *574*(2), 491–507.
Ito, M. (1970). Neurophysiological aspects of the cerebellar motor control system. *International Journal of Neurology*, *7*, 162–176.
Ito, M. (1984). *The cerebellum and neural control*. New York: Raven Press.
Jaeger, H. (2001). *The "echo state" approach to analysing and training recurrent neural networks: GMD report 148*, German National Research Institute for Computer Science.
Kandel, E., Schwartz, J., & Jessell, T. (2013). *Principles of neural science*. The McGraw-Hill Companies, Inc..
Kaplan, J. L., & Yorke, J. A. (1979). Chaotic behavior of multidimensional difference equations. In H. O. Peitgen, & H. O. Walther (Eds.), *Functional differential equations and approximation of fixed points* (pp. 204–227). Berlin, Heidelberg: Springer Berlin Heidelberg.
Katori, Y., Lang, E. J., Onizuka, M., Mitsuo, K., & Aihara, K. (2010). Quantitative modeling of spatio-temporal dynamics of inferior olive neurons with a simple conductance-based model. *International Journal of Bifurcation and Chaos*, *20*(03), 583–603.
Kawai, Y., Park, J., & Asada, M. (2019). A small-world topology enhances the echo state property and signal propagation in reservoir computing. *Neural Networks*, *112*, 15–23.
Kawato, M., Furukawa, K., & Suzuki, R. (1987). A hierarchical neural-network model for control and learning of voluntary movement. *Biological Cybernetics*, *57*(3), 169–185.
Kawato, M., Kuroda, S., & Schweighofer, N. (2011). Cerebellar supervised learning revisited: biophysical modeling and degrees-of-freedom control. *Current Opinion in Neurobiology*, *21*(5), 791–800.
Kehoe, E. J., Horne, P. S., & Horne, A. J. (1993). Discrimination learning using different CS-US intervals in classical conditioning of the rabbit's nictitating membrane response. *Psychobiology*, *21*(4), 277–285.
Korbo, L., Andersen, B. B., Ladefoged, O., & Møller, A. (1993). Total numbers of various cell types in rat cerebellar cortex estimated using an unbiased stereological method. *Brain Research*, *609*(1), 262–268.

Kotani, S., Kawahara, S., & Kirino, Y. (2003). Trace eyeblink conditioning in decerebrate guinea pigs. *European Journal of Neuroscience*, *17*(7), 1445–1454.
Laje, R., & Buonomano, D. V. (2013). Robust timing and motor patterns by taming chaos in recurrent neural networks. *Nature Neuroscience*, *16*(7), 925–933.
Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. In *Proceedings of the IEEE* (pp. 2278–2324).
Li, W.-K., Hausknecht, M. J., Stone, P., & Mauk, M. D. (2013). Using a million cell simulation of the cerebellum: Network scaling and task generality. *Neural Networks*, *47*, 95–102.
Maass, W., Natschläger, T., & Markram, H. (2002). Real-time computing without stable states: A new framework for neural computation based on perturbations. *Neural Computation*, *14*(11), 2531–2560.
Mandelbrot, B. B. (1983). *The fractal geometry of nature* (3rd ed.). New York: W. H. Freeman and Comp..
Marr, D. (1969). A theory of cerebellar cortex. *The Journal of Physiology*, *202*(2), 437–470.
Medina, J. F., Garcia, K. S., Nores, W. L., Taylor, N. M., & Mauk, M. D. (2000). Timing mechanisms in the cerebellum: Testing predictions of a large-scale computer simulation. *Journal of Neuroscience*, *20*(14), 5516–5525.
Morán, A., Frasser, C. F., Roca, M., & Rossellló, J. L. (2020). Energy-efficient pattern recognition hardware with elementary cellular automata. *IEEE Transactions on Computers*, *69*(3), 392–401.
Natschläger, T., Bertschinger, N., & Legenstein, R. (2005). At the edge of chaos: Real-time computations and self-organized criticality in recurrent neural networks. In *Advances in neural information processing systems* (pp. 145–152). MIT Press.
Raymond, J. L., & Medina, J. F. (2018). Computational principles of supervised learning in the cerebellum. *Annual Review of Neuroscience*, *41*(1), 233–253.
Rössert, C., Dean, P., & Porrill, J. (2015). At the edge of chaos: How cerebellar granular layer network dynamics can provide the basis for temporal filters. *PLoS Computational Biology*, *11*(10), 1–28.
Ruigrok, T. J., Sillitoe, R. V., & Voogd, J. (2015). Chapter 9 - cerebellum and cerebellar connections. In G. Paxinos (Ed.), *The rat nervous system (fourth edition)* (4th ed.). (pp. 133–205). San Diego: Academic Press.
Schweighofer, N., Doya, K., Fukai, H., Chiron, J. V., Furukawa, T., & Kawato, M. (2004). Chaos may enhance information transmission in the inferior olive. *Proceedings of the National Academy of Sciences*, *101*(13), 4655–4660.
Schweighofer, N., Lang, E., & Kawato, M. (2013). Role of the olivo-cerebellar complex in motor learning and control. *Frontiers in Neural Circuits*, *7*, 94.
Shimada, I., & Nagashima, T. (1979). A numerical approach to ergodic problem of dissipative dynamical systems. *Progress of Theoretical Physics*, *61*(6), 1605–1616.
Solinas, S., Forti, L., Cesana, E., Mapelli, J., De Schutter, E., & D'Angelo, E. (2007). Computational reconstruction of pacemaking and intrinsic electroresponsiveness in cerebellar golgi cells. *Frontiers in Cellular Neuroscience*, *1*, 2.
Sompolinsky, H., Crisanti, A., & Sommers, H. J. (1988). Chaos in random neural networks. *Physical Review Letters*, *61*, 259–262.
Sudhakar, S. K., Hong, S., Raikov, I., Publio, R., Lang, C., Close, T., Guo, D., Negrello, M., & De Schutter, E. (2017). Spatiotemporal network coding of physiological mossy fiber inputs by the cerebellar granular layer. *PLoS Computational Biology*, *13*(9), 1–35, e1005754.
Sussillo, D., & Abbott, L. F. (2009). Generating coherent patterns of activity from chaotic neural networks. *Neuron*, *63*(4), 544–557.
Tadokoro, S., Yamaguti, Y., Fujii, H., & Tsuda, I. (2011). Transitory behaviors in diffusively coupled nonlinear oscillators. *Cognitive Neurodynamics*, *5*(1), 1–12.
Tanaka, G., Yamane, T., Héroux, J. B., Nakane, R., Kanazawa, N., Takeda, S., Numata, H., Nakano, D., & Hirose, A. (2019). Recent advances in physical reservoir computing: A review. *Neural Networks*, *115*, 100–123.
Thompson, R. F. (2005). In search of memory traces. *Annual Review of Psychology*, *56*(1), 1–23.
Tokuda, I. T., Han, C. E., Aihara, K., Kawato, M., & Schweighofer, N. (2010). The role of chaotic resonance in cerebellar learning. *Neural Networks*, *23*(7), 836–842.
Tokuda, I. T., Hoang, H., & Kawato, M. (2017). New insights into olivo-cerebellar circuits for learning from a small training sample. *Current Opinion in Neurobiology*, *46*, 58–67.
Tokuda, I. T., Hoang, H., Schweighofer, N., & Kawato, M. (2013). Adaptive coupling of inferior olive neurons in cerebellar learning. *Neural Networks*, *47*, 42–50.
Tokuda, K., Katori, Y., & Aihara, K. (2019). Chaotic dynamics as a mechanism of rapid transition of hippocampal local field activity between theta and non-theta states. *Chaos. An Interdisciplinary Journal of Nonlinear Science*, *29*(11), Article 113115.
Tsuda, I., Fujii, H., Tadokoro, S., Yasuoka, T., & Yamaguchi, Y. (2004). Chaotic itinerancy as a mechanism of irregular changes between synchronization and desynchronization in a neural network. *Journal of Integrative Neuroscience*, *3*(02), 159–182.
Turing, A. M. (1952). The chemical basis of morphogenesis. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *237*(641), 37–72.
Vervaeke, K., Lörincz, A., Gleeson, P., Farinella, M., Nusser, Z., & Silver, R. A. (2010). Rapid desynchronization of an electrically coupled interneuron network with sparse excitatory synaptic input. *Neuron*, *67*(3), 435–451.
Watanabe, A. (1958). The interaction of electrical activity among neurons of lobster cardiac ganglion. *The Japanese Journal of Physiology*, *8*, 305–318.
Wolfram, S. (1994). *Cellular automata and complexity*. Boca Raton: CRC Press.
Wolpert, D. M., Miall, R., & Kawato, M. (1998). Internal models in the cerebellum. *Trends in Cognitive Sciences*, *2*(9), 338–347.
Yamada, T., & Kuramoto, Y. (1976). A reduced model showing chemical turbulence. *Progress of Theoretical Physics*, *56*(2), 681–683.
Yamazaki, T., & Igarashi, J. (2013). Realtime cerebellum: A large-scale spiking network model of the cerebellum that runs in realtime using a graphics processing unit. *Neural Networks*, *47*, 103–111.
Yamazaki, T., & Nagao, S. (2012). A computational mechanism for unified gain and timing control in the cerebellum. *PLoS One*, *7*(3), 1–12, e33319.
Yamazaki, T., & Tanaka, S. (2007). The cerebellum as a liquid state machine. *Neural Networks*, *20*(3), 290–297.
Yildiz, I. B., Jaeger, H., & Kiebel, S. J. (2012). Re-visiting the echo state property. *Neural Networks*, *35*, 1–9.
Yilmaz, O. (2015). Machine learning using cellular automata based feature expansion and reservoir computing.. *Journal of Cellular Automata*, *10*, 435–472.