Editors' Suggestion

Physical reservoir computing and deep neural networks using artificial and natural noncollinear spin textures

Haotian Li,[1] Liyuan Li,[1] Rongxin Xiang,[1] Wei Liu,[3] Chunjie Yan,[1] Zui Tao,[1] Lei Zhang,[4,5] and Ronghua Liu[1,2,*]

[1] *National Laboratory of Solid State Microstructures, Jiangsu Provincial Key Laboratory for Nanotechnology, and School of Physics, Nanjing University, Nanjing, 210093, China*

[2] *National Key Laboratory of Spintronics, Nanjing University, Suzhou, 215163, China*

[3] *Institutes of Physical Science and Information Technology, Anhui University, Hefei, 230601, China*

[4] *Anhui Key Laboratory of Low-Energy Quantum Materials and Devices, High Magnetic Field Laboratory, Hefei Institutes of Physical Science, Chinese Academy of Sciences, Hefei, 230031, China*

[5] *The High Magnetic Field Laboratory of Anhui Province, Hefei, 230031, China*

(Received 13 December 2023; revised 14 March 2024; accepted 17 June 2024; published 11 July 2024)

The growing demand for artificial intelligence has motivated research into nontraditional physical devices that enable efficient learning in various tasks. This requires the devices to exhibit natural nonlinear dynamics with minimal power consumption. Here we present the application of artificial spin ice (ASI) and an as-grown chiral helimagnet (CHM) as the nonlinear component in physical reservoir computing (RC) and deep neural networks (DNNs). Their complex nonlinear magnetodynamics can be easily characterized by the broadband coplanar waveguide–based ferromagnetic resonance technique, originating from the specifically geometrical frustration effect and intrinsic multiple magnetic interactions competition, respectively. On the basis of the experimentally obtained nonlinear magnetodynamic response curves of these two noncollinear spin textures, we build ASI- and CHM-based physical reservoirs for RC and use the absorption and differential ferromagnetic resonance spectra as the activation function and its derivatives to perform nonlinear transformation of inputs for DNNs. The results demonstrate that physical RC and DNNs can accomplish time-series prediction and image-recognition tasks, respectively, with high accuracy and low power consumption. Our findings provide valuable insights and a promising pathway toward neuromorphic hardware using abundant artificial or natural nontrivial magnetic systems.

DOI: 10.1103/PhysRevApplied.22.014027

I. INTRODUCTION

Artificial intelligence (AI) is constantly demonstrating its excellent ability in the fields of image processing, speech recognition, and text generation [1–3]. Nevertheless, the implementation of the AI frameworks in traditional von Neumann computing architectures faces problems in terms of power consumption and processing speed [4]. These problems have prompted the search for non-von Neumann-architectures to realize hardware artificial neural networks (ANNs) using special physical devices, such as spintronic, photonic, and memristor devices [5–7]. Among the numerous AI models, reservoir computing (RC) as a kind of recurrent neural network [8,9] and deep neural networks (DNNs) [10] are two ANN paradigms with extensive use and state-of-the-art performance.

The RC model has significant advantages in applications such as time-series prediction and speech recognition. The reservoir in RC can be seen as an unknown nonlinear system, resembling a black box. Within the reservoir, several key properties strongly influence the performance, especially the complexity of the reservoir state. In previous studies [11,12], it was necessary to generate a set of randomly connected nonlinear nodes as a whole nonlinear dynamic system, which remains fixed throughout the training and testing process. However, recent studies have demonstrated that nonlinear dynamical features in spintronic devices can serve as physical reservoirs [13]. In this architecture, physical inputs such as magnetic fields or electric current will be encoded, and the nonlinear response (e.g., power absorption) of the physical reservoir to the encoded input is used as the output for training. In this way, only the weights between the reservoir and the output layer are updated during the training phase, thereby minimizing energy consumption.

In addition to RC, DNNs are another important ANN paradigm. The backpropagation (BP) algorithm is frequently used to train DNNs in software-based DNNs,

*Corntact author: rhliu@nju.edu.cn

but it faces challenges when one is implementing DNNs at the physical hardware level because the error has to be propagated backward layer-by-layer from the output layer. To address this issue, the direct-feedback-alignment (DFA) algorithm was recently proposed for weight updating in hardware-level implementations [14,15]. In the DFA algorithm, the updating weight is replaced by a fixed stochastic matrix, and the error is fed back directly instead of by layer-by-layer backward propagation. However, the DFA still requires the derivative of the nonlinear activation function to update weight matrices in the training process. Therefore, to achieve DFA-based DNNs at the hardware level, we need a physical system with a stable and natural nonlinear response to an external stimulus, and one can easily obtain this nonlinear behavior and its derivative by conventional electrical methods to implement forward and backward propagation at the hardware level.

In recent years, it has been proven that spintronic nanodevices with naturally nonlinear magnetic dynamics, low power consumption, and stochastic behavior are promising hardware systems for hardware-based neuromorphic computing [16,17]. A few reports indicated that some artificial and natural noncollinear-spin-texture systems [18–21] can be used as the nonlinear dynamical physical reservoir or nonlinear transformation part for neuromorphic computing due to their rich nonlinear magnetodynamics. However, there are many algorithms to implement RC or DNNs, especially at the hardware level. Thus, the artificial and natural magnetic systems with nontrivial dynamics used in hardware implementation of neuromorphic computing are far from being well studied yet. Here we use the broadband coplanar waveguide (CPW)–based ferromagnetic resonance (FMR) technique, shown in Fig. 1(a), to fully characterize and understand the complex nonlinear magnetic dynamics of artificial spin ice (ASI) with a hexagonal kagome structure [Fig. 1(b)] and the as-grown chiral helimagnet (CHM) $CrNb_3S_6$ with a naturally occurring superlattice [Fig. 1(c)]. Then we systematically explore the implementation process and performance of ASI- and CHM-based RC and DNNs with the DFA algorithm through the prediction of two chaotic time-series datasets (Mackey-Glass and NARMA10) and a handwritten-digit-recognition task. The results demonstrate that these nonlinear magnetic systems exhibit huge potential to facilitate the implementation of ANNs on hardware with benefits in reducing computational requirements.

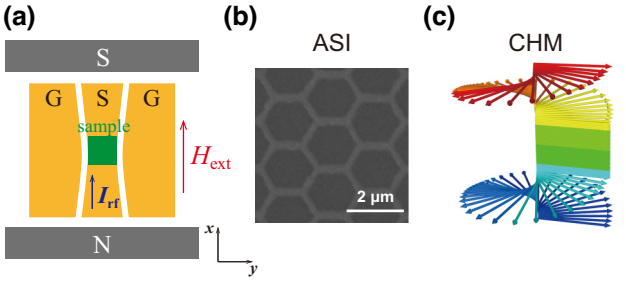

FIG. 1. Differential FMR experimental setup and two nonlinear dynamics of magnetic devices. (a) The sample is fixed to the center signal line of the CPW consisting of two ground parts and one center signal line and the purple and red arrows represent the injection of microwave current and external magnetic field, respectively. (b) Scanning-electron-microscope images of the ASI. (c) Three-dimensional schematic spin configurations of the CSL.

II. RESULTS AND DISCUSSION

A. Dynamic characterization of noncollinear spin textures

1. Artificial spin ice

ASI is a type of geometrically frustrated system, wherein the spin configuration resembles the spatial structure between hydrogen and oxygen atoms in ice. Figure 1(b) shows scanning-electron-microscope images of the hexagonal kagome ASI. The ASI sample is fabricated with use of a hexagonal kagome pattern by permalloy with a thickness of 25 nm on a CPW. Each side of the hexagon as the magnetic pole of the sample has a length of 1 μm and a width of 250 nm. To investigate the characteristics of magnetic dynamics in ASI systems, as shown in Fig. 1(a), we measured the broadband FMR absorption spectra by sweeping the frequency in the range from 0.1 to 10 GHz at a magnetic field H_{ext} parallel to the direction of the CPW. In these kinds of ASI, the dispersion relations of the spin-wave modes are closely related to the external magnetic field, various exchange interactions, and anisotropic field. Because of geometrical frustration and various interactions, ASI exhibits diverse spin dynamical behaviors at different field ranges [22–24]. The magnetic dynamics of ASI can be easily accessed and accurately characterized by the broadband FMR technique based on a coplanar waveguide. Figure 2(a) shows the experimental FMR spectra of this kagome ASI measured by our sweeping the frequency of the excitation radio-frequency (rf) signal in the field range from −1 to 1 kOe along the horizontal axis. The pseudocolor map of the experimental FMR spectra obtained shows several spin-wave modes with distinct H_r-versus-f dispersion relations and spectral intensities. To further analyze these experimentally observed spectra of spin-wave modes, we perform the magnetization simulation of this kagome ASI by using MUMAX^3 [25]. Three representative FMR absorbed power spectra with different fields obtained by simulation are shown in Fig. 2(b). The FMR spectra exhibit four distinct peaks, in agreement with the experiment results [Fig. 2(a)]. The calculated normalized-spatial-power profile of these peaks, as shown in Fig. 2(c), shows that the two high-frequency peaks, with a greater power intensity, are localized in the horizontal

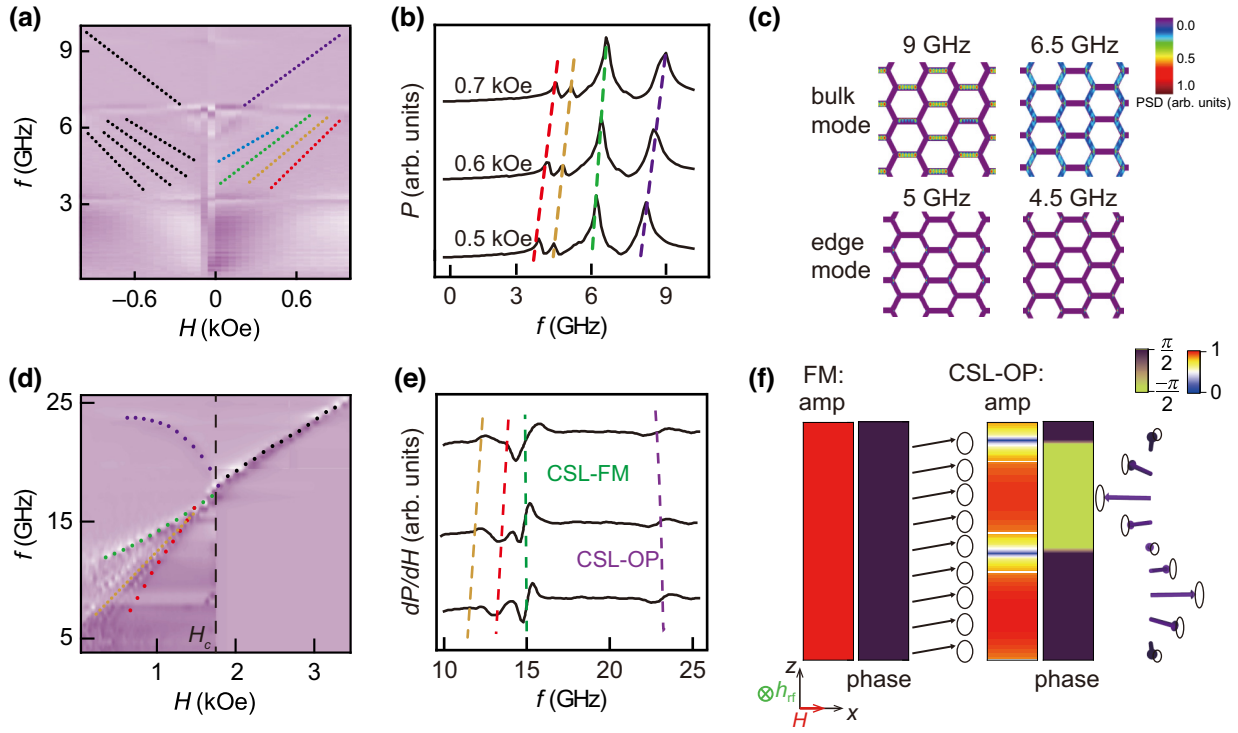

FIG. 2. Magnetic dynamic characteristics of ASI and the CHM. (a) Pseudocolor map of the experimental FMR spectra of ASI at room temperature. Five series of resonance peaks are represented by the dotted lines to guide the eye. (b) Representative FMR absorbed power spectra of ASI obtained by simulation at three selected in-plane fields $H = 0.5$, 0.6 and 0.7 kOe. (c) Spatial profiles of the power spectrum density (PSD) for two high-frequency (upper panel) and two low-frequency (lower panel) peaks correspond to the bulk and edge spin-wave modes, respectively. (d) Pseudocolor map of the experimental FMR spectra of the CHM single crystal $CrNb_3S_6$ obtained at $T = 4$ K. The critical field $H_c = 1.72$ kOe, represented by the vertical dashed line, separates the low-field CSL state from the high-field forced ferromagnetic state in the FMR spectra. The dotted lines are guides for the eye, marking different spin-wave modes. (e) Representative experimental differential FMR spectra of the CSL state measured at three selected in-plane fields $H_{ext} = 1.3$, 1.4, and 1.5 kOe below H_c. These different peaks corresponding to distinct spin-wave modes are represented by "CSL-LF," "CSL-FM," and "CSL-OP", which are defined in the main text. (f) Spatial distributions (amplitude abbreviated as "amp" and phase) and three-dimensional schematic precession trajectory of the ferromagnetic (FM) and CSL-OP modes, respectively.

and vertical bars with a periodic intensity distribution, respectively. These features indicate that the two peaks are the standing backward volume magnetostatic spin waves due to the geometric demagnetization-induced quantum confinement effect, named the "bulk mode." On the other hand, the low-frequency peaks, with a lower power intensity, are mutually exclusive distributions in the corners of the hexagonal kagome lattice due to dipolar interaction. Therefore, they are identified as the edge mode.

2. Chiral helimagnets

In addition to ASI with use of the specific artificial structures for tailoring the dispersion relation of magnetic dynamics, the as-grown CHM with naturally occurring superlattices [e.g., chiral soliton lattice (CSL)] due to crystal chirality and competing various magnetic interactions also exhibit several collective spin-resonance modes with distinct dispersion relations [26]. In addition, they can also be modulated by one changing the temperature and the external magnetic field strength and orientation. We measure the FMR spectra of a 1-mm-square single-crystal $CrNb_3S_6$ sample with a thickness of about 10 μm at a low temperature of 4 K. The direction of H_{ext} is parallel to the direction of the CPW, and the range of excitation frequencies is from 3 to 30 GHz. The CSL consists of an array of chiral soliton kinks (helical spin texture) periodically partitioned by ferromagnetic domains, as illustrated in Fig. 1(c). Because of the competition of magnetic interaction energies (including monoaxial Dzyaloshinskii-Moriya interactions, symmetric Heisenberg exchange, magnetocrystalline anisotropy, and Zeeman energy), the CHM exhibits various field-dependent nontrivial spin textures and is anticipated to possess tunable and reconfigurable magnon dispersion. Figure 2(d) shows representative magnetic field-dependent FMR spectra of $CrNb_3S_6$ single crystal with excitation frequency $f = 4$–25.5 GHz and in-plane field $H = 0$–3.5 kOe at $T = 4$ K. One easily see four distinct spin-dynamic regimes with different spectral characteristics. In the high-field range $H > H_c = 1.72$ kOe,

the single peak in the FMR spectra show the Kittel dispersion relation, consistent with the high-field forced ferromagnetic state. For the low-field CSL state, the spectra exhibit several peaks, as shown in Fig. 2(e). The high-frequency peak, named "CSL-OP," shows the frequency inversely proportional to H [Fig. 2(d)]. This inverse dispersion relation can be described by a collective phonon-resonance mode [27]:

$$f = \frac{nDS^2c}{\hbar \ln\left(\frac{16}{1-H/H_C}\right)}, \tag{1}$$

where $c = 12.018\,\text{Å}$ is the c-axis lattice constant of $CrNb_3S_6$, $\hbar$ is the reduced Planck constant, $D = 2.9\,\text{K}$ is the Dzyaloshinskii-Moriya interaction coefficient, $S = 3/2$ is the total spin of the Cr^{3+} ion, and n is the mode number of the phonon resonance.

The spatial distributions of amplitude and phase for this mode obtained by micromagnetic simulation [left panel in Fig. 2(f)] show that the spins with opposite directions in the CSL show a π phase difference. The behavior is consistent with out-of-phase-precession optical mode characteristics. In addition, the spectrum still contains a middle-frequency peak with a Kittel-like dispersion relation but with a 1-GHz drop compared with the forced ferromagnetic mode. The right panel in Fig. 2(f) shows a uniform spatial profile for both amplitude and phase, consistent with the conventional FMR mode. Thus, this mode is named "CSL-FM." Finally, we still observe low-frequency spectra with a series of stochastic peaks especially for H far below H_c, denoted as "CSL-LF."

Reconfigurable magnonics and spin wave–based neuromorphic computation are promising future technologies reliant on their low power consumption, high speed, and natural complex nonlinearity. Recent work has revealed that nonlinear dynamics of magnetic devices or samples can be used as RC via spin-wave fingerprinting. Therefore, on the basis of our FMR experimental results, we further build ASI- and CHM-based reservoir computing and FMR-based DFA algorithms for DNNs by using their complex field-dependent spin-wave dispersions.

B. Magnetic dynamics–based neuromorphic computing

1. ASI- and CHM-based physical reservoir computing

Figure 3 shows a conceptual construction of the physical RC system based on a nonlinear magnetic dynamic system. The RC system comprises three main components: the input part, the physical nonlinear system as the reservoir, and the output part. The role of the input part is to linearly map the magnitude of the time-series data $\boldsymbol{x}(t)$ to specific physical quantities that drive the physical reservoir. Subsequently, the nonlinear magnetic dynamic response of the physical reservoir (e.g., ASI and CHM) to the input excitation [e.g., the time series of the magnetic field $\mathbf{H}(t)$] is used to expand the time series of input data to a higher dimension $\mathbf{R}_{\text{out}}$. Finally, the higher-dimensional data $\mathbf{R}_{\text{out}}$ with the adjustable weight $\mathbf{W}_{\text{out}}$ are converted back to the final result in the output layer after the training process.

We use the time-series prediction task of a nonlinear dynamic system as an example to specifically illustrate the physical RC system and its computing process. In the input layer [Fig. 3(a)], the time-series prediction task as the input data $\boldsymbol{x}(t) \in \mathbb{R}^N$ is linearly mapped to the appropriate time series of the magnetic field, where N is the number of time steps. The input layer encodes the numbers into the magnetic field by linear mapping. This method

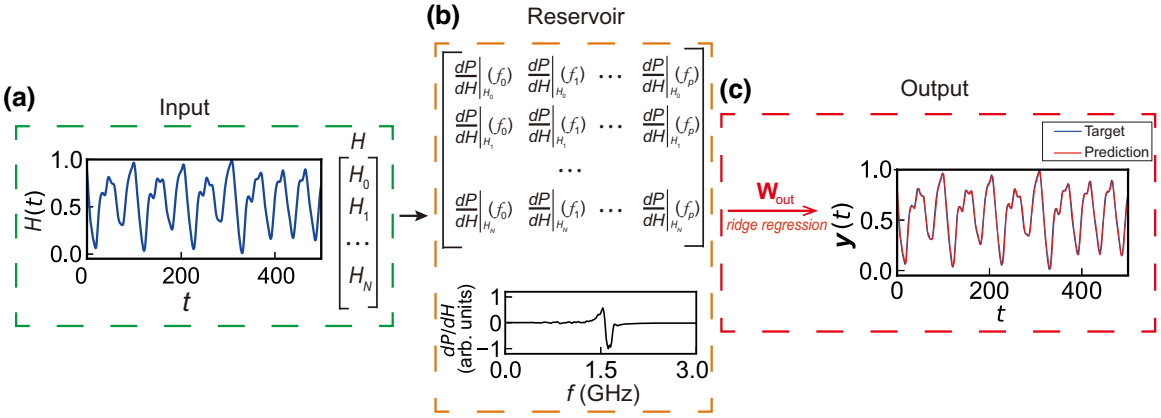

FIG. 3. Reservoir-computing scheme and training results. (a) The input time-series data $\boldsymbol{x}(t) \in \mathbb{R}^N$ are linearly mapped to the time series of the magnetic field $\mathbf{H}(t) \in \mathbb{R}^N$ in the input layer. (b) Magnetic field–dependent differential FMR spectra $dP/dH|_{H_i}(f_j)$ as the physical reservoir. $H(t_i)$ is the input of the physical reservoir at each time step. After the physical reservoir, $\mathbf{H}(t)$ is nonlinearly transformed into a reservoir state matrix $\mathbf{R}_{\text{out}} \in \mathbb{R}^{N\times M}$. (c) $\mathbf{R}_{\text{out}}$ is connected to the output layer by a trainable weight matrix $\mathbf{W}_{\text{out}}$. The error between prediction and target results is used to update $\mathbf{W}_{\text{out}}$ by ridge regression. After the training phase, the $\mathbf{W}_{\text{out}}$ matrix is fixed for the testing phase.

involves mapping the time-series database from 0 to 1 onto a suitable magnetic field range to drive the physical systems as reservoir layers. For ASI, the range of magnetic fields is from 0 to 1 kOe, corresponding to the positive field part in Fig. 2(a). In the case of the CHM, the magnetic field range is from 0.5 to 1.7 kOe, corresponding to the low-field CSL state in Fig. 2(d). Both magnetic systems exhibit multiple spin-wave modes with distinct FMR spectral curves. Here we present a numerical-simulation scheme for physical reservoir computing and the learning progress represented by dimensionless time-step points for widespread usability. In hardware applications, the actual timescale depends mainly on the input-external-field timescale and the response time of the physical reservoir layer. For the time-series prediction task, we use the experimental differential FMR spectral curves from 3 to 6 GHz with $M = 160$ data points for ASI and from 7 to 25 GHz with $M = 92$ data points for the CHM. Here M corresponds to the number of sampling points for the differential FMR curves. All the magnetic dynamic modes of the two magnetic systems can be clearly characterized. Figure 3(b) shows that each $dP/dH|_{H_i}(f_j)$ data point serves as a node of the reservoir in the conventional reservoir model [28], where i is the time step and j is the frequency data points. Therefore, the physical reservoir converts the input $\boldsymbol{x}(t)$ to the reservoir output $\mathbf{R}_{\text{out}} = \boldsymbol{f}_M(\mathbf{H}(t)) \in \mathbb{R}^{N\times M}$, where $\boldsymbol{f}_M()$ is a differential FMR curve with M data points. After nonlinear transformation by the physical reservoir, the dimension of the input data $\boldsymbol{x}(t)$ is expanded to $N \times M$. In addition to nonlinearity, the short-term memory effect in physical devices is also necessary for physical RC. The short-term memory effects in the ASI and CHM magnetic systems are related to the hysteresis behavior due to various nonreversible magnetization processes and the intrinsic relaxation process of magnetization under an alternative magnetic field [29,30]. Finally, $\mathbf{R}_{\text{out}}$ is converted back to a time series with the same dimension as the input data by a suitable weight matrix $\mathbf{W}_{\text{out}}$ with M elements in the output layer. A complete time-series prediction task includes two steps: the training phase and the testing phase. Therefore, the input time-series dataset needs to be divided into the training set and the testing set. The training phase aims to find the optimal weight matrix $\mathbf{W}_{\text{out}}$, which makes the predicted value $\hat{\boldsymbol{y}}$ close to the target value $\boldsymbol{y}(t) = \mathbf{R}_{\text{out}} \times \mathbf{W}_{\text{out}}$ by the ridge-regression method. Subsequently, the test dataset is fed into the physical reservoir to obtain the prediction results with the fixed $\mathbf{W}_{\text{out}}$ obtained by the training phase.

We evaluate the performance of the physical RC system by selecting two nonlinear datasets commonly used for RC benchmarking: the Mackey-Glass dataset [31] and the NARMA10 dataset [32] (see Appendix A for the specific formula). We use the PYTHON reservoirpy package to generate the datasets [33]. During the training phase, we use the time series $x(t)$ at t, and the target value $\boldsymbol{y}(t) = \boldsymbol{x}(t+\tau)$, where τ represents the time step for predicting future data. Firstly, we set τ to 1, meaning that we aim to predict the data after one time step. Specifically, we transform and feed 500 time steps of input data into the physical reservoir (ASI or CHM). At each time step, we update the weight $\mathbf{W}_{\text{out}}$ using the ridge-regression method [34]. Figure 4 shows the target and prediction results of these two nonlinear dynamic systems by ASI- and CHM-based physical RC, where the dotted black line represents the target value and the dotted red line is the predicted result. One can easily see that both nonlinear datasets can be accurately predicted, particularly at jump points. To further quantify the accuracy of the final prediction results, we adopt the normalized root-mean-squared error (NRMSE) $\text{NRMSE} = \sqrt{\sum_{i=0}^{N-1}(y_i - \hat{y}_i)^2/N\lambda}$, where λ is a coefficient depending on the normalization method; here we choose the peak-to-peak-amplitude-normalization method. The NRMSEs are 0.0352 (ASI) and 0.0366 (CHM) for the Mackey-Glass task, and 0.1653 (ASI) and 0.1671 (CHM) for the NARMA10 task, comparable with the results in the digital circuit and optoelectronic reservoirs [35,36]. To further validate the predictive capability of the physical RC, additional forecasting results for more time steps τ are presented in Fig. 5. Figure 5(a) illustrates the predictive performance of two magnetic systems as physical reservoir layers for the Mackey-Glass dataset, while Fig. 5(b) depicts results for the NARMA10 dataset. It can be seen that as the prediction time increases, the physical reservoir layer can realize a high accuracy rate when $\tau \leq 5$.

We also evaluated the information-processing capacity (IPC) [37] and intrinsic nonlinearity of the

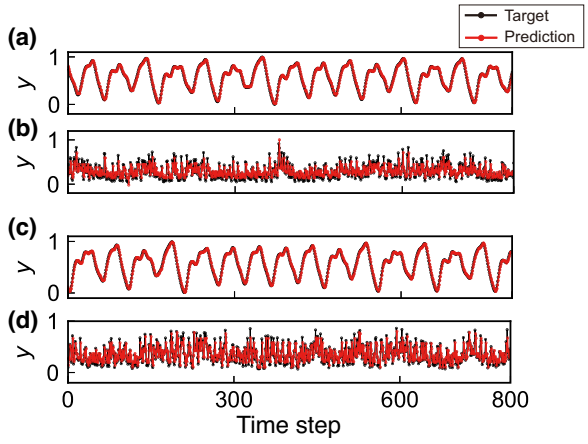

FIG. 4. Prediction results for two nonlinear dynamic systems obtained by ASI- and CHM-based physical RC. (a),(b) Prediction results obtained by ASI-based physical RC for the Mackey-Glass dataset (a) and the NARMA10 dataset (b). (c),(d) Prediction results obtained by CHM-based physical RC for the Mackey-Glass dataset (c) and the NARMA10 dataset (d). The theoretical target results (black) and the prediction results obtained by the physical RC (red) are both in the testing phase.

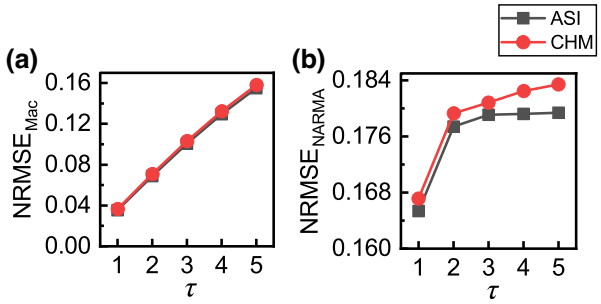

FIG. 5. NRMSE results for (a) the Mackey-Glass dataset and (b) the NARMA10 dataset dependent on prediction time τ for ASI and the CHM.

magnetic systems (see Appendix B for the specific formula).

The nonlinear features were represented with use of the Lyapunov exponent (LE) [38], and the specific expression can be found in Appendix B. The IPC of the system is proposed to quantify the information-processing capability of dynamic systems, and values of IPC closer to 1 signify a robust operational capacity within the dynamical system. Here we analyze the IPC of ASI and the CHM for the Mackey-Glass and NARMA10 datasets. It should be noted that the IPC value of the reservoir layer depends on a specific prediction-task condition. Table I presents all IPC values for two magnetic systems (ASI and CHM) serving as physical reservoir layers for the Mackey-Glass and NARMA10 datasets at prediction steps τ from 1 to 5. For $\tau \leq 5$ prediction tasks, the IPC values of ASI- and CHM-based RC are greater than 0.93 and 0.77 for the Mackey-Glass and NARMA10 datasets, respectively. This quantitative analysis of IPC indicates that the ASI- and CHM-based RC studied can effectively capture the features of the two nonlinear systems.

In addition to IPC, the LE is used to signify the nonlinearity of physical reservoir layers. The largest global LE serves as a key indicator of dynamic sensitivity. A positive maximum LE suggests highly chaotic system behavior, indicative of strong nonlinearity. The LE values of Mackey-Glass dataset and the NARMA10 dataset are 0.0778 and 0.2321, respectively. The LE results of our physical reservoirs based on ASI and CHM are 0.223 and 0.1301, respectively. The LE results indicated that the ASI and CHM systems are suitable for practical implementation within RC.

TABLE I. Information-processing-capacity values for two physical reservoir layers with different prediction datasets and various prediction time steps τ.

	Mackey-Glass		NARMA10	
τ	ASI	CHM	ASI	CHM
1	0.9966	0.9966	0.8156	0.8102
2	0.9872	0.9872	0.7829	0.7854
3	0.9728	0.9727	0.7767	0.7796
4	0.9551	0.9549	0.7745	0.7791
5	0.9359	0.9356	0.7742	0.7788

Because of the short-term memory and complex magnetic dynamic characteristics caused by the special magnetostatic structure, the time-series prediction and handwritten-digit-recognition tasks demonstrated that both devices can be used independently as a reliable reservoir layer in reservoir computing and as the nonlinear transformation component in deep neural networks (as demonstrated below), thereby significantly reducing the processing time required by memory. However, the two devices exhibit different magnetic dynamics behaviors due to distinct magnetic interactions. ASI exhibits a variety of spin-wave modes, including edge modes and bulk modes, due to the special dipolar interaction caused by geometrical frustration. The CSL state of the CHM caused by Dzyaloshinskii-Moriya interactions, symmetric Heisenberg exchange, and other interactions shows various resonance modes in different frequency ranges. Our work shows that although the NRMSE of the CHM is slightly larger than that of ASI, the special nonlinear dynamic characteristics of the CHM enable it to accomplish the same time-series prediction task as ASI.

2. DFA-based deep neural networks

Here we demonstrate how to use the FMR absorption and differential spectra of ASI and CHM systems as the nonlinear activation function and its derivative to achieve hardware implementation of DNNs. As shown in Fig. 6(a), we consider a DNN with four layers, including an input layer, two hidden layers (h_1, h_2), and an output layer. A complete learning cycle is divided into two parts: forward propagation and BP [39,40]. In the forward-propagation process, the input data x propagate forward and satisfy the following relation: $x_n = f(m_n)$, where n is the number of layers from 1 to 3, x_n is the input of the nth layer, $f()$ refers to the nonlinear activation function, $\mathbf{W}_n$ is the weight matrix, and $m_n = \mathbf{W}_n x_{n-1}$. After forward propagation, the BP is essential to learn the characteristics of input data. To evaluate the error of the network, we use the cross-entropy [41] to estimate the loss L and error $r = \hat{Y} - Y$, where $\hat{Y}$ is the DNN output value and Y is the target value. The gradient of each hidden layer can be calculated by one following the formula $\delta m_n = \partial L/\partial m_n = [\mathbf{W}_{n+1}^T r] \odot f'(m_n)$, where the gradient of the last hidden layer δm_3 is the error r, the superscript T is the transpose operation, $\odot$ denotes the Hadamard product, and $f'()$ represents the derived function of $f()$. Finally, the corresponding weight update is computed by the chain rule [42]:

$$\delta \mathbf{W}_n = \frac{\partial L}{\partial \mathbf{W}_n} = -\delta m_n x_{n-1}^T. \tag{2}$$

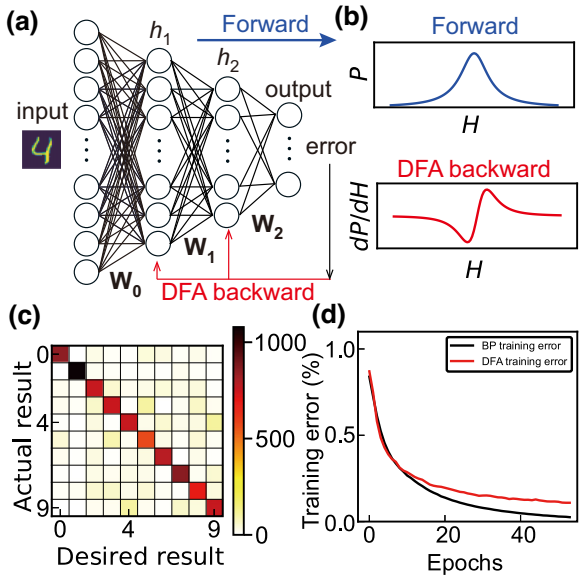

FIG. 6. Architecture diagram of the DFA-based DNN and its performance in the handwritten-digit-recognition task. (a) Schematic network structure of the DFA-based DNN. Differing from the layer-by-layer updating process for the weight in the conventional BP algorithm, the DFA simultaneously updates weights ($\mathbf{W}_0$, $\mathbf{W}_1$, and $\mathbf{W}_2$) on the basis of an arbitrary matrix as the initial weight according to the error between the output value and the target value. (b) Representative FMR absorption spectrum (for forward propagation) and differential FMR spectrum (for backward propagation) are directly obtained by the FMR experiments. (c) Testing results for the physical DFA-based DNN with 10 000 test sets of handwritten-digit images. (d) Comparison of results for the recognition error rate as a function of the number of testing epochs between the conventional BP algorithm and our physical DFA algorithm.

Although the BP algorithm based on gradient descent is the most effective in DNNs, this method requires symmetric weights for both forward propagation and backward propagation. Furthermore, the layer-by-layer BP process does not conform to the data-processing methods observed in the brain, leading to increased computation time and complexity. Therefore, BP is a biologically implausible algorithm and hard to realize at the hardware level [14,43]. To avoid this tedious BP procedure, we use the DFA algorithm proposed recently, where the weight information of the previous layer in the BP process is replaced by a random matrix $\mathbf{B}_n$, which can be directly updated on the basis of the error r with $\delta m_n = (\mathbf{B}_n r) \odot f'(m_n)$ [44].

Next, as shown in Fig. 6(a), we use the handwritten-digit-recognition task as an example to specifically explain the learning process of the physical DFA-based DNN. In our physical DFA-based DNN, the nonlinear dynamical response (the FMR absorption spectra) of the magnetic system is used as the activation function for forward propagation, and the corresponding differential FMR spectrum as the derivative of the activation function DFA-based BP. Both of these can be directly obtained by FMR experimental techniques, as shown in Fig. 6(b). In the initial stage, a random matrix is used as the initial weight. For forward propagation, the input image with 28×28 pixels is linearly mapped to the magnetic field range of the FMR of the physical device [Fig. 6(b)]. In the hidden layers, each neuron node (denoted by a circle) represents the nonlinear operation, i.e., FMR absorption function. Therefore, the input data as the magnetic field with adjustable weights are nonlinearly transformed as the FMR absorption curve of the device. The final output layer has ten neurons, corresponding to the predicted results of ten digits (0–9). Following the forward propagation, the output is compared with the labeled target value to get the loss L and error r. Then the weights can be updated according to Eq. (2) with $\delta m_n = (\mathbf{B}_n r) \odot f'(m_n)$, where the derivative of the activation function is the experimentally obtained differential FMR absorption spectrum. After the training phase, the optimized weights are obtained and fixed. Finial, we use the test set, without it being used in the training phase, as the input to do a validity check of our physical DFA-based DNN. Figure 6(c) shows the testing results with a high accuracy of 90%. Furthermore, we also compare the training accuracy between our physical DFA algorithm and the traditional BP algorithm, as shown in Fig. 6. Although the convergence rate of training error of DFA is slightly slower than that of BP with an increase in the number of epochs, this physical DFA-based DNN still reaches 90% accuracy, a significant achievement in hardware implementation.

III. SUMMARY

In summary, ASI and the CHM possess the advantages of natural nonlinearity of magnetic dynamics and dynamic memory over Si-based devices for non-von Neumann-computing. From the perspective of magnetic dynamics, the CHM exhibits various spin-wave modes in the CSL state, particularly the inverse dispersion relationship in the FMR spectrum due to the collective phonon-resonance modes, which is different from the dispersion described by the Kittel formula for traditional magnetic systems. Compared with ASI, physical RC based on natural CHM materials can realize variable nonlinear transformations of time-series data with nearly the same NRMSE. Therefore, as a two-dimensional van der Waals material, the CHM with complex magnetic dynamics can be directly applied in neuromorphic computing without the need for software-based memory or time-multiplexing components. The application of the CHM simplifies the architecture of physical RC and DNNs. Compared with software-based RC and DNN-based machine-learning algorithms, hardware-based implementation of the physical RC and DFA-based DNNs can yield a huge reduction in computational requirements. We build the physical RC using ASI and CHM materials, which can achieve accurately

predicted results, with NRMSE approximately 0.03 and 0.16 for Mackey-Glass and NARMA10 nonlinear systems, respectively. Additionally, we further use the experimentally obtained FMR absorption spectra and differential FMR spectra of magnetic systems as an activation function and its derivative to build a physical DNN, which can avoid the huge computational requirement for tediously calculating the activation function and its derivatives of data in the numerical DNN model. Furthermore, the DFA algorithm used can simultaneously update weight matrices $\mathbf{W}_i$ rather than the traditional layer-by-layer propagation updating process of $\mathbf{W}_i$, which also can significantly accelerate the speed and with a comparable accuracy rate of 90% in the handwritten-digit-recognition task. Our results demonstrate that the naturally nonlinear magnetic dynamics of ASI and CHM systems can replace the complex nonlinear mathematical operations in traditional machine-learning models, which provides a promising strategy for low-energy analog computing using neuromorphic hardware.

ACKNOWLEDGMENTS

We acknowledge support from the National Key Research and Development Program of China (Grant No. 2023YFA1406603), the National Natural Science Foundation of China (Grant No. 12074178, 12074386, 12374128, 12204006), and the Open Research Fund of Jiangsu Provincial Key Laboratory for Nanotechnology.

APPENDIX A: NONLINEAR TIME-SERIES DATASETS FOR RESERVOIR COMPUTING

We used two nonlinear datasets to validate the predictive capabilities of reservoir computing based on magnetic systems: the Mackey-Glass and NAMAR10 datasets. These datasets are commonly used to assess the predictive performance of systems. The specific expression for the Mackey-Glass dataset used in this work is as follows:

$$\frac{y(t)}{t} = \frac{0.2 \times y(t-17)}{1 + y(t-17)^{10}} - 0.1 \times y(t), \tag{A1}$$

where $y(t)$ is the time-series data.

For the NARMA10 dataset, the expression is

$$y(t+1) = 0.2 \times y(t) + 0.04 \times y(t) \left(\sum_{i=0}^{n-1} y(t-i) \right) + 1.5 \times u(t-(n-1))u(t), \tag{A2}$$

where $n = 10$ and $u(t)$ is sampled following a uniform distribution in [0, 0.5]. We generated these two datasets numerically using the reservoirpy package in PYTHON.

APPENDIX B: INFORMATION-PROCESSING CAPACITY AND LYAPUNOV EXPONENT OF THE MAGNETIC SYSTEM FOR RESERVOIR COMPUTING

The IPC is usually used to quantify the information-processing capability of dynamical systems. The definition of the IPC is as follows:

$$C[R, y(t)] = 1 - \frac{\min_{W_i} M_T[\hat{y}(t)]}{\frac{1}{T}\sum_{t=1}^{T} y(t)^2}, \tag{B1}$$

where R is the state of the reservoir layer, $y(t)$ is the target output, W_i is the weight between the reservoir layer and the output layer, $\hat{y}(t)$ is the actual output, and T is the total time. Here, $M_T[\hat{y}(t)] = \frac{1}{T}\sum_{r-1}^{T}(\hat{y}(t) - y(t))^2$.

The LE measures the rate of exponential growth or decay of nearby trajectories in a dynamical system. A positive LE indicates that nearby trajectories diverge exponentially, meaning even small changes in the initial conditions can lead to drastically different outcomes. The definition of the LE is as follows:

$$|R(t, R_0) - R(t, R_0 + \varepsilon)| = e^{\lambda t}|\varepsilon|, \tag{B2}$$

where $R(t, R_0)$ is the trajectory of the system R starting at the point R_0 in phase space at time t, ϵ is the (infinitesimal) difference vector, and λ is the Lyapunov exponent.

[1] Y. LeCun, Y. Bengio, and G. Hinton, Deep learning, Nature **521**, 436 (2015).
[2] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, Convolutional neural networks: An overview and application in radiology, Insights Imaging **9**, 611 (2018).
[3] P. P. Ray, ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope, Internet Things Cyber-Phys. Syst. **3**, 121 (2023).
[4] A. Prieto, B. Prieto, E. M. Ortigosa, E. Ros, F. Pelayo, J. Ortega, and I. Rojas, Neural networks: An overview of early research, current frameworks and new challenges, Neurocomputing **214**, 242 (2016).
[5] H. Li, L. Li, K. Zhou, C. Yan, Z. Gao, Z. Li, and R. Liu, Stochastic spin-orbit-torque device as the STDP synapse for spiking neural networks, Sci. China Phys. Mech. Astron. **66**, 257512 (2023).
[6] X. Lin, Y. Rivenson, N. T. Yardimci, M. Veli, Y. Luo, M. Jarrahi, and A. Ozcan, All-optical machine learning using diffractive deep neural networks, Science **361**, 1004 (2018).
[7] M. Lanza, A. Sebastian, W. D. Lu, M. Le Gallo, M.-F. Chang, D. Akinwande, F. M. Puglisi, H. N. Alshareef, M. Liu, and J. B. Roldan, Memristive technologies for data storage, computation, encryption, and radio-frequency communication, Science **376**, eabj9979 (2022).
[8] C. Gallicchio, A. Micheli, and L. Pedrelli, Deep reservoir computing: A critical experimental analysis, Neurocomputing **268**, 87 (2017).

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

[10] L. G. Wright, T. Onodera, M. M. Stein, T. Wang, D. T. Schachter, Z. Hu, and P. L. McMahon, Deep physical neural networks trained with backpropagation, Nature **601**, 549 (2022).

[11] D. J. Gauthier, E. Bollt, A. Griffith, and W. A. S. Barbosa, Next generation reservoir computing, Nat. Commun. **12**, 5564 (2021).

[12] M. Cucchi, S. Abreu, G. Ciccone, D. Brunner, and H. Kleemann, Hands-on reservoir computing: A tutorial for practical implementation, Neuromorph. Comput. Eng. **2**, 032002 (2022).

[13] 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).

[14] A. Nøkland, Direct feedback alignment provides learning in deep neural networks, arXiv:1609.01596.

[15] M. Nakajima, K. Inoue, K. Tanaka, Y. Kuniyoshi, T. Hashimoto, and K. Nakajima, Physical deep learning with biologically inspired training method: Gradient-free approach for physical hardware, Nat. Commun. **13**, 7847 (2022).

[16] Q. Zheng, Y. Mi, X. Zhu, Z. Yuan, and K. Xia, Anticipative tracking with the short-term synaptic plasticity of spintronic devices, Phys. Rev. Appl. **14**, 044060 (2020).

[17] W. Wang, Z. Wang, Y. Zhang, B. Sun, and K. Xia, Learning order parameters from videos of skyrmion dynamical phases with neural networks, Phys. Rev. Appl. **16**, 014005 (2021).

[18] C. Castelnovo, R. Moessner, and S. Sondhi, Spin ice, fractionalization, and topological order, Annu. Rev. Condens. Matter Phys. **3**, 35 (2012).

[19] Q. Li, S. Xiong, L. Chen, K. Zhou, R. Xiang, H. Li, Z. Gao, R. Liu, and Y. Du, Spin-wave dynamics in an artificial kagome spin ice, Chin. Phys. Lett. **38**, 047501 (2021).

[20] L. Li, H. Li, K. Zhou, X. Xiao, L. Chen, F. Ma, J. Zhang, L. Wang, L. Zhang, and R. Liu, Observation and characterization of multiple resonance modes in a chiral helimagnet $CrNb_3S_6$, Nano Lett. **23**, 9243 (2023).

[21] J.-i. Kishine and A. S. Ovchinnikov, Theory of spin resonance in a chiral helimagnet, Phys. Rev. B **79**, 220405 (2009).

[22] S. H. Skjærvø, C. H. Marrows, R. L. Stamps, and L. J. Heyderman, Advances in artificial spin ice, Nat. Rev. Phys. **2**, 13 (2020).

[23] T. Dion, D. M. Arroo, K. Yamanoi, T. Kimura, J. C. Gartside, L. F. Cohen, H. Kurebayashi, and W. R. Branford, Tunable magnetization dynamics in artificial spin ice via shape anisotropy modification, Phys. Rev. B **100**, 054433 (2019).

[24] V. S. Bhat, F. Heimbach, I. Stasinopoulos, and D. Grundler, Angular-dependent magnetization dynamics of kagome artificial spin ice incorporating topological defects, Phys. Rev. B **96**, 014426 (2017).

[25] A. Vansteenkiste, J. Leliaert, M. Dvornik, M. Helsen, F. Garcia-Sanchez, and B. Van Waeyenberge, The design and verification of MuMax3, AIP Adv. **4**, 107133 (2014).

[26] O. Lee, T. Wei, K. D. Stenning, J. C. Gartside, D. Prestwood, S. Seki, A. Aqeel, K. Karube, N. Kanazawa, Y. Taguchi, C. Back, Y. Tokura, W. R. Branford, and H. Kurebayashi, Task-adaptive physical reservoir computing, Nat. Mater. **23**, 79 (2024).

[27] K. Hoshi, J.-I. Kishine, and J.-I. Ohe, Coupled-oscillator collective mode of a magnetic chiral soliton lattice, Phys. Rev. B **102**, 134414 (2020).

[28] J. C. Gartside, K. D. Stenning, A. Vanstone, H. H. Holder, D. M. Arroo, T. Dion, F. Caravelli, H. Kurebayashi, and W. R. Branford, Reconfigurable training and reservoir computing in an artificial spin-vortex ice via spin-wave fingerprinting, Nat. Nanotechnol. **17**, 460 (2022).

[29] A. Glavic, B. Summers, A. Dahal, J. Kline, W. Van Herck, A. Sukhov, A. Ernst, and D. K. Singh, Spin solid versus magnetic charge ordered state in artificial honeycomb lattice of connected elements, Adv. Sci. **5**, 1700856 (2018).

[30] M. Ohkuma, M. Mito, N. Nakamura, K. Tsuruta, J. Ohe, M. Shinozaki, Y. Kato, J. Kishine, Y. Kousaka, J. Akimitsu, and K. Inoue, Surface-size and shape dependencies of change in chiral soliton number in submillimeter-scale crystals of chiral magnet $CrNb_3S_6$, AIP Adv. **9**, 075212 (2019).

[31] M. C. Mackey and L. Glass, Oscillation and chaos in physiological control systems, Science **197**, 287 (1977).

[32] T. Kubota, H. Takahashi, and K. Nakajima, Unifying framework for information processing in stochastically driven dynamical systems, Phys. Rev. Res. **3**, 043135 (2021).

[33] N. Trouvain, L. Pedrelli, T. T. Dinh, and X. Hinaut, ReservoirPy: An Efficient and User-Friendly Library to Design Echo State Networks, ICANN 2020 - 29th International Conference on Artificial Neural Networks (2020).

[34] A. E. Hoerl and R. W. Kennard, Ridge regression: Biased estimation for nonorthogonal problems, Technometrics **42**, 80 (2000).

[35] 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).

[36] P. Kumar, M. Jin, T. Bu, S. Kumar, and Y.-P. Huang, Efficient reservoir computing using field programmable gate array and electro-optic modulation, OSA Continuum, OSAC **4**, 1086 (2021).

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

[38] J. P. Eckmann, S. O. Kamphorst, D. Ruelle, and S. Ciliberto, Liapunov exponents from time series, Phys. Rev. A **34**, 4971 (1986).

[39] A. F. Agarap, Deep learning using rectified linear units (ReLU), arXiv:1803.08375.

[40] T. P. Lillicrap, A. Santoro, L. Marris, C. J. Akerman, and G. Hinton, Backpropagation and the brain, Nat. Rev. Neurosci. **21**, 335 (2020).

[41] A. Mao, M. Mohri, and Y. Zhong, Cross-entropy loss functions: Theoretical analysis and applications, arXiv:2304.07288.

[42] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, Learning representations by back-propagating errors, Nature **323**, 533 (1986).

[43] M. Refinetti, S. d'Ascoli, R. Ohana, and S. Goldt, Align, then memorise: The dynamics of learning with feedback alignment*, J. Phys. A: Math. Theor. **55**, 044002 (2022).

[44] T. P. Lillicrap, D. Cownden, D. B. Tweed, and C. J. Akerman, Random synaptic feedback weights support error backpropagation for deep learning, Nat. Commun. **7**, 13276 (2016).