Learning Trajectories from Spin-Wave Dynamics

Stuart Watt* and Mikhail Kostylev†

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

(Received 11 October 2022; revised 7 March 2023; accepted 12 May 2023; published 8 June 2023)

The efficacy of a physical reservoir computer model based on traveling spin waves in a spin-wave delay-line active-ring resonator was demonstrated recently. In the present work, we investigate how this neuromorphic device can be adapted for sensing applications. In this "reservoir computing for sensing" framework, we exploit strong coupling of the physical reservoir to its environment to utilize the reservoir as a sensing element. The dynamics of traveling spin waves in delay-line active rings are strongly dependent on the magnetic field and carrier frequency of those spin waves. Treating the spin-wave frequency as an environmental variable, we excite the active ring into different dynamical states by modulating the carrier frequency of a drive signal of microwave pulses injected into the ring. Training a linear regression on the time-multiplexed output from the ring allows the periodic amplitude patterns of the spin waves to be mapped reproducibly onto two-dimensional trajectories, representing periodic "behavioral" targets. Our work demonstrates the versatility of a magnonic resonator as a multipurpose computing and sensing device.

DOI: 10.1103/PhysRevApplied.19.064029

I. INTRODUCTION

The reservoir computer (RC) model is a type of recurrent neural network, renowned for its simplicity in construction and training [1–3]. The key component of an RC model is a nonlinear dynamical system, called the "reservoir," which performs nonlinear transformations on temporal data. The design constraints for the reservoir are quite relaxed and allow the possibility for physical systems to implement the reservoir. A physical RC is a type of unconventional computing paradigm that has received much interest in recent years for its promise of fast and energy-efficient hardware for edge-computing applications. Proposals for physical RC substrates cover a wide range of research fields and is a very open research topic [4,5].

One facet of physical reservoir computing that has not received as much attention is the so-called "reservoir computing for sensing." In 2016, Konkoli introduced the state-weaving environment-echo tracker (SWEET) algorithm, which exploits the reservoir as a sensing element when the reservoir and its environment are coupled [6]. In this system, the dynamical reservoir is driven by a preengineered drive signal, and the read-out is trained on the reservoir state to predict some environmental conditions. The SWEET concept, shown schematically in Fig. 1(a), is fairly straightforward. To make use of this framework, the reservoir must be coupled to its environment, such that the response of the reservoir to the drive signal is dependent on the environmental conditions. For physically implemented reservoirs, this condition is easily met. Using an RC for sensing does not require any fundamental changes to the RC itself, only a change in the read-out training method. In subsequent theoretical work [7], the SWEET algorithm is applied to an environment-sensitive memristor network, showing how the network is driven to different regions of its state space, depending on the environmental conditions. The reservoir state is then used to infer these environmental conditions via the trained read-out.

Here, a distinction should be made between applying an RC to process sensor information and the SWEET algorithm, which uses the RC as the sensor itself. In the former case, the RC is not an integral part of the sensor, but rather is used to postprocess the output from some distinctly separate sensor. The design and functionality of the sensor is then independent of the RC used to make the predictions of the environmental conditions. In contrast, the SWEET algorithm places the reservoir as the sensing element itself, making it an integral component of the sensing device. The RC then performs sensing and data processing in one device, at the same time.

Here, we demonstrate experimentally how the SWEET framework can be applied to the magnonic based RC concept. Recently, we demonstrated how a spin-wave active-ring resonator (SWARR) could effectively implement a reservoir and be used to store and process information encoded onto the amplitude of traveling spin waves [8].

*stuart.watt@research.uwa.edu.au
†mikhail.kostylev@uwa.edu.au

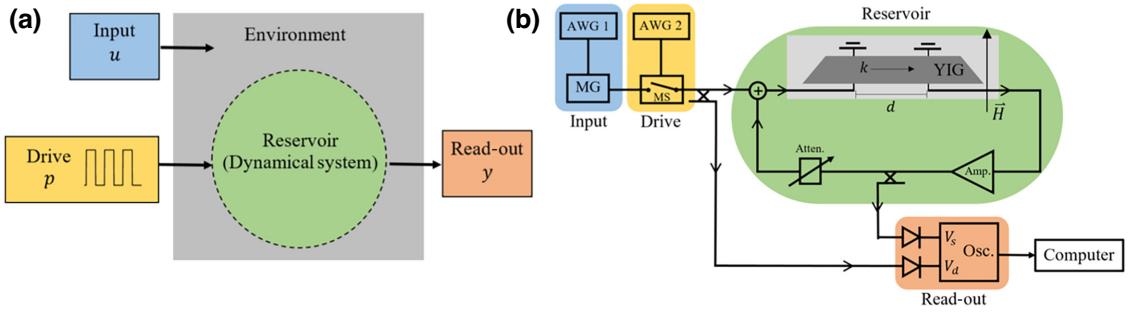

FIG. 1. (a) Schematic diagram of the SWEET algorithm. Preengineered "drive" signal maintains the reservoir in a complex dynamical state. (b) Schematic diagram of the SWARR RC configured for magnetostatic surface spin-wave transmission. Components are described in the Appendix.

That work demonstrated the efficacy of using traveling spin waves for reservoir computing by evaluating the system on some benchmark machine-learning tasks. Here, we use the same physical system for sensing rather than for computation.

The dynamics of traveling spin waves in a SWARR device are highly nonlinear, showing a host of nonlinear behavior, such as soliton formation, modulational instability, and chaos [9]. The spin-wave amplitude profile around the ring can form complex patterns, which are strongly dependent on the strength and orientation of the applied magnetic field. We propose that by implementing a SWARR as a reservoir and analyzing the complex response of the spin waves when driven by some preengineered microwave signal, we can infer changes to the reservoir's environment.

This paper is structured as follows. In the next section, we introduce the design of the SWARR-based RC. Then we discuss the principle behind how this neuromorphic device can be repurposed as a sensor of its environment. We evaluate the efficacy of such a sensing device in the following section by performing the task of mapping changes in the carrier frequency of the spin waves to periodic trajectories. This demonstration shows how the SWARR RC can be trained to provide meaningful outputs based on changes to its environment. Following this is a discussion of how such a sensor can be optimized and how errors in the mappings can be reduced by operating multiple SWARR RC devices in parallel. Finally, we discuss some future prospects for reservoir computing for sensing in magnonic reservoir computing.

II. MAGNONIC RESERVOIR COMPUTER

Spintronic and magnonic devices make promising candidates for practical physical RC applications due to their low power usage, strong nonlinearity arising from magnetization dynamics, scalability, and compatibility with existing technologies [10,11] The use of spintronic phenomena in physical reservoir computing is reported in Ref. [12], where the magnetization dynamics of a spin-torque nano-oscillator (STNO) is used to implement a single-node reservoir. This simple device satisfies all the requirements of a suitable reservoir, including nonlinear dynamics and a fading memory. There are many other studies of the computing capabilities of STNO devices aiming to improve some aspect of the physical RC. The use of consistent benchmark machine-learning tasks across studies allows the systematic evaluation of the effect of subsequent improvements to the performance of the physical RC. In addition to STNOs, RC architectures are also proposed using magnetic skyrmion memristors [13], magnetic skyrmion fabrics [14,15], dipole-coupled nanomagnets [16,17], and spin-wave interference in garnet films [18,19]. The majority of these works are carried out theoretically or through simulations, with only a few works showing experimental results [20–23]. For a recent review of spintronic based RC concepts, see Ref. [24].

In our previous work [8], we developed a physical reservoir-computer design, based on the dynamics of spin waves (the quanta of which are called "magnons") in a SWARR device. The schematic diagram of this "magnonic reservoir computer" is shown in Fig. 1(b) and described in the Appendix. The main component of the RC is the SWARR, with microwave inputs and outputs for injecting and reading data. The group velocity of spin waves in thin magnetic films of yttrium iron garnet (YIG) is orders of magnitude lower than that for electromagnetic waves, and so a strip of thin-film YIG can be placed into the path of a microwave signal to effectively introduce a time delay to the signal (hence, the name "spin-wave delay line"). In doing so, the microwave signal picks up an additional phase shift determined by the spin-wave wave number and the length of the spin-wave propagation path. The SWARR is constructed by amplifying the output signal of the spin-wave delay line and feeding it back into the input [25–27]. This resonator will have resonant modes that satisfy the condition $k_{\mathrm{res}}d = 2\pi m$, where k_{res} is the spin-wave wave number for the resonant mode, d is the length of spin-wave propagation path in the delay line, and m is an integer [28].

In addition to a time delay, the delay line inserts microwave loss into the loop. The loss is associated with natural magnetic damping of spin waves. Above a certain feedback-gain threshold, where the gain is sufficient to compensate for the delay-line losses, thermally excited magnons with wave numbers satisfying the above resonance condition will be resonantly amplified, which will trigger the development of auto-oscillations in the SWARR. We define the ring gain, G, as the difference between the delay-line losses and the feedback gain. Thus, $G = 0$ defines the threshold for auto-oscillations. The chaotic regime for the SWARR occurs for gains significantly above the threshold. For $G < 0$, losses dominate and the spin waves gradually attenuate with each circulation of the ring.

In our recent work, we demonstrated how this system could implement an RC model [8]. Briefly, information is injected into the ring by encoding onto the spin-wave amplitude and is stored along the delay line. As the spin waves circulate in the ring, spin-wave nonlinearity [29–31], dispersion, and interference form complex patterns in spin-wave amplitude. The purpose of the reservoir here is to nonlinearly map the low-dimensional input data to higher-dimensional state spaces to improve the linear separability of inputs. Our work demonstrated that the dynamics of spin waves propagating in the SWARR can effectively perform this task. Although the SWARR essentially represents a single nonlinear node with feedback, the higher dimensionality can be obtained through the time-multiplexing method [8,32]. Here, the one-dimensional output is sampled in time to create a vector of "virtual" neurons.

To fully benefit from this method, a temporal mask is applied to the input signal to increase the complexity of the reservoir dynamics. This mask converts the constant-amplitude microwave input to a train of pulses, where information is then encoded onto the amplitude of those pulses. The choice of a train of pulses is made due to the simplicity of implementation in our experiments; however, the exact shape of the temporal mask used in the time-multiplexing method is not a critical factor in the design of the RC [33]. In our work on the SWARR RC, the reservoir state is measured by sampling part of the spin-wave signal in the ring and rectifying it using a fast microwave diode. This means we measure the amplitude of the spin waves as the dynamical variable in our experiments. The temporal mask of a train of microwave pulses is used to ensure that the amplitude of the spin waves traveling in the ring forms complex patterns, which vary across each output.

III. MAGNONIC RESERVOIR FOR SENSING

In the present work, we propose that the SWARR RC described above can be adapted as a sensor. In a physical RC device, it is important that the parameters of the reservoir remain fixed and only the input is modulated, allowing reproducibility of the mapping from input to output. Converting the physical RC from a computational device into a sensing device involves reversing the above condition, keeping the input consistent and modulating the parameters of the reservoir. To do so, we replace the varying-amplitude microwave input of the RC with a constant-amplitude sequence of pulses, representing the preengineered driving signal called the "drive" in Fig. 1(a), the purpose of which is to maintain the internal state of the reservoir in some complex dynamical regime. When the reservoir dynamics are strongly coupled to the environment, changes in certain environmental conditions result in shifts of the dynamical state of the reservoir. A linear regression can then be trained on the SWARR output to make predictions about those environmental conditions, instead of the variations of the input microwave signal in the case of neuromorphic computing.

The dynamics of spin waves in the SWARR are dependent on a variety of environmental conditions, such as temperature, magnetic field strength, and magnetic field orientation. One important peculiarity of spin waves in magnetic films is that their frequency depends on the magnetic field applied to the film [see Eq. (A1) in the Appendix]. To demonstrate the use of this system for RC sensing, we set the input microwave carrier frequency as the "environmental variable" to be modulated and sensed, which is functionally equivalent to sensing the magnetic field environment of the delay line. Since the SWARR operates with a circulation period on the order of hundreds of nanoseconds, it is practically easier to modulate the frequency, which can be modulated on the same very fast scale (on the order of tens of nanoseconds), rather than relying on a much slower rate of modulation of the magnetic field achievable with electromagnets (typically milliseconds or seconds).

The condition for resonance in the SWARR, $k_{\mathrm{res}}d = 2\pi m$, is satisfied for a number of frequencies, and we choose to modulate the input frequency between adjacent resonant modes. Thus, for each different value of the carrier frequency, the circulating spin waves will gather different amounts of phase with each circulation of the ring and will interfere differently with the incoming microwaves at the SWARR input. We operate the ring *below* the auto-oscillation threshold ($G < 0$). Under these conditions, a spin-wave signal injected into the ring can circulate many times before fading away. As a result, the amplitude pattern of spin waves in the ring becomes complex and strongly dependent on the frequency of the input.

To ensure maximum complexity of the spin-wave amplitude in the SWARR, a preengineered amplitude mask is applied to the input microwave signal. We use a similar masking procedure to that in Ref. [8] to create a repeating sequence of random binary microwave pulses with a

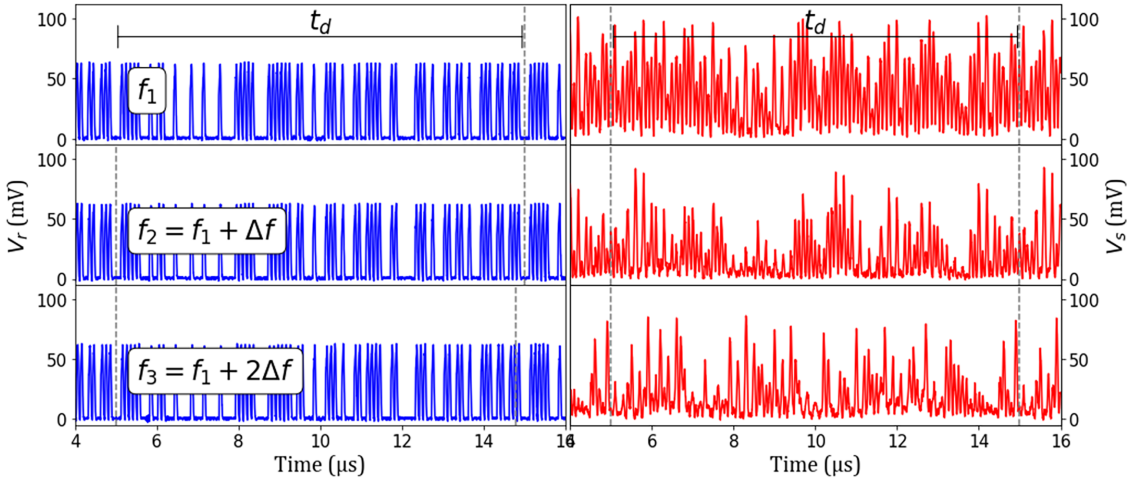

FIG. 2. Examples of the drive signal (left column) and output (right column) of the SWARR RC, measured using a microwave diode. For these measurements, the drive signal is a random sequence of $P = 100$ binary pulses, which repeat. Here, the carrier frequency is $f_1 = 7097$ MHz and $\Delta f = 1$ MHz.

repetition period of t_d. The amplitudes of the pulses in each period of the sequence are determined by a mask of length P, where P is the number of binary pulses in the sequence. Construction of the drive signal is described in the Appendix.

The left-hand column of Fig. 2 shows examples of the input drive sequence measured using a microwave diode. The microwave diode is a passive broadband device, which returns the slowly varying amplitude of a microwave signal passing through it. Using such a device allows for the simple detection of microwave signals for which it is not necessary to know the carrier frequency. The diode voltage of the input drive signal, as measured before injection into the SWARR, is denoted as V_d. Modulating the carrier frequency of these pulses has a negligible effect on their shape and amplitude. This confirms that the microwave carrier frequency itself does not represent the "preengineered driving signal" [or the drive in Fig. 1(a)]. The low-frequency binary-pulse sequence acts as the drive signal in our case. Hence, the microwave carrier frequency represents an environmental condition rather than the drive. Similarly, the SWARR output is passed through another microwave diode, removing the carrier frequency from the signal and resulting in a low-frequency signal, denoted as V_s.

Below the auto-oscillation threshold ($G < 0$), the SWARR dynamics are strongly coupled to the driving signal and strongly dependent on the carrier frequency of the input driving signal. The right-hand column of Fig. 2 shows the diode voltage of the SWARR output signal corresponding to the inputs on the left-hand panel. It shows that different carrier frequencies of the drive signal change the SWARR amplitude pattern significantly. In the following section, we demonstrate how the changing reservoir output can be mapped to different targets, effectively representing a "sensing" of the carrier frequency, and hence, the magnetic environment.

It is important to note that these patterns are periodic, with the same repetition period, t_d, as the drive signal, *regardless* of the circulation period of the SWARR. This is a very useful property, as it allows the user to synchronize the drive signal to the output of the SWARR device, regardless of other environmental conditions that may influence the SWARR circulation period (such as changes to magnetic field strength or field orientation). We further elaborate on this idea in the Sec. V where we drive multiple SWARR devices simultaneously, each with a different circulation period. More importantly for this work, the stable periodicity of the SWARR output can also be mapped to periodic trajectories.

Above the auto-oscillation threshold, the onset of a self-generated microwave signal and more complex behavior (such as chaos) weakens the coupling of the SWARR to the drive signal, and the SWARR output loses its reproducibility and periodicity. This makes repeatable mapping to learned trajectories impossible. For these reasons, the SWARR is operated below the auto-oscillation threshold.

IV. MAPPING RESERVOIR DYNAMICS TO PERIODIC TRAJECTORIES

Here, we take the concept of sensing a step further. Somewhat different to the SWEET algorithm, which uses the reservoir-environment coupling to provide an estimate for the environment variables themselves (for example, the magnetic field strength), we instead map the complex dynamics of the SWARR to periodic trajectories in two-dimensional phase space. This allows the system to replicate certain output "behaviors" rather than individual states. Inspiration for this work comes from Ref. [34],

which theoretically shows how the dynamics of a chaotic reservoir can be mapped to quasiattractors. We emulate this idea experimentally using a physically implemented reservoir. In contrast to Ref. [34], however, we do not operate the SWARR in its chaotic regime in this proof-of-concept work. The main reason for this is the impossibility of measurement reproducibility in the chaotic regime.

As we demonstrate in the previous section, by driving the SWARR with a periodic drive signal, the pattern of spin-wave amplitude is also periodic with the same repetition rate. The periodicity of these amplitude patterns leads naturally to mapping to periodic target trajectories, rather than static targets. In this work, we choose two-dimensional Lissajous figures as the periodic trajectories to map to because of their convenience in visualization; however, this idea can easily be applied to arbitrary patterns in higher dimensions. Using this method ensures that the sensor output remains bounded in phase space over the sensing-time duration.

One more advantage of periodic trajectories is as follows. The strong dependence of the spin-wave dynamics in the ring to changes in the environment causes the spin-wave amplitude pattern to change significantly when the environment condition is altered. We employ a time-delay sampling method to create the reservoir state vector at each time step and a linear read-out to map these state vectors to coordinates along the target trajectories. If trajectories are periodic, the output will not need synchronization to the sensor, as the linear read-out will always be able to pick up the trajectory at some arbitrary point along its path upon changes to the environment.

To test the efficacy of the SWARR as a frequency sensor, we give the system the task of mapping scalar frequency inputs to the periodic two-dimensional trajectories. The number of unique trajectories that the read-out is trained to recreate is defined by M, and each trajectory is discretized to a length of L points. The coordinates for each target trajectory are defined as $\hat{s}_i^x$ and $\hat{s}_i^y$ and are constructed from sinusoids with an amplitude of one but varying angular frequency. Note that the actual scale of the Lissajous curves does not matter for the purposes of this experiment. Here, the subscript index $1 \leq i \leq L$ indicates the position along the trajectory. Figure 3(a) shows the coordinates of the unique trajectories set as targets in this work. Here, we use up to a maximum of $M = 6$ different input frequencies to map to six unique trajectories.

For each frequency, the drive signal is applied to the reservoir for a duration equal to $50 \times t_d$ before switching to the next carrier frequency, creating a training set of duration $300 \times t_d$. This process is repeated to construct the testing set, using a duration of $25 \times t_d$ for each frequency interval. The entire testing set then has a duration of $150 \times t_d$. The testing data are measured immediately after the training set, so the complete measured trace has a temporal length of $450 \times t_d$.

The entire series is injected into the SWARR and a fast oscilloscope records the system output. The output of the SWARR at any point in time is just a one-dimensional scalar-amplitude time series. To create a high-dimensional representation of the output, we embed the time series into a higher-dimensional phase space using delayed values of the output as the additional coordinates. This method is similar to time multiplexing used in time-delay RC [32]. The number of embedding dimensions is defined as N, which also determines the number of virtual neurons in the RC model. The SWARR output, V_s, at some time T, is embedded as a N-dimensional vector defined as

$$\vec{x}(T) = (V_s(T), V_s(T - \theta_N), \ldots, V_s(T - (N-1)\theta_N)), \tag{1}$$

where θ_N is the embedding delay. This is equivalent to sampling the SWARR output over a window of time equal to $N\theta_N$, creating the reservoir state column vector $\vec{x}$. The embedding delay and embedding dimension are chosen such that $N\theta_N = t_d$.

The system read-out is performed in the same manner as that for a standard RC application. A linear regression of the reservoir states is used to reconstruct the RC output as

$$\vec{y}_{\text{out}} = \mathbf{W}^{\text{out}}\vec{x}. \tag{2}$$

$\vec{y}_{\text{out}}$ is a two-dimensional column vector containing the predicted coordinates of the trajectory, and $\mathbf{W}^{\text{out}}$ is the read-out matrix with dimensions $2 \times N$. This process of sampling the SWARR output voltage, V_s, and taking a linear combination of the states is shown schematically in Fig. 3(b). Successive outputs are obtained by shifting the sampling window forward in time by a step size of $\theta_L = t_d/L$ and repeating the sampling process. This "sliding-window" sampling process results in L output vectors for each period of the drive signal, with each output being mapped to a point along the target trajectory. In this way, the periodic SWARR output will be mapped to a periodic target. An example is shown in Fig. 3(c), which shows how each window is mapped sequentially to a different coordinate along a trajectory. After L sample windows, the SWARR output pattern repeats and the mapping begins at the start of the trajectory again.

The purpose of the training procedure is to determine the optimal $\mathbf{W}^{\text{out}}$ that maps the reservoir dynamics to the target trajectories to minimize the error between the predictions and targets:

$$\mathbf{W}^{\text{out}}\vec{x} = \vec{y}_{\text{out}} \approx \hat{y}_{\text{out}}. \tag{3}$$

Here, we note that quantities denoted with a circumflex represent the task targets (the actual trajectories), while those without a circumflex represent the predictions of those targets.

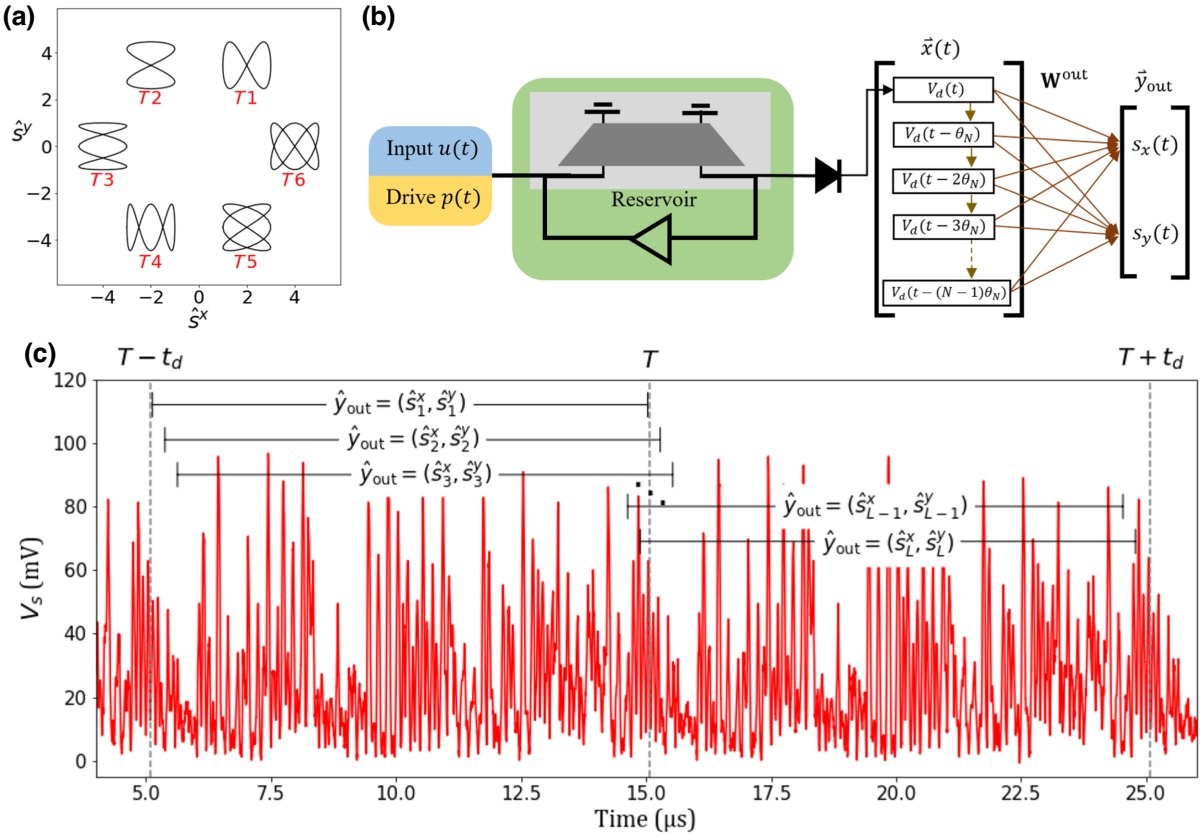

FIG. 3. (a) $M = 6$ target trajectories. (b) Schematic showing how V_s is mapped to the task trajectories. (c) Example of the sliding-window sampling procedure. Each step of the sample window forward in time creates a new reservoir state vector, which is mapped to the next point on the target trajectory.

From each oscilloscope measurement, samples from the training set are selected to train the weighted model. For each frequency interval, the first ten periods of the driving signal are dropped to remove any transient behavior when the spin-wave dynamics transition from one regime to another. The SWARR output takes approximately one period of the drive signal to transition to a new output pattern, so dropping the first ten periods of each interval ensures that these transition regions are clearly left out of the training set. This way, the read-out is only trained on regions that are stable and periodic. This is equivalent to the "washout" region in conventional physical RC applications, where an initial period of reservoir dynamics is disregarded so that any influence of initial conditions is "washed out" of the system. This washing-out process can only occur if the memory in the system is "fading;" this is another reason why the SWARR is operated below the auto-oscillation threshold.

Training the linear read-out is a straightforward process of matrix multiplication and can be performed in one shot (as described in the Appendix). To test the read-out, a matrix of all testing instances is constructed in the same manner as that for the training procedure. In each frequency interval, the first 5 periods of the drive signal are dropped to remove the transition region, keeping the remaining 20 periods for testing the read-out. The testing matrix is then premultiplied by the read-out matrix, returning the predicted coordinates for each instance. To determine how well the SWARR RC can reconstruct the target trajectories, we utilize the root-mean-square error (RMSE) as the metric of accuracy in this work. Each point along the trajectory can be considered a unique target. We calculate the RMSE between the target trajectory points, $\hat{y}_{\text{out}} = (\hat{s}_i^x,\ \hat{s}_i^y)$, and the reconstructed trajectory points, $\vec{y}_{\text{out}} = \mathbf{W}^{\text{out}}\vec{x} = (s_i^x,\ s_i^y)$, as

$$E = \sqrt{\sum((s_i^x - \hat{s}_i^x)^2 + (s_i^y - \hat{s}_i^y)^2)/(20ML)}. \quad (4)$$

The sum above is over all $20 \times M \times L$ instances in the testing set. During the transition region between frequency intervals, the read-out maps the SWARR output to some unlearned coordinates, which contribute significantly to error. This is the reason that these regions are dropped for the evaluation of the system.

The drive signal in this work is a random binary sequence of $P = 100$ bits. Each pulse has a width of 100 ns and, as such, the period of the drive signal is

$t_d = 10\,\mu$s. The spin-wave group velocity is determined predominantly by the magnetic field strength applied to the YIG film and the separation of the spin-wave antennas, for a given frequency range. With a magnetic field of $H = 1730$ Oe, the SWARR circulation period (i.e., the time for a spin wave to travel around the ring exactly once) is 196 ns for a frequency range centered on 7097 MHz. This value is determined experimentally by measuring the period of a single pulse injected into the SWARR and left to circulate multiple times.

The microwave carrier frequency is modulated between the values $f = f_1$ and $f = f_1 + M\Delta f$, where $f_1 =$ 7097 MHz and $\Delta f = 1$ MHz. All carrier frequencies in this work reside in the transmission band of the delay line. For each series, the diode voltage of both the drive signal, V_d, and the SWARR output, V_s, are recorded. A separate $\mathbf{W}^{\text{out}}$ is trained on either case to demonstrate the improved trajectory reconstruction when utilizing the SWARR RC. Examples of the predictions are shown in Figs. 4(a) and 4(b) when the read-out is trained to learn M different number of trajectories.

When $M = 1$, the reconstructed trajectory matches very well to the target for both the drive and SWARR signals. However, when there is more than one trajectory to learn, the complexity of the SWARR dynamics is required to separate the different frequency inputs. The plots in Fig. 4(c) compare the reconstruction of $M = 3$ targets in the test set when the read-out is trained on V_d and V_s. The gray line in these plots shows how the read-out behaves in the transition region between frequency intervals. As the transition occurs, and the output pattern evolves from one stable pattern to another, the read-out is mapped to some unlearned coordinates before settling to a learned trajectory. These plots clearly demonstrate how the complexity brought about by the interference of spin waves in the SWARR makes the reconstruction of the target trajectories possible.

The plots in Figs. 4(d)–4(f) show how the error in the prediction depends on the number of virtual neurons sampled for the read-out, the length and number of unique trajectories, and the gain of the SWARR, respectively. For these measurements, ten oscilloscope traces are recorded for each set of parameters and postprocessed independently. The RMSE values presented here are averages over the ten traces and the error bars give the standard deviation.

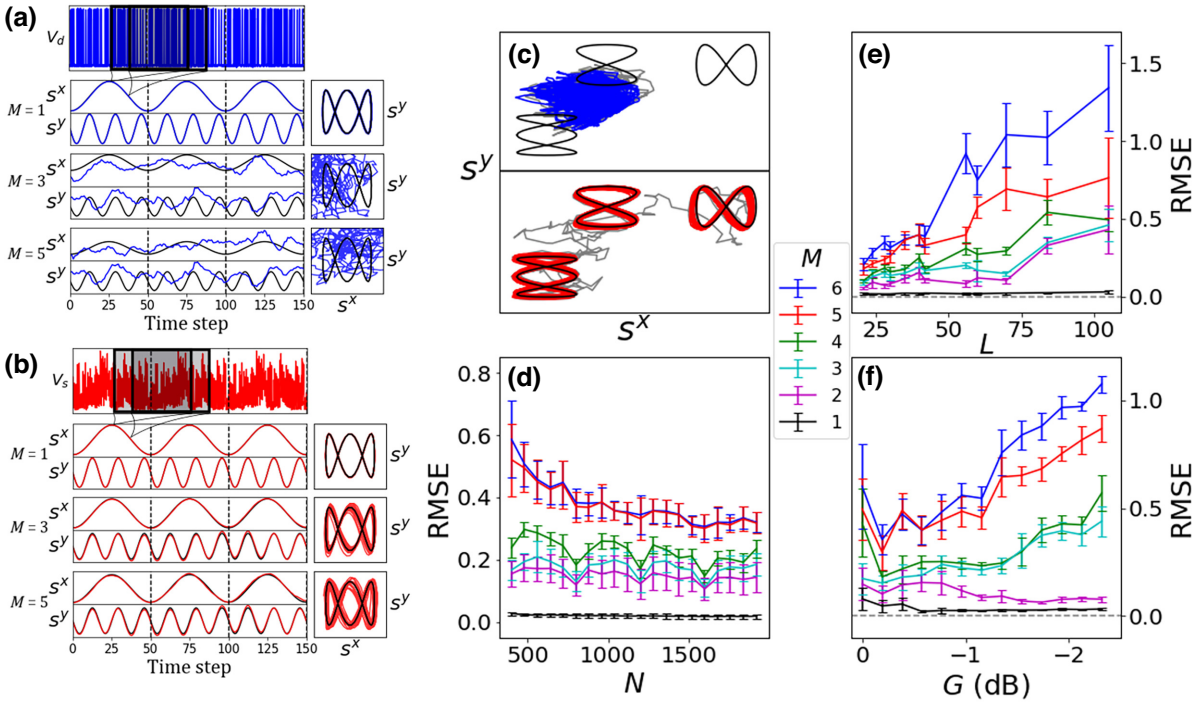

FIG. 4. Comparing prediction against target for the $T4$ trajectory. (a) Top panel displays the neuron values sampled from the V_d trace and how the sliding-sample window corresponds to each point on the predicted trajectory. Bottom three panels compare the predicted trajectory to the target trajectory when the read-out is trained to learn $M = 1$, 3, and 5 trajectories with $(L,\ N) = (42,\ 1260)$. Panels to the right show the predicted trajectories across all testing instances of the $T4$ trajectory. (b) Same as (a), but the read-out is trained on V_s. (c) Examples of test-set mapping for $M = 3$. Top (bottom) panel shows the case when the read-out is trained on $V_d(V_s)$. Gray lines indicate the transition regions between different frequency intervals. (d) Prediction RMSE versus the number of virtual neurons, N, for a gain of -0.58 dB and $L = 40$. (e) Prediction RMSE versus trajectory length, L, for a gain of -0.58 dB and $N = 840$. (f) Prediction RMSE versus the gain, G, for $N = 840$ and $L = 40$.

The trends shown in Figs. 4(d) and 4(e) are intuitive. Each coordinate along each of the trajectories can be considered as a distinct target, meaning that there is a total of LM targets for the system to learn. Increasing either L or M reduces the overall mapping performance, as shown in Fig. 4(e). This can be offset by increasing the virtual-neuron sampling rate [Fig. 4(d)]. The higher the dimension of the reservoir, the higher the linear separability of different targets, and the resulting RMSE decreases. However, improvement of the RMSE saturates as N is further increased, since the samples become close enough that no additional variation from the output is captured.

How the RMSE depends on the gain of the SWARR is also intuitive. Close to the threshold of auto-oscillation, the system is unstable and fluctuations in the output cause significant changes between training and testing sets, significantly hindering the prediction quality. As the gain is decreased slightly, the SWARR signal stabilizes and the RMSE drops. Each spin-wave pulse will circulate multiple times in the ring and the output pattern is at its most complex while maintaining stability and synchronization to the driving signal. The RMSE then reaches its minimum. As the gain is further decreased, the fading memory of the system decreases, and the output pattern becomes less complex, as there is less pulse interference and weaker nonlinearity of the spin-wave dynamics. In the case where there is no feedback at all, the SWARR output pattern approaches that of the drive signal, with only small differences in the pulse amplitude due to different delay-line losses associated with different carrier frequencies. However, these small differences are not sufficient to result in separation of the different frequencies.

V. REDUCING MAPPING ERROR

In this section, we discuss some considerations for improving the mapping performance of the SWARR system, some applications for this system, and some avenues for further research. One limitation of the current idea is that there is an upper limit to the number of virtual neurons one can sample from the output before no new information is being added to the reservoir state vector. The use of pulses to construct the driving signal is based on the idea that the spin-wave nonlinearity is amplitude dependent. Narrow pulses provide the largest variation in the driving-signal amplitude to extract the most nonlinearity of the SWARR response. However, the sampling of virtual neurons is not restricted by the number of pulses. Here, we employ $N > P$, meaning that multiple virtual neurons are sampled for each pulse of the drive signal. Such a drive results in a SWARR output that varies in amplitude on the order of tens of nanoseconds. With virtual neurons spaced at nanosecond or sub-nanosecond intervals, the variation between adjacent neurons is too small to contribute additional value to the linear read-out.

One solution to this is to simply extend the length of the drive signal by adding additional pulses to the sequence. Increasing the number of binary pulses, P, increases the amount of potential dynamics that the read-out has access to. Using a constant $N = 600$ and $L = 40$, we compare different drive signal lengths in Fig. 5(a). As the number of target trajectories to learn increases, more information is required to linearly separate the different coordinates. Our results indicate that simply increasing P can lead to improved performance (of course, at the expense of increasing sensing time). The key here is that the drive signal is "preengineered" to bring the SWARR into a complex dynamical regime. In this work, we implement a binary sequence of pulses as the drive; however, a more complex drive signal may induce more-complex reservoir dynamics. Designing the drive signal to optimize the performance is likely a task-specific operation and is beyond the scope of this work.

An alternative solution is to drive multiple SWARR devices in parallel, boosting the dimensionality of the entire system by combining the outputs of each SWARR to the reservoir state vector. To demonstrate this, we implement three separate SWARR devices that differ only in their antenna separation, which dictates the length, d, of the path of spin-wave signal propagation in a YIG film, as shown in Fig. 5(b). In a laboratory setting, it is faster and easier to emulate a system of three separate delay lines instead of building it as hardware. This can be done by measuring a single delay line sequentially for three different antenna separations and combining the outputs in the postprocessing stage. This approach emulates three SWARRs connected in parallel and fed from the same input-signal source with the help of a microwave-power divider that splits the input-drive-signal power equally between the SWARRs. Importantly, the three effective SWARR devices are not coupled further down the microwave path via an analogue channel and their outputs are independent from each other. The outputs are combined first in digital form after sampling the output signals. Together with the simplicity of accurate emulation of the action of the microwave power divider, this system design ensures that the emulation delivers valid results.

The three SWARR devices are each driven by the same drive signal. Since the spin-wave amplitude pattern is synchronized to the drive signal and *not* to the circulation period of the spin waves, each SWARR is then synchronized with each other. This is an important point, as it would allow many SWARR devices to operate in parallel with the same drive signal and each with its own parameters. On the other hand, the different circulation periods due to different antenna separations allows each SWARR device to have completely different spin-wave amplitude patterns (but with the same period). So, each device can contribute new information to the read-out. The read-out itself is constructed in the same manner as that

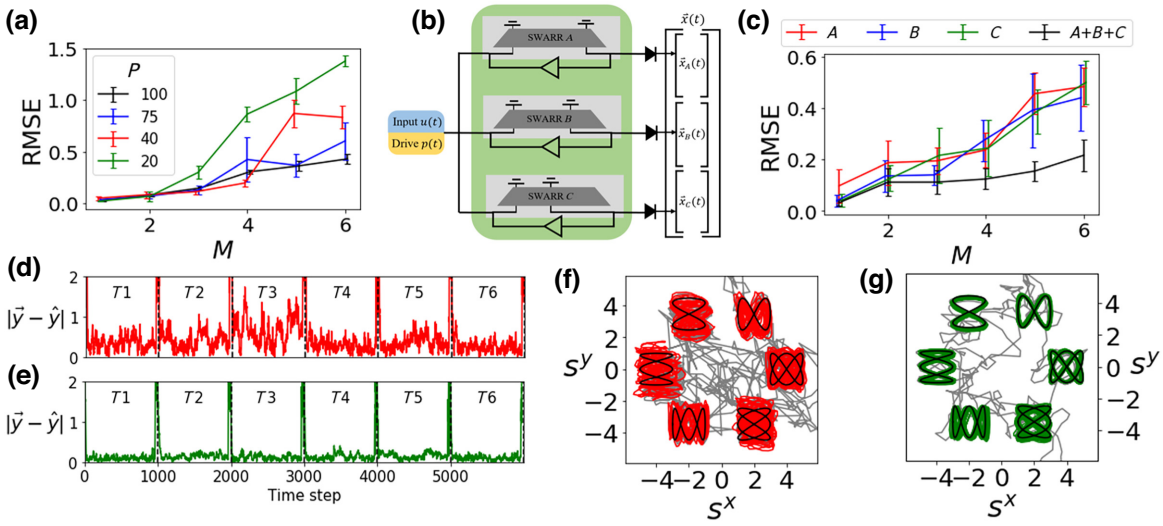

FIG. 5. (a) RMSE against M for different lengths of random binary-pulse drive signal ($P = 100$, 75, 40, and 25). Here, $N = 600$ and $L = 40$. (b) Schematic showing dimensionality boosting by driving three reservoirs simultaneously. Each SWARR differs only in the antenna separation ($d = 6.8$, 5.4, and 4.5 mm). (c) RMSE against M for each individual SWARR RC (labeled "A", "B," and "C") and the combined SWARR RC (labeled "$A+B+C$"). In all four cases, $N = 600$ (for the combined case, each SWARR contributes 200 neurons to the read-out, totaling 600) and $L = 40$. (d) Root-squared-error of the predictions on the testing set for a single SWARR RC (labeled A). Vertical dashed lines indicate the transition between frequency intervals (with $L = 40$ and 25 periods of the drive signal in each frequency interval; each interval corresponds to 1000 coordinates). (e) Same as for (d) but for the combined SWARR RC (labeled $A+B+C$). (f) Predictions versus targets for the single SWARR RC (labeled A). Gray lines indicate the transition regions between different frequency intervals. (g) Same as for (f) but for the combined SWARR RC (labeled $A+B+C$).

in the previous section, with the reservoir state vectors of each individual SWARR concatenated together.

To test this idea, we compare the single SWARR RC to the combined SWARR RC. Figure 5(c) shows how the RMSE grows with the number of learned trajectories, M, for each individual SWARR and for the combined case. To maintain consistency, each individual SWARR output is sampled using $N = 600$ and $L = 40$. For the combined case, each SWARR contributes 200 virtual neurons to the read-out, totaling $N = 600$. The combined case performs significantly better than the individual SWARR devices. So, even though each individual SWARR contributes less information, in total, there is more complexity in the system when each SWARR contributes to the read-out. The plots in Figs. 5(d) and 5(e) show the Euclidean distance between the predicted trajectory and the target across the entire testing set for the single and combined SWARR RC, respectively. Again, it is clear that the reconstruction of the targets is significantly improved in the combined SWARR case. The sharp peaks in these plots correspond to the transition regions between different microwave carrier frequencies, showing how the system output stabilizes within one period of the drive signal. The plots in Figs. 5(f) and 5(g) are visual aids to compare the predictive capabilities of the two cases when $M = 6$. The gray lines in these plots show how the read-out maps the SWARR output when the microwave carrier frequency transitions to a new state. It is clear from Fig. 5(g) that the mapped output stabilizes to a learned trajectory much faster for the combined SWARR case than for the single SWARR case.

VI. FUTURE PROSPECTS

We now discuss some possible extensions of this concept. What we have proposed is a simple system that performs sensing and data processing in one device. In addition, since the linear read-out is computationally inexpensive, we expect that the mapping of the SWARR output to learned trajectories could be performed in real time. A typical use of a machine-learning model for sensing involves applying that model to the output of some sensor to process the readings. The present system can perform sensing and processing at the same time and the linear read-out provides the reading straightaway. This effectively cuts out a step from the sensing process.

In a sufficiently complex system (such as having multiple SWARR devices operating in parallel), the sensor could effectively multitask. One of the benefits of a reservoir-computer model is that the reservoir itself is never trained, only the read-out. This means one reservoir could be applied to a wide range of machine-learning tasks simply by changing the linear read-out. In our previous works [8,35], we make use of this idea to perform two benchmark machine-learning tasks simultaneously on the same physical device. We propose that this idea could

be applied to the "reservoir computer for sensing" concept, whereby the same reservoir is used to sense any of the environmental variables in which the reservoir is coupled to. A different read-out could be trained to predict each variable. If the dynamics were sufficiently complex and very high dimensional, the system could sense a wide range of parameters *simultaneously*, and thus, implement a general-purpose sensor. In the present case of the SWARR RC, one can think of simultaneously measuring environmental factors, such as microwave carrier frequency, magnetic field strength, and temperature, all of which contribute to the spin-wave dynamics in the ring.

The conversion from reservoir dynamics to read-out is done through training, not any specific rules-based approach. Thus, the device requires calibration. Each new SWARR device would require a separate calibration to account for manufacturing variations. However, the training is very fast and always stable. In addition, since the read-out is performed external to the reservoir, the calibration could be performed many times to adjust for new unseen conditions. An important consideration is how the SWARR dynamics are mapped when the system is exposed to environmental conditions it has not been trained on. In this work, we use a maximum of six distinct carrier frequencies; however, in real-world applications, environmental conditions are continuous, and it is important not to introduce unwanted behaviors that result when the system enters a dynamical regime that it has not been trained on. Figure 6(a) shows how the SWARR dynamics are mapped when the microwave carrier frequency is set to intermediate values between "learned" frequencies. In other words, carrier frequencies that the read-out has not been trained on.

In this experiment, the read-out is trained on two output patterns, produced from two different microwave carrier frequencies with a difference of 1 MHz, to recreate two trajectories. The system is then tested on a series of intermediate frequencies in between. What we observe is that the predicted trajectory evolves across the output space from one trajectory to another as the frequency sweeps between the two "learned" frequencies. With only two trajectories to learn, the evolution from one to another is fairly straightforward. The addition of another trajectory results in a more complex evolution of the predicted output [as shown in Fig. 6(b) for $M = 3$]. An important feature that these results illustrate is that the SWARR amplitude pattern varies smoothly as the input carrier frequency is modulated. The implication of these results is that the system can somewhat interpolate between trajectories when the environmental conditions are unknown.

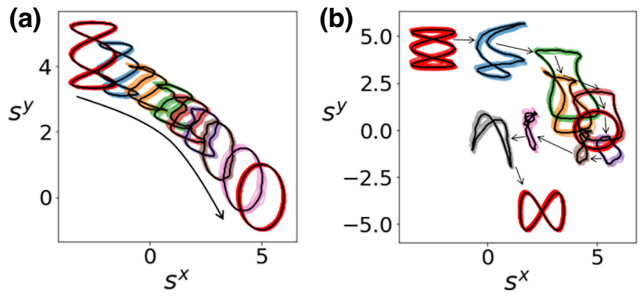

FIG. 6. (a) Read-out is trained to recreate $M = 2$ trajectories. "Learned" frequencies have a difference of 1 MHz. Arrow indicates direction of successive frequency steps. (b) Same as for (a) but for $M = 3$.

The SWARR RC concept is still very much in its infancy. While the device presented in Ref. [8] meets the basic requirements of a reservoir computer, its performance on benchmark tasks falls below that of some other spintronic concepts (for example, Ref. [36]). Further work is required to determine how its performance might be improved and how the nonlinearity of spin waves can be further exploited in the device for computing. One significant limitation of the SWARR RC, in terms of computation, is the relatively large size and slow operation of the experimental setup. For a physical RC to be of any practical use as a computing device, it needs to be fast, low-power, small, and robust. Creating such a physical device, which outperforms the modern computer chip, with its gigahertz operating frequency, nanometer-scale components, and applicability to a vast variety of tasks, is a significant challenge. On the other hand, the physical RC may not face the same limitations in sensing applications. For one, sensing occurs on much slower timescales than computation, relaxing the high-speed requirements of a reservoir. The RC is naturally integrated into its environment, performing sensing and computation in the same component, in real time. This has the potential to reduce the complexity, power usage, and sensing time of the device.

Lastly, it should be noted that the general concepts we demonstrate here are not restricted to the SWARR RC alone. Physical reservoir computing has become a popular field of research in recent years. Since the first demonstrations of a physically implemented reservoir [37], the field has grown across a wide variety of research disciplines, with reservoirs implemented on optical, mechanical, quantum mechanical, biological, electronic, and spintronic substrates [4], among others. Many of these reservoir substrates are strongly coupled to their environment. This is especially true in the case of spintronic and magnonic reservoirs. A recent review [24] covers a range of "nanomagnetic devices" that have so far been proposed as reservoir-computer substrates. The designs vary between spin-torque oscillators, skyrmion textures, domain-wall structures, magnonic systems, and others. One thing they all have in common is they are all sensitive to the magnetic field in which they operate. Other variables one must also consider when experimenting with

these devices are temperature, atmospheric conditions, or electrical signals.

The SWEET method [6] is a general approach to reservoir computing for sensing and does not impose any restrictions on the design of the reservoir itself. Apart from the sensitivity of the spin-wave dynamics in the SWARR to the external magnetic field, there is no aspect of the SWARR itself that makes it more applicable to implementing the SWEET method any more than the other spintronic RC devices.

In the seminal work of Ref. [12], for example, a simple RC was constructed from a single STNO. This simple RC places a STNO in a static magnetic field, and inputs data by modulating the voltage applied to the STNO along with a dc input current, which drives the oscillation. The oscillation amplitude is read as the reservoir output and used to reconstruct the targets. Both the oscillation amplitude and frequency are dependent on the strength of the magnetic field applied to the STNO. The SWEET method could be applied to this device simply by modulating the magnetic field, instead of the input voltage, to change the oscillator dynamics. Then the oscillator amplitude could be mapped to the magnetic field strength, or some derivation of it. In this example, the STNO RC may become a form of sensor of magnetic field fluctuation, similar to our present work.

VII. CONCLUSION

We demonstrate how a reservoir computer built on a substrate consisting of a spin-wave delay-line active-ring resonator can be used as a sensing element by exploiting the sensitivity of spin-wave dynamics to environmental conditions. We employ the SWEET algorithm to the resonator and show how the spin-wave amplitude at the output can be linearly mapped to different targets, depending on the input microwave frequency. The period of the spin-wave amplitude patterns in the active-ring resonator is highly coupled to the period of the driving signal, allowing us to map the output to periodic trajectories in phase space, regardless of the input microwave frequency of the driving signal. This effectively represents a frequency and magnetic field sensor.

ACKNOWLEDGMENTS

The work of S.W. is supported by the Australian Government Research Training Program.

M.K. came up with the general idea that a SWARR could be an efficient hardware platform for implementing a reservoir computer. S.W. designed the concept of the SWARR RC and realized that a good way to experimentally demonstrate its usage for sensing was by modulating the microwave frequency and mapping the frequency-dependent output of a SWARR RC to a trajectory. He carried out the experiments and drafted the paper. The draft was edited jointly by S.W. and M.K.

APPENDIX: METHODS

1. DESCRIPTION OF SETUP

The experimental setup is shown schematically in Fig. 1(b). The SWARR (with a green background) consists of a spin-wave delay line (shaded rectangle) with amplified feedback. The delay line is assembled from a 2-mm-wide 5.7-μm-thick YIG film (dark-gray trapezoid), sitting on top of two 50-μm-wide short-circuited microstrip transducers ("spin-wave antennas"). These antennas are spaced $d = 6.8$ mm apart and are used to excite and detect spin waves.

The spin-wave antennas excite spin waves when a microwave current flowing through them induces a microwave magnetic field localized about the antenna, which drives precession of the magnetic moment under the input antenna (left-hand side of the delay line). Due to the dipole-dipole and exchange interactions between neighboring magnetic moments, the precessional energy will be carried away from the antenna as a spin wave. The magnetic dipole field produced by the precessing moments will then induce an electromotive force in the output antenna (right-hand side of the delay line) and the energy will carry on as a microwave signal.

A broadband solid-state microwave amplifier ("Amp") provides a constant 40-dB amplification of the delay line output. A variable attenuator ("Atten.") controls the feedback gain. A −10-dB directional coupler samples the amplified signal and a tunnel diode rectifies the sampled microwave signal to a dc voltage, which is measured and saved using a fast oscilloscope ("Osc"). The oscilloscope data are processed on an external computer (i.e., offline). The SWARR allows for injection of an external microwave signal via a microwave combiner (plus symbol).

The SWARR is configured to allow the excitation of magnetostatic surface spin waves (MSSWs), the angular frequency of which is determined by the dispersion relationship [38]:

$$\omega(H,k) = \gamma\sqrt{H(H+4\pi M_s) + \frac{(4\pi M_s)^2}{4}(1-e^{-2kw})}. \tag{A1}$$

Here, k is the spin-wave wave number, $|\gamma|/2\pi = 2.8$ MHz/Oe is the gyromagnetic ratio, $4\pi M_s = 1750$ Oe is the saturation magnetization for YIG, and $w = 5.7\,\mu$m is the YIG film thickness. Importantly, the frequency depends on a magnetic field, H, applied to the sample. The choice of MSSW in this work has three benefits. First, MSSWs are efficiently excited by microstrip spin-wave antennas. Second, the MSSWs are unidirectional and will only be excited efficiently in one direction, preventing energy loss 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

magnetic losses, allowing long propagation times at room temperature.

2. GENERATION OF THE DRIVE SIGNAL

A microwave source [labeled "MG" in Fig. 1(b)] produces a constant-amplitude microwave signal, the frequency of which is modulated using an arbitrary waveform generator ("AWG 1"). A second arbitrary waveform generator ("AWG 2") controls a microwave *p-i-n* diode switch ("MS") to create a repeating sequence of microwave pulses with a period of t_d. The amplitudes of the pulses in each period of the sequence are determined by a mask of length P, where P is the number of pulses in the sequence. Here, the mask is a randomly generated binary series. A mask value of "1" corresponds to a pulse, while a value of "0" corresponds to the absence of a pulse. AWG 2 uses square-shaped voltage pulses with a width of 50 ns and a repetition period of 100 ns to create microwave pulses with Gaussian-like shape at the MS output. The period of the driving signal is then $t_d = 100\ P$ ns.

3. TRAINING THE LINEAR READ-OUT

Defining $\hat{\mathbf{Y}}$ 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 $\vec{x}$ for all instances in the training set, concatenated horizontally; similarly, $\hat{\mathbf{Y}}$ is the concatenation of all targets in the training set), the optimal read-out matrix, $\mathbf{W}^{\text{out}}$, is obtained by

$$\mathbf{W}^{\text{out}} = \hat{\mathbf{Y}}\mathbf{X}^{-1},$$

with L points in each trajectory and 40 periods of the driving signal in each frequency interval of the training set; there are a total of $40LM$ training instances in the entire training set, with LM unique pairs of coordinates. Each training instance is a column vector of neuron values with length N. Therefore, the matrix $\mathbf{X}$ has dimensions $N(40LM)$. Similarly, $\hat{\mathbf{Y}}$ has dimensions $2(40LM)$.

With a large number of neurons, the linear regression can be prone to overfitting to the training set. To prevent this, one can use a longer training set, or introduce a regularization parameter, α. In this work, we utilize Ridge regression, also called $L2$ regularization, which is a useful form of regularization for models with a large number of variables. Regularization prevents any read-out weight from becoming too dominant over others, and hence, reducing overfitting. Using Ridge regression, the read-out weights are obtained from data by

$$\mathbf{W}^{\text{out}} = \hat{\mathbf{Y}}\mathbf{X}^{T}(\mathbf{X}\mathbf{X}^{T} + \alpha\mathbf{I})^{-1}.$$

Here, $\mathbf{I}$ is the N-dimensional identity matrix. A low value of α results in overfitting, while a high value results in underfitting. An appropriate value for the regularization parameter is determined to be $\alpha = 0.01$ to give a compromise between the two.

[1] 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 Networks **126**, 191 (2020).

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

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

[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 Networks **115**, 100 (2019).

[5] K. Nakajima, Physical reservoir computing—an introductory perspective, Jpn. J. Appl. Phys. **59**, 060501 (2020).

[6] Z. Konkoli, On developing theory of reservoir computing for sensing applications: The state weaving environment echo tracker (SWEET) algorithm, Int. J. Parallel, Emergent Distrib. Syst. **33**, 121 (2018).

[7] V. Athanasiou and Z. Konkoli, On using reservoir computing for sensing applications: Exploring environment-sensitive memristor networks, Int. J. Parallel, Emergent Distrib. Syst. **33**, 367 (2018).

[8] S. Watt, M. Kostylev, A. B. Ustinov, and B. A. Kalinikos, Implementing a Magnonic Reservoir Computer Model Based on Time-Delay Multiplexing, Phys. Rev. Appl. **15**, 064060 (2021).

[9] M. Wu, Nonlinear spin waves in magnetic film feedback rings, Solid State Phys. **62**, 163 (2010).

[10] J. Grollier, D. Querlioz, K. Y. Camsari, K. Everschor-Sitte, S. Fukami, and M. D. Stiles, Neuromorphic spintronics, Nat. Electron. **3**, 360 (2020).

[11] D. Pinna, G. Bourianoff, and K. Everschor-Sitte, Reservoir Computing with Random Skyrmion Textures, Phys. Rev. Appl. **14**, 054020 (2020).

[12] 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).

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

[14] 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).

[15] 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).

[16] 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).

[17] 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).

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

[19] T. Ichimura, R. Nakane, G. Tanaka, and A. Hirose, in *2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, United Kingdom, 1–8* (2020).

[20] 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).

[21] 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).

[22] 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).

[23] 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).

[24] D. A. Allwood, M. O. A. Ellis, D. Griffin, T. J. Hayward, L. Manneschi, M. F. K. H. Musameh, S. O'Keefe, S. Stepney, C. Swindells, M. A. Trefzer, E. Vasilaki, G. Venkat, I. Vidamour, and C. Wringe, A perspective on physical reservoir computing with nanomagnetic devices, Appl. Phys. Lett. **122**, 040501 (2023).

[25] J. Castera, Tunable magnetostatic surface-wave-oscillators, IEEE Trans. Magn. **14**, 826 (1978).

[26] W. Ishak, 4–20 GHz magnetostatic-wave delay-line oscillator, Electron. Lett. **19**, 930 (1983).

[27] Y. K. Fetisov, P. Kabos, and C. E. Patton, Active magnetostatic wave delay line, IEEE Trans. Magn. **34**, 259 (1998).

[28] We assume that the phase shift associated with the microwave electrical components in the ring is negligible and that all the time delay of the microwave signal occurs in the YIG strip.

[29] 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).

[30] A. B. Ustinov and B. A. Kalinikos, A microwave nonlinear phase shifter, Appl. Phys. Lett. **93**, 102504 (2008).

[31] 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).

[32] 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).

[33] 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).

[34] K. Inoue, K. Nakajima, and Y. Kuniyoshi, Designing spontaneous behavioral switching via chaotic itinerancy, Sci. Adv. **6**, eabb3989 (2020).

[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**, 044902 (2021).

[36] 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 (2021).

[37] C. Fernando and S. Sojakka, Pattern recognition in a bucket, Adv. Artif. Life **2801**, 588 (2003).

[38] D. D. Stancil and A. Prabhakar, *Spin waves: Theory and applications* (Springer New York, NY, 2009).