Computing with Injection-Locked Spintronic Diodes

Luciano Mazza,[1] Vito Puliafito,[1,†] Eleonora Raimondo,[2] Anna Giordano,[3] Zhongming Zeng,[4] Mario Carpentieri,[1] and Giovanni Finocchio[2,*]

[1] *Department of Electrical and Information Engineering, Politecnico di Bari, via Orabona 4, 70125 Bari, Italy*
[2] *Department of Mathematical and Computer Sciences, Physical Sciences and Earth Sciences, University of Messina, 98166, Messina, Italy*
[3] *Department of Engineering, University of Messina, c.da Di Dio, 98166 Messina, Italy*
[4] *Suzhou Institute of Nano-tech and Nano-bionics, Chinese Academy of Sciences, Ruoshui Road 398, Suzhou 215123, People's Republic of China*

(Received 2 June 2021; revised 25 November 2021; accepted 23 December 2021; published 31 January 2022)

Spintronic diodes (STDs) are emerging as a technology for the realization of high-performance microwave detectors. The key advantages of such devices are their high sensitivity, capability to work at low input power, and compactness. In this work, we show a possible use of STDs for neuromorphic computing to expand the realm of their functionalities for the implementation of analog multiplication, which is a key operation in convolutional neural networks (CNNs). In particular, we introduce the concept of degree of rectification (DOR) in injection-locked STDs. Micromagnetic simulations are used to design and identify the working range of the STDs for the implementation of the DOR. Previous experimental data confirm the applicability of the proposed solution, which is tested in image processing and in a CNN that recognizes handwritten digits.

DOI: 10.1103/PhysRevApplied.17.014045

I. INTRODUCTION

The use of deep convolutional neural networks (CNNs) has grown exponentially, impacting on the development of the emerging area of artificial intelligence. The realization of low-power and compact building blocks (neurons and synapses) for such networks is a key requirement for their hardware realization and mass integration into consumer electronics and Internet of things (IOT) nodes [1].

One of the most critical operations for CNNs is multiplication, which is implemented to calculate the inputs of each neuron and to perform the convolution operation needed for feature extractions.

Digital multiplication is realized with multiple adders, multiplexers, and carryover systems. This approach is scalable but critical in terms of energy dissipation, computing time, and area occupancy [2].

An analog electronic implementation of convolution may overcome today's bottlenecks for the realization of an efficient device, in terms of computational time and energy consumption, but it turns out to be very difficult to obtain because of the susceptibility to noise and voltage offsets [3,4].

A possible direction for analog solutions is the use of photonic tensor cores based on phase-change materials [5]. This approach is parallelizable, since the inputs may be multiplexed in the frequency domain, but because of its size it is not suitable for compact systems. Another solution is based on memristors [6,7], where multiplication is obtained using Ohm's law. For applications in CNNs, however, the main problems of memristors are device-to-device variability and conductance degradation, which can be corrected with continuous *in situ* training of the CNN weights [7].

From a theoretical point of view, the analog multiplication of two values, F and G, can also be implemented by considering an observable, P, characterized by a parabolic input-output relation, $P(X) = aX^2 + bX + c$, where a, b, and c are characteristic parameters of the physical system and X is the input. Combining three measurements, where $X = F - G$, $X = F$, and $X = -G$, it can be easily demonstrated that the multiplication, FG, is given by

$$FG = \frac{P(F-G) - P(F) - P(-G) + c}{-2a}. \tag{1}$$

A description of the use of Eq. (1) is presented in Sec. IV.

The degree of match (DOM) between two interacting nonlinear oscillators is defined as

$$z_{\mathrm{DOM}}(t) = \frac{1}{2}|z_1(t) + z_2(t)| \ , \tag{2}$$

*gfinocchio@unime.it
†vito.puliafito@poliba.it

where $z_1(t)$ and $z_2(t)$ are the complex oscillating variable describing the behavior of each oscillator, which is characterized by power $p = |z|^2$ and phase $\phi = \arg(z)$. As already demonstrated theoretically in previous works, pairs of tunable complementary metal oxide semiconductor (CMOS) [8] and spintronic [9] oscillators exhibit a time-independent DOM in the locking region, which can be used for the computation of multiplication by using Eq. (1).

In particular, calculation of the DOM using spintronic devices seems to be promising in terms of time and energy [10], compactness (nanoscale size [11]), and CMOS compatibility (spin-transfer-torque magnetoresistive random access memory has already been integrated with CMOS processes by the main developers [12]). For two equal spintronic oscillators with an output power approximately constant in the locking region, the DOM is given by

$$z_{\mathrm{DOM}} \cong \left|\sqrt{p}\right| \cos\frac{\Delta\phi}{2} \quad , \tag{3}$$

where $\Delta\phi = \phi_1 - \phi_2$ is the phase difference between the time-domain traces of the two oscillators [8,9]. When at the working point of the oscillators, which can be set and controlled by a bias current [9], $\Delta\phi$ is close to zero and the cosine can be approximated by a second-order polynomial function applying the Taylor expansion; therefore, the DOM is eligible as observable P in Eq. (1) (X is the oscillator phase difference). Similar remarks can be generalized in the case of two different oscillators [13]. However, the read-out mechanism of the DOM is a key problem because the amplitude and phase of each oscillator are not easy to determine by on-chip evaluation.

Recent developments of spintronic diodes (STDs) gave rise to room-temperature solutions that overcame the thermodynamic limits of Schottky diodes [14–18].

The self-oscillation state of the magnetoresistive signal in STDs is driven by a sufficiently large dc current. When the device is locked to an external ac current, it gives rise to a rectification voltage, V_{dc}, with sensitivities exceeding 0.2 MV/W [19,20], which concretely means that an order of mV rectification voltage can be achieved at an input power of tens of nW or lower. This sensitivity has been recently improved by an order of magnitude by combining a bolometric effect with injection locking [20].

This work introduces the concept of degree of rectification (DOR), which is based on the idea to use Eq. (1) to compute analog multiplication, having the rectified voltage, V_{dc}, of injection-locked STDs as an observable P and the bias current, I_{dc}, as an input. The applicability of the DOR is studied through micromagnetic simulations, which show the fundamental aspects of the parabolic relationship between V_{dc} and I_{dc} [19], in comparison with experimental data from Ref. [15]. Such an approach has the intrinsic advantages of spintronic technology (CMOS compatible, scalable, and low power dissipation), simplifies the read-out mechanism compared with the DOM, and can be realized with a single device. To test the robustness of DOR-based multiplication, we realize, in software, a simple CNN that recognizes the digits of the Modified National Institute of Standards and Technology (MNIST) dataset [21,22], where convolution is obtained with DOR-based and ideal multiplication. Then, we emulate the effect of transfer learning in hardware by computing the performance of the network, in terms of recognition accuracy, for the same dataset. The results show that the transfer of knowledge is robust and, introducing DOR multiplication in the training phase of the fully connected (FC) layer of the CNN, the accuracy of the ideal CNN is recovered. We show that the use of the DOR is a reliable method for the extraction of dark knowledge used for knowledge distillation [23,24], and it is also robust to thermal fluctuations and device-to-device variability.

II. DEVICE CONCEPT AND MICROMAGNETIC MODELING

The device idea is a hybrid magnetic tunnel junction (MTJ), as sketched in Fig. 1(a), having an out-of-plane free layer (FL, 1.63-nm-thick $Co_{20}Fe_{60}B_{20}$) and an in-plane polarizer [PL, synthetic antiferromagnet $Co_{70}Fe_{30}$ (2.3 nm)/Ru (0.85 nm)/$Co_{40}Fe_{40}B_{20}$ (2.4 nm)] exchange biased by a PtMn (15 nm) layer. The device is patterned with an elliptical cross section (150×60 nm^2) and its resistances in the parallel and antiparallel states are $R_P = 640\ \Omega$ and $R_{\mathrm{AP}} = 1200\ \Omega$, respectively. The further advantage of this device is zero-field operation, as already demonstrated [25].

To explain the concept of the DOR, we perform micromagnetic simulations of the MTJ's FL magnetization by solving numerically the Landau-Lifshitz-Gilbert-Slonczewski equation [15,25,26]:

$$(1+\alpha^2)\frac{d\mathbf{m}}{d\tau} = -(\mathbf{m}\times\mathbf{h}_{\mathrm{eff}}) - \alpha\mathbf{m}\times(\mathbf{m}\times\mathbf{h}_{\mathrm{eff}}) \tag{4}$$

$$+\sigma I g_T[\mathbf{m}\times(\mathbf{m}\times\mathbf{m}_p) - q(\mathbf{m}\times\mathbf{m}_p)]. \tag{5}$$

On the right side of Eq. (4), the three terms represent the conservative dynamics, the Gilbert dissipation, and the spin-transfer torque. α is the damping parameter; $\mathbf{m} = \mathbf{M}/M_S$ is the normalized magnetization of the FL, where M_S is its saturation magnetization; and $d\tau = \gamma_0 M_S dt$ is the dimensionless time step, where γ_0 is the gyromagnetic ratio. The effective field, $\mathbf{h}_{\mathrm{eff}}$, includes standard micromagnetic contributions from the exchange, anisotropy, external, and demagnetizing field. The torque term is proportional to $\sigma = \frac{g|\mu_B|}{|e|\gamma_0 M_S^2 V_{\mathrm{FL}}}$, where g is the gyromagnetic splitting factor, μ_B is the Bohr

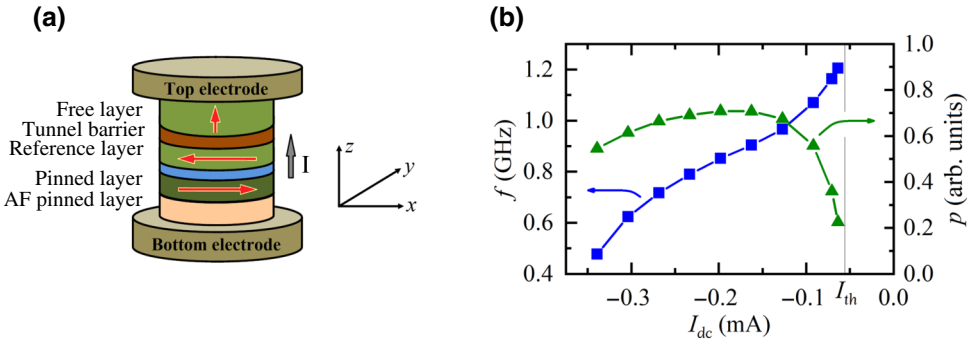

FIG. 1. (a) Sketch of the STD with the indication of free and polarizer layers. Structure is hybrid, free layer is perpendicular, and polarizer is aligned along the $-x$ direction. Cartesian coordinate system is also included as an inset. (b) Frequency (blue squares) and power (green triangles) of FL magnetization oscillation as a function of applied dc current, obtained by means of micromagnetic simulations. Vertical line represents threshold current $|I_{th}| = 0.056$ mA.

magneton, e is the electron charge, V_{FL} is the FL volume. I is the total current flowing into the MTJ given by $I = I_{dc} + I_{ac,max}\sin(2\pi f_{ac} + \varphi_{ac})$. For the scalar function g_T, we consider the expression $g_T(\theta) = 2\eta_T[1 + \eta_T^2\cos(\theta)]^{-1}$, where η_T is the polarization efficiency, and $\cos(\theta) = \mathbf{m}\cdot\mathbf{m}_p$, [26,27] where $\mathbf{m}_p = \mathbf{M}/M_{SP}$ is the normalized magnetization of the PL and M_{SP} its saturation magnetization. The micromagnetic parameters are $M_S = 9.5 \times 10^5$ A/m, perpendicular anisotropy constant $k_U = 5.45 \times 10^5$ J/m^3, exchange constant $A = 2.0 \times 10^{-11}$ J/m, $\alpha = 0.02$, and $\eta_T = 0.66$. The spin-transfer torque includes both dampinglike and fieldlike torques, where q is the ratio between the two torque contributions and it is equal to 0.1 [28]; the results achieved for $q = 0$ and 0.2 are qualitatively the same with a variation in the oscillation frequency of less than 3%. $\mathbf{m}_p$ is aligned along the $-x$ direction.

III. SIMULATION RESULTS

First, we characterize the properties of self-oscillations, frequency f_0 (blue squares), and power p_0 (green triangles), as a function of the dc current, as summarized in Fig. 1(b). We find a threshold current of $|I_{th}| = 0.056$ mA and a nonlinear frequency shift of $N/2\pi = df_0/dp_0 \cong -411$ MHz; the latter is a parameter that links the oscillation frequency and power of a free-running spin-transfer-torque oscillator near the threshold current: $f_0 = f_0(I_{th}) + Np_0/2\pi$ [29]. As discussed later in the text, N is a crucial parameter for the properties of magnetization dynamics in the injection-locking regime, which is achieved when f_{ac} approaches the self-oscillation frequency, f_0. In the locking regime, the magnetoresistance oscillates at the same frequency as that of the ac current; hence, it is possible to observe a rectification voltage [23,30]. The variation of the input dc current, within a specific locking range, does not lead to a change in frequency, but it can modify the amplitude of the oscillating magnetization, $dm_X(I_{dc})$, which is linked to the oscillator power, p $\left(dm_X = \sqrt{p}\right)$, and the intrinsic phase shift, $\varphi(I_{dc})$, between the ac current and the oscillating magnetoresistive signal [31]. The output voltage can be computed from [15]

$$V_{dc} = \frac{(R_{AP} - R_P)\sqrt{p}}{4} I_{ac,max}\cos[\varphi(I_{dc})]. \tag{6}$$

Figure 2(a) shows an example of rectification voltage obtained for $I_{ac,max} = 70.7$ μA, $f_{ac} = 800$ MHz, and $\varphi_{ac} = 0$. The rectification voltage has a maximum at about $I_{dc,0} = -0.134$ mA, corresponding to a phase shift close to zero [see Fig. 2(b), where the additional phase shift introduced by the direction of the polarizer is not taken into account]. Figure 2(b) shows dm_X (blue diamonds) and φ (green squares) for the simulations reported in Fig. 2(a). The intrinsic phase shift is a quasi-linear function of the dc current, with a slight deviation close to the edge of the locking region, similar to what is also observed in Refs. [31,32]. Instead, the amplitude of magnetization is weakly dependent on the current. This result is expected for an oscillator with a large nonlinear frequency shift. In fact, the power, p, of the injection-locked oscillator is given by $\frac{p}{p_0} = 1 + \frac{\sigma I_{ac,max}}{\sqrt{1+(N/\Gamma_P\xi)^2}}$, where $\xi = I_{dc}/I_{th}$ is the supercriticality of the dc bias current (or voltage) and Γ_P is the effective damping rate [29]. For the device we study, N/Γ_P is larger than 15; hence, a reduced dependence of the oscillator power as a function of I_{dc} is achieved [29], as shown in Fig. 2(b) (blue diamonds), where a change of less than 3% is observed in dm_X.

The quasi-linear behavior of the intrinsic phase shift shown in Fig. 2(b) can be approximated by $\varphi(I_{dc}) = mI_{dc} + n$. The fitting parameters can be identified directly from the rectified voltage as follows. $V_{dc,max} \cong (1/4)(R_{AP} - R_P)I_{ac,max}\sqrt{p}$ is obtained at $I_{dc,0}$ [see Fig. 2(a)]; for this value of current, the argument of the cosine is zero, and hence, $n = -mI_{dc,0}$.

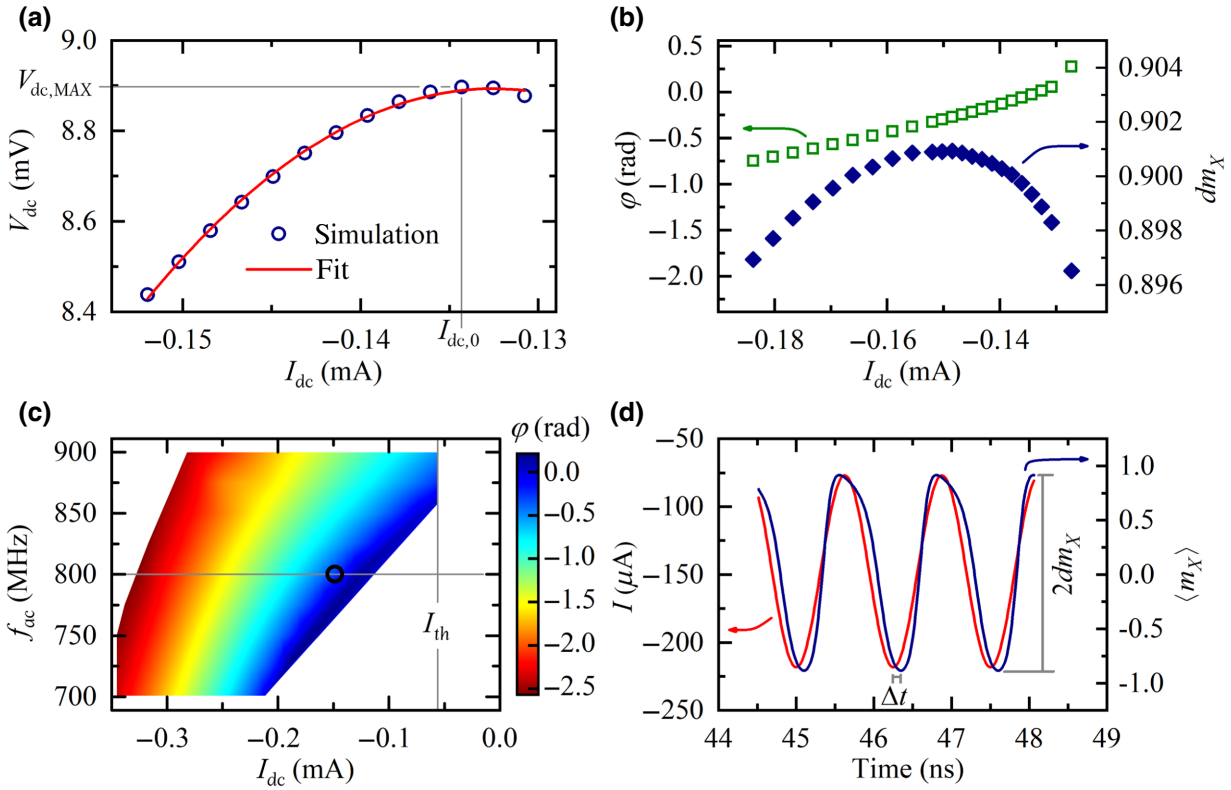

FIG. 2. (a) Plot of rectified dc voltage as a function of dc current applied to the spin-torque diode, for $I_{ac,max} = 70.7\ \mu$A and $f_{ac} = 800$ MHz; blue circles are the results of micromagnetic simulations, and red line is the parabolic fit. (b) Intrinsic phase shift (empty squares) and amplitude of magnetization along the x axis (full diamonds) as a function of dc current for the same $I_{ac,max}$ and f_{ac} as those in (a). (c) Phase diagram of the intrinsic phase shift as a function of microwave frequency and dc current for $I_{ac,max} = 70.7\ \mu$A. Vertical line is the auto-oscillation current threshold, $|I_{th}| = 0.056$ mA. Horizontal line represents the microwave frequency value used in (a),(b). (d) Time traces of applied current (left y axis) and spatially averaged x component of magnetization $\langle m_X \rangle$ (right y axis) for the working point indicated with a circle in (c). Time shift, Δt, between two time traces is also indicated.

In addition, the value of m can be estimated from the second derivative of the rectified voltage with respect to the dc current evaluated at $I_{dc,0}$, $\left.\frac{d^2V_{dc}}{dI_{dc}^2}\right|_{I_{dc}=I_{dc,0}} = -m^2 V_{dc,max}$. This approach can also be used with experimental data to extract information about the intrinsic phase shift. Equation (5) can be then rewritten as the equation of a parabola, $V_{dc}(I_{dc}) = aI_{dc}^2 + bI_{dc} + c$, where the expressions of the coefficients are $a = -(1/2)V_{dc,max}m^2$, $b = V_{dc,max}m^2 I_{dc,0}$, and $c = V_{dc,max}\{1 - [(mI_{dc,0})^2/2]\}$. This parabolic relation, which is a key result of this work, can be used to estimate the DOR and give results very close to the direct parabolic fit of data. Figure S1(a) within the Supplemental Material [33] shows a comparison between the parabolic fit of micromagnetic data observed in Fig. 2(a) and parabola obtained with the analytically evaluated parameters [see Fig. S1(b) within the Supplemental Material [33]], highlighting excellent agreement.

It is clear that, to implement a multiplier with spintronic diodes, we need the devices to work with currents and microwave input frequencies to driving an intrinsic phase shift, φ, close to zero or π. Figure 2(c) summarizes the results of a systematic study of φ as a function of I_{dc} and f_{ac} for $I_{ac,max} = 70.7\ \mu$A, with an indication of the threshold-current value, I_{th}. The horizontal line is the working point for data in Figs. 2(a) and 2(b). For this device geometry, the value $\varphi = 0$ is achieved close to the edge of the locking range.

Figure 2(d) shows an example of time-domain evolution of the spatially averaged x component of magnetization, $\langle m_X \rangle$, obtained for $I_{dc} = -0.148$ mA and $f_{ac} = 800$ MHz [circle in Fig. 2(c)] together with the ac current and indication of dm_X.

It can be noted that a constant time shift can be identified by comparing the time traces; however, the magnetization dynamics is characterized by a first harmonic, which has about 76% of the energy and high-order harmonics and the other 24% (as shown in Fig. S2 within the Supplemental Material [33]) may influence the measurement of the intrinsic phase shift directly in the

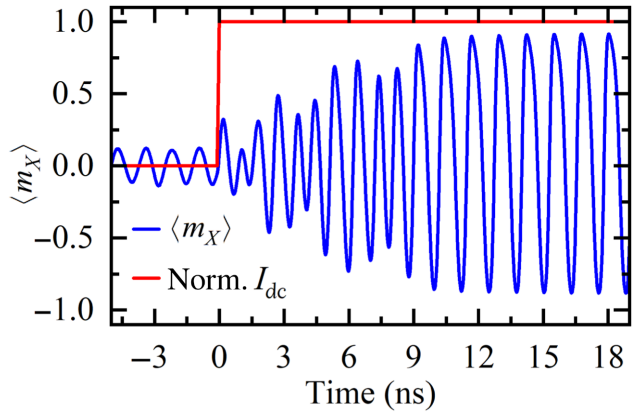

FIG. 3. An example of the time domain trace of the injection locking of the x-component of the magnetization (blue solid line) achieved after the application of a dc current step from 0 to −0.148 mA (normalized dc current is shown in red solid line). The ac current has an amplitude $I_{ac,max} = 70.7\ \mu$A and $f_{ac} = 800$ MHz.

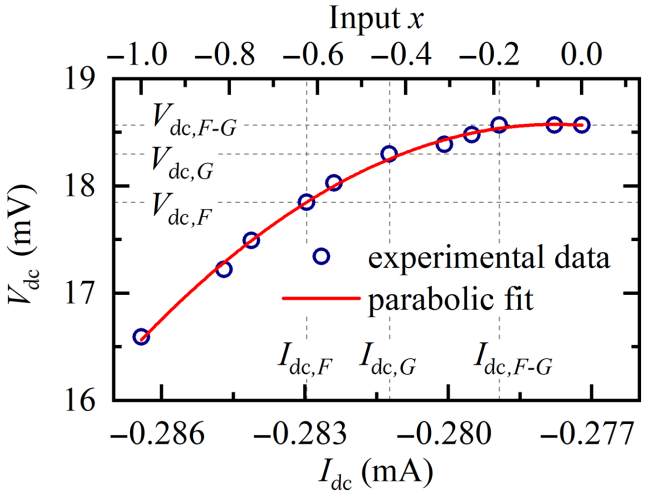

FIG. 4. Experimental data (blue circles), from Ref. [15], of rectified voltage as a function of bias current of the injection-locked STD (bottom axis) and numeric input (top axis) for the even parabolic equation. Solid red line is the parabolic fit. Points corresponding to the input values $F = -0.62$, $G = -0.44$, and $F - G = -0.18$ are also indicated.

time-domain traces. For this reason, this parameter is computed in the Fourier space. Figure 3 illustrates an example of
magnetization evolution when a dc step is applied to achieve the injection-locking regime. The transient time is about 10 ns and represents a good estimation of the speed of the multiplication computation.

Figure S3(a) within the Supplemental Material [33] shows a simulation of the same device with $I_{ac,max} = 70.7\ \mu$A and $f = 543$ MHz at room temperature. As can be noted, in the presence of a thermal field, the frequency of self-oscillation is smaller, as expected by the reduction of saturation magnetization, and the time required for locking is reduced.

Interestingly, experimental data of rectification voltages as a function of bias current in high-sensitivity spin-torque diodes also exhibit a trend similar to the theoretical curves proposed in this work for the evaluation of the DOR [see Fig. 5(e) in Ref. [15], these data are partially reported in Fig. 4 and in Fig. S3(b) within the Supplemental Material [33]]. To achieve high sensitivity, there is an additional term in the rectified voltage that is proportional to the dc current and will give rise to a linear shift of the parabolic equation, $V_{dc}(I_{dc}) = aI_{dc}^2 + (b + \Delta R_{dc})I_{dc} + c$, where ΔR_{dc} is the variation of the dc resistance induced by the microwave input [15]. In the rest of the paper, we will show calculations performed with the DOR estimated from either theoretical or experimental curves.

IV. DOR-BASED MULTIPLICATION

From the device's input-output relationship, it is possible to identify the parameters a, b, and c, which satisfy the relation $V_{dc}(I_{dc}) = aI_{dc}^2 + bI_{dc} + c$, linking the bias current, I_{dc}, and the rectified dc voltage, V_{dc}. The range of input currents is then scaled to the range of desired input x (we consider the range $[-1, 0]$ for simplicity) with the following linear transformation: $I_{dc} = |I_{dc,0} - I_{dc,-1}|x + I_{dc,0}$, where $I_{dc,0}$ and $I_{dc,-1}$ are the current values associated with the numeric inputs 0 and −1. In this way, we get an even parabolic equation and $V_{dc}(x) = V_{dc}(-x)$. The new parabolic relation is then given by $V_{dc}(x) = a'x^2 + c'$, where $a' = a|I_{dc,0} - I_{dc,-1}|^2$ and $c' = V_{dc,max}$ (see Fig. 4). The final calculation to multiply F and G relies on the evaluation of the voltages for $x = F$, G, $(F - G)$, combined as described in Eq. (1).

As an example, we consider Fig. 4, where experimental values of rectified V_{dc} are reported versus both I_{dc} and input x. If we consider $F = -0.62$, $G = -0.44$, and $F - G = -0.18$, the respective V_{dc} values, $V_{dc,F} = 17.85$ mV, $V_{dc,G} = 18.30$ mV, and $V_{dc,F-G} = 18.57$ mV, can be used in Eq. (1), considering parameters $a' = -2.0476$ mV and $c' = 18.565$ mV. In this way, the product obtained is equal to 0.241, which is very close to the desired value of $FG = 0.273$.

The speed of multiplication depends on which of the two following scenarios is implemented in hardware:

(1) Maximum velocity. This is achieved by considering three diodes for each multiplication and the CMOS circuitry to perform the addition. The time required is the time necessary to achieve locking plus that necessary to perform addition (division is achieved together with considering proper gain for the analog adder).

(2) Minimum-area occupancy. The three DOR operations are performed with the same diode. In this

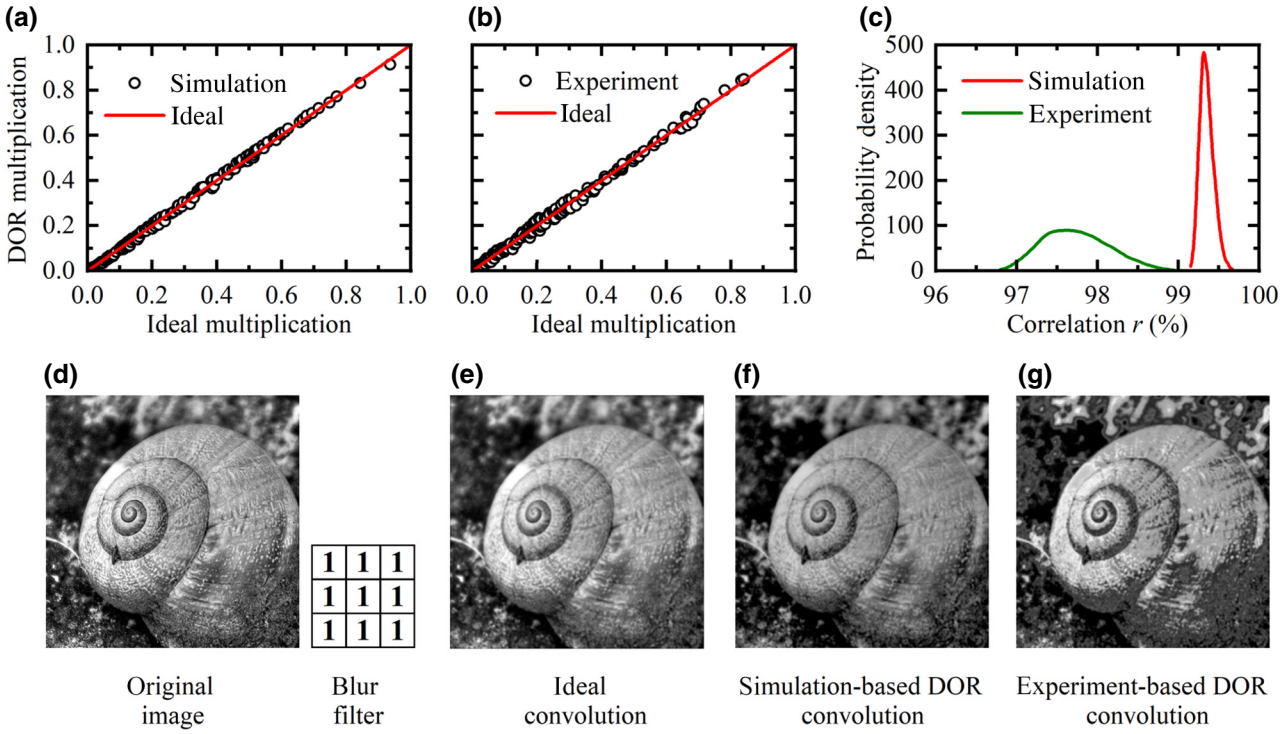

FIG. 5. (a) Comparison between DOR multiplications based on micromagnetic simulations (black circles) and the ideal multiplication (red line, bisector of the first). (b) Same comparison as that in (a) but with the use of the experimental curve from Ref. [15] for DOR multiplication. (c) Correlation probability density functions for the convolution of 10 000 random filters, considering DOR multiplication through simulation (red curve) and experimental data (green line). (d) Image of a snail, extracted from ImageNet dataset; inset, 3×3 blur filter used for convolution. (e) Ideal convolution; (f) DOR-based convolution obtained through micromagnetic data; (g) DOR-based convolution obtained through experimental data.

case, the time lengthens by at least 3 times and additional memory elements are needed to store data before performing the sum.

V. PERFORMANCE EVALUATIONS

As a first step, we compare the micromagnetic and experimental DOR-based multiplications with the ideal one. Figures 5(a) and 5(b) show 200 multiplications obtained with the DOR, as computed from numerical data in Fig. 2(a) (blue circles) and experimental data of Fig. 4 compared with the output of the ideal multiplication (red line). The results show that the correlation between the ideal case and the micromagnetic (experimental) DOR multiplication is 99.93% (99.83%).

The second test is the calculation of convolution between an image of a snail (extracted from the ImageNet dataset [34]) with 3×3 filters. Figure 5(c) illustrates the probability density functions (PDFs) of the correlation coefficients, r, computed by considering 10 000 random instances of filters. The average correlation coefficients are $\bar{r}_{\mathrm{sim}} = 99.41\%$ and $\bar{r}_{\mathrm{expt.}} = 97.87\%$ for the simulation and experimental data, respectively. The smaller average correlation and larger dispersion achieved with experimental data is due to less-accurate parabolic behavior. It is important to highlight that micromagnetic calculations suggest that MTJs can be very effective in the implementation of convolution, reaching correlation values with an ideal multiplication larger than 99%. As an example, we show the convolution of the snail image, represented in Fig. 5(d), with a 3×3 blurring filter, consisting of equal weights. This filter is not included in the random statistical analysis of Fig. 5(c), but it represents a limit case since multiplication errors are similar and they constructively accumulate in convolution. Figures 5(e)–5(g) show the results of convolution in the case of ideal (e) and DOR-based [(f) simulation and (g) experimental] multiplication. The correlation coefficients are $r_{\mathrm{sim}} = 99.07\%$ and $r_{\mathrm{expt.}} = 96.64\%$, which, in fact, are beyond the lower tails in the PDFs of Fig. 5(c). Similar results are also achieved with other images from the same database.

Today, there is growing interest in developing precision-scalable architectures [2,35], including binary neural networks. With this in mind, we benchmark the impact of DOR-based multiplication in a simple CNN, using for this application experimental data which give the worst performance. Even if the DOR-based multiplication gives rise to

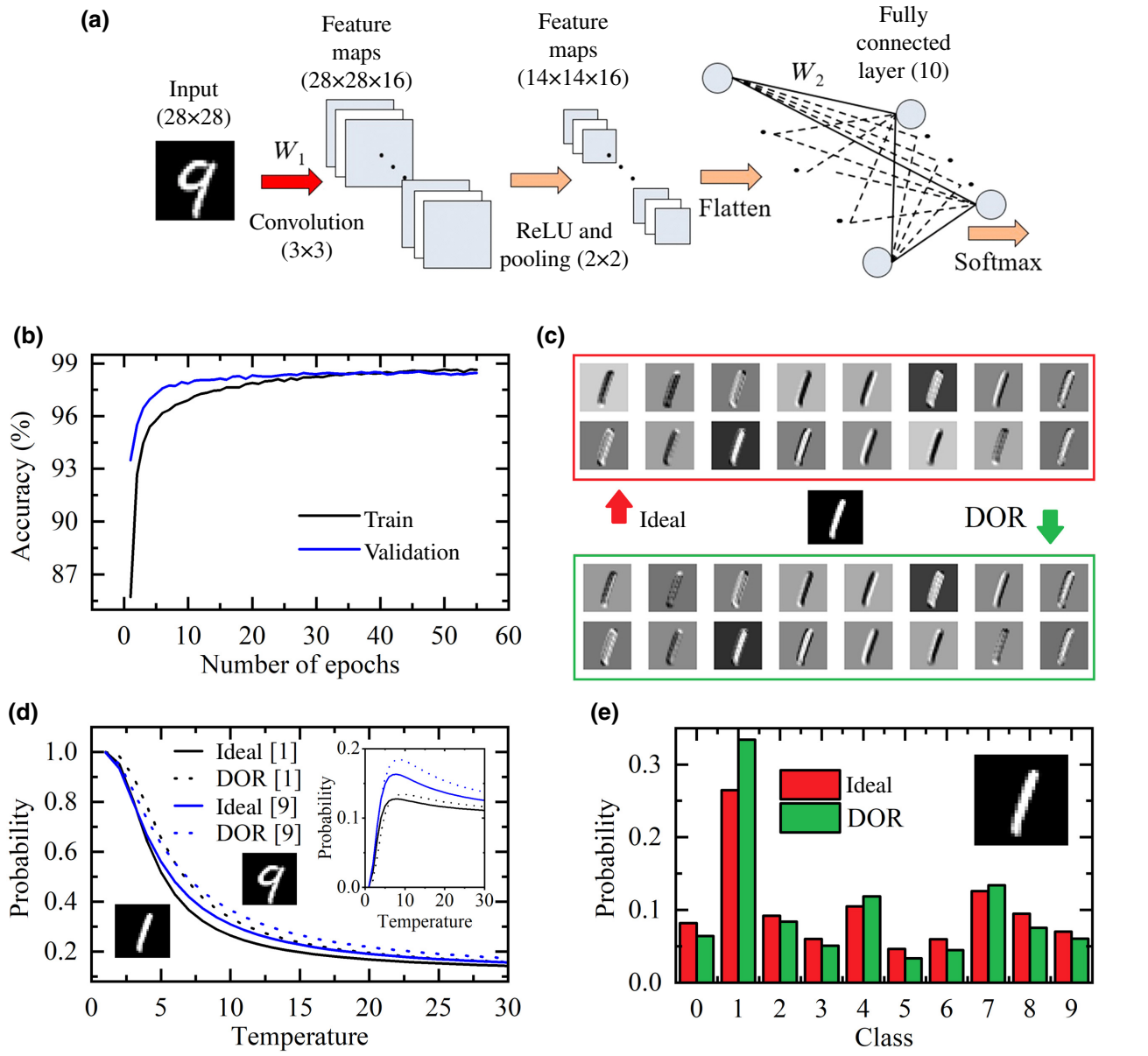

FIG. 6. (a) Structure of the CNN composed of convolutional layer, ReLU activation function, pooling layer, FC layer of classifications. (b) Percentage of recognition accuracy as a function of the number of epochs. Black (blue) line corresponds to the results for the training (validation) dataset. (c) Feature maps of a test image obtained with ideal multiplication (top) and DOR multiplication (bottom). (d) Probability of the most-probable class as a function of the temperature coefficient for the two represented test images (black line, image of handwritten digit one; blue line, image of handwritten digit nine) obtained from CNN based on ideal multiplication (solid line) and CNN with multiplication based on DOR for the convolutional layer with additional training of the FC layer (dashed line); graph in the inset shows the probability of the second-most-probable class as a function of temperature. (e) Probability of all classes for $T = 10$ for the represented image of handwritten digit one, obtained from CNN based on ideal multiplication (red) and CNN with DOR-based multiplication applied to the convolutional layer with the additional training of the FC layer (green).

a reduced precision for the results, our study shows that there is a low impact on the accuracy of neural networks.

In detail, we consider a vanilla CNN with the architecture shown in Fig. 6(a). It is composed of a single convolutional layer with 16 filters of 3×3 in size. Its output feature maps propagate through a layer of neurons having the rectified linear unit (ReLU) activation function and a pooling layer that halves the results' spatial dimension with the max-pooling operation. The CNN then has a flattened layer and a FC layer with 10 neurons with the softmax activation function.

The CNN is trained in Python with TensorFlow on the MNIST dataset, considering a training set of 48 000 images, and a validation set of 12 000 images, and

TABLE I. Summary of the results on the test set for the CNN trained, respectively, with ideal multiplication (Ideal), with DOR-based multiplication applied to the convolutional layer ($Conv_{DOR}$), with DOR-based multiplication applied to both the convolutional and FC layers ($Conv_{DOR}+FC_{DOR}$), and finally the results of the implementation of an additional training of the FC layer when DOR-based multiplication is applied to the convolutional one ($Conv_{DOR}+$trainFC). (a) Results obtained with the main curve [represented in Fig. 4 and Fig. S3(b) within the Supplemental Material, red curve]; (b) results obtained considering device-to-device variations of the STDs; (c) results obtained using the curve with a larger input-current range [Fig. S3(a), blue curve].

	Test accuracy (%)			
	Ideal	$Conv_{DOR}$	$Conv_{DOR}+FC_{DOR}$	$Conv_{DOR}+$trainFC
a	98.57	96.83	94.72	98.40
b		85.51	51.18	98.33
c		97.07	93.11	98.35

testing is performed on a test set containing 10 000 images. To prevent overfitting, dropout layers [36] and early stopping are used. The recognition accuracy reached is 98.64% on the training set and 98.57% on the test set. Then, we use the trained weights to evaluate the accuracy on the same test set, considering DOR-based multiplication for the convolutional layer ($Conv_{DOR}$) (the recognition accuracy achieved is 96.83%) and both the convolutional and FC layers ($Conv_{DOR}+FC_{DOR}$) (94.72%), see Table I (row a). Considering that this numerical test simulates a possible effect of the hardware implementation of multiplication with STDs, the accuracy reduction (<4%) can be improved with few training iterations of the FC layer, when the DOR-based multiplication is applied to the convolutional one ($Conv_{DOR}+$trainFC); in this case, we use the DOR-based convolution for the convolutional layer and the ideal multiplication for the fully connected one. After this additional training, the recognition accuracy on the test set increases to 98.40%, a value comparable to that achieved initially (98.57%).

In addition, we simulate the device-to-device variations of the STDs. We carry out simulations by adding ±2.5% random variation of the parameters of the parabola performing the DOR-based multiplication (see Supplemental Material Note 1 [33] for more details). Table I (row b) summarizes the results obtained considering this nonideality of the device for each of the cases described above.

Further simulations are performed by considering a DOR-based multiplication computed with a parabola having a larger-current region for the input [see Fig. S3(b) within the Supplemental Material [33]]. Table I (row c) shows the accuracies obtained with this configuration. We wish to highlight that the additional training phase is able to fix the error introduced by using this curve, which is not well approximated by a parabola.

Figure 6(c) shows some feature maps of a test image obtained with ideal multiplication (top) and DOR-based multiplication (bottom).

In the last part of this work, we also investigate how the DOR-based multiplication can affect the extraction of dark knowledge. Dark knowledge is the basis of the distillation process in neural networks and is developed primarily with the aim to transfer knowledge from a large network model to a smaller one more suitable for deployment [23,37]. Dark knowledge is revealed through soft probabilities, which effectively smooth out the probability distribution and reveal interclass relationships. To obtain it, an increase of the temperature coefficient (T) of the softmax function, $\mathrm{softmax}(x_i) = e^{z_i/T}/\sum_j e^{z_j/T}$, where z_i are the logits (where usually $T = 1$ is used) [23] is necessary. Figure 6(d) shows the output probability of the most-probable class for two representative images (similar results are achieved for most of the other images) as a function of temperature, while the inset shows the probability of the second-most-probable class as a function of temperature. It can be observed that the extraction of dark knowledge remains robust when using DOR-based multiplication. Figure 6(e) summarizes the probability for all classes for $T = 10$.

VI. SUMMARY AND CONCLUSIONS

Here, we show a strategy to implement analog multiplication with injection-locked STDs by introducing the concept of the DOR. In particular, these devices exhibit a parabolic rectification curve as a function of dc near the current region, where the intrinsic phase shift is close to zero; hence, the DOR, which is nothing more than the rectified voltage, can be used as an observable function for the implementation of analog multiplication. Thanks to micromagnetic simulations and experimental data, we show that DOR-based multiplication is robust, can be achieved at zero-bias field, and can be used for neuromorphic applications. In addition, future studies can take advantage of a properly designed in-plane field originating from the polarizer for an additional degree of freedom to improve the DOR curve. Our results open a path for low power consumption (tens of μW per operation) and potentially high-speed implementation of the multiplication operation (<15 ns), paving the way towards processing in mobile devices and IOT nodes. However, we estimate that the performance of the approach can be improved in terms of velocity and power consumption by using MTJs with a lower resistance-area product [38]; larger oscillation frequencies, which result in faster injection locking [39]; and reduced currents, lowering the dissipation power per operation. Further advancements can be potentially achieved with future solutions based on antiferromagnets, where injection locking is achieved at

tens of picoseconds, as predicted by micromagnetic simulations [40–42] and shown experimentally for switching [43]. High-speed multiplication is crucial not only for neuromorphic computing but also for real-time signal-processing applications.

ACKNOWLEDGMENTS

This work is supported by Grant No. 2019-1-U.0 ("Diodi spintronici rad-hard ad elevate sensitività – DIOSPIN") funded by the Italian Space Agency (ASI) within the call "Nuove idee per la componentistica spaziale del futuro," PON R&I 2014-2020 Project No. ARS01_01215 ("New satellites generation components – NSG") funded by the Italian Ministry of University and Research, and Project No. PRIN 2020LWPKH7 funded by the Italian Ministry of University and Research. The work is also supported by the Petaspin Association [44]. The authors thank Caterina Ciminelli for fruitful discussions.

[1] K. S. N. Yengade and A. P. P. R. Indurkar, Review on design of low power multiply and accumulate unit using baugh-wooley based multiplier, Int. Res. J. Eng. Technol. **4**, 638 (2017).

[2] V. Camus, L. Mei, C. Enz, and M. Verhelst, Review and benchmarking of precision-scalable multiply-accumulate unit architectures for embedded neural-network processing, IEEE J. Emerg. Sel. Top. Circuits Syst. **9**, 697 (2019).

[3] N. Mirchandani and A. Shrivastava, in *2019 IEEE 62nd International Midwest Symposium on Circuits and Systems (MWSCAS)* (IEEE, 2019), Vol. 2019-Augus, pp. 1139–1142.

[4] C. Chen and Z. Li, A low-power CMOS analog multiplier, IEEE Trans. Circuits Syst. II: Express Br. **53**, 100 (2006).

[5] 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 convolutional processing using an integrated photonic tensor core, Nature **589**, 52 (2021).

[6] C. Li, D. Belkin, Y. Li, P. Yan, M. Hu, N. Ge, H. Jiang, E. Montgomery, P. Lin, Z. Wang, W. Song, J. P. Strachan, M. Barnell, Q. Wu, R. S. Williams, J. J. Yang, and Q. Xia, Efficient and self-adaptive in-situ learning in multilayer memristor neural networks, Nat. Commun. **9**, 2385 (2018).

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

[8] D. M. Chiarulli, B. Jennings, Y. Fang, A. Seel, and S. P. Levitan, in *2015 IEEE Computer Society Annual Symposium on VLSI* (IEEE, 2015), Vol. 07-10-July, pp. 125–130.

[9] K. Yogendra, D. Fan, Y. Shim, M. Koo, and K. Roy, in *2016 21st Asia and South Pacific Design Automation Conference (ASP-DAC)* (IEEE, 2016), Vol. 25-28-Janu, pp. 312–317.

[10] B. Dieny, *et al.*, Opportunities and challenges for spintronics in the microelectronics industry, Nat. Electron. **3**, 446 (2020).

[11] Z. Zeng, G. Finocchio, B. Zhang, P. K. Amiri, J. A. Katine, I. N. Krivorotov, Y. Huai, J. Langer, B. Azzerboni, K. L. Wang, and H. Jiang, Ultralow-current-density and bias-field-free spin-transfer nano-oscillator, Sci. Rep. **3**, 1426 (2013).

[12] D. Zhang, L. Zeng, Y. Zhang, J. O. Klein, and W. Zhao, Reliability-enhanced hybrid CMOS/MTJ logic circuit architecture, IEEE Trans. Magn. **53**, 1 (2017).

[13] L. Mazza, V. Puliafito, M. Carpentieri, and G. Finocchio, Robustness of using degree of match in performing analog multiplication with spin-torque oscillators, Solid State Electron. **183**, 108045 (2021).

[14] S. Miwa, S. Ishibashi, H. Tomita, T. Nozaki, E. Tamura, K. Ando, N. Mizuochi, T. Saruya, H. Kubota, K. Yakushiji, T. Taniguchi, H. Imamura, A. Fukushima, S. Yuasa, and Y. Suzuki, Highly sensitive nanoscale spin-torque diode, Nat. Mater. **13**, 50 (2014).

[15] 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. K. Amiri, G. Finocchio, and Z. Zeng, Giant spin-torque diode sensitivity in the absence of bias magnetic field, Nat. Commun. **7**, 11259 (2016).

[16] A. S. Jenkins, R. Lebrun, E. Grimaldi, S. Tsunegi, P. Bortolotti, H. Kubota, K. Yakushiji, A. Fukushima, G. De Loubens, O. Klein, S. Yuasa, and V. Cros, Spin-Torque resonant expulsion of the vortex core for an efficient radiofrequency detection scheme, Nat. Nanotechnol. **11**, 360 (2016).

[17] G. Finocchio, R. Tomasello, B. Fang, A. Giordano, V. Puliafito, M. Carpentieri, and Z. Zeng, Perspectives on spintronic diodes, Appl. Phys. Lett. **118**, 160502 (2021).

[18] N. Leroux, D. Marković, E. Martin, T. Petrisor, D. Querlioz, A. Mizrahi, and J. Grollier, Radio-frequency Multiply-and-Accumulate Operations with Spintronic Synapses, Phys. Rev. Appl. **15**, 034067 (2021).

[19] L. Zhang, B. Fang, J. Cai, M. Carpentieri, V. Puliafito, F. Garescì, P. K. Amiri, G. Finocchio, and Z. Zeng, Ultrahigh detection sensitivity exceeding 10 5 V/W in spin-torque diode, Appl. Phys. Lett. **113**, 102401 (2018).

[20] M. Goto, Y. Yamada, A. Shimura, T. Suzuki, N. Degawa, T. Yamane, S. Aoki, J. Urabe, S. Hara, H. Nomura, and Y. Suzuki, Uncooled sub-GHz spin bolometer driven by auto-oscillation, Nat. Commun. **12**, 536 (2021).

[21] E. Kussul and T. Baidyk, Improved method of handwritten digit recognition tested on MNIST database, Image Vis. Comput. **22**, 971 (2004).

[22] E. Raimondo, A. Giordano, A. Grimaldi, V. Puliafito, M. Carpentieri, Z. Zeng, R. Tomasello, and G. Finocchio, Reliability of neural networks based on spintronic neurons, IEEE Magn. Lett. **12**, 6102805 (2021).

[23] S. Louis, V. Tyberkevych, J. Li, I. Lisenkov, R. Khymyn, E. Bankowski, T. Meitzler, I. Krivorotov, and A. Slavin, Low power microwave signal detection with a spin-torque nano-oscillator in the active self-oscillating regime, IEEE Trans. Magn. **53**, 1400804 (2017).

[24] H. Zhang, Z. Hu, W. Qin, M. Xu, and M. Wang, Adversarial Co-distillation learning for image recognition, Pattern Recognit. **111**, 107659 (2021).

[25] G. Finocchio, B. Azzerboni, G. D. Fuchs, R. A. Buhrman, and L. Torres, Micromagnetic modeling of magnetization switching driven by spin-polarized current in magnetic tunnel junctions, J. Appl. Phys. **101**, 063914 (2007).

[26] J. C. Slonczewski, Current-Driven excitation of magnetic multilayers, J. Magn. Magn. Mater. **159**, L1 (1996).

[27] J. C. Slonczewski, Currents, torques, and polarization factors in magnetic tunnel junctions, Phys. Rev. B **71**, 024411 (2005).

[28] M. Carpentieri, E. Martinez, and G. Finocchio, High frequency spin-torque-oscillators with reduced perpendicular torque effect based on asymmetric vortex polarizer, J. Appl. Phys. **110**, 093911 (2011).

[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. 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).

[31] Y. Zhou, J. Persson, and J. Åkerman, Intrinsic phase shift between a spin torque oscillator and an alternating current, J. Appl. Phys. **101**, 09A510 (2007).

[32] Y. Zhou, V. Tiberkevich, G. Consolo, E. Iacocca, B. Azzerboni, A. Slavin, and J. Åkerman, Oscillatory transient regime in the forced dynamics of a nonlinear auto oscillator, Phys. Rev. B **82**, 012408 (2010).

[33] See the Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevApplied.17.014045 for a comparison of the parabolic fit of numerical results and parabola obtained with the analytical model, the FFT of magnetization and the contribution of its components, numerical simulations at room temperature, and an evaluation of the robustness of the system to device-to-device variability.

[34] L. Fei-Fei, J. Deng, and K. Li, Imagenet: Constructing a large-scale image database, J. Vis. **9**, 1037 (2010).

[35] X. Lin, C. Zhao, and W. Pan, Towards accurate binary convolutional neural network, Adv. Neural Inf. Process. Syst. **2017-Decem**, 345 (2017).

[36] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, Dropout: A simple way to prevent neural networks from overfitting, J. Mach. Learn. Res. **15**, 1929 (2014).

[37] Y. Zhang, T. Xiang, T. M. Hospedales, and H. Lu, in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (IEEE, 2018), pp. 4320–4328.

[38] L. Zhang, J. Cai, B. Fang, B. Zhang, L. Bian, M. Carpentieri, G. Finocchio, and Z. Zeng, Dual-band microwave detector based on magnetic tunnel junctions, Appl. Phys. Lett. **117**, 072409 (2020).

[39] L. Zhang, B. Fang, J. Cai, W. Wu, B. Zhang, B. Wang, P. K. Amiri, G. Finocchio, and Z. Zeng, Enhanced broadband radio frequency detection in nanoscale magnetic tunnel junction by interface engineering, ACS Appl. Mater. Interfaces **11**, 29382 (2019).

[40] V. Puliafito, R. Khymyn, M. Carpentieri, B. Azzerboni, V. Tiberkevich, A. Slavin, and G. Finocchio, Micromagnetic modeling of terahertz oscillations in an antiferromagnetic material driven by the spin Hall effect, Phys. Rev. B **99**, 024405 (2019).

[41] A. Safin, V. Puliafito, M. Carpentieri, G. Finocchio, S. Nikitov, P. Stremoukhov, A. Kirilyuk, V. Tyberkevych, and A. Slavin, Electrically tunable detector of THz-frequency signals based on an antiferromagnet, Appl. Phys. Lett. **117**, 222411 (2020).

[42] V. Puliafito, L. Sanchez-Tejerina, M. Carpentieri, B. Azzerboni, and G. Finocchio, Modulation, injection locking, and pulling in an antiferromagnetic spin-orbit torque oscillator, IEEE Trans. Magn. **57**, 4100106 (2021).

[43] T. Kampfrath, T. Jungwirth, V. Novák, P. Wadley, R. P. Campion, P. Kužel, T. Seifert, P. Němec, M. Baumgartner, Z. Kašpar, J. Sinova, P. Gambardella, J. Wunderlich, M. Müller, and K. Olejník, Terahertz electrical writing speed in an antiferromagnetic memory, Sci. Adv. **4**, eaar3566 (2018).

[44] www.petaspin.com