scientific reports

OPEN

Spintronic reservoir computing without driving current or magnetic field

Tomohiro Taniguchi[1✉], Amon Ogihara[2], Yasuhiro Utsumi[2] & Sumito Tsunegi[1,3]
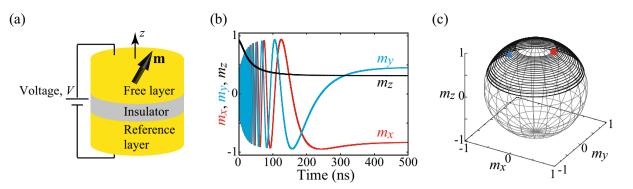
Recent studies have shown that nonlinear magnetization dynamics excited in nanostructured ferromagnets are applicable to brain-inspired computing such as physical reservoir computing. The previous works have utilized the magnetization dynamics driven by electric current and/or magnetic field. This work proposes a method to apply the magnetization dynamics driven by voltage control of magnetic anisotropy to physical reservoir computing, which will be preferable from the viewpoint of low-power consumption. The computational capabilities of benchmark tasks in single MTJ are evaluated by numerical simulation of the magnetization dynamics and found to be comparable to those of echo-state networks with more than 10 nodes.
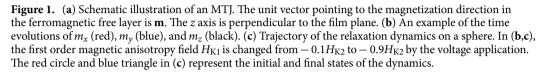
Recent development of neuromorphic computing with spintronics devices[1–4], such as pattern recognition and associative memory, has provided a bridge between condensed matter physics, nonlinear science, and information science, and become of great interest from both fundamental and practical viewpoints. In particular, an application of nonlinear magnetization dynamics in ferromagnets to physical reservoir computing[5–19] is an exciting topic[1,20–30]. Physical reservoir computing is a kind of recurrent neural network, which has recurrent interaction among large number of neurons in artificial neural network and, for example, recognizes a time sequence of the input data, such as human voice and movie, from the dynamical response in nonlinear physical systems[19]. In reservoir computing, only the weights between neurons and output are trained, whereas the weights among neurons are randomly given and fixed, and therefore, low calculation cost of training is expected. It has been shown that several kinds of physical systems, such as optical circuit[10], soft matter[12], quantum matter[15], fluid[18], and spintronics devices, can be used as reservoir for information processing[19].

In physical reservoir computing with spintronics devices, nonlinear magnetization dynamics has been excited in nanostructured ferromagnets by applying electric current and/or magnetic field. For example, spin-transfer effect[31,32] has been frequently used to excite an auto-oscillation of the magnetization in magnetic tunnel junctions (MTJs)[1,20–22,24–26,28–30], where the spin angular momentum from conducting electrons carrying electric current is transferred to ferromagnet and excites magnetization dynamics. It is, however, preferable to excite magnetization dynamics without driving current and magnetic field from the viewpoints of low-power consumption and simple implementation.

In this work, we propose that physical reservoir computing can be performed by magnetization dynamics induced by voltage control of magnetic anisotropy in solid devices[33–50]. The voltage control of magnetic anisotropy is a fascinating technology as the low-power information writing scheme in magnetoresistive random access memory, instead of using spin-transfer torque effect. An application of electric voltage to a metallic ferromagnet/insulator interface modifies electron states near the interface[34,36,37] and/or induces magnetic moment[46], and changes magnetic anisotropy. The magnetization in the ferromagnetic metal changes its direction to minimize the magnetic anisotropy energy. Therefore, the voltage application can cause the relaxation dynamics of the magnetization in the ferromagnet. In the practical application of nonvolatile random access memory, an external magnetic field is necessary to achieve a deterministic magnetization switching guaranteeing reliable writing[40–43]. On the other hand, we notice that the magnetization switching, as well as magnetic field, is not a necessary condition in physical reservoir computing. Accordingly, the voltage control of magnetic anisotropy can be used to realize physical reservoir computing by spintronics devices without driving current or magnetic field. Here, we perform numerical simulation of the Landau-Lifshitz-Gilbert (LLG) equation and find that the computational

[1]Research Center for Emerging Computing Technologies, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Ibaraki 305-8568, Japan. [2]Department of Physics Engineering, Faculty of Engineering, Mie University, Tsu, Mie 514-8507, Japan. [3]PRESTO, Japan Science and Technology Agency (JST), Saitama, Japan. ✉email: tomohiro-taniguchi@aist.go.jp

Figure 1. (**a**) Schematic illustration of an MTJ. The unit vector pointing to the magnetization direction in the ferromagnetic free layer is **m**. The z axis is perpendicular to the film plane. (**b**) An example of the time evolutions of m_x (red), m_y (blue), and m_z (black). (**c**) Trajectory of the relaxation dynamics on a sphere. In (**b**,**c**), the first order magnetic anisotropy field H_{K1} is changed from $-0.1H_{K2}$ to $-0.9H_{K2}$ by the voltage application. The red circle and blue triangle in (**c**) represent the initial and final states of the dynamics.

capabilities of benchmark tasks in single spintronics device are comparable to those of echo-state networks with more than 10 nodes.

Model

LLG equation. The system under investigation is a cylinder-shaped MTJ schematically shown in Fig. 1a, where the z axis is perpendicular to the film plane. The MTJ consists of ferromagnetic free layer, MgO insulator, and ferromagnetic reference layer. The ferromagnetic free layer has the perpendicular magnetic anisotropy, where the magnetic energy density is given by

$$\varepsilon = \sum_{i=x,y,z} 2\pi M^2 N_i m_i^2 + K_1 \left(1 - m_z^2\right) + K_2 \left(1 - m_z^2\right)^2. \tag{1}$$

The first term on the right-hand side in Eq. (1) represents the shape magnetic anisotropy energy density with the saturation magnetization M and the demagnetization coefficients N_i. Since we assume the cylinder shape, $N_x = N_y$. The unit vector pointing in the magnetization direction of the free layer is denoted as $\mathbf{m} = (m_x, m_y, m_z)$. The second and third terms are the first and second order magnetic anisotropy energy densities with the coefficients K_1 and K_2. Note that the energy density relates to the magnetic field inside the free layer as

$$\mathbf{H} = \left[H_{K1} + H_{K2}\left(1 - m_z^2\right)\right] m_z \mathbf{e}_z, \tag{2}$$

where $H_{K1} = (2K_1/M) - 4\pi M(N_z - N_x)$ and $H_{K2} = 4K_2/M$; see also "Methods". The magnetization in the reference layer points to the z direction, and therefore, m_z is experimentally measured through tunnel magnetoresistance effect.

The first order magnetic anisotropy energy coefficient K_1 consists of the bulk and interfacial contributions, K_v and K_i, and the voltage-controlled magnetic anisotropy effect described as $K_1 d = K_v d + K_i - \eta \mathscr{E}$. The thickness of the ferromagnetic free layer is d, whereas $\mathscr{E} = V/d_I$ is the electric field with the voltage V and the thickness of the insulator d_I. In typical MTJs consisting of CoFeB free layer and MgO insulator, K_i dominates in K_1, where K_i increases with the increase of the composition of Fe[51–53]. It can reach on the order of 1.0 mJ/m^2 at maximum, which in terms of magnetic field, $2K_i/(Md)$, is typically on the order of 1 T. Note that the magnitude of the shape magnetic anisotropy field $-4\pi M(N_z - N_x)$ is also on the order of 1 T, where a typical value of the saturation magnetization in CoFeB, i.e., M of about 1000 emu/c.c., is assumed. As a result of the competition between them, the ferromagnetic free layer in the absence of the voltage application can be either in-plane or perpendicular-to-plane magnetized[51–53]. The voltage control of magnetic anisotropy also modifies the magnetic anisotropy field H_{K1} through the modification of the electron occupation states near the ferromagnetic interface[34,36,37] and/or the generation of the magnetic dipole moment[46]. The coefficient of the voltage-controlled magnetic anisotropy effect, η, is recently achieved in the experiment to be about 300 fJ/(Vm)[45,50], whereas the thickness of the insulator is about 2.5 nm. A typical values of the applied voltage is about 0.5 V at maximum[48]. Thus, the tunable range of the magnetic anisotropy by the voltage application in terms of the magnetic field, $(2|\eta|V)/(Mdd_I)$, is about 1.0 kOe, where we assume that $M = 1000$ emu/c.c., $d = 1$ nm, $d_I = 2.5$ nm, and $|\eta| = 250$ fJ/(V m). Note that the sign of the voltage-controlled magnetic anisotropy effect depends on that of the voltage. Summarizing these contributions, H_{K1} in the presence of the voltage can also be either positive or negative, depending on the materials and their compositions, as well as the magnitude and sign of the applied voltage. For example, Ref.[38] uses an in-plane magnetized ferromagnet, i.e., $H_{K1} < 0$ for $V = 0$. The voltage control of magnetic anisotropy in Ref.[38] enhances the perpendicular anisotropy K_1 and makes H_{K1} positive at nonzero V. On the other hand, perpendicularly magnetized free layers where $H_{K1} > 0$ for $V = 0$ have been used in Ref.[43]. Contrary to H_{K1}, the dependence of $H_{K2} \propto K_2$ on the applied voltage is still unclear, where Ref.[47] reports that H_{K2} is approximately independent of

the voltage while Ref.[48] observes the voltage dependence of H_{K2}. Throughout this paper, for simplicity, we assume that only H_{K1} depends on the voltage. As mentioned in the following, we performed numerical simulation by changing the value of H_{K1}. It means that we do not specify the size (the thickness and cross-section area) of MTJ explicitly because H_{K1} includes the information of the shape of MTJ through the demagnetization coefficients N_i. It is, however, useful to mention that macrospin model has been proven to work well to describe the magnetization dynamics for MTJ whose typical size is 1-2 nm in thickness and the diameter less than 200 nm[40,42,49].

In typical experiments on voltage control of magnetic anisotropy, a relatively thick (typically 1.5-2.5 nm) MgO barrier is used as an insulator[42,43,49], compared with MTJ manipulated by spin-transfer torque, where the thickness of the barrier is about 1.0 nm[54]. As a result, the resistance of MTJ used for experiments of voltage control of magnetic anisotropy, on the order of 10-100 kΩ, is two or three orders of magnitude larger than that used for spin-transfer torque experiments. On the other hand, the maximum voltage used in both experiments is almost identical. Accordingly, current flowing in MTJ used for experiments of voltage control of magnetic anisotropy is two or three orders of magnitude smaller than that used for spin-transfer torque experiments (see also "Methods"). In this sense, we mention that the driving force of magnetization dynamics is voltage control of magnetic anisotropy effect, although current cannot be completely zero in experiments. As mentioned in "Methods", typical value of current I flowing in MTJ is on the order of 1 μA, while the current used in physical reservoir computing utilizing spin-transfer torque is on the order of 1 mA[29]. On the other hand, the magnitude of the voltage V applied to MTJ is nearly the same for both experiments on voltage control of magnetic anisotropy and spin-transfer effects. Accordingly, using the voltage control of magnetic anisotropy effect could reduce energy consumption by three orders.

The magnetization in equilibrium points to the direction at which the energy density is minimized. For example, when $H_{K1} > (<)0$ and $H_{K2} = 0$, the energy is minimized when the magnetization is parallel (perpendicular) to the z axis. Another example is studied in Ref.[55], where, if $H_{K1} < 0$ and $|H_{K1}| < H_{K2}$, the energy density ε is minimized when $m_z = \pm\sqrt{1 - (|H_{K1}|/H_{K2})}$. When the voltage is applied to the MTJ and the minimum energy state is changed as a result, the magnetization relaxes to the state. The relaxation dynamics is described by the LLG equation,

$$\frac{d\mathbf{m}}{dt} = -\gamma\mathbf{m} \times \mathbf{H} + \alpha\mathbf{m} \times \frac{d\mathbf{m}}{dt}, \tag{3}$$

where γ and α are the gyromagnetic ratio and the Gilbert damping constant, respectively. Note that the macrospin model works well to describe the magnetization dynamics driven by the voltage application[40,42,44]. The values of the parameters used in the following are derived from typical experiments[35,38–44,47,48]. The gyromagnetic ratio and the Gilbert damping constant are $\gamma = 1.764 \times 10^7$ rad/(Oe s) and $\alpha = 0.01$. The second order magnetic anisotropy field H_{K2} is 500 Oe.
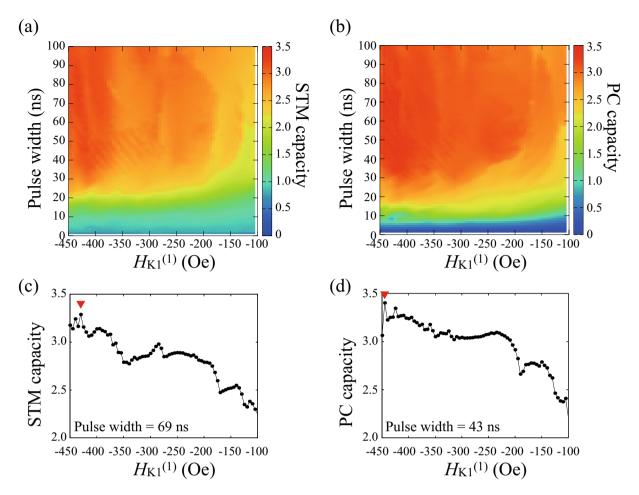
Let us show an example of the magnetization dynamics driven by the voltage control of magnetic anisotropy. We firstly set H_{K1} to be $H_{K1}^{(0)} = -0.1H_{K2} = -50$ Oe and solve the LLG equation with an arbitrary initial condition. The magnetization saturates to a certain point where m_z saturates to $m_z \to m_z^{(0)} \simeq 0.95$. We use this state as a new initial state and solve the LLG equation by changing H_{K1} to $H_{K1}^{(1)} = -0.9H_{K2} = -450$ Oe. Then, the magnetization starts to move to a new stable state where m_z saturates to $m_z \to m_z^{(1)} \simeq 0.32$. Figure 1b,c show time evolution of $\mathbf{m}$ and its spatial orbit from the initial state of $m_z^{(0)}$ to the final state $m_z^{(1)}$. We confirm that the initial and final states are those expected from the minimum energy state mentioned above, i.e., $m_z^{(0)} = \sqrt{1 - |H_{K1}^{(0)}|/H_{K2}} = \sqrt{1 - 0.1} \simeq 0.95$ and $m_z^{(1)} = \sqrt{1 - |H_{K1}^{(1)}|/H_{K2}} = \sqrt{1 - 0.9} \simeq 0.32$. We emphasize that m_z monotonically changes with respect to the change of H_{K1}. Since the value of H_{K1} can be manipulated by the voltage application, the time evolution of m_z can be used to identify the value of the applied voltage. The estimation of the input data, which is the sequence of the applied voltage in the present case, from the dynamical response of physical system is the aim of physical reservoir computing. Therefore, the magnetization dynamics driven by the voltage control of magnetic anisotropy is applicable to physical reservoir computing. In the following, we evaluate its computational ability.
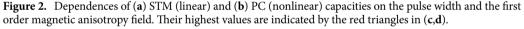
Results

Memory capacity. The ability in physical system for reservoir computing has been quantified by memory capacity[15,18,20,21,25,28,30]. The memory capacity corresponds to the number of past data physical reservoir can store. For example, let us imagine injecting random binary input $b = 0$ or 1 to reservoir, as done in experiments[21,25]. The input data are often injected as pulses with the pulse width of t_p, i.e., the value of b is constant during time t_p. Therefore, it is convenient to add a suffix $k = 1, 2, \ldots$ to b as b_k to distinguish the order of the input data. We also introduce an integer $D = 0, 1, 2, \ldots$, called delay, characterizing the order of the past input data. In this case,

$$y_{k,D}^{\text{STM}} = b_{k-D}, \tag{4}$$

are called target data of short-term memory (STM) task. We predict the value of the target data from the output of the reservoir and evaluate the reproducibility. The predicted data are called system output. The reproducibility is quantified by the correlation coefficient between the target data and system output. Roughly speaking, if the reservoir can reproduce the past data up to D, the STM capacity is defined as D. There is another kind of memory capacity, called parity-check (PC) capacity, where the target data are defined as

(a)

(b)

(c)

(d)

Figure 2. Dependences of (**a**) STM (linear) and (**b**) PC (nonlinear) capacities on the pulse width and the first order magnetic anisotropy field. Their highest values are indicated by the red triangles in (**c,d**).

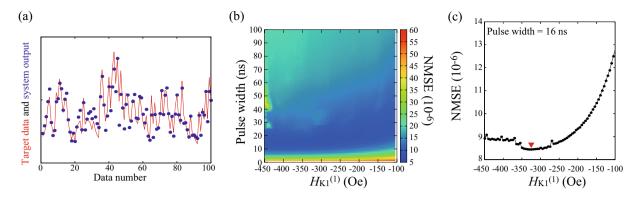$$y_{k,D}^{\text{PC}} = \sum_{\ell=0}^{D} b_{k-D+\ell} \;(\text{mod}2). \tag{5}$$

According to their definitions, the STM and PC capacities quantify the number of the target data the reservoir can store, where the target data are defined as linear and nonlinear transformations of the input data, respectively. A large memory capacity means that reservoir can store, recognize, and/or predict large data. See also "Methods" for the detail of the evaluation method of these capacities.
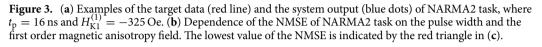
In the present system, the random binary inputs are injected as voltage pulses, which change the first order magnetic anisotropy field H_{K1} as

$$H_{\text{K1}} = H_{\text{K1}}^{(0)}(1 - b_k) + H_{\text{K1}}^{(1)}b_k. \tag{6}$$

Accordingly, when the input is $b_k = 0\ (1)$, the value of H_{K1} is $H_{\text{K1}}^{(0)}\ [H_{\text{K1}}^{(1)}]$. In the following, we fix $H_{\text{K1}}^{(0)} = -50$ Oe, i.e., $H_{\text{K1}}^{(0)}/H_{\text{K2}} = -0.1$, whereas $H_{\text{K1}}^{(1)}$ varies in the range of $-450 \leq H_{\text{K1}}^{(1)} \leq -100$ Oe, i.e., $-0.9 \leq H_{\text{K1}}^{(1)}/H_{\text{K2}} \leq -0.2$. Figure 2a,b show the STM and PC capacities as a function of $H_{\text{K1}}^{(1)}$ and the pulse width of the input data. The highest value of the STM capacity, 3.29, is found at the conditions of $H_{\text{K1}}^{(1)} = -430$ Oe and $t_{\text{p}} = 69$ ns, as shown in Fig. 2c. On the other hand, the highest value of the PC capacity, 3.40, is found at the conditions of $H_{\text{K1}}^{(1)} = -445$ Oe and $t_{\text{p}} = 43$ ns, as shown in Fig. 2d.

NARMA task. Another benchmark task to quantify the computational ability of physical system to reservoir computing is nonlinear autoregressive moving average (NARMA) task[12,15,18,30,56]. NARMA task is a function-approximation task to reproduce a nonlinear function defined from input data by using output data in recurrent neural networks. The task is classified as NARMAD with $D = 2, 5, 10$ and so on, where D represents the delay included in the nonlinear function. In other words, the target data of NARMAD task consist of data defined until D times before from the present data. For example, in NARMA2 task, the system is aimed to reproduce the target data,

Figure 3. (**a**) Examples of the target data (red line) and the system output (blue dots) of NARMA2 task, where $t_p = 16$ ns and $H_{K1}^{(1)} = -325$ Oe. (**b**) Dependence of the NMSE of NARMA2 task on the pulse width and the first order magnetic anisotropy field. The lowest value of the NMSE is indicated by the red triangle in (**c**).

$$y_{k+1}^{\text{NARMA2}} = 0.4 y_k^{\text{NARMA2}} + 0.4 y_k^{\text{NARMA2}} y_{k-1}^{\text{NARMA2}} + 0.1 z_k^3 + 0.1, \qquad (7)$$

from output data, where $z_k = 0.2 r_k$ is defined from uniform random input data r_k at a discrete time k; see "Methods". The computational ability of NARMA task is evaluated from normalized mean-square error (NMSE) defined as

$$\text{NMSE} = \frac{\sum_k \left(y_k^{\text{NARMA2}} - v_k^{\text{NARMA2}} \right)^2}{\sum_k \left(y_k^{\text{NARMA2}} \right)^2}, \qquad (8)$$

where v_k^{NARMA2} is the data reproduced from the output data (see also "Methods"). A low NMSE corresponds to high reproducibility of the target data. Figure 3a shows an example of the target data (red line), y_k^{NARMA2}, and the system output (blue dots). By evaluating the difference between the target data and the system output as such, the NMSE is obtained as shown in Fig. 3b. The NMSE is on the order of $10^{-6} - 10^{-5}$ and is minimized to be 8.43×10^{-6} at $t_p = 16$ ns and $H_{K1}^{(1)} = -325$ Oe; see Fig. 3c.

Discussion

We have developed theoretical analysis of the magnetization dynamics in nanostructured ferromagnetic multilayers driven by the voltage control of magnetic anisotropy, and showed that the dynamics is applicable to physical reservoir computing through the evaluations of the memory capacity and the NMSE of NARMA task. Neither electric current nor external magnetic field is introduced in the computation, contrary to the previous works focusing on the application to nonvolatile memory, because magnetization switching is unnecessary. This fact will be preferable for reducing power consumption in reservoir computing.

Figure 2a,b show that the memory capacity increases with the difference between $H_{K1}^{(0)}$ and $H_{K1}^{(1)}$ increasing. This is because when the difference between $H_{K1}^{(0)}$ and $H_{K1}^{(1)}$ is large, the range of the dynamical response of m_z also becomes large, which makes it easy to identify the input data from the change of m_z. Due to a similar reason, the memory capacity increases with the increase of the pulse width. When the pulse width is relatively long, the change of m_z during a pulse injection becomes large, which again makes it easy to identify the input data. However, when the pulse width is sufficiently long, m_z finally saturates to a stable state, and becomes approximately constant, as implied from Fig. 1b. When m_z becomes constant, it becomes impossible to estimate the past input from the present output. Therefore, the memory capacity does not increase monotonically with the pulse width increasing. As written above, the STM and PC capacities are maximized at the pulse width of 69 and 43 ns, respectively. A similar trend is found in NARMA2 task, where low NMSEs are achieved in a relatively large $H_{K1}^{(1)}$ region. Note that the memory capacity at the maximum was found to be about 3, which is comparable to the computational ability of echo-state network with approximately 10 nodes[20,28]. The value is also comparable or larger than that obtained by the other single spintronics reservoirs without additional circuits[20,21,29], driven by electric current and/or magnetic field. This might be due to a matching between the relaxation time of the output signal and the pulse width. Another possible reason is a large change in the dynamical amplitude, compared with an oscillator system[29]. The NMSE of NARMA2 task, minimized to be on the order of 10^{-6}, is also comparable or lower than that found in soft robot[12] and echo-state network with nodes more than 10^{18}. These results indicate the potential applicability of an MTJ driven by the voltage-controlled magnetic anisotropy effect to physical reservoir computing.

An empirical rule shared among the research community is that the computational ability of physical reservoir computing is maximized at the edge of chaos[13,30,57]. Simultaneously, an existence of chaos might lose the reproducibility of the computation due to the sensitivity to initial states. Note that chaos is prohibited in the present system when random inputs are absent. This is because the magnetization dynamics are described by two variables, $\theta = \cos^{-1} m_z$ and $\varphi = \tan^{-1}(m_y/m_x)$, whereas the Poincaré-Bendixson theorem argues that chaos does not appear in a two dimensional system. When the random input are injected to the MTJ, the system becomes nonautonomous due to the presence of time-dependent torque. In this case, the number of the dimension in

the phase space becomes three, and the possibility to induce chaos becomes finite. For example, Ref.[30] reported the appearance of chaos in a spin-torque oscillator due to the injection of random input current. However, we should notice that the presence of time-dependent input does not necessarily guarantee the presence of chaos. The identification of chaos is done by, for example, evaluating the Lyapunov exponent. The Lyapunov exponent quantifies the time evolution of an infinitesimal difference given at the initial state. The positive Lyapunov exponent implies the presence of chaos. On the other hand, when the Lyapunov exponent is negative, the dynamics saturate to fixed points. When the Lyapunov exponent is zero, the dynamics is periodic. The dynamics with negative or zero Lyapunov exponent are classified as ordered dynamics. Since the LLG equation describes the relaxation dynamics to stable states, one might consider that the largest Lyapunov exponent of an MTJ in the absence of random inputs is negative. However, notice that the axial symmetry of the present system enables us to move the magnetization rotating around the z axis without energy injection. In fact, the energy density, as well as the equation of motion for m_z depends on m_z only, as explained in "Methods"; in other words, the equation of motion for θ is independent of φ. As a result, an infinitesimal difference given to the phase φ is not shortened by the LLG equation. Therefore, the largest Lyapunov exponent in the absence of the random input is zero. The fact that the equation of motion for θ depends on θ only also implies the absence of homoclinic bifurcations, as well as chaos, even when the pulse data, independent of θ and φ, are injected; in fact, the numerically evaluated Lyapunov exponent was zero, as explained in "Methods". The absence of chaos indicates the reproducibility of the computation in the present reservoir.

In summary, we perform numerical experiments of the magnetization dynamics in an MTJ driven by the voltage control of magnetic anisotropy. Injecting the voltage pulse to the MTJ, the magnetization changes its direction to minimize the magnetic anisotropy energy. The time evolution of the relaxation dynamics reflects the value of the input voltage, and therefore, can be used to reproduce the time sequence of the input data. We evaluate the computing abilities, such as the memory capacity and the error in the reproducibility, of common benchmark task, and show that even a single MTJ can show high computing performance comparable to echo-state network consisting of multiple nodes more than 10. Since neither electric current nor external magnetic field is necessary, the proposal here will be of interest for energy-saving computing technologies.

Methods

Definition of magnetic field and relaxation time.   The magnetic field **H** relates to the energy density ε as $\mathbf{H} = -\partial \varepsilon / \partial (M\mathbf{m})$, and therefore, is obtained from Eq. (1) as

$$\mathbf{H} = \begin{pmatrix} -4\pi M N_x m_x \\ -4\pi M N_y m_y \\ [(2K_1/M) - 4\pi M N_z]m_z + (4K_2/M)\left(1 - m_z^2\right)m_z \end{pmatrix}. \tag{9}$$

We should note that the magnetization dynamics described by the LLG equation is unchanged by adding a term proportional to **m** to **H** because the LLG equation conserves the magnitude of **m**. Adding a term as such corresponds to shifting the origin of the energy density ε by a constant. In the present case, we added a term $4\pi M N_x \mathbf{m}$ to **H** and obtained Eq. (2), where we should remind that $N_x = N_y$ because we assume a cylinder shaped MTJ. The added term to **H** shifts the origin of the energy density ε by the constant $-2\pi M^2 N_x \mathbf{m}^2 = -2\pi M^2 N_x$ and makes it depend on m_z only.

The LLG equation in the present system can be integrated as

$$t = \frac{1 + \alpha^2}{\alpha \gamma (H_{K1} + H_{K2})} \left[\log \left(\frac{\cos \theta_f}{\cos \theta_i} \right) - \frac{H_{K1} + H_{K2}}{H_{K1}} \log \left(\frac{\sin \theta_f}{\sin \theta_i} \right) + \frac{H_{K2}}{2H_{K1}} \log \left(\frac{H_{K1} + H_{K2} \sin^2 \theta_f}{H_{K1} + H_{K2} \sin^2 \theta_i} \right) \right], \tag{10}$$

where θ_i and θ_f are the initial and final values of $\theta = \cos^{-1} m_z$. Equation (10) provides the relaxation time from $\theta = \theta_i$ to $\theta = \theta_f$. Note that the relaxation time is scaled by $\alpha \gamma H_{K1}/(1 + \alpha^2)$ and H_{K2}/H_{K1}, which can be manipulated by the voltage control of magnetic anisotropy. We also note that Eq. (10) has logarithmic divergence due to asymptotic behavior in the relaxation dynamics.

Role of spin-transfer torque.   We neglected spin-transfer torque in the main text because the current magnitude in typical MTJ used for voltage control of magnetic anisotropy effect is usually small. For example, when using typical values[47,49] for the voltage (0.4 V), resistance (60 kΩ), and cross-section being $\pi \times 60^2$ nm^2, the value of the current density is about 0.06 MA/cm^2 (6.7 μA in terms of current). Such a value is sufficiently small compared with that used in spin-transfer torque switching experiments[54]. To verify the argument, we perform numerical simulations, where spin-transfer torque, $-H_s \mathbf{m} \times (\mathbf{p} \times \mathbf{m})$, is added to the right-hand side of Eq. (3). We fix the values of $H_{K2} = 500$ Oe and $H_{K1} = -0.1 H_{K2} = -50$ Oe. The unit vector **p** along the direction of the magnetization in the reference layer points to the positive z direction. Spin polarization P in the spin-transfer torque strength, $H_s = \hbar P j/(2eMd)$, is assumed to be 0.5. Figure 4a shows time evolution of **m** for the current density j of 0.06 MA/cm^2. Although the magnetization slightly moves from the initial (stable) state due to spin-transfer torque, the change of the magnetization direction is small compared with that shown in Fig. 1b. Therefore, we do not consider that spin-transfer torque plays a major role in physical reservoir computing, although current cannot be completely zero in experiments.

For comprehensiveness, however, we also show the magnetization dynamics when the current density j is increased by one order Figure 4b shows the dynamics for $j = 0.6$ MA/cm^2, where the magnetization switching by spin-transfer torque is observed. We note that the current density is sufficiently small compared with that used in typical MTJs in nonvolatile memory[54]. Nevertheless, the switching is observed because of a small value of the
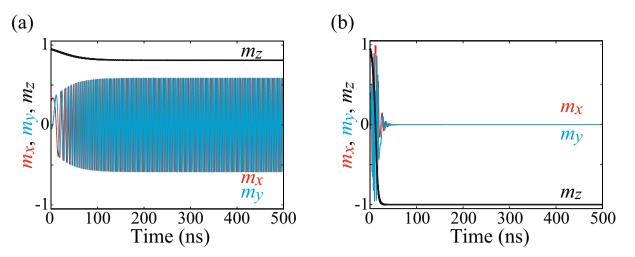
Figure 4. Examples of the time evolutions of m_x (red), m_y (blue), and m_z (black) in the presence of spin-transfer torque, where the current density is (**a**) 0.06 MA/cm^2 and (**b**) 0.6 MA/cm^2.

magnetic anisotropy field in the present system. We assume that H_{K2} is finite and $|H_{K1}| < H_{K2}$ to make a tilted state of the magnetization $[m_z = \pm\sqrt{1 - (|H_{K1}|/H_{K2})}]$ stable due to the following reason. Remind that there are other stable states, such as $m_z = \pm1$ for $H_{K1} > 0$ and $m_z = 0$ for $H_{K1} < 0$, when $H_{K2} = 0$. Note that these states ($m_z = \pm1$ or $m_z = 0$) are always local extrema of energy landscape. Accordingly, once the magnetization saturates to these states, it cannot change the direction even if another input is injected. This conclusion can be understood in a different way, where the relaxation time given by Eq. (10) shows a divergence when $\theta_i = 0$ ($m_z = +1$), π ($m_z = -1$), or $\pi/2$ ($m_z = 0$) is substituted. On the other hand, for a finite H_{K2}, the magnetization can move from the state $m_z = \pm\sqrt{1 - (|H_{K1}|/H_{K2})}$ when an input signal changes the value of H_{K1} and makes the state no longer an extremum. We note that the assumption $|H_{K1}| < H_{K2}$ restricts the magnitude of the magnetic field. In fact, the magnitude of **H** is small due to a small value of $H_{K2} = 500$ Oe found in experiments[47,48] and the restriction of $|H_{K1}| < H_{K2}$. Since a critical current density destabilizing the magnetization by spin-transfer effect is proportional to the magnitude of the magnetic field, a small **H** implies that a small current mentioned above might induce a large-amplitude magnetization dynamics.

In summary, the magnitude of the current density is sufficiently small, and the magnetization dynamics are mainly driven by voltage control of magnetic anisotropy effect. The condition to stabilize a tilted state, however, might make the magnitude of the magnetic field, as well as the critical current density of spin-transfer torque switching, small. Thus, even a small current may cause nonnegligible dynamics. Simultaneously, however, it is practically difficult to increase the current magnitude by one order, and therefore, in the present study, we still consider that voltage control of magnetic anisotropy effect is the main driving force of the magnetization dynamics.

Evaluation method of memory capacity. The memory capacity corresponds to the number of data which can be reproduced from the output data, as mentioned in the main text. The evaluation of the memory capacity consists of two processes. During the first process called training (or learning), weights are determined to reproduce the target data from the output data. In the second process, the reproducibility of the target data defined from other input data is evaluated.

Let us first describe the training process. We inject the random binary input $b_k = 0$ or 1 into MTJ as voltage pulse. The number of the random input is N. The input b_k is converted to the first order magnetic anisotropy field through the voltage control of magnetic anisotropy, which is described by Eq. (6). We choose m_z as output data, which can be measured experimentally through magnetoresistance effect. Figure 5a shows an example of the time evolution of m_z in the presence of several random binary inputs, where the values of the parameters are those at the maximum STM capacity conditions, i.e., the pulse width and the first order magnetic anisotropy field are $t_p = 69$ ns and $H_{K1}^{(1)} = -430$ Oe. As can be seen, the injection of the random input drives the dynamics of m_z.

The dynamical response $m_z(t)$, during the presence of the kth input b_k, is divided into nodes, where the number of nodes is N_{node}. We denote the $i(=1, 2, \ldots, N_{node})$th output with respect to the kth input as $u_{k,i} = m_z(t_0 + (k-1)t_p + i(t_p/N_{node}))$, where t_0 is time for washout. The output $u_{k,i}$ is regarded as the status of the ith neuron at a discrete time k. Figure 5b shows an example of the time evolution of m_z with respect to an input pulse, whereas the dots in the inset of the figure are the nodes $u_{k,i}$ defined from m_z. The method to define such virtual neurons is called time-multiplexing method[15,20,21]. We also introduce bias term $u_{k,N_{node}+1} = 1$. In the training process, we introduce weight $w_{D,i}$ and evaluate its value to minimize the error,

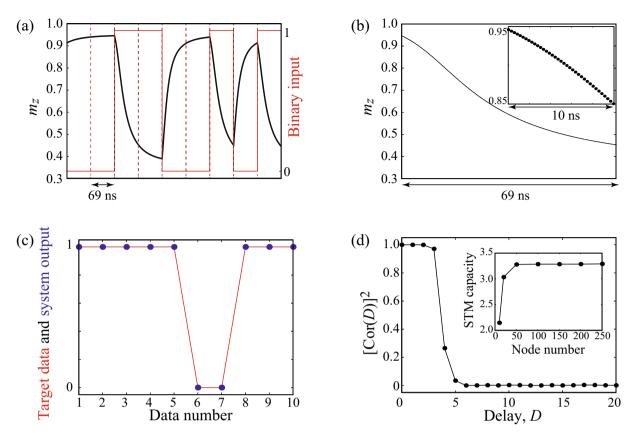$$\sum_{k=1}^{N} \left(\sum_{i=1}^{N_{node}+1} w_{D,i} u_{k,i} - y_{k,D} \right)^2, \tag{11}$$

Figure 5. (**a**) An example of the time evolution of m_z (black) in the presence of several binary pulses (red). The dotted lines distinguish the input pulse. The pulse width and the first order magnetic anisotropy field are 69 ns and − 430 Oe, respectively, where the STM capacity is maximized. (**b**) An example of m_z in the presence of a random input. The dots in the inset shows the definition of the nodes $u_{k,i}$ from m_z during a part of an input pulse. The node number is $N_{\text{node}} = 250$. (**c**) Examples of the target data $y'_{n,D}$ (red line) and the system output $v'_{n,D}$ (blue dots) of STM task with $D = 1$. (**d**) Dependence of $[\text{Cor}(D)]^2$ on the delay D for STM task. The node number is $N_{\text{node}} = 250$. The inset shows the dependence of the STM capacity on the node number.

where, $y_{k,D}$ are the target data defined by Eqs. (4) and (5). For simplicity, we omit the superscripts such as "STM" and "PC" in the target data because the difference in the evaluation method of the STM and PC capacities is merely due to the definition of the target data. In the following, we add superscripts or subscripts, such as "STM" and "PC", when distinguishing quantities related to their capacities are necessary. The weight should be introduced for each target data. According to the above statement, we denote the weight to evaluate the STM (PC) capacity as $w_{D,i}^{\text{STM(PC)}}$, when necessary. Also, we note that the weights are different for each delay D.

Once the weights are determined, we inject other random binary inputs b'_k to the reservoir, where the number of the input data is N'. Note that N' is not necessarily the same with N. Here, we use the prime symbol to distinguish the input data from those used in training. Similarly, we denote the output and target data with respect to b'_k as $u'_{n,i}$ and $y'_{n,D}$, respectively, where $n = 1, 2, \ldots, N'$. From the output data $u'_{n,i}$ and the weight $w_{D,i}$, we define the system output $v'_{n,D}$ as

$$v'_{n,D} = \sum_{i=1}^{N_{\text{node}}+1} w_{D,i} u'_{n,i}. \tag{12}$$

Figure 5c shows an example of the comparison between the target data $y'_{n,D}$ (red line) and the system output $v'_{n,D}$ (blue dots) of STM task with $D = 1$. It is shown that the system output well reproduces the target data. The reproducibility of the target data is quantified from the correlation coefficient $\text{Cor}(D)$ between $y'_{n,D}$ and $v'_{n,D}$ defined as

$$\text{Cor}(D) \equiv \frac{\sum_{n=1}^{N'} \left(y'_{n,D} - \langle y'_{n,D} \rangle \right) \left(v'_{n,D} - \langle v'_{n,D} \rangle \right)}{\sqrt{\sum_{n=1}^{N'} \left(y'_{n,D} - \langle y'_{n,D} \rangle \right)^2 \sum_{n=1}^{N'} \left(v'_{n,D} - \langle v'_{n,D} \rangle \right)^2}}, \tag{13}$$

where $\langle \cdots \rangle$ represents the averaged value. Note that the correlation coefficients are defined for each delay D. We also note that the correlation coefficients are defined for each kind of capacity, as in the case of the weights and target data. In general, $[\text{Cor}(D)]^2 \leq 1$, where $[\text{Cor}(D)]^2 = 1$ holds only when the system output completely reproduces the target data. Figure 5d shows an example of the dependence of $[\text{Cor}(D)]^2$ for STM task on the delay

D. The results implies that the reservoir well reproduces the target data until $D = 3$, whereas the reproducibility drastically decreases with the delay D increasing. The STM and PC capacities, C_{STM} and C_{PC}, are defined as

$$C = \sum_{D=1}^{D_{\text{max}}} [\text{Cor}(D)]^2. \tag{14}$$

Note that the definition of the memory capacity obeys, for example, Refs.[18,20,21,25], where the memory capacity in Eq. (14) is defined by the correlation coefficients starting from $D = 1$. In some papers such as Refs.[15,30], however, the square of the correlation coefficient at $D = 0$ is added to the right-hand side of Eq. (14).

In the present study, we introduce $N_{\text{node}} = 250$ nodes and use $N = 1000$ and $N' = 1000$ random binary pulses for training of the weight and evaluation of the memory capacity, respectively. The number of nodes is chosen so that the value of the capacity saturates with the number of nodes increasing; see the inset of Fig. 5d. We also use 300 random binary pulses before the training and between training and evaluation for washout. The maximum delay D_{max} is 20. Note that the value of each node should be sampled within a few hundred picosecond: specifically, in the case of an example shown in Fig. 2c, it is necessary to sample data within $t_{\text{p}}/N_{\text{node}} = 69\text{ns}/250 \simeq 276$ ps. We emphasize that it is experimentally possible to sample data within such a short time. For example, in Ref.[21], $t_{\text{p}} = 20$ ns and $N_{\text{node}} = 200$ were used, where the sampling step is 100 ps.

NARMA task. The evaluation procedure of the NMSE in NARMA task is similar to that of the memory capacity. The binary input data, $b_k = 0$ or 1, in the evaluation of the memory capacity are replaced by uniform random number r_k in (0, 1). The variable z_k in Eq. (7) is generally defined as $z_k = \mu + \sigma r_k$[30], where the parameters μ and σ are determined to make z_k be in (0, 0.2)[15]. As in the case of the evaluation of the memory capacity, the evaluation of the NMSE consists of two procedures. The first procedure is the training, where the weight is determined to reproduce the target data from the output data $u_{k,i}$. Secondly, we evaluate the reproducibility of another set of the target data from the system output v_n^{NARMA2} defined from the weight and the output data. Then, the NMSE can be evaluated. Note that some papers[13,27,30] define the NMSE in a slightly different way, where $\sum_{n=1}^{N'} \left(y_n^{\text{NARMA2}}\right)^2$ in the denominator of Eq. (8) is replaced by $\sum_{n=1}^{N'} \left(y_n^{\text{NARMA2}} - \overline{y}^{\text{NARMA2}}\right)^2$, where $\overline{y}^{\text{NARMA2}}$ is the average of the target data y_n^{NARMA2}. In this work, we use the definition given by Eq. (8), which is used, for example, in Refs.[12,15,18].

Evaluation of Lyapunov exponent. We evaluated the conditional Lyapunov exponent as follows[58]. The LLG equation was solved by the fourth-order Runge-Kutta method with time increment of $\Delta t = 1$ ps. We added perturbations $\delta\theta$ and $\delta\phi$ with $\varepsilon = \sqrt{\delta\theta^2 + \delta\varphi^2} = 10^{-5}$ to $\theta(t)$ and $\varphi(t)$ at time t. Let us denote the perturbed $\theta(t)$ and $\varphi(t)$ as $\theta'(t)$ and $\varphi'(t)$, respectively. Solving the LLG equation from time t to $t + \Delta t$, the time evolution of the perturbation is obtained as $\varepsilon'(t) = \sqrt{[\theta'(t + \Delta t) - \theta(t + \Delta t)]^2 + [\varphi'(t + \Delta t) - \varphi(t + \Delta t)]^2}$. A temporal Lyapunov exponent is obtained as $\lambda(t) = (1/\Delta t) \log[\varepsilon'(t)/\varepsilon]$. Repeating the procedure, the temporal Lyapunov exponent is averaged as $\lambda(\mathcal{N}) = (1/\mathcal{N}) \sum_{i=1}^{\mathcal{N}} \lambda(t_i) = [1/(\mathcal{N}\Delta t)] \sum_{i=1}^{\mathcal{N}} \log\{\varepsilon'[t_0 + (i-1)\Delta t]/\varepsilon\}$, where t_0 is time at which the first random input is injected, whereas $\mathcal{N}$ is the number of averaging. The Lyapunov exponent is given by $\lambda = \lim_{\mathcal{N} \to \infty} \lambda(\mathcal{N})$. In the present study, we used the time range same as that used in the evaluations of the memory capacity and the NMSE and added uniform random input. Hence, notice that $\mathcal{N} = \mathcal{M} t_{\text{p}}/\Delta t$ depends on the pulse width, where $\mathcal{M}$ is the total number of the random inputs including washout, training, and evaluation. We confirmed that $\lambda(\mathcal{N})$ monotonically saturates to zero; at least, $|\lambda(\mathcal{N})|$ is one or two orders of magnitudes smaller than $1/t_{\text{p}}$. Thus, the expansion rate of the perturbation, $1/\lambda(\mathcal{N})$, is much slower than the injection rate of the input signal. Considering these facts, we concluded that the largest Lyapunov exponent can be regarded as zero, and therefore, chaos is absent. Note that the absence of chaos in the present system relates to the facts that the free layer is axially symmetric and the applied voltage modifies the perpendicular anisotropy only. When there are factors breaking the symmetry, such as spin-transfer torque with an in-plane spin polarization, chaos will appear[30].

Data availability
The datasets used and/or analysed during the current study available from the corresponding author on reasonable request.

References
1. Torrejon, J. *et al.* Neuromorphic computing with nanoscale spintronic oscillators. *Nature* **547**, 428 (2017).
2. Borders, W. A. *et al.* Analogue spin-orbit torque device for artificial-neural-network-based associative memory operation. *Appl. Phys. Express* **10**, 013007 (2017).
3. Kudo, K. & Morie, T. Self-feedback electrically coupled spin-Hall oscillator array for pattern-matching operation. *Appl. Phys. Express* **10**, 043001 (2017).
4. Grollier, J. *et al.* Neuromorphic spintronics. *Nat. Electron.* **3**, 360 (2020).
5. Maas, W., Natschläger, T. & Markram, H. Real-time computing without stable states: A new framework for neural computation based on perturbations. *Neural Comput.* **14**, 2531 (2002).
6. Jaeger, H. & Haas, H. Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. *Science* **304**, 78 (2004).

7. Verstraeten, D., Schrauwen, B., D'Haene, M. & Stroobandt, D. An experimental unification of reservoir computing methods. *Neural Netw.* **20**, 391 (2007).
8. Hermans, M. & Schrauwen, B. Memory in linear recurrent neural networks in continuous time. *Neural Netw.* **23**, 341 (2010).
9. Appeltant, L. *et al.* Information processing using a single dynamical node as complex system. *Nat. Commun.* **2**, 468 (2011).
10. Paquot, Y. *et al.* Optoelectronic reservoir computing. *Sci. Rep.* **2**, 287 (2012).
11. Brunner, D., Soriano, M. C., Mirasso, C. R. & Fischer, I. Parallel photonic information processing at gigabyte per second data rates using trasient states. *Nat. Commun.* **4**, 1364 (2013).
12. Nakajima, K., Hauser, H., Li, T. & Pfeifer, R. Information processing via physical soft body. *Sci. Rep.* **5**, 10487 (2015).
13. Nakayama, J., Kanno, K. & Uchida, A. Laser dynamical reservoir computing with consistency: an approach of a chaos mask signal. *Opt. Express* **24**, 8679–8692 (2016).
14. der Sande, G. V., Brunner, D. & Soriano, M. C. Advances in photonic reservoir computing. *Nanophotonics* **6**, 561–576 (2017).
15. Fujii, K. & Nakajima, K. Harnessing disordered-ensemble quantum dynamics for machine learning. *Phys. Rev. Appl.* **8**, 024030 (2017).
16. Dion, G., Mejaouri, S. & Sylvestre, J. Reservoir computing with a single delay-coupled non-linear mechanical oscillator. *J. Appl. Phys.* **124**, 152132 (2018).
17. Nakajima, K. Physical reservoir computing—An introductory perspective. *Jpn. J. Appl. Phys.* **59**, 060501 (2020).
18. Goto, K., Nakajima, K. & Notsu, H. Twin vortex computer in fluid flow. *New J. Phys.* **23**, 063051 (202).
19. Nakajima, K. & Fischer, I. (eds) *Reservoir Computing: Theory, Physical Implementations, and Applications* (Springer, Singapore, 2021).
20. Furuta, T. *et al.* Macromagnetic simulation for reservoir computing utilizing spin dynamics in magnetic tunnel junctions. *Phys. Rev. Appl.* **10**, 034063 (2018).
21. Tsunegi, S. *et al.* Evaluation of memory capacity of spin torque oscillator for recurrent neural networks. *Jpn. J. Appl. Phys.* **57**, 120307 (2018).
22. Bourianoff, G., Pinna, D., Sitte, M. & Everschor-Sitte, K. Potential implementation of reservoir computing models based on magnetic skyrmions. *AIP Adv.* **8**, 055602 (2018).
23. Nakane, R., Tanaka, G. & Hirose, A. Reservoir computing with spin waves excited in a garnet film. *IEEE Access* **6**, 4462 (2018).
24. Marković, D. *et al.* Reservoir computing with the frequency, phase, and amplitude of spin-torque nano-oscillators. *Appl. Phys. Lett.* **114**, 012409 (2019).
25. Tsunegi, S. *et al.* Physical reservoir computing based on spin torque oscillator with forced synchronization. *Appl. Phys. Lett.* **114**, 164101 (2019).
26. Riou, M. *et al.* Temporal patter recognition with delayed-feedback spin-torque nano-oscillators. *Phys. Rev. Appl.* **12**, 024049 (2019).
27. Nomura, H. *et al.* Reservoir computing with dipole-coupled nanomagnets. *Jpn. J. Appl. Phys.* **58**, 070901 (2019).
28. Yamaguchi, T. *et al.* Periodic structure of memory function in spintronics reservoir with feedback current. *Phys. Rev. Res.* **2**, 023389 (2020).
29. Yamaguchi, T. *et al.* Step-like dependence of memory function on pulse width in spintronics reservoir computing. *Sci. Rep.* **10**, 19536 (2020).
30. Akashi, N. *et al.* Input-driven bifurcations and information processing capacity in spintronics reservoirs. *Phys. Rev. Res.* **2**, 043303 (2020).
31. Slonczewski, J. C. Current-driven excitation of magnetic multilayers. *J. Magn. Magn. Mater.* **159**, L1 (1996).
32. Berger, L. Emission of spin waves by a magnetic multilayer traversed by a current. *Phys. Rev. B* **54**, 9353 (1996).
33. Weisheit, M. *et al.* Electric field-induced modification of magnetism in thin-film ferromagnets. *Science* **315**, 349 (2007).
34. Duan, C.-G. *et al.* Surface magnetoelectric effect in ferromagnetic metal films. *Phys. Rev. Lett.* **101**, 137201 (2008).
35. Maruyama, T. *et al.* Large voltage-induced magnetic anisotropy change in a few atomic layers of iron. *Nat. Nanotechnol.* **4**, 158 (2009).
36. Nakamura, K. *et al.* Giant modification of the magnetocrystalline anisotropy in transition-metal monolayers by an external electric field. *Phys. Rev. Lett.* **102**, 187201 (2009).
37. Tsujikawa, M. & Oda, T. Finite electric field effects in the large perpendicular magnetic anisotropy surface Pt/Fe/Pt(001): A first-principles study. *Phys. Rev. Lett.* **102**, 247203 (2009).
38. Shiota, Y. *et al.* Voltage-assisted magnetization switching in ultrahin Fe80 Co20 alloy layers. *Appl. Phys. Express* **2**, 063001 (2009).
39. Nozaki, T., Shiota, Y., Shiraishi, M., Shinjo, T. & Suzuki, Y. Voltage-induced perpendicular magnetic anisotropy change in magnetic tunnel junctions. *Appl. Phys. Lett.* **96**, 022506 (2010).
40. Shiota, Y. *et al.* Induction of coherent magnetization switching in a few atomic layers of FeCo using voltage pulses. *Nat. Mater.* **11**, 39 (2011).
41. Wang, W.-G., Li, M., Hageman, S. & Chien, C. L. Electric-field-assisted switching in magnetic tunnel junctions. *Nat. Mater.* **11**, 64 (2011).
42. Shiota, Y. *et al.* Pulse voltage-induced dynamic magnetization switching in magnetic tunnel junctions with high resistance-area product. *Appl. Phys. Lett.* **101**, 102406 (2012).
43. Kanai, S. *et al.* Electric field-induced magnetization reversal in a perpendicular-anisotropy CoFeB-MgO magnetic tunnel junction. *Appl. Phys. Lett.* **101**, 122403 (2012).
44. Grezes, C. *et al.* Ultra-low switching energy and scaling in electric-field-controlled nanoscale magnetic tunnel junctions with high resistance-area product. *Appl. Phys. Lett.* **108**, 012403 (2016).
45. Nozaki, T. *et al.* Highly effcient voltage control of spin and enhanced interfacial perpendicular magnetic anisotropy in iridium-doped Fe/MgO magnetic tunnel junctions. *NPG Asia Mater.* **9**, e451 (2017).
46. Miwa, S. *et al.* Voltage controlled interfacial magnetism through platinum orbits. *Nat. Commun.* **8**, 15848 (2017).
47. Okada, A., Kanai, S., Fukami, S., Sato, H. & Ohno, H. Electric-field effects on the easy cone angle of the easy-cone state in CoFeB/MgO investigated by ferromagnetic resonance. *Appl. Phys. Lett.* **112**, 172402 (2018).
48. Sugihara, A. *et al.* Evaluation of higher order magnetic anisotropy in a perpendicularly magnetized epitaxial ultrathin Fe layer and its applied voltage dependence. *Jpn. J. Appl. Phys.* **58**, 090905 (2019).
49. Yamamoto, T. *et al.* Improvement of write error rate in voltage-driven magnetization switching. *J. Phys. D Appl. Phys.* **52**, 164001 (2019).
50. Nozaki, T. *et al.* Voltage-cotrolled magnetic anisotropy in an ultrathin Ir-doped Fe layer with a CoFe termination layer. *APL Mater.* **8**, 011108 (2020).
51. Yakata, S. *et al.* Influence of perpendicular magnetic anisotropy on spin-transfer switching current in CoFeB/MgO/CoFeB magnetic tunnel junctions. *J. Appl. Phys.* **105**, 07D131 (2009).
52. Ikeda, S. *et al.* A perpendicular-anisotropy CoFeB-MgO magnetic tunnel junction. *Nat. Mater.* **9**, 721 (2010).
53. Kubota, H. *et al.* Enhancement of perpendicular magnetic anisotropy in FeB free layers using a thin MgO cap layer. *J. Appl. Phys.* **111**, 07C723 (2012).
54. Yakushiji, K., Fukushima, A., Kubota, H., Konoto, M. & Yuasa, S. Ultralow-voltage spin-transfer switching in perpendicularly magnetized magnetic tunnel junctions with synthetic antiferromagnetic reference layer. *Appl. Phys. Express* **6**, 113006 (2013).
55. Matsumoto, R., Arai, H., Yuasa, S. & Imamura, H. Spin-transfer-torque switching in a spin-valve nanopillar with a conically magnetized free layer. *Appl. Phys. Express* **8**, 063007 (2015).

56. Atiya, A. F. New results on recurrent network training: Unifying the algorithms and accelerating convergence. *IEEE Trans. Neural. Netw.* **11**, 697 (2000).
57. Bertschinger, N. & Natschläger, T. Real-time computation at the edge of chaos in recurrent neural networks. *Neural. Comput.* **16**, 1413 (2004).
58. Taniguchi, T. Synchronization and chaos in spin torque oscillator with two free layers. *AIP Adv.* **10**, 015112 (2020).

Author contributions

T.T. designed the project with help from S.T. A.O., Y.U. and T.T. developed the codes and performed the simulations. T.T. wrote the manuscript and prepared the figures. All authors contributed to discussing the results.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to T.T.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.