Reservoir Computing Using a Spin-Wave Delay-Line Active-Ring Resonator Based on Yttrium-Iron-Garnet Film

Stuart Watt* and Mikhail Kostylev
School of Physics and Astrophysics, University of Western Australia, Crawley, W.A. 6009, Australia

(Received 11 December 2019; revised manuscript received 24 February 2020; accepted 2 March 2020; published 23 March 2020)

We demonstrate the use of propagating spin waves for implementing a reservoir-computing architecture. Our concept utilizes an active-ring resonator comprising a magnetic thin-film delay line with an integrated feedback loop. These systems exhibit strong nonlinearity and delayed response, two important properties required for an effective reservoir-computing implementation. In a simple design, we exploit the electric control of feedback gain to inject input data into the active-ring resonator and use a microwave diode to read out the amplitude of the spin waves circulating in the ring. We employ two baseline tasks, namely the short-term memory and parity-check tasks, to evaluate the suitability of this architecture for processing time-series data.

DOI: 10.1103/PhysRevApplied.13.034057

I. INTRODUCTION

The reservoir computer is a computational framework based on the recursive neural network [1,2] and as such is suitable for modeling complex dynamical systems. In a reservoir-computing (RC) architecture, a time-dependent data stream is fed into a dynamical system called the "reservoir," the internal state of which is governed by nonlinear activation functions and recursive feedback loops. The function of the "reservoir" is to map the input data to a higher-dimensional state space in which complex patterns present in the input become linearly separable and easily determinable using simpler models. Unlike traditional neural-network frameworks, training a RC model does not involve modifying the internal properties of the reservoir and only the readout layer is trained using simple linear-regression techniques. Due to this relative ease of training and its remarkable computational power, RC has received great interest in the past decade.

In order to be suitable as a RC architecture, the dynamical system must possess two important properties. The first is a nonlinear mapping of the input data to a higher-dimensional state space. The second is a "fading" memory, where the system state depends not only on the current input but also on past inputs, with the influence of these past inputs "fading" in time. Since RC architectures require only simple training of the readout layer, they are not restricted to software implementations and in recent years it has been demonstrated that a wide variety of physical dynamical systems can be exploited for RC implementation with the goal of increasing computation speed and reducing power usage. For a review of RC implementations, see Ref. [3].

Among these, spintronic-based architectures are promising candidates for practical RC applications due to their low power usage, their strong nonlinearity arising from magnetization dynamics, and their ability to be scaled down to small sizes. In particular, many studies have already been conducted on spin-torque nano-oscillators, with experimental results [4–8] demonstrating promising performance as RC implementations. Further simulations [9,10] indicate the potential for improved performance by using arrays of coupled oscillators. Other proposed magnetism-based RC implementations are systems utilizing magnetic skyrmion memristors [11], magnetic skyrmion fabrics [12,13], dipole-coupled nanomagnets [14,15], and spin-wave interference in garnet films [16]. In the present work, we propose an alternative spintronic RC architecture also based on spin-wave propagation. The physical setup, shown in Fig. 1(a), is a magnetic-film active-ring resonator consisting of a spin-wave delay line with an amplified feedback loop. This system utilizes the delay and nonlinear behavior of traveling spin waves in magnetic films and has been used in many studies to explore nonlinear behavior, such as soliton formation, modulational instability, and chaotic behavior [17]. First, we describe the spin-wave active-ring resonator system and demonstrate how it can be used for RC implementation. We then briefly cover the principles of RC followed by experimental evaluation of the system's performance as a reservoir computer using benchmark tasks.

*stuart.watt@research.uwa.edu.au

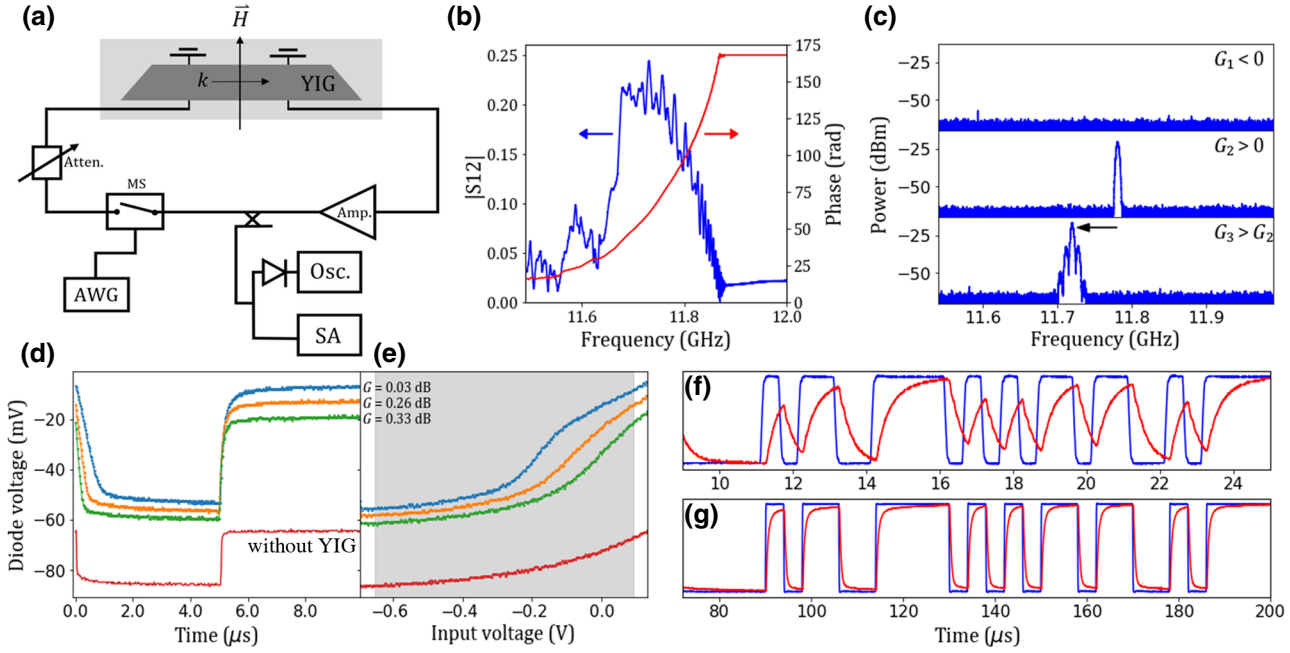

FIG. 1. (a) A schematic diagram of the spin-wave delay-line active-ring resonator system. The components are described in the text. (b) The amplitude and phase of the transmission (S12 parameter) characteristic for the spin-wave delay line, taken with a microwave vector network analyzer. (c) The frequency spectrum of the active ring for various gains. (d) The resonator state response to sudden change (step function) in the microwave-switch input bias voltage for various gains. (e) The resonator state response to the microwave-switch input bias voltage for various gains. The gray region indicates the computing range of the input bias voltage. (f) The system input (blue) and the diode voltage output (red) in response to a binary input sequence with an input time interval of 0.5 μs. (g) The same as for (f) but with an input time interval of 4 μs.

II. SPIN-WAVE DELAY-LINE ACTIVE-RING RESONATOR

A spin wave is a fundamental type of magnetization dynamics in magnetic materials, representing collective precession of elementary magnetic moments (spins of localized electrons), whereby the phase of precession varies harmonically in one direction in space. Typical precession frequencies lie in the microwave range and depend on a static magnetic field applied to the material and the material's magnetic parameters (its saturation magnetization, magnetic anisotropies, and gyromagnetic ratio). The frequency also depends on the sample's geometry due to the collective character of precession resulting from the coupling of spins at neighboring crystal-lattice sites by exchange and magnetic dipole-dipole interactions.

The dipole-dipole interactions are dominant for the long-wavelength spin waves observed in this work. In the geometry of a thin ferro- or ferrimagnetic film, dipole-dominated spin waves propagating in the film plane are characterized by large group velocities, a dispersion law that strongly depends on the direction of the film magnetization with respect to the film plane, and the direction of the spin-wave propagation with respect to the static magnetic field projection onto the film plane. In this work, we employ the configuration of the magnetostatic surface spin wave (MSSW), which is characterized by the static magnetic field lying in the film plane and spin waves propagating perpendicular to the field direction.

The MSSW angular frequency, ω, is determined by the MSSW dispersion relation [18]:

$$\omega(H,k) = \gamma\sqrt{H(H+4\pi M_s) + \frac{(4\pi M_s)^2}{4}(1-e^{-2kL})}. \tag{1}$$

Here, H is the magnetic field and k is the spin-wave wave number. The constants γ, $4\pi M_s$, and L are the gyromagnetic ratio, the saturation magnetization, and the film thickness. For the material employed in this work, they have values of $2\pi \cdot (2.8\ \mathrm{MHz/Oe})$, 1750 Oe, and 43 μm, respectively.

The choice of this configuration is due to a unique combination of properties of the MSSW. First, this wave is efficiently excited by microwave currents in stripline transducers. Second, the excitation is almost unidirectional—the amplitude of the wave excited in one direction perpendicular to the transducer's longitudinal axis [Fig. 1(a)] is much larger than in the opposite direction. This minimizes transduction losses for the input and output transducers of spin waves and thus the total losses inserted by the spin-wave device (the "spin-wave delay line") are

minimized. Third, the free-propagation path for MSSW is relatively large. The path length is determined by the magnetic loss parameter for the material and the spin-wave group velocity. Epitaxially grown films of yttrium iron garnet (YIG) possess remarkably low magnetic losses. For the film employed in this work, the ferromagnetic-resonance linewidth is 0.4 Oe, which translates into a free path length on the order of about 1 cm.

Importantly, due to the small magnetic losses, YIG films are characterized by very low thresholds for nonlinear wave processes, such as modulation instability and envelope soliton formation. The four-wave processes of modulation instability and formation of bright solitons are prohibited for MSSW, because the Lighthill criterion is not satisfied for this spin-wave configuration [19]. Only dark solitons can be formed in a lossless medium if this criterion is not satisfied [20].

The spin-wave active-ring resonator represents a loop consisting of a spin-wave delay line based on a YIG film and a microwave amplifier that provides positive feedback. The spin-wave delay line [highlighted by the light-gray background in Fig. 1(a)] is a device that consists of a strip of YIG film (the dark-gray trapezoid) and two microstrip transducers ("spin-wave antennas"). The antennas are shown as straight lines with grounding signs at their ends. The left-hand antenna is an input transducer, which launches a spin wave into the film. It converts the power of an electromagnetic wave propagating in the stripline into MSSW. The output antenna picks up the signal carried by the MSSW and converts it back into an electromagnetic wave. In the present experiment, the MSSW delay line is made up of a 4-mm-wide and 22-mm-long YIG film, with a thickness of 43 μm. The film sits on top of the two short-circuited 0.5-mm-wide microstrip antennas, with a 12-mm separation between the antennas. A magnetic field of 3250 Oe is applied parallel to the antennas and perpendicular to the spin-wave propagation direction, which allows the excitation of the MSSW.

Importantly, the spin-wave delay line represents a broadband device with a continuous frequency spectrum within the transmission band. The frequency width of the transmission band is determined by the filtering properties of the spin-wave antennas and the loss inserted in the maximum of transmission is determined by the efficiency of the antennas and the magnetic damping constant for the film. The transmission characteristic for the present delay line, taken with a microwave vector network analyzer, is shown in Fig. 1(b). One sees that the transmission bandwidth is approximately 150 MHz and the minimum insertion losses are -12.3 dB. This narrow bandwidth is due to a larger width of our antennas (0.5 mm). Narrower antennas would have produced a wider transmission band for the device.

The broadband passive device is converted into an auto-oscillator generating at a single frequency first after a positive delayed feedback has been added to the system [21–23]. In the active-ring configuration, the microwave signal from the output antenna is amplified and then injected back into the input antenna. The spin-wave delay line introduces a time delay of the signal passing from the input to the output antenna, resulting in a phase shift of the microwave signal across the YIG strip. If the feedback gain, G, is sufficient to compensate the losses occurring in the YIG film, then a signal can circulate in the ring, thus forming a ring resonator with a discrete set of frequencies. Assuming that the phase shift resulting from the electrical components is negligible, the resonant-frequency eigenmodes satisfy the condition $k_{\mathrm{res}}d = 2\pi n$, where k_{res} is the spin-wave wave number and d is the antenna separation. Which eigenmodes are actually excited depends on the frequency characteristic of the delay line—the transduction frequency band of the spin-wave antennas and the MSSW dispersion relation.

Shown graphically in Fig. 1(c) is the process of ring-eigenmode excitation, measured using a spectrum analyzer (denoted "SA"). At low feedback gain, delay-line losses are dominant and the spin waves do not have enough power to propagate between the two antennas. Above a certain threshold (the "auto-oscillation threshold," defined as $G = 0$), where the gain is sufficient to compensate the losses, thermally excited magnons with wave numbers corresponding to the eigenmode with the lowest loss will be resonantly amplified in the ring [G2 in Fig. 1(c)]. The gain threshold for the onset of auto-oscillations is determined by the linear damping coefficient for MSSW and the efficiency of transduction by the spin-wave antennas (also the transmission coefficient for the microwave switch in our ring implementation). As the gain increases, the next eigenmode with the lowest loss is excited and so on (G3). Furthermore, the magnetization precession angle increases with the spin wave power, effectively reducing $4\pi M_s$ and shifting the dispersion curve downward [see Eq. (1)], resulting in a shift of the eigenmode frequency [see Fig. 1(c)]. The onset of more complex behavior (first, the generation of multiple spectral lines and, ultimately, chaotic behavior) occurs at amplification gains much higher than the auto-oscillation threshold [24,25].

An important aspect of the active-ring resonator functionality is how the amplitude of auto-oscillations stabilizes above the auto-oscillation threshold to produce a steady state of single-frequency generation (fixed point). This question is important because the time of transition to the steady state determines the period for which the "fading" memory of our RC device retains information. The nonlinearity of MSSW damping plays the central role in the transition to the steady state. That damping of MSSW in YIG films is highly nonlinear has been found in Ref. [26]. Due to this damping nonlinearity, the system is able to saturate to the steady state for a relatively broad range of smaller amplification gains. If the nonlinear damping were absent, one would expect unstable (chaotic) behavior of

the signal generated in the ring at smaller gains [27]. For a review of the nonlinear behavior in magnetic film feedback rings, see Ref. [17].

As mentioned previously, these systems exhibit a host of nonlinear behavior, which we propose can be exploited for physical RC. In particular, in this work we exploit the nonlinearity of spin-wave damping at high microwave power mentioned previously, for which the active-ring resonator can still be operated in the single-mode regime [G2 in Fig. 1(c)]. The parameters that control the dynamics of this system are the ring gain and the magnetic field strength. In this work, we employ a static magnetic field and use a variable ring gain as the method of data input. This is done by using a PIN-diode microwave switch [denoted as "MS" in Fig. 1(a)], which varies the ring gain by attenuating the microwave signal according to the dc bias voltage supplied by an arbitrary waveform generator ("AWG"). In addition, an adjustable attenuator and a directional coupler are inserted into the feedback loop. The attenuator is used to set the base amplification gain level in the ring and the directional coupler extracts a portion of microwave power from the ring that enables observation of the time evolution of the signal circulating in the ring with a microwave diode and an oscilloscope ("Osc."). In the frequency domain, this signal is observed with a microwave spectrum analyzer.

Figure 1(e) shows the ring state in response to the microwave-switch input voltage at various levels of the base ring gain G. The gray-shaded region shows the input-voltage range supplied by the AWG used in this work. The resonator state depends nonlinearly on the input voltage, satisfying the first condition required for RC implementation. Furthermore, the nonlinearity depends on the value of G, with the more nonlinear behavior occurring closer to the auto-oscillation threshold ($G = 0$). It is found that the nonlinearity has contributions from both the YIG film and also the microwave-switch component. The bottom curve [labeled "without YIG" in Fig. 1(e)] shows the response of the switch acting alone on a constant microwave signal with the same frequency as the ring eigenmode, showing that it too results in a nonlinear behavior of the diode voltage.

The second property required by a RC implementation is the "fading" memory property. In this system, the "fading" memory is achieved through the delayed response of the system to a given input. The resonator-state response to a step function of the input voltage is shown in Fig. 1(d). The time for the resonator to reach a steady state, where the circulating ring signal has a constant amplitude, depends on the value of G, with longer relaxation times occurring at gains closer to the auto-oscillation threshold. If the amplification gain only slightly exceeds this threshold, the signal may circulate the ring many times before the amplitude reaches a level at which the nonlinearity of spin-wave damping becomes significant and ultimately saturates the system.

The round-trip time of the signal in the ring, t_d, is determined by the delay in the YIG strip (assuming that the electrical feedback loop is practically delayless) and is approximately 50 ns. The total time, t_0, of stabilizing single-mode generation (i.e., the time of onset of the steady auto-oscillation) in the ring is expected to amount to multiple round trips of the signal and thus, t_0 can be large enough in order to enable processing of time-series data with duration times for individual series elements smaller than t_0. This time of approach, t_0, to steady amplitude of the eigenmode is governed by the coefficients of the linear and nonlinear damping for spin waves [26], the spin-wave delay time for the film, and the amplification gain. For the system controlled with a microwave switch as employed in the present work, it also depends on the nonlinearity of the switch response. Thus, control of the amplification gain in the ring allows partial tuning of both the nonlinearity and the response time.

III. RESERVOIR-COMPUTING PERFORMANCE

The traditional RC model has three components—the input, the reservoir, and the readout [28]. The system state at a discrete time step T is a vector containing the values of each node in the reservoir, determined by the update equation

$$\vec{x}(t) = f\,[\mathbf{W}^{\text{in}}\vec{u}(t) + \mathbf{W}\vec{x}(t-1)]. \tag{2}$$

The matrices $\mathbf{W}^{\text{in}}$ and $\mathbf{W}$ determine the weights for the input layer and internal connections, respectively. $\vec{u}(t)$ is the vector of input values and f is the nonlinear function that governs the response of the system. Here f and $\mathbf{W}$ are determined by the physical parameters of the RC implementation and the output of the system is obtained as a weighted sum of the reservoir states:

$$\vec{y}_{\text{out}}(t) = \mathbf{W}^{\text{out}}\vec{x}(t). \tag{3}$$

During the training phase, the weights in $\mathbf{W}^{\text{out}}$ are adjusted to reduce the mean-squared error between the RC output and the target output. Since, in RC, only the output weight matrix $\mathbf{W}^{\text{out}}$ is trained, all of the input data can be passed through the reservoir at once and then the training and evaluation are performed offline later. Defining $\mathbf{Y}$ and $\mathbf{X}$ as matrices containing the target and reservoir states for all time steps, the optimal $\mathbf{W}^{\text{out}}$ is obtained by taking the product $\mathbf{YX}^{-1}$.

In order to process continuous time signals, one must first sample the input and then each discrete value is exposed to the system for a time interval of θ^{int}. This time interval can be chosen to control the level of "fading" memory in the system. Examples of the active-ring resonator-state output to the input binary sequence are shown in Figs. 1(f) and 1(g) for input intervals of 0.5 μs

and 4 μs, respectively. For larger values of θ^{int}, the system has sufficient time to reach a steady state that is less dependent on the past inputs, reducing the "fading" memory. Unlike traditional RC, where the reservoir is comprised of many nodes, the proposed system is effectively a single node with a single input and a single recursive feedback loop. In order to increase the dimensionality of the reservoir, the resonator output is sampled into N "virtual" nodes [29] for each input time interval with separation θ^{int}/N, creating the reservoir state vector $\vec{x}(t)$. As is typical in RC applications, the input value for a given time step and an additional bias term (equal to -0.2 for all time steps) is also added to the state vector and used for training and prediction. Thus there is a total of $N+2$ values used to compute the system output for each time step.

We employ two popular metrics to evaluate the performance of the active-ring resonator system as a reservoir computer. These are the short-term memory (STM) [30] and parity-check (PC) [9,31] tasks. For both, the system state at a given time T is measured in response to a random binary input $u(T) \in [0,1]$. A sequence of 2100 time steps is input to the system, where the first 100 steps are used to "wash out" any transient behavior in the system that may depend on the arbitrary initial state. The next 1000 time steps are used for training $\mathbf{W}^{\text{out}}$ and the final 1000 steps are used for evaluation. In order to input the sequence to the active-ring resonator, the inputs [0, 1] are converted to corresponding input bias voltages [-650 mV, 90 mV], which correspond to the boundaries of the input range shown in Fig. 1(e). For each step in the input sequence, the resonator output is split into $N = 20$ virtual nodes.

The STM task provides a measure of the system's ability to predict previous inputs given the current output. In this task, the target output for each time step is simply the input at some previous time in the past, $y_{\text{STM}}(T,\tau) = u(T-\tau)$, where τ is the time delay. Following the standard training procedure, the output weight matrix $\mathbf{W}^{\text{out}}$ is determined using the training set. The system then makes predictions on the test set and the square of the correlation coefficient between the ideal targets and the model predictions is determined as [9]

$$r_{\text{STM}}(\tau)^2 = \frac{\text{Cov}[y_{\text{out}}(T), y_{\text{STM}}(T,\tau)]^2}{\text{Cov}[y_{\text{out}}(T)] \times \text{Cov}[y_{\text{STM}}(T,\tau)]}. \quad (4)$$

Here, $\text{Cov}(A,B)$ is the covariance between some vectors A and B and $\text{Cov}(A) \cong \text{Cov}(A,A)$. $r_{\text{STM}}(\tau)^2$ takes values between 0 and 1, where the value 1 indicates perfect replication of the targets. The STM capacity is then calculated by taking the sum of $r_{\text{STM}}(\tau)^2$ over the range of delays:

$$C = \sum_{\tau=1}^{\tau_{\max}=20} r_{\text{STM}}(\tau)^2. \quad (5)$$

While the STM task is a good characterization of a reservoir computer's capability to store memory, it does not require a nonlinear system in order to yield high STM capacities and thus is not a good metric to evaluate the nonlinearity performance of the reservoir computer. The PC task is a nonlinearly separable task that requires both "fading" memory and nonlinearity. The target outputs used in this task are determined by taking the parity of the sum of the consecutive binary inputs up to some time delay in the past:

$$y_{\text{PC}}(T,\tau) = \text{PARITY}[u(T-\tau), u(T-\tau+1), \ldots, u(T)]. \quad (6)$$

Here, the PARITY operation returns the parity (0 for even, 1 for odd) of the sum of the parameters in the brackets and, as such, the target itself is also a sequence of binary values. It should be noted that the evaluation of the nonlinear characteristics is not restricted to the PC task; one may also choose a task based on some other arbitrary binary operation acting on past inputs. Evaluation of the performance on the PC task is carried out in the same manner as for the STM task, with the PC capacity similarly determined using Eq. (5). Since the STM and PC tasks both use binary sequences, only one input sequence is needed to compute both C_{STM} and C_{PC}.

The results of our experiments are shown in Fig. 2. From the figure, one sees that as the delay time increases, the ability of the system to correctly predict the targets reduces. This can best be described using "forgetting" curves [30], as shown in Figs. 2(a) and 2(b), which plot the square of the correlation coefficient against the delay time. The respective capacity is then calculated as the area under the curves. The plot of C_{STM} versus the input time interval is shown in Fig. 2(c). One sees that the active-ring system performs better than using electronic components alone for all θ^{int}, indicating the importance of the feedback loop for realizing the "fading" memory. The C_{STM} increases for smaller input time intervals, which can be explained intuitively since the system cannot reach a steady state and the final output depends on the previous system state. The minor improvement in C_{STM} with increasing gain is likely the result of reduced noise as the ring resonator moves away from the auto-oscillation threshold (G=0). At the lowest input time interval, the C_{STM} reaches its maximum value of 3.16.

The performance of the active ring evaluated using the PC task [Fig. 2(d)] shows a significant dependence on the gain, with C_{PC} reaching its maximum at larger values of θ^{int} for lower feedback gains. Similar to the STM task, the active ring performs significantly better than the electronic components alone except at higher gains and time intervals. We obtain a maximum C_{PC} of 1.27 for an input time of 1 μs. These STM and PC results are comparable to those from Ref. [6] for a single spin-torque oscillator using a

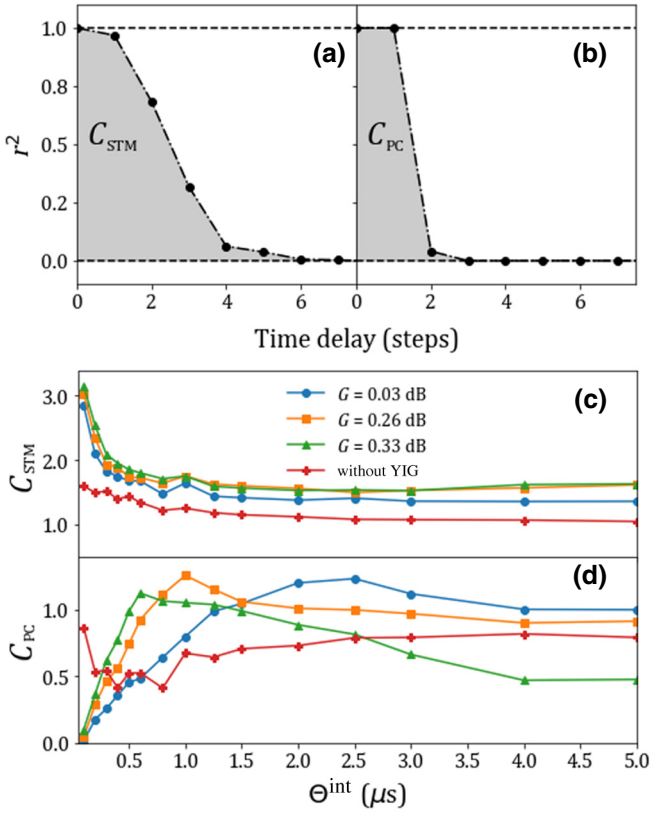

FIG. 2. (a),(b) The square of the correlation coefficient for increasing delay (forgetting curves) for the STM and PC tasks. The data shown here are for θ^{int} equal to 0.3 and 1 μs, respectively. (c) The STM capacity versus θ^{int} for different gain levels. The data labeled "without YIG" correspond to the STM capacity of the electronic components alone. (d) The same as (c) but for the PC task.

single-binary-sequence output. Reference [6] shows how the results can be further improved by averaging over multiple repeated sequences.

IV. CONCLUSION

The results that we obtain suggest that the system we propose has potential for development in the growing field of physical RC implementation. As well as being a simple design, this concept requires little additional pre- or postprocessing of the input or output. Furthermore, YIG is one of the most nonlinear materials available and is extremely robust. The closest competitor to our concept, the implementation based on the spin-torque nano-oscillator, requires injection of currents with high densities for their operation and can be damaged by excess currents. This technical problem is naturally solved for our concept of a dielectric-based device. Furthermore, low-magnetic-loss YIG films are now available in a very large range of thicknesses—from several nanometers to several tens of micrometers. With a decrease in the film thickness, the spin-wave group velocity decreases, leading to an increase in the delay time per unit film length. In addition, for thinner films, one may also expect that smaller microwave powers (i.e., smaller amplification gains in the feedback loop) are needed to reach the nonlinearity threshold for spin waves. This shows that our concept can easily be scaled down for integration on chips and one may even expect improvement of performance with scaling down.

ACKNOWLEDGMENTS

We acknowledge the Research Collaboration Award and the Vice Chancellor's Senior Research Award from the University of Western Australia. The work of S.W. was supported by the Australian Government Research Training Program.

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

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

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

[5] 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).

[6] 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).

[7] M. Riou, J. Torrejon, B. Garitaine, F. Abreu Araujo, P. Bortolotti, V. Cros, S. Tsunegi, K. Yakushiji, A. Fukushima, H. Kubota, S. Yuasa, D. Querlioz, M. D. Stiles, and J. Grollier, Temporal Pattern Recognition with Delayed-Feedback Spin-Torque Nano-Oscillators, Phys. Rev. Appl. **12**, 024049 (2019).

[8] 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).

[9] 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).

[10] 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).

[11] W. Jiang, L. Chen, K. Zhou, L. Li, Q. Fu, Y. Du, and R. H. Liu, Physical reservoir computing using magnetic skyrmion memristor and spin torque nano-oscillator, Appl. Phys. Lett. **115**, 192403 (2019).

[12] 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).

[13] G. Bourianoff, D. Pinna, M. Sitte, and K. Everschor-Sitte, Potential implementation of reservoir computing models based on magnetic skyrmions, AIP Adv. **8**, 055602 (2018).

[14] H. Nomura, T. Furuta, K. Tsujimoto, Y. Kuwabiraki, F. Peper, E. Tamura, S. Miwa, M. Goto, R. Nakatani, and Y. Suzuki, Reservoir computing with dipole-coupled nanomagnets, Jpn. J. Appl. Phys. **58**, 070901 (2019).

[15] H. Nomura, K. Tsujimoto, M. Goto, N. Samura, R. Nakatani, and Y. Suzuki, Reservoir computing with two-bit input task using dipole-coupled nanomagnet array, Jpn. J. Appl. Phys. **59**, SEEG02 (2020).

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

[17] M. Wu, in *Solid State Physics—Advances in Research and Applications* Vol. 62 (Elsevier, USA, 2010), Nonlinear Spin Waves in Magnetic Film Feedback Rings, p. 163.

[18] D. D. Stancil and A. Prabhakar, *Spin Waves: Theory and Applications* (Springer, New York, 2009), p. 1.

[19] A. K. Zvezdin and A. F. Popkov, Contribution to the nonlinear theory of magnetostatic spin waves, Sov. Phys. JETP **57**, 350 (1983).

[20] M. Chen, M. A. Tsankov, J. M. Nash, and C. E. Patton, Microwave Magnetic-Envelope Dark Solitons in Yttrium Iron Garnet Thin Films, Phys. Rev. Lett. **70**, 1707 (1993).

[21] Note the fundamental difference of our system to those in references 22 and 23, where the systems naturally represent auto-oscillators and the positive feedback is added artificially as a distortion of the natural auto-oscillation process.

[22] J. Williame, A. Difini Accioly, D. Rontani, M. Sciamanna, and J. V. Kim, Chaotic dynamics in a macrospin spin-torque nano-oscillator with delayed feedback, Appl. Phys. Lett. **114**, 232405 (2019).

[23] T. Taniguchi, N. Akashi, H. Notsu, M. Kimura, H. Tsukahara, and K. Nakajima, Chaos in nanomagnet via feedback current, Phys. Rev. B **100**, 174425 (2019).

[24] M. Wu, B. A. Kalinikos, and C. E. Patton, Self-Generation of Chaotic Solitary Spin Wave Pulses in Magnetic Film Active Feedback Rings, Phys. Rev. Lett. **95**, 237202 (2005).

[25] Z. Wang, A. Hagerstrom, J. Q. Anderson, W. Tong, M. Wu, L. D. Carr, R. Eykholt, and B. A. Kalinikos, Chaotic Spin-Wave Solitons in Magnetic Film Feedback Rings, Phys. Rev. Lett. **107**, 114102 (2011).

[26] M. M. Scott, C. E. Patton, M. P. Kostylev, and B. A. Kalinikos, Nonlinear damping of high-power magnetostatic waves in yttrium-iron-garnet films, J. Appl. Phys. **95**, 6294 (2004).

[27] M. P. Kostylev, Self-generation of bright spin-wave envelope solitons in active ferromagnetic-film rings, J. Commun. Technol. Electron. **50**, 313 (2005).

[28] M. Lukoševičius, in *Neural Networks: Tricks of the Trade: Second Edition*, edited by G. Montavon, G. B. Orr, and K.-R. Müller (Springer, Berlin, 2012), p. 659.

[29] 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).

[30] H. Jaeger, A tutorial on training recurrent neural networks, covering BPPT, RTRL, EKF and the "echo state network" approach, ReVision **2002**, 1 (2005).

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