Radio-Frequency Multiply-and-Accumulate Operations with Spintronic Synapses

Nathan Leroux,[1,*] Danijela Marković,[1] Erwann Martin,[2] Teodora Petrisor,[2] Damien Querlioz,[3] Alice Mizrahi,[1] and Julie Grollier[1]

[1] *Unité Mixte de Physique, CNRS, Thales, Université Paris-Saclay, 91767 Palaiseau, France*
[2] *Thales Research and Technology, 91767 Palaiseau, France*
[3] *Université Paris-Saclay, CNRS, Centre de Nanosciences et de Nanotechnologies, 91120 Palaiseau, France*

(Received 13 November 2020; revised 23 February 2021; accepted 2 March 2021; published 23 March 2021)

Exploiting the physics of nanoelectronic devices is a major lead for implementing compact, fast, and energy-efficient artificial intelligence. In this work, we propose a strategy in this direction, where assemblies of spintronic resonators used as artificial synapses can classify analogue radio-frequency signals directly without digitalization. The resonators convert the radio-frequency input signals into direct voltages through the spin-diode effect. In the process, they multiply the input signals by a synaptic weight, which depends on their resonance frequency. We demonstrate through physical simulations with parameters extracted from experimental devices that frequency-multiplexed assemblies of resonators implement the cornerstone operation of artificial neural networks, multiply and accumulate (MAC), directly on microwave inputs. The results show that, even with a nonideal realistic model, the outputs obtained with our architecture remain comparable to that of a traditional MAC operation. Using a conventional machine-learning framework augmented with equations describing the physics of spintronic resonators, we train a single-layer neural network to classify radio-frequency signals encoding 8×8 pixel handwritten-digit pictures. The spintronic neural network recognizes the digits with an accuracy of 99.96%, equivalent to purely software neural networks. This MAC implementation offers a promising solution for fast low-power radio-frequency classification applications and another building block for spintronic deep neural networks.

DOI: 10.1103/PhysRevApplied.15.034067

I. INTRODUCTION

Radio-frequency (rf) signals are ubiquitous today [1]. Finding ways to automatically recognize and classify these signals is important for numerous applications, such as medicine [2–4], rf fingerprinting [5], gesture sensing [6], radar applications [7], and aerial vehicle detection and identification [8]. Artificial neural networks are proven to be more accurate and more resilient to real-world conditions (noisy electromagnetic environment, imperfect rf components or antennas etc.) than more conventional algorithms that rely on specific feature extractors and complex analysis tools [1]. Currently, applying artificial neural networks to rf signals requires first digitizing the signal sensed by the antenna and then using and running a neural network on conventional CMOS-based hardware (such as a central processing unit, graphics processing unit (GPU), field-programmable gate array, or application-specific integrated circuit). Both stages of the process are computationally heavy, leading to delays (a few miliseconds) and high power and energy consumption (hundreds of watts) [9]. To decrease the size and dependency on cloud computing of embedded rf devices, it is thus essential to build fast and low-power systems that integrate both rf signal analyzers and *in situ* artificial intelligence accelerators.

Presently, the most promising artificial intelligence algorithms are based on deep neural networks [10], which contain several layers of artificial neurons, each of them linked by synaptic connections: in each layer of an artificial neural network, the neuron signals are multiplied by synaptic weights, summed, and injected into a neuron of the following layer [see Fig. 1(a)]. This elementary operation is called multiply and accumulate (MAC). In a computer using the von Neumann architecture, weight multiplications and sums are performed by processing units, whereas synaptic weight values are stored in spatially separated memory units. In such an architecture, the data flow between the processing and memory units induces a slowdown and excess energy consumption [11] that can be avoided by implementing the MAC operation in hardware, using *in situ* memory devices emulating neurons and synapses [12–17]. Neurons that take dc inputs and convert them into microwave signals are demonstrated using spintronic nano-oscillators [18–23] and CMOS ring oscillators [24,25]. However, to date, there is no demonstration

*nathan.leroux@cnrs-thales.fr

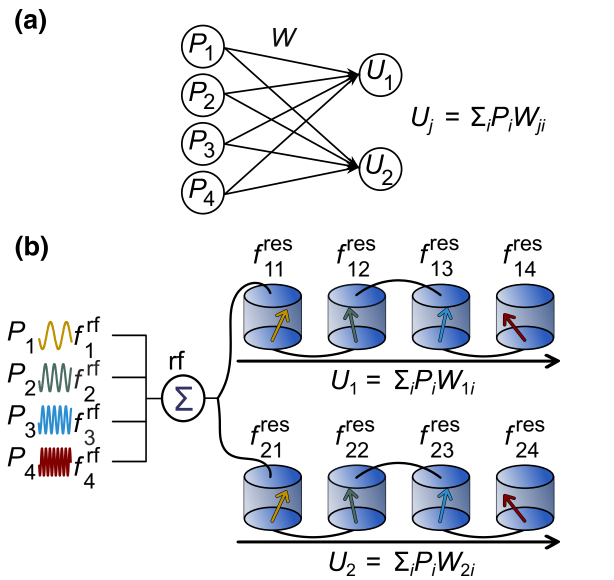

FIG. 1. (a) Multiply-and-accumulate operation: neural signals P_1, P_2, P_3, and P_4 are multiplied by different synaptic weights W_{ji} and summed. (b) Multiply-and-accumulate operation with different radio-frequency signals sent simultaneously in two chains of resonators: each resonator rectifies mostly one of the input signals, hence multiplying it by a weight. Chain voltages are the sum of all their resonator voltages.

of tunable artificial synapses that directly perform MAC operations on microwave signals.

Here, we show through theoretical analysis and numerical simulations that spintronic resonators, which are devices similar to spintronic oscillators, may be used as rf nanosynapses. These resonators apply synaptic weights to microwave signals through the spin-diode effect [26–28]: because of magnetic resonance, when a rf current passes through a spintronic resonator, the resistance of the resonator is forced to oscillate at the same frequency as the rf current, thus rectifying the input power. The amplitude of the rectified voltage depends on the difference between the frequency of the rf signal and the frequency of the device resonance [26–29]. We can thus use spintronic resonators as artificial synapses that take microwave power as an input, rectify direct voltage as an output, and store synaptic weights encoded in their resonance frequency.

We first explain the rf neural-network architecture with inputs encoded in microwave powers and show how frequency multiplexing can be used to make a simple and compact implementation of the MAC operation. We explain how the spintronic resonators can implement tunable synaptic weights through the spin-diode effect. To be realistic, we have to consider the nonlinear behavior of spintronic resonators that may, in principle, represent a challenge. Physical simulations using parameters extracted from experimental devices demonstrate that our architecture is equivalent to a MAC operation, even when the resonator nonlinearities are considered. Furthermore, the response of a resonator under the superimposition of multiple rf signals with different frequencies is nontrivial. We demonstrate through dynamical simulations that we can simulate this response with a simple analytical model. This allows us to highlight the fundamental requirements for this rf MAC proposal. Finally, we show through simulations of a complete network that a single layer of our rf neural network can be used to classify microwave signals encoding handwritten-digit images, with an accuracy equivalent to software MAC.

II. PRINCIPLE OF RESONATOR-BASED MAC OPERATIONS ON RADIO-FREQUENCY SIGNALS: STUDY IN IDEALIZED CONDITIONS

As represented in Fig. 1(a), the multiply-and-accumulate operation is a weighted sum of N input values. In our proposal, these N input values are encoded in the microwave powers P_i of N rf signals of index i, each with a different frequency. A MAC operation is performed by sending these N rf signals simultaneously to a chain of N resonators, indexed by k, wired in series [Fig. 1(b)]. A neural network with M outputs requires M different resonator chains, indexed by j. The goal of this section is to show that the voltage across each chain j can be seen as a weighted sum of the input microwave powers P_i:

$$U_j = \sum_{i=0}^{N-1} P_i W_{ji}, \tag{1}$$

where W_{ji} is the synaptic weight between the input i and the output j, as determined by the physics of the spin-diode effect. All resonators in a chain j can contribute to W_{ji}, depending on their frequency distribution and the frequency of the input signals.

Following the universal and experimentally validated [30,31] auto-oscillator model described by Slavin and Tiberkevich [29], the rectification voltage of an ideal spintronic resonator k in chain j submitted to a rf power P_i with angular frequency $\omega_i^{\rm rf}$ due to the spin-diode effect, as displayed in Fig. 2(a), is

$$\begin{aligned} v_{kji} &= P_i G(\Delta f_{kji}) = P_i \frac{\omega_i^{\rm rf} - \omega_{kj}^{\rm res}}{\Gamma_{kj}^{\rm res2} + (\omega_i^{\rm rf} - \omega_{kj}^{\rm res})^2}\beta, \\ \omega_{kj}^{\rm res} &= 2\pi f_{kj}^{\rm res}, \\ \Gamma_{kj}^{\rm res} &= \alpha\omega_{kj}^{\rm res}, \end{aligned} \tag{2}$$

where $\omega_{kj}^{\rm res}$ is the angular frequency of the resonance; $\Gamma_{kj}^{\rm res}$ is the resonance line width; α is the magnetic damping; Δf_{kji} is the frequency mismatch $\Delta f_{kji} = f_i^{\rm rf} - f_{kj}^{\rm res}$; and β is a factor that depends on several characteristics of the resonators, for instance, the magnetoresistance, if they are magnetic tunnel junctions [18].

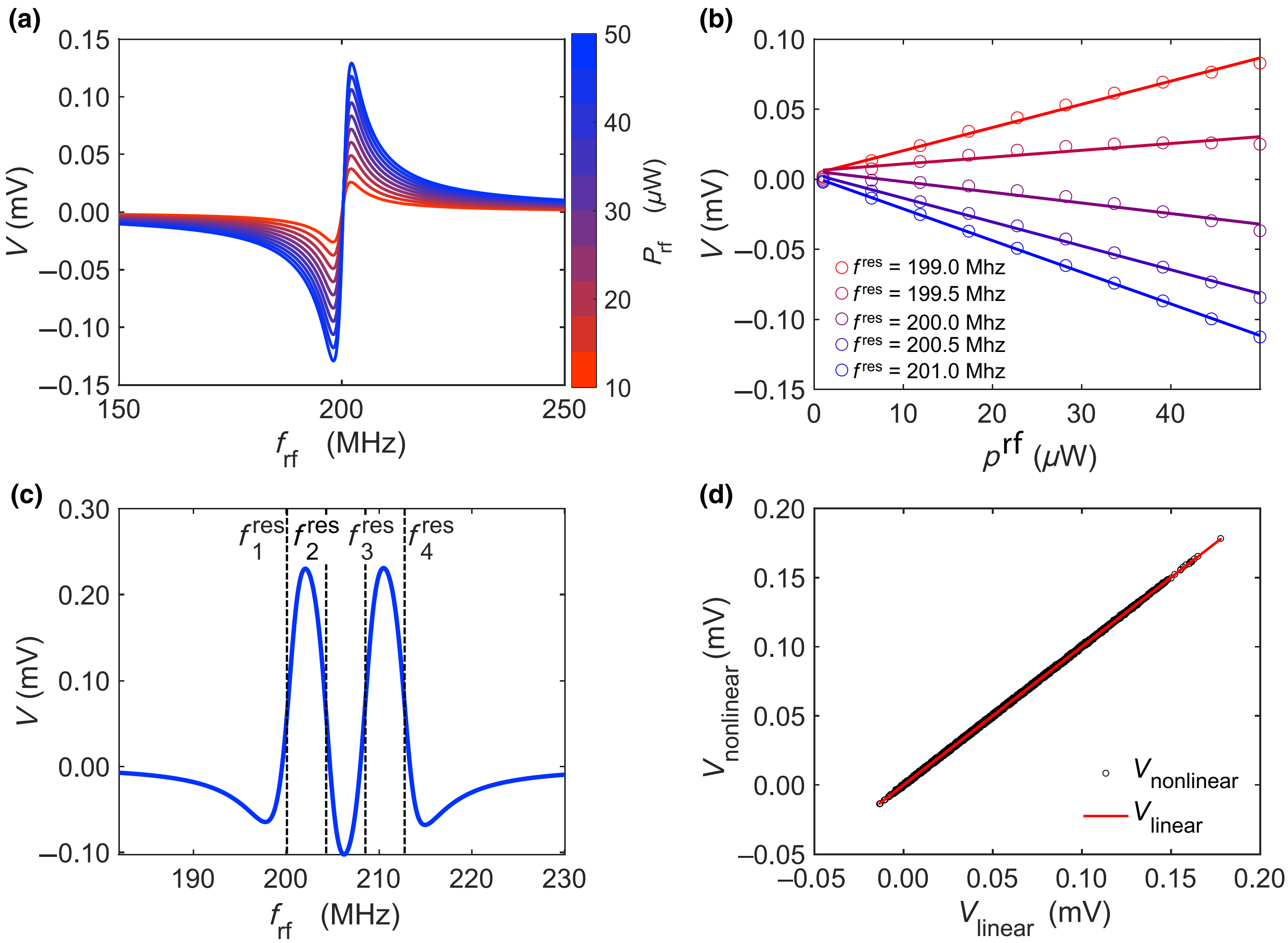

FIG. 2. Analytical calculations based on the realistic spin-diode model described by Eqs. (2) and (6). (a) Spin-diode rectified voltage of a spintronic resonator with resonance frequency $f^{\rm res} = 200$ MHz versus the frequency of an input radio-frequency signal for microwave powers between 10 and 50 μW (in color scale). (b) Circles are the spin-diode voltage of a spintronic resonator with different resonance frequencies versus the microwave power of the rf signal at 200 MHz. Solid lines are linear fits. (c) Spin-diode voltage of a chain of four spintronic resonators wired head-to-tail with resonance frequencies $f^{\rm res} = 200$ Hz, 204.0 MHz, 208.2 MHz, and 212.4 MHz, versus the frequency of a radio-frequency signal with a power of 50 μW. (d) Calculations for the same chain with four different radio-frequency signals for 6561 different combinations of microwave powers for the radio-frequency signals (5, 10, and 15 μW) and different resonance frequencies for the resonators. Scatter dots are the voltages of the calculations with nonlinear resonators plotted against the voltages of the calculations with ideal linear resonators. Red solid line corresponds to calculated voltages with ideal linear resonators plotted against themselves. Root-mean-square deviation between scatter dots and the red solid line is 0.58 μV and the correlation is 99.98%.

Because there are often different torques acting on the magnetization dynamics, the spin-diode effect usually leads to the sum of a symmetric and an antisymmetric part in the rectified voltage [27]. For instance, in Magnetic Tunnel Junctions, the fieldlike torque usually leads to an antisymmetric component and the spin-transfer torque usually leads to a symmetric part. For clarity, we choose to focus on the antisymmetric part in this paper. Results of simulations, including the symmetric part, are presented in Sec. V.

Each resonator receives simultaneously the N rf input signals. We show in Sec. IV that the resulting rectified voltage is the sum of the dc voltages they generate when they receive each rf signal individually. Therefore, the rectified voltage across each chain is

$$U_j = \sum_{k=0}^{N-1}\sum_{i=0}^{N-1} v_{kji}. \tag{3}$$

In this work, we propose to wire the resonators of a chain in a head-to-tail configuration, as depicted in Fig. 1(b), to cancel the voltage offsets at low frequency [$\omega_i^{\rm rf} \to 0$ in Eq. (2)]. Indeed, the rf signal of index i is rectified into a positive voltage by all resonators of index $k > i$. All these offset voltages would accumulate and become larger than the resonant signals useful for synaptic operations. But, if we wire the resonators head-to-tail, then the rf signal of index i is rectified into a positive voltage by all resonators of even indexes $k > i$ and into a negative voltage by all resonators of odd indexes $k > i$. Hence, the offset generated by the rf signal of index i is approximately compensated for, and it remains mostly the voltage rectified by the resonator of index $k = i$. The combination of Eqs. (2) and (3) therefore gives

$$U_j = \sum_{k=0}^{N-1} \sum_{i=0}^{N-1} P_i G(\Delta f_{kji})(-1)^k, \tag{4}$$

where the factor $(-1)^k$ accounts for the head-to-tail wiring. This naturally leads to Eq. (1), with synaptic weights equal to

$$W_{ji} = \sum_{k=0}^{N-1} G(\Delta f_{kji})(-1)^k. \tag{5}$$

Spintronic resonators are frequency selective. As can be seen in Fig. 2(a) and Eq. (2), the rectification voltage drops to zero when ω_i^{rf} tends toward infinity and to a small offset (2α times smaller than the maximum voltage) when ω_i^{rf} tends toward zero. For operating a resonator chain as a useful neural network, each resonator in a synaptic chain should be chosen to have a resonance frequency matching the frequency of one of the input signals, so that this resonator features a greater rectification effect on this matching signal. For instance, in Fig. 1(b), the resonator with resonance frequency f_{12}^{res} receives the four rf signals, but rectifies the signal with frequency f_2^{rf} most effectively. When using the synaptic chain in this configuration, each synaptic weight can be approximated, leading to a simplified expression of Eq. (5):

$$W_{ji} = G(\Delta f_{iji})(-1)^i.$$

This simplified equation highlights that it is possible to tune each synaptic weight W_{ji} by tuning the resonance frequency of the resonator indexed by $k = i$, which plays the role of a synaptic connection between input i and output j [see Fig. 1(a)]. Zahedinejad *et al,* [32] demonstrated voltage-gated memristive control of the perpendicular magnetic anisotropy at a ferromagnetic/oxide interface, leading to nonvolatile control of the oscillating properties of a spin Hall nano-oscillator. Such memristive control of the magnetic properties of a spintronic resonator could allow tuning of the resonance frequencies in future experimental implementations. To send the sum of the rf signals into different chains of resonators, like in Fig. 1(b), it is possible to build micrometer-scale rf combiners with low-power dissipation, as reported in Ref. [24].

III. MULTIPLY-AND-ACCUMULATE SIMULATION RESULTS INCORPORATING DEVICE NONLINEARITIES

We now quantify the accuracy of the spintronic-resonator-based MAC operation compared with an ideal one. Spintronic resonators have an intrinsic nonlinear dependence of their frequency and line width on the magnetization oscillation amplitude, which is typically expressed as [29]

$$\begin{aligned} f^{\mathrm{res}}(p) &= f^{\mathrm{res}}(0)(1 + Np), \\ \Gamma^{\mathrm{res}}(p) &= 2\pi f^{\mathrm{res}}(0)(1 + Qp), \end{aligned} \tag{6}$$

where N and Q are nonlinear parameters, and p is the normalized magnetization oscillation power, equal to the square of the magnetization oscillation amplitude. According to Ref. [29], p can be expressed as

$$p = \frac{P^{\mathrm{rf}}}{\Gamma^{\mathrm{res}}(p)^2 + [\omega^{\mathrm{rf}} - \omega^{\mathrm{res}}(p)]^2}\gamma^2, \tag{7}$$

where γ is a proportionality factor between the amplitude of the rf signal and the amplitude of the torque acting on the resonator magnetization. This means that, in real devices, the dependence of v_{kji} with the input power $P^{\mathrm{rf}} = P_i$ in Eq. (2) is not perfectly linear, as $\omega_{kj}^{\mathrm{res}}$ and $\Gamma_{kj}^{\mathrm{res}}$ both depend on P_i. In other words, the weights W_{ji} depend on the inputs; this does not correspond to the usual mathematical description of neural networks. This effect, called weight nonlinearity, is an issue for learning in hardware neural networks utilizing nanodevices, such as memristors, to implement MAC operations through physical phenomena.

To quantify this effect, we choose parameters extracted from experiments on similar structures as the resonators. We take nonlinear coefficients $N = 0.1$ and $Q = 1$, which are close to values determined experimentally in studies on spin-torque nano-oscillators [29,33–35]. We choose frequencies close to 200 MHz and parameters $\beta = 1.7 \times 10^6\ C^{-1}$ and $\gamma = 7.1 \times 10^7$ Hz $\mathrm{W}^{-1/2}$, according to experimental values from prior work [33]. Finally, we take a damping parameter of $\alpha = 0.01$, corresponding to Permalloy. It is important to note that these parameters are only used as a guideline, as we can also use spin-orbit torque or an Oersted field instead of spin-transfer torque, and the materials and the geometry of the spintronic resonators can also be different. These different devices would follow the same model, but with different parameters. It is even possible to reduce the nonlinear parameters N and Q using a specific fabrication process [34] or geometry [36].

We plot in Fig. 2(b) the dependence of the spin-diode voltage of a single resonator as a function of the microwave power of a rf signal, for different resonance frequency values using the analytical model of Eqs. (2)–(8). We see that, despite the nonlinearities N and Q, the voltage response can be fitted linearly with the microwave power, indicating that the dependence of the corresponding weight on the input power remains small. Figure 2(b) also shows that the slope can be controlled by tuning the resonance frequency of each resonator. This result confirms that, with realistic devices, the synaptic weights of a resonator-based neural network can be tuned by changing the resonance frequencies of the resonators.

We now compare the resonator-based MAC operation to a perfectly linear MAC operation. We consider four input rf signals of frequencies $f^{\rm rf} = 200.0$, 204.0, 208.2, and 212.4 MHz and simulate a chain of four different resonators, as illustrated in Fig. 2(c), with $N_{\rm powers} = 3$ different input powers (5, 10, and 15 μW) for each rf signal and $N_{\rm frequencies} = 3$ different resonance frequencies for each resonator, resulting in a set of $N_{\rm powers}^{N^{\rm rf}} N_{\rm frequencies}^{N^{\rm res}} = 6561$ different combinations.

This result gives us the performance of the nonlinear MAC operation. We then need to define a reference ideal MAC operation to evaluate the results. For this purpose, for every point, we compute the magnetization oscillation powers of the four diodes with Eq. (7) and store the maximum for each resonator and for each rf signal: $p_{ki}^{\max} = \max_{\rm realMAC}(p_{ki})$. These maximum oscillation power values serve as a reference to simulate a MAC operation with a chain of four ideal linear spintronic resonators, the resonance frequency and line width of which does not depend on the input rf power. This approach gives the following linear reference model for the MAC operation:

$$U^{\rm linear} = \sum_{k=0}^{N-1} \sum_{i=0}^{N-1} P_i \frac{\omega_i^{\rm rf} - \omega_k^{\rm res}(0)(1 + N p_{ki}^{\max})}{\Gamma_{kj}^{\rm res}(0)^2 (1 + Q p_{ki}^{\max})^2 + [\omega_i^{\rm rf} - \omega_k^{\rm res}(0)(1 + N p_{ki}^{\max})]^2} \beta. \tag{8}$$

For each diode and for each rf signal, using a single value of the magnetization oscillation power makes the model linear. We choose these values to be the maximum $p_{ki}^{\max}$ to make the output of the linear model match, as far as possible, the output of the realistic model. We repeat the same set of 6561 different calculations to those performed with the realistic model (same sweeps of power for the four rf signals and same sweeps of resonance frequencies for each resonator), but this time with the linear model described by Eq. (8). We can then compare the realistic model, including nonlinearities, to a model where the synaptic weights do not depend at all on the input of the synaptic layer.

In Fig. 2(d), we plot the voltage of the nonlinear MAC simulations as a function of the voltage of the linear MAC. We see that the scatter plot thus created is aligned with the $y = x$ curve with a root-mean-square deviation of 0.58 μV. This result shows that the MAC implemented by a chain of spintronic resonators is comparable to a linear MAC when the nonlinear coefficients N and Q are less than or equal to 0.1 and 1, respectively, and thus, that spintronic resonators can be used as artificial synapses for neural networks.

IV. VALIDATION OF THE MODEL FOR MULTIPLE RF SIGNAL SUPERIMPOSITION

To simulate the effect of sending multiple rf signals simultaneously through a chain of spintronic resonators, we suppose that the voltage rectified by each resonator is the sum of the voltages it would rectify for each rf signal received individually. This assumption is based on the hypothesis that a spintronic resonator can oscillate simultaneously at different frequencies if it receives different rf signals. To demonstrate this assumption, we analyze the magnetization motion of a spintronic resonator under the influence of different rf signals. We use an ordinary differential equation (ODE) solver to compute the solution of the system of equations of magnetization dynamics [29]:

$$\begin{aligned} \frac{dp^{\rm res}}{dt} &= -2\Gamma^{\rm res}(p^{\rm res})p^{\rm res} \\ &\quad + 2\sqrt{p^{\rm res}} \sum_i^N F_i^{\rm rf} \cos\left(\varphi^{\rm res} - \psi_i^{\rm rf} + \omega_i^{\rm rf} t\right), \\ \frac{d\varphi^{\rm res}}{dt} &= -\omega^{\rm res}(p^{\rm res}) - \frac{1}{\sqrt{p^{\rm res}}} \sum_i^N F_i^{\rm rf} \sin(\varphi^{\rm res} - \psi_i^{\rm rf} + \omega_i^{\rm rf} t), \end{aligned} \tag{9}$$

where $p^{\rm res}$ and $\varphi^{\rm res}$ are the normalized oscillation power and the phase of the resonator, respectively; $\psi_i^{\rm rf}$ are the frequencies and phases of incoming rf signals; and $F_i^{\rm rf}$ is the amplitude of the torque exerted on magnetization.

We simulate one spintronic resonator with a frequency of $f^{\rm res} = 200$ MHz, a random initial phase of $\psi^{\rm res}$, and four rf signals with the same amplitude of microwave torque of $F^{\rm rf} = 0.2 \times 2\pi$ rad MHz; four different frequencies of $f_1^{\rm rf} = 200.0$ MHz, $f_1^{\rm rf} = 204.0$ MHz, $f_1^{\rm rf} = 208.2$ MHz, and $f_1^{\rm rf} = 212.4$ MHz; and four random initial phases of $\psi_1^{\rm rf}$, $\psi_2^{\rm rf}$, $\psi_3^{\rm rf}$, and $\psi_4^{\rm rf}$. We then compare the horizontal component of magnetization, $m_x(t) = \sqrt{p^{\rm res}(t)}\cos[\varphi^{\rm res}(t)]$, with our model,

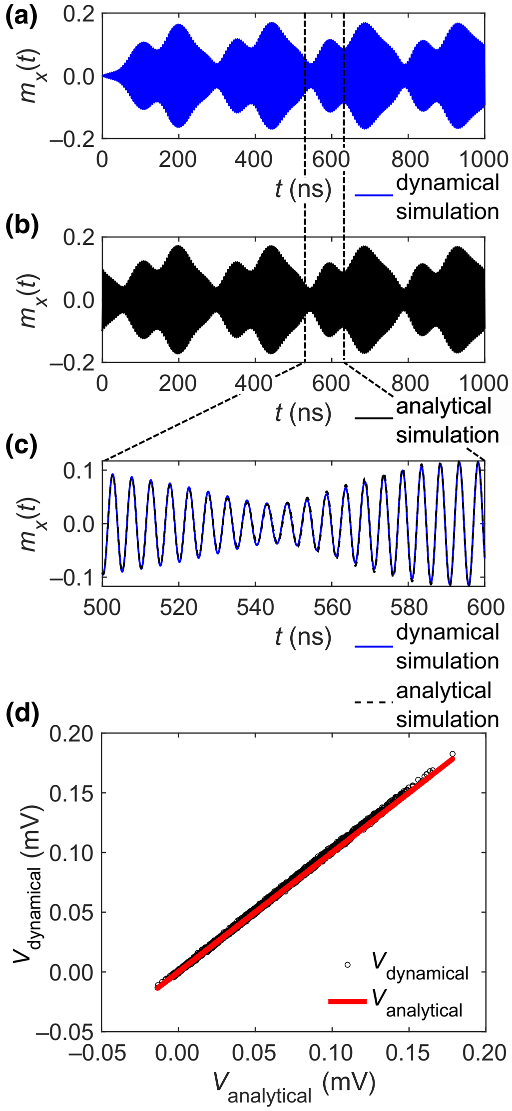

FIG. 3. (a) Horizontal component of a spintronic resonator magnetization with four different radio-frequency signals simulated with an ordinary differential equation solver. (b) Theoretical model $m'_x(t) = \sum_i^N \sqrt{p_i^{\prime \mathrm{res}}}\cos(\psi^{\mathrm{rf}} - \omega_i^{\mathrm{rf}} t)$. (c) ODE simulations (blue solid line) and theoretical model (black dashed line). (d) Simulation of four different radio-frequency signals sent through a chain of four different spintronic resonators for 6561 different combinations of microwave powers for the radio-frequency signals (5, 10, and 15 μW) and different resonance frequencies for the resonators. Scatter dots are the voltages of the ODE simulations plotted against the voltages of the calculations with theoretical model. Simulations are averaged over 20 repetitions, each with a random initialization for the rf signal phases. Red solid line corresponds to voltages of simulations realized with the theoretical model plotted against themselves. Root-mean-square deviation between the scatter dots and the red solid line is 1.72 μV and the correlation is 99.94%.

$m'_x(t) = \sum_i^N \sqrt{p_i^{\prime \mathrm{res}}}\cos(\psi_i^{\mathrm{rf}} + \psi_i^{\mathrm{relaxation}} - \omega_i^{\mathrm{rf}} t)$. Here, each of the normalized oscillation powers is calculated using Eq. (7) for a spintronic resonator receiving a single rf signal with microwave power P_i^{rf}, frequency f_i^{rf}, and initial phase ψ_i^{rf}. We consider that the resonator in resonance oscillates at the frequency of the rf signal it receives, f_i^{rf}. The phases $\psi_i^{\mathrm{relaxation}}$ result from the transient dynamics that occurs during the relaxation period (period until magnetization is periodic); they are determined by fitting the dynamical simulation results to the analytical model. In Fig. 3(a), we see that, after this relaxation period, the magnetization dynamics of a spintronic resonator with multiple rf signals corresponds perfectly to our model. Then, the resistance oscillations mixed with the rf signals give

$$
\begin{aligned}
V\left(\sum_i^N S_i^{\mathrm{rf}}\right) &= \sum_i^N S_i^{\mathrm{rf}} r(t) \\
&= \sum_i^N I_i^{\mathrm{rf}} \cos(\psi_i^{\mathrm{rf}} - \omega_i^{\mathrm{rf}} t) \sum_i^N R_{P-\mathrm{AP}} \sin[\varphi_i^{\mathrm{res}}(t)] \\
&\approx \sum_i^N R_{P-\mathrm{AP}} I_i^{\mathrm{rf}} \sin[\varphi_i^{\mathrm{res}}(t) - \varphi_i^{\mathrm{rf}}(t)] \\
&= \sum_i V(S_i^{\mathrm{rf}}),
\end{aligned}
$$

which confirms our assumption. We also repeat all simulations from Sec. III with the ODE method. In Fig. 3(b), we compare the voltage of a chain of four resonators simulated with the ODE method and the voltage of the same chain simulated using the analytical model. The voltages are averaged over 20 repetitions of simulations; each time the initial rf signal phases are initialized randomly. The results show that the two models are correlated at 99.94%. These simulations show that it is valid to consider that the effects of multiple input rf signals simply sum at the resonator level. They validate the use of the analytical model from Sec. III in neural-network simulations.

V. HANDWRITTEN-DIGIT RECOGNITION WITH A SINGLE-LAYER MICROWAVE NEURAL NETWORK

In this last section, we prove that we can teach a radio-frequency-based neural network to classify microwave-encoded inputs by tuning the resonance frequencies of spintronic resonators, hence demonstrating that the system is able to process rf signals and to directly apply MAC operations on them. To test the efficiency of our implementation of MAC operations for neural networks, we choose a standard task of image classification, for which the goal is to recognize handwritten digits from zero to nine.

We first consider a dataset called “digits” comprising of 1797 images of $8 \times 8 = 64$ pixels. The dataset is split in two: three quarters of the images are used for neural-network training and one quarter is used for testing. The goal for the network is to classify each image between zero and nine. The network inputs are encoded into 64 rf signals: the brighter the pixel, the higher the rf signal power. The sum of the 64 rf signals is sent into 10 chains of 64 resonators, and the voltages of the 10 synaptic chains are the outputs of the network. The choice of frequencies of the rf signals, the microwave power scaling, and the initialization of the spintronic resonator frequencies are discussed in the Appendix.

To train the network to classify these handwritten-digit images, we use PyTorch, software that allows the implementation of backpropagation, which is the most commonly used algorithm for neural-network training [10]. This supervised algorithm propagates the gradient of a loss function across a neural network, so that, for each iteration, the weight updates $\mathbf{W} \leftarrow \mathbf{W} - \eta(\partial L/\partial \mathbf{W})$ reduce the loss L, which is the error between the predictions of the network and the targets, i.e., the classification labels assigned to each input. η is a learning-rate coefficient that we initialize empirically at $\eta = 10^{-4}$. We use the optimizer Adam [37]. The loss is calculated by simulating the voltages, U_j, of the 10 synaptic chains and applying the cross-entropy loss function

$$L(\mathbf{y}, \mathbf{U}) = -\sum_{j=0}^{9} y_j \ln\left[\frac{\exp(U_j)}{\sum_{j=0}^{9}\exp(U_j)}\right], \quad (10)$$

where y_j are the targets.

At each iteration, we present a batch of 16 pictures to the network and use Eqs. (2) and (5) with resonator nonlinearities to compute the network output. The loss for each picture of the batch is computed and averaged. We then compute the gradient of the loss with respect to the 64×10 weights. To find the updates for the resonance frequencies, using the full nonlinear equations leads to an inefficient backpropagation algorithm because of the dependencies between the synaptic weights and the inputs. However, as the weight changes provoked by backpropagation are by construction small, it is possible to compute them using linearized equations. Therefore, instead of using the model with nonlinear resonators, we use the model with linear resonators defined by Eq. (8), which is initialized with the same parameters as those of the model with nonlinear resonators. To define the reference $p^{\max}$, we compute the maximum of magnetization oscillation power for each resonator at initialization for a maximum input (white image, i.e., all the pixel values are one). Then, we update the resonance frequencies of the linear model resonators using the weight gradient with respect to the resonance frequencies:

$$f^{\text{res}} \leftarrow f^{\text{res}} - \eta \frac{\partial W}{\partial f^{\text{res}}}\frac{\partial L}{\partial W}. \quad (11)$$

In the next iteration, we take the resonance frequencies that are updated for the linear model and use them in the realistic model.

To complete the training procedure, we perform 20 epochs, meaning that we present the entire dataset (training on three quarters and testing on one quarter) 20 times, and we repeat the entire procedure 10 times to gather statistics. To compute the success rate, i.e., the proportion of images in the dataset that the network is able to classify, we take the class that corresponds to the chain index j, the output of which is maximum, and we compare it with the target class of the dataset. The purple line in Fig. 4(d) shows the mean success rate as a function of the epoch number, and the mean deviation is indicated by purple shading. The mean success rate at the end of training reaches 99.96% for both the test and training sets. Looking at the standard deviation in purple shading, we see that the result is reproducible: if the result is stochastic for the first epochs, the outcome always converges. We perform exactly the same simulations, but include the symmetric part for the spin-diode effect with a ratio of 0.5; this means that in Eq. (2) there will be a symmetric part with half the amplitude of the antisymmetric part. For this configuration, we achieve 99.84% success rate. We perform classification for the same task with classical software for a neural network trained with backpropagation on an equivalent architecture (64 inputs fully connected by synapses to the 10 outputs). The success rate of the neural-network software [blue line in Fig. 4(d)] is equivalent to classification with the resonator network. This result shows that it is possible to train a network made of chains of spintronic resonators by tuning their resonance frequency to classify microwave-encoded signals. The training algorithm we develop can also be used to train an experimentally constructed spintronic-resonator-based neural network.

Finally, we want to prove that we can scale this type of network to chains comprising of several hundreds of spintronic resonators that are very large. The goal is to check that cross talk due to the superimpositions of many rf signals does not prevent learning. To do so, we solve the "Modified National Institute of Standards and Technology (MNIST)" dataset, which is similar to the digits dataset, but with 70 000 pictures of $28 \times 28 = 784$ pixels. Then the network is made of 10 chains of 784 spintronic resonators. We use the same algorithm to train this network as that we use for the digits dataset, but with a learning rate of $\eta = 5 \times 10^{-6}$ and batches of 500 pictures. Due to the large RAM the simulations require, it is not possible to implement the nonlinear behavior of these spintronic resonators for this study.

We arrange the 784 rf frequencies between 50 MHz and 20 GHz as explained in the Appendix. For a fixed-frequency arrangement, the cross talk increases with the line width of the spintronic resonator, which is proportional to magnetic damping [see Eq. (2)]. On the other hand, higher magnetic damping helps to stabilize the orbit of magnetic oscillations faster and leads to higher computational speed. It is possible to engineer the spintronic

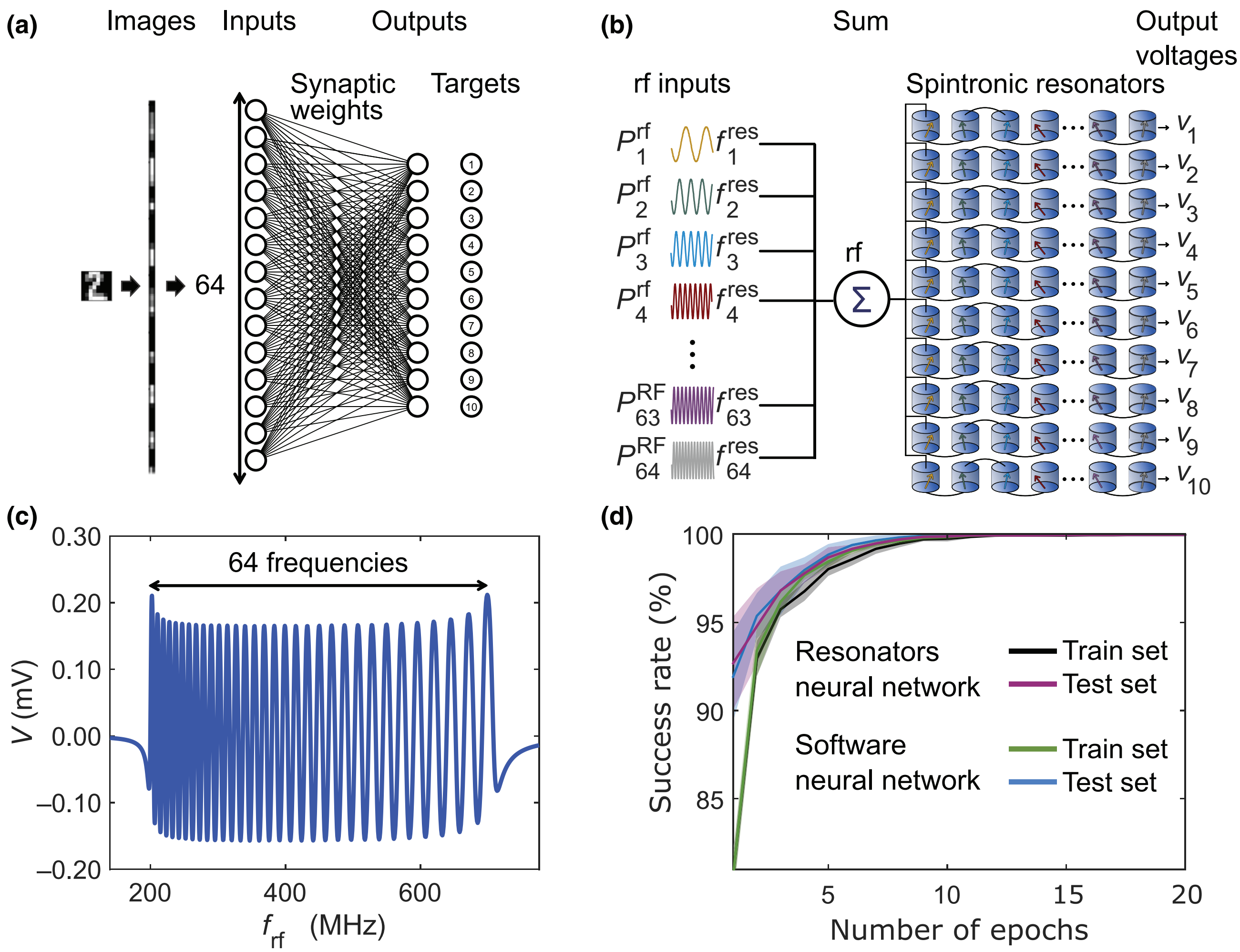

FIG. 4. (a) Classical neural-network architecture to solve the digits dataset. From left to right: 8×8 pixel input images, 64×1 flattened input layer, synaptic layer connecting the input to 10 outputs, and a comparison of outputs with the targets. (b) Equivalent radio-frequency spintronic-synapse-based neural-network architecture. From left to right: 8×8 pixel input images and 64×1 flattened input layer; each input is encoded in the microwave power of a radio-frequency signal with a different frequency. All 64 signals are summed and sent to 10 chains of 64 resonators wired in series head to tail. Each resonator rectifies its matching frequency signal, thus applying a synaptic weight to it. Output voltages are compared with targets. (c) Analytical simulations of the spin-diode voltage of a chain of 64 spintronic resonators wired head to tail versus the frequency of a rf signal with a power of 50 μW. First resonator has a resonance frequency of $f_0^{\mathrm{res}} =$ 200 MHz and others are arranged following Eq. (A1) of the Appendix. (d) Percentage of successful classifications versus number of epochs. Black (purple) corresponds to the results for the training (test) set for the resonator neural network and green (blue) is for the training (test) set for the equivalent regular neural-network software. Lines (dashed lines for neural-network software) represent the mean success rates and the shading indicates standard deviations. Success rate reaches 99.96% for both for the neural-network software and resonator-based neural network for training and test sets.

resonators with different materials to change their magnetic damping.

Figure 5, we plot the success rate on the MNIST dataset with a layer of these spintronic resonators for different magnetic damping after 20 epochs. The results are averaged over 10 repetitions. For the results on the training set (again one quarter of the dataset), we find that the maximum recognition rate is 99.40% for a magnetic damping of $\alpha = 0.0188$. For this magnetic damping value, the line width is comparable to the separation coefficient (see the Appendix): cross talk is not dominant and does not prevent the network from learning and separating different classes of inputs. For comparison, we solve the same dataset with neural-network software of the same architecture and achieve, at best, 92.27% recognition. We see that the accuracy of the resonator neural network decreases strongly for $\alpha > 0.1$. This is because, if the magnetic damping is far greater than the separation coefficient, different resonance curves overlap and cross talk degrades the classification.

To discuss the relationship between this magnetic damping and the computational speed, we have to consider that each iteration is limited by the speed of the slowest

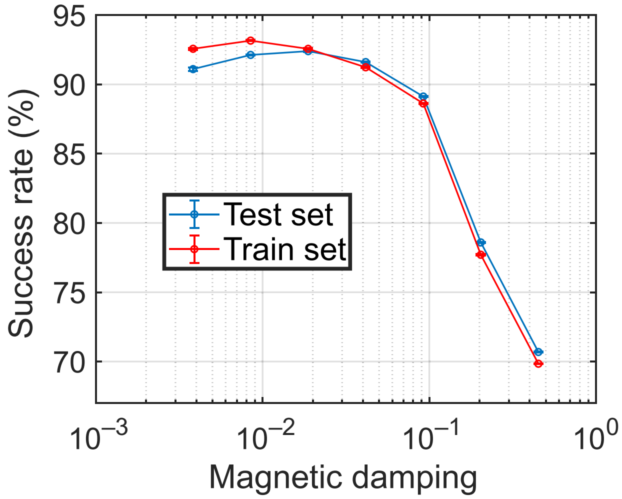

FIG. 5. Percentage of successful classifications of a single-layer resonator network on the MNIST dataset versus magnetic damping of the material used for spintronic resonators ($\log_{10}$ scale). In blue (red) are plotted the results for the test (training) set. Results are averaged over 10 repetitions and the error bar corresponds to the root-mean-square deviation.

spintronic resonator, which, in this case, is 50 MHz. To estimate the relaxation time of this spintronic resonator, we use an ordinary differential equation solver, as we did in Sec. IV to solve Eq. (9). We fit the horizontal component of the magnetization dynamics to an exponential decay model:

$$m_x^{\text{decay}}(t) = \sqrt{p^{\text{res}}(t)}\left(1 - e^{-t/\tau}\right)\cos[\varphi^{\text{res}}(t)].$$

Using this technique, we extract that the relaxation time, τ, is equal to 194 ns when $\alpha = 0.01$ (Permalloy) and τ is equal to 20 ns when $\alpha = 0.1$, which sets the limiting speed for the computation.

It is currently difficult to compute the power consumption of such a system because it would depend strongly on electronic implementation and on the choice of spintronic resonators. We can still estimate that the minimum consumption per device due to Joule heating is 1 μW [35]. Since here we use $10 \times 784 = 7840$ spintronic resonators, the minimum consumption of this resonator neural network is about 8 mW. It is three orders of magnitude less than the power consumption of a traditional GPU to solve the MNIST dataset [38].

VI. CONCLUSION

This work shows theoretically and numerically that it is possible to build a synaptic layer made of chains of spintronic resonators, with each resonator emulating a synapse and storing a synaptic weight in its resonance frequency. We demonstrate that the MAC operation thus created is equivalent to a typical software MAC operation and is able to classify analog rf signals directly without digitalization. We verify the validity of these results with a realistic model, considering the nonlinear behavior of the resonators. We prove that it is possible to train a network of these resonators receiving microwave-encoded inputs by changing their resonance frequencies and we achieve software-equivalent recognition on the digits database. Electric field control of magnetism allows the possibility of having a nonvolatile-voltage control of these resonance frequencies [32]. Finally, we discuss the issues of cross talk due to superimposition of multiple rf signals. Thanks to simulations on the MNIST dataset, we show that our network can scale at least to 784 resonators per chain, which shows that our architecture is competitive with memristor crossbar arrays. Building such a large array of spintronic resonators will necessitate optimization of the transmission of rf signals in the resonator chains and the addressing of impedance mismatches. Our concept of using spintronic resonators and frequency multiplexing provides a fast, compact, and low-power solution to process radio-frequency-encoded information with artificial intelligence methods.

ACKNOWLEDGMENTS

This work is supported by the European Research Council ERC under Grant No. bioSPINspired 682955, the French ANR project SPIN-IA (Grant No. ANR-18-ASTR-0015), and the French Ministry of Defense (DGA). The authors thank Axel Laborieux and Tifenn Hirtzlin for their scientific support.

APPENDIX: FREQUENCY ARRANGEMENT AND AMPLIFICATION FOR FREQUENCY MULTIPLEXING

To prevent the resonators from rectifying several rf signals simultaneously, we have to space their frequencies. They should be arranged in a manner in which the whole frequency range is not too wide (spintronic resonators can cover a finite frequency range between a few tens of MHz and a few tens of GHz [30,36,39]) but with resonances overlapping with each other as little as possible. We have to consider that, for a specific type of spintronic resonator, a higher frequency will result in a wider line width of the resonance [see Eq. (2)]. This is because the line width scales with the frequency and the magnetic damping, α. We want to choose a coefficient of separation that optimizes the spacing between the resonance frequency of the resonators. We define the coefficient of separation, μ, so that $f_{i+1} - \mu f_{i+1} = f_i + \mu f_i$. This way, the spacing between the resonators increases with their line width (see Fig. 6).

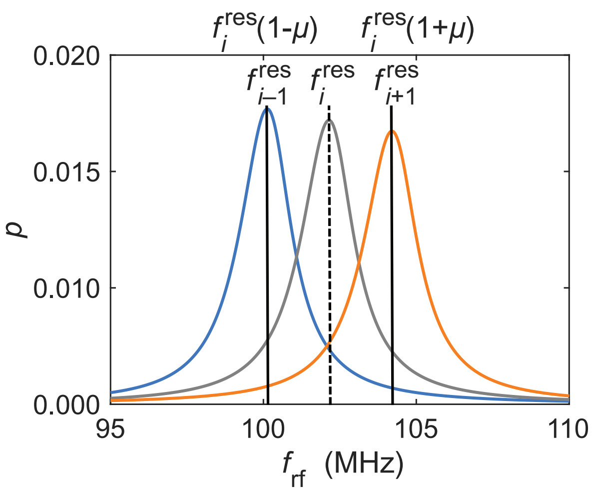

FIG. 6. Magnetization-normalized oscillation power versus frequency of a rf signal for three spintronic resonators of different frequencies following Eq. (A1).

Hence, the rf frequencies follow the law

$$f_i = f_0\left(\frac{1+\mu}{1-\mu}\right)^i, \tag{A1}$$

where f_0 is the lowest frequency. We choose $f_0 = f_{\min} =$ 50 MHz to solve the MNIST dataset. To space the rf frequencies as far as possible, we choose $f_{\max} = 20$ GHz as the resonance frequency. If the number of rf signals is N, then $f_{N-1} = f_{\max}$ and $f_{\max} = f_{\min}[(1+\mu)/(1-\mu)]^{N-1}$. We can then compute that the optimum spacing coefficient μ is

$$\mu = \frac{(f_{\max}/f_{\min})^{1/(N-1)} - 1}{(f_{\max}/f_{\min})^{1/(N-1)} + 1}. \tag{A2}$$

In the case of MNIST $N = 784$, which gives $\mu = 0.0038$. It is important to note that in Eq. (A2) the coefficient of separation does not depend directly on the frequency of the resonators, but on the ratio between the highest frequency and the lowest one. To solve the digits dataset, we choose $f_0 = 100$ MHz and $\mu = \alpha = 0.01$. In this situation, cross talk is not dominant because each curve of the resonance starts when the previous one ends, as represented in Fig. 6 ($\mu = \alpha = 0.01$)

The initialization resonance frequencies of the resonators of each synaptic chain also follow Eq. (A1), but with a random shift following a normal distribution with a standard deviation of $f_k(0.001/\sqrt{64})$ for the digits dataset and that of $f_k(0.001/\sqrt{784})$ for the MNIST dataset.

In Eq. (2), for the spin-diode voltage, we see that the resonator voltage decreases as $1/f^{\,\mathrm{res}}$. Hence, to obtain comparable signals for all resonators of a chain, we scale the microwave powers of the input layer to increase the signal emitted by high-frequency signals: $P_i \to P_i(f_i^{\,\mathrm{res}}/f_0^{\,\mathrm{res}})$.

[1] T. J. O'Shea, T. Roy, and T. C. Clancy, Over-the-air deep learning based radio signal classification, IEEE J. Sel. Top. Signal Process. **12**, 168 (2018).

[2] Y. H. Yoon, S. Khan, J. Huh, and J. C. Ye, Efficient B-mode ultrasound image reconstruction from sub-sampled RF data using deep learning, IEEE Trans. Med. Imaging **38**, 325 (2019).

[3] M. Dai, S. Li, Y. Wang, Q. Zhang, and J. Yu, Post-processing radio-frequency signal based on deep learning method for ultrasonic microbubble imaging, Biomed. Eng. Online **18**, 95 (2019).

[4] E. Besler, Y. C. Wang, and A. V. Sahakian, Real-time radiofrequency ablation lesion depth estimation using multi-frequency impedance with a deep neural network and tree-based ensembles, IEEE Trans. Biomed. Eng. **67**, 1890 (2020).

[5] K. Merchant, S. Revay, G. Stantchev, and B. Nousain, Deep learning for RF device fingerprinting in cognitive communication networks, IEEE J. Sel. Top. Signal Process. **12**, 160 (2018).

[6] J. Lien, N. Gillian, M. E. Karagozler, P. Amihood, C. Schwesig, E. Olson, H. Raja, and I. Poupyrev, Soli: Ubiquitous gesture sensing with millimeter wave radar, ACM Trans. Graph. **35**, 142 (2016).

[7] Y. Kim, in 2018 International Symposium on Antennas and Propagation, (2018) pp. 1–2.

[8] M. F. Al-Sa'd, A. Al-Ali, A. Mohamed, T. Khattab, and A. Erbad, RF-based drone detection and identification using deep learning approaches: An initiative towards a large open source drone database, Future Generation Comput. Syst. **100**, 86 (2019).

[9] E. García-Martín, C. F. Rodrigues, G. Riley, and H. Grahn, Estimation of energy consumption in machine learning, J. Parallel Distrib. Comput. **134**, 75 (2019).

[10] Y. LeCun, Y. Bengio, and G. Hinton, Deep learning, Nature **521**, 436 (2015).

[11] Big data needs a hardware revolution, Nature **554**, 145 (2018).

[12] C. D. Schuman, T. E. Potok, R. M. Patton, J. D. Birdwell, M. E. Dean, G. S. Rose, and J. S. Plank, A survey of neuromorphic computing and neural networks in hardware, arXiv 1705.06963 (2017).

[13] S. Ambrogio, P. Narayanan, H. Tsai, R. M. Shelby, I. Boybat, C. di Nolfo, M. Giordano, M. Bodini, N. C. P. Farinha, B. Killeen, C. Cheng, Y. Jaoudi, and G. W. Burr, Equivalent-accuracy accelerated neural-network training using analogue memory, Nature **558**, 60 (2018).

[14] F. Cai, J. M. Correll, S. H. Lee, Y. Lim, V. Bothra, Z. Zhang, M. P. Flynn, and W. D. Lu, A fully integrated reprogrammable memristor–CMOS system for efficient multiply–accumulate operations, Nat. Electron. **2**, 290 (2019).

[15] P. Yao, H. Wu, B. Gao, J. Tang, Q. Zhang, W. Zhang, J. Joshua Yang, and H. Qian, Fully hardware-implemented memristor convolutional neural network, Nature **577**, 641 (2020).

[16] R. Hamerly, L. Bernstein, A. Sludds, M. Soljačić, and D. Englund, Large-Scale Optical Neural Networks Based

on Photoelectric Multiplication, Phys. Rev. X **9**, 021032 (2019).

[17] J. Feldmann, N. Youngblood, M. Karpov, H. Gehring, X. Li, M. Stappers, M. Le Gallo, X. Fu, A. Lukashchuk, A. S. Raja, J. Liu, C. D. Wright, A. Sebastian, T. J. Kippenberg, W. H. P. Pernice, and H. Bhaskaran, Parallel convolution processing using an integrated photonic tensor core, Nature **589**, 82 (2021).

[18] J. Torrejon, M. Riou, F. Abreu 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).

[19] 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).

[20] 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).

[21] M. Zahedinejad, A. A. Awad, S. Muralidhar, R. Khymyn, H. Fulara, H. Mazraati, M. Dvornik, and J. Åkerman, Two-dimensional mutually synchronized spin hall nano-oscillator arrays for neuromorphic computing, Nat. Nanotechnol. **15**, 47 (2020).

[22] M. Koo, M. R. Pufall, Y. Shim, A. B. Kos, G. Csaba, W. Porod, W. H. Rippard, and K. Roy, Distance Computation Based on Coupled Spin-Torque Oscillators: Application to Image Processing, Phys. Rev. Appl. **14**, 034001 (2020).

[23] H. Arai and H. Imamura, Neural-Network Computation Using Spin-Wave-Coupled Spin-Torque Oscillators, Phys. Rev. Appl. **10**, 024040 (2018).

[24] D. E. Nikonov, P. Kurahashi, J. S. Ayers, H.-J. Lee, Y. Fan, and I. A. Young, Convolution inference via synchronization of a coupled CMOS oscillator array, IEEE J. Explor. Solid-State Comput. Devices Circuits **6**, 170 (2020).

[25] D. E. Nikonov, G. Csaba, W. Porod, T. Shibata, D. Voils, D. Hammerstrom, I. A. Young, and G. I. Bourianoff, Coupled-oscillator associative memory array operation for pattern recognition, IEEE J. Explor. Solid-State Comput. Devices Circuits **1**, 85 (2015).

[26] A. A. Tulapurkar, Y. Suzuki, A. Fukushima, H. Kubota, H. Maehara, K. Tsunekawa, D. D. Djayaprawira, N. Watanabe, and S. Yuasa, Spin-torque diode effect in magnetic tunnel junctions, Nature **438**, 339 (2005).

[27] B. Fang, M. Carpentieri, X. Hao, H. Jiang, J. A. Katine, I. N. Krivorotov, B. Ocker, J. Langer, K. L. Wang, B. Zhang, B. Azzerboni, P. Khalili Amiri, G. Finocchio, and Z. Zeng, Giant spin-torque diode sensitivity in the absence of bias magnetic field, Nat. Commun. **7**, 11259 (2016).

[28] J. Cai, L. Zhang, B. Fang, W. Lv, B. Zhang, G. Finocchio, R. Xiong, S. Liang, and Z. Zeng, Sparse neuromorphic computing based on spin-torque diodes, Appl. Phys. Lett. **114**, 192402 (2019).

[29] A. Slavin and V. Tiberkevich, Nonlinear auto-oscillator theory of microwave generation by spin-polarized current, IEEE Trans. Magn. **45**, 1875 (2009).

[30] A. Dussaux, A. V. Khvalkovskiy, P. Bortolotti, J. Grollier, V. Cros, and A. Fert, Field dependence of spin-transfer-induced vortex dynamics in the nonlinear regime, Phys. Rev. B **86**, 014402 (2012).

[31] A. Litvinenko, V. Iurchuk, P. Sethi, S. Louis, V. Tyberkevych, J. Li, A. Jenkins, R. Ferreira, B. Dieny, A. Slavin, and U. Ebels, Ultrafast sweep-tuned spectrum analyzer with temporal resolution based on a spin-torque nano-oscillator, Nano Lett. **20**, 6104 (2020).

[32] M. Zahedinejad, H. Fulara, R. Khymyn, A. Houshang, S. Fukami, S. Kanai, H. Ohno, and J. Åkerman, Memristive Control of Mutual SHNO Synchronization for Neuromorphic Computing, ArXiv:2009.06594 [Physics] (2020).

[33] D. Marković, N. Leroux, A. Mizrahi, J. Trastoy, V. Cros, P. Bortolotti, L. Martins, A. Jenkins, R. Ferreira, and J. Grollier, Detection of the Microwave Emission from a Spin-Torque Oscillator by a Spin Diode, Phys. Rev. Appl. **13**, 044050 (2020).

[34] S. Jiang, R. Khymyn, S. Chung, T. Quang Le, L. H. Diez, A. Houshang, M. Zahedinejad, D. Ravelosona, and J. Åkerman, Reduced spin torque nano-oscillator linewidth using He+ irradiation, Appl. Phys. Lett. **116**, 072403 (2020).

[35] M. Romera, P. Talatchian, S. Tsunegi, F. Abreu Araujo, V. Cros, P. Bortolotti, J. Trastoy, K. Yakushiji, A. Fukushima, H. Kubota, S. Yuasa, M. Ernoult, D. Vodenicarevic, T. Hirtzlin, N. Locatelli, D. Querlioz, and J. Grollier, Vowel recognition with four coupled spin-torque nano-oscillators, Nature **563**, 230 (2018).

[36] B. Divinskiy, S. Urazhdin, S. O. Demokritov, and V. E. Demidov, Controlled nonlinear magnetic damping in spin-Hall nano-devices, Nat. Commun **10**, 5211 (2019).

[37] D. P. Kingma and J. Ba, Adam: A Method for Stochastic Optimization, ArXiv:1412.6980 (2014).

[38] V. Joseph and C. Nagarajan, MADONNA: A Framework for Energy Measurements and Assistance in designing Low Power Deep Neural Networks.

[39] S. Bonnetti, P. Muduli, F. Mancoff, and J. Åkerman, Spin torque oscillator frequency versus magnetic field angle: The prospect of operation beyond 65 GHz, Appl. Phys. Lett. **94**, 102507 (2009).