Neuromorphic spintronics simulated using the Thiele equation approach

Anatole MOUREAUX,[1, a)] Simon DE WERGIFOSSE,[1] Chloé CHOPIN,[1] Jimmy WEBER,[1] and Flavio ABREU ARAUJO[1, b)]
Institute of Condensed Matter and Nanosciences, UCLouvain, Louvain-la-Neuve, Belgium

A hardware neural network based on a single spin-torque vortex nano-oscillator (STVO) is designed using time-multiplexing and simulated with our latest and most accurate models based on the Thiele equation approach (TEA). Different mathematical and numerical adaptations are brought in order to increase the accuracy and the speed of the simulations. A benchmark task of waveform detection is designed to assess the performance of the resulting neural network in the framework of reservoir computing and compare the different models. The obtained results allow to conclude on the ability of the system to effectively classify sine and square signals with high recognition rates and low root-mean-squares deviations. Given the high throughput of the simulations, two unprecedented parametric studies on the dc bias current and the level of noise in the input signal are performed to demonstrate the value of our models. The efficiency of our system is also tested during speech recognition and shows the agreement between these models and the corresponding experimental measurements.

I. INTRODUCTION

The need for low power and efficient hardware dedicated to machine learning has led to a new type of electronics called neuromorphic computing[1]. By taking inspiration from the brain, it tries to overcome the von Neumann bottleneck by proposing artificial neurons or synapses that are highly interconnected with a parallel architecture. Different systems are under study like photonics[2], memristors[3] and spintronics[4]. Among the latter, spin torque vortex nano-oscillators (STVOs) have already been proved to be choice candidates to implement hardware neurons for machine learning applications. Thanks to their highly non-linear behavior and short-term memory, several machine learning tasks such as waveform and speech recognition have been performed successfully with STVO-based neural networks[5–7]. Their small size, low power consumption and CMOS-compatibility reinforce their potential interest concerning the development of neuromorphic computing systems.

The simulation of such neural networks is of prior importance to better understand the underlying phenomena involved in their cognitive properties, and to optimize these systems before the actual fabrication. Thus, a fast and quantitative model is needed. Indeed, it would require a huge computational power and an extensive amount of time to study with micromagnetic simulations the dynamics of such oscillators as well as to test several neuromorphic architectures or input parameters that could be optimized for experimental measurements. Several solutions are proposed like using a model based on non-linear magnetic oscillator theory[8,9], using machine learning to predict the dynamics of STVOs with Neural ODEs[10] or using analytical models based on the Thiele equation approach (TEA)[11,12] for simulating STVOs.

Abreu Araujo *et al*[9] have already compared experimental results[5] and results from simulated STVO with the non-linear magnetic oscillator theory[8] for the recognition of spoken digit with reservoir computing. The parameters needed for the non-linear magnetic oscillator model are extracted experimentally and there is an excellent agreement between experimental and simulated results. However, the experimental results are surprisingly better than the simulated results. On the contrary, the function given by Neural ODEs[10] with the addition of noise allows to predict perfectly the experimental results of Torrejon *et al.*[5] with an acceleration of 200 compared to micromagnetic simulations and in two hours instead of a week of experiments. Still, the function given by the neural network is a black box and one does not have access to the underlying physics of the oscillator. Thus, TEA models which are fully analytical are an interesting solution. However, they give quantitative results only for the damping regimes[13] (*i.e.* when the vortex core damps back to the magnetic dot center) and for the steady-state regime[14] (*i.e.* when the vortex core reaches a stable orbit). For the transient regimes of the vortex core (*i.e.* when the vortex core evolve towards the magnetic dot center or a stable steady-state orbit) which is the regime of interest for reservoir computing application, TEA models are only qualitative. A recent data-driven TEA (DD-TEA) model[15] described quantitatively both the damping and steady-state regime as well as the transient dynamics of the STVO's vortex core dynamics with, in addition, an acceleration of six order of magnitude compared to micromagnetic simulation. It allows to simulate days or month of simulations in a few minutes. This quantitative and ultra-fast model is simply based on TEA and a few micromagnetic simulations.

The speed of this DD-TEA model can be further improved as shown by the two analytical models proposed below. Indeed, the *s*-analytical low-order (LO) TEA model and the *s*-analytical high-precision (HP) TEA model described later reach a nine order of magnitude acceleration compared to micromagnetic simulation. Theses two models are then applied to three tasks. The first task is a proof of concept of pattern recognition where sine and square functions have to be differentiate. The second task is a parametric study of STVO results depending on the dc input current and the SNR of the sine and square functions. The last task is the comparison with experimental results from Torrejon *et al.*[5], the phenomenological model from Abreu Araujo *et al.*[9], and the results from the two *s*- LO and HP analytical models.

[a)]anatole.moureaux@student.uclouvain.be
[b)]flavio.abreuaraujo@uclouvain.be; https://flavio.be

II. METHODS

The vortex core dynamics can by described by a simple harmonic oscillator equation[13] when the pulsation ω is constant and the transient dynamic factor $\Gamma \to 0$:

$$\begin{bmatrix} X(t) \\ Y(t) \end{bmatrix} = ||\mathbf{X}||e^{\Gamma t} \begin{bmatrix} \sin(\omega t) \\ -\cos(\omega t) \end{bmatrix} \tag{1}$$

where $X(t)$ and $Y(t)$ are the Cartesian coordinates of the vortex core as a function of time. As both $X(t)$ and $Y(t)$ vary quickly in time, Eq. 1 can be rewritten by considering the reduced vortex core position $s(t) = \frac{\sqrt{X(t)^2+Y(t)^2}}{R}$ where R is the radius of the magnetic dot. This gives:

$$s(t) = s_\infty e^{\Gamma t} \tag{2}$$

where s_∞ is the final vortex orbit and Γ is a constant. In reality, Γ is a function which depends on $s(t)$ and Eq. 1 has a general solution:

$$s(t) = s_\infty \exp \int_0^t \Gamma(s(t'))dt' \tag{3}$$

After a few developments not detailed here, Eq. 3 can be expressed as the following ordinary differential equation (ODE):

$$\dot{s}(t) = \Gamma(s(t))s(t) \tag{4}$$

The fully analytical expression of Γ as a function of s based on TEA is given in a previous publication[13]. Here, we can truncate Γ to the second order:

$$\Gamma(s) = \alpha + \beta s^2 \tag{5}$$

Both α and β are functions depending on the input current density J and a, a_J, b, b_J are fully analytically described (not shown here).

$$\alpha(J) = a_J J + a \tag{6}$$

$$\beta(J) = b_J J + b \tag{7}$$

Finally, by injecting Eq. 5 in Eq. 4, the following equation appears:

$$\dot{s}(t) = \alpha s(t) + \beta s^3(t) \tag{8}$$

Eq. 8 is a Bernoulli differential equation and it has the following solution:

$$s(t) = \frac{s_0}{\sqrt{\left(1 + \frac{s_0^2}{\alpha/\beta}\right)\exp(-2\alpha t) - \frac{s_0^2}{\alpha/\beta}}} \tag{9}$$

Where s_0 is the initial reduced vortex core position at $t = 0$. The final position of s, *i.e.* when $t \to \infty$, depends on the input current density J:

$$s_\infty(J) = \sqrt{\frac{-\alpha(J)}{\beta(J)}} \tag{10}$$

Guslienko *et al.*[14] have reported a similar expression of the transient regime obtained with a different method. This s-analytical LO-TEA is two order of magnitude faster than DD-TEA as no ODEs need to be numerically solved. In addition, the data-driven method used in DD-TEA[15] can be adapted to this model. Indeed, $\alpha(J)$ and $\beta(J)$ can be obtained by fitting a few micromagnetic simulations, leading to a precise description of the transient regime as well as the damping and steady-state regimes. This model combine a physical origin and the precision of a data-driven method.

The second model is an extension of the first one, allowing to account for any additional non-linearity in the dynamics of the vortex core that would not be accounted for in Eq. 9 due to the truncation to the second order of the $\Gamma(s)$ parameter. To do so, Eq. 9 is generalized to the order n (Eq. 11), where $n(J)$ is a fifth-order polynomial of the current density J injected into the STVO, whose coefficients have been determined by fitting to micromagnetic simulations results. This model was hence called the *s-analytical High-Precision Thiele equation approach* model (s-analytical HP-TEA).

$$s(t) = \frac{s_0}{\sqrt[n]{\left(1 + \frac{s_0^n}{\alpha/\beta}\right)\exp(-n\alpha t) - \frac{s_0^n}{\alpha/\beta}}} \tag{11}$$

These two models, on top of being extremely fast to compute using a standard computer compared to micromagnetic simulations, yield accurate results thanks to the combination with micromagnetic simulations results. They were hence benchmarked using a simple machine learning task of waveform recognition.

The benchmark task consisted in the classification of sine and square periods composed of 8 samples (Fig. 1). To do so, a single-STVO neural network composed of 24 different neurons was designed using time multiplexing as described in ref.[5]. The intrinsic dynamics of the STVO allowing to generate the output signal was modeled using the two models (Eq. 9 and Eq. 11) successively, allowing to compare their performance.

To train the network, a database of 80 sine periods and 80 square periods arranged in an alternate fashion was used, and the training procedure described in ref.[9] was followed. For the testing, a similar yet randomly arranged database was preferred. The target related to a sine sample is the value 1, while the target related to a square sample is the value -1. To perfectly classify a given period, the 8 samples of the input signal must be classified correctly by the neural network. As any intermediate value between 1 and -1 can be returned by the neural network, any positive output value was considered as the estimation of a sine sample, while any negative value was considered as the estimation of a square sample. In the cases were the neural network was not able to yield a conclusive estimation (*i.e.*, the returned value was close to 0, making the decision uncertain), the output value was chosen randomly between 1 and -1 to implement random choice. This method is an update to the technique used previously in ref.[9] to simulate the behavior of the neural network during a machine learning

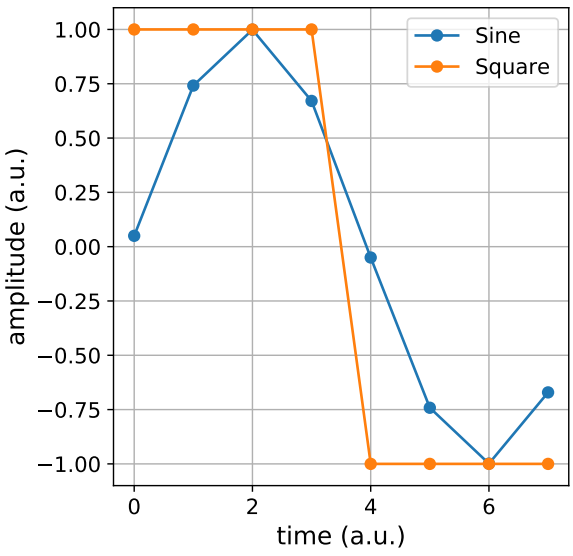

FIG. 1. Sine and square input signals used for benchmarking. Note that the sine period is slightly offset in order to avoid having two samples at $y = 0$.

task. This allowed to get rid of a bias in the results as it will be shown later.

The estimated value yielded by the neural network was constructed in two different ways, as presented in ref.[9]. The τ-Wise (TW) approach, consists in treating each of the 8 samples of a period independently. The recognition score is thus a value between 0% if none of the samples was correctly recognized, and 100% if all the samples were correctly recognized (see Fig. 2, top). This allows to assess the performance of the neural network at recognizing (sometimes very small) partial inputs of data (in this case $1/8$ of a period). The Winner-Takes-All (WTA) method involves averaging the estimation of the 8 samples of a period before choosing the value between 1 and -1 that is the closest to the average. This leads to a recognition score of 1 if the final value corresponds to the expected target, or 0 if not (see Fig. 2, bottom). This allows to absorb any small inaccuracies that may occur during the recognition of a period.

The diameter of the STVO was fixed to $d = 200$ nm, while its dc resistance R_{osc} was set to 140.6Ω, accordingly to the values presented in ref.[5,6,9]. Concerning the splitting phenomenon presented in ref.[13], only the C^+ configuration was considered, but all the calculations and results are extendable to the noOe or the C^- configurations.

The dc bias current, or working current I_{w} is used to trigger the gyrotropic motion of the vortex core. The input signal is added to it before entering the STVO. The value of I_{w} is important as it defines the dynamical regime in which the STVO will operate, and will be used to non-linearly process the data. I_{w} is represented by the black line in Fig. 3, while the dynamical regime accessible for the STVO is the colored range. Most of the time, I_{w} was set to 1.986 mA, but this value can actually be swept across a range of values in order to assess the influence of I_{w} on the results of the neural network during the benchmark task. The amplitude of the input signal was scaled up to $\Delta V = 150$ mV (centered on the working voltage related to I_{w}). It is represented by the red interval around the I_{w} in Fig. 3. The sampling rate was characterized with a time constant D_{t} of 50 nanoseconds. This parameter must also be chosen wisely as the transient state of the dynamics of the vortex core is the main contribution to the non-linear treatment of the data. Hence, a too small D_{t} will not allow to explore the transient state sufficiently, while a too long D_{t} will saturate the dynamics into the steady-state, hence decreasing the efficiency of the data treatment. Finally, a certain level of additive Gaussian white noise was added to the input signal to simulate experimental conditions. This noise is responsible for an additional range of voltage to be added on the signal (the blue intervals in Fig. 3). This value was most of the time fixed to $\Delta V_{\text{noise}} = 30$ mV, but it can also be swept on a range to assess the influence of the noise on the accuracy of the recognition. The addition of all these contributions to the effective current density injected into the STVO was sanity-checked before every simulation as it should not exceed J_{cr2}, the second critical current density representing the expulsion of the vortex state out of the STVO, introduced in ref.[13].

The results of the recognition were analyzed with two indicators : the recognition rate (or score) and the RMS deviation. The first indicator is the average of the recognition score for all the signals injected into the network. Due to the low complexity of the benchmark task, this ratio was equal to 100% in the large majority of the cases, if not in the totality. Hence, the RMS deviation was used to get a better insight of the quality of the recognition. This is the root-mean-square of the distance between the expected target for a given sample (*i.e.*, 1 for a sine period or -1 for a square period) and the estimation yielded by the neural network. This gives an estimation of the quality of the recognition as high RMS values correspond to non-optimal recognition even if the final estimation is correct. These indicators were averaged over all the 160 signals of the testing database.

Due to the high throughput of the simulations allowed by the new models, two parametric studies have been performed. The first one consisted in a sweep of the working current I_{w} to investigate the influence of the operating regime and the corresponding non-linearity on the treatment of the data. The working current was swept from $1.001 \times I_{\text{cr1}}$ (with I_{cr1} the first critical current introduced as in ref.[13]) to I_{cr2}, the second critical current. The other operating parameters are listed in Tab. I. For each value in the range swept, the recognition rate and the RMS deviation were recorded for both TW and WTA approaches, and so for both *s*-analytics LO-TEA and *s*-analytics HP-TEA models. To avoid the random fluctuations introduced with the background noise, the results were averaged over 200 simulations for each value of the parametric sweep.

The second parametric study consisted in a sweep of the signal-to-noise ratio (SNR). For a given value of the SNR (in dB), a corresponding value of additive Gaussian white noise voltage was added to the signal. Hence, a SNR of 0 dB corresponds to the case where the spectral power density of the input signal is equal to that of the added white noise, while

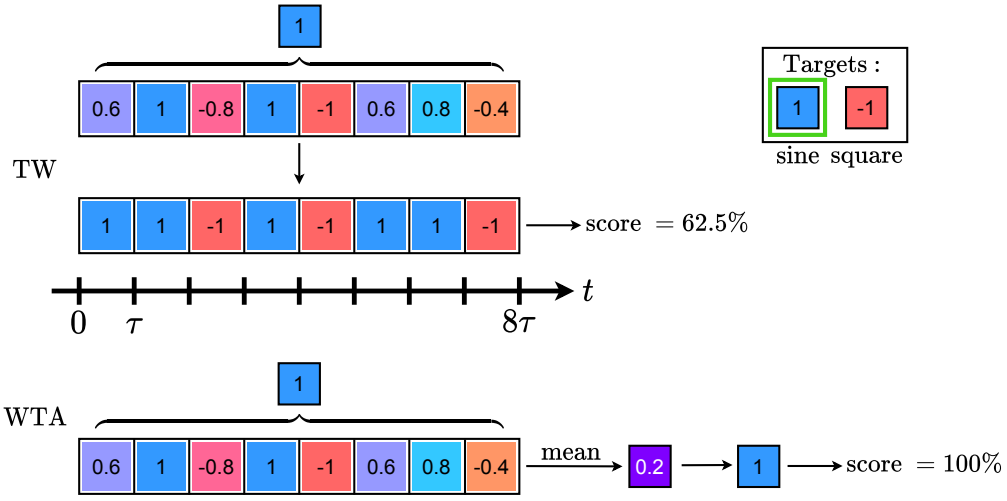

FIG. 2. Distinction between the TW and WTA approaches. The first and last arrays represent the result obtained by applying a final linear transformation on the output of the reservoir for a sine signal of length 8τ. The TW approach treats each sample of the signal independently, leading to a score between 0% and 100%. The WTA approach absorbs the errors occurring in the recognition of the whole signal by averaging the results, leading to a score of 100%.

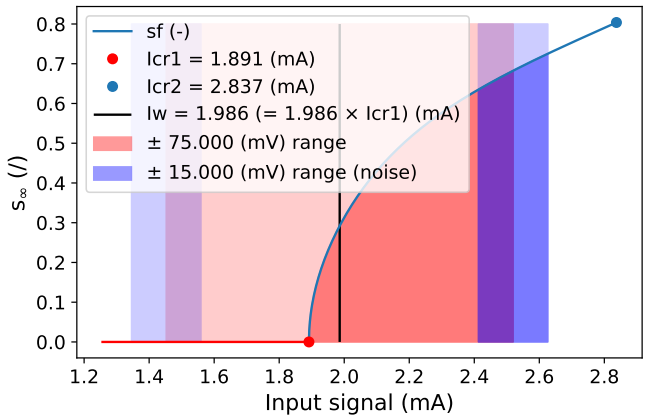

FIG. 3. Example of the current map and the range sounded by the dynamics of the oscillator. The black line corresponds to the working current I_w , the red range corresponds to a input voltage range of $\Delta V = 150$ mV, and the blue ranges correspond to an additional noise voltage range of $\Delta V_{noise} = 30$ mV. The curve is the steady-state reduced position of the vortex core s_∞ for each input current intensity.

positive and negative values corresponds respectively to the case were the power spectral density of the signal is higher and lower than that of the noise. The SNR was swept from -40 dB to 100 dB, and the other operating parameters are the same as the one listed in Tab. I (except for ΔV_{noise}, which is replaced with $I_w = 1.05 \times I_{cr1} = 1.986$ mA). The same variables as in the first parametric study were recorded, and were also averaged on 200 simulations.

The third and last task to which the models from Eq. 9 and Eq. 11 were applied was the same speech recognition task as in ref.[9] to compare the results with experimental measurements. To do so, the same database and code were used, and a similar neural network was designed. Only the non-linear treatment of the data was adapted with the new models.

R_{osc}	140.6 Ω
Chirality	C^+
D_t	50 ns
ΔV	150 mV
d	200 nm
ΔV_{noise}	50 mV

TABLE I. Input parameters for the working current parametric sweep

III. RESULTS AND DISCUSSION

The results of the comparison between the *s*-analytics LO-TEA and the *s*-analytics HP-TEA models, averaged over 1000 simulations, are listed in Tab. II. It can be seen that the two models are practically equally fast. Concerning the performance at the benchmark task, one can observe that in most of the cases, the *s*-analytics HP-TEA model yields slightly better results than the *s*-analytics LO-TEA model (higher recognition scores and lower RMS values). Given the high number of runs, these differences are unlikely to be due to the noise in the input signal, but rather to the models themselves, and more specifically to the better description of the complexity of the dynamics of the vortex core in the HP-TEA model. The *s*-analytics HP-TEA model, in addition to being the most accurate so far, also yields the better results while performing the benchmark task. This highlights the contribution of the non-linearity of the physical phenomena involved in the dynamics of the STVOs into the performance of the latter as hardware neurons. In a more general way, it can be seen that the scores are extremely close to a perfect recognition, while needing a minimal amount of time (the durations displayed in

		s-analytics LO-TEA	*s*-analytics HP-TEA
Training	Score (WTA)	99.99%	100%
	Score (TW)	99.77%	99.94%
	RMS (WTA)	0.235	0.198
	RMS (TW)	0.350	0.300
Testing	Score (WTA)	99.82%	99.73%
	Score (TW)	99.26%	99.39%
	RMS (WTA)	0.278	0.245
	RMS (TW)	0.402	0.348
Average elapsed time		0.684 s	0.713 s

TABLE II. Results of the benchmarking task for the *s*-analytics LO-TEA and the *s*-analytics HP-TEA models

II correspond to the training of the AI with 160 signals and the recognition with 160 other signals). It thus allows to conclude on the ability of the system to perform automatic classification.

The results obtained for the parametric sweep of the working current intensity with the two models are displayed in 4. It can be seen that for higher working current intensities, the performance of the neural network gets progressively better *i.e.*, the recognition score reaches an upper asymptote (at 100%), and the RMS deviation reaches a lower asymptote. These observations are valid for both models. It means that the non-linear regime in which the STVO operates at higher current intensities is beneficial for the recognition *i.e.*, it is less linear than at lower current intensities. It can also be noticed that the WTA approach systematically yields better results (higher recognition scores and lower RMS values) than the TW approach, thanks to the beneficial role of the average involved in the process that allows to get rid of small recognition errors. Furthermore, it can be seen that the asymptotical values obtained with the *s*-analytics HP-TEA model are slightly better than that obtained with the *s*-analytics LO-TEA model, confirming the conclusions drawn from Tab II. Although other interesting observations and conclusions could be made, this kind of study demonstrates the use of the new models to guide the design of an actual physical system, by specifying the best operating regime (for our system, I_w must lie above ~ 2.05 mA).

Concerning the parametric study on the SNR, the results are displayed in Fig. 5. It can be noticed that for the recognition scores, a sigmoid-like figure is obtained. For positive SNRs, the recognition reaches 100% due to the high-quality of the signal. As the SNR decreases, the score reaches 50%, which corresponds to random choice between the two categories (sine and square). This is due to the noise decreasing the quality of the signal to the point where no useful feature in the signal can be used by the neural network to classify the signals. Furthermore, the curves being averaged on 200 simulations, empirical logistic approximations could be fitted to the results (Eq. 12 for the recognition scores obtained with the LO-TEA model and Eq. 13 for the recognition scores obtained with the HP-TEA model). These relations allow to state that the average recognition score at 0 dB (*i.e.*, when the noise and the input signal have the same spectral power density) is 88.31% with the *s*-analytics LO-TEA model and 90.56%

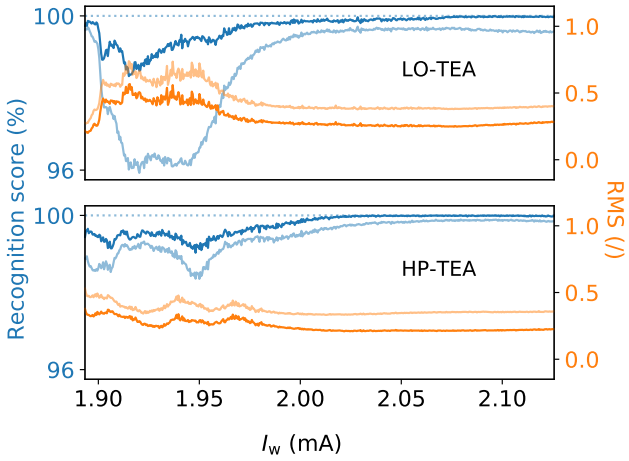

FIG. 4. Recognition score and RMS deviation of the recognition of the neural network emulated by a STVO simulated using the *s*-analytics LO-TEA model (top) and the *s*-analytics HP-TEA model (bottom) during a parametric sweep of the working current intensity, averaged over 200 simulations. The darker curves are obtained with the WTA approach, and the lighter ones are obtained with the TW approach.

with the *s*-analytics HP-TEA model (see Fig. 5), once again confirming the ability of the system. These relations, which would not have been possible to express accurately without the new models, could help to estimate precious information about a physical prototype, like the SNR once the recognition score has been properly determined. Unfortunately, no such conclusive observations could be made using the RMS, except the decrease to 0 as the SNR increases.

$$\mathrm{SCORE}(\mathrm{SNR}) \approx \frac{50}{1+\exp\left(\frac{-(\mathrm{SNR}+4.75)}{4}\right)} + 50 \quad (12)$$

$$\mathrm{SCORE}(\mathrm{SNR}) \approx \frac{50}{1+\exp\left(\frac{-(\mathrm{SNR}+4.43)}{3.04}\right)} + 50 \quad (13)$$

Finally, the two new models were used to perform the same speech recognition task as in ref.[9]. Indeed, it was observed that in some cases, the simulated results (with the phenomenological model based on non-linear magnetic oscillator theory) were lower than the experimental ones, hence motivating the development of the more accurate models based on physical relations presented in this work. However, the deletion of the bias presented earlier (occurring when the neural network is not able to classify the data) led to the decrease of this gap. That is why the difference between the experimental results and the results obtained with the phenomenological model is lower in this work than in ref.[9]. Still, the large difference between the phenomenological model and the experimental measurements when no acoustic filter is applied on the data motivates the comparison with the new models. The results

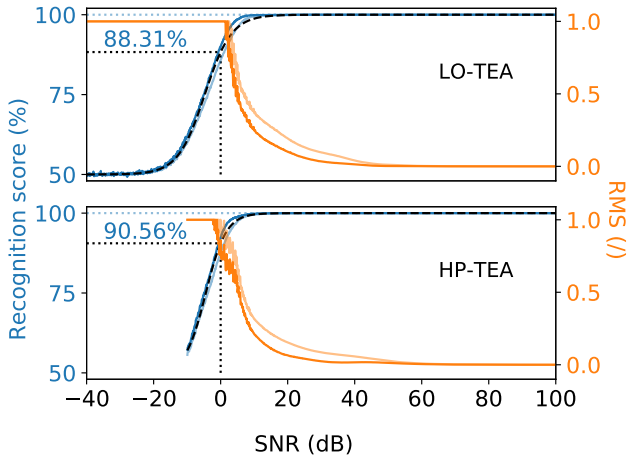

FIG. 5. Recognition score and RMS deviation of the emulated neural network with respect to the SNR using the *s*-analytics LO-TEA model (top) and the *s*-analytics HP-TEA model (bottom) averaged over 200 simulations. The RMS values were truncated to 1.00 due to high RMS deviations for negative SNRs. The darker curves are obtained with the WTA approach, and the lighter ones are obtained with the TW approach. The black dashed lines represent the approximations of Eq. 12 and Eq. 13.

obtained for the training and testing phases are displayed in Fig. 6 and Fig. 7. It can be seen that in all the cases where an acoustic filter is applied on the input, the results obtained with the two new models are higher or equal to the experimental ones, which is expected and supports the validity of the new models. When no acoustic filter is applied on the data, one can see that the gap between the LO-TEA and HP-TEA models and the experimental measurements is lower than the difference between the phenomenological model and the experiment during training (left column in Fig. 6). This reinforces the credibility of the two new models. However, the gap with the experiment is well higher during testing (left column in Fig. 7). This corresponds to a high overfitting phenomenon. Hence, that means that the physical sample is somewhat able to extract the useful features in the signal while the new models are struggling.

IV. CONCLUSION

Two new models were developed based on the Thiele equation approach. The combination of the mathematical expression with numerical results obtained with micromagnetic simulations allowed to obtain a physical description of the dynamics of the vortex core of STVOs under a given input signal with the same level of accuracy as MMS. Additional mathematical and numerical adjustments were brought to allow the analytical resolution of the models and the acceleration of the simulations. One of the model (LO-TEA) truncates the non-linearity of the dynamics to the second order, while the other model (HP-TEA) takes into account any additional non-linearity. A benchmark task consisting in the classification of sine and square periods was designed. It was observed that the simulations related to the two models require a similar amount of time, but the HP-TEA model yields slightly better results due to the consideration of the higher-order complexity of the dynamics non-linearity. The speed of the simulations allowed to design two parametric studies, concerning the working current intensity I_w and the level of noise in the input signal. The first study allowed to determine an optimal operating range for the use of the actual system, while the second study allowed to drawn unprecedented relations between the recognition rate and the input signal-to-noise ratio. These studies would have been impossible to make in practice using micromagnetic simulations, hence highlighting the value of analytical models. Finally, a comparison was made between the two new models and the values obtained for a speech recognition task in Refs.[5,9]. In most of the cases, a very good agreement is obtained between the new models and the experiment. However, it seems that the experimental setup performs much better than the new models at selecting the relevant features in the input signal, leading to strong overfitting.

V. ACKNOWLEDGEMENTS

Computational resources have been provided by the Consortium des Équipements de Calcul Intensif (CÉCI), funded by the Fonds de la Recherche Scientifique de Belgique (F.R.S.-FNRS) under Grant No. 2.5020.11 and by the Walloon Region. F.A.A. is a Research Associate of the F.R.S.-FNRS. S.d.W. aknowledges the Walloon Region and UCLouvain for FSR financial support. The experimental data originates from the measurements by Jacob Torrejon, Mathieu Riou, and F.A.A. while working in Julie Grollier's group at Unité Mixte de Physique CNRS/Thales in France and the samples were produced by Sumito Tsunegi from AIST in Japan.

VI. AUTHOR'S CONTRIBUTION

The study was designed by F.A.A. who created the analytical foundation related to the Thiele equation approach with the assistance of S.d.W. and C.C.. F.A.A. developed the two analytical models and combined them with numerical data obtained by micromagnetic simulations performed by S.d.W.. A.M. designed and performed the benchmark tasks, the parametric studies, and compared the numerical results with the experiment data. A.M. wrote the core of the manuscript and all the other co-authors (F.A.A., S.d.W., C.C., and J.W.) contributed to the text as well as to the analysis of the results.

VII. REFERENCES

[1]G. Indiveri and T. K. Horiuchi, Frontiers in neuroscience **5**, 118 (2011).
[2]B. J. Shastri, A. N. Tait, T. F. de Lima, W. H. Pernice, H. Bhaskaran, C. D. Wright, and P. R. Prucnal, Nature Photonics **15**, 102 (2021).
[3]S. H. Jo, T. Chang, I. Ebong, B. B. Bhadviya, P. Mazumder, and W. Lu, Nano letters **10**, 1297 (2010).

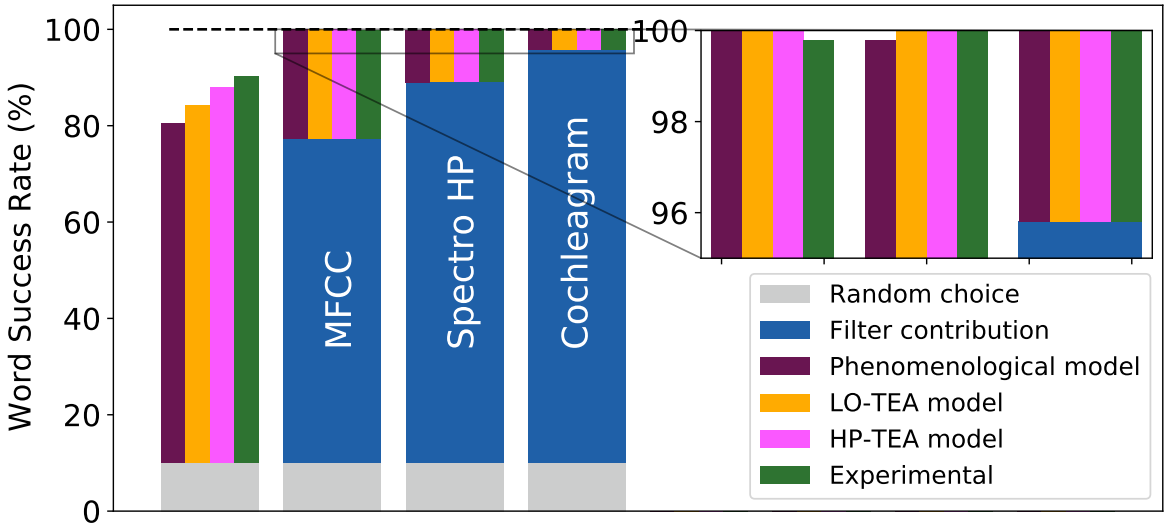

FIG. 6. Comparison between the different contributions to the recognition score (WSR), obtained with the WTA approach during training, with different acoustic filters. The first column represents the case with linear acoustic filtering (linear spectrogram). The green and purples columns are from ref.[9]. The orange columns represent the score obtained with the *s*-analytics LO-TEA model, and the pink columns represent the score obtained with the HP-TEA model.

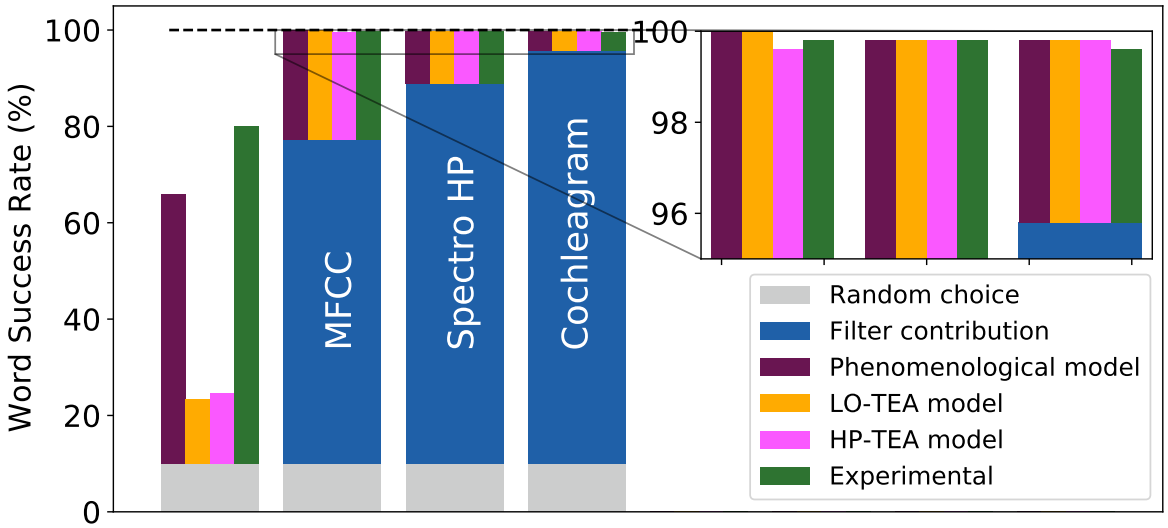

FIG. 7. Comparison between the different contributions to the recognition score (WSR), obtained with the WTA approach during testing, with different acoustic filters. The first column represents the case with linear acoustic filtering (linear spectrogram). The green and purples columns are from ref.[9]. The orange columns represent the score obtained with the *s*-analytics LO-TEA model, and the pink columns represent the score obtained with the HP-TEA model.

[4]J. Grollier, D. Querlioz, K. Camsari, K. Everschor-Sitte, S. Fukami, and M. D. Stiles, Nature electronics **3**, 360 (2020).
[5]J. Torrejon, M. Riou, F. Abreu Araujo, S. Tsunegi, G. Khalsa, D. Querlioz, P. Bortolotti, V. Cros, K. Yakushiji, A. Fukushima, *et al.*, Nature **547**, 428–431 (2017).
[6]M. Riou, J. Torrejon, B. Garitaine, F. Abreu Araujo, P. Bortolotti, V. Cros, S. Tsunegi, K. Yakushiji, A. Fukushima, H. Kubota, *et al.*, Physical Review Applied **12** (2019), 10.1103/physrevapplied.12.024049.
[7]D. Marković, N. Leroux, M. Riou, F. Abreu Araujo, J. Torrejon, D. Querlioz, A. Fukushima, S. Yuasa, J. Trastoy, P. Bortolotti, *et al.*, Applied Physics Letters **114**, 012409 (2019).
[8]A. Slavin and V. Tiberkevich, IEEE Transactions on Magnetics **45**, 1875 (2009).
[9]F. Abreu Araujo, M. Riou, J. Torrejon, S. Tsunegi, D. Querlioz, K. Yakushiji, A. Fukushima, H. Kubota, S. Yuasa, M. D. Stiles, *et al.*, Scientific Reports **10** (2020), 10.1038/s41598-019-56991-x.

[10] X. Chen, F. Abreu Araujo, M. Riou, J. Torrejon, D. Ravelosona, W. Kang, W. Zhao, J. Grollier, and D. Querlioz, Nature communications **13**, 1 (2022).
[11] A. Thiele, Physical Review Letters **30**, 230 (1973).
[12] D. Huber, Physical Review B **26**, 3758 (1982).
[13] F. Abreu Araujo, C. Chopin, and S. de Wergifosse, Scientific Reports **12** (2022), 10.1038/s41598-022-14574-3.
[14] K. Y. Guslienko, O. V. Sukhostavets, and D. V. Berkov, Nanoscale research letters **9**, 1 (2014).
[15] F. Abreu Araujo, C. Chopin, and S. de Wergifosse, arXiv:2206.13596 (2022), 10.48550/arXiv.2206.13596.