Performance Enhancement of a Spin-Wave-Based Reservoir Computing System Utilizing Different Physical Conditions

Ryosho Nakane,[1,*] Akira Hirose,[1] and Gouhei Tanaka[1,2,3]

[1] *Department of Electronic Engineering and Information Systems, Graduate School of Engineering, The University of Tokyo, 7-3-1 Hongo Bunkyo-ku, Tokyo 113-8656, Japan*

[2] *International Research Center for Neurointelligence (IRCN), The University of Tokyo, 7-3-1 Hongo Bunkyo-ku, Tokyo 113-0033, Japan*

[3] *Department of Mathematical Informatics, Graduate School of Information Technology and Science, The University of Tokyo, Tokyo 113-8656, Japan*

(Received 13 September 2022; revised 10 January 2023; accepted 25 January 2023; published 14 March 2023)

We numerically study how to enhance reservoir computing performance by thoroughly extracting the spin-wave device potential for higher-dimensional information generation. The reservoir device has a 1-input exciter and 120-output detectors on the top of a continuous magnetic garnet film for spin-wave transmission. For various nonlinear and fading-memory dynamic phenomena distributing in the film space, small in-plane magnetic fields are used to prepare stripe domain structures and various damping constants at the film sides and bottom are explored. The ferromagnetic resonant frequency and relaxation time of spin precession clearly characterizes the change in spin dynamics with the magnetic field and damping constant. The common input signal for reservoir computing is a 1-GHz cosine wave with random 6-valued amplitude modulation. A basic 120-dimensional reservoir output vector is obtained from time-series signals at the 120-output detectors under each of three magnetic field conditions. Then, 240- and 360-dimensional reservoir output vectors are also constructed by concatenating two and three basic ones, respectively. In nonlinear autoregressive moving average (NARMA) prediction tasks, the computational performance is enhanced as the dimension of the reservoir output vector becomes higher and a significantly low prediction error is achieved for the tenth-order NARMA task using the 360-dimensional vector and optimum damping constant. The results are clear evidence that the collection of diverse output signals efficiently increases the dimensionality of the integrated reservoir state vector (i.e. reservoir-state richness) and thereby contributes to high computational performance. This paper demonstrates that performance enhancement through various configuration settings is a practical approach for on-chip reservoir computing devices with small numbers of real output nodes.

DOI: 10.1103/PhysRevApplied.19.034047

I. INTRODUCTION

Reservoir computing (RC) is an emerging machine-learning scheme suitable for information processing of time-series data [1,2]. Recently, RC utilizing a nonlinear physical system, called physical RC [3,4], has received much attention from electronics, because its low training cost and high adaptability for various information processing allow us to implement low-power and real-time computing hardware needed for edge computing on the internet of things (IoT) era [5,6]. So far, physical RC systems with various on-chip reservoir devices have been widely studied, which include device proposals, numerical experiments, and experimental demonstrations [3,4,7–15]. One of the requirements for high-performance RC is to extract high-dimensional output node states through rich physical dynamics in response to a time-series input signal, i.e., to obtain vast numbers of diverse temporal output signals from a reservoir device. In this respect, spin waves are very promising, since nonlinear phenomena, such as nonlinear propagation and interference, can induce space-distributing nonlinear dynamics in a continuous magnetic film [16–19], which can also exclude an intractable technical problem in fabricating enormous internal wirings. Furthermore, even after the fabrication of an on-chip device, physical dynamics can be reproducibly changed with the condition of the transmission medium, i.e., the magnetic material, through sensitive adjustment in

*Corresponding author. nakane@cryst.t.u-tokyo.ac.jp

configuration settings, such as the magnitude of an external magnetic field.

We proposed a spin-wave-based RC device on a chip and demonstrate its basic capabilities for information processing, using numerical experiments [9,20,21]. More-recent experimental RC with spin waves also indicates high potentials of this methodology in actual device operations [22]. At the present stage, high dimensionality in output node states is a key to further boost the computational performance. Thus, a practical approach for this purpose is strongly needed, where the most desirable strategy is to realize diverse dynamics using a single device structure.

In this study, we develop an efficient method to realize higher dimensionality in output node states for RC, by collecting output signals in response to a time-series input signal while the physical condition of the device is changed with the magnitude of an external magnetic field.

II. SIMULATION METHOD, MATERIAL PARAMETERS, AND DEVICE STRUCTURE

Figure 1 shows our RC system that is composed of a preprocessing part, a reservoir part, and a readout part. The role of the reservoir part is to nonlinearly transform a time-series input signal to higher-dimensional spatiotemporal output signals. First, we explain the spin-wave device shown as a top view, where a 320-nm-thick $Y_3Fe_5O_{12}$ (YIG) film with 12×12 μm^2 in area is used for the spin-wave signal transmission, a Cartesian coordinate system is defined with the origin at the center of the surface plane, one input exciter with 800 nm $\times$ 4 μm in area and 40 nm in depth is located at the center of the surface, and 120 output detectors with 200 nm $\times$ 200 nm in area and 40 nm in depth are regularly arranged along the x and y axes on the surface. Figure 2 shows the detailed arrangement of the exciter (red rectangle) and detectors (blue squares) on the YIG film. Hereafter, a specific detector is expressed by the following indexes in this order: R (right area) or L (left area), A–E (row), and 1–12 (column), e.g., R-A6.

In numerical experiments, we use a micromagnetic simulator Mumax 3 [23] based on the Landau-Lifshitz-Gilbert (LLG) equation [24]:

$$\frac{d\mathbf{M}}{dt} = \frac{\gamma}{1+\alpha^2}(\mathbf{M} \times \mathbf{H}^{\text{eff}} + \frac{\alpha}{M_S}(\mathbf{M} \times (\mathbf{M} \times \mathbf{H}^{\text{eff}})), \quad (1)$$

where $\mathbf{M}$ distributing in the YIG film is a spin having a saturation magnetization M_S and $\alpha \leq 1$ is a dimensionless damping constant. As explained below, $\mathbf{H}^{\text{eff}}$ is the effective field acting on $\mathbf{M}$ at a specific xyz position and it has magnetic field components of exchange magnetic fields from neighboring spins, an external magnetic field, a magnetocrystalline anisotropic field, and magnetic dipole fields from all other spins [25,26]. After the set of a simulation time step Δt, the time evolution of $\mathbf{M}$ is iteratively calculated with Eq. (1) using a Runge-Kutta method and a finite-difference discretization of space. The mesh inside the YIG film is a cubic with $40 \times 40 \times 40$ nm^3 along the Cartesian coordinates, $\mathbf{M}$ is located at every mesh corner, Δt is 1×10^{-12} s, and temperature is set at 0 K. We use material parameters consistent with those estimated from epitaxial YIG(111) films on a $Gd_3Ga_5O_{12}$(111) substrate: $M_S = 190$ kA/m (at low temperatures [27]), the uniaxial anisotropy along the z axis $K_U^H = 5.0$ KJ/m^3 [28], the exchange stiffness constant $A_{\text{ex}} = 3.7 \times 10^{-12}$ J/m [29], and the cubic magnetocrystalline anisotropy approximately 0. Hereafter, the x, y, and z components of the normalized spin s are

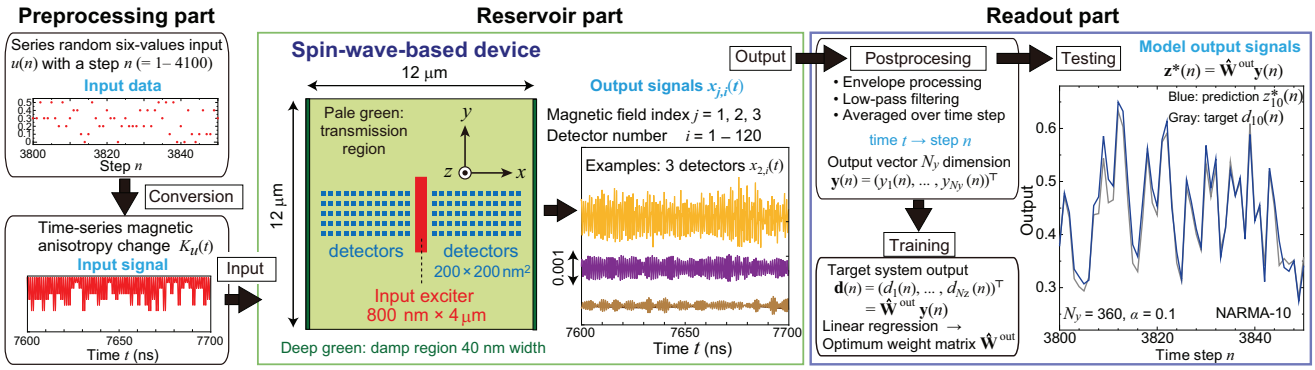

FIG. 1. Spin-wave-based RC system that consists of a preprocessing part, a reservoir part with a spin-wave-based device, and a readout part with adjustable weights. The spin-wave-based device has an input exciter with 800 nm $\times$ 4 μm $\times$ 40 nm and 120 detectors with 200 nm $\times$ 200 nm $\times$ 40 nm, where the input exciter is placed at the center of the xy plane and 12×5 detectors are regularly arranged in the right and left areas of the x-y plane (see Fig. 2). In the reservoir part, orange, purple, and brown curves are the output signals obtained at R-C1, R-C6, and R-C12 detectors, respectively (the detector index is defined in Fig. 2), for which an external magnetic field $\mu_0 H^{\text{ex}}$ is 0.5 mT (j index is 2) and the damping constant α in the damping region is 0.1. In the readout part, gray and blue curves show the tenth-order nonlinear autoregressive moving average (NARMA-10) target $d_{10}(n)$ and prediction $z^*_{10}(n)$ signals, respectively, in the testing phase of RC.

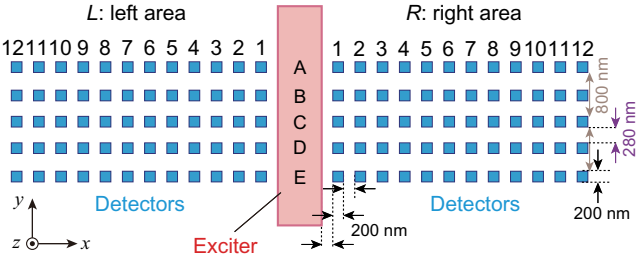

FIG. 2. Detailed arrangement of the input exciter and detectors on the surface of the YIG film, where a red rectangle and blue squares illustrate the exciter and detectors, respectively. A specific detector is expressed using the following indexes in this order: R (right area) or L (left area), A–E, and 1–12.

expressed by s_x, s_y, and s_z, respectively. The YIG film is divided into two parts: the signal transmission and damping regions. The damping region having a 40-nm width is located at the left and right sides (the boundaries at $x = \pm 6$ μm) and the bottom plane (the boundary at $z = -320$ nm). The damping constant α is set at $\alpha_0 = 0.001$ in the transmission region, whereas it is varied from 0.001 to 1 in the damping region to change the spin-wave dynamics. An external magnetic field H^{ex} is applied along the $+y$ direction and its magnitude is constant during a simulation batch. This study uses $\mu_0 H^{\mathrm{ex}} = 2 - 8$ mT to explore various spin-wave dynamics under smaller magnetic fields toward practical electronics. In the device operation, an input value is expressed by a corresponding K_U value in the exciter, while K_U^H remains unchanged in the rest region. The output value at each detector is determined by the averaged s_z inside.

III. FUNDAMENTAL PROPERTIES OF FERRIMAGNETIC GARNET FILM

A. Magnetic domain structure

In preparation for numerical experiments of spin waves, an initial spin distribution is formed using the following H^{ex} sequence. First, a uniform distribution along the y direction is set under $\mu_0 H^{\mathrm{ex}} = 100$ mT. Then, $\mu_0 H^{\mathrm{ex}}$ is decreased from 100 to 20 mT in steps of 10 mT and successively from 20 to 5 mT in steps of 1 mT, while the spin distribution is sufficiently relaxed at each $\mu_0 H^{\mathrm{ex}}$ value. This spin distribution is used for experiments under $\mu_0 H^{\mathrm{ex}} = 5$ mT. After that, spin distributions under $\mu_0 H^{\mathrm{ex}} = 2$ and 8 mT are prepared with decreasing and increasing $\mu_0 H^{\mathrm{ex}}$, respectively, in steps of 1 mT, and those are also used for experiments.

Figures 3(a), 3(b), and 3(c) show the top-view distributions of s_z under $\mu_0 H^{\mathrm{ex}} = 2$, 5, and 8 mT, respectively, where the Cartesian coordinate system is shown and the color bar represents the magnitude of s_z. The stripe domain structures are basically similar to each other, whereas the absolute magnitude of s_z decreases with the increase of

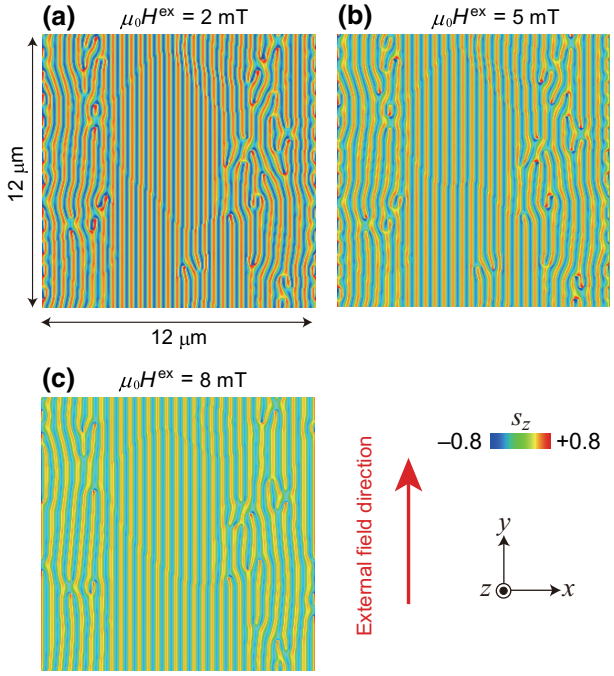

FIG. 3. Top-view distributions of the z component s_z of the normalized spin s in the magnetic garnet film under $\mu_0 H^{\mathrm{ex}}$ = (a) 2, (b) 5, and (c) 8 mT, respectively, where a Cartesian coordinate system is shown and a color bar represents the magnitude of s_z.

$\mu_0 H^{\mathrm{ex}}$. Our previous paper revealed that such domain structures are effective for RC [21]. Stripe domain structures under a low (or zero) $\mu_0 H^{\mathrm{ex}}$ along an in-plane direction are frequently observed in epitaxial oxide magnetic films with a large easy anisotropy perpendicular to the film plane, which have been experimentally and theoretically studied [26,30–35]. From the past knowledge, stripe domain structures having long-range strips without isolated domains, such as bubbles, are typically obtained when a spin distribution is relaxed under a slowly decreasing magnetic field along an in-plane direction [26]. Hence, stripe domain structures like those in Figs. 3(a), 3(b), and 3(c) can be reproducibly realized under normal gradual-relaxation procedures and our approach is feasible in practical devices. On the other hand, in Figs. 3(a), 3(b), and 3(c), the widths of red and blue strips decrease as $\mu_0 H^{\mathrm{ex}}$ is increased from 2 to 8 mT. This feature is consistent with those observed in experiments [26,30].

B. Characterization of spin dynamics

Fundamental properties of spin dynamics are characterized by the ferromagnetic resonant (FMR) frequency f_0 and the relaxation time τ_0 of the spin precession under $\mu_0 H^{\mathrm{ex}}$ (= 2 − 8 mT in steps of 1 mT). After the relaxation of a magnetic state under a specific $\mu_0 H^{\mathrm{ex}}$ value, $\mu_0 H^z = 0.1$ mT along the $+z$ direction is added at $t = 0$ and the averaged s_z over the entire film is recorded at

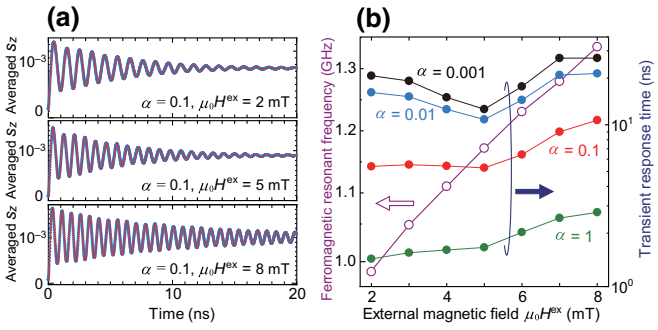

FIG. 4. (a) Time evolution of the averaged s_z after the application of a magnetic field along the $+z$ direction $\mu_0 H^z = 0.1$ mT at $t = 0$, where $\alpha = 0.1$, $\mu_0 H^{ex} = 2$, 5, and 8 mT (from top to bottom), blue dots represent calculated values, and a red curve at each $\mu_0 H^{ex}$ is the fitting curve. (b) Ferromagnetic resonant frequency f_0 and relaxation time τ_0 plotted against $\mu_0 H^{ex}$, where purple open circles represent f_0 values for all the α values and black, blue, red, and green closed circles represent the transient response time τ_0 values for $\alpha = 0.001$, 0.01, 0.1, and 1, respectively.

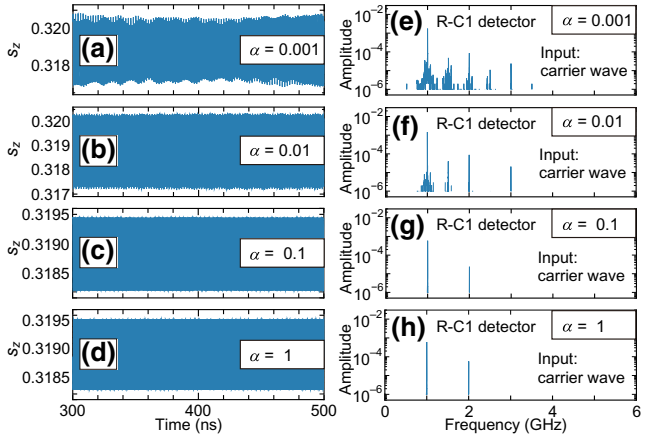

FIG. 5. (a)–(d) Output signals at the R-C1 detector under $\mu_0 H^{ex} = 5$ mT and (e)–(h) their DFT spectra, when a time-series 1-GHz cosine wave (carrier wave) having an amplitude $\Delta K_U = 500$ J/m^3 is fed into the input exciter: α = (a),(e) 0.001, (b),(f) 0.01, (c),(g) 0.1, and (d),(h) 1.

every 1×10^{-11} s. Finally, f_0 and τ_0 are estimated by fitting $C + A\sin(2\pi f_0 t + \phi_0)\exp(-t/\tau_0)$ to the signal, where C, A, and ϕ_0 are constants. Figure 4(a) shows time evolution of the averaged s_z after the application of $\mu_0 H^z = 0.1$ mT at $t = 0$, where $\alpha = 0.1$, $\mu_0 H^{ex} = 2$, 5, and 8 mT (from top to bottom), blue dots represent calculated values, and a red curve at each $\mu_0 H^{ex}$ is the fitting curve. Figure 4(b) shows f_0 and τ_0 plotted against $\mu_0 H^{ex}$, where purple open circles represent f_0 values for all the α values and black, blue, red, and green closed circles represent τ_0 values for $\alpha = 0.001$, 0.01, 0.1, and 1, respectively. Whereas f_0 monotonically increases with increasing $\mu_0 H^{ex}$, τ_0 varies with both $\mu_0 H^{ex}$ and α. These complicated properties are likely related with the stripe domain structure changing with $\mu_0 H^{ex}$, as shown in Figs. 3(a), 3(b), and 3(c). Those f_0 and τ_0 results indicate that the spin-wave dynamics is changed with $\mu_0 H^{ex}$ and α.

Based on the f_0 results, a cosine wave with 1 GHz is used as a carrier wave for input signals. To characterize fundamental properties of the carrier wave, a cosine wave with a constant amplitude $\Delta K_U = 500$ J/m^3 is fed into the input exciter and output waveforms at various detectors are observed under $\alpha = 0.001$, 0.01, 0.1, and 1. Figures 5(a)–5(d) show output signals under $\mu_0 H^{EX} = 5$ mT and $\alpha = 0.001$, 0.01, 0.1, and 1, respectively, which are obtained at the R-C1 detector. Figures 5(e)–5(h) show discrete Fourier transform (DFT) spectra calculated from Figs. 5(a)–5(d), respectively, for which the time length of an input signal is 1.3 μs, the output signals in the first 100-ns range are discarded, and the Hanning window is used. In all the α cases, the linear response to the input signal is characterized by an intense peak at 1 GHz, whereas nonlinear phenomena are characterized by components at integer harmonics (2 and 3 GHz) in the DFT spectra. When $\alpha = 0.001$ and 0.01, there are fluctuations in the amplitude of the waveforms, as seen in Figs. 5(a) and 5(b), which appear as a spread of the components at 1 and 2 GHz and interharmonic components in the DFT spectra in Figs. 5(e) and 5(f). In addition, the decrease of α from 0.01 to 0.001 leads to larger amplitudes of interharmonic components, including appearance of higher-order fractional hamonics at 2.5 and 3.5 GHz. Those features indicate that the strength of nonlinear phenomena increases with the decrease of α.

On the other hand, when $\alpha = 0.1$ and 1, the amplitudes of the waveforms in Figs. 5(c) and 5(d) are constant and the corresponding DFT spectra in Figs. 5(g) and 5(h) have components at 1 and 2 GHz. In consequence, the features are similar between these two α cases. For further observation, Appendix B provides waveforms and DFT spectra at R-C6 and R-C12 detectors under $\mu_0 H^{ex} = 5$ mT and $\alpha = 0.001$, 0.01, and 0.1, which allows confirmation that the main feature determined by α does not significantly change with the detector. The output signals of the carrier wave are closely related with RC, in particular, interharmonic components are found to have a significant impact on computational performance, as discussed in Sec. IV D. Note that the lower limit of the accuracy in DFT spectra is set at 1×10^{-6} in this study.

IV. RESERVOIR COMPUTING

A. Preprocessing and reservoir parts

In the preprocessing part, a sequential random input data $u(n)$ with a discrete time step $n(= 1, 2, \ldots, 4100)$ is converted to a time-series input signal $K_U(t)$ with a continuous time t: an input value (= 0, 0.1, 0.2, 0.3, 0.4, and 0.5) is converted to the corresponding $K_q(q = 0, 1, 2, 3, 4,$

and 5) value with a constant interval $\Delta K_q = 100$ J/m^3. The amplitude of the carrier wave is modulated by K_q with a time-step length $X = 2$ ns, as shown by a red curve in Fig. 6(a), where the amplitude in $(n-1)X \le t < nX$ is $K_{\mathrm{U}}^{\mathrm{H}} - K_q(t)$. In the reservoir part, $K_U(t)$ is fed into the input exciter and the response output signal $x_i(t)$ at the ith detector ($i = 1, 2, \ldots, 120$) is recorded at every 5×10^{-11} s. The correspondence between the output node number i and detector position in Fig. 2 is listed in Appendix A.

In the reservoir part in Fig. 1, three output signals are shown, where orange, purple, and brown curves are obtained at R-C1, R-C6, and R-C12 detectors, respectively, $\mu_0 H^{\mathrm{ex}} = 5$ mT, and $\alpha = 0.1$. Those results confirm that various $x_i(t)$ waveforms are generated through spatially distributed spin dynamics. Figure 6(b) shows DFT spectra calculated from the orange, purple, and brown signals in Fig. 1, for which the time length of an input signal is 8.2 μs, the output signals in the first 200 ns range are discarded, and the Hanning window is used. Unlike the spectrum in Fig. 5(g), all the spectra in Fig. 6(b) have interharmonic components, despite the same α value of 0.1. Thus, the information about the random 6-valued amplitude modulation in the input signal appears as interharmonic components in the output signals.

To analyze the effect of $\mu_0 H^{\mathrm{ex}}$ on output signals, waveforms at R-C1 detector and their DFT spectra are observed for $\alpha = 0.1$ and $\mu_0 H^{\mathrm{ex}} = 2$, 5, and 8 mT, as shown in Figs. 6(c) and 6(d), respectively, where blue, orange, green curves represent the results for $\mu_0 H^{\mathrm{ex}} = 2$, 5, and 8 mT, respectively, and a red curve in (c) represents the waveform of the input signal. Those results clearly characterize that the output signal at the same detector differs with $\mu_0 H^{\mathrm{ex}}$, whereas a peaked DFT component always appears at the forced oscillation frequency 1 GHz, irrespective of f_0 changing with $\mu_0 H^{\mathrm{ex}}$. Thus, it is found that the change in $\mu_0 H^{\mathrm{ex}}$ works well to efficiently generate different $x_i(t)$ at the single ith detector. Note that the amplitude of the waveform in Fig. 6(c) is the largest at $\mu_0 H^{\mathrm{ex}} = 2$ mT, which is probably due to the fact that 1 GHz is nearly the same as the f_0 value for this condition [Fig. 4(b)]. To characterize the output signals further, waveforms and their DFT spectra at all the detectors are observed to find that they vary with detector and $\mu_0 H^{\mathrm{ex}}$ (several selected DFT spectra are shown in Appendix C). The relation between computational performance and diversity in the output signals is discussed in Sec. V B.

It is noteworthy that this study uses a 1-GHz cosine wave as a carrier wave instead of time-series triangular pulses in our previous paper [21]. From the analyses in both cases, harmonic components are not necessary in an input signal to realize diverse nonlinear dynamics in output signals. Besides, nonlinear phenomena are also obtained even when f_0 significantly differs from the input carrier frequency ($\mu_0 H^{\mathrm{ex}} = 5$ and 8 mT conditions). Thus, it is expected that a wide range of carrier-wave parameters can be used. This scalability can be a technical advantage of the spin-wave-based device.

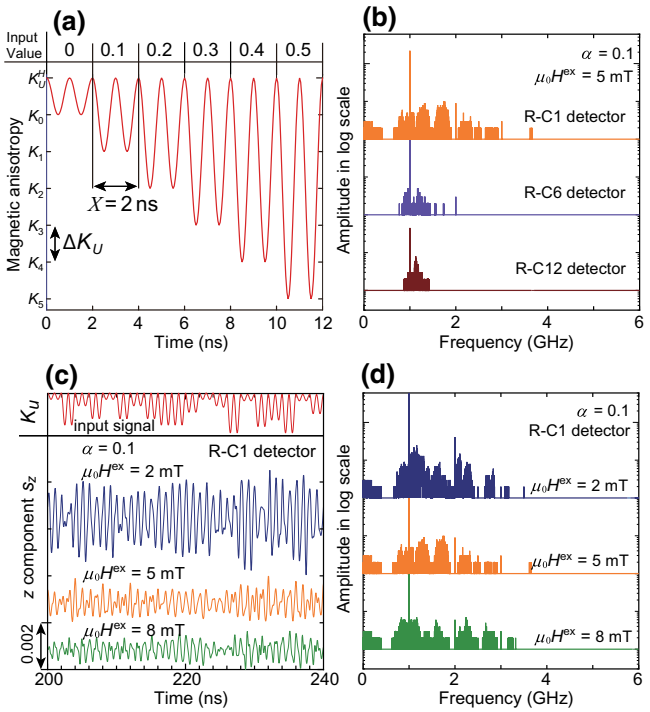

FIG. 6. (a) Input value ($= 0 - 0.5$) and the corresponding input signal that is a time-series K_U, where the carrier wave is a 1-GHz cosine wave, $X = 2$ ns is the time step, the maximum $K_U^H = 5$ kJ/m^3, and $\Delta K_U = 100$ J/m^3. (b) DFT spectra calculated from three output signals (Fig. 1), where orange, purple, and brown curves are obtained at R-C1, RC-6, and RC-12 detectors, respectively, $\mu_0 H^{\mathrm{ex}} = 5$ mT, and $\alpha = 0.1$. (c) Waveforms and (d) corresponding DFT spectra at the R-C1 detector, where $\alpha = 0.1$, dense-blue, orange, and green curves are the results under $\mu_0 H^{\mathrm{ex}} = 2$, 5, and 8 mT, respectively, and a red curve in (c) represents the waveform of the input signal. In (b),(c),(d), each curve is vertically shifted to facilitate visualization.

B. Readout part

In the readout part, $x_i(t)$ are converted to the time-step reservoir output state $y_i(n)$ ($n = 1, 2, \ldots, 4100$) through postprocessing. In our simulation, the 6-valued input information is encoded by the 1-GHz carrier wave with the 6-level amplitude, as shown in Fig. 6(a). Thus, it is expected that the amplitudes of the output waveforms at the detectors significantly vary depending on the time-series input information. This consideration also comes from common amplitude modulation and demodulation methods in wireless communication and the physical RC study using a spin-torque oscillator [8]. To extract temporal amplitude, three methods are used: an envelope processing (Env), a low-pass filtering (LPF), and the combination of Env and LPF. In those processing methods, the time window is set at 1 ns so that nonlinear postprocessing does not affect

short-term memory over $X\,(=2\,\mathrm{ns})$ and the capability of the RC device can be simply exploited. Another worthwhile note is that the delay time (0.5 ns) after LPF is calibrated to exclude the influence of short-term memory. Hereafter, some combinations of processing methods are explored and the processing sequence is shown by a series expression in the processing order. For example, Env-LPF means successive Env and LPF (first Env and then LPF) and our expectation for this combination is to enhance the capability of the information extraction.

Figures 7(a) and 7(b) show the waveforms after three methods, where a gray curve represents the output waveform at R-C1 detector and pale-blue, pale-green, and pale-red narrow curves represent the signals after Env, LPF, and Env-LPF, respectively. The resultant signals are not so gentle because of the short time window in the processing. Whereas the difference between the signals after Env and LPF is less significant, the signal after Env-LPF has a more gradual change than the others. After those processing methods, $y_i(n)$ (one value at each time step n) is extracted using three processing methods: averaging signal over X (Ave), taking a value at the middle of the time step (Mid), and taking the maximum value in the time step (Max). For visual evaluation, $y_i(n)$ is drawn over the time-step duration X in Figs. 7(a) and 7(b), where blue, green, and red steplike curves represent the $y_i(n)$ values after Env-Ave, LPF-Ave, and Env-LPF-Ave, respectively. The changes in the curves are seemingly similar to each other, nonetheless those possibly lead to different computational performances. In this section, RC using $y_i(n)$ after Env-LPF-Ave is performed, whereas the effects of the postprocessing methods on computational performance will be discussed in Sec. V C.

Hereafter, $x_i(t)$ and $y_i(n)$ obtained under $\mu_0 H^{\mathrm{ex}} = 2$, 5, and 8 mT are denoted by $x_{j,i}(t)$ and $y_{j,i}(n)$, corresponding to $j = 1$, 2, and 3, respectively. To exclude unstable responses, $y_{j,i}(n)$ in the first 100 steps are discarded, $y_{j,i}(n)$ in $101 \le n \le 3598$ are used for training, and the remaining $y_{j,i}(n)$ in $3599 \le n \le 4098$ are used for testing. Note that $y_{j,i}(n)$ in the last two n steps ($n = 4099$ and 4100) are also discarded because of higher error in the envelope processing at the end of the sequence. Next, the reservoir output vector $\mathbf{y}_j(n)$ is given through a row-wise collection of $y_{j,i}(n)$:

$$\mathbf{y}_j(n) = (y_{j,1}(n), \ldots, y_{j,120}(n))^{\mathsf{T}} \in \mathbb{R}^{120}. \tag{2}$$

Then, combination of $\mathbf{y}_j(n)$ with different j values are considered to form a reservoir state vector as follows:

$$\mathbf{y}(n) = \begin{cases} \mathbf{y}_1(n), \mathbf{y}_2(n), \text{or } \mathbf{y}_3(n) & \text{for } N_y = 120 \\ [\mathbf{y}_1(n); \mathbf{y}_2(n)], [\mathbf{y}_2(n); \mathbf{y}_3(n)], \text{or } [\mathbf{y}_3(n); \mathbf{y}_1(n)] & \\ \qquad\qquad \text{for } N_y = 240\,, & \\ [\mathbf{y}_1(n); \mathbf{y}_2(n); \mathbf{y}_3(n)] & \text{for } N_y = 360 \end{cases}$$

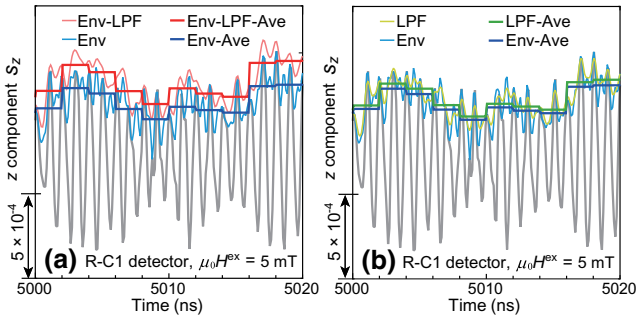

FIG. 7. (a),(b) Output waveform and waveforms after three postprocessing methods, where a gray curve represents the output waveform at R-C1 detector and pale-blue, pale-green, and pale-red narrow curves represent the signals after an envelope processing (Env), a low-pass filtering (LPF), and successive Env and LPF processings (Env-LPF), respectively. After those processings, $y_i(n)$ (one value at each time step n) is extracted using three methods (one example is averaging signal over time-step duration X (Ave)). In the figures, $y_i(n)$ drawn over the time duration X is also shown, where blue, green, and red steplike curves represent the $y_i(n)$ values after Env-Ave, LPF-Ave, and Env-LPF-Ave, respectively.

where N_y is the dimension of a vector and $[\mathbf{a}; \mathbf{b}]$ represents a vertical concatenation of vectors. Using the system output at the lth node $z_l(n)$, a system output vector $\mathbf{z}(n)$ is given by

$$\mathbf{z}(n) = (z_1(n), \ldots, z_{N_z}(n))^{\mathsf{T}} \in \mathbb{R}^{N_z}. \tag{3}$$

The output weight matrix $\mathbf{W}^{\mathrm{out}}$ is defined as follows:

$$\mathbf{W}^{\mathrm{out}} = (w_{li}^{\mathrm{out}}) \in \mathbb{R}^{N_z \times N_y}. \tag{4}$$

Then, the system output is given by

$$\mathbf{z}(n) = \mathbf{W}^{\mathrm{out}} \mathbf{y}(n). \tag{5}$$

In the training phase, the target of the system output vector $\mathbf{d}(n)$ is given by a collection of the target (desired) signal $d_l(n)$ for the lth system output:

$$\mathbf{d}(n) = (d_1(n), \ldots, d_{N_z}(n))^{\mathsf{T}} \in \mathbb{R}^{N_z}. \tag{6}$$

The teacher collection matrix $\mathbf{D}$ and reservoir state collection matrix $\mathbf{Y}$ are formed by column-wise collections of $\mathbf{d}(n)$ and $\mathbf{y}(n)$ for $101 \le n \le 3598$, respectively. The optimum weight matrix $\hat{\mathbf{W}}^{\mathrm{out}}$ is calculated by [36]

$$\hat{\mathbf{W}}^{\mathrm{out}} = \mathbf{D}\mathbf{Y}^{\dagger}, \tag{7}$$

where $\mathbf{Y}^{\dagger}$ is the Moore-Penrose pseudoinverse of $\mathbf{Y}$. In the testing phase, the system output vector $\mathbf{z}^*(n)$ is obtained by

$$\mathbf{z}^*(n) = \hat{\mathbf{W}}^{\mathrm{out}} \mathbf{y}(n). \tag{8}$$

C. Benchmark tasks

We perform two time-series benchmark tasks that are widely used for evaluating RC systems. The first one is to predict time-series data generated from the mth-order nonlinear autoregressive moving average (NARMA-m) model ($m = 2, \ldots, 10$) derived from Eq. (86) in Ref. [37]. A time-series target $d_m(n+1) = z_m(n+1)$ is iteratively generated using the following relation:

$$z_m(n+1) = 0.3 z_m(n) + 0.5 z_m(n) \sum_{s=0}^{m-1} z_m(n-s) + 1.5 u(n-m-1) u(n) + 1. \quad (9)$$

Since the state of the model depends on the past m states, an RC system for the prediction task requires more short-term memory of past inputs as the order m increases. The performance is evaluated by the normalized mean square error (E_m) for the NARMA-m task:

$$E_m = \frac{1}{L} \frac{\sum_{n=1}^{L} (d_m(n) - z_m^*(n))^2}{\mathrm{Var}(d_m(n))}, \quad (10)$$

where Var denotes variance and $L = 500$ is the step length in the testing duration. The second one is delay tasks whose targets are given by

$$d_k(n) = u(n-k) \quad \text{for delay } k = 1, \ldots, 120. \quad (11)$$

The performance is evaluated by memory capacity [38]:

$$C_k = \frac{\mathrm{Cov}^2(u(n-k), z^*(n,k))}{\mathrm{Var}(u(n))\mathrm{Var}(z^*(n,k))}, \quad (12)$$

where Cov and $z^*(n,k)$ denote covariance and the predicted value for $d_k(n)$, respectively, and the time-step length is 400. The total MC is the summation of C_k over $k = 1 - 120$.

D. Results and analysis

In Fig. 1, time-evolutional NARMA-10 signals are shown in the readout part, where gray and blue curves represent the target $d_{10}(n)$ and predicted $z_{10}^*(n)$ signals in the testing phase, respectively, under $N_y = 360$ and $\alpha = 0.1$. In the figure, $z_{10}^*(n)$ reasonably reproduces $d_{10}(n)$. Hence, many $x_{j,i}(t)$ have both short-term memory and nonlinearity effective for RC and the post signal processing extracted such properties for high dimensionality in $\mathbf{y}(n)$.

Figures 8(a) and 8(b) show NARMA-5 and NARMA-8 signals under the same $N_y = 360$ and $\alpha = 0.1$ condition, respectively, where gray and blue curves represent the target $d_m(n)$ and prediction $z_m^*(n)$ signals, respectively. For lower m values, $z_m^*(n)$ more accurately reproduces $d_m(n)$. This is reasonable since the prediction becomes harder for the task with higher m. Figures 8(c) and 8(d) show NARMA-10 signals under $N_y = 360$ and $\alpha = 0.001$ and 0.01, respectively. It is found that $z_m^*(n)$ more accurately reproduces $d_m(n)$ with the increase of α from 0.001 to 0.1. In the same manner, NARMA-m signals are calculated for $N_y = 360$ and E_m are plotted against m, as shown in Fig. 9(a), where black squares, blue triangles, red circles, and green diamonds are estimated values for $\alpha = 0.001$, 0.01, 0.1, and 1, respectively. In the entire m range, E_m decreases as α is increased from 0.001 to 0.1, whereas it increases as α is increased from 0.1 to 1.

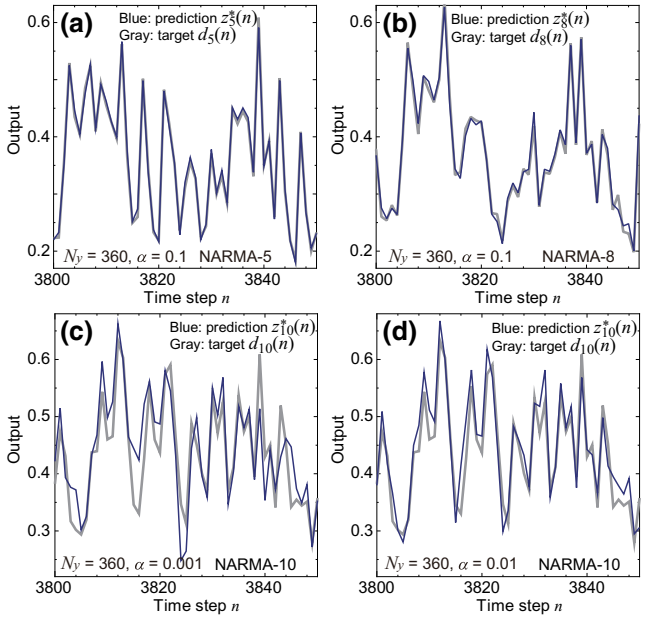

FIG. 8. (a) NARMA-5 and (b) NARMA-8 signals under $N_y = 360$ and $\alpha = 0.1$. (c),(d) NARMA-10 signals under $N_y = 360$ and $\alpha =$ (c) 0.001 and (d) 0.01. The gray and blue curves represent the target $d_m(n)$ and prediction $z_m^*(n)$ signals, respectively.

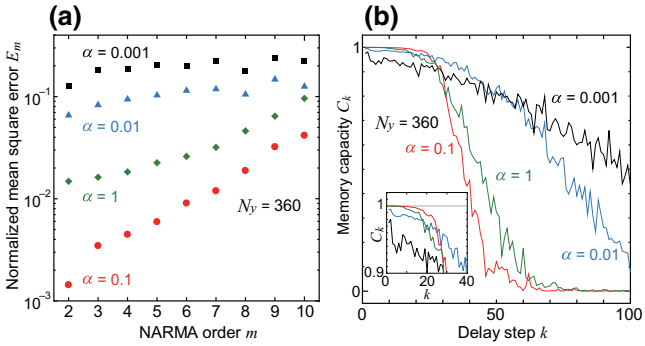

FIG. 9. (a) Normalized mean square error for mth-order NARMA (E_m) plotted against m and (b) memory capacity C_k plotted against time-step delay k when $N_y = 360$, where black, blue, red, and green colors are estimated values for $\alpha = 0.001$, 0.01, 0.1, and 1, respectively. The inset is a close-up view of $C_k \sim 1$ at around $k = 0$.

To analyze the results in Fig. 9(a), C_k is plotted against k, as shown in Fig. 9(b), where black, blue, red, and green curves are estimated memory curves for $\alpha = 0.001$, 0.01, 0.1, and 1, respectively. As α is decreased from 0.1 to 0.001, the k length for $C_k \sim 1$ decreases and the slope of the curve becomes more gradual. This trend is very similar to that of the memory curve with increasing noise in an input sequence in Ref. [38]. This can be interpreted as follows: fluctuation of the carrier wave increases with the decrease of α from 0.01 to 0.001 and it acts as noise in the derivation of $\hat{\mathbf{W}}^{\text{out}}$. This is because output behavior characterized by interharmonic components in a DFT spectrum can arise from both the fluctuation of the carrier wave [Figs. 5(e) and 5(f)] and the random 6-valued input information [Figs. 6(b) and 6(d)]. This "noise effect" is probably related with the large E_m values.

On the other hand, the difference between the results with $\alpha = 0.1$ and 1 possibly originates from the difference in τ_0: $\tau_0 = 5.3 - 10.7$ ns for $\alpha = 0.1$ and $\tau_0 = 1.5 - 2.7$ ns for $\alpha = 1$ [Fig. 4(b)]. It is expected that the response signals are significantly relaxed during $X = 2$ ns when $\alpha = 1$. For RC, nonlinear responses with short-term memories of past inputs are necessary. Thus, although it has not been fully clarified, many studies revealed that the time step X moderately smaller than τ_0 is useful for high computing capabilities [3,7,8]. In this respect, the $\alpha = 1$ condition does not meet the requirement. We consider that the highest performance with $\alpha = 0.1$ is achieved through the above two beneficial conditions: no interharmonics components in the carrier wave and $X < \tau_0$.

The relation between E_m in Fig. 9(a) and C_k in Fig. 9(b) is further analyzed. When the total MC for the α values are compared with each other, the $\alpha = 0.1$ condition has the smallest value of 38. However, when C_k at around smaller k is focused, as shown in the inset, the $\alpha = 0.1$ condition has the longest k length for $C_k \sim 1$. Hence, the highest performance in the NARMA tasks for $\alpha = 0.1$ can also be characterized by the highly accurate memory for the longest k length in C_k. From the results thus far, the $\alpha = 0.1$ condition is focused below.

Figures 10(a) and 10(b) show NARMA-10 signals under $\alpha = 0.1$ and $N_y = 120$ and 240, respectively, where gray and blue curves represent the target $d_{10}(n)$ and prediction $z^*_{10}(n)$ signals, respectively. Those two figures confirm that $z^*_{10}(n)$ reasonably reproduces $d_{10}(n)$ even under $N_y = 120$ and the difference in the computational performance between $N_y = 120$ and 240 is seemingly insignificant. On the other hand, there are three different E_m and C_k values when $N_y = 120$ and 240. In each case, the discrepancy between the results is within the error margin of approximately 20%. Hereafter, the averaged result is shown.

Figure 10(c) shows E_m plotted against m, where black squares, blue triangles, and red circles represent the values for $N_y = 120$, 240, and 360, respectively. In the entire m range, E_m decreases as N_y is increased, indicating that

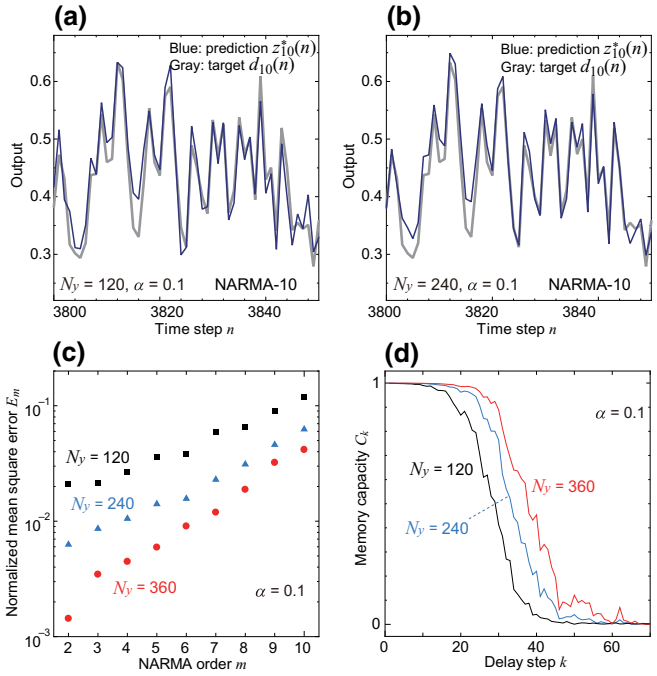

FIG. 10. (a),(b) NARMA-10 signals under $\alpha = 0.1$ and $N_y =$ (a) 120 and (b) 240, where gray and blue curves represent the target $d_{10}(n)$ and prediction $z^*_{10}(n)$ signals, respectively. (c) E_m plotted against m and (d) C_k plotted against k, when $\alpha = 0.1$, where black blue, and red colors represent the results estimated for $N_y = 120$, 240, and 360, respectively.

the effective dimensionality of the reservoir state vector increases with N_y in the NARMA-m tasks. The effect of N_y on C_k is also analyzed by $C_k - k$ plot, as shown in Fig. 10(d), where black, blue, and red curves are estimated memory curves for $N_y = 120$, 240, and 360, respectively. It is found that both the total MC and the k length for $C_k \sim 1$ increase with increasing N_y. Thus, N_y is probably the dominant parameter for high computational performance in various computational tasks.

V. DISCUSSION

A. Comparison with other RC systems

At $m = 10$ (NARMA-10), the E_{10} values for $N_y = 120$, 240, and 360 are 0.120, 0.063, and 0.042, respectively, when $\alpha = 0.1$. The minimum $E_{10} = 0.042$ is comparable to those obtained in the most advanced physical RC that utilizes optical systems with a feedback loop [39]. On the other hand, NARMA-10 prediction has not been demonstrated by an on-chip RC device, even in numerical experiments (excluding an electronic circuit simulation [7]). To further gain insight into the performance enhancement with increasing N_y, the E_{10} values are discussed through the comparison with those estimated using echo state networks (ESNs) with some network sizes in Ref. [40]: E_{10} values are 0.095, 0.051, and 0.042 for $N_y = 100$,

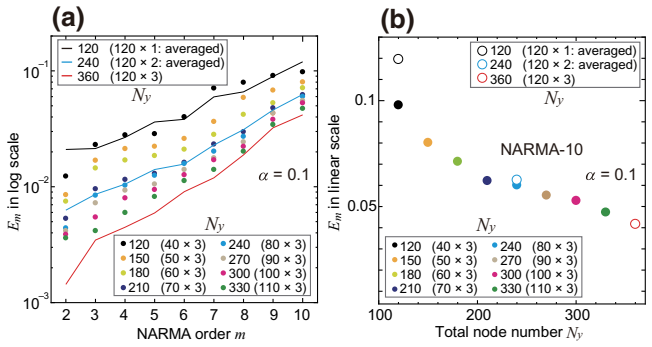

FIG. 11. (a) E_m plotted against m when $\alpha = 0.1$, where colored circles are the results for $\mathbf{y}(n)$ having $y_{1,i}(n)$, $y_{2,i}(n)$, and $y_{3,i}(n)$ ($i \leq N_y/3$). The black-, blue-, and red-solid lines are the results for $\mathbf{y}(n)$ with $N_y = 120$, 240, and 360 in Fig. 10(c), respectively. (b) NMSE_{10} plotted against N_y, where colored circles correspond to those in (a) and open circles correspond to those in Fig. 10(c).

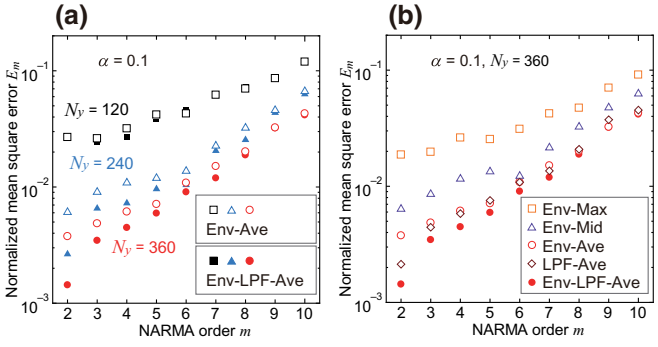

FIG. 12. (a) E_m plotted against m when Env-Ave postprocessing is used to extract $y_i(n)$, where $\alpha = 0.1$ and open-black squares, open-blue triangles, and open-red circles are the results under $N_y = 120$, 240, and 360, respectively. For comparison, the results with Env-LPF-Ave [Fig. 10(c)] are also plotted by closed-black squares, closed-blue triangles, and closed-red circles for $N_y = 120$, 240, and 360, respectively. (b) E_m plotted against m when several postprocessing methods are used to extract $y_i(n)$, where $\alpha = 0.1$, $N_y = 360$, open-orange squares, open-purple triangles, open-red circles, and open-brown diamonds are the results with Env-Max, Env-Mid, Env-Ave, and LPF-Ave, respectively, and closed-red circles are the results with Env-LPF-Ave [Fig. 10(c)].

150, and 200, respectively. Although our E_{10} value at similar N_y values is a little larger than that for the ESN, E_{10} tends to saturate with increasing N_y in both cases. Thus, it is a general property that the enhancement rate decreases with increasing N_y. Note that here we do not consider the difference in the input data between ESNs and our RC system: the former and latter use uniformly distributed and 6-valued random input data, respectively.

B. Computational performance with various output node numbers

For further study, the vectors $\mathbf{y}(n)$ with various $N_y (= 120 - 330)$ values are constructed by collecting $y_{j,i}(n)$ with $j (= 1, 2, 3)$ and $i (= 1, 2, \ldots N_y/3)$. Figure 11(a) shows E_m plotted against m, where colored circles are the results for $\mathbf{y}(n)$ having $y_{1,i}(n)$, $y_{2,i}(n)$, and $y_{3,i}(n)$ ($i \leq N_y/3$). The results for $\mathbf{y}(n)$ with $N_y = 120$, 240, and 360 in Fig. 10(c) are also plotted by black-, blue-, and red-solid lines, respectively. With the increase of N_y, E_m decreases at all the m values. At $N_y = 120$ and 240, the circles and lines are remarkably similar to each other regardless of the selection rule, indicating that each $y_{j,i}(n)$ plays a comparable role in computing. Figure 11(b) shows E_{10} plotted against N_y, where colored circles correspond to those in Fig. 11(a) and open circles correspond to those in Fig. 10(c). As N_y is increased, E_{10} continuously decreases and its rate of change seemingly decreases. The features in Figs. 11(a) and 11(b) clearly reveal that every $y_{j,i}(n)$ contributes to the computational performance, since N_y simply determines E_m. Therefore, it can be concluded that the performance enhancement in our system originates from the increase in the dimensionality in $\mathbf{y}(n)$. In addition, our analysis demonstrates that DFT spectra are helpful for the analysis of dimensionality of the reservoir state vector in physical RC, although they cannot fully characterize time-series output signals including transient temporal responses to a randomly changing input signal.

Considering that our method employs the spin dynamics under the different conditions in parallel, the whole system can be interpreted as a multireservoir system such as the grouped echo-state networks [41,42]. The explanation in terms of system architecture is our future interest.

C. RC systems with several postprocessing methods

As described in Sec. IV D, the high computational performances are achieved using $y_i(n)$ after the postprocessing method Env-LPF-Ave that was developed in our previous paper [21] addressing basic benchmark tasks with binary inputs. In physical RC systems, the postprocessing of reservoir output signals is needed unless virtual node states [7] are used, and it can have influence on computational performance. Toward the establishment of the system design for high computational performance, the NARMA-m tasks are also performed using $y_i(n)$ extracted through several postprocessing methods, Env-Ave, LPF-Ave, Env-Mid, and Env-Max, as introduced in Sec. IV B. In what follows, we analyze how E_m values change from those in Fig. 10(c).

Figure 12(a) shows E_m plotted against m, where open-black squares, open-blue triangles, and open-red circles are the results with Env-Ave under $N_y = 120$, 240, and 360, respectively, and closed-black squares, closed-blue

triangles, and closed-red circles are the results with Env-LPF-Ave [Fig. 10(c)]. When $N_y = 120$, the E_m values are unchanged. On the other hand, $N_y = 240$ and 360, the E_m values are almost unchanged in larger m range, whereas those become a little higher in smaller m range, seemingly in $m \leq 6$. Nonetheless, the difference is on the order of 10^{-3}.

Figure 12(b) shows E_m plotted against m when $N_y = 360$, where open-orange squares, open-purple triangles, open-red circles, and open-brown diamonds are the results with Env-Max, Env-Mid, Env-Ave, and LPF-Ave, respectively, and closed-red circles are the results with Env-LPF-Ave [Fig. 10(c)]. The E_m values with Env-Ave are almost identical with those with LPF-Ave. This is reasonable because the $y_i(n)$ curves (the steplike curves) in Fig. 7(b) are very similar between Env-Ave and LPF-Ave. On the other hand, in the entire m range, the E_m values with Env-Max and Env-Mid are significantly larger than those with Env-LPF-Ave. The results in Figs. 12(a) and 12(b) reveal that averaging over the time-step duration X is effective and successive Env and LPF is preferable to extract useful reservoir node states for the NARMA-m tasks. In our system, this knowledge can be a basic guideline for the postprocessing method in other tasks.

VI. CONCLUSION

We numerically study the performance enhancement in RC utilizing the single spin-wave-based reservoir device with the 1-input exciter and 120-output detectors. The input signal is the 1-GHz cosine wave whose amplitude is modulated following a series random 6 values and the output signals at the detectors are found to have nonlinear phenomena changing with the spatial position, external magnetic field, and damping constant at the sides and bottom of the magnetic garnet film. The reservoir output vectors with various dimensions, including higher than 120 dimension, are constructed by collecting the output signals under the three different magnetic fields. With the optimum damping constant, the computational performance in NARMA tasks is enhanced with the increase in the dimension of the reservoir output vector. The result is evidence that our method efficiently increases the dimensionality effective for NARMA tasks.

We demonstrate that the single spin-wave-based device with the $\mu_0 H^{\mathrm{ex}}$-controlled different physical conditions can generate diverse output signals that are essential for higher computational performance. This strategy is also very useful for other physical RC, along with the virtual node method, to tackle nonlinear time-series prediction tasks. In particular, since it is quite strict to implement a vast number of real output nodes on a chip device, performance enhancement through various configuration settings is a practical approach.

TABLE I. Correspondence table.

Output node number i	Detector position
1 – 12	R-C1 – 12
13 – 24	L-C1 – 12
25 – 36	R-A1 – 12
37 – 48	L-A1 – 12
49 – 60	R-E1 – 12
61 – 72	L-E1 – 12
73 – 84	R-B1 – 12
85 – 96	L-B1 – 12
97 – 108	R-D1 – 12
109 – 120	L-D1 – 12

ACKNOWLEDGMENTS

This work is supported in part by the New Energy and Industrial Technology Development Organization (NEDO) under Project JPNP16007.

APPENDIX A: CORRESPONDENCE BETWEEN THE OUTPUT NODE NUMBER AND DETECTOR POSITION

The output node number $i(= 1 - 120)$ is used in $x_{j,i}(t)$ and $y_{j,i}(n)$. The detector position indexes are shown in Fig. 2. Table I shows the correspondence between them.

APPENDIX B: OUTPUT SIGNALS IN RESPONSE TO THE INPUT CARRIER WAVE

In Sec. III B, the waveform and DFT spectrum at R-C1 detector are observed while the carrier wave is fed into the exciter under $\mu_0 H^{\mathrm{ex}} = 5$ mT, and their changes with α are explained. For the discussion in Sec. V B, it is necessary that the main feature determined by α does not significantly change with detector and $\mu_0 H^{\mathrm{ex}}$. In this Appendix, we show waveforms and DFT spectra at R-C6 and R-C12 detectors under $\mu_0 H^{\mathrm{ex}} = 5$ mT and $\alpha = 0.001$, 0.01, and 0.1.

Figures 13(a) and 13(b) show waveforms at R-C6 and R-C12 detectors, respectively, under $\alpha = 0.001$, and Figs. 13(c) and 13(d) show DFT spectra calculated from (a) and (b), respectively. As in the case of R-C1 detector

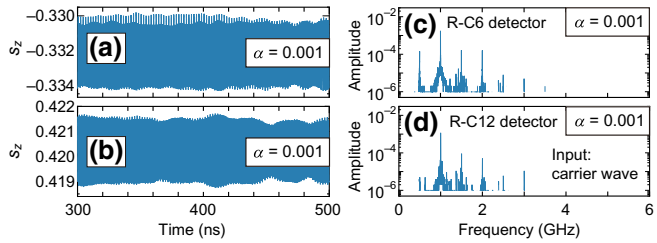

FIG. 13. (a),(b) Waveforms and (c),(d) DFT spectra of the carrier wave under $\mu_0 H^{\mathrm{ex}} = 5$ mT and $\alpha = 0.001$.

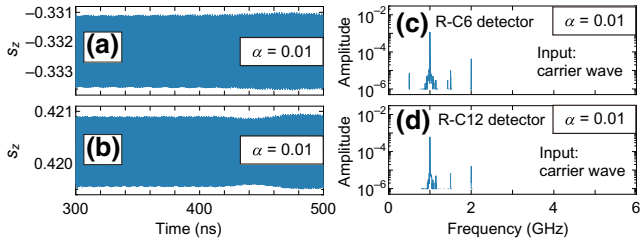

FIG. 14. (a)(b) Waveforms and (c),(d) DFT spectra of the carrier wave under $\mu_0 H^{\rm ex} = 5$ mT and $\alpha = 0.01$.

[Fig. 5(a)], there are fluctuations in the amplitude of the waveforms. In addition, the mean amplitude and DFT spectrum are little changed in these three detectors, which comes from extremely low damping of spin waves.

Figures 14(a) and 14(b) show waveforms at R-C6 and R-C12 detectors, respectively, under $\alpha = 0.01$, and Figs. 14(c) and 14(d) show DFT spectra calculated from (a) and (b), respectively. In the DFT spectra, there are interhamonic components as in the case of R-C1 detector [Fig. 5(f)], whereas there is no peak at 3 GHz. Observations of the DFT spectra at all the detectors found a trend that a peak component at 3 GHz appears in the DFT spectrum when the column number is less than 6 (the distance from the exciter is shorter than approximately 1 μm). This fact indicates slight damping of spin waves to suppress nonlinear phenomena as the traveling distance becomes longer.

Figures 15(a) and 15(b) show waveforms at R-C6 and R-C12 detectors, respectively, under $\alpha = 0.1$, and Figs. 15(c) and 15(d) show DFT spectra calculated from (a) and (b), respectively. There is no fluctuation in the amplitude of the waveforms, which is reflected in no interharmonic components in the DFT spectra. On the other hand, there is a trend that a peak component at 2 GHz appears in the DFT spectrum when the column number is less than 6. Under $\alpha = 1$, similar features to those under $\alpha = 0.1$ are obtained, whereas the amplitude of the waveform decays faster as the distance from the exciter increases.

In the same manner, we observe waveforms and DFT spectra at all the detectors under $\mu_0 H^{\rm ex}$ (= 2, 5, and 8 mT), and confirm that α mostly determines the main feature,

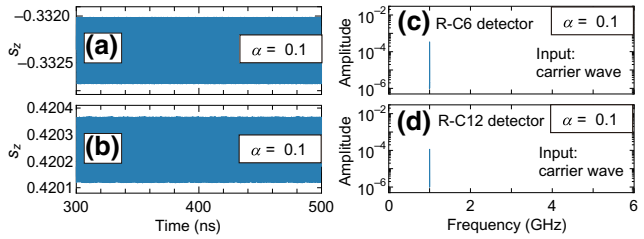

FIG. 15. (a),(b) Waveforms and (c),(d) DFT spectra of the carrier wave under $\mu_0 H^{\rm ex} = 5$ mT and $\alpha = 0.1$.

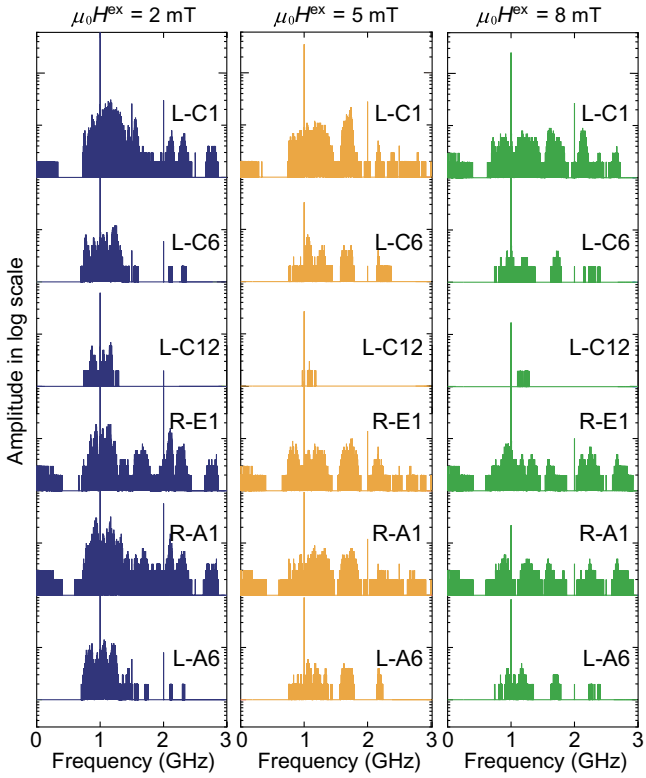

FIG. 16. DFT spectra of the output signals at several detectors under $\alpha = 0.1$ while the 6-valued random input signal is fed into the exciter, where blue, orange, and green lines represent the results under $\mu_0 H^{\rm ex} = 2$, 5, and 8 mT, respectively.

e.g., appearance of interharmonic components only when $\alpha = 0.001$ and 0.01.

APPENDIX C: DFT SPECTRA OF THE OUTPUT SIGNALS IN RESPONSE TO THE 6-VALUED RANDOM INPUT SIGNAL

In Sec. IV A, the waveforms and DFT spectra at some detectors are observed while the 6-valued random input signal is fed into the exciter under $\mu_0 H^{\rm ex} = 2$, 5, and 8 mT. To further support diversity in the output waveforms, here we show DFT spectra at other detectors under $\alpha = 0.1$ and $\mu_0 H^{\rm ex} = 2$, 5, and 8 mT.

Figure 16 shows DFT spectra of the output signals at several detectors, which characterize a unique output signal for each detector.

[1] H. Jaeger, The "echo state" approach to analysing and training recurrent neural networks-with an erratum note, Bonn, Germany: German National Research Center for Information Technology GMD Technical Report **148**, 13 (2001).

[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] G. Tanaka, T. Yamane, J. B. Héroux, R. Nakane, N. Kanazawa, S. Takeda, H. Numata, D. Nakano, and A. Hirose, Recent advances in physical reservoir computing: A review, Neural Netw. **115**, 100 (2019).
[4] K. Nakajima, Recent advances in physical reservoir computing: A review, Jpn. J. Appl. Phys. **59**, 060501 (2020).
[5] N. Abbas, Y. Zhang, A. Taherkordi, and T. Skeie, Mobile edge computing: A survey, IEEE Internet Things J. **5**, 450 (2017).
[6] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, Edge computing: Vision and challenges, IEEE Internet Things J. **3**, 637 (2016).
[7] 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).
[8] J. Torrejon, M. Riou, F. A. Araujo, S. Tsunegi, G. Khalsa, D. Querlioz, P. Bortolotti, V. Cros, K. Yakushiji, A. Fukushima, Hitoshi Kubota, Shinji Yuasa, Mark D. Stiles, and Julie Grollier, Neuromorphic computing with nanoscale spintronic oscillators, Nature **547**, 428 (2017).
[9] R. Nakane, G. Tanaka, and A. Hirose, Reservoir computing with spin waves excited in a garnet film, IEEE Access **6**, 4462 (2018).
[10] J. Moon, W. Ma, J. H. Shin, F. Cai, C. Du, S. H. Lee, and W. D. Lu, Temporal data classification and forecasting using a memristor-based reservoir computing system, Nat. Electron. **2**, 480 (2019).
[11] K. Toprasertpong, E. Nako, Z. Wang, R. Nakane, M. Takenaka, and S. Takagi, Reservoir computing on a silicon platform with a ferroelectric field-effect transistor, Commun. Eng. **1**, 21 (2022).
[12] T. Kanao, H. Suto, K. Mizushima, H. Goto, T. Tanamoto, and T. Nagasawa, Reservoir Computing on Spin-Torque Oscillator Array, Phys. Rev. Appl. **12**, 024052 (2019).
[13] G. Tanaka and R. Nakane, Simulation platform for pattern recognition based on reservoir computing with memristor networks, Sci. Rep. **12**, 9868 (2022).
[14] D. Pinna, G. Bourianoff, and K. Everschor-Sitte, Reservoir Computing with Random Skyrmion Textures, Phys. Rev. Appl. **14**, 054020 (2020).
[15] M.-K. Lee and M. Mochizuki, Reservoir Computing with Spin Waves in a Skyrmion Crystal, Phys. Rev. Appl. **18**, 014074 (2022).
[16] A. G. Gurevich and G. A. Melkov, *Magnetization Oscillations and Waves* (CRC press, London, 2020).
[17] D. D. Stancil and A. Prabhakar, *Spin Waves* (Springer, New York, 2009).
[18] A. Mahmoud, F. Ciubotaru, F. Vanderveken, A. V. Chumak, S. Hamdioui, C. Adelmann, and S. Cotofana, Introduction to spin wave computing, J. Appl. Phys. **128**, 161101 (2020).
[19] Á. Papp, W. Porod, and G. Csaba, Nanoscale neural network using non-linear spin-wave interference, Nat. Commun. **12**, 1 (2021).
[20] T. Ichimura, R. Nakane, G. Tanaka, and A. Hirose, A numerical exploration of signal detector arrangement in a spin-wave reservoir computing device, IEEE Access **9**, 72637 (2021).
[21] R. Nakane, A. Hirose, and G. Tanaka, Spin waves propagating through a stripe magnetic domain structure and their applications to reservoir computing, Phys. Rev. Res. **3**, 033243 (2021).
[22] S. Watt, M. Kostylev, A. B. Ustinov, and B. A. Kalinikos, Implementing a Magnonic Reservoir Computer Model Based on Time-Delay Multiplexing, Phys. Rev. Appl. **15**, 064060 (2021).
[23] A. Vansteenkiste, J. Leliaert, M. Dvornik, M. Helsen, F. Garcia-Sanchez, and B. Van Waeyenberge, The design and verification of MuMax3, AIP Adv. **4**, 107133 (2014).
[24] T. L. Gilbert, Applications of the phenomenological theory of ferromagnetism (Ph. D. thesis), Illinois Institute of Technology, Chicago, 1956, partly reprinted in IEEE, IEEE Trans. Magn. **40**, 3443 (2004).
[25] C. Kittel, Physical theory of ferromagnetic domains, Rev. Mod. Phys. **21**, 541 (1949).
[26] A. Hubert and R. Schäfer, *Magnetic Domains: The Analysis of Magnetic Microstructures* (Springer Science & Business Media, Berlin, 2008).
[27] E. E. Anderson, Molecular field model and the magnetization of YIG, Phys. Rev. **134**, A1581 (1964).
[28] S. A. Manuilov, S. Khartsev, and A. M. Grishin, Pulsed laser deposited $Y_3Fe_5O_{12}$ films: Nature of magnetic anisotropy I, J. Appl. Phys. **106**, 123917 (2009).
[29] S. Klingler, A. V. Chumak, T. Mewes, B. Khodadadi, C. Mewes, C. Dubs, O. Surzhenko, B. Hillebrands, and A. Conca, Measurements of the exchange stiffness of YIG films using broadband ferromagnetic resonance techniques, J.Phys. D: Appl. Phys. **48**, 015001 (2014).
[30] A. Bobeck, Properties and device applications of magnetic domains in orthoferrites, Bell Syst. Tech. J. **46**, 1901 (1967).
[31] C. Kooy and U. Enz, Experimental and theoretical study of the domain configuration in thin layers of $BaFe_{12}O_{19}$, Phil. Res. Rept. **15**, 7 (1960).
[32] A. Bobeck, E. Spencer, L. Van Uitert, S. Abrahams, R. Barns, W. Grodkiewicz, R. Sherwood, P. Schmidt, D. Smith, and E. Walters, Uniaxial magnetic garnets for domain wall"bubble" devices, Appl. Phys. Lett. **17**, 131 (1970).
[33] W. Xia, Y. Chun, S. Aizawa, K. Yanagisawa, K. M. Krishnan, D. Shindo, and A. Tonomura, Investigation of magnetic structure and magnetization process of yttrium iron garnet film by Lorentz microscopy and electron holography, J. Appl. Phys. **108**, 123919 (2010).
[34] P. Ghising, Z. Hossain, and R. Budhani, Stripe magnetic domains in $CeY_2Fe_5O_{12}$ (Ce: YIG) epitaxial films, Appl. Phys. Lett. **110**, 012406 (2017).
[35] I. Zavislyak, V. Sohatsky, M. Popov, and G. Srinivasan, Electric-field-induced reorientation and flip in domain magnetization and light diffraction in an yttrium-iron-garnet/lead-zirconate-titanate bilayer, Phys. Rev. B **87**, 134417 (2013).
[36] M. Lukoševičius and H. Jaeger, Reservoir computing approaches to recurrent neural network training, Comput. Sci. Rev. **3**, 127 (2009).

[37] 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).

[38] H. Jaeger, Short term memory in echo state networks. gmd-report 152, GMD-German National Research Institute for Computer Science GMD Report 152 (2002).

[39] Y. K. Chembo, Machine learning based on reservoir computing with time-delayed optoelectronic and photonic systems, Chaos **30**, 013111 (2020).

[40] A. Rodan and P. Tino, Minimum complexity echo state network, IEEE Trans. Neural Netw. **22**, 131 (2010).

[41] C. Gallicchio, A. Micheli, and L. Pedrelli, Deep reservoir computing: A critical experimental analysis, Neurocomputing **268**, 87 (2017).

[42] Z. Li and G. Tanaka, Multi-reservoir echo state networks with sequence resampling for nonlinear time-series prediction, Neurocomputing **467**, 115 (2022).