Reservoir Computing on Spin-Torque Oscillator Array

Taro Kanao ⊙,[*] Hirofumi Suto, Koichi Mizushima, Hayato Goto, Tetsufumi Tanamoto, and
Tazumi Nagasawa

*Corporate Research and Development Center, Toshiba Corporation, 1, Komukai-Toshiba-cho, Saiwai-ku,
Kawasaki 212-8582, Japan*

We numerically study reservoir computing on a spin-torque oscillator (STO) array, describing the magnetization dynamics of the STO array by a nonlinear oscillator model. The STOs exhibit synchronized oscillation due to coupling by magnetic dipolar fields. We show that reservoir computing can be performed using the synchronized oscillation state. The performance can be improved by increasing the number of STOs. The performance becomes highest at the boundary between the synchronized and disordered states. Using an STO array, we can achieve higher performance than that of an echo-state network with a similar number of units. This result indicates that STO arrays are promising for hardware implementation of reservoir computing.

I. INTRODUCTION

Reservoir computing [1–3], a framework for machine learning, has attracted much attention for realizing high-performance real-time computing for applications such as time-series modeling and prediction, signal processing, and pattern classification. Reservoir computing utilizes the complex dynamics of interacting nonlinear units. In reservoir computing, only output weights are learned in an external computer. This feature allows hardware implementations of reservoir computing, as well as reduction of the computational costs of the learning process. Hardware implementations have been proposed in several physical systems, such as analog electronic circuits [4], optical systems [5,6], quantum-bit systems [7], and magnetic systems [8–10]. Although these systems have shown good potential for reservoir computing, further improvements in computational capacity, operation speed, hardware size, energy consumption, and balance are required for practical applications.

Recently, reservoir computing on spin-torque oscillators (STOs) [11,12] has been reported [13–16]. (Applications of STOs to other types of computing have also been proposed [17–23].) An STO is a nonlinear auto-oscillator with a power-dependent frequency shift. An STO exhibits microwave oscillation induced by a spin-polarized direct current via a spin-transfer torque, yielding microwave magnetic and electromagnetic fields. Compared with the other systems above, STOs have the advantages of nanosecond-scale fast dynamics, nanometer-scale size, and low energy consumption. The microwave fields

can lead STOs to interact with each other in an array. Although the use of many interacting STOs is expected to achieve higher performance due to complex dynamics with many degrees of freedom, a *single* STO has been assumed in previous studies of STO reservoir computing [13–16].

In this paper, we investigate reservoir computing on an array of STOs coupled by magnetic dipolar fields, as shown in Fig. 1, where couplings naturally appear when the STOs are integrated closely with each other. We perform numerical experiments on reservoir computing utilizing a nonlinear oscillator model [12] corresponding to an STO array. As a result, we show that reservoir computing can be performed in the synchronized oscillation state of the array and that its performance can be improved by increasing the number of STOs. The highest performance is obtained at the boundary between the synchronized and disordered states. The performance can become higher than that of a standard neural-network model known as an echo-state network (ESN) [1,2].

II. RESERVOIR COMPUTING ON STO ARRAY

A. Model of STO array

We place N STOs on a square lattice with a size of $M \times M \, (= N)$ as shown in Fig. 1. In this paper, we assume that the STOs exhibit out-of-plane precession [24]. These STOs emit large microwave dipolar fields [25] and output signals due to the magnetoresistive effect [26,27], which lead to strong couplings between the STOs and a high signal-to-noise ratio in measurements, respectively. The oscillations of the STOs are synchronized, with the same phase [28–30]. [We show in Appendix A an example of
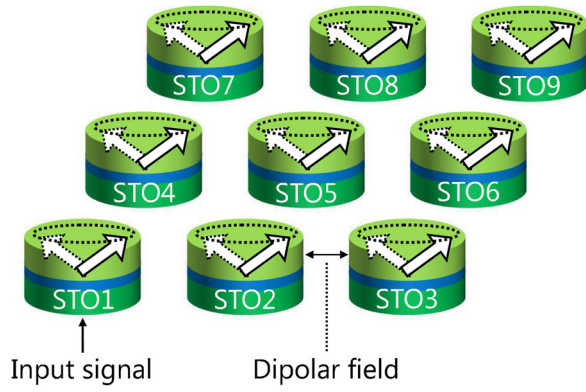
*taro.kanao@toshiba.co.jp

FIG. 1.   Schematic of a spin-torque oscillator array.

such synchronized oscillations by use of a numerical simulation based on the Landau-Lifshitz-Gilbert-Slonczewski (LLGS) equation [31].]

For reservoir computing, we utilize the dynamics of the oscillation powers and phases (or frequencies) of the STOs [13–16]. To calculate this dynamics, we use a corresponding nonlinear oscillator model [12,32], which can capture the essential features of the dynamics of the STOs, including the synchronization [33], irrespective of the detailed magnetization configuration of the STOs. The nonlinear oscillator model can be derived from the LLGS equation under certain assumptions. (See Appendix B for details.) The powers and frequencies of the STOs usually vary slowly compared with the oscillation itself [12,25], and fast oscillation components can be eliminated in the nonlinear oscillator model. Thus this model allows us to reduce the computational costs of the simulation.

The nonlinear oscillator model for N STOs is written as [12]

$$\frac{dc_i}{dt} = -\mathrm{i}\,\delta\omega_i(p_i)c_i - \Gamma_i(p_i)c_i - \mathrm{i}\sum_{j=1,j\neq i}^{N} \Omega_{ij}\,c_j, \quad (1)$$

where $t, c_i, \delta\omega_i(p_i), \Gamma_i(p_i)$, and Ω_{ij} $(i = 1, 2, \ldots, N)$ are the time, complex amplitudes representing the slowly varying magnetization dynamics, the nonlinear frequency shifts originating from magnetic anisotropy, the effective damping rates taking account of Gilbert damping and the spin-transfer torque, and the coupling strengths between the STOs caused by the dipolar fields, respectively. Here, a uniform high-frequency oscillation factor has been eliminated (see Sec. B 1). $\delta\omega_i(p_i)$ and $\Gamma_i(p_i)$ are supposed to depend on a dimensionless oscillation power $p_i = |c_i|^2$. Assuming a linear dependence of $\delta\omega_i(p_i)$ and $\Gamma_i(p_i)$ on p_i [12,32], we express them as

$$\delta\omega_i(p_i) = N_f\,(p_i - p_{i0}), \quad (2)$$

$$\Gamma_i(p_i) = \frac{\Gamma_p}{p_{i0}}\,(p_i - p_{i0}) - \xi_i(t), \quad (3)$$

where p_{i0}, N_f, Γ_p, and $\xi_i(t)$ are the stationary oscillation power, the nonlinear frequency shift, the rate of damping of power, and the modulation by an input signal, respectively, in units of frequency [12]. $\delta\omega_i(p_i)$ and $\Gamma_i(p_i)$ describe the properties of a single STO, and $\delta\omega_i(p_{i0}) = \Gamma_i(p_{i0}) = 0$ holds for its stationary oscillation state for a constant current. We add the input signal to the current, which, via the spin-transfer torque, leads to a modulation of $\Gamma_i(p_i)$. This modulation is represented by $-\xi_i(t)$. It is supposed that $\delta\omega_i(p_i)$ does not directly depend on the current, and instead is affected through the changes in p_i [12]. Equations (2) and (3) can be regarded as linear approximations of $\delta\omega_i(p_i)$ and $\Gamma_i(p_i)$ for small deviations of p_i from p_{i0} and for an input signal in the current. (See Sec. B 2 for details.)

In the array, the dipolar fields between the nearest-neighbor STOs, Ω, and between the next-nearest-neighbor STOs, $\Omega/\sqrt{8}$, are taken into account, as the dipolar field is inversely proportional to the cube of the distance. Differences between the STOs are modeled by taking p_{i0} as a Gaussian random variable with a mean of p_0. In this work, the input signal is applied to an STO at the corner of the array, namely STO1 in Fig. 1, and thus only $\xi_1(t)$ is modulated, and $\xi_i(t) = 0$ for $i \neq 1$. Based on reported values [32,34], we choose the following values: $p_0 = 0.5$, $N_f/(2\pi) = 0.1$ GHz, $\Gamma_p = 0.1$ ns^{-1}, $|\xi_1(t)| < 0.1$ ns^{-1}, and $\Omega = -0.1$ ns^{-1}, unless otherwise stated. The value of Ω is estimated in Sec. B 1. The standard deviation of p_{i0} is set to 0.01. The denominator p_{i0} in Eq. (3) is approximated by p_0. Equation (1) is numerically integrated with the initial condition $c_i = \sqrt{p_0}$, which is equivalent to all the magnetizations of a free layer of STOs being in the same direction. This situation can be realized in a strong uniform external field.
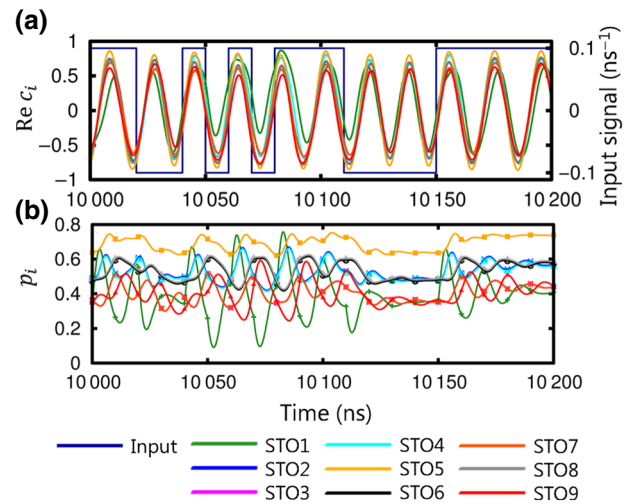

FIG. 2.   (a) Waveforms of input signal $\xi_1(t)$ and real parts of c_i for nine STOs with a sampling time $T_s = 10$ ns. (b) Oscillation powers p_i under the same conditions. Measured values of p_i are indicated by symbols.

Reservoir computing utilizes the dynamics of the waveforms induced by the input signals. Figure 2 shows examples of waveforms of the input signal $\xi_1(t)$ and of nine STOs obtained from Eq. (1). The waveforms after a sufficiently long time are shown. The oscillation of Re c_i in Fig. 2(a) is induced by the terms $\delta\omega_i(p_i)$ and Ω_{ij} in Eq. (1). It can be seen that the oscillation phases are nearly synchronized. Fluctuations of the oscillations are induced by the input signal. Since the input signal is applied to STO1, the power of STO1 is directly modulated by the input signal, as shown in Fig. 2(b). The effects of the input signal propagate to the other STOs. When the input signal is constant, the dynamics of the powers is damped in several tens of nanoseconds. The timescale of the damping of the powers is determined by $1/\Gamma_p$ $(= 10$ ns). The waveforms of the STOs at symmetric positions (see Fig. 1), i.e., (STO2 and 4), (STO3 and 7), and (STO6 and 8), almost overlap, but not completely, because of the deviation of p_{i0}.

B. Reservoir computing

Next, we explain reservoir computing on the STO array. We evaluate its performance by following the setting used in Ref. [7]. The kth input signal, s_k, is introduced into STO1 at a rate $1/T_s$ during the time interval between $T_s(k-1)$ and $T_s k$. The powers are measured at the end of each interval, in order to fully incorporate the effects of the latest signal. The measured values are denoted by $x_{ki} = p_i(T_s k)$. The output data are $y_k = \sum_{i=0}^{N} x_{ki} w_i$, where w_i is the ith weight for reservoir computing, and $x_{k0} = 1$ is introduced to treat a constant term. The weights are determined in a training process by minimizing $\sum_k (y_k - \bar{y}_k)^2$, so that a target series $\bar{y}_k$ is reproduced. These w_i are given by $\mathbf{w} = X^+ \bar{\mathbf{y}}$, where X is a matrix and $\mathbf{w}$ and $\bar{\mathbf{y}}$ are column vectors, with components x_{ki}, w_i, and $\bar{y}_k$, respectively, and where X^+ is the Moore-Penrose pseudoinverse matrix of X [7]. The dynamics of the STO array is calculated for a time of $5000T_s$. The first $1000T_s$ is discarded, and the subsequent $3000T_s$ and the last $1000T_s$ are used for training and evaluation, respectively.

We evaluate the performance by using short-term memory (STM) and parity-check (PC) tasks [7,10,14,16], because these tasks can estimate capacities for short-term memory and nonlinearity, which are necessary for applications to real-time computing. The STM task measures the STM capacity by outputting delayed inputs. The PC task can estimate nonlinearity because the required function is a nonlinear mapping. [We also evaluate a nonlinear autoregressive moving average (NARMA) task [7,35], and show the result in Appendix C.] For the STM and PC tasks, s_k is a random sequence of 0s and 1s. We convert s_k to $\xi_1(t)$ by use of a square wave with value $\xi_1(t) = 0.1(2s_k - 1)$ ns^{-1}. The targets of the STM and PC tasks are $\bar{y}_k = s_{k-\tau}$ and $\bar{y}_k = Q\left(\sum_{l=0}^{\tau} s_{k-l}\right)$, respectively, for $\tau = 0, 1, 2, \ldots$. The parity function $Q(S)$ is 0 for even S and 1 for odd S.

The capacities are measured by $C = \sum_{\tau=0}^{\tau_{\max}-1} C_\tau$, where C_τ is the square of the correlation coefficient of $\bar{y}_k$ and y_k, called a τ-delay capacity $(0 \le C_\tau \le 1)$. Here, $\tau_{\max} = 500$. A larger C_τ means higher performance, and C_τ reaches its maximum value of 1 when $\bar{y}_k = y_k$. The performances are evaluated for 20 sets of (p_{i0}, s_k), and the average values of the performances and standard deviations are calculated.

III. RESULTS AND DISCUSSION

A. Performance improvement with increasing N

Figures 3(a) and 3(b) show the STM and PC capacities as a function of the number of STOs N and the sampling time T_s. The STM capacity increases with N in the range shown $(N \le 400)$. The PC capacity saturates at large N. The behaviors of the capacities at larger N are discussed below. The STM and PC capacities depend on T_s in opposite ways; namely, the STM capacity decreases with increasing T_s, while the PC capacity increases. The STM capacity decreases with increasing T_s because the effect of the input signal decays when T_s is long. On the other hand, the PC capacity increases with increasing T_s because a longer T_s allows larger variations of the powers such that nonlinear effects appear. These behaviors occur in the range of 2 ns $\lesssim T_s \lesssim$ 50 ns, which can be understood from the timescale of the damping of power, $1/\Gamma_p = 10$ ns. Outside this range of T_s, the powers are no longer susceptible to influence from the input signals.

The data in Figs. 3(a) and 3(b) are summarized in Fig. 3(c). The capacities are distributed below a curve on which the product of the STM and PC capacities is a constant (equal to 85), implying a trade-off relation between the STM and PC capacities. This kind of relation has been reported as a memory-nonlinearity trade-off [36].

Figure 4 shows the capacities at each τ (the τ-delay capacities) for $N = 400$. The capacities decrease from 1 to 0 with increasing τ for $T_s \ge 10$ ns $= 1/\Gamma_p$. The capacities for $T_s < 1/\Gamma_p$ are small compared with 1 even at small τ, which indicates that T_s is too short for the powers to fully respond to the input signals. The STM capacities for $T_s \le 2$ ns show finite values at large τ, indicating that the effects of the input signals remain.

The largest values of the STM and PC capacities obtained in the present study are 46.4 at $N = 400$ and $T_s = 2$ ns, and 6.4 at $N = 196$ and $T_s = 50$ ns, respectively. These values are larger than the values for an ESN with 500 units reported in Ref. [7], which are about 17.6 and 6.0, respectively. Remarkably, the largest value of the STM capacity here is larger than twice the value for the ESN.

A possible reason for this large STM capacity is that the STOs have a phase degree of freedom, where the phases are time integrations of the oscillation frequencies in the past. In the STO array, the phases affect the time evolution of the powers, and a memory for longer time can be
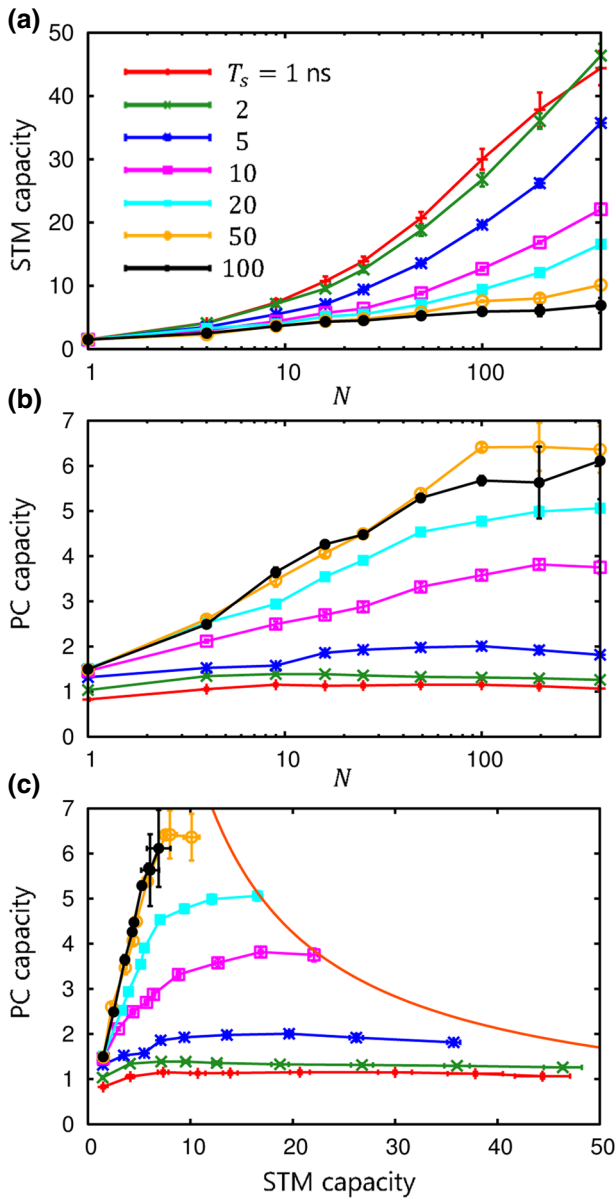
FIG. 3.   (a) Short-term memory and (b) parity-check capacities as a function of number of STOs N and sampling time T_s. T_s is indicated by common symbols and colors in (a)–(c). (c) STM capacity in (a) vs PC capacity in (b).

obtained from the powers. In an ESN, on the other hand, each unit has only one variable and shows damped motion in reservoir computing, which suppresses the past memory. Studies of reservoir computing on general nonlinear oscillator networks are an important topic for future work. Another difference which can lead to higher performance of an STO array than of an ESN is in the network structure. The STOs here are coupled by short-range interaction. This kind of weak connection allows various cooperative dynamics suitable for reservoir computing [1], which is discussed in the next section (Sec. III B). On the other
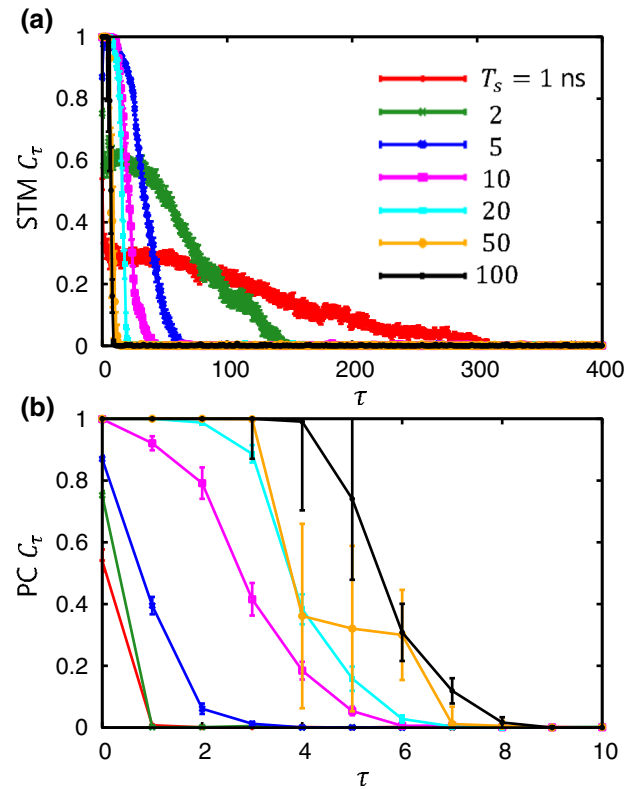

FIG. 4.   (a) STM and (b) PC capacities C_τ as a function of τ for $N = 400$.

hand, the ESN used for comparison is coupled by all-to-all connections, which tend to suppress the diversity of the dynamics.

B. Performance enhancement near transition

Finally, we investigate the dependence of the STM and PC capacities on the coupling, and find that the capacities depend strongly on the coupling strength. Figures 5(a) and 5(b) show the STM and PC capacities as a function of the coupling strength $|\Omega|$ and of N for $T_s = 5$ ns. The largest capacities appear around $|\Omega| = 0.14$ ns^{-1}.

To clarify the reason for these dependences of the capacities, we examine the oscillation states of the STO array under the corresponding conditions. The oscillation states are characterized by an order parameter ρ [37,38] and an average power $\bar{p}$,

$$\rho = \frac{1}{N}\left|\sum_{i=1}^{N} \frac{c_i}{|c_i|}\right|, \qquad (4)$$

$$\bar{p} = \frac{1}{N}\sum_{i=1}^{N} |c_i|^2, \qquad (5)$$

which are shown in Figs. 5(c) and 5(d), respectively. The order parameter ρ characterizes the uniformity of the
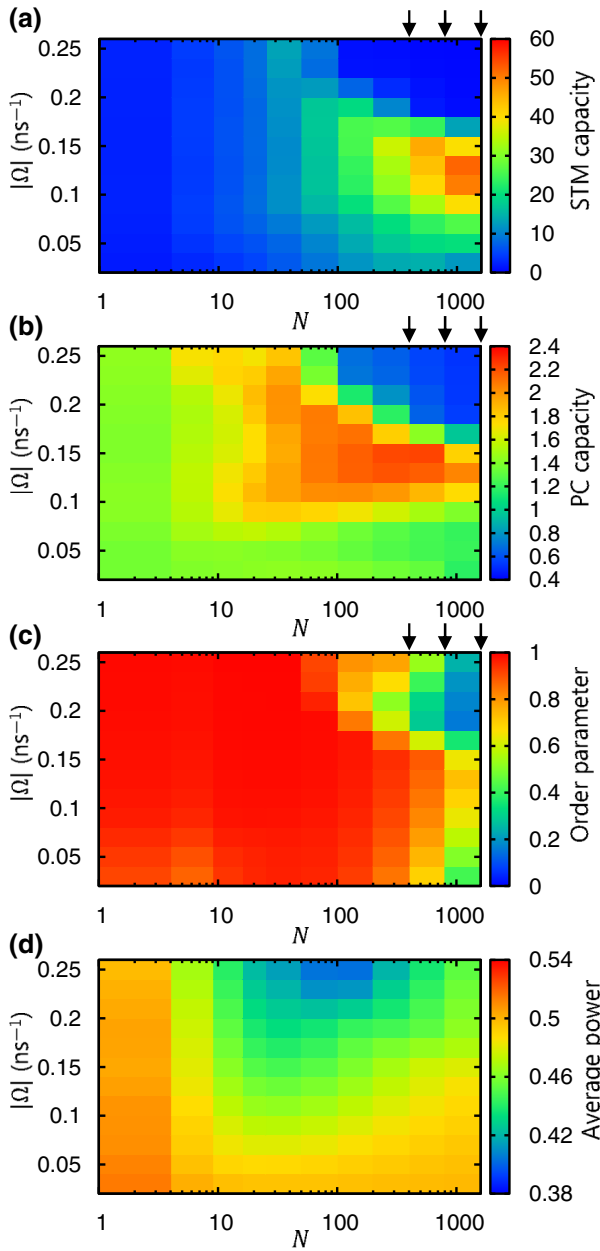
FIG. 5. (a) STM and (b) PC capacities, and (c) order parameter ρ and (d) average power $\bar{p}$ as a function of N and coupling strength $|\Omega|$ for $T_s = 5$ ns. Arrows indicate the values of N for which data are plotted in Fig. 6. The order parameter and the average power are averaged in the same way as for the STM and PC tasks.

oscillation phase, and reaches its largest value of 1 when all the STOs are perfectly synchronized.

Figure 5(c) shows that ρ is nearly 1 in most of the region, indicating that all c_i have nearly the same phases and thus the oscillations are synchronized. ρ decreases as N is increased for the same $|\Omega|$. The regions where the STM and PC capacities are finite in Figs. 5(a) and 5(b) are included in the region where ρ is rather large, $(\rho \gtrsim 0.5)$, in Fig. 5(c).

To see more clearly the relation between the performance and the oscillation state, we plot the capacities and the order parameter as a function of $|\Omega|$ in Fig. 6, where $N = 400 (= 20^2)$, $784 (= 28^2)$, and $1600 (= 40^2)$. As shown in Figs. 6(a) and 6(b), with increasing $|\Omega|$ the STM and PC capacities increase almost monotonically for $|\Omega| \lesssim 0.15$ ns^{-1} and then suddenly decrease. As shown in Fig. 6(c), the order parameter ρ exhibits similar behavior, which indicates a transition from the synchronized oscillation state to a disordered state around $|\Omega| = 0.15$ ns^{-1} (see below for details). From the results in Figs. 5 and 6, we conclude that reservoir computing can be performed in
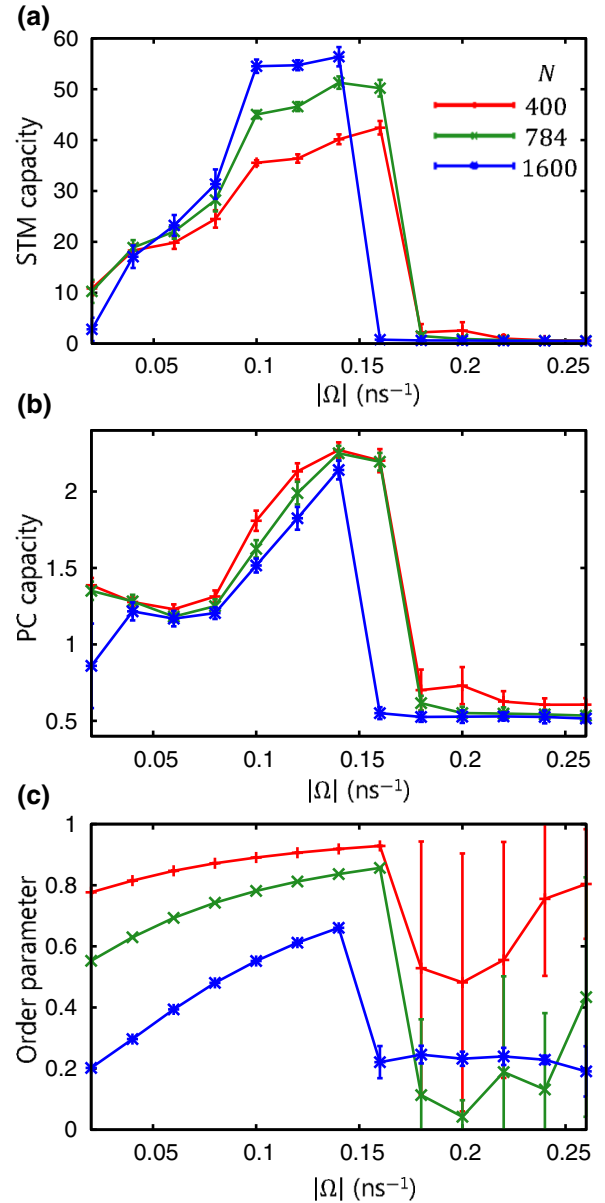

FIG. 6. (a) STM and (b) PC capacities, and (c) order parameter ρ as a function of $|\Omega|$ for $T_s = 5$ ns. Error bars indicate standard deviations for 20 sets of (p_{i0}, s_k).

the synchronized oscillation state of the STO array and the performance is highest near the transition to the disordered state, which is consistent with the reported behavior that the capacities of reservoir computing can be enhanced at the edge of this kind of transition [3,4].

The dependence of the oscillation state on the coupling strength observed in Fig. 6(c) can be understood as follows. The explanation is divided into two parts, for small and large $|\Omega|$. For small $|\Omega|$, the power is almost constant, $\bar{p} \simeq p_0 = 0.5$, as seen in Fig. 5(d), and thus the phase freedom is important. Then, the model of Eq. (1) can be reduced to a two-dimensional Kuramoto model with a short-range interaction [12,37,38]. In this Kuramoto model, it is known that stronger couplings lead to more ordered synchronization, while a larger number of oscillators causes a more disordered oscillation state [30,39]. These results for the Kuramoto model can explain the dependence of ρ on $|\Omega|$ and N for $|\Omega| \lesssim 0.15$ ns^{-1} in Fig. 6(c), where ρ increases with $|\Omega|$ and decreases with N. For large $|\Omega|$, on the other hand, the average power deviates from $p_0 = 0.5$, as seen in Fig. 5(d), and thus nonlinear effects of $\delta\omega_i(p_i)$ and $\Gamma_i(p_i)$ in Eq. (1) can appear. In this situation, it is known that the power degree of freedom causes disordered unsteady states, including chaotic states [38,40]. In these states the synchronization is suppressed, which corresponds to low ρ for $|\Omega| \gtrsim 0.15$ ns^{-1} in Fig. 6(c). These disordered unsteady states [38,40] can also explain the large standard deviations observed for the same region in Fig. 6(c), which indicates that the oscillation states are sensitive to small changes in the conditions. Between the synchronized state and the disordered state, for small and large $|\Omega|$, respectively, a transition occurs, causing a sharp change in ρ.

The decrease of ρ with increasing N is due to the short-range interaction between the STOs. By designing the STOs in the array to couple through a long-range interaction, e.g., by using a common electrode [21], one can make ρ finite for large N [37,38], allowing the synchronization of a large number of STOs.

IV. SUMMARY AND OUTLOOK

We investigate reservoir computing on an STO array coupled by dipolar fields. Numerical experiments on reservoir computing are performed by using a corresponding nonlinear oscillator model, which exhibits synchronized oscillation. As a result, we show that the performance of reservoir computing is improved by increasing the number of STOs in the synchronized oscillation state. The values of the STM and PC capacities for $N \leq 400$ can be larger than those for an ESN with 500 units. In particular, the maximum STM capacity is more than twice that of an ESN. The largest capacities are obtained at the boundary between the synchronized and disordered states. We

believe that the results can be explained by an enhancement of the capacities at the edge of chaos. The present results indicate that STO arrays are promising for the hardware implementation of reservoir computing. The performance can be further improved by introducing methods in reservoir computing not included here, such as time multiplexing. In this paper, the effects of thermal fluctuations [34], which are large in nanometer-sized magnetizations, are not included. This issue is left for future work.

To realize reservoir computing on an STO array, further considerations of several factors are necessary. We expect that an array of pillar-shaped magnetic tunnel junctions (MTJs) like those in magnetoresistive random access memories [41] or an array of giant magnetoresistive elements can be utilized as the STO array. MgO-based MTJs can exhibit large output signals due to the tunneling magnetoresistive effect. Typical ferromagnetic materials based on elements such as Fe and Co will yield enough coupling by dipolar fields, as estimated in Appendixes A and B. At the current time, a few STOs have been coupled by dipolar fields [29]. To increase the number of STOs, a one-dimensional array is the first candidate because the design of the electrodes which apply currents and extract signals will be simpler. With an increasing number of lines, the array approaches a two-dimensional one. Three-dimensional design of the electrodes will be necessary for integrating a large number of STOs into a two-dimensional array. Instead, reservoir computing may be possible when the number of measured signals is smaller than the number of STOs, for example by measuring signals from a fraction of the STOs or measuring sums of signals from several STOs, which reduces the complexity of the electrode configuration. In this case, estimation of the performance degradation caused by the reduced number of measured signals is necessary. Although we have assumed that the input signals are applied to one STO, reservoir computing can be performed by applying input signals to several STOs, which increases the freedom of design of the array. Input signals with waveforms other than a square wave, such as a distorted square wave or a triangle wave, will introduce similar fluctuations of the powers when the rates and amplitudes are suitable. In the measurement of the powers as the outputs, signal-processing technologies for amplitude modulation can be used [42].

ACKNOWLEDGMENTS

APPENDIX A: SYNCHRONIZATION OF STOS IN LLGS EQUATION

We show an example of the magnetization dynamics of STOs by solving the LLGS equation. The LLGS equation for the magnetizations of a free layer of STOs $\mathbf{m}_i$ ($i =$
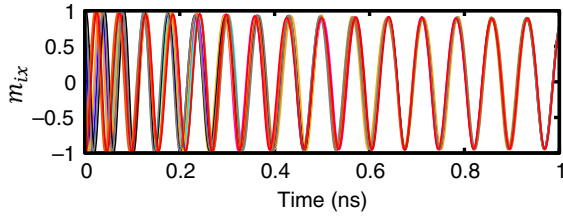
FIG. 7. Waveforms of m_{ix} for free layers of nine STOs obtained from the LLGS equation. $M_s = 1.6 \times 10^3$ emu/cm^3, $\alpha = 0.02$, and $H_z = 9$ kOe. Owing to H_z, $\mathbf{e}_p = (\cos 5°, 0, \sin 5°)$ is assumed. Neighboring STOs are separated by 50 nm. The mean and standard deviation of I_i are 20 mA and 0.2 mA, respectively. The initial directions of $\mathbf{m}_i$ are randomized in the plane of $m_{iz} = 0$.

$1, 2, \ldots, N$) normalized by the saturation magnetization M_s is given by

$$\frac{\partial \mathbf{m}_i}{\partial t} = -\gamma \mathbf{m}_i \times \mathbf{H}_i^{\mathrm{eff}} + \alpha \mathbf{m}_i \times \frac{\partial \mathbf{m}_i}{\partial t} + \mathbf{T}_i^S, \quad \text{(A1)}$$

where N, t, γ, and α are the number of STOs, the time, the gyromagnetic ratio, and the Gilbert damping, respectively. $\mathbf{H}_i^{\mathrm{eff}}$ and $\mathbf{T}_i^S = \sigma I_i \mathbf{m}_i \times (\mathbf{m}_i \times \mathbf{e}_p)$ are an effective field and a spin-torque term with efficiency σ, current I_i, and unit vector in the direction of the spin polarization of the current $\mathbf{e}_p$.

In the example, we choose an array of typical STOs with in-plane magnetized free and fixed layers, having a rectangular shape. A small ($50^2 \times 10$ nm^3) and thus uniform $\mathbf{m}_i$ is assumed. A perpendicular field H_z is applied, yielding an out-of-plane precession (OPP) for large I_i [24]. The STOs are placed on a square lattice, and all the magnetizations $\mathbf{m}_i$ are coupled by dipolar fields [43,44]. I_i and the initial directions of $\mathbf{m}_i$ are randomized. Figure 7 shows an example of the waveforms. Owing to the randomness, the waveforms are initially disordered. After several oscillation periods, the oscillations are synchronized. This state is obtained for other initial conditions, indicating the robustness of the synchronized state. This kind of synchronization has been reported in vortex-type STOs coupled by dipolar fields [28–30].

APPENDIX B: FROM THE LLGS EQUATION TO THE NONLINEAR OSCILLATOR MODEL

1. Coupling between STOs

Here, we show the correspondence between the LLGS equation and the nonlinear oscillator model. In the OPP, the x and y components of $\mathbf{m}_i$ are mainly oscillating, and we introduce complex amplitudes $\tilde{c}_i = m_{ix} - i m_{iy}$. From Eq. (A1), the equation of motion of $\tilde{c}_i$ is

$$\frac{d\tilde{c}_i}{dt} = -i\gamma H_{iz}^{\mathrm{eff}}\tilde{c}_i + i\gamma m_{iz}\left(H_{ix}^{\mathrm{eff}} - iH_{iy}^{\mathrm{eff}}\right) - \Gamma_i\tilde{c}_i, \quad \text{(B1)}$$

where the term including $\mathbf{H}_i^{\mathrm{eff}}$ is transformed exactly, while the damping and spin-torque terms are represented by using an effective damping rate Γ_i, which is discussed in Sec. B 2. Here, we assume uniform magnetizations in each STO. In the OPP, H_{iz}^{eff} mainly determines the precession axis and frequency $\tilde{\omega}_i = \gamma H_{iz}^{\mathrm{eff}}$. Besides, $H_{ix}^{\mathrm{eff}} - iH_{iy}^{\mathrm{eff}}$ is due to the dipolar fields from the other STOs. Although the prefactor includes m_{iz}, this prefactor depends little on $\tilde{c}_i$, because m_{iz} is almost constant for the OPP. The term including $H_{ix}^{\mathrm{eff}} - iH_{iy}^{\mathrm{eff}}$ reduces to the coupling term as follows. We approximate, for simplicity, the magnetization of a free layer of volume V by a point magnetic moment $\boldsymbol{\mu} = M_s V \mathbf{m}$. The dipolar field at the position $\mathbf{r} = (x, y, z)$ arising from this moment at the origin is given by

$$\mathbf{H} = 3\frac{(\boldsymbol{\mu} \cdot \mathbf{r}) \mathbf{r}}{r^5} - \frac{\boldsymbol{\mu}}{r^3}. \quad \text{(B2)}$$

Then,

$$H_x - iH_y = \frac{M_s V}{2r^5}\left[\left(x^2 + y^2 - 2z^2\right)\tilde{c} + 3(x - iy)^2\tilde{c}^* + 6(x - iy)zm_z\right]. \quad \text{(B3)}$$

On the same plane, i.e., $\mathbf{r} = (x, y, 0)$,

$$H_x - iH_y = \frac{M_s V}{2r^3}\left[\tilde{c} + 3\frac{(x - iy)^2}{r^2}\tilde{c}^*\right]. \quad \text{(B4)}$$

Thus Eq. (B1) can be written as

$$\frac{d\tilde{c}_i}{dt} = -i\tilde{\omega}_i\tilde{c}_i - i\sum_{j=1,j\neq i}^N \left(\Omega_{ij}\tilde{c}_j + \Xi_{ij}\tilde{c}_j^*\right) - \Gamma_i\tilde{c}_i, \quad \text{(B5)}$$

where $\Omega_{ij} = -\gamma m_{iz}M_s V/\left(2|\mathbf{r}_i - \mathbf{r}_j|^3\right)$ and $\Xi_{ij} = -3\gamma m_{iz}M_s V[x_i - x_j - i(y_i - y_j)]^2/\left(2|\mathbf{r}_i - \mathbf{r}_j|^5\right)$. Here $\mathbf{r}_i$ represents the position of STO i.

We assume that the STOs generate a common frequency ω when isolated in the steady state, and that the variation of the frequencies is caused by the nonlinear frequency shifts $\delta\omega_i$. Thus, the oscillation frequencies are represented by $\tilde{\omega}_i = \omega + \delta\omega_i$. Then, the uniform oscillation with frequency ω is decoupled by $\tilde{c}_i = c_i e^{-i\omega t}$ as

$$\frac{dc_i}{dt} = -i\,\delta\omega_i\,c_i - i\sum_j \left(\Omega_{ij}c_j + \Xi_{ij}c_j^* e^{2i\omega t}\right) - \Gamma_i c_i. \quad \text{(B6)}$$

Note that an oscillating factor with a frequency 2ω appears in the coefficient of c_j^*. This term does not affect the dynamics of c_i when c_i varies slowly, as shown in the following. We assume that the timescale of the change of c_i is much longer than the oscillation period $2\pi/\omega$ and that c_i

can be approximated by a constant during $2\pi/\omega$. Then, by averaging Eq. (B6) for a time π/ω, the term including c_j^* vanishes owing to the factor $e^{2i\omega t}$, and hence the nonlinear oscillator model,

$$\frac{dc_i}{dt} = -i\,\delta\omega_i\,c_i - \Gamma_i c_i - i\sum_{j=1,j\neq i}^{N}\Omega_{ij}c_j, \qquad \text{(B7)}$$

is obtained. This averaging is equivalent to using a variable

$$\bar{c}_i(t) = \frac{\omega}{\pi}\int_0^{\pi/\omega}dt'c_i(t+t'). \qquad \text{(B8)}$$

The numerical order of magnitude of $\left|\Omega_{ij}\right|$ can be estimated as follows. In the typical case of $m_{iz}^2 = 0.5$, $M_s = 1.6\times10^3$ emu/cm^3, and $V = 50^2\times10$ nm^3, $\left|\Omega_{ij}\right|$ is 0.031 ns^{-1} and 0.25 ns^{-1} for center-to-center distances of 200 nm and 100 nm, respectively, indicating that $\left|\Omega_{ij}\right|$ is of the order of 0.01 ns^{-1} to 0.1 ns^{-1}.

2. Effective damping rate and nonlinear frequency shift

The effective damping rate Γ_i is assumed to be a function of the current I_i, while the nonlinear frequency shift $\delta\omega_i$ is not [12]. The reason for this can be illustrated as follows. Since Γ_i and $\delta\omega_i$ describe the properties of a single STO, we focus on the single-STO case, in which the LLGS equation (A1) is written as

$$\frac{\partial\mathbf{m}}{\partial t} \simeq -\gamma\mathbf{m}\times\mathbf{H}^{\text{eff}} - \alpha\gamma\mathbf{m}\times\left(\mathbf{m}\times\mathbf{H}^{\text{eff}}\right)$$
$$+ \sigma I\mathbf{m}\times\left(\mathbf{m}\times\mathbf{e}_p\right). \qquad \text{(B9)}$$

Here, the Gilbert damping term is transformed to this form [the second term in the right-hand side of Eq. (B9)] by assuming $\alpha^2\ll1$ and $|\alpha\sigma I|\ll1$, which is valid for typical STOs. We consider the OPP and assume $\mathbf{H}^{\text{eff}}$ in the z direction. By supposing $\mathbf{e}_p = e_{px}\hat{\mathbf{x}} + e_{pz}\hat{\mathbf{z}}$ (where $\hat{\mathbf{x}}$ and $\hat{\mathbf{z}}$ are unit vectors in the x and z directions), the spin-torque term can be divided into two terms. The term including $e_{pz}\hat{\mathbf{z}}$ is collinear with the Gilbert damping term and can balance and cancel that term, allowing a stationary oscillation state. From these terms, the effective damping rate can be expressed as $\Gamma = \left(\alpha\gamma H_z^{\text{eff}} - \sigma I e_{pz}\right)m_z$, which depends explicitly on I and is almost constant for the OPP. On the other hand, the term including $e_{px}\hat{\mathbf{x}}$ can have a component parallel to the first torque term $-\gamma\mathbf{m}\times\mathbf{H}^{\text{eff}}$, which determines the oscillation frequency. The total contribution of the $e_{px}\hat{\mathbf{x}}$ term during one precession, however, is small because the component in the direction of the first torque term changes sign during the precession and cancels out [45]. The component in the direction of the first torque is proportional

to $-\left(\mathbf{m}\times\mathbf{H}^{\text{eff}}\right)\cdot\left[\mathbf{m}\times\left(\mathbf{m}\times e_{px}\hat{\mathbf{x}}\right)\right] = H_z^{\text{eff}}e_{px}m_y$, and the sign of m_y changes for the OPP during one precession. Thus the dependence of the oscillation frequency on the current can be neglected, and it is assumed that $\delta\omega$ is not an explicit function of I. Note that $\delta\omega$ is indirectly affected by I as a result of changes in the power p.

The forms of Γ_i and $\delta\omega_i$ used in Eqs. (2) and (3) are obtained as follows. In a stationary oscillation state of an individual STO with power p_{i0} and current I_{i0}, Γ_i is zero because the spin-transfer torque and the Gilbert damping are balanced and cancel each other. When a small modulation of I_i is introduced as an input signal, p_i deviates from p_{i0} as a result of time evolution. Then Γ_i can be linearly approximated as

$$\Gamma_i \simeq \frac{\partial\Gamma_i}{\partial p_i}\left(p_i - p_{i0}\right) + \frac{\partial\Gamma_i}{\partial I_i}\left(I_i - I_{i0}\right) \qquad \text{(B10)}$$

$$= \frac{\Gamma_p}{p_{i0}}\left(p_i - p_{i0}\right) - \xi_i(t), \qquad \text{(B11)}$$

where the first term is expressed by using a damping rate of the power [12]

$$\Gamma_p = p_{i0}\frac{\partial\Gamma_i}{\partial p_i}. \qquad \text{(B12)}$$

We express the second term yielded by the input signal as

$$\frac{\partial\Gamma_i}{\partial I_i}\left(I_i - I_{i0}\right) = -\xi_i(t). \qquad \text{(B13)}$$

The minus sign is because the spin-transfer torque works as a negative damping, which is proportional to the current. As we assume that $\delta\omega_i = 0$ at $p_i = p_{i0}$ and $\delta\omega_i$ does not depend explicitly on I_i, $\delta\omega_i$ is approximated as $\delta\omega_i \simeq N_f\left(p_i - p_{i0}\right)$. This kind of description of the current modulation has been developed in Ref. [12].

APPENDIX C: NARMA TASK

Here, we show the performance in the NARMA task, where series defined by nonlinear recurrence relations are emulated, and the ability to predict nonlinear time series is measured. We perform a NARMA task defined by the following equation: $y_{k+1} = 0.4y_k + 0.4y_k y_{k-1} + 0.6s_k^3 + 0.1$ for s_k consisting of uniform random numbers in $(0, 0.2)$, called the NARMA2 task [7,35]. We convert s_k to $\xi_1(t)$ by use of a square wave with value $\xi_1(t) = s_k - 0.1$ ns^{-1}. The performance is quantified by the normalized mean-squared error $\epsilon = \left[\sum_k\left(\bar{y}_k - y_k\right)^2\right]/\left(\sum_k\bar{y}_k^2\right)$, which is smaller when $\bar{y}_k$ and y_k are closer, indicating higher performance.

The results for the NARMA2 task are shown in Fig. 8. The error for $T_s = 20$ ns decreases with increasing N by three orders of magnitude, and the value of the error is comparable to that in previous studies [7].
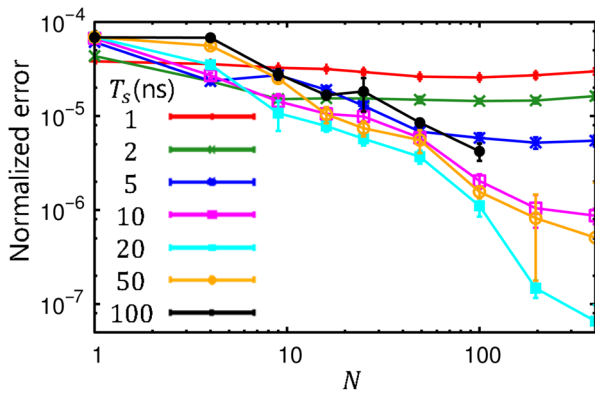
FIG. 8.   Normalized mean-squared error ϵ in NARMA2 task as a function of N and T_s.

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

[2] 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).

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

[4] 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).

[5] L. Larger, M. C. Soriano, D. Brunner, L. Appeltant, J. M. Gutierrez, L. Pesquera, C. R. Mirasso, and I. Fischer, Photonic information processing beyond Turing: An optoelectronic implementation of reservoir computing, Opt. Express **20**, 3241 (2012).

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

[7] K. Fujii and K. Nakajima, Harnessing Disordered-Ensemble Quantum Dynamics for Machine Learning, Phys. Rev. Appl. **8**, 024030 (2017).

[8] D. Prychynenko, M. Sitte, K. Litzius, B. Krüger, G. Bourianoff, M. Kläui, J. Sinova, and K. Everschor-Sitte, Magnetic Skyrmion as a Nonlinear Resistive Element: A Potential Building Block for Reservoir Computing, Phys. Rev. Appl. **9**, 014034 (2018).

[9] R. Nakane, G. Tanaka, and A. Hirose, Reservoir computing with spin waves excited in a garnet film, IEEE Access **6**, 4462 (2018).

[10] T. Furuta, K. Fujii, K. Nakajima, S. Tsunegi, H. Kubota, Y. Suzuki, and S. Miwa, Macromagnetic Simulation for Reservoir Computing Utilizing Spin Dynamics in Magnetic Tunnel Junctions, Phys. Rev. Appl. **10**, 034063 (2018).

[11] S. I. Kiselev, J. C. Sankey, I. N. Krivorotov, N. C. Emley, R. J. Schoelkopf, R. A. Buhrman, and D. C. Ralph, Microwave

[12] oscillations of a nanomagnet driven by a spin-polarized current, Nature (London) **425**, 380 (2003).

A. N. Slavin and V. Tiberkevich, Nonlinear auto-oscillator theory of microwave generation by spin-polarized current, IEEE Trans. Magn. **45**, 1875 (2009).

[13] J. Torrejon, M. Riou, F. Abreu 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 (London) **547**, 428 (2017).

[14] S. Tsunegi, T. Taniguchi, S. Miwa, K. Nakajima, K. Yakushiji, A. Fukushima, S. Yuasa, and H. Kubota, Evaluation of memory capacity of spin torque oscillator for recurrent neural networks, Jpn. J. Appl. Phys. **57**, 120307 (2018).

[15] D. Marković, N. Leroux, M. Riou, F. Abreu Araujo, J. Torrejon, D. Querlioz, A. Fukushima, S. Yuasa, J. Trastoy, P. Bortolotti, and J. Grollier, Reservoir computing with the frequency, phase, and amplitude of spin-torque nano-oscillators, Appl. Phys. Lett. **114**, 012409 (2019).

[16] S. Tsunegi, T. Taniguchi, K. Nakajima, S. Miwa, K. Yakushiji, A. Fukushima, S. Yuasa, and H. Kubota, Physical reservoir computing based on spin torque oscillator with forced synchronization, Appl. Phys. Lett. **114**, 164101 (2019).

[17] F. Macià, A. D. Kent, and F. C. Hoppensteadt, Spin-wave interference patterns created by spin-torque nano-oscillators for memory and computation, Nanotechnology **22**, 095301 (2011).

[18] T. Shibata, R. Zhang, S. P. Levitan, D. E. Nikonov, and G. I. Bourianoff, in *Proceedings of the 13th International Workshop on Cellular Nanoscale Networks and their Applications* (IEEE, New York, 2012), p. 1.

[19] G. Csaba and W. Porod, Computational study of spin-torque oscillator interactions for non-Boolean computing applications, IEEE Trans. Magn. **49**, 4447 (2013).

[20] J. Grollier, D. Querlioz, and M. D. Stiles, Spintronic nanodevices for bioinspired computing, Proc. IEEE **104**, 2024 (2016).

[21] K. Kudo and T. Morie, Self-feedback electrically coupled spin-Hall oscillator array for pattern-matching operation, Appl. Phys. Express **10**, 043001 (2017).

[22] D. Vodenicarevic, N. Locatelli, F. Abreu Araujo, J. Grollier, and D. Querlioz, A nanotechnology-ready computing scheme based on a weakly coupled oscillator network, Sci. Rep. **7**, 44772 (2017).

[23] M. Romera, P. Talatchian, S. Tsunegi, F. Abreu Araujo, V. Cros, P. Bortolotti, J. Trastoy, K. Yakushiji, A. Fukushima, H. Kubota, S. Yuasa, M. Ernoult, D. Vodenicarevic, T. Hirtzlin, N. Locatelli, D. Querlioz, and J. Grollier, Vowel recognition with four coupled spin-torque nano-oscillators, Nature (London) **563**, 230 (2018).

[24] S. I. Kiselev, J. C. Sankey, I. N. Krivorotov, N. C. Emley, M. Rinkoski, C. Perez, R. A. Buhrman, and D. C. Ralph, Current-Induced Nanomagnet Dynamics for Magnetic Fields Perpendicular to the Sample Plane, Phys. Rev. Lett. **93**, 036601 (2004).

[25] T. Kanao, H. Suto, K. Kudo, T. Nagasawa, K. Mizushima, and R. Sato, Transient magnetization dynamics of spin-torque oscillator and magnetic dot coupled by magnetic

dipolar interaction: Reading of magnetization direction using magnetic resonance, J. Appl. Phys. **123**, 043903 (2018).

[26] Z. Zeng, G. Finocchio, B. Zhang, P. Khalili Amiri, J. A. Katine, I. N. Krivorotov, Y. Huai, J. Langer, B. Azzerboni, K. L. Wang, and H. Jiang, Ultralow-current-density and bias-field-free spin-transfer nano-oscillator, Sci. Rep. **3**, 1426 (2013).

[27] H. Kubota, K. Yakushiji, A. Fukushima, S. Tamaru, M. Konoto, T. Nozaki, S. Ishibashi, T. Saruya, S. Yuasa, T. Taniguchi, H. Arai, and H. Imamura, Spin-torque oscillator based on magnetic tunnel junction with a perpendicularly magnetized free layer and in-plane magnetized polarizer, Appl. Phys. Express **6**, 103003 (2013).

[28] A. D. Belanovsky, N. Locatelli, P. N. Skirdkov, F. Abreu Araujo, J. Grollier, K. A. Zvezdin, V. Cros, and A. K. Zvezdin, Phase locking dynamics of dipolarly coupled vortex-based spin transfer oscillators, Phys. Rev. B **85**, 100409(R) (2012).

[29] N. Locatelli, A. Hamadeh, F. Abreu Araujo, A. D. Belanovsky, P. N. Skirdkov, R. Lebrun, V. V. Naletov, K. A. Zvezdin, M. Muñoz, J. Grollier, O. Klein, V. Cros, and G. de Loubens, Efficient synchronization of dipolarly coupled vortex-based spin transfer nano-oscillators, Sci. Rep. **5**, 17039 (2015).

[30] V. Flovik, F. Macià, and E. Wahlström, Describing synchronization and topological excitations in arrays of magnetic spin torque oscillators through the Kuramoto model, Sci. Rep. **6**, 32528 (2016).

[31] J. C. Slonczewski, Current-driven excitation of magnetic multilayers, J. Magn. Magn. Mater. **159**, L1 (1996).

[32] T. Kanao, T. Nagasawa, K. Kudo, H. Suto, M. Yamagishi, K. Mizushima, and R. Sato, Effects of power fluctuation on fast magnetic field detection using a spin-torque oscillator, Appl. Phys. Express **9**, 113001 (2016).

[33] A. N. Slavin and V. S. Tiberkevich, Theory of mutual phase locking of spin-torque nanosized oscillators, Phys. Rev. B **74**, 104401 (2006).

[34] M. Quinsat, D. Gusakova, J. F. Sierra, J. P. Michel, D. Houssameddine, B. Delaet, M.-C. Cyrille, U. Ebels, B. Dieny, L. D. Buda-Prejbeanu, J. A. Katine, D. Mauri, A. Zeltser, M. Prigent, J.-C. Nallatamby, and R. Sommet,

Amplitude and phase noise of magnetic tunnel junction oscillators, Appl. Phys. Lett. **97**, 182507 (2010).

[35] A. F. Atiya and A. G. Parlos, New results on recurrent network training: Unifying the algorithms and accelerating convergence, IEEE Trans. Neural Netw. **11**, 697 (2000).

[36] J. Dambre, D. Verstraeten, B. Schrauwen, and S. Massar, Information processing capacity of dynamical systems, Sci. Rep. **2**, 514 (2012).

[37] Y. Kuramoto and I. Nishikawa, Statistical macrodynamics of large dynamical systems. Case of a phase transition in oscillator communities, J. Stat. Phys. **49**, 569 (1987).

[38] J. A. Acebrón, L. L. Bonilla, C. J. Pérez Vicente, F. Ritort, and R. Spigler, The Kuramoto model: A simple paradigm for synchronization phenomena, Rev. Mod. Phys. **77**, 137 (2005).

[39] T. E. Lee, H. Tam, G. Refael, J. L. Rogers, and M. C. Cross, Vortices and the entrainment transition in the two-dimensional Kuramoto model, Phys. Rev. E **82**, 036202 (2010).

[40] P. C. Matthews and S. H. Strogatz, Phase Diagram for the Collective Behavior of Limit-Cycle Oscillators, Phys. Rev. Lett. **65**, 1701 (1990).

[41] S. Bhatti, R. Sbiaa, A. Hirohata, H. Ohno, S. Fukami, and S. N. Piramanayagam, Spintronics based random access memory: A review, Mater. Today **20**, 530 (2017).

[42] Y. Nakamura, M. Nishikawa, H. Osawa, Y. Okamoto, T. Kanao, and R. Sato, Envelope detection using temporal magnetization dynamics of resonantly interacting spin-torque oscillator, AIP Adv. **8**, 056512 (2018).

[43] K. Kudo, T. Nagasawa, H. Suto, T. Yang, K. Mizushima, and R. Sato, Influence of dynamical dipolar coupling on spin-torque-induced excitations in a magnetic tunnel junction nanopillar, J. Appl. Phys. **111**, 07C906 (2012).

[44] K. Kudo, H. Suto, T. Nagasawa, K. Mizushima, and R. Sato, Resonant magnetization switching induced by spin-torque-driven oscillations and its use in three-dimensional magnetic storage applications, Appl. Phys. Express **8**, 103001 (2015).

[45] T. Taniguchi, H. Arai, S. Tsunegi, S. Tamaru, H. Kubota, and H. Imamura, Critical field of spin torque oscillator with perpendicularly magnetized free layer, Appl. Phys. Express **6**, 123003 (2013).