Study of the robustness of neural networks based on spintronic neurons

Eleonora Raimondo¹, Anna Giordano¹, Andrea Grimaldi¹, Vito Puliafito², Mario Carpentieri³*, Zhongming Zeng⁴, Riccardo Tomasello⁵, and Giovanni Finocchio¹*

¹ Department of Mathematical and Computer Sciences, Physical Sciences and Earth Sciences, University of Messina, I-98166, Messina, Italy
² Department of Engineering, University of Messina, I-98166, Messina, Italy
³ Department of Electrical and Information Engineering, Polytechnic of Bari, via Orabona 4, 70125 Bari, Italy
⁴ Key Laboratory of Multifunctional Nanomaterials and Smart Systems, Suzhou Institute of Nano-tech and Nano-bionics, Chinese Academy of Sciences, Ruoshui Road 398, Suzhou 215123, PR China
⁵ Institute of Applied and Computational Mathematics, FORTH, GR-70013 Heraklion, Crete, Greece
* Senior Member, IEEE

Abstract

Spintronic technology is emerging as a direction for the hardware implementation of neurons and synapses of neuromorphic architectures. In particular, a single spintronic device can be used to implement the nonlinear activation function of neurons. Here, we propose how to implement spintronic neurons with a sigmoidal and ReLU-like activation functions. We then perform a numerical experiment showing the robustness of neural networks made by spintronic neurons all having different activation functions to emulate device-to-device variations in a possible hardware implementation of the network. Therefore, we consider a vanilla neural network implemented to recognize the categories of the Mixed National Institute of Standards and Technology database, and we show an average accuracy of 98.87 % in the test dataset which is very close to the 98.89% as obtained for the ideal case (all neurons have the same sigmoid activation function). Similar results are also obtained with neurons having a ReLU-like activation function.

Corresponding authors: Riccardo Tomasello; Giovanni Finocchio (e-mail: rtomasello@iacm.forth.gr; gfinocchio@unime.it)
I. INTRODUCTION

The implementation of neuromorphic computing, paradigm based on a combination of synapses and neurons, is very appealing for the potential to process large amount of data and to perform complex operations using low-power electronic circuits. In this respect, the research is very active both at the software level with the development of Graphics Processing Unit (GPU)-based accelerators [1], and at the hardware level with the direct implementation of synaptic and neuronal functionalities with Field Programmable Gate Arrays [2], Application Specific Integrated Circuit [3], memristors [4] and spintronic technology. In particular, spintronic devices, such as magnetic tunnel junctions (MTJs), promise to reduce significantly computational costs and energy consumption for the hardware implementation of synapses and neurons [5–9]. In addition, Spintronics has the advantage to offer integrability with CMOS systems, high working speeds (GHz or even THz) and intrinsic nonlinearity. The latter has been the successful property exploited in reservoir computing for recognizing spoken digits [10], as well as spoken vowels [11]. The contribution of this work is twofold. First, we propose MTJ-based neurons having sigmoid and ReLU-like activation functions (AFs) improving the ideas proposed in previous works [12–14]. However, when a specific software neural network architecture is transferred to the hardware level, differences in the synaptic weight and neuron AFs can occur because of the geometrical and physical device-to-device variations. Therefore, the second part of this work studies the robustness of the accuracy of a simple neural network architecture in presence of the abovementioned variations considering neuron-to-neuron variations in the fully-connected (FC) layer.

II. SPINTRONIC NEURONS

A. Sigmoidal activation function

An experimental implementation of a neuron with the sigmoid AF proposed in [12] is based on a superparamagnetic MTJ. An external out-of-plane field controls the jump rate coded in an average resistance value, which shows a sigmoidal behavior (see Fig. 2(d) in [12]), between two resistance states, parallel (R_P) or antiparallel (R_AP) depending on the relative orientations of the magnetizations of the free and polarizer layers. At zero external field, the output resistance is (R_P + R_AP)/2, while at large-enough positive (negative) fields, it is R_P (R_AP). However, a drawback of this configuration is that long times of the order of ms are necessary to obtain a reliable value of the output resistance. To solve this limit, here, we propose a neuron with a deterministic sigmoid AF. This can be implemented using a hybrid MTJ where the free layer has a perpendicular magnetization \( \mathbf{m} \) and the polarizer magnetization \( \mathbf{p} \) is in-plane (see
sketch in the bottom-right inset of Fig. 1(a)). The AFs are calculated with micromagnetic computations by numerically integrating the Landau-Lifshitz-Gilbert (LLG) equation [15,16]:

\[
\frac{dm}{d\tau} = -(m \times h_{eff}) + \alpha_G \left( m \times \frac{dm}{d\tau} \right),
\]

where \( \alpha_g \) is the Gilbert damping, \( m = M / M_s \) is the normalized magnetization of the MTJ free layer, and \( \tau = \gamma_0 M_s t \) is the dimensionless time, with \( \gamma_0 \) being the gyromagnetic ratio, and \( M_s \) the saturation magnetization. \( h_{eff} \) is the normalized effective magnetic field, which includes the exchange, perpendicular anisotropy, magnetostatic and external fields.

We analyze an elliptical 20x18x1 nm\(^3\) MTJ with a discretization cell of 1x1x1 nm\(^3\). The nominal physical parameters are \( M_s=1100 \text{ kA/m} \), exchange constant \( A=20 \text{ pJ/m} \), and \( \alpha_g =0.02 \).

At zero external field the free layer magnetization is out-of-plane (perpendicular anisotropy constant \( K_u=0.6215 \text{ MJ/m}^3 \)), therefore the x-component is zero and so it is the output resistance \( R(\theta) = R_p + (R_{sp} - R_p)[1 - \cos(\theta)]/2 \), where \( \theta \) is the angle between free layer and polarizer magnetization vectors. An external field \( H_x \) applied along the x-axis controls the equilibrium direction of the magnetization, which will eventually align with \( H_x \). The simulation results are shown in the main panel of Fig. 1(a) (red circles) and are well-fitted by an ideal sigmoid \( \sigma(x) = \frac{1}{1 + e^{-\alpha x}} \) with \( \alpha =1 \), thus confirming that this MTJ configuration can be used as a deterministic neuron with a sigmoid AF. The \( H_x \) can be generated by a DC current flowing in a nanowire on top of the MTJ as sketched in the bottom-right area of Fig. 1(a). Specifically, in our model, a positive (negative) field is generated by a negative (positive) current \( I_{DC} \), therefore promoting the parallel (antiparallel) state.

To study the effect of device-to-device variations in implementing the FC layer, we have performed simulations considering a maximum variation of the anisotropy constant ranging from \( K_u = 0.6225 \text{ to } 0.6200 \text{ MJ/m}^3 \). The previous anisotropy values still yield a sigmoid AF with coefficient \( \alpha =0.91 \) and \( \alpha =1.30 \), respectively, as shown in the top-left inset of Fig. 1(a). The use of a larger \( K_u \), while maintaining the other parameters fixed, ensures a stable perpendicular easy-axis also in the presence of thermal fluctuations at room temperature. In this case, we still find a sigmoidal activation function (not shown) which is linked to the time-averaged value of the magnetization x-component. We wish to highlight that it is possible to control the anisotropy by finely tuning it with a voltage applied across the MTJ [17–20]. This is the reason we are considering small variations. This aspect must be considered important otherwise the neurons can work at very different input fields.

An alternative implementation of a deterministic neuron with a sigmoid AF can also be realized by
taking advantage of the spin-transfer torque \cite{21,22} due to a spin-polarized current (see Fig. 1(b)):

\[
\tau_{oof} = \frac{g \mu_B J_{\text{DC}}}{\gamma_0 e M^2 S_{\text{FM}} L_{\text{FM}}} \left[ \frac{2 \eta}{1 + \eta^2 (\mathbf{m} \cdot \mathbf{p})} \right] \left[ \mathbf{m} \times (\mathbf{m} \times \mathbf{p}) \right]
\]  

(2)

with \( g \) being the Landé factor, \( \mu_B \) the Bohr magneton, \( J_{\text{DC}} \) the current density flowing through the MTJ, \( e \) the electron charge, \( L_{\text{FM}} \) the thickness of the free layer, and \( \eta = 0.66 \) the spin polarization.

This solution does not require an external field at the cost of a larger area occupancy. Indeed, it is necessary to use two hybrid MTJs in parallel with two diodes that select which MTJ to bias depending on the current sign (see inset in Fig. 1(b)). This dual-branch scheme is needed because, for one sign of the current either a self-oscillation state or switching dynamics is excited depending on the current amplitude \cite{23–26}.

B. ReLU-like activation function

Spintronic diodes have been demonstrated to be very promising for microwave detection \cite{27–29} overcoming the thermodynamics limit of Schottky diodes. In addition, those devices can be used for neuromorphic applications. For example, in \cite{13}, it has been shown how to implement a neuron with a ReLU-like AF. The key aspect is to consider the ascending branch of the rectified signal obtained as a function of \( I_{\text{DC}} \) (see Fig. 3(a) in \cite{13}).
Fig. 1. (a) Example of a magnetoresistance curve of the MTJ device as a function of $I_{\text{DC}}$ as computed from micromagnetic simulations for $K_u = 0.6215$ MJ/m$^3$ (red dots) fitted by the ideal sigmoid function (black line). Top-left inset: magnetoresistance curves of the MTJ device for $K_u = 0.6225$ MJ/m$^3$ and 0.6200 MJ/m$^3$ which can be described by a sigmoid with a slope coefficient $a = 0.91$ and 1.3, respectively. Bottom-right inset: a sketch of the deterministic neuron sigmoid AF composed of an MTJ and a current line. The $I_{\text{DC}}$ current generates the in-plane magnetic field which aligns the free layer magnetization $m$ parallel or antiparallel to the polarizer magnetization $p$. (b) Magnetoresistance curve of the neuron concept involving two MTJs and two diodes (inset). The red dots represent the micromagnetic simulations results fitted by an ideal sigmoid function (solid black line). (c) Experimental rectified voltage $V_{\text{DC}}$ as a function of DC current. The red dots represent the experimental data from Ref. [30] fitted by a third-order polynomial function (red line). Top-left inset: fitted experimental data with different spreading angles as indicated in the legend. Bottom-right inset: sketch of the MTJ spintronic diode where the rectification effect is achieved via the injection locking due to a locally-injected AC current $I_{\text{AC}}$ [30].

Such a behavior suggests the idea to design spintronic diodes exhibiting asymmetric rectified curves with the ascending branch as large as possible. We found that this characteristic has been already
experimentally observed in [30], where a spintronic diode with a sensitivity exceeding $10^5$ V/W at zero bias field has been measured. The device structure is still an MTJ having a tilted free layer and in plane polarizer (see inset Fig. 1(c)) which exhibits the abovementioned asymmetric rectification response (see Fig. 3(d) in Ref. [30]). Such an experimental $V_{DC} - I_{DC}$ curve reproduces a ReLU-like AF that can be described by a third-order polynomial function (see Fig. 1(c)). For this device, we have accounted device-to-device variations by performing a linear transformation (rotation of $\theta = \pm 5^\circ$) of the diode response, as shown in the top-left inset of Fig. 1(c).

![Fig. 2. Sketch of the structure of the CNN we construct for the recognition task](image)

### III. APPLICATION IN CONVOLUTIONAL NEURAL NETWORKS

In this last part of the work, we have trained a vanilla convolutional neural network (CNN) to recognize the categories of the Mixed National Institute of Standards and Technology (MNIST) database (indeed, we also used the Fashion-MNIST dataset obtaining a slightly smaller accuracy larger than 91%), consisting of grayscale images with 28x28 size organized in ten categories. Figure 2 depicts the network whose feature learning part is composed of a single convolutional layer with 64 filters of size 3x3. Its output feature maps propagate through the ReLU activation function and a pooling layer that halves the results’ spatial dimension with a max pooling operation. The CNN is then characterized by a flatten layer and an FC layer with 128 neurons. Finally, there is an FC layer with 10 neurons (softmax activation function) which returns the 10 probability outputs, thus obtaining the class to which the input belongs. The first step of the study is the identification of the neural network parameters while considering all the neurons of the FC layer having the same AF and, to prevent the overfitting, dropout layers and early-stopping are used. Secondly, with the weights of the ideal CNNs (all neurons of the FC layer have the same AFs), we also create 10000 different instances of the CNN each of them characterized by a nonideal FC layer (all the neurons of the FC have different AFs). Below, we detailly describe the two studies performed.
A. Results for the sigmoid AF

We train the network using the FC with ideal sigmoid AF $\alpha = 1$, and we obtain, for the validation set, a percentage of accuracy of 98.93% (Fig. 3(a)) and, for the test set, 98.89% (ideal accuracy in Fig 3(b)). We also found a negligible effect on the accuracy when two MTJs with different perpendicular anisotropy are used for the implementation of the AF with current.

As already explained above, we have used this trained network to test the effect of the device-to-device variation in the form of different magnetic anisotropy values. We set different slopes for the sigmoid, sampling 128 random values from a Gaussian distribution with a unitary mean (slope of the ideal sigmoid) and a standard deviation of 0.2 (all the negative values are discarded, and the values are re-sampled). We consider 10000 realizations to build the statistics of the accuracy evaluated with the test data. Figure 3(b) summarizes the results, as it can be seen the accuracy can be well described by a Gaussian distribution with mean of 98.87%.

B. Results for the ReLU-like AF
We also train the network using the FC with the ReLU-like AF shown in Fig. 1(c), and we obtain, for the validation set, a percentage of accuracy of 98.95% (Fig. 3(c)) while, for the test set, 98.92% (ideal accuracy in Fig. 3(d)). These results are comparable with the 98.87% accuracy obtained with the ideal ReLU AF which returns the maximum between 0 and the input value. Similarly to the case of the sigmoid AF, we have also studied the effect of device-to-device variation considering AF curves between the boundary shown in the top-left inset of Fig. 1(c) (−5° < θ < 5°) using a Gaussian distribution with zero mean and standard deviation of 3. The results achieved from 10000 realizations are shown in Fig. 3(d). The values are well-described by a Gaussian distribution with mean of 98.91%.

IV. CONCLUSION

In summary, we have proposed device concepts for spintronic neurons based on the use of hybrid MTJs (perpendicular free layer and in-plane polarizer) where it is possible to achieve either a sigmoid or ReLU-like AF. We have also studied the robustness with respect to device-to-device variations of a simple neural network, where the training is performed in an ideal scenario with all neurons having the same AF, while the test accuracy is computed considering the FC layer of the network with all neurons having a different AF. We have obtained promising results with the test accuracy exhibiting a small variation of less than few % for both sigmoid and ReLU-like AFs.

ACKNOWLEDGMENT

R.T. and G.F. thank the project ThunderSKY funded from the Hellenic Foundation for Research and Innovation (HFRI) and the General Secretariat for Research and Technology (GSRT) under Grant No. 871 and the support by PETASPIN association.

This work was also supported under Grant No. 2019-1-U.0. (“Diodi spintronici rad-hard ad elevate sensitivitá — DIOSPIN”) funded by the Italian Space Agency (ASI) within the call “Nuove idee per la componentistica spaziale del futuro”.

REFERENCES

[1] S. Chetlur, C. Woolley, P. Vandermersch, J. Cohen, J. Tran, B. Catanzaro, and E. Shelhamer, CuDNN: Efficient Primitives for Deep Learning, (2014).
[2] C. Zhang, P. Li, G. Sun, Y. Guan, B. Xiao, and J. Cong, Optimizing FPGA-Based Accelerator Design for Deep Pdf, ACM/SIGDA Int. Symp. Field-Programmable Gate Arrays(FPGA) 161 (2015).
[3] S. Han, X. Liu, H. Mao, J. Pu, A. Pedram, M. A. Horowitz, and W. J. Dally, EIE: Efficient Inference Engine on Compressed Deep Neural Network, in 2016 ACM/IEEE 43rd Annual International Symposium on Computer Architecture (ISCA), Vol. 16 (IEEE, 2016), pp. 243–254.
[4] 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. (2018).

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

[6] G. Finocchio, M. Di Ventra, K. Y. Camsari, K. Everschor-Sitte, P. Khalili Amiri, and Z. Zeng, *The Promise of Spintronics for Unconventional Computing*, J. Magn. Magn. Mater. 521, 167506 (2021).

[7] Y. Shao, S. L. Sinaga, I. O. Sunmola, A. S. Borland, M. J. Carey, J. Katine, V. Lopez-Dominguez, and P. Khalili Amiri, *Implementation of Artificial Neural Networks Using Magnetoresistive Random-Access Memory-Based Stochastic Computing Units*, IEEE Magn. Lett. 11, 1 (2021).

[8] R. Tomasello, A. Giordano, F. Garescì, G. Siracusano, S. De Caro, C. Ciminelli, M. Carpentieri, and G. Finocchio, *Role of Magnetic Skyrmions for the Solution of the Shortest Path Problem*, J. Magn. Magn. Mater. 532, 167977 (2021).

[9] S. Liu, C. Bennett, J. Friedman, M. Marinella, D. Paydarfar, and J. A. Incorvia, *Controllable Reset Behavior in Domain Wall–Magnetic Tunnel Junction Artificial Neurons for Task-Adaptable Computation*, IEEE Magn. Lett. 12, 1 (2021).

[10] J. Torrejon, M. Riou, F. A. Araujo, S. Tsunegi, G. Khalsa, D. Querlioz, P. Bortolotti, V. Cros, K. Yakushiji, A. Fukushima, H. Kubota, S. Yuasa, M. D. Stiles, and J. Grollier, *Neuromorphic Computing with Nanoscale Spintronic Oscillators*, Nature 547, 428 (2017).

[11] 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. Vodenicairevic, T. Hirtzlin, N. Locatelli, D. Querlioz, and J. Grollier, *Vowel Recognition with Four Coupled Spin-Torque Nano-Oscillators*, Nature 563, 230 (2018).

[12] J. Cai, B. Fang, L. Zhang, W. Lv, B. Zhang, T. Zhou, G. Finocchio, and Z. Zeng, *Voltage-Controlled Spintronic Stochastic Neuron Based on a Magnetic Tunnel Junction*, Phys. Rev. Appl. 11, 1 (2019).

[13] 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).

[14] Y. B. Chen, X. K. Yang, T. Yan, B. Wei, H. Q. Cui, C. Li, J. H. Liu, M. X. Song, and L. Cai, *Voltage-Driven Adaptive Spintronic Neuron for Energy-Efficient Neuromorphic Computing*, Chinese Phys. Lett. (2020).

[15] A. Giordano, G. Finocchio, L. Torres, M. Carpentieri, and B. Azzerboni, *Semi-Implicit Integration Scheme for Landau–Lifshitz–Gilbert-Slonczewski Equation*, J. Appl. Phys. 111, 07D112 (2012).

[16] L. Lopez-Diaz, D. Aurelio, L. Torres, E. Martinez, M. a Hernandez-Lopez, J. Gomez, O. Alejos, M. Carpentieri, G. Finocchio, and G. Consolo, *Micromagnetic Simulations Using Graphics Processing Units*, J. Phys. D. Appl. Phys. 45, 323001 (2012).

[17] J. Zhu, J. A. Katine, G. E. Rowlands, Y.-J. Chen, Z. Duan, J. G. Alzate, P. Upadhyaya, J. Langer, P. K. Amiri, K. L. Wang, and I. N. Krivorotov, *Voltage-Induced Ferromagnetic Resonance in Magnetic Tunnel Junctions*, Phys. Rev. Lett. 108, 197203 (2012).

[18] W.-G. Wang, M. Li, S. Hageman, and C. L. Chien, *Electric-Field-Assisted Switching in Magnetic Tunnel Junctions*, Nat. Mater. 11, 64 (2012).

[19] M. Carpentieri, R. Tomasello, M. Ricci, P. Burrascano, and G. Finocchio, *Micromagnetic Study of Electrical-Field-Assisted Magnetization Switching in MTJ Devices*, IEEE Trans. Magn. 50, 4 (2014).

[20] A. Fassatoui, J. P. Garcia, L. Ranno, J. Vogel, A. Bernard-Mantel, H. Béa, S. Pizzini, and S. Pizzini,
Reversible and Irreversible Voltage Manipulation of Interfacial Magnetic Anisotropy in Pt/Co/Oxide Multilayers, Phys. Rev. Appl. 14, 064041 (2020).

[21] J. C. Slonczewski and J. Z. Sun, Theory of Voltage-Driven Current and Torque in Magnetic Tunnel Junctions, J. Magn. Magn. Mater. 310, 169 (2007).

[22] R. Tomasello, V. Puliafito, B. Azzerboni, and G. Finocchio, Switching Properties in Magnetic Tunnel Junctions with Interfacial Perpendicular Anisotropy: Micromagnetic Study, IEEE Trans. Magn. 50, 7100305 (2014).

[23] A. Slavin and V. Tiberkevich, Auto-Oscillator Theory of Microwave Generation by Spin-Polarized Current, IEEE Trans. Magn. 45, 1875 (2009).

[24] 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).

[25] C. Fowley, V. Sluka, K. Bernert, J. Lindner, J. Fassbender, W. H. Rippard, M. R. Pufall, S. E. Russek, and A. M. Deac, Zero-Field Spin-Transfer Oscillators Combining in-Plane and out-of-Plane Magnetized Layers, Appl. Phys. Express 7, 043001 (2014).

[26] M. Kazemi, E. Ipek, and E. G. Friedman, Adaptive Compact Magnetic Tunnel Junction Model, IEEE Trans. Electron Devices 61, 3883 (2014).

[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. K. Amiri, G. Finocchio, and Z. Zeng, Giant Spin-Torque Diode Sensitivity in the Absence of Bias Magnetic Field, Nat. Commun. 7, 11259 (2016).

[28] B. Fang, M. Carpentieri, S. Louis, V. Tiberkevich, A. Slavin, I. N. Krivorotov, R. Tomasello, A. Giordano, H. Jiang, J. Cai, Y. Fan, Z. Zhang, B. Zhang, J. A. Katine, K. L. Wang, P. K. Amiri, G. Finocchio, and Z. Zeng, Experimental Demonstration of Spintronic Broadband Microwave Detectors and Their Capability for Powering Nanodevices, Phys. Rev. Appl. 11, 014022 (2019).

[29] R. Tomasello, B. Fang, P. Artemchuk, M. Carpentieri, L. Fasano, A. Giordano, O. V. Prokopenko, Z. M. Zeng, and G. Finocchio, Low-Frequency Nonresonant Rectification in Spin Diodes, Phys. Rev. Appl. 14, 024043 (2020).

[30] L. Zhang, B. Fang, J. Cai, M. Carpentieri, V. Puliafito, F. Garesci, 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).