Macromagnetic Simulation for Reservoir Computing Utilizing Spin Dynamics in Magnetic Tunnel Junctions

Taishi Furuta,[1] Keisuke Fujii,[2,3,*] Kohei Nakajima,[3,4] Sumito Tsunegi,[5] Hitoshi Kubota,[5] Yoshishige Suzuki,[1,6] and Shinji Miwa[1,6,7,†]

[1] *Graduate School of Engineering Science, Osaka University, Toyonaka, Osaka 560-8531, Japan*
[2] *Graduate School of Faculty of Science, Kyoto University, Sakyo, Kyoto 606-8502, Japan*
[3] *PRESTO, Japan Science and Technology Agency (JST), Kawaguchi, Saitama 332-0012, Japan*
[4] *Graduate School of Information Science and Technology, The University of Tokyo, Bunkyo, Tokyo 113-8656, Japan*
[5] *National Institute of Advanced Industrial Science and Technology (AIST), Spintronics Research Center, Tsukuba, Ibaraki 305-8568, Japan*
[6] *Center for Spintronics Research Network, Osaka University, Toyonaka, Osaka 560-8531, Japan*
[7] *The Institute for Solid State Physics, The University of Tokyo, Kashiwa, Chiba 277-8581, Japan*

(Received 25 May 2018; revised manuscript received 21 August 2018; published 27 September 2018)

The figures-of-merit for reservoir computing (RC), using spintronics devices called magnetic tunnel junctions (MTJs), are evaluated. RC is a type of recurrent neural network (RNN). The input information is stored in certain parts of the reservoir and computation can be performed by optimizing a linear transform matrix for the output. While all the network characteristics should be controlled in a general RNN, such optimization is not necessary for RC. The reservoir only has to possess a nonlinear response with memory effect. In this paper, macromagnetic simulation is conducted for the spin dynamics in MTJs for RC. It is determined that the MTJ system possesses the memory effect and nonlinearity required for RC. With RC using 5–7 MTJs, high performance can be obtained, similar to an echo-state network with 20–30 nodes, even if there are no magnetic and/or electrical interactions between the magnetizations.

DOI: 10.1103/PhysRevApplied.10.034063

I. INTRODUCTION

The magnetization direction of ferromagnetic metallic film is determined by the magnetic anisotropy energy, which causes nonvolatility. This property can be used for magnetic random access memory devices [1]. In magnetic tunnel junction (MTJ) devices consisting of ferromagnetic and dielectric thin films, the magnetization direction in the ferromagnet can be detected by the change in device resistance originating from the tunneling magnetoresistance (TMR) effect [2–5]. Moreover, the magnetization direction can be electrically controlled by the spin torque [6–9]. Therefore, MTJ devices are suitable for constructing nonvolatile high-density memory devices. In addition to a *long-term* memory effect, the magnetization precessional dynamics appear to possess a *short-term* memory effect with nonlinear behavior. Such additional magnetization dynamics properties may be suitable for computation using MTJ devices.

The recurrent neural network (RNN) [10,11] is a machine learning method. It is a mathematical model, which emulates the nerve system in the human brain. The RNN concept is depicted in Fig. 1(a). The model consists of three layers, input, middle (node), and output. In the RNN, the information of the middle layer recursively propagates in itself. The middle-layer state is determined by the present input and past middle-layer state, i.e., the middle layer in the RNN possesses the memory effect. All the weight matrices for the input ($\mathbf{W}_{\text{in}}$), middle ($\mathbf{W}$), and output ($\mathbf{W}_{\text{out}}$) should be precisely trained to obtain the desired output. However, when the middle layer has sufficient memory effect and nonlinearity, it is feasible to perform computation by optimizing only the output matrix ($\mathbf{W}_{\text{out}}$). This type of simple RNN is called reservoir computing (RC) [12–14]. In RC, as system training is simple, it is easy to construct large-scale systems. The RC concept is depicted in Fig. 1(b). It has been reported that RC can be implemented in real physical systems, such as atomic switches [15–18], optoelectronic architecture [19–21], and the mechanical bodies of soft and compliant robots [22–24]. While it is possible to perform RC with such classical systems, RC using quantum dynamics can show higher figures-of-merit [25]. Recently, voice recognition by RC using an MTJ [26] was reported; however, the figures-of-merit for RC using MTJs are not quantitatively

*fujii@qi.t.u-tokyo.ac.jp
†miwa@issp.u-tokyo.ac.jp

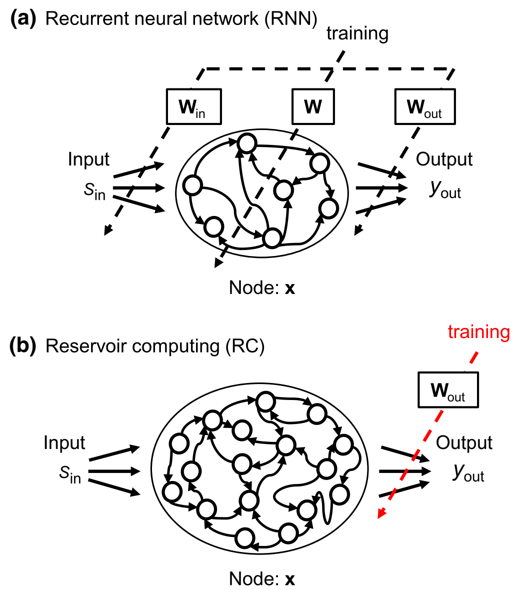

FIG. 1. Concept of (a) recurrent neural network (RNN) and (b) reservoir computing (RC).

understood. In this paper, we report the quantitative analysis of the figures-of-merit for RC [27,28] using MTJ devices. We employ macromagnetic simulation for the study.

II. METHODS

A. Reservoir computing

The RC method is as follows. A Boolean-type input $s_{in}(T)$ is employed. T is an integer variable that represents time. $s_{in}(T)$ randomly assumes "0" or "1" in every time step:

$$s_{\rm in}(T) = 0 \text{ or } 1. \tag{1}$$

The middle-layer state is defined as a node vector $\mathbf{x}(T)$ consisting of N elements:

$$\mathbf{x}(T) = \begin{pmatrix} x_1(T) \\ \vdots \\ x_N(T) \end{pmatrix}. \tag{2}$$

The time evolution of the node is determined from the input at the present time and the past state of the node:

$$\mathbf{x}(T+1) = f\,[\mathbf{x}(T), s_{\rm in}(T+1)]. \tag{3}$$

Here, f is a function for time evolution. Then the output $y_{out}(T)$ is defined as the inner product of the time-independent weight vector $\mathbf{W}_{out}$, and the node vector $\mathbf{x}(T)$:

$$y_{\rm out}(T) = \sum_{i=1}^{N} W_i x_i(T) = \mathbf{W}_{\rm out}\mathbf{x}(T) = (W_1 \cdots W_N)\begin{pmatrix} x_1(T) \\ \vdots \\ x_N(T) \end{pmatrix}, \tag{4}$$

$$\mathbf{W}_{\rm out} = (W_1 \cdots W_N). \tag{5}$$

The training data $y_{train}(T)$ is prepared to optimize the system. $\mathbf{W}_{out}$ is determined for $y_{out}(T)$, reproducing $y_{train}(T)$, and is selected to minimize the mean squared error (*MSE*) between $y_{out}(T)$ and $y_{train}(T)$. The *MSE* is expressed as follows:

$$MSE = \frac{1}{L}\sum_{T=1}^{L}[y_{\rm train}(T) - y_{\rm out}(T)]^2 = \frac{1}{L}\sum_{T=1}^{L}[y_{\rm train}(T) - \mathbf{W}_{\rm out}\mathbf{x}(T)]^2. \tag{6}$$

$\mathbf{W}_{out}$ is optimized using L time steps. In this paper, L is 2000. A pseudoinverse matrix $\mathbf{x}^{-1}$ is used for optimization.

$$\mathbf{W}^t_{\rm out} = \mathbf{x}^{-1}\mathbf{y}_{\rm train}. \tag{7}$$

The optimization of the output weight vector $\mathbf{W}_{out}$ is called *learning*. In this paper, a time-independent constant is added to $x_{N+1}(T)$ as a bias term, in addition to $x_1(T)$ to $x_N(T)$.

B. Figures-of-merit for reservoir computing

In this paper, two types of tasks are employed for learning. One is a short-term memory (STM) task [28] for characterizing the memory effect in the system. The training data for the short-term memory task is expressed as follows:

$$y_{\rm train,\ STM}(T, T_{\rm delay}) = s_{\rm in}(T - T_{\rm delay}). \tag{8}$$

Here, $s_{in}(T)$ are random pulses, which are described later. It is feasible to obtain a finite memory effect, even if the system is completely linear. Therefore, we need another task to characterize the computing capability. In this paper, we additionally employ the parity check (PC) task [28]. The training data for the PC task requests the parity of the input sum. PC is used for characterizing the type of nonlinearity in the system and is expressed as follows:

$$y_{\rm train,\ PC}(T, T_{\rm delay}) = s_{\rm in}(T - T_{\rm delay}) + s_{\rm in}(T - T_{\rm delay} + 1) + \cdots + s_{\rm in}(T) \pmod 2. \tag{9}$$

After learning with the training data, the correlation between the output and training data is evaluated using the following equation:

$$\mathrm{Cor}(T_{\mathrm{delay}})^2 = \frac{\mathrm{Cov}[y_{\mathrm{train}}(T, T_{\mathrm{delay}}), y_{\mathrm{out}}(T)]^2}{\mathrm{Var}[y_{\mathrm{train}}(T, T_{\mathrm{delay}})]\mathrm{Var}[y_{\mathrm{out}}(T)]}. \quad (10)$$

Here, Cor, Cov, and Var are the correlation, covariance, and variance, respectively. In this paper, Cor^2 is evaluated during 500 time steps, after learning with 2000 time steps. Cor^2 assumes values from 0 to 1, with a larger Cor^2 indicating better learning. Moreover, the capacity, C, is defined as the integration of Cor^2, which can be used for evaluating the figures-of-merit for RC.

$$\text{Capacity}: \quad C \equiv \sum_{T_{\mathrm{delay}}=1}^{T_{\mathrm{delay,max}}} \mathrm{Cor}(T_{\mathrm{delay}})^2. \quad (11)$$

$T_{\mathrm{delay,\,max}}$ should be sufficiently large. In our calculation, Cor is always less than 0.01, when T_{delay} is more than 10. Therefore, we set $T_{\mathrm{delay,\,max}} = 30$. In this paper, we define C_{STM} as the capacity for short-term memory and C_{PC} as the capacity for PC.

C. Magnetic tunnel junction system

Figure 2 shows the schematics of the RC simulation using MTJs. An MTJ contains an insulating tunneling barrier layer with two ferromagnetic layers. For ferromagnetic layer 1 called the reference layer, the magnetization direction is designed to be fixed. This can be done with the exchange bias effect using antiferromagnetic materials, such as PtMn and IrMn [29], or magnetic anisotropy energy. In this study, the magnetization direction of layer 1 is fixed perpendicular to the film plane. For ferromagnetic layer 2 called the free layer, the magnetization direction is not fixed and can be controlled by the current [6,7]- or voltage [8]-driven spin torque. The MTJ device resistance reflects the magnetization direction of $\mathbf{s}_2$.

The spin dynamics in ferromagnetic layer 2 follows the Landau-Lifshitz-Gilbert (LLG) equation with spin-transfer torque [30], where a thermal fluctuation in ferromagnetic layers [31] is not included:

$$\frac{d\mathbf{s}_2}{dt} = -\gamma_0 \mathbf{s}_2 \times \mathbf{H}_{\mathrm{eff}} - \alpha \mathbf{s}_2 \times \frac{d\mathbf{s}_2}{dt} + \frac{P}{1+P^2 \mathbf{s}_1 \cdot \mathbf{s}_2} \frac{I}{-e} \frac{\hbar}{2} \mathbf{s}_2 \times (\mathbf{s}_1 \times \mathbf{s}_2). \quad (12)$$

Here, $\mathbf{s}_1$ and $\mathbf{s}_2$ represent the unit spin vectors for ferromagnetic layers-1 and -2, respectively. γ_0 (<0) is the gyro magnetic ratio. α is the Gilbert damping constant. P is the spin polarization in the vicinity of the Fermi level in the ferromagnetic layers. I ($=V_{\mathrm{in}}/R$) is the electric current, where V_{in} and R are the input voltage and device resistance of the MTJ, respectively. $\mathbf{H}_{\mathrm{eff}}$ is the effective magnetic field in $\mathbf{s}_2$:

$$\mathbf{H}_{\mathrm{eff}} = -\frac{1}{\gamma_0} \nabla U. \quad (13)$$

Here, U is the magnetization energy for ferromagnetic layer 2, which includes the external magnetic field $\mathbf{H}_{\mathrm{ext}}$ and the magnetic anisotropy tensor $\hat{\mathbf{H}}_{\mathrm{ani}}$:

$$\begin{aligned} U &= \mu_0 M_S A \left(\mathbf{H}_{\mathrm{ext}} \cdot \mathbf{s}_2 + \frac{1}{2} \mathbf{s}_2^t \cdot \hat{\mathbf{H}}_{\mathrm{ani}} \cdot \mathbf{s}_2 \right) \\ &= \frac{1}{2} \mu_0 M_S A\, \mathbf{s}_2^t \cdot \hat{\mathbf{H}}_{\mathrm{ani}} \cdot \mathbf{s}_2 \; (\because \mathbf{H}_{\mathrm{ext}} = \mathbf{0}), \end{aligned} \quad (14)$$

$$\hat{\mathbf{H}}_{\mathrm{ani}} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & H_{\mathrm{azz}} \end{pmatrix}. \quad (15)$$

Here, μ_0, M_S, and A are the magnetic permeability in a vacuum, saturation magnetization, and volume of the ferromagnetic layer 2, respectively. In the simulation, the external magnetic field is not applied. We assume uniaxial anisotropy perpendicular to the film plane. Here, $H_{\mathrm{azz}} > 0$ ($H_{\mathrm{azz}} < 0$) shows in-plane (perpendicular) magnetic anisotropy. The device resistance of the MTJ (R) varies as a function of the relative angle between the spins

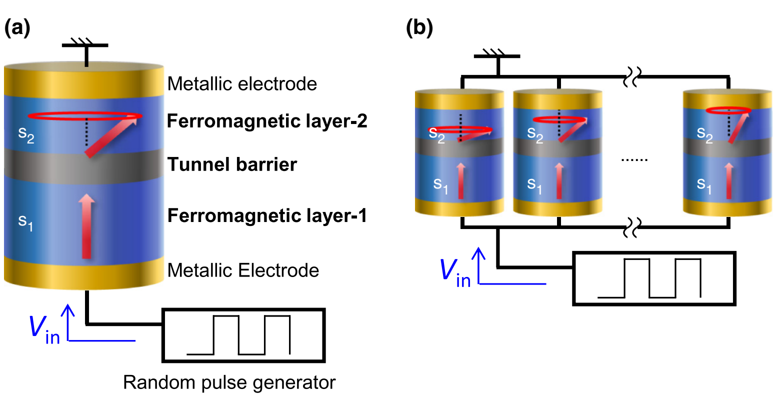

FIG. 2. (a) Schematic of a RC system using the spin dynamics in a magnetic tunnel junction (MTJ) and (b) system with multiple MTJs. In the MTJs, the spin direction of the ferromagnetic layer 2 ($\mathbf{s}_2$) can be controlled by the input bias voltage V_{in}, whereas that of the ferromagnetic layer 1 ($\mathbf{s}_1$) is fixed.

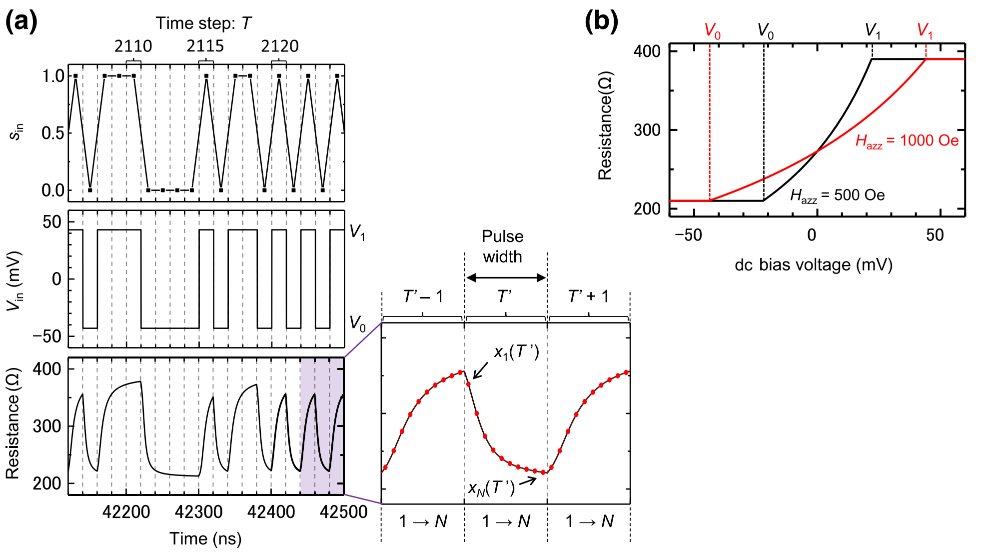

FIG. 3. (a) Input $s_{\text{in}}(T)$, input bias voltage to the MTJ device V_{in}, and MTJ device resistance as a function of time. Typical characteristics during the learning and evaluation processes. We define the virtual nodes $[x_1(T)$ to $x_N(T)]$ as shown in the inset and the (b) MTJ device resistance as a function of the static input dc bias voltage. The black and red plots indicate the resistances when the values of the uniaxial anisotropy fields are 500 and 1000 Oe, respectively. V_0 and V_1 are voltages that render the device resistance constant.

in the free and pinned layers:

$$R = \frac{R_{\text{AP}} R_{\text{P}}}{(R_{\text{AP}} + R_{\text{P}}) + (R_{\text{AP}} - R_{\text{P}})(\mathbf{s}_1 \cdot \mathbf{s}_2)}. \tag{16}$$

R_{AP} and R_{P} are the resistances when $\mathbf{s}_1$ and $\mathbf{s}_2$ are parallel and antiparallel, respectively. The time evolution of the MTJ resistance is characterized by sequential calculation using the fourth Runge-Kutta method. For evaluating the STM and PC capacities, an input pulse voltage, V_{in}, corresponding to the computational input, $s_{\text{in}}(T)$, is applied to the MTJs, as depicted in Fig. 3(a). Figures 2(a) and 2(b) display the schematics of circuits with single and multiple MTJs, respectively. In this paper, the physical parameters listed in Table I are employed. It almost follows our previous experimental research [32].

D. Reference calculation with echo-state network

Additionally, an echo-state network [13] is introduced for comparison with the system using MTJs, where the following function is employed for Eq. (3):

$$\mathbf{x}(T) = \tanh[\mathbf{W}\mathbf{x}(T-1) + \mathbf{W}_{\text{in}} s_{\text{in}}(T)]. \tag{17}$$

The tanh function is used for component-wise projection. $\mathbf{W}$ and $\mathbf{W}_{\text{in}}$ are matrices whose components are time-independent random values from (-1) to 1. We normalize by dividing each component of $\mathbf{W}$ by the spectral radius, r, obtained by singular value decomposition [14].

III. RESULTS AND DISCUSSION

Figure 3(a) depicts an example of the MTJ device resistance under an input voltage, V_{in}. We employ a pulse voltage with binary values of V_0 $(=-44$ mV) and V_1 $(=+44$ mV) as V_{in}. These binary values of V_{in} correspond to 0 and 1 in $s_{\text{in}}(T)$, respectively, in the RC learning and evaluation processes [Eq. (1)]. The pulse width [20 ns in Fig. 3(a), for instance] corresponds to the discrete unit time step T. Because the device resistance is scalar, the node dimension is only one. However, the number of nodes can be increased by employing virtual nodes [15,33]. As shown in the inset of Fig. 3(a), the virtual nodes x_1 to x_N are defined; these virtual nodes are further defined as a node vector $\mathbf{x}(T')$.

Figure 3(b) depicts the dc bias voltage dependence of the static MTJ device resistance. Under a dc bias voltage, the MTJ device resistance is collected after the spin dynamics are damped. Under a positive bias voltage, the spin-polarized current flows from the free layer $\mathbf{s}_2$, to the pinned layer $\mathbf{s}_1$. Then the spin-transfer effect induces auto-oscillation [34,35] in $\mathbf{s}_2$. The relative magnetization angle between $\mathbf{s}_1$ and $\mathbf{s}_2$ increases, and an antiparallel-like magnetization configuration is realized. Therefore, the device resistance increases when a positive bias voltage is applied. Under a negative bias voltage, a parallel-like magnetization configuration is induced, and the device resistance

TABLE I. Physical parameters' set of ferromagnetic layer 2 for RC with a single MTJ [Fig. 2(a)].

Parameter	Value
Gilbert damping constant (layer 2): α	0.009
Uniaxial anisotropy (layer 2): H_{azz}	1000 Oe
Saturation magnetization (layer 2): M_{S}	1375 emu/cc
Volume (layer 2): A	23 500 nm^3 (ϕ122 nm × 2 nm)
Resistance in parallel: R_{P}	210 Ω
Resistance in antiparallel: R_{AP}	390 Ω

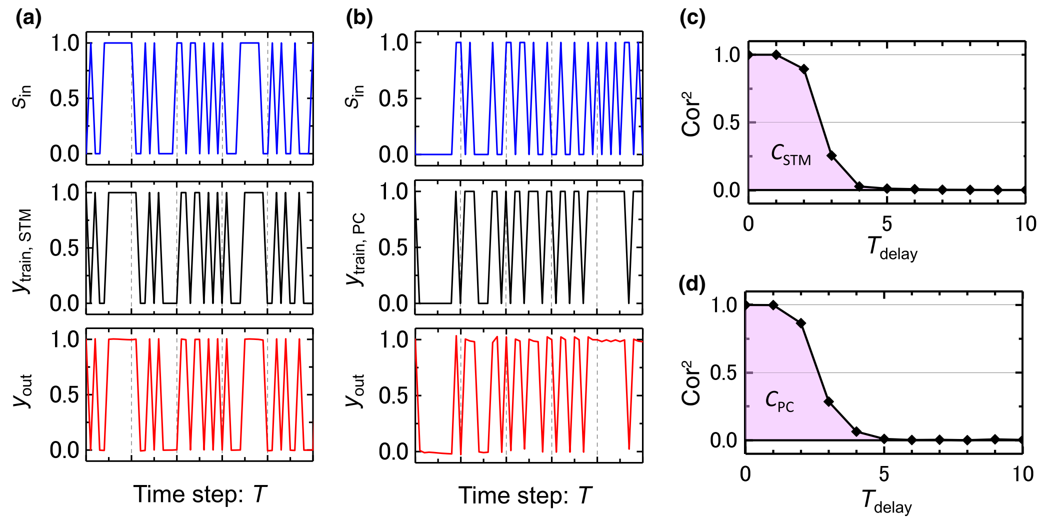

FIG. 4. (a) Input s_{in}, output for training $y_{train,\,STM}$ [Eq. (8) with $T_{delay} = 1$], and trained output y_{out} for evaluating the short-term memory task, (b) Input s_{in}, output for training $y_{train,\,PC}$, [Eq. (9) with $T_{delay} = 1$], and trained output y_{out} for evaluating the parity check task, (c) Correlation using Eq. (10); the integrated values are defined as the short-term memory capacity (C_{STM}), (d) Correlation using Eq. (10); the integrated values are defined as the parity check capacity (C_{PC}). The input-voltage pulse width = 20 ns and the number of virtual nodes $N = 50$.

decreases. For the input pulse voltage in Fig. 3(a), the binary values V_0 and V_1 are defined as voltages that render the device resistance constant. As shown in Fig. 3(b), V_0 and V_1 vary as a function of the uniaxial magnetic anisotropy H_{azz}.

A. Figures-of-merit for RC using a single MTJ

In this section, we present the figures-of-merit for RC using a single MTJ device. The uniaxial magnetic anisotropy of the free layer $\mathbf{s}_2$ is fixed as $H_{azz} = 1000$ Oe. Here, the positive value of H_{azz} shows the magnetic cell in MTJ is in-plane magnetized. Figure 4 shows the simulated data for evaluating the short-term memory and parity check capacities for a single MTJ. In Fig. 4, the input-voltage pulse width is 20 ns and the number of virtual nodes, N, is 50. Figure 4(a) shows the typical simulation results for an input $s_{in}(T)$, training data for the short-term memory task $y_{train,\,STM}(T)$, and trained output $y_{out}(T)$ as a function of the time step. Similarly, Fig. 4(b) shows the input $s_{in}(T)$, training data for PC task $y_{train,\,PC}(T)$, and trained output $y_{out}(T)$. Here, training data for the short-term memory task $y_{train,\,STM}(T)$ and PC task $y_{train,\,PC}(T)$ are defined using Eqs. (8) and (9) at $T_{delay} = 1$, respectively. The output is calculated using the simulated MTJ resistance (see Fig. 3) and W_{out} using Eq. (4). W_{out} is trivially calculated using the definitions given by Eqs. (5)–(7).

Figures 4(c) and 4(d) depict the correlations [Eq. (10)] between y_{out} and the training data as a function of T_{delay}. We use $y_{train,\,STM}$ as the training data for short-term memory and $y_{train,\,PC}$ for the parity check. C_{STM} and C_{PC} are

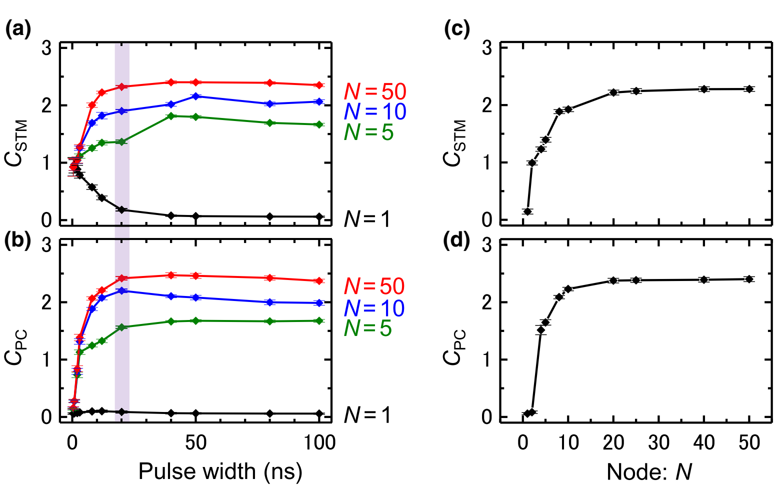

FIG. 5. Results of RC using single MTJ. (a) Short-term memory capacity (C_{STM}) as a function of the input-voltage pulse width, (b) Parity check capacity (C_{PC}) as a function of the input-voltage pulse width; N is number of virtual nodes in the MTJ, (c) C_{STM}, and (d) C_{PC} as functions of the virtual-node number, where the input-voltage pulse width is fixed to 20 ns.

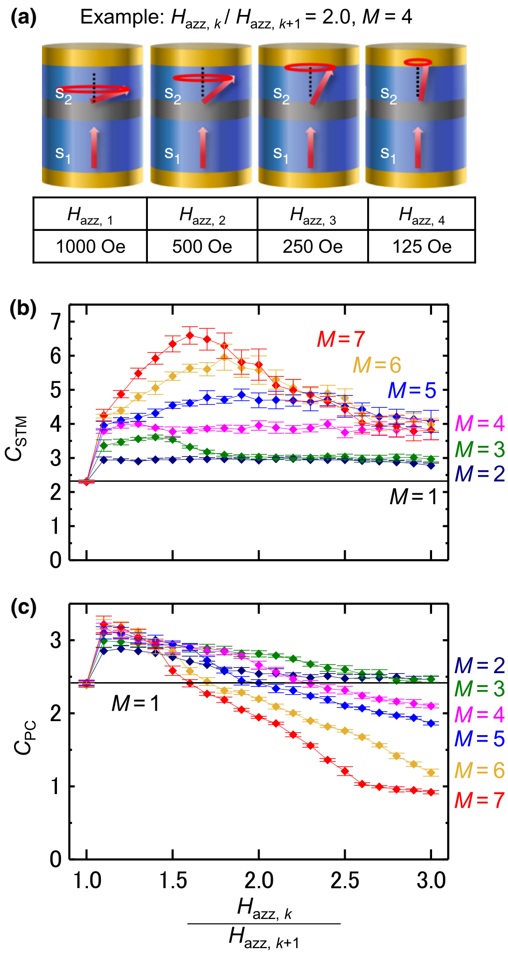

FIG. 6. Results of RC using multiple MTJs. (a) Example of a parameter set for multiple MTJs, when the ratio of the uniaxial magnetic anisotropy in each MTJ ($H_{azz,\,k}/H_{azz,\,k+1}$) = 2 and the number of MTJs (M) = 4, (b) Short-term memory capacity (C_{STM}), and (c) parity check capacity (C_{PC}) as functions of $H_{azz,\,k}/H_{azz,\,k+1}$. The input-voltage pulse width = 20 ns and the virtual nodes for each MTJ (N) = 50.

defined as the numerical integration of the correlation and the capacity using training data for the short-term memory and parity check, respectively.

Figure 5 shows the C_{STM} and $C_{PC,}$ respectively, as functions of the input-voltage pulse width [Figs. 5(a) and 5(b)] and the number of virtual nodes N [Figs. 5(c) and 5(d)]. From Figs. 5(a) and 5(b), both C_{STM} and C_{PC} increase as the pulse width increases. When the pulse width is greater than 20 ns, C_{STM} and C_{PC} are nearly constant because when the pulse width is less than 20 ns, the change in the magnetization direction is very small and the spin dynamics cannot work as a reservoir. In Figs. 5(c) and 5(d), the dependence of C_{STM} and C_{PC}, respectively, on the number of virtual nodes N, are displayed when the pulse width is fixed at 20 ns. From Figs. 5(c) and 5(d), we find that it is better to set the number of virtual nodes greater than 20.

TABLE II. Variation for uniaxial magnetic anisotropy for RC with multiple MTJs [Fig. 2(b)].

$H_{azz,\,k}/H_{azz,\,k+1}$	$H_{azz,\,1}$	$H_{azz,\,2}$	$H_{azz,\,3}$	...	$H_{azz,\,7}$
1.0	1000 Oe	1000 Oe	1000 Oe	...	1000 Oe
1.1	1000 Oe	909.1 Oe	826.4 Oe	...	564.5 Oe
1.2	1000 Oe	833.3 Oe	694.4 Oe	...	334.9 Oe
...	...	...	...	...	...
2.9	1000 Oe	344.8 Oe	118.9 Oe	...	1.7 Oe
3.0	1000 Oe	333.3 Oe	111.1 Oe	...	1.4 Oe

B. Figures-of-merit for RC using multiple MTJs

When multiple MTJs are employed for RC, higher figures-of-merit can be obtained. A schematic of a multiple MTJ circuit for RC is depicted in Fig. 2(b). Multiple MTJs are placed in parallel, and an identical pulse voltage is applied to all the MTJs. To construct nodes for RC, spatial multiplexing [36] is employed. The node vector $\mathbf{x}(T)$ is defined as a vector with $M \times N$ elements, where M is the number of MTJs and N is the number of virtual nodes in a MTJ:

$$\mathbf{x}_1(T) = \begin{pmatrix} x_{11}(T) \\ \vdots \\ x_{1N}(T) \end{pmatrix}, \mathbf{x}_2(T) = \begin{pmatrix} x_{21}(T) \\ \vdots \\ x_{2N}(T) \end{pmatrix}, \ldots, \mathbf{x}_M(T) = \begin{pmatrix} x_{M1}(T) \\ \vdots \\ x_{MN}(T) \end{pmatrix}$$
$$\rightarrow \mathbf{x}(T) \equiv [x_{11}(T) \cdots x_{1N}(T)\, x_{21}(T) \cdots x_{2N}(T) \cdots x_{M1}(T) \cdots x_{MN}(T)]^t. \quad (18)$$

Figure 6 shows the C_{STM} and C_{PC} with multiple MTJs. The uniaxial anisotropy $H_{azz,}$ of ferromagnetic layer 2 in each MTJ is listed in Table II. For instance, when four MTJs and $H_{azz,\,k}/H_{azz,\,k+1} = 2$ are employed, the uniaxial anisotropies of the MTJs are 1000, 500, 250, and 125 Oe, respectively, as shown in Fig. 6(a). Such variations in the anisotropies can be obtained by voltage-controlled magnetic anisotropy in the MTJs [37]. In this study, thermal fluctuation in ferromagnetic layer 2 is not included. For instance, thermal fluctuation energy at room temperature (26 meV) is negligibly small when compared to the magnetization energy from Eq. (14) (approximately 10 eV) when the magnetic anisotropy energy is $H_{azz} = 1000$ Oe. Therefore, thermal fluctuation can be comparable or less than the magnetization energy of ferromagnetic layer 2 at $H_{azz} < 3$ Oe. In such a region, simulations assuming the ground state are not very correct, and a random magnetic field to reproduce the thermal fluctuation [31] should be included in the simulation.

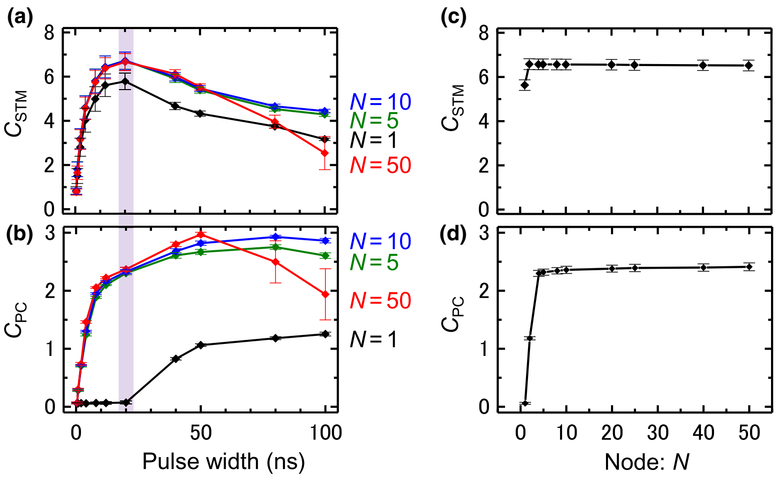

FIG. 7. Results of RC using multiple MTJs. (a) Short-term memory capacity (C_{STM}), and (b) Parity check capacity (C_{PC}) with seven MTJs ($M = 7$) as functions of the input-voltage pulse width, (c) $C_{STM,}$ and (d) C_{PC} as functions of the virtual node number of each MTJ (N) under an input-voltage pulse width of 20 ns. The uniaxial magnetic anisotropy ratio ($H_{azz,\ k}/H_{azz,\ k+1}$) = 1.6.

Similar to a single MTJ, the binary values V_0 and V_1 for the input voltage are determined as shown in Fig. 3(b). Note that V_0 and V_1 vary as a function of the uniaxial anisotropy field, and the smallest absolute values of the saturation voltages are employed as V_0 and V_1 for RC with multiple MTJs; i.e., V_0 and V_1 are determined for the MTJ with the smallest uniaxial magnetic anisotropy field. Figures 6(b) and 6(c) display the C_{STM} and C_{PC}, respectively, as functions of the anisotropy ratio $H_{azz,k}/H_{azz,k+1}$. In the simulation, the input-voltage pulse width is 20 ns and the number of virtual nodes for each MTJ is 50 for all MTJs. The maximum value of C_{STM} increases as the number of MTJs (M) increases. Because each MTJ has a different uniaxial magnetic anisotropy field H_{azz}, it has a different response speed to an external voltage/current. This variation in the response speed increases the C_{STM} of the system. On the other hand, the increase in C_{PC} is insignificant compared to that of the C_{STM}. Note that as there is no electric and/or magnetic interaction between the free layers of the MTJs, the C_{PC} is insignificant. In Fig. 6(c), when $H_{azz,k}/H_{azz,k+1}$ is large, the C_{PC} using multiple MTJs is less than that using a single MTJ. This is because the input-voltage pulse width of 20 ns is the best condition only for the parameters of a single MTJ ($H_{azz} = 1000$ Oe, $V_1 = 44$ mV, $V_0 = -44$ mV).

Figure 7 shows the C_{STM} and C_{PC} under various conditions. In Fig. 7, $H_{azz,k}/H_{azz,k+1}$ is fixed to 1.6. This is the best condition for the C_{STM} with $M = 7$. From Fig. 7(a), the C_{STM} is maximum, around a pulse width of 20 ns. When the pulse width is less than 20 ns, the change in the magnetization direction by the spin-transfer torque is too small to perform as a reservoir. When the pulse width is greater than 20 ns, the spin dynamics are almost damped during a unit time step, and such a condition is not preferable for RC. From Fig. 7(b), the best conditions for the C_{STM} and C_{PC} are not identical. This is because a relatively long pulse is required to induce nonlinearity in the spin dynamics in multiple MTJs. Figures 7(c) and 7(d) depict the N dependences of the C_{STM} and C_{PC}, respectively, when $M = 7$, $H_{azz,\ k}/H_{azz,\ k+1} = 1.6$, and the pulse width = 20 ns. When N is greater than four, both the C_{STM} and C_{PC} are nearly constant.

C. Comparison with the echo-state network

The C_{STM} and C_{PC}, using a multiple MTJ system, are summarized in Fig. 8(a); the pulse width = 20 ns and the virtual node number $N = 50$ for each MTJ. The data points from top to bottom are the data when $H_{azz,\ k}/H_{azz,\ k+1}$

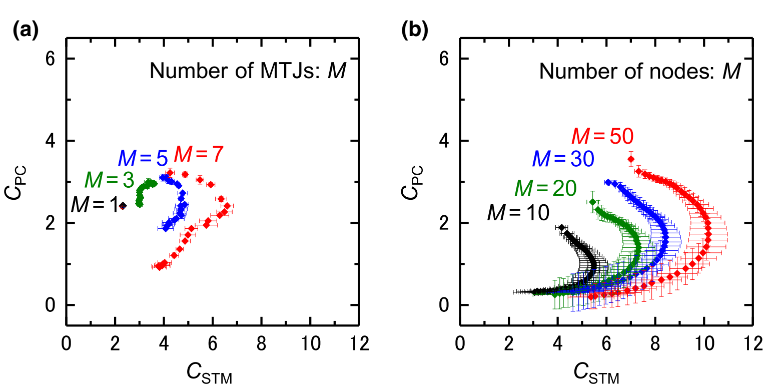

FIG. 8. Plots showing C_{STM} vs C_{PC} in the (a) MTJ system and (b) echo-state network with Eq. (17). In the MTJ system, the pulse width and virtual nodes of each MTJ (N) are fixed to 20 ns and 50, respectively.

changes from 1.1 to 3.0. The result of the echo-state network, using the tanh function, is shown in Fig. 8(b). The data points from top to bottom are the data when the spectrum radius, r, of $\mathbf{W}$ [see Eq. (17)] varies from 0.05 to 2.0. From Fig. 8, it can be observed that a high performance can be obtained for RC using 5–7 MTJs, similar to an echo-state network with 20–30 nodes. In terms of the total number of virtual nodes in the system ($M \times N$), 35 nodes (7×5) of an MTJ system are comparable to 20–30 nodes of an echo-state network [see Figs. 7(c) and 7(d)]. Although the C_{PC} increases slightly as M increases, we can obtain a large C_{PC} if there are magnetic and/or electrical interactions between the free layers in each MTJ.

IV. CONCLUSION

In this research, we demonstrate RC using the spin dynamics in MTJs. With RC using 5–7 MTJs, we can obtain a high performance similar to that of an echo-state network using tanh functions with 20–30 nodes. If there are magnetic and/or electrical interactions between the free layers in each MTJ, higher performance can be obtained.

ACKNOWLEDGMENTS

We thank E. Tamura, K. Shimose, S. Hasebe, and M. Goto of Osaka University and T. Taniguchi of AIST for the discussions. Part of this work was supported by JSPS KAKENHI (Grants No. JP18H03880, No. JP26103002, No. JP16H02211, No. JP16KT0019, No. JP15K16076, and No. JP26880010) and the Ministry of Internal Affairs and Communications. This work was also supported by JST-ERATO (Grant No. JPMJER1601), JST-CREST (Grant No. JPMJCR1673), and JST-PRESTO (Grants No. JPMJPR1668 and No. JPMJPR15E7), Japan.

[1] H. Yoda, T. Kishi, T. Nagase, M. Yoshikawa, K. Nishiyama, E. Kitagawa, T. Daibou, M. Amano, N. Shimomura, S. Takahashi, T. Kai, M. Nakayama, H. Aikawa, S. Ikegawa, M. Nagamine, J. Ozeki, S. Mizukami, M. Oogane, Y. Ando, S. Yuasa, K. Yakushiji, H. Kubota, Y. Suzuki, Y. Nakatani, T. Miyazaki, and K. Ando, High efficient spin transfer torque writing on perpendicular magnetic tunnel junctions for high density MRAMs, Curr. Appl. Phys. **10**, 87 (2010).

[2] T. Miyazaki and N. Tezuka, Giant magnetic tunneling effect in Fe/Al_2O_3/Fe junction, J. Magn. Magn. Mater. **139**, 231 (1995).

[3] J. S. Moodera, L. R. Kinder, T. M. Wong, and R. Meservey, Large Magnetoresistance at Room Temperature in Ferromagnetic Thin Film Tunnel Junctions, Phys. Rev. Lett. **74**, 3273 (1995).

[4] S. Yuasa, T. Nagahama, A. Fukushima, Y. Suzuki, and K. Ando, Giant room-temperature magnetoresistance in single-crystal Fe/MgO/Fe magnetic tunnel junctions, Nat. Mater. **3**, 868 (2004).

[5] S. S. P. Parkin, C. Kaiser, A. Panchula, P. M. Rice, B. Hughes, M. Samant, and S. Yang, Giant tunnelling magnetoresistance at room temperature with MgO (100) tunnel barriers, Nat. Mater. **3**, 862 (2004).

[6] J. A. Katine, F. J. Albert, and R. A. Buhrman, Current-Driven Magnetization Reversal and Spin-Wave Excitations in Co/Cu/Co Pillars, Phys. Rev. Lett. **84**, 3149 (2000).

[7] L. Liu, C. Pai, Y. Li, H. W. Tseng, D. C. Ralph, and R. A. Buhrman, Spin-torque switching with the giant spin Hall effect of tantalum, Science **336**, 555 (2012).

[8] Y. Shiota, T. Nozaki, F. Bonell, S. Murakami, T. Shinjo, and Y. Suzuki, Induction of coherent magnetization switching in a few atomic layers of FeCo using voltage pulses, Nat. Mater. **11**, 39 (2012).

[9] N. Locatelli, V. Cros, and J. Grollier, Spin-torque building blocks, Nat. Mater. **13**, 11 (2014).

[10] R. J. Williams and D. Zipser, A learning algorithm for continually running fully recurrent neural networks, Neural Comput. **1**, 270 (1989).

[11] P. J. Werbos, Backpropagation through time: What it does and how to do it, Proc. IEEE **78**, 1550 (1990).

[12] W. Maass, T. Natschläger, and H. Markram, Real-time computing without stable states: A new framework for neural computation based on perturbations, Neural Comput. **14**, 2531 (2002).

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

[14] D. Verstraeten, B. Schrauwen, M. D'Haene, and D. Stroobandt, An experimental unification of reservoir computing methods, Neural Netw. **20**, 391 (2007).

[15] L. Appeltant, M. C. Soriano, G. Van der Sande, J. Danckaert, S. Massar, J. Dambre, B. Schrauwen, C. R. Mirasso, and I. Fischer, Information processing using a single dynamical node as complex system, Nat. Commun. **2**, 468 (2011).

[16] A. Z. Stieg, A. V. Avizienis, H. O. Sillin, C. Martin-Olmos, M. Aono, and J. K. Gimzewski, Emergent criticality in complex turing B-type atomic switch networks, Adv. Mater. **24**, 286 (2012).

[17] H. O. Sillin, R. Aguilera, H. Shieh, A. V. Avizienis, M. Aono, A. Z. Stieg, and J. K. Gimzewski, A theoretical and experimental study of neuromorphic atomic switch networks for reservoir computing, Nanotechnology **24**, 384004 (2013).

[18] E. C. Demis, R. Aguilera, H. O. Sillin, K. Scharnhorst, E. J. Sandouk, M. Aono, A. Z. Stieg, and J. K. Gimzewski, Atomic switch networks-nanoarchitectonic design of a complex system for natural computing, Nanotechnology **26**, 204003 (2015).

[19] Y. Paquot, F. Duport, A. Smerieri, J. Dambre, B. Schrauwen, M. Haelterman, and S. Massar, Optoelectronic reservoir computing, Sci. Rep. **2**, 287 (2012).

[20] R. Martinenghi, S. Rybalko, M. Jacquot, Y. K. Chembo, and L. Larger, Photonic Nonlinear Transient Computing with Multiple-Delay Wavelength Dynamics, Phys. Rev. Lett **108**, 244101 (2012).

[21] K. Vandoorne, P. Mechet, T. Van Vaerenbergh, M. Fiers, G. Morthier, D. Verstraeten, B. Schrauwen, J. Dambre, and P. Bienstman, Experimental demonstration of reservoir

computing on a silicon photonics chip, Nat. Commun. **5**, 3541 (2014).

[22] K. Nakajima, H. Hauser, R. Kang, E. Guglielmino, D. G. Caldwell, and R. Pfeifer, A soft body as a reservoir: case studies in a dynamic model of octopus-inspired soft robotic arm, Front. Comput. Neurosci. **7**, 1 (2013).

[23] K. Nakajima, T. Li, H. Hauser, and R. Pfeifer, Exploiting short-term memory in soft body dynamics as a computational resource, J. R. Soc. Interface **11**, 20140437 (2014).

[24] K. Nakajima, H. Hauser, T. Li, and R. Pfeifer, Information processing via physical soft body, Sci. Rep. **5**, 10487 (2015).

[25] K. Fujii and K. Nakajima, Harnessing Disordered-Ensemble Quantum Dynamics for Machine Learning, Phys. Rev. Appl. **8**, 24030 (2017).

[26] J. Torrejon, M. Riou, F. A. Araujo, S. Tsunegi, G. Khalsa, D. Querlioz, P. Bortolotti, V. Cros, K. Yakushiji, A. Fukushima, H. Kubota, S. Yuasa, M. D. Stiles, and J. Grollier, Neuromorphic computing with nanoscale spintronic oscillators, Nature **547**, 428 (2017).

[27] H. Jaeger, GMD Report No 148, German National Research Institute for Computer Science, 2001.

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

[29] W. H. Meiklejohn and C. P. Bean, New magnetic anisotropy, Phys. Rev. **102**, 1413 (1956).

[30] J. C. Slonczewski, Current-driven excitation of magnetic multilayers, J. Magn. Magn. Mater. **159**, 1 (1996).

[31] W. F. Brown, Thermal fluctuations of a single-domain particle, Phys. Rev. **130**, 1677 (1963).

[32] S. Miwa, S. Ishibashi, H. Tomita, T. Nozaki, E. Tamura, K. Ando, N. Mizuochi, T. Saruya, H. Kubota, K. Yakushiji, T. Taniguchi, H. Imamura, A. Fukushima, S. Yuasa, and Y. Suzuki, Highly sensitive nanoscale spin-torque diode, Nat. Mater. **13**, 50 (2014).

[33] K. Nakajima, H. Hauser, T. Li, and R. Pfeifer, Exploiting the dynamics of soft materials for machine learning, Soft Robt. **5**, 339 (2018).

[34] S. I. Kiselev, J. C. Sankey, I. N. Krivorotov, N. C. Emley, R. J. Schoelkopf, R. A. Buhrman, and D. C. Ralph, Microwave oscillations of a nanomagnet driven by a spin-polarized current, Nature **425**, 380 (2003).

[35] D. Sameddine, U. Ebels, B. Delaët, B. Rodmacq, I. Firastrau, F. Ponthenier, M. Brunet, C. Thirion, J.-P. Michel, L. Prejbeanu-buda, M.-C. Cyrille, O. Redon, and B. Dieny, Spin-torque oscillator using a perpendicular polarizer and a planar free layer, Nat. Mater. **6**, 447 (2007).

[36] K. Nakajima, K. Fujii, M. Negoro, K. Mitarai, and M. Kitagawa, Boosting computational power through spatial multiplexing in quantum reservoir computing, arXiv:1803.04574.

[37] T. Maruyama, Y. Shiota, T. Nozaki, K. Ohta, N. Toda, M. Mizuguchi, A. A. Tulapurkar, T. Shinjo, M. Shiraishi, S. Mizukami, Y. Ando, and Y. Suzuki, Large voltage-induced magnetic anisotropy change in a few atomic layers of iron, Nat. Nanotechnol. **4**, 158 (2009).