


Implementing a Magnonic Reservoir Computer Model Based on Time-Delay Multiplexing

Stuart Watt¹, Mikhail Kostylev^{1,2,*}, Alexey B. Ustinov², and Boris A. Kalinikos²

¹*Department of Physics, University of Western Australia, Crawley, Western Australia 6009, Australia*

²*Department of Physical Electronics and Technology, St. Petersburg Electrotechnical University, St. Petersburg 197376, Russia*

 (Received 15 October 2020; revised 26 May 2021; accepted 7 June 2021; published 24 June 2021)

In the present paper, we propose and experimentally verify a concept of a magnonic reservoir computer. The system utilizes the nonlinear behavior of propagating magnetostatic surface spin waves in a yttrium-iron garnet thin film and the time delay inherent in the active ring configuration to process time-dependent data streams. Higher reservoir dimensionality is obtained through the time-multiplexing method, whereby inputs to the system are multiplied by a mask to drive complex dynamics in the ring and the output is sampled in time to create a series of “virtual” neurons for each sample. The input mask is implemented as a train of microwave pulses injected to the system. To demonstrate the efficacy of the concept, the reservoir computer is evaluated on the short-term memory and parity-check benchmark tasks, and the physical system parameters are tuned to optimize performance. By incorporating a reference line to mix the input signal directly onto the ring-resonator output, both the amplitude and phase nonlinearity of the spin waves can be exploited. The addition of a second spin-wave delay line increases the delay time of the ring and enhances the fading memory capacity of the system. Configuring the second delay line to transmit backward volume spin waves also partly compensates the dispersive pulse broadening that is present because of the large delay time.

DOI: [10.1103/PhysRevApplied.15.064060](https://doi.org/10.1103/PhysRevApplied.15.064060)

I. INTRODUCTION

A reservoir computer (RC) is a brain-inspired machine-learning model, which aims to simplify the construction and training of recurrent neural networks by outsourcing learning to a readout layer only [1,2]. A RC is characterized by the “reservoir,” which is a collection of randomly interconnected nonlinear neurons with input and output connections. The purpose of the reservoir is to apply a nonlinear kernel mapping of some low-dimensional input data to a higher-dimensional output space, where the input data becomes linearly separable. Recurrent connections also allow the RC to account for temporal dependencies present in the input. The output is reconstructed from the reservoir states by a linear readout, making training of a RC very simple.

Despite the simplicity, RCs have shown to be very powerful tools in temporal data analysis [3]. Learning is outsourced to the readout layer, making the design of the reservoir arbitrary and not limited to simulated neurons on a computer. Any dynamical system that satisfies certain properties can be implemented as the reservoir. The first is a nonlinear mapping of the input to a higher-dimensional

state space to induce linear separability. The second is a “fading” memory where the system state in response to the current input depends also on the recent history of inputs, whose influence over the reservoir state decays with time.

These ideas lead to the utilization of physical systems to implement the RC model, where the complex, nonlinear dynamics already present in nature can be substituted for simulated dynamics. The field of physical RC is fast growing and covers many disciplines due to the vast range of nonlinear dynamical systems available in nature. Physical RC offers a possible solution to obtaining fast and more energy-efficient computing. Over the past several years, many different physical systems have been proposed and successfully demonstrated in order to implement the physical RC model (see, e.g., the review papers by Tanaka *et al.* [4] and Nakajima [5] and references therein).

Among the different implementations, spintronic based architectures are promising candidates due to the high nonlinearity of magnetization dynamics, low power usage, scalability, and compatibility with existing computing technology. Many recent studies have proposed spintronic based implementations including RC architectures based on individual or arrays of coupled spin-torque nano-oscillators [6–13], magnetic skyrmion memristors [14], magnetic skyrmion fabrics [15,16],

*mikhail.kostylev@uwa.edu.au

dipole-coupled nanomagnets [17,18], and spin-wave interference in garnet films [19]. The majority of works on spintronic RC have been carried out theoretically or through simulations, with only a few works [6,8–11] showing experimental results. In our recent publication [20], we added another RC concept based on the spin-wave delay-line active-ring auto-oscillator and experimentally demonstrated adequate performance on some benchmark tasks. This system utilizes the delay and nonlinear behavior of traveling spin waves in magnetic film feedback rings, and naturally satisfies the required properties of a suitable reservoir implementation.

In the present work, we propose an alternative RC implementation based on the spin-wave delay-line active ring. The proposed device represents an effectively passive ring resonator of traveling spin waves and makes use of the time delay of the spin-wave delay line to implement the time-delay RC model [21]. As in Ref. [20], the core of the present physical system is the spin-wave delay-line active ring. However, the ring is now operated below the auto-oscillation threshold and the delay time of the signal in the open ring is much larger. This allows us to employ methods of data input and output that are unavailable in the auto-oscillation regime. More specifically, we make use of sampling the output signal *along the delay line*, thus creating multiple virtual nodes simultaneously circulating in the ring (that is the essence of time multiplexing). We also make use of the faster reaction times of the ring with an external microwave input source to apply an analog temporal mask to the reservoir inputs, increasing the complexity of the reservoir dynamics. As we show below, these features result in qualitatively different dynamics of the reservoir and significantly improved reservoir performance with respect to Ref. [20].

Similar work on physical RC using spintronic devices with time-delayed feedback has been carried out in Refs. [11,13]. Here the addition of an external feedback loop to a spin-torque nano-oscillator was shown to improve the memory capacity of the reservoir. In this work we demonstrate that the intrinsic delay time of traveling spin waves can be used to implement the same model without the need for external delay.

The structure of this paper is as follows. In Sec. II A we introduce the experimental system and describe the physical dynamics behind spin-wave delay lines and the ring construction. The RC model is described mathematically in Sec. II B and we demonstrate how the spin-wave delay-line ring resonator can be utilized to implement the time-delay-based reservoir. Section II C describes the two benchmark tasks we employ to evaluate the performance of this system as a RC implementation. Results are presented in Sec. III where the STM and PC performance is measured compared to several parameters that characterize the experimental setup. In Sec. IV we present a modified version of our RC where a second spin-wave delay line

is added to the ring, which yields further improvement of performance on the benchmark tasks.

II. METHODS

A. Description of experimental setup

The experimental setup is shown schematically in Fig. 1(a). The ring (with green background) consists of a spin-wave delay line (shaded rectangle) with a feedback loop. The delay line is assembled from a 2-mm-wide, 5.7- μm -thick yttrium-iron garnet (YIG) film (dark-gray trapezoid), sitting on top of two 50- μm -wide short-circuited microstrip transducers (“spin-wave antennas”). These antennas are spaced $d_1 = 8.2$ mm apart and are used to excite and detect spin waves.

A spin wave is a fundamental type of magnetization dynamic in magnetic materials, representing collective precession of elementary magnetic moments (spins of localized electrons), whereby spins at neighboring crystal lattice sites are coupled via magnetic exchange and dipole-dipole interactions and the precession phase varies linearly in one direction in space. The spin-wave frequency depends on the strength of the static magnetic field applied to the film, the material’s magnetic parameters (saturation magnetization, magnetic anisotropies, and gyromagnetic ratio) and also on the sample’s geometry with respect to the magnetic field. In this work, the static magnetic field is applied in the plane of the YIG strip and parallel to the spin-wave antennas. In this configuration, the spin waves are called magnetostatic surface spin waves (MSSW) and travel perpendicular to the antennas. The angular frequency of the MSSW is determined by the dispersion relation [22]

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

Here H is the external magnetic field, applied perpendicular to the spin-wave propagation direction, and k is the spin-wave wave number. The constants $|\gamma|/2\pi = 2.8$ MHz/Oe, $4\pi M_s = 1750$ G, and $L = 5.7$ μm are the gyromagnetic ratio, the saturation magnetization, and the film thickness, respectively.

The spin-wave delay line works as follows. A microwave current injected into the input antenna (left-hand side) induces a microwave Oersted field, which drives precession of the magnetic moments localized near the antenna. Neighboring moments will precess due to exchange and dipole-dipole interactions, and precession is carried away from the antennas. Energy is carried along the YIG strip in the form of a spin wave. The inverse process occurs where the dynamic magnetic dipole field produced by the precessing moments induces an electromotive force in the output antenna (right-hand side).

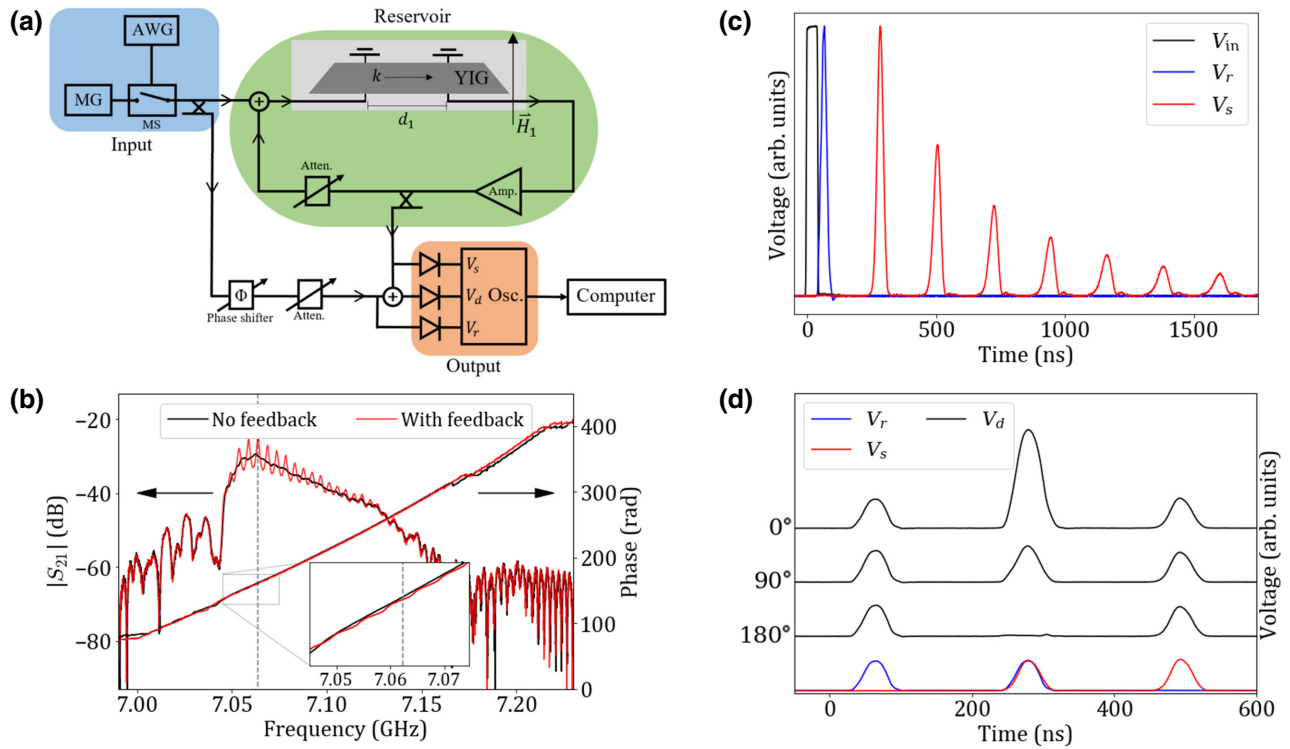


FIG. 1. (a) Schematic diagram of the spin-wave delay-line active-ring resonator system. Components are described in the text. (b) The amplitude and phase of the transmission (S_{21} parameter) characteristic for the MSSW delay line, measured with a vector network analyzer. The red trace shows resonance mode formation when feedback is added. (c) Time evolution of a single pulse injected into the ring. The 45-ns voltage pulse from the AWG (V_{in}) applied to the MS creates a microwave pulse, which travels through the reference line (V_r) and through the ring (V_s). The traces have been normalized. Time separation between pulses is $T_r = 215$ ns. (d) Double pulse interference of reference (blue) and delay line signals (red) for various reference signal phases.

The choice of MSSW has three benefits. First, MSSW are efficiently excited by spin-wave antennas. Second, the MSSW are unidirectional and will only be excited efficiently in one direction. This prevents energy lost to spin waves excited in the opposite direction. Finally, MSSWs have a large free propagation length in epitaxially grown monocrystalline YIG films due to very low microwave losses allowing long propagation times.

Placing the spin-wave delay line into the microwave signal path introduces a time delay and phase shift of the signal determined by the spin-wave group and phase velocities and the spin-wave antenna separation. Amplifying the output signal and feeding back into the input antenna creates a ring resonator [23–25] with resonant frequencies satisfying the condition $k_{res}d = 2\pi m$. Here k_{res} is the spin-wave wave number, d the antenna separation, and m an integer. We assume that the phase shift associated with the electrical components in the ring is negligible.

Spin waves in YIG films are highly nonlinear. For the frequencies employed in this work, the four-wave nonlinear processes determine the wave nonlinearity. The processes manifest themselves as nonlinear spin-wave damping [26] and a nonlinear shift of frequency or accumulated phase, depending on the experimental

conditions [27,28]. Because of this, YIG-film delay-line active rings have been extensively studied for the host of nonlinear behaviors they exhibit such as soliton formation, modulational instability, and chaos [29]. We recently demonstrated how these rings can effectively be constructed to implement the physical RC model [20]. That work exploited another useful feature of active rings called auto-oscillation. At low feedback gain, delay-line losses dominate and the amplitude of the ring signal drops below the noise level after a couple of cycles of circulation. However, above a threshold when the feedback gain is sufficient to compensate the delay-line losses, thermally excited magnons with wave numbers satisfying the above resonance condition will be resonantly amplified. We define the ring gain G as the difference between the delay-line losses and the feedback gain. Thus, at the auto-oscillation threshold, $G = 0$.

The work in Ref. [20] made use of the nonlinear damping of spin waves and the slow transient response of the ring signal to changes in G above the auto-oscillation threshold. In the present work, a similar ring structure is used but operated below the auto-oscillation threshold so that the input data, injected using an external microwave signal fixed to a resonance frequency of the ring, will

gradually lose power with each circulation. This gradual decay of signal amplitude implements the fading memory. The operation of the ring below the auto-oscillation threshold is one of the key differences between this work and that of Ref. [20]. This work exploits the physics of short spin-wave pulses traveling in dispersive and nonlinear media, and whose spatial duration is shorter than the length of the delay line [30]. The dynamics of spin-wave pulses circulating in the ring are more complex than the transient switching between stable states of an auto-oscillation signal performed in Ref. [20], and the RC performance is improved as a result.

A microwave generator (“MG”) produces a constant signal, which is converted into a series of pulses using a p - i - n diode microwave switch (“MS”) controlled using an arbitrary waveform generator (“AWG”). The data to be processed is encoded onto this input by varying the pulse amplitude. The signal is fed into the ring using a microwave combiner [plus symbol in Fig. 1(a)] after part of the signal is first sampled using a directional coupler (−6 dB) to act as a reference. The microwave pulses excite MSSW pulses at the input antenna, which travel along the YIG strip and convert back to microwave pulses at the output antenna. The signal is amplified with a low noise amplifier (“Amp.”) and fed back into the input antenna. Since the microwave carrier frequency matches a resonance frequency of the ring, the signal will interfere constructively with each circulation in the ring.

The ring signal is sampled with another directional coupler (−10 dB), placed after the amplifier, and combined with the reference signal. A tunnel diode rectifies the combined microwave signal. An oscilloscope (“Osc.”) measures the resulting voltage (labeled V_d). A phase shifter in the reference line adjusts the relative phase between the signal from the ring and the reference, and a variable attenuator adjusts the reference amplitude. Finally, a second variable attenuator in the ring feedback loop controls the amount of feedback amplification in the ring.

A static magnetic field of 1772 Oe is produced by two permanent neodymium magnets on a cast-iron yoke. Figure 1(b) shows the amplitude-frequency and phase-frequency characteristics of the delay line. When feedback is added to the delay line, ring resonance modes are formed. As the ring gain is raised above the auto-oscillation threshold, the resonance mode with the lowest loss is self-excited first, followed by the second lowest loss resonance mode, and so on. In this case, the frequency of the resonance mode with the lowest loss is 7.06 GHz and the external microwave signal is set to this frequency. As mentioned above, the ring is operated *below* the auto-oscillation threshold and so all inputs will gradually fade over time. To demonstrate this, Fig. 1(c) shows the time evolution of a 45-ns input pulse as it circulates in the ring below the auto-oscillation threshold. The diode voltage of the reference signal (labeled V_r) is delayed by about 40 ns

from the AWG control voltage (labeled V_{in}). The output from the ring (labeled V_s) shows how the pulse circulates multiple times, its amplitude decreasing with each successive cycle. YIG is a dispersive medium, which results in dispersive pulse broadening. The pulse energy spreads out over time, reducing the peak amplitude. The dispersion coefficient is determined as

$$D = \frac{\partial^2 \omega(H, k)}{\partial k^2}. \quad (2)$$

The time delay between the reference pulse and the pulse after one period of circulation in the ring is 215 ns, which matches the separation between successive circulations of the pulse. This indicates that the time delay in the feedback loop is negligible and that the circulation time is due predominantly to the spin-wave delay line.

B. RC operation

The RC model has received much attention due to its simple structure of just three components [31]—the input, reservoir, and readout. When constructed using software, the reservoir comprises a series of neurons with weighted connections between the input layer, output layer, and internal reservoir neurons. The reservoir state at some time step, T , is described by a vector, $\vec{x}(T)$, containing the values of each reservoir node and is updated as

$$\vec{x}(T) = f[\mathbf{W}^{in}\vec{u}(T) + \mathbf{W}\vec{x}(T-1)]. \quad (3)$$

The matrices \mathbf{W}^{in} and \mathbf{W} determine the input and internal connection weights, respectively, and $\vec{u}(T)$ is the vector of input values. The reservoir dimensionality is usually much higher than that of the input, with a large number of neurons characterized by a nonlinear activation function, f . The purpose of the reservoir then is to perform a kernel mapping from lower to higher dimensions. Unlike the recurrent neural networks, where all the connection weights are trained using the back-propagation-through-time method, \mathbf{W}^{in} and \mathbf{W} are randomly initialized and remain fixed throughout the training. Only the readout weights are trained. The output of the system is obtained as a weighted sum (linear regression) of the reservoir states

$$\vec{y}_{out}(T) = \mathbf{W}^{out}\vec{x}(T). \quad (4)$$

During the training phase, the weights \mathbf{W}^{out} are adjusted to reduce the mean-squared error between $\vec{y}_{out}(T)$ and some target output. When implemented physically, \mathbf{W} and f are determined by the physical parameters of the system, while \mathbf{W}^{in} is determined by the input mechanism to the system. Presently, the dynamics of the microwave signal in the ring determine \mathbf{W} and f , while the AWG control voltage applied to the MS determines \mathbf{W}^{in} .

While conventional RC models make use of large numbers of neurons to create complex reservoir dynamics, the present physical system essentially acts as a single nonlinear neuron with feedback onto itself. The spin-wave delay line performs nonlinear operations onto the input microwave signal and the amplified feedback loop provides a recurrent connection. While the dynamics of spin waves in these rings are complex, the system still only has a single input and output. Inputs are injected via the externally applied microwave signal, which excites spin-wave dynamics in the ring. For each input, the ring signal is then measured, providing the output from the reservoir. This output is a scalar amplitude, which varies in time.

Despite its inherent low dimensionality, this physical system is suitable to fully implement the time-multiplexing method developed in Ref. [21]. That work showed how a single nonlinear node can be used to emulate an entire reservoir by substituting many spatially separate neurons with a series of “virtual” neurons separated temporally. The time-multiplexing method boosts the dimensionality of a single node reservoir. The RC architectures, which implement this method, are referred to as time-delay RC (TDRC) models. TDRC models have been extensively studied in optical and optoelectronic circuits [32,33]. The main benefit of TDRC is in the simplicity of fabrication, requiring only a single nonlinear node in place of an entire array of neurons, each with their own inputs and outputs.

To implement this model, each input, $u(T)$, is applied to the reservoir for a time interval of duration θ^{int} . This input time interval is split into N equal-length subintervals of duration θ_N , such that $N\theta_N = \theta^{\text{int}}$. The output from the reservoir is sampled at each subinterval to create a vector of output values for each input. In this way the scalar output from the single nonlinear node is converted into a vector of length N , where each value represents a “virtual” neuron in the network.

In order to benefit from this time-multiplexing method, the output should be complex enough that its value varies sufficiently across the input time interval, θ^{int} . To ensure this, a temporal mask is often applied to the input to induce more complex reservoir dynamics than just a constant input signal.

There are a variety of ways to construct the mask such as sampling from a random binary distribution, noise or even a chaotic time series [34], with the general aim of increasing the variation of the reservoir dynamics. In this work, we make use of an analog mask to convert the constant input microwave signal into a train of microwave pulses. These pulses have a width of 45 ns and a repetition period of 71.7 ns. This means that three consecutive pulses match the circulation period of the ring (i.e., $3 \times 71.7 = T_r$). The purpose of the mask is to induce complex dynamics in the reservoir. Ideally, the input mask would vary much more

rapidly, and hence be more complex, but we are limited by the capabilities of the available electronic equipment. A train of pulses gives the optimal input complexity with the available electronic resources. A variable input is achieved by varying the amplitude of the pulses. For this, the control voltage is set within the range $V_{\text{in}} = [-300, 200]$ mV. Figure 2(a) shows how the pulse amplitude varies approximately linearly from maximum to zero across the range of V_{in} .

The input mask is chosen to be equal to four pulses in length, making the input time interval $\theta^{\text{int}} = 4 \times 71.7 = 286.8$ ns. This number of pulses is chosen empirically to give the best compromise between fading memory and complexity of reservoir dynamics. The total input time interval, θ^{int} , should be on the order of the circulation period of the time-delay reservoir. This enables maximum overlap of virtual neurons from different inputs. Importantly, this ensures that the virtual neuron spacing, θ_N , is smaller than the delay time for the system. Multiple virtual neurons are then stored in the fading memory of the system as physical signals that travel along the delay line. However, it is not necessary that $\theta^{\text{int}} = T_r$. As was demonstrated in Ref. [35], desynchronization of the inputs to the circulation period of the ring leads to a richer interaction of virtual neurons and higher computational performance. How the performance of the RC depends on the length of the temporal mask is discussed in the Appendix.

To give an example of the temporal masking procedure, Fig. 2(b) shows how a continuous input is sampled into discrete values for each time step and encoded as a sequence of four microwave pulses with varying amplitude. The inclusion of the masking procedure is one of the differences of the present work from the work in Ref. [20]. In that work, the delay line with feedback provided a means to generate auto-oscillation and provide a slow transient response of the system to external stimuli. In that regard, the ring could be considered as a single, lumped element, whose output experienced a transient response to changes in the input, and *not* as a storage ring for virtual neurons. The relatively slow transient response of the system in Ref. [20] (many times greater than the ring circulation period) also meant that the masking procedure is not feasible since the system could not relax fast enough to capture the variations within each input time interval. The present system responds much faster due to an external microwave input and its operation below the threshold of auto-oscillation, allowing masking of the input. The additional complexity afforded by this feature results in improved RC performance.

Each input is injected into the system and the reservoir state vector, $\vec{x}(T)$, for each input is measured using the fast microwave diode. Due to the time delay introduced by the spin-wave propagation, the ring output signal at time T is not influenced immediately by the current input, $\vec{u}(T)$, for that interval. The addition of the reference signal resolves

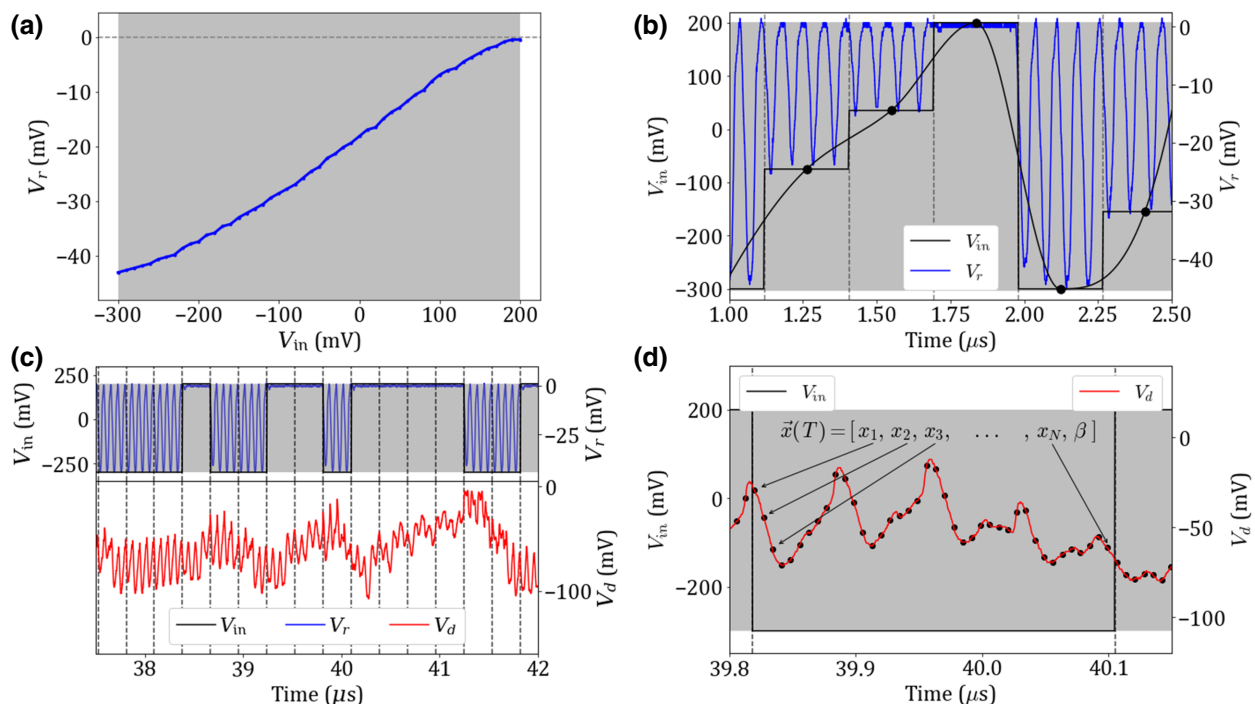


FIG. 2. (a) Amplitude of the pulses created using the MS as a function of V_{in} , measured using the reference signal. (b) Example of sampling and masking of a continuous signal to create the RC input. V_{in} represents the amplitude of the voltage pulses from the AWG, and V_r shows the microwave pulses created by the MS and measured with the microwave diode. (c) Example of control voltage (V_{in}), RC input (measured as V_r from the reference line) and the RC output (V_d) for the binary input of the STM and PC tasks. (d) Example of sampling each interval into N virtual neurons to create the system state vector $\vec{x}(T)$. To each output a constant $\beta = 0.1$ is concatenated, making the output dimension $N + 1$.

this problem by adding the input directly to the output. The RC output equation then becomes

$$\vec{y}_{out}(T) = \mathbf{W}^{out}[\vec{x}(T) + \alpha \vec{u}(T)]. \quad (5)$$

When constructing a RC model using software the input can be directly included into the linear regression in the postprocessing stage. By feeding part of the input off using the reference line, this step is done physically, simplifying the postprocessing without any degradation in processing time. Furthermore, adding the reference signal to the RC output adds more tunable parameters to the system; notably the relative phase of the reference signal with respect to the ring signal and the input scaling factor α .

The scaling factor α represents an attenuation of the reference-signal amplitude before combining it with the ring output signal. The attenuation is constant in all measurements such that the pulse from the reference signal and the pulse after one pass through the delay line are equal in amplitude. Figure 1(d) shows this case. With the feedback line disconnected, two pulses are injected into the system with a separation equal to the delay time (215 ns). The blue and red traces show the reference signal and delay-line signal, respectively, while the black traces show the combined signal. The pulses interfere constructively or destructively

depending on their relative phase. This interference provides an extra layer of complexity and nonlinearity to the diode output.

Figure 2(c) shows an example of the system output in response to a binary input. For each input interval, the output diode voltage is sampled into N equispaced points to obtain the virtual node values, as shown in Fig. 2(d). This creates an output vector, $\vec{x}(T)$, for each input with a dimensionality of N . An additional bias term, $\beta = 0.1$, is concatenated to $\vec{x}(T)$ for all time intervals in the postprocessing stage [7,8]. This term corresponds to one additional “virtual” neuron, which adds a constant to the linear combination of the output. The nonzero value of β is chosen arbitrarily and does not affect the computational performance. The total output dimensionality is then $N + 1$.

C. Benchmark performance tasks

To evaluate the performance of the ring-resonator system as a RC, we employ the short-term memory task (abbreviated earlier as STM) [36] and the parity-check task (abbreviated earlier as PC) [7,37]. The employed procedure is the same as in our previous work [20]. However, we believe that it is useful to briefly describe it here for

the sake of completeness of the paper. In both tasks, the system state at a given time T is measured in response to a random binary input $u(T) \in [0, 1]$.

The STM task provides a measure of the fading memory present in the system. The target for each time step is simply the input at some delay (τ steps) in the past determined by

$$\hat{y}_{\text{STM}}(T, \tau) = u(T - \tau). \quad (6)$$

The PC task is a nonlinearly separable task, which requires both fading memory and nonlinearity. The target for each time step is determined by taking the parity of the sum of the consecutive inputs up to some delay in the past

$$\hat{y}_{\text{PC}}(T, \tau) = \text{PARITY}[u(T - \tau), u(T - \tau + 1), \dots, u(T)]. \quad (7)$$

Here the PARITY operation returns the parity (0 for even, 1 for odd) of the sum of the values in the brackets. In both tasks, the targets are also binary values.

A sequence of 2200 random binary inputs is fed into the RC to perform the tasks. The inputs $[0, 1]$ are converted to corresponding AWG input control voltages [200 mV, -300 mV] (no pulse or maximum amplitude pulse), respectively. Figure 2(c) shows an example of the input and output for these tasks.

Since both tasks require the same input, they can be performed simultaneously. Conventional recurrent neural networks are trained for specific tasks and are then not generalizable. In RC, only the readout weights are trained, allowing the same reservoir to be adapted to a wide range of tasks. This multitasking property is especially useful when implementing the RC physically, since the physical parameters of the system need not be altered to execute a specific task.

The entire 2200-value sequence is fed into the ring and 2200 output vectors are recorded in one batch. Training of \mathbf{W}^{out} and evaluation can all be done “offline” later. The first 200 outputs are discarded. \mathbf{W}^{out} is trained on the following 1000 outputs and evaluated on the remaining 1000 outputs. Defining \mathbf{Y} and \mathbf{X} as matrices containing the targets and reservoir states for all time steps in the training set [i.e., \mathbf{X} is the matrix of reservoir state vectors $\vec{x}(T)$ for all T concatenated horizontally, similarly for \mathbf{Y}], the optimal \mathbf{W}^{out} is obtained by taking the product $\mathbf{Y}\mathbf{X}^{-1}$. This singular training step is stable and very fast. A different \mathbf{W}^{out} is trained for each value of τ . The success of the linear regression to reconstruct the desired target is measured by calculating the square of the correlation coefficient between the reconstructed target, y_{out} , and the actual target, \hat{y} [7].

Figures 3(a) and 3(b) show examples of the comparison between the reconstructed target and the actual target for the STM and PC tasks, respectively. The x axis is denoted

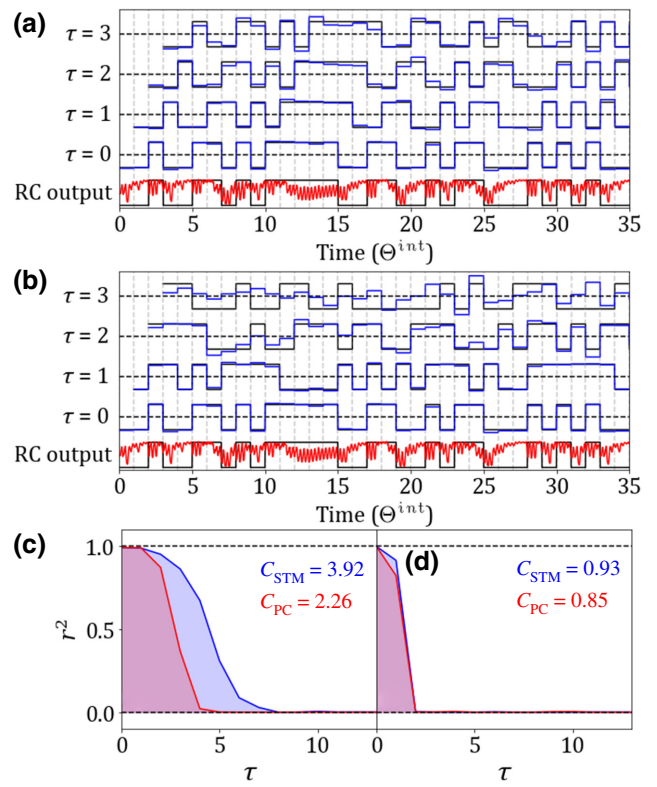


FIG. 3. (a) Comparison of the targets, $\hat{y}_{\text{STM}}(T)$ (black), and reconstructed targets, y_{out} (blue), for the linear regressions trained on the RC output (red) for the STM task. (b) Same as for (a) but for the PC task. (c) $r(\tau)^2$ for increasing delay (forgetting curve) for the STM and PC tasks. The data shown here is for a ring loss of 7.68 dB. The reference signal is in antiphase to the ring output. (d) Forgetting curve for the STM and PC tasks evaluated using only the reference signal, V_r .

in units of θ^{int} . The black traces show the targets, \hat{y} , for each input interval while the blue traces show the reconstructed (predicted) targets, y_{out} , after the readout weights have been trained. As τ is increased, the ability to correctly predict the target from the current output is reduced.

This behavior is summarized using “forgetting” curves [shown in Figs. 3(c) and 3(d)], which plot $r(\tau)^2$ against τ . These curves are visual aids to show how the reconstruction performance depends on the delay. Finally, taking the sum of $r(\tau)^2$ over the range of τ (equivalent to the area under the forgetting curves) returns the STM and PC capacities as

$$C_{\text{STM-PC}} = \sum_{\tau=1}^{\tau_{\text{max}}=20} r(\tau)^2. \quad (8)$$

Ten different 2200-value binary sequences are passed through the system and the capacities are calculated for each trial and averaged. The uncertainties in C_{STM} and C_{PC} are calculated as the standard deviation across the ten trials.

III. RESULTS AND DISCUSSION

In this section, we explore how the STM and PC task performance depends on the various tunable parameters in the system. In order to obtain meaningful estimations of the system performance, the contribution from the ring must be isolated from that of the preprocessing and input generation steps (nonlinearities and transient behaviors arising from the MS). The STM and PC capacities are first measured using only the signal from the reference line, essentially removing the ring completely. The output equation becomes

$$y_{\text{out}}(T) = \mathbf{W}^{\text{out}}[\alpha u(T)]. \quad (9)$$

Since \mathbf{W}^{out} is linear and the reference signal is instantaneous and memoryless, the STM and PC capacities should ideally be 0. This is not the case however, as can be seen in the forgetting curve in Fig. 3(d).

The reason that the $r(\tau)^2$ do not drop to zero for $\tau = 1$ is that the pulses are not completely separated, and the fourth pulse of the previous input merges slightly into the first pulse of the current input [see Fig. 2(b)], hence there is a slight fading memory that only goes back one unit of delay. The forgetting curve in Fig. 3(d) supports this claim. The linear regression can accurately predict the current input and also adequately predict the previous input. The prediction accuracy drops to 0 for a delay of $\tau = 2$ or more.

These benchmark measurements allow us to evaluate the influence of the ring alone on the computation by isolating the contribution to computation due to the preprocessing steps and input generation at the MS. In the figures that follow, these benchmark capacities are displayed as horizontal dashed lines.

A. Performance against ring gain

Tuning the ring gain controls the fading memory of the system. The low-noise amplifier provides a constant amplification and the total gain of the ring is controlled with the variable attenuator. We thus define the ring gain in terms of the loss introduced by the attenuator relative to the auto-oscillation threshold. As ring loss is increased, the inputs decay more rapidly with each circulation of the ring. This is equivalent to a reduction in memory.

Figure 4 shows the STM and PC capacities measured against the ring loss. Two different operational schemes are compared. In the first, the reference line is disconnected and the output is sampled from V_s . In the second, the reference line is connected and the system output depends on both the input and the ring signals, as in Eq. (5). In this case, the output is sampled from V_d . The relative phase of the reference signal is chosen to be completely in phase ($\Phi = 0^\circ$) and in antiphase ($\Phi = 180^\circ$) with the ring signal.

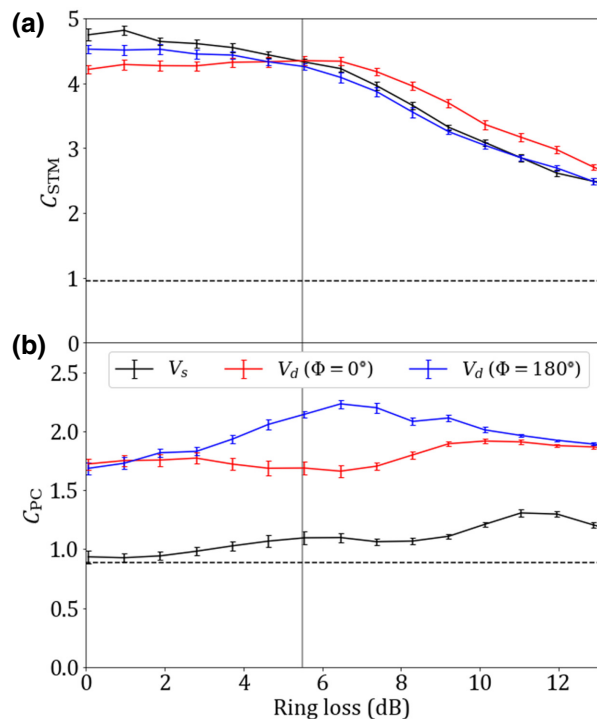


FIG. 4. (a) C_{STM} measured for decreasing G . The black data are measured without the reference signal (i.e., V_s). The red (blue) data are measured with the reference line connecting in phase (antiphase) with the ring signal (i.e., V_d for $\Phi = 0^\circ$ and $\Phi = 180^\circ$). The dashed black line represents the RC performance evaluated using only the reference signal (i.e., V_r). (b) Same as for (a) but with the PC task.

C_{STM} decreases monotonically as the ring loss is increased due to the reduced pulse persistence. Note that *qualitatively* the behavior is similar between the three different cases presented. The addition of the reference signal to the output of the RC should not affect the memory capacity in the ring because the fading memory originates in the ring alone. There is, however, a slight reduction in the maximum STM capacity reached when the reference signal is included, perhaps being the result of the reference signal partly obscuring the ring signal. The maximum C_{STM} reached in the first scheme is 4.82 ± 0.09 . For the second scheme, the maximum C_{STM} is $4.35(4.53) \pm 0.08(0.08)$ when the reference is in phase (antiphase).

As to the PC task, the addition of the reference line significantly improves C_{PC} due to the nonlinear nature of the task, reaching a maximum of 2.23 ± 0.05 in the antiphase case and only 1.31 ± 0.05 in the case without the reference signal. There are two reasons for this. First, the system output is not actually influenced by the current input if the reference line is disconnected, apart from the transient merging of one input into the next. The second reason is the interference of the two signals when the reference line is connected. This interference adds a degree

of complexity to the output and thus helps the computation. The amplitude of an interference pattern is a nonlinear (harmonic) function of the phase difference between the interfering signals adding more nonlinearity to the system. Furthermore, by combining the ring signal with a reference one, both amplitude and phase information from the ring signal are imparted to the output. In addition to the nonlinear damping [26] exploited in Ref. [20], spin waves also exhibit a nonlinear power-dependent phase shift [27,28]. This additional source of nonlinearity in the reservoir aids in the quantitative improvement of the computation performance from our previous work.

B. Performance against reference phase

As shown in Fig. 4(b), C_{PC} is significantly improved when comparing whether the reference signal is in phase or in antiphase with the ring output. To explore this, the ring loss is set to 5.49 dB (vertical gray line in Fig. 4) and C_{STM} and C_{PC} are measured against the reference phase. At this point, all three cases in Fig. 4(a) returned the same C_{STM} . Indeed this is confirmed in Fig. 5(a) where C_{STM} is unchanged for all reference phases. On the other hand, the PC results show a strong dependence on the reference phase, C_{PC} being maximized when the reference is in antiphase to the ring output.

The traces at the bottom of Fig. 5(a) show the mixed pulse at the output. Note that due to dispersive broadening of the spin-wave pulses, the output pulse from the ring is slightly wider than the pulse from the reference line, and the shoulders of the pulse remain even when the two pulses are in antiphase. This fact may actually benefit the computational performance of the ring RC, as the pulses get wider with each circulation of the ring. The shape of the output trace will then become more complex. Note that for this complexity to be captured, one needs to sample a sufficient number of virtual neurons from the output.

The data presented so far use $N = 40$ virtual neurons. Figure 5(b) shows how C_{STM} and C_{PC} depend on N . After approximately $N = 40$, the performance saturates. With the addition of more neurons, the nodal separation becomes shorter than the timescale of variations in the output and these extra neurons do not add any information to the linear readout.

C. Discussion

From an engineering standpoint, this system is very simple and advantageous. Little preprocessing is required in order to inject data into the ring and the linear readout can be performed directly on the microwave diode voltage, requiring no postprocessing. By utilizing the time-multiplexing method, the total device has much fewer components that a full network of spatially distinct nodes, especially considering the time delay and nonlinearity are combined within the same element (i.e., the delay line).

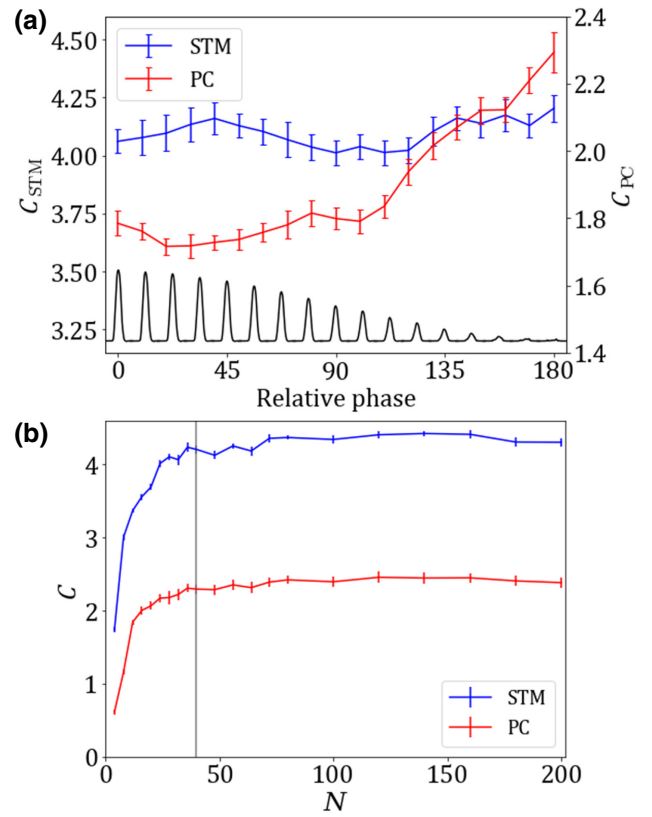


FIG. 5. (a) C_{STM} and C_{PC} measured against the relative phase between reference and ring signals, Φ . The black traces show the interference pattern of two pulses. (b) Dependence of C_{STM} and C_{PC} on the number of virtual neurons, N , sampled from the output. The vertical line designates $N = 40$.

The ring itself comprises only the YIG film, an amplifier, a single input connection and a single output connection. While the system presently is quite large, scaling to microscopic sizes can easily be achieved.

By using thinner YIG films, the spin-wave group velocity can be reduced significantly, allowing the antenna separation to be decreased to micrometer scales. Although the losses associated with exciting spin waves in much thinner films are larger, device optimization (for example, by fabricating coplanar antennas directly on the film surface [38] and impedance matching to the feeding line) could minimize additional losses in a microscopic device. On the other hand, exciting nonlinear spin-wave dynamics in a microscopic device would become much more energy efficient than in the present macroscale device.

While the RC performance of the present system is currently below that of other implementations of the TDRC model, such as in optical and optoelectronic circuits [32,33], it has advantages in terms of scaling and simplicity. One method of boosting the computational performance is to increase the complexity of the system by combining multiple rings together, either in parallel, in

series, or a combination of both. For instance, the in-series connection can be achieved by driving the input of one ring from the output of another. As for the parallel connections, one can think of inserting microwave directional couplers into the rings to couple dynamics of nearest-neighbor ones. By coupling multiple rings together, the total dynamics may be much richer, at the expense of a more complicated construction. These considerations are beyond the scope of this paper.

IV. EXTENDED DELAY LINE

In this section, the delay time of the ring is increased by the addition of a second delay line connected in series with the first. The experimental setup is depicted schematically in Fig. 6(a) where the output of the first delay line is amplified and then injected into the input of the second. The additional delay line is constructed in the same way as the first. The antenna separation is $d_2 = 4.9$ mm. The defining difference between the two delay lines is the orientation with respect to the applied magnetic field. The first delay line is orientated such that the spin-wave propagation direction is perpendicular to the

magnetic field, exciting MSSW. The second delay line is oriented such that the spin-wave propagation is parallel to the in-plane magnetic field, exciting backward-volume spin waves (BVSW). These spin waves are characterized by the dispersion relation [22,39]

$$\omega(H, k) = \gamma \sqrt{H \left[H + 4\pi M_s \left(\frac{1 - e^{-kL}}{kL} \right) \right]}. \quad (10)$$

The BVSW delay line is placed in a magnetic field of 1813 Oe to align the transmission band to that of the MSSW delay line. The transmission characteristics of each individual delay line as well as that of the combined delay line are shown in Fig. 6(b).

Delay-line rings configured using either MSSW or BVSW operate in the same manner. Thus the RC performance of a ring configured for the transmission of BVSW would yield results qualitatively similar to the preceding sections. Where the two forms of spin wave differ is their excitation efficiency and dispersion coefficient. The excitation of BVSW is not as efficient as MSSW due to the way that the components of the dynamic magnetization vector couple to the driving magnetic field. For the same

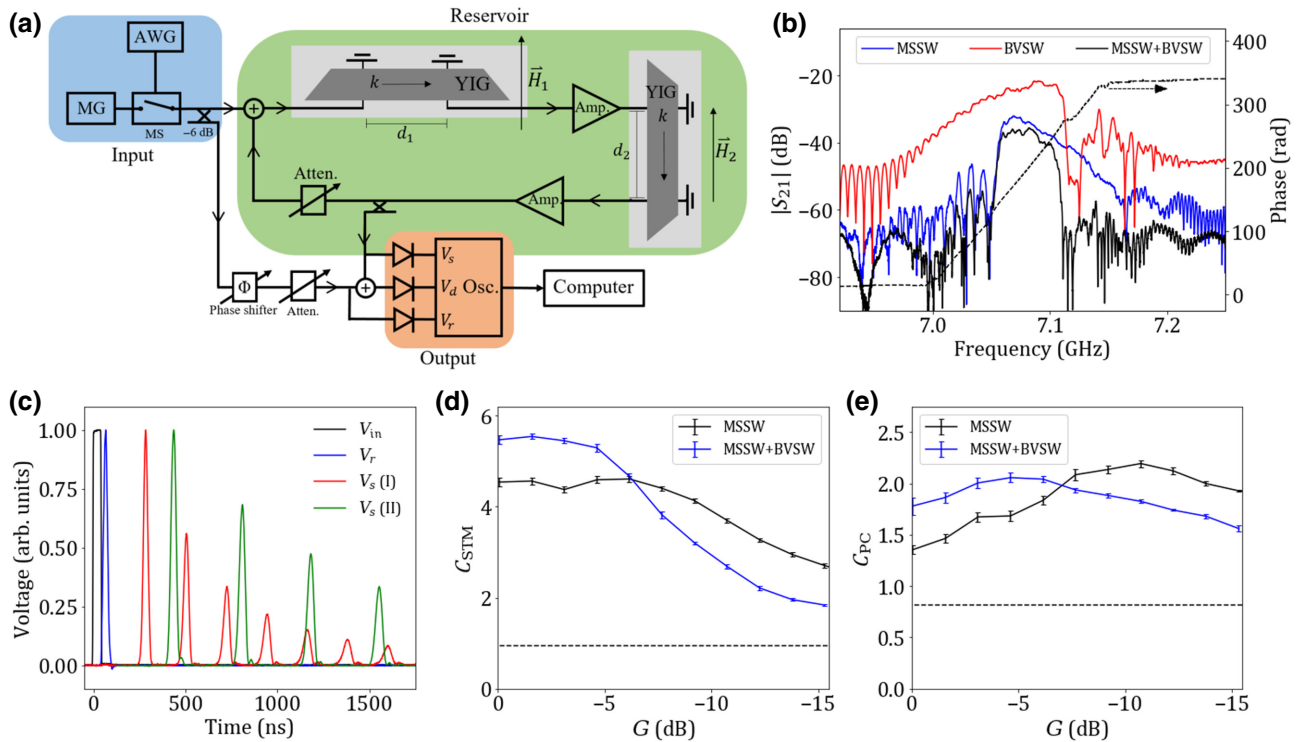


FIG. 6. (a) Schematic diagram of the extended spin-wave delay-line ring-resonator system. (b) Amplitude transmission characteristics for the MSSW, BVSW, and combined (MSSW+BVSW) ring configurations. The dashed black line shows the transmission phase of the combined delay line. (c) Time evolution of a single pulse injected into the ring. The 45-ns voltage pulse from the AWG (V_{in}) creates a microwave pulse at the MS, which travels through the reference line (V_r) and through the MSSW [V_s (I)] and MSSW+BVSW [V_s (II)] delay-line rings. The traces have been normalized. (d) C_{STM} measured for increasing ring loss (decreasing G) with $\Phi = 180^\circ$ for the MSSW and combined MSSW+BVSW delay-line rings. The dashed black line represents the RC performance evaluated using only the reference signal, V_r . (e) Same as (d) but for PC task.

microwave power, the MSSW are excited more strongly, and at the output antenna, the inverse process is likewise more efficient. Furthermore, the excitation of BVSW is not unidirectional, so half power of the input is instantly lost to BVSW traveling away from the second antenna. Thus, from an engineering viewpoint, the use of the MSSW is favorable because smaller powers are needed for the ring operation. Alternatively, for the same available input power, the MSSW-based ring will operate in a significantly more nonlinear regime than the BVSW-based one. For these reasons, the MSSWs are chosen as the information carrier in the preceding sections.

On the other hand, the choice of using a BVSW delay line instead of a second MSSW delay line is due to the opposing dispersion coefficients [Eq. (2)] of BVSW and MSSW. MSSWs have a negative dispersion coefficient. With the addition of a second MSSW delay line, the pulses would broaden more with each circulation of the ring and their amplitude would decay faster. Since the ring state is measured using the amplitude of the microwave signal in the ring, the faster broadening of the pulses effectively reduces the amount of memory of the system.

BVSW have a positive dispersion coefficient, which can partly compensate the broadening of the pulses due to the first MSSW delay line. In this way, the total ring circulation time can be extended, without the addition of excessive pulse broadening. Figure 6(c) compares the evolution of a single pulse traveling around the single or double delay-line ring. The BVSW delay line introduces an additional 152-ns delay. In addition to this, the pulse, which circulates in the double-delay-line ring, decreases in amplitude at a lower rate than in the single-delay-line case. Indeed, the width of the pulse in the single-delay-line ring grows faster than in the double-delay-line case, due to the opposite dispersion coefficients.

The operation of the double-delay-line ring is functionally the same. The input is similarly masked to produce a train of 45-ns-wide pulses. The repetition period is set to 73.4 ns, such that five pulses match the circulation period of the double-delay line ring. Using the results for optimizing the ring RC performance determined in the previous section, C_{STM} and C_{PC} are measured against the ring loss with the following parameters. The reference line is connected in antiphase with the ring. The temporal mask is set to have a duration of six pulses (one greater than the number of pulses, which simultaneously circulate in the ring to avoid input synchronization with the ring). Finally the output is again sampled using $N = 40$ neurons.

The double-delay-line ring RC shows the same qualitative performance as the single-delay-line ring, with some differences. The first is a significant improvement of C_{STM} at a low ring loss [as shown in Fig. 6(d)]. This is most likely due to the slower decay rate of the pulse amplitude during each round trip of the ring.

The PC capacities [shown in Fig. 6(e)] do not show the same improvement, but instead sees a translation of the curve to lower ring loss so that the maximum C_{PC} now coincides with the larger C_{STM} values. Typical RC models show a trade-off between fading memory and nonlinearity [40], however this trade-off is not as severe in the double-delay-line ring RC. C_{STM} and C_{PC} reach optimal values of 5.43 ± 0.08 and 2.05 ± 0.05 , respectively, for the same set of system parameters.

V. CONCLUSION

In conclusion, we show experimentally the physical implementation of the time-delay reservoir computing model using the spin-wave delay-line ring resonator operating with large values of delay time. We show that this system naturally satisfies the required reservoir properties of fading memory and nonlinearity due to the dynamics of the spin-wave delay line. By incorporating the reference line into the design of the RC, the system can exploit both the amplitude (nonlinear damping) and phase (nonlinear phase shift) nonlinearities of spin waves, which is shown to significantly improve its nonlinear computing capacity without drastic diminishing of the fading memory capacity. We evaluate performance using two benchmark tests and show improved performance with regards to our previous RC concept. Finally, we extend the delay time of the ring by inserting a second delay line configured for the BVSW resulting in a partial compensation of the dispersive pulse broadening, improvement of the linear memory of the system and much better trade-off between linear memory and nonlinearity.

ACKNOWLEDGMENTS

The work of S. Watt is supported by the Australian Government Research Training Program. The work was supported by the Ministry of Science and Higher Education of the Russian Federation, Grant No. 075-15-2021-609.

APPENDIX: PERFORMANCE AGAINST INPUT MASK LENGTH

Here C_{STM} and C_{PC} are measured against the ring loss when the input temporal mask duration is equal to one to four pulses. We define the number of pulses that constitute the mask as n . The input time interval is then given by $\theta^{int} = n \times 71.3$ ns. For these measurements, the ring contains a single MSSW delay line and the reference signal is kept in antiphase with the ring signal. A constant virtual neuron subinterval is used across all measurements of $\theta_N = 7.17$ ns. This means that as the length of the mask increases, the dimensionality of the output is likewise increased.

Plots of C_{STM} against the feedback loss are shown in Fig. 7(a). For each encoding, the qualitative behavior is

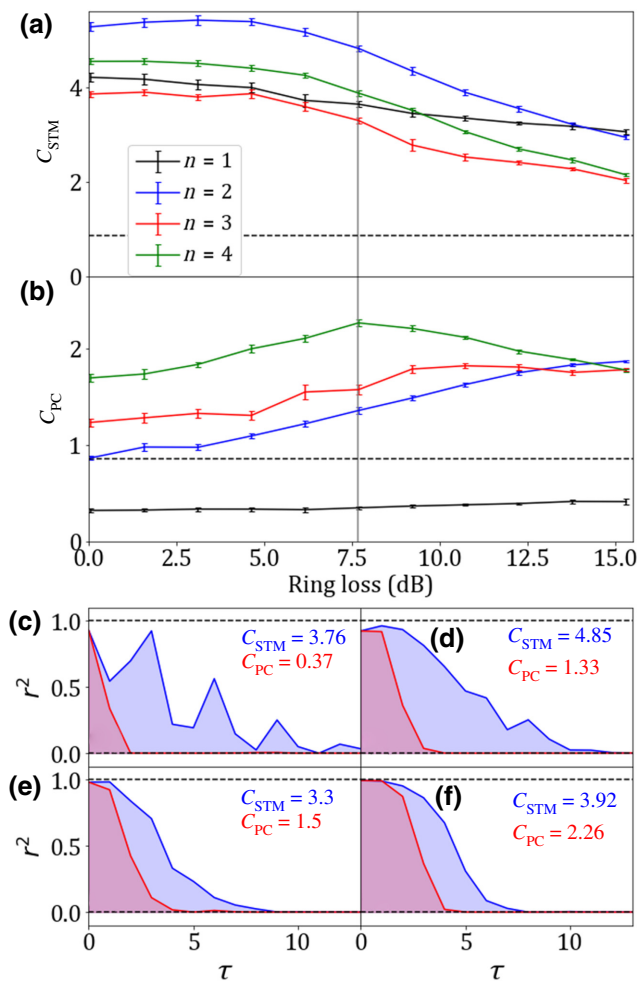


FIG. 7. (a),(b) C_{STM} and C_{PC} measured for increasing ring loss with $n = 1$ to 4. (c)–(f) Forgetting curves corresponding to a ring loss of 7.68 dB [vertical gray line in (a),(b)] for $n = 1$ to 4.

the same as that described in Sec. III A, decreasing monotonically as the fading memory is reduced. The same can be said for C_{PC} shown in Fig. 7(b). To help elucidate the behavior shown in the plots, the forgetting curves for each encoding are shown in Figs. 7(c)–7(f).

Given that three pulses circulate simultaneously in one cycle of the ring, when the input is encoded using $n = 1$, there is only mixing of every third input. This is evident in the forgetting curve in Fig. 7(c). The short-term memory recall peaks when τ is a multiple of 3 and is poor for other delays. On the other hand, the shorter input duration means that inputs are injected to the ring much faster and will persist in the ring for more input intervals before fading out.

Since θ^{int} is less than the ring round trip time, there is no influence at all of $\vec{u}(T)$ on the delay-line output. The lack of mixing between consecutive inputs significantly hinders performance on the PC task. Recall that the PC task involves a binary operation over all the inputs up to

some τ in the past, and so the lack of mixing between consecutive inputs and the lower number of virtual neurons has drastic effects on C_{PC} .

The case is different once $n = 2$. Since the circulation period is equivalent to three pulses long, each input partially overlaps with the previous input as well as with the reference signal. This prevents any periodicity in the forgetting curve and C_{STM} reaches its maximum values. Performance on the PC task is also significantly improved due to this overlap as well as double the number of virtual neurons ($N = 20$) available for computing. However, because of the partial overlap, only half of the output neurons are influenced by the preceding input.

For $n = 3$, there is a complete overlap between consecutive inputs. The PC and STM forgetting curves [Fig. 7(e)] reach almost 1 for $\tau = 0$. Combining this overlap with the larger number of virtual neurons ($N = 30$) again improves C_{PC} from the $n = 2$ case. On the other hand, C_{STM} is reduced as the inputs are injected slower relative to the ring circulation time and thus their influence does not persist for as many input intervals. Additionally, it has been shown that the memory capacity in delay-based RCs is degraded when the delay time is an integer multiple of the input time interval [13]. This is related to the interaction of different virtual neurons between time steps when the delay time and input time interval are not synchronized.

In the final case of $n = 4$, there is a partial overlap between the fourth pulse in the reference signal (current input) and the first pulse of the delayed input through the ring. This overlap significantly increases the complexity of the output. For this reason and the greater number of virtual neurons ($N = 40$), C_{PC} is significantly improved. One may also expect a significant improvement in C_{STM} also due to the desynchronization of the delay time to the input, however this improvement is partially compensated by the longer input duration and reduced input persistence in the ring (evidenced by the sharper drop off of the forgetting curve [Fig. 7(f)] compared to the previous three cases).

One can expect reductions in both C_{STM} and C_{PC} as n is further increased, due to the slower injection time relative to the signal attenuation. An input mask of $n = 4$ pulses provides the optimal compromise between fading memory and nonlinear computing capability.

-
- [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] P. R. Vlachas, J. Pathak, B. R. Hunt, T. P. Sapsis, M. Girvan, E. Ott, and P. Koumoutsakos, Backpropagation algorithms and reservoir computing in recurrent neural

- networks for the forecasting of complex spatiotemporal dynamics, *Neural Netw.* **126**, 191 (2020).
- [4] 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).
- [5] K. Nakajima, Physical reservoir computing—an introductory perspective, *Jpn. J. Appl. Phys.* **59**, 060501 (2020).
- [6] 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).
- [7] 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).
- [8] 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).
- [9] 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).
- [10] 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).
- [11] M. Riou, J. Torrejon, B. Garitainė, 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).
- [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] T. Yamaguchi, N. Akashi, S. Tsunegi, H. Kubota, K. Nakajima, and T. Taniguchi, Periodic structure of memory function in spintronics reservoir with feedback current, *Phys. Rev. Res.* **2**, 023389 (2020).
- [14] 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).
- [15] 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).
- [16] 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).
- [17] 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).
- [18] 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).
- [19] R. Nakane, G. Tanaka, and A. Hirose, Reservoir computing with spin waves excited in a garnet film, *IEEE Access* **6**, 4462 (2018).
- [20] S. Watt and M. Kostylev, Reservoir Computing Using a Spin-Wave Delay-Line Active-Ring Resonator Based on Yttrium-Iron-Garnet Film, *Phys. Rev. Appl.* **13**, 034057 (2020).
- [21] 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**, 466 (2011).
- [22] D. D. Stancil and A. Prabhakar, *Spin Waves: Theory and Applications* (Springer, New York, 2009), p.1.
- [23] J. Castera, Tunable magnetostatic surface-wave-oscillators, *IEEE Trans. Magn.* **14**, 826 (1978).
- [24] W. Ishak, 4–20 GHz magnetostatic-wave delay-line oscillator, *Electron. Lett.* **19**, 930 (1983).
- [25] Y. Fetisov, P. Kabos, and C. Patton, Active magnetostatic wave delay line, *IEEE Trans. Magn.* **34**, 259 (1998).
- [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] A. B. Ustinov and B. A. Kalinikos, Power-dependent switching of microwave signals in a ferrite-film nonlinear directional coupler, *Appl. Phys. Lett.* **89**, 172511 (2006).
- [28] A. B. Ustinov and B. A. Kalinikos, A microwave nonlinear phase shifter, *Appl. Phys. Lett.* **93**, 102504 (2008).
- [29] M. Wu, *Solid State Physics—Advances in Research and Applications*, Chap. Nonlinear Spin Waves in Magnetic Film Feedback Rings, Vol. 62 (Elsevier, San Diego, 2010), p. 163.
- [30] M. Chen, M. A. Tsankov, J. M. Nash, and C. E. Patton, Backward-volume-wave microwave-envelope solitons in yttrium iron garnet films, *Phys. Rev. B* **49**, 12773 (1994).
- [31] M. Lukoševičius, in *Neural Networks: Tricks of the Trade: Second Edition*, edited by G. Montavon, G. B. Orr, and K.-R. Müller (Springer Berlin Heidelberg, Berlin, Heidelberg, 2012), p. 659.
- [32] Y. Paquot, F. Duport, A. Smerieri, J. Dambre, B. Schrauwen, M. Haelterman, and S. Massar, Optoelectronic reservoir computing, *Sci. Rep.* **2**, 287 (2012).
- [33] D. Brunner, B. Penkovsky, B. A. Marquez, M. Jacquot, I. Fischer, and L. Larger, Tutorial: Photonic neural networks in delay systems, *J. Appl. Phys.* **124**, 152004 (2018).
- [34] J. Nakayama, K. Kanno, and A. Uchida, Laser dynamical reservoir computing with consistency: An approach of a chaos mask signal, *Opt. Express* **24**, 8679 (2016).
- [35] S. Watt, M. Kostylev, and A. B. Ustinov, Enhancing computational performance of a spin-wave reservoir computer with input synchronization, *J. Appl. Phys.* **129**, 44902 (2021).

- [36] H. Jaeger, A tutorial on training recurrent neural networks, covering BPPT, RTRL, EKF and the “echo state network” approach, *ReVision* **2002**, 1 (2005).
- [37] N. Bertschinger and T. Natschläger, Real-time computation at the edge of chaos in recurrent neural networks, *Neural Comput.* **16**, 1413 (2004).
- [38] C. S. Chang, M. Kostylev, E. Ivanov, J. Ding, and A. O. Adeyeye, The phase accumulation and antenna near field of microscopic propagating spin wave devices, *Appl. Phys. Lett.* **104**, 032408 (2014).
- [39] B. A. Kalinikos, Excitation of propagating spin waves in ferromagnetic films, *IEE Proc. H—Microw. Opt. Antennas* **127**, 4 (1980).
- [40] J. Dambre, D. Verstraeten, B. Schrauwen, and S. Massar, Information processing capacity of dynamical systems, *Sci. Rep.* **2**, 514 (2012).