



HAL
open science

Temporal pattern recognition with delayed feedback spin- torque nano-oscillators

M. Riou, J. Torrejon, B. Garitain, F. Abreu Araujo, P. Bortolotti, V. Cros,
S. Tsunegi, K. Yakushiji, A. Fukushima, H. Kubota, et al.

► **To cite this version:**

M. Riou, J. Torrejon, B. Garitain, F. Abreu Araujo, P. Bortolotti, et al.. Temporal pattern recognition with delayed feedback spin- torque nano-oscillators. *Physical Review Applied*, 2019, 12 (2), 10.1103/PhysRevApplied.12.024049 . hal-02400740

HAL Id: hal-02400740

<https://hal.science/hal-02400740>

Submitted on 9 Dec 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Temporal pattern recognition with delayed feedback spin-torque nano-oscillators

M. Riou,¹ J. Torrejon,¹ B. Garitainé,¹ F. Abreu Araujo,¹ P. Bortolotti,¹ V. Cros,¹ S. Tsunegi,³ K. Yakushiji,³ A. Fukushima,³ H. Kubota,³ S. Yuasa,³ D. Querlioz,⁴ M. D. Stiles,² J. Grollier¹

¹*Unité Mixte de Physique CNRS, Thales, Université Paris-Sud, Université Paris-Saclay, 91767 Palaiseau, France*

²*Center for Nanoscale Science and Technology, National Institute of Standards and Technology, Gaithersburg, Maryland 20899-6202, USA*

³*National Institute of Advanced Industrial Science and Technology (AIST), Spintronic Research Center, Tsukuba, Ibaraki 305-8568, Japan*

⁴*Center for Nanoscience and Nanotechnology, CNRS, Université Paris-Sud, Université Paris-Saclay, 91405, Orsay, France*

The recent demonstration of neuromorphic computing with spin-torque nano-oscillators has opened a path to energy efficient data processing. The success of this demonstration hinged on the intrinsic short-term memory of the oscillators. In this study, we extend the memory of the spin-torque nano-oscillators through time-delayed feedback. We leverage this extrinsic memory to increase the efficiency of solving pattern recognition tasks that require memory to discriminate different inputs. The large tunability of these non-linear oscillators allows us to control and optimize the delayed feedback memory using different operating conditions of applied current and magnetic field.

I. INTRODUCTION

Recurrence can play a powerful role in information processing. It is thought to provide a source of memory in the brain and allows recurrent artificial neural networks to process sequences [1,2]. When a network is recurrent, input data can remain present in the network for an extended time creating dynamical memory. The state of the network not only depends on the current input value, but also on past values. For neuromorphic computing, recurrence has been adapted in hardware by combining a feedback loop with the physical component that plays the role of the neuron. This approach has been tested in different systems (electronic, optical, photonic) for tasks that require memory such as chaotic series prediction [3–11].

Here, we implement recurrence in lower power nano-devices called spin-torque nano-oscillators, [12,13] which are composed of two ferromagnetic layers separated by a non-magnetic material. In these devices, magnetization dynamics is driven by spin-polarized current through an effect known as spin-transfer torque. When current flows through the magnetic multilayer, the current becomes spin polarized and transfers angular momentum between the magnetic layers, resulting in a torque on the magnetization. For high enough current density, this torque can destabilize the magnetization of the free layer, which is then set in a sustained precessional state. These magnetic oscillations are converted into resistance oscillations due to magnetoresistance effects: the device then functions as an electrical

nano-oscillator. The frequency of these oscillators can be tuned from a few hundreds of megahertz to tens of gigahertz by varying the applied electrical current and magnetic field.

This control makes spin-torque nano-oscillators promising as nanoscale microwave sources. Moreover, these oscillators have the all essential features needed as hardware neurons. They can natively interact through their emitted magnetic fields [14–16] or electrical currents [17] to emulate synaptic coupling. They are ultra-compact: their lateral size can be reduced as small as ten nanometers and their power consumption below one microwatt [18]. Their structure is similar to the magnetic memory cells so that can be produced by the hundreds of millions on the back end of the line of Complementary Metal Oxide Semiconductor chips. The amplitude of the oscillation depends non-linearly on the input current, and they have an intrinsic memory related to the relaxation of magnetization, combining in a single nanodevice the two most important properties of neurons: non-linearity and memory. Alternative implementations would require assembling several lumped components and would occupy a much larger area on chip using conventional electronics or photonic technology [19].

These features formed the basis of the recent demonstration of neuromorphic reservoir computing with spin-torque oscillators [19–23]. This work showed that the nanoscale oscillator could classify spoken digits with an accuracy similar to macroscopic hardware neurons, and that it could improve classification performance by several tens of percent [20]. In this experimental demonstration, however, the intrinsic memory of these oscillators derived solely from their relaxation time, which is quite short: tens to hundreds of nanoseconds for vortex gyration [20,21] and even less for other types of magneto-dynamical modes such as uniform precession, bullet modes, or spin waves. This short time limits the number of cognitive tasks that can be performed [23]. In particular, sequence problems such as chaotic series prediction or natural language processing require longer memory to recognize temporal patterns. Delayed feedback spin torque oscillators have been studied both theoretically [24,25] and experimentally [26,27]. But the delay in these studies was much smaller (tens of nanoseconds) and applications were mainly related to the reduction of phase noise as well as the optimization of emitted power and linewidth.

In this study, we use a delayed-feedback spin-torque nano-oscillator to remember and recognize patterns. In Sec. II, we show that adding delayed-feedback to the oscillator can enhance the range of the memory from one relaxation time (≈ 200 ns) to tens of relaxation times. In Sec. III, we emulate a reservoir computer with a single, time-multiplexed oscillator [28,29], and demonstrate the performance of the extrinsic memory by recognizing temporal patterns composed of random sine and square waveforms. Finally, in Sec. IV, we vary the current and magnetic field applied to the nano-oscillator to find the optimal operating regime of the oscillator for pattern recognition.

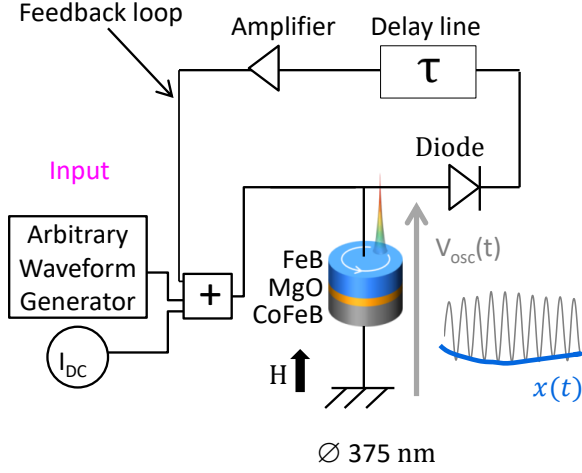


FIG 1: Experimental set up. The spin-torque oscillator is subjected to DC current and perpendicular magnetic field which set the operating point. It emits an oscillating voltage $V_{\text{osc}}(t)$. The time-varying input is generated by an arbitrary waveform generator. A diode allows measuring directly the amplitude of the oscillations $x(t)$, which is used for computation. The feedback loop consists of an electronic delay line ($\tau = 4.3 \mu\text{s}$) and an amplifier (the total amplification in the line is +20 dB). The signals are added with power splitters.

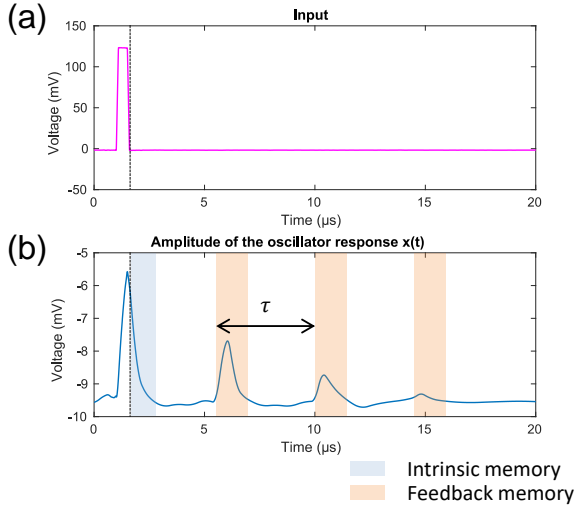


FIG 2: (a) Input spike (in magenta) sent by the arbitrary waveform generator to the spin-torque oscillator. (b) Blue curve: variation in the amplitude of the delayed feedback oscillator response $x(t)$. The shaded areas in blue and orange indicate the intrinsic and feedback memory respectively. The operating point is 600 mT and -6.5 mA.

II. EXPERIMENTAL IMPLEMENTATION OF DELAYED FEEDBACK SPIN TORQUE OSCILLATOR

The experimental set up is illustrated in Fig 1. It shows the spin-torque nano-oscillator, which receives as input the sum of the static DC current that it needs to oscillate, the rapidly varying input provided through the arbitrary waveform generator, and the feedback. The feedback is obtained by passing the envelope of the oscillator signal through a diode, applying a delay then amplifying it before reinjection.

The nano-oscillator is a cylindrical magnetic tunnel junction of diameter 375 nm. The pinned layer is a CoFeB-based synthetic antiferromagnet, the tunnel barrier is MgO and the free layer is a 3 nm thick FeB layer. For this aspect ratio of the free layer, the vortex magnetization is the stable configuration in the ground state. Such oscillators exhibit a high power (a few microwatts) over noise ratio (≈ 100), low phase noise [30], and show optimal microwave properties for neuromorphic computing [19–22]. The oscillator is connected to a DC current source and is subjected to an external magnetic field to induce gyrotropic motion of vortex core via spin transfer torque mechanism. These two tunable parameters set the regime where the oscillator operates.

The classification process used in the next section starts by encoding the different patterns in time-varying electrical input signals generated by an arbitrary waveform generator. This electrical input signal moves the oscillator out of its operating point, thus changing the amplitude and phase of the oscillator voltage response $V_{\text{osc}}(t)$ (in grey in Fig. 1). For simplicity, only the amplitude $x(t)$ (in blue in Fig. 1(a)) of the oscillations is used for computing. The amplitude is measured by placing a diode after the oscillator to capture the envelope of the oscillating signal. The delayed feedback loop is composed of an electronic delay line and an amplifier. The electronic delay line is passive and implements a delay of $\tau = 4.3 \mu\text{s}$. It is composed of a series of high performance inductors and shunt capacitors. For our experiments with inputs varying at about 10 MHz, the attenuation is less than 5 %. The amplifier placed after the delay line works from kHz to GHz regime and removes the DC part (offset) of the voltage response before reinjection. The total amplification of the delayed feedback loop is about 20 dB. The addition of the delayed feedback to the input is made with a power splitter. The emitted voltage of the oscillator (10 mV to 15 mV) is reinjected by the delay line. The reinjected voltage level depends on the signal amplitude emitted by the oscillator itself $x(t)$ and thus depends on the operating point (magnetic field and current). For the highest emitted voltage (at low magnetic field), the reinjected amplitude can reach approximately 50 % of the amplitude of the input signal (250 mV peak to peak). Finally, the oscillation amplitude is recorded with an oscilloscope.

The intrinsic memory of the oscillator and the feedback memory are illustrated in Fig. 2. A current pulse of 200 ns duration is sent to the oscillator (in magenta in Fig. 2(a)) and the voltage amplitude of oscillator $x(t)$ is recorded by an oscilloscope (in blue in Fig. 2(b)) during a much longer time (20 μs) to observe the reinjection effects and the memory induced by the delay line. The input spike creates a perturbation in the oscillator and modifies the orbit of gyrotropic motion of the vortex core. This changes the amplitude of the signal emitted by the oscillator $x(t)$. After the spike, the vortex core returns to the

orbit defined by the fixed magnetic field and DC current: the amplitude of the oscillations $x(t)$ returns progressively to its initial level. During this time, which corresponds to the relaxation, the oscillator still remembers the input because the oscillation amplitude has not returned to its initial level. The relaxation time of magnetization dynamics, roughly defined by the magnetic damping and the frequency of oscillator ($T_{relax} \sim \frac{1}{\alpha f}$), is around 200 ns, which corresponds to the range of the intrinsic memory of the oscillator (except when the oscillator operates in the immediate proximity of the critical current, where the relaxation time increases but the emission level is very low). This intrinsic memory, highlighted in blue in Fig. 2(b), has a duration of a few hundreds of nanoseconds.

The feedback is implemented by propagating the perturbation of the oscillation amplitude in the delay line for $\tau = 4.3 \mu\text{s}$ and reinjecting it into the oscillator. This injection induces new variations in the amplitude of the oscillator response $x(t)$. Indeed, echoes in the oscillator response are observed every τ after the end of the input signal in Fig. 2(b). These echoes are the manifestation of the external memory provided by the delayed feedback. Since after each τ the echo is more and more attenuated, the feedback memory is a fading memory. The fact that memory fades is important to process temporal sequence for which only the recent history is important [3]. For a magnetic field of 600 mT and a DC current of -6.5 mA, we observe variations of the oscillator output even after 13 μs (approximately 60 relaxation times) using the same type of measurements as shown in Fig. 2. The range of the feedback memory depends of the delay in the line. Here we observe a memory of 13 μs , which corresponds to three times the delay of the line but choosing a longer delay in the line would have resulted in a longer range of memory. Moreover, the feedback memory is sparse, that is, the information that the oscillator received as input in the past is only accessible every time τ (4.3 μs in our case) in the areas highlighted in orange in Fig. 2(b). In between, this information cannot be extracted from the measured trace. If the input is discrete, the time step of the input and the delay in the line are often chosen to be equal so the oscillator remembers the input at previous time steps.

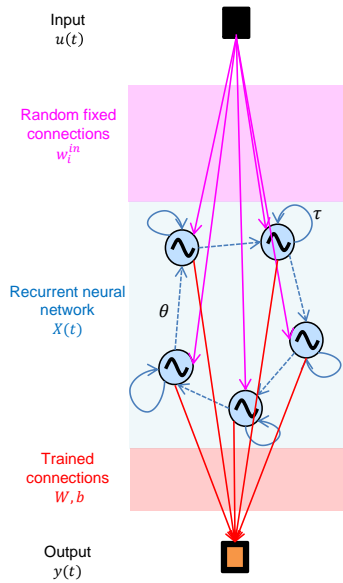


FIG 3: Principle of reservoir computing. The input $\mathbf{u}(t)$ is connected via fixed connections \mathbf{w}_i^{in} to a recurrent neural network, called a reservoir. The reservoir maps the input $\mathbf{u}(t)$ into a higher dimensional state $\mathbf{X}(t)$, that is, each neuron output is a coordinate of the projected input. The output $\mathbf{y}(t)$ is obtained by combining the neuron responses with trained connections \mathbf{W}, \mathbf{b} .

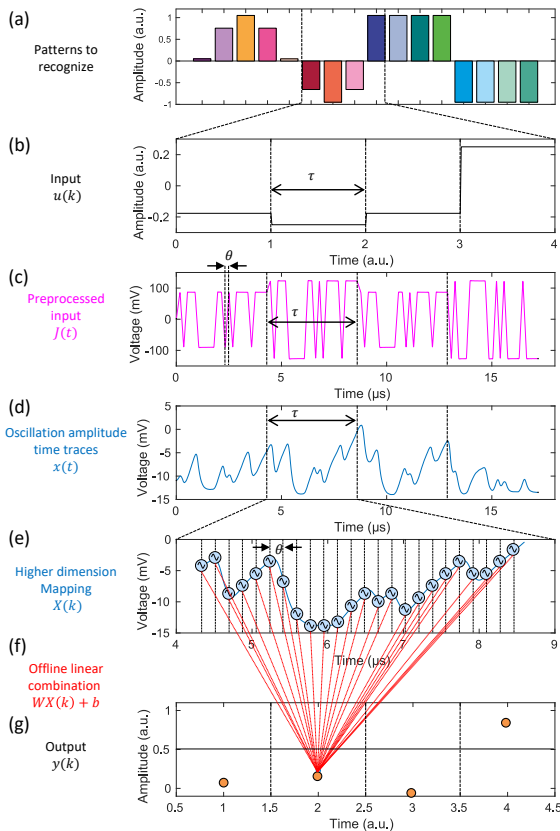


FIG 4: Implementation of reservoir computing with a single oscillator. (a) The patterns to recognize are discretized sine and square periods. (b) The discrete input $\mathbf{u}(k)$ is a time sequence of these sine and square periods, randomly arranged. (c) Preprocessed input $\mathbf{J}(t)$ obtained by multiplying $\mathbf{u}(k)$ with a fast varying sequence called a mask. This mask varies every θ and represents the coupling \mathbf{w}_i^{in} between the input and each virtual neuron of the reservoir. (d) $\mathbf{x}(t)$ is the oscillator response in oscillation

amplitude to the preprocessed input $\mathbf{J}(\mathbf{t})$. The oscillator plays the role of all virtual neurons one after the other, every θ . (e) The higher dimension mapping $\mathbf{X}(\mathbf{k})$ of the input $\mathbf{u}(\mathbf{k})$ can be retrieved in the different sequences of $\mathbf{x}(\mathbf{t})$ of duration τ . In our experiment the oscillator emulates $n = 24$ neurons, $\theta = 180 \text{ ns}$, $\tau = 4.3 \mu\text{s}$. (f) Then the output $\mathbf{y}(\mathbf{k})$ is reconstructed offline on a computer by combining linearly the responses of the virtual neurons $\mathbf{y}(\mathbf{k}) = \mathbf{W} \cdot \mathbf{X}(\mathbf{k}) + \mathbf{b}$. (g) If the output $\mathbf{y}(\mathbf{k})$ is less than 0.5, the input $\mathbf{u}(\mathbf{k})$ is classified as a sine, otherwise it is classified as a square.

III. NEUROMORPHIC COMPUTING APPROACH: RESERVOIR COMPUTING

To test the efficiency of the extrinsic memory induced by reinjection for computation, we perform pattern recognition of sine and square waveforms. To perform this task, we have adapted the delayed feedback spin torque oscillator to a recurrent neural network approach, called reservoir computing. Different hardware implementations have been identified to be suitable for reservoir computing: optical [3,4], photonic [5,11] and spintronic devices [20–22]. These hardware implementations exploit the intrinsic dynamics and variability of these devices, the latter particularly relevant in the case of spintronic devices with nanoscale magnetic volumes. Fig. 3 shows the architecture of a general reservoir computing scheme. The input $u(t)$ (in black in Fig. 3) is coupled with fixed connections w_i^{In} (in magenta in Fig. 3) to a recurrent neural network referred to as reservoir (in blue in Fig. 3) where the internal connections between neurons are chosen to be fixed and random. The reservoir maps the input $u(t)$ in a higher dimensional state referred as reservoir state $\mathbf{X}(t)$, where the output of each neuron represents a transformation of the projected input $u(t)$ (note that the dimension of $\mathbf{X}(t)$ is equal to the number of neurons n). If the reservoir is complex enough (which means that it is on the one hand of high enough dimension, with sufficient non-linearity and on the other it has sufficient memory), the transformed problem becomes linearly separable in the reservoir state. The output $y(t)$ is obtained by combining the neuron outputs ($x_i(t)$ values) with trained connections \mathbf{W}, \mathbf{b} .

The experimental implementation of reservoir computing with multiple interconnected devices is challenging [16], but a simplified approach has been successfully demonstrated [3] based on constructing a reservoir from a time-multiplexed single device. Here, the entire recurrent network is replaced by a single non-linear node which serves as each virtual neuron one after the other. In our case, the non-linear node is a spin-torque oscillator. The input is projected to higher dimension in time instead of in multiple devices (see figures Fig. 4(a-g)). The input stream u is an aggregation of discretized sine and square periods (patterns in Fig. 4(a)). The goal is to return an output 0 if the input value $u(k)$ belongs to a sine and an output of 1 if the input value $u(k)$ belongs to a square. The system processes a discretized version of the input stream $u(t)$ with each input value $u(k)$ (in black in Fig. 4(b)) being applied n times into the reservoir (where n is the number of virtual neurons) with a time step θ . The varying connections between the input and the virtual nodes of the reservoir are captured by multiplying

$u(k)$ by a sequence that varies every time step θ and repeats every $\tau = n\theta$, the time between different input values. We choose the connection between the input and the virtual neurons to be either +1 or -1. The resulting preprocessed input $J(t)$ (see Fig. 4(c)) is applied to the oscillator.

The oscillator then emits a transient amplitude response $x(t)$ (Fig. 4(d)). The responses of the n neurons $x_i(t)$ in a spatial reservoir are here replaced by the response of the single oscillator $x(t)$ sampled n times over intervals of length θ during a time τ (Fig. 4(e)). If the time θ is shorter than the relaxation time, $x(t)$ depends on $x(t - \theta)$. This situation is analogous to a neural network consisting of multiple connected devices. Here the coupling between the devices is provided by the influence of previous states of the oscillator on its current state. If a delayed feedback loop with a delay $\tau = n\theta$ is added to the oscillator, the oscillator also receives the signal $x(t - \tau)$ at time t . The virtual neurons x_i therefore receive their own past output, so they are recurrently coupled to themselves. The equivalent multiple-device neuron network architecture is represented in blue in Fig. 3. Dashed arrows are the connections emulated by the relaxation and solid line feedback arrows are the connections emulated by the delayed feedback loop. $\mathbf{X}(k)$ (dimension $(n,1)$) represents the mapping $u(k)$ into a higher dimensional space that is obtained from discrete values sampled every θ in each interval of length $\tau = n\theta$ in the time trace $x(t)$ (highlighted by circles in Fig. 3(d)). The output value is obtained by the linear combination $y(k) = \mathbf{W} \cdot \mathbf{X}(k) + b$ (Fig. 4(f)) (with \mathbf{W} of dimension $(1, n)$). If the output $y(k)$ is less than 0.5, the input $u(k)$ is classified as a sine. Otherwise, it is classified as a square (Fig. 4(g)).

In our experiment, we use $n = 24$ virtual neurons, which we found was the minimum number of neurons required to classify sines and squares [21]. The oscillator acts as each virtual neuron every $\theta = 180$ ns time interval. To emphasize the role of the memory effect brought by delayed feedback, we choose a larger time θ (comparable with the relaxation time) than that used in the previous spintronic implementation without delayed feedback [20,21]. The longer interval attenuates the intrinsic oscillator memory effect so as to focus on the memory provided by the feedback. The input $u(k)$ is processed with a time step $\tau = 4.3 \mu\text{s}$, which corresponds to the delay time in the feedback circuit. The preprocessed input $J(t)$ is created by the arbitrary waveform generator. The oscillator amplitude response $x(t)$ is recorded with an oscilloscope and the higher dimension mappings $\mathbf{X}(k)$ are extracted from $x(t)$. The optimal output weights \mathbf{W} , b , which are determined during the learning phase, and the output $y(k)$ are computed offline on a computer [3].

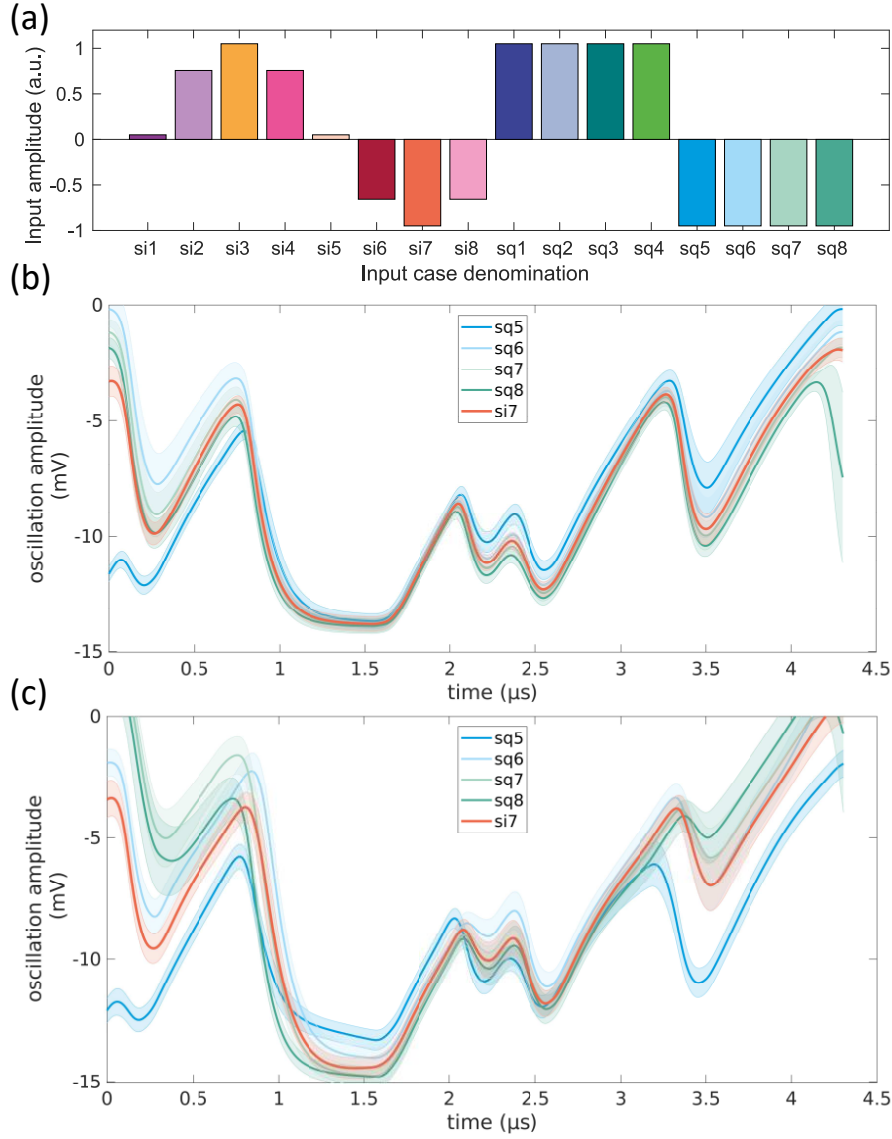


FIG 5 : (a) Pattern to recognize: the time-varying input $\mathbf{u}(\mathbf{k})$ used to evaluate the delayed feedback spin torque oscillator is composed of sine and square sequences. The 8 inputs from sine are designated as si1-8 and the 8 inputs of a square are designated as sq1-8. (b) Comparison of the average experimental time traces of $\mathbf{x}(\mathbf{t})$ without feedback for si7 (represented in orange) and sq5-8 (represented in cool colors, blues and greens). The shaded regions indicate twice the standard deviation of $\mathbf{x}(\mathbf{t})$ for sq5, sq8, and si7. The operating point is 300 mT and -6.5 mA. (c) Similar comparison with feedback.

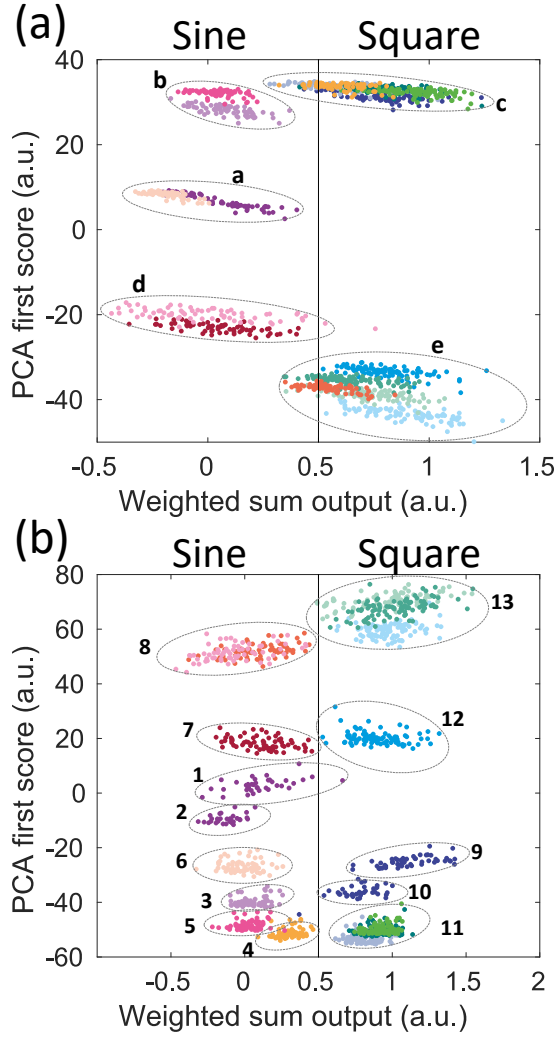


FIG 6: (a) 2D visualization of the higher dimension mapping $\mathbf{X}(\mathbf{k})$ obtained from experiment for all the different inputs $\mathbf{u}(\mathbf{k})$ without feedback. The operating point is 300 mT and -6.5 mA. The separation between sine and square cases is the vertical black solid $x = 0.5$. For a perfect recognition, points corresponding to sine input should be left to this line and points corresponding to square input should be to the right. Without feedback, the points for si7 (orange) and si3 (yellow) are mixed up with the corresponding square cases. The error rate is 10.8 % in the test phase (69 errors). Five clusters are observed corresponding to the five different input values. These clusters are highlighted by grey ellipses and denominated by the letters a to e. (b) 2D visualization of the higher dimension mapping $\mathbf{X}(\mathbf{k})$ obtained from experiment for all the different inputs $\mathbf{u}(\mathbf{k})$ with feedback. The operating point is 300 mT and -6.5 mA. Thirteen clusters are observed confirming that feedback enables the oscillator to separate inputs by remembering sequences of two consecutive inputs. These clusters highlighted by grey ellipses are denominated by the numbers 1 to 13. The points for si7 and si3 are different from the square cases and fall to the left of the vertical separation line. The final classification is almost perfect with only 0.16 % error rate (one error) during the test phase.

IV. PATTERN RECOGNITION USING DELAYED FEEDBACK SPIN TORQUE NANO-OSCILLATOR

We choose to classify sine and square waveforms, as done in previous studies [4,10,20,21], to evaluate the performance of our delayed-feedback reservoir computing based on a spin torque oscillator. Each period of the waveforms is discretized in 8 points giving 16 different inputs (Fig. 5(a)). We refer to the 8 points of the sine waveform as $si1-8$, which are represented in warm colors (yellow, red, and orange), and the 8 points of the square waveform as $sq1-8$, which are represented in cool colors (green and blue) in Fig. 5(a). Because of degeneracies, the inputs take 5 different values in a sine waveform and two different values in a square waveform. To return the same output value for 5 different input values of a sine (or 2 in the case of the square), the reservoir must be non-linear. In the sine period, the 3rd and the 7th point (referred as $si3$ and $si7$ see Fig. 5(a)) have the values +1 and -1, which are the same values as in the square waveform. In the absence of memory, it is impossible to classify the input value of +1 or -1 as belonging either to a sine or to square waveform. Therefore, this temporal pattern recognition task needs both non-linearity and memory in the neural network. The input $u(k)$ is composed of 1280 points (160 periods of sine or square randomly arranged). The first half of the points is used for training (to find optimum output weights \mathbf{W} and b) and the second half for testing. The operating point of the oscillator is 300 mT and -6.5 mA.

The lines in Figures 5(b) and 5(c) show the average $x(t)$ response over all the cases where input is the same for the $si7$, $sq5$, and $sq8$ cases with and without feedback. The colored regions centered around the lines indicate the dispersion of the dynamical response, their width corresponds to twice the standard deviation of responses to the same input. First, without feedback, we can see that time traces $x(t)$ for $si7$ overlap within the two-standard-deviation uncertainties with time traces $x(t)$ for $sq5$ to $sq8$ (Fig. 5(b)). When the feedback memory is included (Fig. 5(c)), the time trace $si7$ differs from the respective sq time traces. The feedback also breaks the symmetry between the sq points with the same input values. The traces for $sq5$ becomes significantly different from the traces for $sq6$ to $sq8$ (see Fig. 5(c)).

It is common in machine learning to use principal component analysis (PCA) to help with visualization of high-dimensional data by reducing its dimensionality. PCA rotates the high-dimensional data along a new set of axes, in a way that identifies the axes along which the data varies the most. The first axis is the one along which the projection of the data has the maximum variance among all possible axes. Each succeeding axis is chosen in turn to be orthogonal to all previous axes and have the maximum variance of the data of all the possible remaining directions. As the variance along each succeeding axis gets smaller and smaller, the data becomes more and more independent of the remaining directions. In this manner, the axes are sorted from most important to least important. Keeping only the projection of the data along the first few axes reduces the number of variables needed to describe the data, while keeping the most possible information from the data. In particular, projecting the data along the two first axes chosen by PCA allows for visualization of two-dimensional clusters in the data. We use this technique to analyze the ability of the oscillator to cluster data in the absence or presence of delayed

feedback. The 2D maps shown in Fig. 6(a) and (b) correspond to transformations of the oscillator response $\mathbf{X}(k)$.

As mentioned in the previous section, each τ long time trace corresponds to the mapping in 24 dimensions $\mathbf{X}(k)$ of an input value $u(k)$. To visualize the separation, the mappings $\mathbf{X}(k)$ are projected linearly in two dimensions in (Fig. 6(a) and (b)). To see the separation between the sine and the square, the first coordinate (x axis) of the 2D representation is given by the linear combination $\mathbf{W} \cdot \mathbf{X} + b$. Here b is the offset bias weight, trained offline together with the other weights [3]. It can be seen geometrically as a projection of the data along the weight vector \mathbf{W} . In this 2D projection, the separation between sine and square is defined by the vertical line at $x = 0.5$ in Fig 6(a-b). Ideally the sine cases (si1-8) should all fall on the left of this line and square cases (sq1-8) on the right, respectively.

The second coordinate (y axis) of the projected data points is the first component from a principal component analysis (PCA) of the data reduced in the space orthogonal to \mathbf{W} . The PCA separates the clusters (collections of points that are neighbors in the 24-dimension reservoir state) in the data. The PCA is performed in the space orthogonal to \mathbf{W} to project the data along two orthogonal vectors. Without feedback, 5 clusters denominated “a” to “e” are observed in the 2D map. They correspond respectively to the 5 different values taken by the input: 0 (si1 and si5), 0.71 (si2 and si4), 1 (si3 and sq1-sq4), -0.71 (si6 and si8) and -1 (si7 and sq5-sq8). The ambiguous cases corresponding to si3 and si7 are completely mixed up with sq1-4 and sq5-8, respectively. Since the time step θ has been chosen close to the relaxation time of the oscillator, the intrinsic memory is too small to remember a two-point sequence. With no feedback, the error rate for the identification of sine and square is 10.8 % for this value of θ . All the 69 errors during test phase are due to bad classification between si3 and sq1-4 or si7 and sq5-8. These results are in excellent agreement with the prediction and the qualitative analysis of the time traces.

Without memory, we expect five clusters, because the input takes only five different values. With memory, new clusters appear and the number of clusters depends of the range of the memory. If the reservoir remembers one time step, the clusters in the higher dimension mapping should correspond to the different sequences of two consecutive input points at time k and $k-1$. We expect nine different clusters for the sine and five different clusters for the square. Similarly, if the reservoir can remember two time steps in the past, eighteen clusters are expected.

With delayed feedback to provide memory, the 2D map shows 13 clusters instead of the 5 found without delayed feedback. This symmetry breaking can be explained by a memory of one time step in the past. We observed 13 (and not 14 clusters as expected) on the 2D projection because points for si7 and si8 overlap (cluster 8 in Fig. 6(b)). The first component of the PCA may have not conserved all the clusters present in the initial 24-dimension reservoir state. However, all the other 13 expected clusters are clearly separated in the 2D projection, confirming that the delayed feedback spin torque nano-oscillator remembers sequence of two consecutive inputs. With delayed feedback, si3 and si7 become different from the square traces and a linear separation can be found with a very small error rate (0.16

% errors on the test set which corresponds to only one error). Adding delayed feedback to the oscillator allows it to distinguish si3 and si7 from square points because the feedback adds a memory of the previous input point. For the optimal operating point (300 mT and -6.5 mA), the feedback is very efficient and suppresses 98.6 % of the errors (Fig 6(a-b), Fig 9), demonstrating the efficiency of the feedback memory for computation.

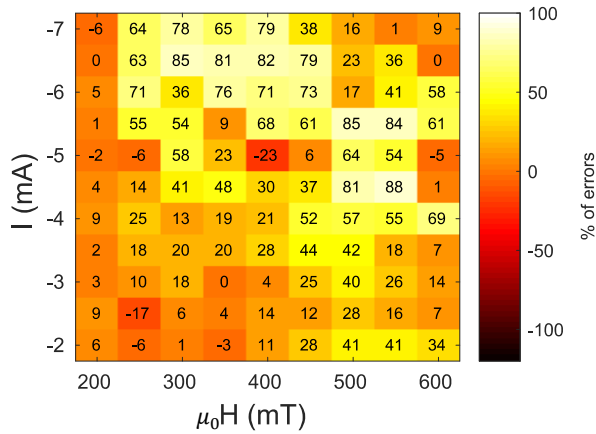


FIG 7: Reduction of errors on the ambiguous cases si3 and si7 due to feedback depending of the operating point, Eq. (1): this reduction of the error is evaluated on the training set. The number of suppressed errors is renormalized by the total number of errors on the training set without feedback. Brighter colors indicate high reduction of the error. The magnetic field is swept from 200 mT to 600 mT and the DC current is swept from -2 mA to -7 mA. In 90 % of operating points, feedback helps distinguishing si3 and si7 during the training phase. Error reduction is high for high DC current (larger than -4 mA) and magnetic field between 250 mT and 400 mT.

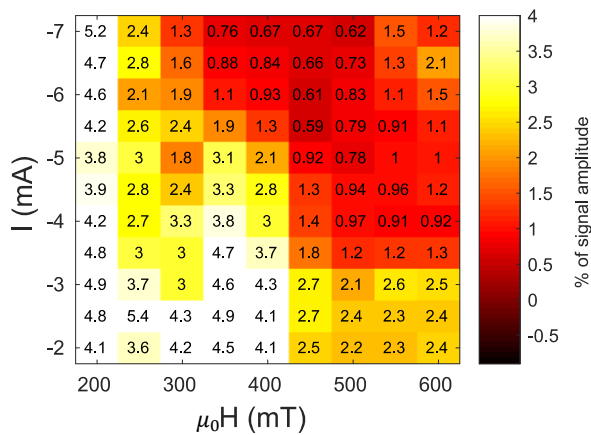


FIG 8: Normalized noise level as a function of the operating point: the magnetic field is swept from 200 mT to 600 mT and the DC current is swept from -2 mA to -7 mA. At low field (200 mT to 400 mT) and

low current (-2 mA to -5 mA) the noise level in the oscillation amplitude response is high. Areas of high noise (brighter color) correspond to areas where the feedback is not efficient at suppressing errors.

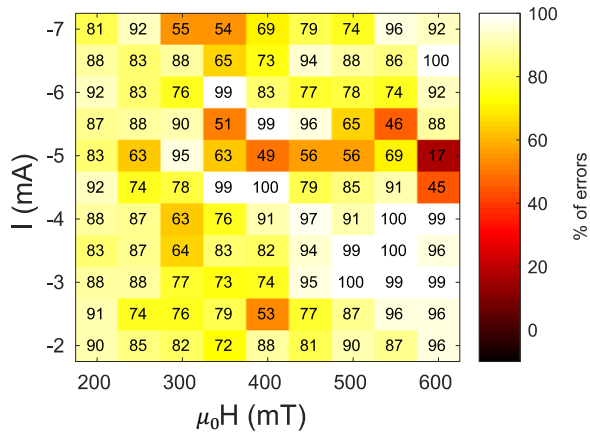


FIG 9: Proportion of si_3 and si_7 misclassified on training set without feedback, Eq. (2): the magnetic field is swept from 200 mT to 600 mT and the DC current is swept from -2 mA to -7 mA. For 600 mT and -5 mA, si_3 and si_7 are well classified even without feedback (only 17 % error). Other operating points exhibit good performance (orange and red areas).

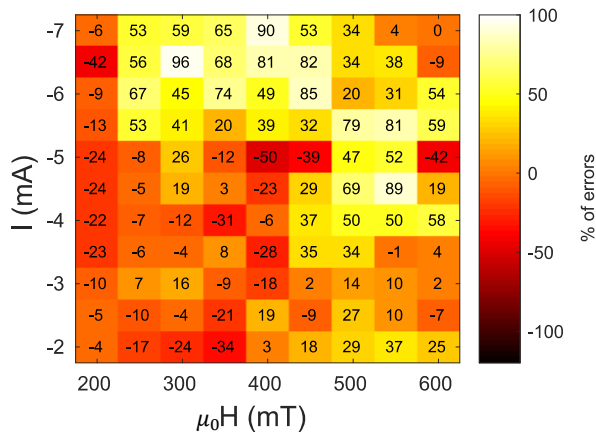


FIG 10: Global error reduction on the training set due to the feedback as a function of the operating point, Eq. (3): the number of suppressed errors is renormalized by the total number of errors on the training set without feedback. The magnetic field is swept from 200 mT to 600 mT and the DC current is swept from -2 mA to -7 mA. Feedback reduces the error rate on the training set only in 66 % of the cases. The feedback thus generates new errors in at least 24 % of the bias points. Brighter colors indicate high reduction of the error. Error reduction is high for high DC current (larger than -4 mA) and fields between 250 mT and 400 mT.

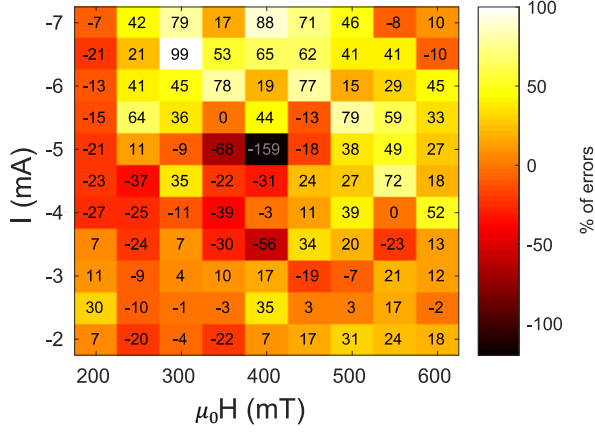


FIG 11: Global error rate reduction due to feedback on the test set as a function of the operating point, Eq. (4): the number of suppressed errors is renormalized by the total number of errors on the test set without feedback. The magnetic field is swept from 200 mT to 600 mT and the DC current is swept from -2 mA to -7 mA. Global error rate reduction during test phase is in good agreement with the improvement during training phase. Feedback improves the result in 60 % of the cases during the test phase. Brighter colors indicate high reduction of the error. Error reduction is high for high DC current (larger than -4 mA) and fields between 250 mT and 400 mT.

Both the magnetic field and the DC current change the non-linear dependence of the voltage oscillation amplitude with the input current (see figure 3 in Ref. 1) with the DC current acting like an offset for the input current. The operating point fixed by these two parameters changes the effectiveness of the feedback memory. As seen previously the reduction of the errors classifying the si3 and si7 cases shows that the memory of the oscillator was enhanced. This improvement is quantified by the proportion

$$Pr_1 = \frac{(E_{si3}^{no\ fb} + E_{si7}^{no\ fb}) - (E_{si3}^{fb} + E_{si7}^{fb})}{2(N_{si3}^{train} + N_{si7}^{train})}, \quad (1)$$

, where $E_{si3}^{no\ fb}$ and $E_{si7}^{no\ fb}$ are respectively the number of misclassified si3 and si7 examples during training in the case without feedback, E_{si3}^{fb} and E_{si7}^{fb} are respectively the number of misclassified si3 and si7 examples during training phase in the case with feedback and N_{si3}^{train} and N_{si7}^{train} are the total number of si3 and si7 examples during training phase. Fig. 7 is a 2D map of this quantity as a function of the operating point, where bright colors indicate high improvements and dark colors low improvement or, in the case where Pr_1 is negative, that more cases si3 and si7 are misclassified with feedback. In 90% of the operating point conditions the feedback improves the recognition of the si3 and si7 cases (see Fig. 7), showing the memory it brings breaks the degeneracies between inputs from square and sine periods. The low improvements are mainly at low magnetic fields: 200 mT and between 250 mT and 400 mT and between -2.0 mA to -5.0 mA. This area correlates well with regions where the amplitude noise of the oscillator is high compared to the signal amplitude (see Fig 8, bright region).

Some other cases where Pr_1 is negative or positive but low are observed on isolated operating points such as 600 mT and -4.5 mA, 600 mT and -5.0 mA, 400 mT and -5.0 mA, 450 mT and -5.0 mA. Fig. 9 shows

$$Pr_2 = \frac{(E_{si3}^{no\ fb} + E_{si7}^{no\ fb})}{(N_{si3}^{train} + N_{si7}^{train})}, \quad (2)$$

which is the proportion of si3 and si7 inputs misclassified during the training phase in the case without feedback. For these operating points, where higher relaxation time are measured, using the oscillator without feedback already leads to partly good classification of the si3 and si7 cases (see Fig. 9). Indeed with a 180 ns θ time step, the intrinsic memory is not completely negligible close to these operating points, resulting in good classification even without feedback. In particular, at 600 mT and -5.0 mA, 87% of the si3 and si7 cases are well classified without feedback. For 400 mT and -5.0 mA, the feedback generates new mistakes on the si3 and si7 cases. In this particular case the feedback works against the intrinsic memory of the oscillator.

The total error improvement for the training phase is computed as

$$\Delta\epsilon_{train} = \frac{\epsilon_{no\ fb} - \epsilon_{fb}}{\epsilon_{no\ fb}}, \quad (3)$$

where $\epsilon_{no\ fb}$ and ϵ_{fb} are the total error rates respectively without feedback and with feedback, such that $\Delta\epsilon_{train} > 0$ corresponds to an improvement of classification due to the feedback. For the training phase, the feedback improves the result in 66% of the operating point conditions (Fig. 10). Areas where the feedback is detrimental ($\Delta\epsilon_{train} < 0$) correspond to areas where Pr_1 was low or negative. If, in the large majority of cases, feedback improves the classification of si3 and si7 cases, it may also generate new errors that in some operating point conditions, overcome the benefit of external memory. These new errors could be due to the increase of dispersion in the reservoir state. Indeed, if feedback suppresses unwanted degeneracies between inputs from square and inputs from sine, it also generates new clusters as seen at the beginning of this section. This effect is reinforced in areas where dispersion is high due to stochastic behavior at threshold current and field for auto-oscillation (for 200 mT and between 250 mT and 400 mT and -2 mA to -5 mA).

Computing the error gain on the testing data

$$\Delta\epsilon_{test} = \frac{\epsilon_{no\ fb} - \epsilon_{fb}}{\epsilon_{no\ fb}}, \quad (4)$$

where $\epsilon_{no\ fb}$ and ϵ_{fb} are the error rates for the testing data respectively in the case without feedback and with feedback, we found that the dependency with the operating point is similar to $\Delta\epsilon_{train}$ with the feedback improving in 60% of the cases (Fig. 11). Some values of $\Delta\epsilon_{test}$ are worse than $\Delta\epsilon_{train}$ because

of bad generalization such as observed for instance at 400 mT and -5.0 mA. The feedback brings memory in a vast majority of the operating point conditions and even though it may bring new types of errors due to an increased dispersion of the data in the reservoir state, it still improves the error rate in a large range of operating point conditions. Best improvements are obtained in areas of low noise when compared to the amplitude and of low intrinsic memory.

CONCLUSION

Spin-torque nano-oscillators demonstrate clear memory effects when feedback is added. The feedback oscillator can remember the effect of the input signal tens of times longer than it can just with the intrinsic memory that is defined by its relaxation time. This feedback can be adapted to the time steps of the input by tuning the delay line. We evaluate the efficiency of feedback as a memory using a single spin torque oscillator with feedback as a reservoir computing recurrent network. The single oscillator with time multiplexing projects the initial problem in a higher dimensional state, turning it into a linearly separable problem. We tested this computation scheme on a temporal pattern recognition task, sine and square waveform classification, which requires memory and non-linearity. Choosing the optimal operating point, the error rate drops from 10.8 % without feedback to 0.16% with feedback. Analyzing the different clusters that appear in the data with a 2D projection for the optimal operating point shows that the delayed feedback remembers one time step in the past allowing a nearly perfect separation of the data. We identify the optimal operating point to be when the amplitude of the reinjected level is high and the noise of the oscillator is low (300 mT and -6.5 mA). Here we demonstrated the value of adding delayed feedback to a single node reservoir computing scheme. To have significant economic impact, neuromorphic computing will need to be massively parallel to be fast and energy efficient. Spin-torque nano-oscillators are promising for this purpose because they can be fabricated by the hundreds of millions on chip and they can be natively coupled through their electromagnetic emissions [15,31,17] which then play the roles of synapses [32]. Delayed feedback can be used to extend the memory for a collection of many such oscillators. In this work, we overcome the intrinsic memory limitation of spin-torque oscillators, opening the path to solve complex sequence recognition with networks of spin-torque oscillators.

Acknowledgements

Research supported as part of the Q-MEEN-C, an Energy Frontier Research Center funded by the U.S. Department of Energy (DOE), Office of Science, Basic Energy Sciences (BES), under Award DE-SC0019273 (design of experiments, numerical simulations) and by the European Research Council ERC under Grant bioSPINspired 682955 (experiments). F.A.A. is a Research Fellow of the F.R.S.-FNRS.

Appendix: Characterization of the oscillator microwave properties

The resistance of the magnetic tunnel junction as a function of in plane field is shown in Fig. 12. The observed hysteresis cycle is typical of vortex nucleation, motion and expulsion [33]. The magnetoresistance is close to 100%.

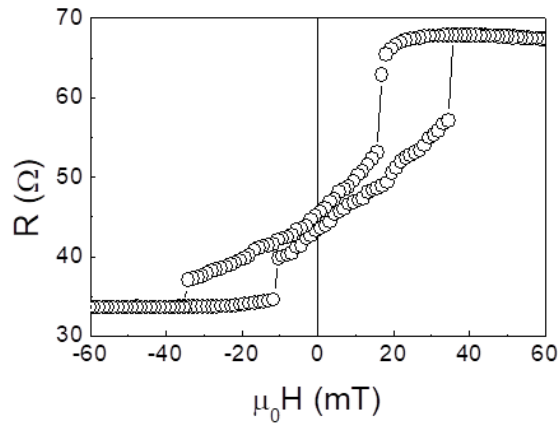


FIG 12: Resistance as a function of the in-plane applied magnetic field.

The evolution of oscillator frequency, linewidth and emitted microwave power as a function of dc current and applied perpendicular magnetic field is shown in Fig. 13.

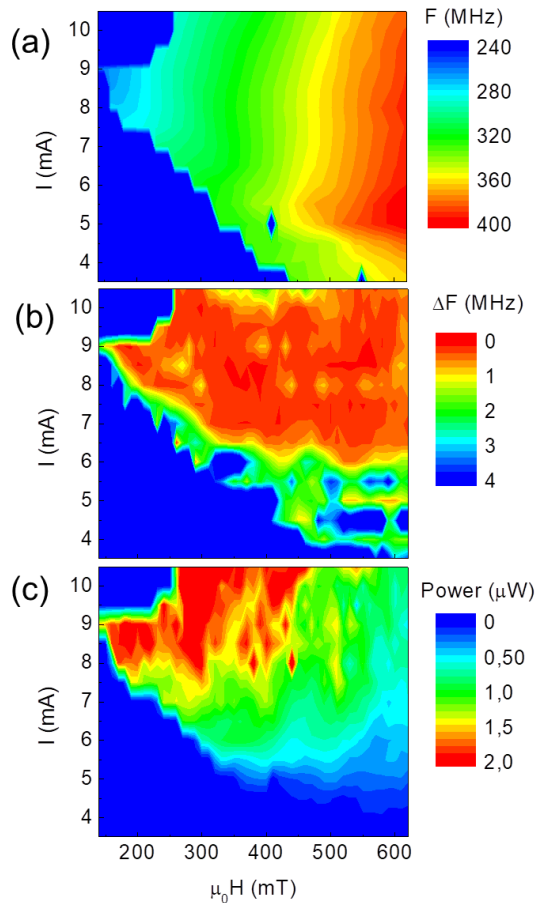


FIG 13: Color maps of the (a) frequency, (b) linewidth and (c) emitted microwave power of the oscillator as a function of dc current and applied perpendicular magnetic field.

- [1] S. Hochreiter and J. Schmidhuber, *Neural Comput.* **9**, 1735 (1997).
- [2] M. Schuster and K. K. Paliwal, *IEEE Trans. Signal Process.* **45**, 2673 (1997).
- [3] L. Appeltant, M. C. Soriano, G. V. der Sande, J. Danckaert, S. Massar, J. Dambre, B. Schrauwen, C. R. Mirasso, and I. Fischer, *Nat. Commun.* **2**, 468 (2011).
- [4] Y. Paquot, F. Duport, A. Smerieri, J. Dambre, B. Schrauwen, M. Haelterman, and S. Massar, *Sci. Rep.* **2**, 287 (2012).
- [5] D. Brunner, M. C. Soriano, C. R. Mirasso, and I. Fischer, *Nat. Commun.* **4**, 1364 (2013).
- [6] A. Dejonckheere, F. Duport, A. Smerieri, L. Fang, J.-L. Oudar, M. Haelterman, and S. Massar, *Opt. Express* **22**, 10868 (2014).
- [7] Q. Vinckier, F. Duport, A. Smerieri, K. Vandoorne, P. Bienstman, M. Haelterman, and S. Massar, *Optica* **2**, 438 (2015).
- [8] F. Duport, B. Schneider, A. Smerieri, M. Haelterman, and S. Massar, *Opt. Express* **20**, 22783 (2012).
- [9] F. Duport, A. Smerieri, A. Akrouf, M. Haelterman, and S. Massar, *Sci. Rep.* **6**, 22381 (2016).
- [10] K. Vandoorne, W. Dierckx, B. Schrauwen, D. Verstraeten, R. Baets, P. Bienstman, and J. V. Campenhout, *Opt. Express* **16**, 11182 (2008).
- [11] K. Vandoorne, P. Mechet, T. V. Vaerenbergh, M. Fiers, G. Morthier, D. Verstraeten, B. Schrauwen, J. Dambre, and P. Bienstman, *Nat. Commun.* **5**, 3541 (2014).

- [12] S. I. Kiselev, J. C. Sankey, I. N. Krivorotov, N. C. Emley, R. J. Schoelkopf, R. A. Buhrman, and D. C. Ralph, *Nature* **425**, 380 (2003).
- [13] W. H. Rippard, M. R. Pufall, S. Kaka, S. E. Russek, and T. J. Silva, *Phys. Rev. Lett.* **92**, (2004).
- [14] A. D. Belanovsky, N. Locatelli, P. N. Skirdkov, F. Abreu Araujo, K. A. Zvezdin, J. Grollier, V. Cros, and A. K. Zvezdin, *Appl. Phys. Lett.* **103**, 122405 (2013).
- [15] A. A. Awad, P. Dürrenfeld, A. Houshang, M. Dvornik, E. Iacocca, R. K. Dumas, and J. Åkerman, *Nat. Phys.* **13**, 292 (2017).
- [16] T. Kanao, H. Suto, K. Mizushima, H. Goto, T. Tanamoto, and T. Nagasawa, *ArXiv190507937 Cond-Mat Physicsphysics* (2019).
- [17] R. Lebrun, S. Tsunegi, P. Bortolotti, H. Kubota, A. S. Jenkins, M. Romera, K. Yakushiji, A. Fukushima, J. Grollier, S. Yuasa, and V. Cros, *Nat. Commun.* **8**, 15825 (2017).
- [18] H. Sato, E. C. I. Enobio, M. Yamanouchi, S. Ikeda, S. Fukami, S. Kanai, F. Matsukura, and H. Ohno, *Appl. Phys. Lett.* **105**, 062403 (2014).
- [19] M. Romera, P. Talatchian, S. Tsunegi, F. A. Araujo, V. Cros, P. Bortolotti, J. Trastoy, K. Yakushiji, A. Fukushima, H. Kubota, S. Yuasa, M. Ernoult, D. Vodenicarevic, T. Hirtzlin, N. Locatelli, D. Querlioz, and J. Grollier, *Nature* **563**, 230 (2018).
- [20] 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, *Nature* **547**, 428 (2017).
- [21] M. Riou, F. A. Araujo, J. Torrejon, S. Tsunegi, G. Khalsa, D. Querlioz, P. Bortolotti, V. Cros, K. Yakushiji, A. Fukushima, H. Kubota, S. Yuasa, M. D. Stiles, and J. Grollier, in *2017 IEEE Int. Electron Devices Meet. IEDM* (2017), pp. 36.3.1-36.3.4.
- [22] D. Marković, N. Leroux, M. Riou, F. Abreu Araujo, J. Torrejon, D. Querlioz, A. Fukushima, S. Yuasa, J. Trastoy, P. Bortolotti, and J. Grollier, *Appl. Phys. Lett.* **114**, 012409 (2019).
- [23] S. Tsunegi, T. Taniguchi, K. Nakajima, S. Miwa, K. Yakushiji, A. Fukushima, S. Yuasa, and H. Kubota, *Appl. Phys. Lett.* **114**, 164101 (2019).
- [24] G. Khalsa, M. D. Stiles, and J. Grollier, *Appl. Phys. Lett.* **106**, 242402 (2015).
- [25] V. S. Tiberkevich, R. S. Khymyn, H. X. Tang, and A. N. Slavin, *Sci. Rep.* **4**, 3873 (2014).
- [26] S. Tsunegi, E. Grimaldi, R. Lebrun, H. Kubota, A. S. Jenkins, K. Yakushiji, A. Fukushima, P. Bortolotti, J. Grollier, S. Yuasa, and V. Cros, *Sci. Rep.* **6**, 26849 (2016).
- [27] D. Kumar, K. Konishi, N. Kumar, S. Miwa, A. Fukushima, K. Yakushiji, S. Yuasa, H. Kubota, C. V. Tomy, A. Prabhakar, Y. Suzuki, and A. Tulapurkar, *Sci. Rep.* **6**, 30747 (2016).
- [28] W. Maass, T. Natschläger, and H. Markram, *Neural Comput.* **14**, 2531 (2002).
- [29] H. Jaeger and H. Haas, *Science* **304**, 78 (2004).
- [30] S. Tsunegi, K. Yakushiji, A. Fukushima, S. Yuasa, and H. Kubota, *Appl. Phys. Lett.* **109**, 252402 (2016).
- [31] N. Locatelli, A. Hamadeh, F. Abreu Araujo, A. D. Belanovsky, P. N. Skirdkov, R. Lebrun, V. V. Naletov, K. A. Zvezdin, M. Muñoz, J. Grollier, O. Klein, V. Cros, and G. de Loubens, *Sci. Rep.* **5**, 17039 (2015).
- [32] J. Grollier, D. Querlioz, and M. D. Stiles, *Proc. IEEE* **104**, 2024 (2016).
- [33] R. P. Cowburn, D. K. Koltsov, A. O. Adeyeye, M. E. Welland, and D. M. Tricker, *Phys. Rev. Lett.* **83**, 1042 (1999).