

Reservoir computing leveraging the transient non-linear dynamics of spin-torque nano-oscillators

Mathieu Riou, Jacob Torrejon, Flavio Abreu Araujo, Sumito Tsunegi, Guru Khalsa, Damien Querlioz, Paolo Bortolotti, Nathan Leroux, Danijela Marković, Vincent Cros, Kay Yakushiji, Akio Fukushima, Hitoshi Kubota, Shinji Yuasa, Mark D. Stiles and Julie Grollier

Abstract Present artificial intelligence algorithms require extensive computations to emulate the behavior of large neural networks, operating current computers near their limits, which leads to high energy costs. A possible solution to this problem is the development of new computing architectures, with nanoscale hardware components that use their physical properties to emulate the behavior of neurons. In spite of multiple theoretical proposals, there have been only a limited number of experimental demonstrations of brain-inspired computing with nanoscale neurons. Here we describe such demonstrations using nanoscale spin-torque oscillators, which exhibit key features of neurons, in a reservoir computing approach. This approach offers an interesting platform to test these components, because a single component can emulate a whole neural network. Using this method, we classify sine and square

Mathieu Riou, Jacob Torrejon, Flavio Abreu Araujo, Paolo Bortolotti, Nathan Leroux, Danijela Marković, Vincent Cros and Julie Grollier

Unité mixte de physique CNRS/Thales, Université Paris-Sud, Université Paris-Saclay, 91767 Palaiseau, France, e-mail: mathieu.riou@yahoo.fr, jtorrejon81@gmail.com, flavio.abreuaraujo@uclouvain.be, paolo.bortolotti@thalesgroup.com, nathan.leroux@cnrs-thales.fr, danijela.markovic@cnrs-thales.fr, vincent.cros@cnrs-thales.fr and julie.grollier@cnrs-thales.fr

Sumito Tsunegi, Kay Yakushiji, Akio Fukushima, Hitoshi Kubota and Shinji Yuasa
National Institute of Advanced Industrial Science and Technology (AIST), Spintronics Research Center, Tsukuba, Ibaraki 305-8568, Japan, e-mail: tsunegi.sb@aist.go.jp, k-yakushiji@aist.go.jp, akio.fukushima@aist.go.jp, hit-kubota@aist.go.jp, yuasa-s@aist.go.jp

Guru Khalsa
Department of Material Science and Engineering, Cornell University, Ithaca, NY 14853, USA, e-mail: guru.khalsa@cornell.edu

Mark D. Stiles
Center for Nanoscale Science and Technology, National Institute of Standards and Technology, Gaithersburg, Maryland 20899-6202, USA, e-mail: mark.stiles@nist.gov

Damien Querlioz
Centre de Nanosciences et de Nanotechnologies, CNRS, Université Paris-Sud, Université Paris-Saclay, 91405 Orsay, France, e-mail: damien.querlioz@u-psud.fr

waveforms perfectly and achieve spoken digit recognition with state of the art results. We illustrate optimization of the oscillator's operating regime with sine/square classification.

1 Context

Artificial intelligence has attracted interest because it offers the possibility of machines outperforming humans at cognitive tasks such as image or speech recognition. In a data driven society, artificial intelligence will become more and more essential as many industries begin to automate the analysis of ambiguous situations. The algorithms at the base of this progress are artificial neural networks, which take inspiration from the plasticity and non-linearity of biological neural networks. They originate in the nineteen fifties, when the first algorithm allowing a machine to learn abstract representations was developed [1]. Artificial neural networks are now becoming popular, because the computation capabilities of microprocessors have improved enough to run these complex algorithms [2] to solve useful tasks. Tasks such as image recognition require computing the response of millions of formal neurons and tuning tens of millions of parameters. Even though these algorithms are still far from the complexity of the human brain, which has one hundred billion neurons, running them on classical computer architectures already costs significantly more energy than appropriate for the range of possible applications.

Modern computers are based on the Von Neuman architecture where the processing unit is separated from the memory and data has to move sequentially back and forth between these two units through a shared bus. While this architecture has been responsible for the dramatic growth in computational capabilities, it is not well adapted for energy efficient cognitive computing. While artificial neural networks can perform comparably to humans for some particular tasks, they do so consuming three to four orders of magnitude more energy than the human brain. The energy efficiency of the brain comes in part from its entanglement of processing and memory, with biological neurons that process the information densely interconnected by synapses that hold the memory. This observation motivates building brain-inspired chips with analog components whose physics mimics the behavior of neurons. Biological neurons encode their information in the spikes they emit. One branch of neuroscience models neurons as non-linear auto-oscillators and the brain as a large assembly of interconnected oscillators that compute through their complex dynamics. A brain-inspired chip based on these principles requires integrating millions of non-linear oscillators in an area as small as one square centimeter. Simple arithmetic shows that this scale requires nanoscale non-linear oscillators.

While there are proposals for neurons based on memristors and Josephson junctions, these have not been experimentally demonstrated yet. Here we describe an experimental realization of nanoscale neurons based on spin-torque oscillators [3, 4] in the approach of reservoir computing. Reservoir computing offers a convenient platform to test the potential of spin-torque oscillators since a single oscillator ex-

cited by time-multiplexed inputs can generate a transient signal equivalent to the response of a whole neural network. This approach greatly simplifies the experimental set-up, and has been successfully used, notably in optics, to perform complex cognitive tasks such as speech recognition.

In this chapter, we summarize our work on reservoir computing using spin-torque oscillators [5, 6]. Here, we emphasize the measurement system, the results of pattern recognition tasks, and the optimization of the operating regime. We only touch lightly on the single-node reservoir computing approach and the physics of the spin-torque oscillators. Interested readers can find more details in our original papers [5, 6] or the other chapters in this volume. Sec. 2 presents our experimental implementation of single node reservoir computing based on a single vortex spin-torque nano-oscillator. Then, Sec. 3 presents the classification results obtained using this approach [5] for two tasks, sine versus square classification (Sec. 3.1.1) and spoken digit recognition (Sec. 3.2.2). Sec. 4 describes the optimization of the different parameters that can be tuned to improve the recognition rate on the sine and square classification task.

2 Hardware implementation

2.1 Measurement set-up

The samples used for this demonstration are magnetic tunnel junctions (MTJs) with a vortex magnetization in the free layer. The TMR of the junction is about 135% for a resistance of 42Ω . For the dimensions used here (thickness $L=6$ nm and diameter $\Phi=375$ nm), the FeB layer has a remanent vortex magnetization. Under DC current injection, the core of the vortex steadily gyrates around the center of the dot with a frequency in the range 250 MHz to 400 MHz. Vortex dynamics driven by spin torque are well understood [7], well controlled and have been shown to be particularly stable [8], and in this case have an excellent signal to noise ratio (for more details see the chapter by Taniguchi et al and reference [9]).

The measurement set-up is shown in Figure 1a. The experimental pre-processed input signal V_{in} is generated by a high-frequency arbitrary-waveform generator and injected as a current through the magnetic nano-oscillator. The preprocessed input varies with a time step θ . The sampling rate of the source is set to 200 MHz, which corresponds to 20 points per interval of discretized time (θ) in the reservoir computing scheme for the spoken-digit recognition task and 500 MHz (50 points per interval) for the classification of sines and squares. The peak-to-peak voltage variation in the input signal is 500 mV, which corresponds to peak-to-peak current variations of 6 mA (part of the incoming signal is reflected due to impedance mismatch between the sample and the circuit). The bias conditions of the oscillator are set by a DC current source and an electromagnet that applies a field perpendicular to the plane of the magnetic layers. These bias conditions determine the operating point of the oscillator. The oscillating voltage emitted by the nano-oscillator is rectified by a planar tunnel microwave diode, with a bandwidth of 0.1 GHz to 12.4 GHz and a

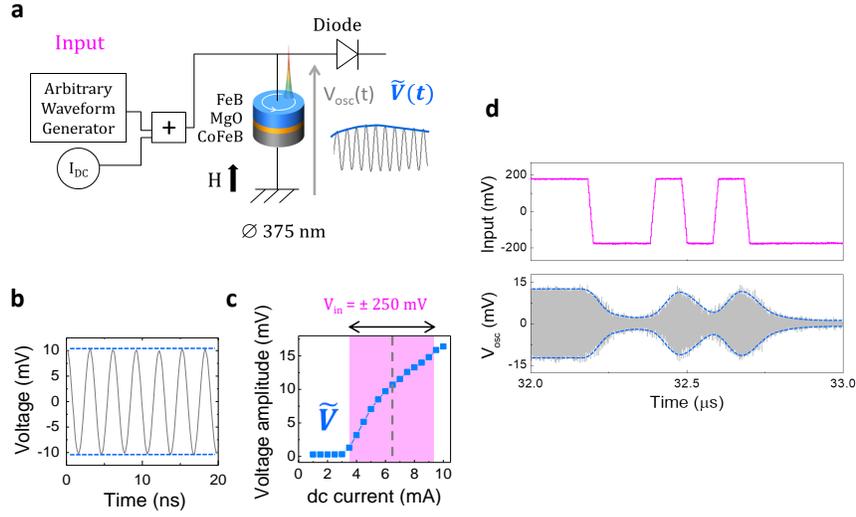


Fig. 1 a. The measurement set-up: A magnetic field delivered by an electromagnet (not shown) and DC source fix the operating point of the oscillator. The signal to analyze is sent by an arbitrary waveform generator. The magnetic tunnel junction emits an oscillating voltage. A diode allows extraction of the oscillation amplitude. b. Oscillating response in grey and oscillation amplitude figured in blue. c. Non-linear variation of the oscillation amplitude \tilde{V} as a function of the input current I_{DC} at $\mu_0 H = 430$ mT. The purple shaded area highlights the typical excursion in the voltage amplitude that results when an input signal of $V_{in} = \pm 250$ mV is injected. Here $I_{DC} = 6.5$ mA (vertical dotted line). d. Upper graph: input signal sent by the arbitrary waveform generator (magenta). Lower graph: transient response of the oscillator. The emitted oscillating voltage is plotted in grey. The amplitude of the oscillation is figured by dashed blue lines. Adapted from Torrejon et al [5].

response time of 5 ns. For the range of power and frequency used here, the response of the diode can be considered as linear. The input dynamic range of the diode is between 1 μ W and 3.15 mW, corresponding to a DC output level of 0 mV to 400 mV. We use an amplifier to adjust the emitted power of the nano-oscillator to the working range of the diode. The output signal is then recorded by a real-time oscilloscope.

2.2 Physical properties of the oscillator used for computation

The two main properties of the spin-torque nano-oscillator used for computation are the non-linearity of the oscillation amplitude with the input DC current and the relaxation of the oscillation amplitude. These two properties are necessary to ensure the separation property and the fading memory needed for reservoir computing [10].

2.2.1 Non-linearity of the oscillations amplitude

A reservoir realizes a non-linear transformation of an input signal. This non-linearity allows the reservoir to project the initial problem into a different space in which classes that were not linearly separable in the original space become linearly separable. The variable used for computation is the amplitude of the oscillations (Figure 1b). The corresponding non-linearity is thus the non-linearity of the amplitude level as a function of the DC current. Figure 1c shows this non-linearity for a magnetic field of 400 mT. Below a threshold current (which is around 3 mA for this sample at this magnetic field), the oscillator does not emit any signal because the spin torque is not sufficient to compensate the damping and thus the spin polarized current does not move the vortex core. Above this threshold current, the vortex core gyrates. The amplitude of the oscillation is proportional to the radius of the vortex orbit s . The non-linearity of the amplitude evolves approximately like $\lambda(I, H_{\perp})\sqrt{I - I_{th}}$ [7].

Figure 1c illustrates the effect of the bias current on the oscillation amplitude response \tilde{V} to an AC current. As mentioned in Sec. 2.2.1, the input signal delivered by the arbitrary waveform generator has an amplitude of 500 mV peak to peak which corresponds to a 6 mA variation of injected current. Thus the input induces variations on a limited part of the non-linear amplitude curve. This area is represented in magenta in Figure 1c for a DC current of 6 mA. By changing the DC current, one would move the pink area (which is centered on the bias DC current value) and thus the typical excursion in the voltage amplitude which is explored. For computation, it is not necessary that this excursion includes the threshold current (unlike with the classical rectified linear activation function). Indeed the square root-like behavior of the amplitude provides the non-linearity for separating the inputs in the framework of reservoir computing.

The influence of the magnetic field is also crucial because it changes the shape of the non-linear function itself. It changes the threshold current (for smaller fields the threshold current is smaller) and the amplitude of the non-linearity. The optimum choices for DC currents and magnetic fields is be discussed in Sec 4.

2.2.2 Relaxation of the oscillation amplitude

The second essential property for single node reservoir computing is a form of memory. This property is important both for ensuring the connectivity between temporal neurons and to ensure a fading memory. This demonstration only uses the intrinsic memory due to the relaxation of the oscillator's amplitude. When the excitation of the vortex core changes suddenly, the amplitude of its orbit changes more slowly. Thus, the orbit radius and the proportional amplitude also change slowly. Figure 1d shows the variation of the oscillation amplitude when the oscillator is subjected to a varying input signal. The response of the oscillator is plotted in grey and the amplitude of the oscillation is highlighted in blue. Transitions in the input signal are much more abrupt than in the oscillation amplitude signal. The characteristic time of these changes is the relaxation time of the oscillator which is

roughly inversely proportional to the frequency and the damping factor ($T_{relax} \approx \frac{1}{\alpha f}$ [11]). For our sample, the relaxation time is around 200 ns, except when the current is close to the threshold. For the later regime, the relaxation time is larger but the oscillation amplitude is low and very noisy, making it difficult to exploit this regime for computation.

The relaxation is important to determine the discretization time θ which is used to define the state of one temporal neuron in a ring of similar neurons. In order to have connections between the temporal neurons, we should choose $\theta < T_{relax}$. For the work described in this chapter, $\theta = 100$ ns. This choice is discussed in Sec. 4.

3 Results on classification tasks

3.1 Results on sine/square recognition task

3.1.1 Task

The sine and square waveform classification task has been used in other studies [12] to evaluate the performance of reservoir computing based on an oscillator with delayed feedback. The goal is to classify each point of the input as part of a sine by returning an output 0, or as part of a square by returning an output 1. Each period of sine and square is discretized into eight points giving 16 different cases to classify. The input $u(k)$ is composed of 1280 points (160 randomly arranged periods of sine or square). The first half of the points are used for training (to find optimum output weights W) and the second half for testing.

The different inputs to classify are shown as red dots in Figure 2a. The inputs take 5 different values in a sine period and two different values in a square input. 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 addition, in the sine period the 3rd and the 7th point have a value +1 and -1 that corresponds to the values taken by the input in the square. So in the absence of memory, when the input value is +1 or -1 it is impossible to know whether these points belong to a sine or to a square. Therefore this temporal pattern recognition task is not trivial because it needs both the non-linearity and the memory of a neural network.

3.1.2 Protocol

Different steps of the protocol are represented in Figures 2a-c. Figure 2 shows how the oscillator is driven by the input signal. The input $u(k)$ (Figure 2a) is a sequence of discretized sine and square periods. This discrete input is then preprocessed. This preprocessed input is a time-multiplexed version of the inputs that different neurons should receive in a standard spatial neural network. Here the neural inputs are the value of $u(k)$ multiplied by a coefficient ± 1 . This allows the oscillator response to

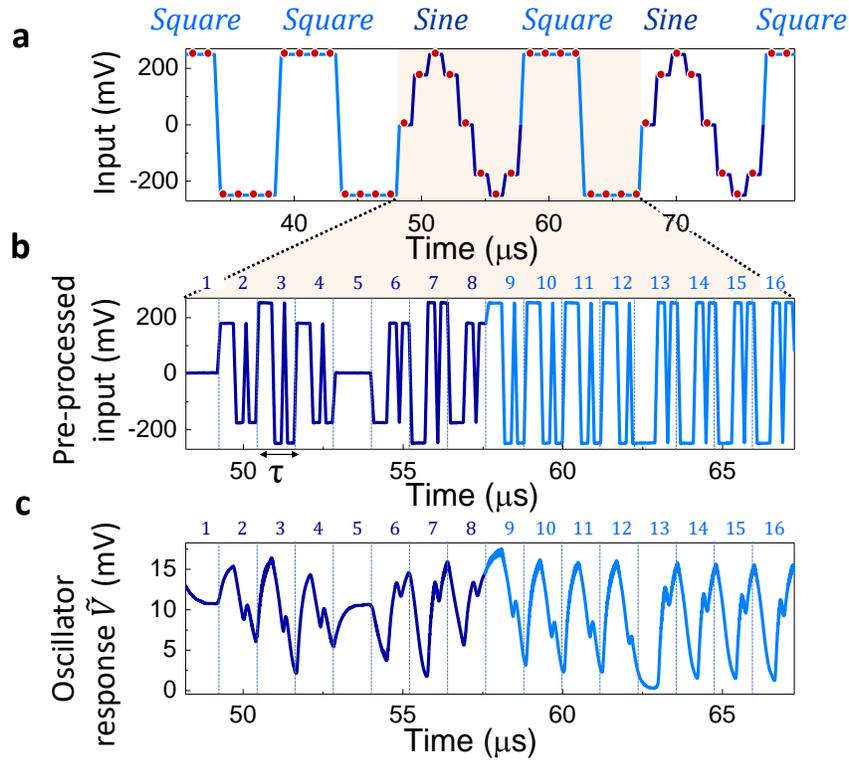


Fig. 2 Sine/square classification process: a. The signal to classify composed of periods of sine and square with the points to identify as belonging to a sine or to a square in red. b. The preprocessed input signal for a sine (dark blue) and square (light blue) period for a mask with 12 values used. c. Oscillation amplitude transient response to the oscillator signal. Figure extracted from Riou et al. [6].

be a time-multiplexed version of a spatial neural-network response. Later we will refer to these equivalent neurons as temporal neurons. If the preprocessed input varies faster than the relaxation time of the oscillator, it creates connections between the temporal neurons because the saturation is never reached so the oscillator state always depends of its previous state. The preprocessed input $J(t)$ for a sine and square period is shown in Figure 2b. An input with only 12 temporal neurons is chosen here. The reservoir emulates 12 temporal neurons with a mask containing only binary values +1 and -1. The number of neurons is smaller than the optimum number for better visualization in the figure. The results in Sec. 3.1.3 are based on a 24 temporal neurons reservoir. The time allocated to each neuron, θ , is taken here to be 100 ns, which we found to be optimal (see Sec. 4). When the oscillator receives this preprocessed input, it emits a transient response. The amplitude of the emitted oscillation measured experimentally is plotted in Figure 2c.

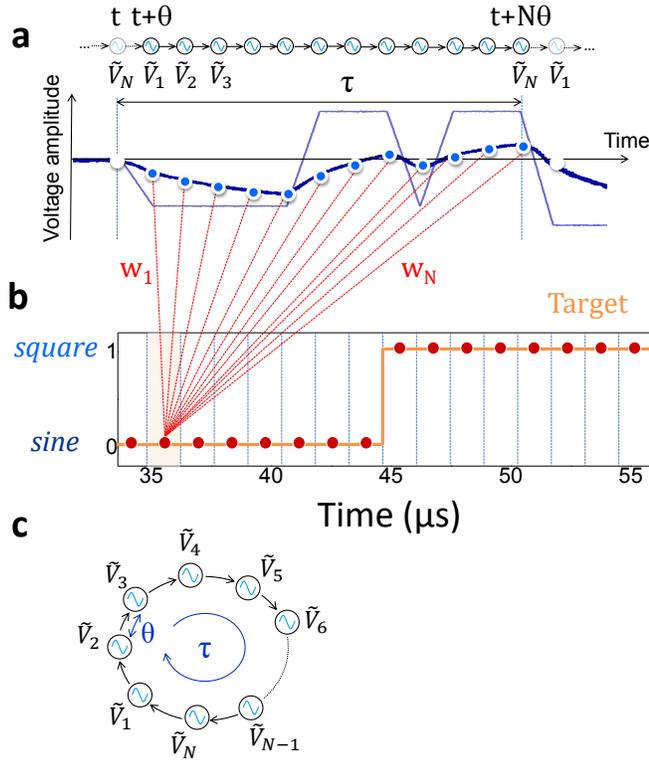


Fig. 3 Mapping and reconstruction of the output: a. The oscillator voltage amplitude \tilde{V} during a single time segment $\tau = N\theta$. Here, $N = 12$ neurons (12 samples \tilde{V}_i separated by the time step θ) are used to construct the output. b. Target for the output, $\sum_{i=1}^N w_i \tilde{V}_i$, reconstructed from the output voltages \tilde{V}_i and the weights w_i in each time segment τ . c. The transient states of the oscillator give rise to a chain reaction emulating a neural network with a ring structure. Figure extracted from Riou et al. [6].

Figure 3a shows how to retrieve the mapping of the preprocessed input by the single oscillator. This step is done off line on a computer. Discrete points are sampled at every time step θ . The values of these points correspond to the response of the temporal neurons. Measuring these values gives the reservoir state. Sampling the temporal traces of the oscillation amplitude properly requires aligning them with the preprocessed input (misalignment can result in bad classification). The rest of the process is a standard reservoir computing procedure. The neuron responses are linearly combined to reconstruct the output. This step is shown in Figure 3b. The coefficients of the linear combination are obtained by linear regression method. This approach using a single oscillator with time multiplexing emulates the response of a ring shape recurrent neural network. This architecture is represented in Figure 3c. Response of this equivalent neural network is used to classify the sine and square inputs.

3.1.3 Set point dependent results

Figure 4a shows the best reconstructed output obtained by experimentally emulating a 24-neuron network. The root mean square (RMS) deviation between target and output is 11 %, which is small enough to distinguish between sines and squares without any error (perfect classification) for the chosen choice of parameters: DC current $I_{DC} = 7.2$ mA, magnetic field $\mu_0 H = 447$ mT, input amplitude $V_{in} = 500$ mV (equivalent to 6 mA peak to peak). Figure 4a shows that if we trace a threshold line (in blue) at 0.5, all the outputs for square inputs are over this threshold and all the points for sine inputs are under this threshold. This perfect classification is achieved if the RMS deviation between the target and the output is small enough, here, close to 10 %.

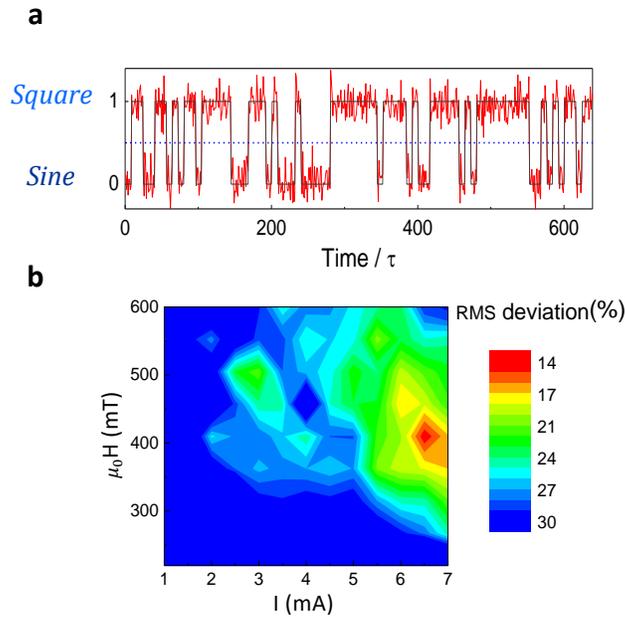


Fig. 4 a, Reconstructed output (red) and target (grey) in response to an input waveform with 80 randomly arranged sines and squares. The magnetic field is 447 mT, and the applied current 7.2 mA. The results are based on 24 neurons separated by $\theta = 100$ ns. b, Root mean-square deviation of output-to-target deviations: map as a function of DC current I and magnetic field H .

Perfect classification is obtained for operating points (the values of the bias DC current and the bias magnetic field) that give a small enough RMS deviation between the target and the output. As seen in Figure 4b, the RMS deviation varies from 10 % to more than 30 % depending on the bias conditions. We interpret the optimal operating point conditions in Sec. 4. After identifying this region of magnetic field and DC current leading to high performance of the oscillator for sine/square classification, we discuss the more complex task of spoken digit recognition.

3.2 Results on spoken digit recognition

3.2.1 Task

Spoken digit recognition is a widely used benchmark task in the hardware reservoir computing community [12, 13, 14, 15, 16]. The goal of the task is to recognize digits from audio waveforms produced by different speakers. For this task, the inputs are taken from the NIST TI-46 data corpus. The input consists of isolated english spoken digits said by five different female speakers. Each speaker pronounces each digit ten times. The set of ten digits said by all of the speakers is called an "utterance". The 500 audio waveforms are sampled at a rate of 12.5 kHz and have variable time lengths. Only female speakers are chosen to benchmark with the literature. Adding male and children increases the complexity of the task because it increases the variety of voice tones and thus increases the dispersion of the data to classify.

3.2.2 Protocol

The recognition of audio waveforms requires an additional time-domain to time-frequency-domain transformation, before inputting the signal into the reservoir. For this purpose we used two different methods: a spectrogram and a cochlear model. Both methods divide each word in several time intervals N_τ . Each of these time sequences has a fixed length τ and undergoes a frequency analysis through either Fourier transform (spectrogram model; 65 channels, $N_\tau \in 24, \dots, 67$) or a more complicated nonlinear approach (cochlear model; 78 channels, $N_\tau \in 14, \dots, 41$), which uses several different notch filters followed by non-linear automatic gain controllers [17].

After the transformation from time-domain to time-frequency domain, each word is represented by a matrix with $N_f = 65$ or $N_f = 78$ rows representing the frequency channels and N_τ columns representing the time (Figure 5b). These inputs are then preprocessed being multiplied by a $N_f \times N_\theta$ matrix (where N_θ is the number of neurons in the reservoir) called a mask, containing binary values. The resulting input for the oscillator is $N_\theta \times N_\tau$ matrix. Here we are emulating $N_\theta = 400$ temporal neurons, each of which is connected to all of the frequency channels for each time interval with the binary input weights defined by the mask.

Each preprocessed input value is consecutively applied to the oscillator as a constant current for a time interval of $\theta \approx 100$ ns (Figure 5c). This time is short enough to guarantee that the oscillator is maintained in its transient regime so the emulated neurons are connected to each other, but is long enough to let the oscillator respond to the input excitation. The amplitude of the AC voltage across the oscillator is recorded for offline post-processing (Figure 5d).

As for the previous task, the post-processing is then separated in two different phases: training and testing. During the training phase, a part of the spoken digits is used to determine the optimal sets of output weights $w_{i,\theta}$, where i indexes the desired digit. The recorded traces are multiplied by the output weights $w_{i,\theta}$ and

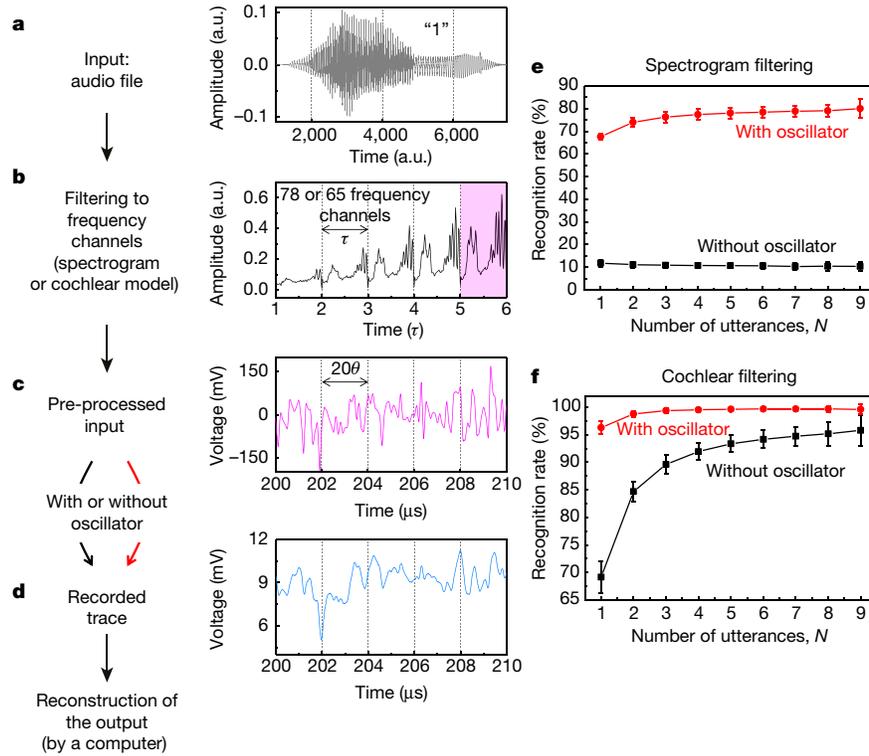


Fig. 5 Protocol for spoken digit recognition. a-d, Principle of the experiment. a, Audio waveform corresponding to the digit 1 pronounced by speaker 1. b, Filtering to the frequency channels for acoustic feature extraction. The audio waveform is divided in intervals of duration τ . The cochlear model filters each interval into 78 frequency channels (65 for the spectrogram model), which are then concatenated as 78 (65) values for each interval, to form the filtered input. c, Pre-processed input (transformed from the purple shaded region in b). The filtered input is multiplied by a randomly filled binary matrix (masking process), resulting in 400 points separated by a time step θ of 100 ns in each interval of duration τ ($\tau = 400\theta$). d, Oscillator output. The envelope $\tilde{V}(t)$ of the emitted voltage amplitude of the experimental oscillator is shown ($\mu_0 H = 430$ mT, $I = 6$ mA). The 400 values of $\tilde{V}(t)$ per interval τ (\tilde{V}_i , sampled with a time step θ) emulate 400 neurons. The reconstructed output '1', corresponding to this digit, is obtained by linearly combining the 400 values of \tilde{V}_i , sampled from each interval τ . e,f, Spoken-digit recognition rates in the testing set as a function of the number of utterances N used for training for the spectrogram filtering (a; $H = 430$ mT, $I = 6$ mA) and for the cochlear filtering (b; $\mu_0 H = 448$ mT, $I = 7$ mA). Because there are many ways to pick the N utterances, the recognition rate is an average over all $10! / [(10 - N)! N!]$ combinations of N utterances out of the 10 in the dataset. The red curves are the experimental results using the magnetic oscillator. The black curves are control trials, in which the pre-processed inputs are used for reconstructing the output on a computer directly, without going through the experimental set-up. The error bars correspond to the standard deviation of the recognition rate, based on training with all possible combinations. Figure extracted from Torrejon et al. [5].

then averaged over the N_τ time steps. It results in 10 output values y_i , which should ideally be equal to the target values $y_i = 1.0$ for the appropriate digit and 0.0 for

the rest. In the training process, a fraction of the utterances are used to train these weights; the rest of the utterances are used in the classification process to test the results. The optimum weights are found by minimizing the difference between \tilde{y}_i and y_i for all of the words used in the training. The reconstruction of the outputs y_i is a linear combination and thus finding the optimal weights for the training examples is a linear regression process. If we consider the target matrix \tilde{Y} , which contains the targets \tilde{y}_i for all of the time steps τ used for the training, and the response matrix S , which contains all neuron responses for all of the time steps τ used for the training, then the matrix W , which contains the optimal weights, is given by $W = \tilde{Y}^\dagger S$, where the symbol \dagger represents the Moore-Penrose pseudo inverse [18].

During the testing phase, the weights are fixed and applied to the remaining recorded traces. The ten reconstructed outputs corresponding to one digit are averaged over all of the time steps τ of the signal, and the digit is identified by taking the maximum value of the ten averaged reconstructed outputs. The efficiency of the recognition is evaluated by the word success rate, which is the rate of digits that are correctly identified.

Training is achieved with training sets which can have variable size (number of utterances used for the training) and composition (which combination of utterances is used for the training). Different training sets led to different performances during the testing phase. We trained the system using the ten digits spoken by the five speakers. The only parameter that we changed is the number of utterances used for the training. If we use N utterances for training, then we use the remaining $10 - N$ utterances for testing. However, some utterances are very well pronounced whereas others are hardly distinguishable. As a consequence, the resulting recognition rate depends on which N utterances are picked for training in the set of ten (for example, if $N = 2$, then the utterances picked for training could be the first and second, but also the second and third, or the sixth and tenth, or any other of the $10!/(8!2!)$ combinations of 2 picked out of 10). To avoid this bias, the recognition rates that we present here are the average of the results over all possible combinations. The error bars correspond to the standard deviation of the word recognition rate.

In order to see the contribution of the spin-torque oscillator in the recognition process we compare the results obtained from the oscillator time traces with a control trial. During the control trial, the pre-processed inputs are used for reconstructing the output on a computer directly, without going through the experimental set-up.

3.2.3 Preprocessing dependant results

The improvement shown in the experimental results over the control results (see Figure 5e) indicates that the spin-torque nano-oscillator greatly improves the quality of spoken-digit recognition, despite the added noise that is concomitant to its nanometre-scale size. In this case, the extraction of acoustic features, achieved by Fourier transforming the audio waveform over finite time windows, plays a minimal part in classification. Without the oscillator (black line), the recognition rates are consistent with random choices; with the oscillator (red line), the recognition rate

is improved by 70 %, reaching values of up to 80 %. This example highlights the crucial role of the oscillator in the recognition process.

Using the cochlear filtering (Figure 5f), which is the standard in reservoir computing and has been optimized on the basis of the behavior of biological ears, we achieve recognition rates of up to 99.6 %, as high as the state-of-the-art. Compared to the control trial, the oscillator reduces the error rate by a factor of up to 15. Our results with a spin-torque nano-oscillator are therefore comparable to the recognition rates obtained with more complicated electronic or macroscopic optical systems (between 95.7 % and 99.8 % for the same task with cochlear filtering).

3.3 Conclusion

In this section, two tasks were used to evaluate the performance of reservoir computing using the dynamics of a spin-torque nano-oscillator: sine/square classification and spoken digit recognition. Sine/square classification is a simpler task but allows testing the non-linear behavior and memory of the reservoir which are the key features for good classification on more complex tasks. Using 24 temporal neurons, a systematic study of classification for different magnetic fields and DC current bias conditions was conducted. In the best case a 10 % root-mean-square deviation between the reconstructed output and the target was obtained, which allows perfect classification of sine and square inputs. Best bias conditions were selected to move on to the more complex task of spoken digit recognition.

Spoken digit recognition requires frequency filtering of audio files prior to sending the input to the oscillator. Two filtering methods were studied: first a simple linear spectrogram method and a more complex cochlear decomposition to benchmark our result with the existing literature. Using the spectrogram method, the oscillator improves the recognition up to 70 % and the overall success rate is 80 % by training on 90 % of the data. This result stresses the critical role of the nano-oscillator for recognition. Using a cochlear decomposition, 99.6 % success rate was reached which is state of the art results for both numerical and hardware methods. Also it is important to note that training a linear classifier directly on a cochleogram enables a success rate of 96 %. Thus cochlear decomposition already separates a part of the input. The results in this chapter demonstrate neuromorphic computing performed with a nanoscale "neuron".

For this demonstration, experimental parameters such as the temporal time scale θ and the operating point (DC current and field) were specially selected to achieve good classification results. The next section elucidates the influence of these different parameters and gives guidelines to choose them appropriately.

4 Optimizing the experimental parameters and data processing for improved classification

4.1 Input sampling rate and amplitude

The memory and connectivity in single oscillator reservoir computing are obtained by driving the oscillator with a quickly varying preprocessed input. The time interval of the input variation is θ . In this part, we discuss how to optimize θ to obtain the best classification performance. Figures 6a-c illustrate the dynamical response of the oscillation envelope for different values of θ and different amplitudes of the input.

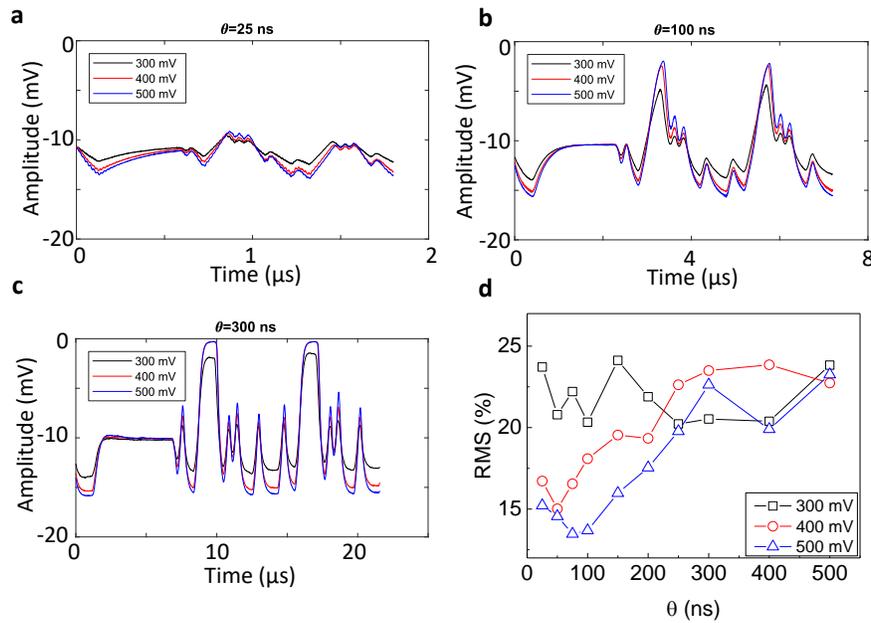


Fig. 6 a,b,c, \tilde{V} oscillation amplitude variation of the oscillator response for three levels of input 300 mV (black), 400 mV (red) and 500 mV (blue) in the case of time step $\theta = 25$ ns (a), $\theta = 100$ ns (b) and $\theta = 300$ ns (c). d, Classification performance depending on θ : Root mean square of output-to-target deviations as a function of the time step θ (separation between transient states of the oscillator \tilde{V}_i) for different amplitudes of the input signal (300, 400 and 500) mV when the target is in exact phase with the input.

The input signal sent to the oscillator by the arbitrary waveform generator (AWG) takes three different amplitude values (300 mV, 400 mV and 500 mV peak to peak). Figures 6a-c represent the oscillation amplitude response for respectively $\theta = 25$ ns (Figure 6a), $\theta = 100$ ns (Figure 6b) and $\theta = 300$ ns (Figure 6c). For $\theta = 25$ ns, the time rate of the input is much shorter than the intrinsic relaxation time of the oscillator measured around 200 ns ($\theta = T_{relax}/8$) so the oscillator stays in the transient regime

all the time. The memory effect is in this case strong but the variation of the emitted oscillation amplitude signal is quite small. The oscillator does not have the time to reach higher values. The amplitude of the oscillation also depends on the peak-to-peak input amplitude and for an input amplitude of 300 mV, the signal is particularly small.

At the other extreme for $\theta = 300$ ns, which is greater than the oscillator relaxation time ($\theta = 1.5T_{relax}$), the measured amplitude response reaches its saturation regime and the oscillator loses the memory of what happened one time step θ in the past. But for such high θ , the amplitude of the signal is much larger because the oscillator reaches its saturation regime.

For an intermediate value $\theta = 100$ ns (approximately $T_{relax}/2$), the oscillator still remains in a transient regime with a large amplitude. Intuitively, we expect that the trade off is to choose a value of θ that allows for a significant amplitude (so a decent signal over noise ratio) and still has memory. It is important to note that a higher input amplitude for a fixed θ always improves the signal to noise ratio.

Figure 6d presents the experimental RMS error result for sine and square classification as a function of θ . For 500 ns, increasing θ first decreases the RMS error until an optimum after which the RMS error increase again. This optimum seems to be around 100 ns. In the first part, the error decreases because the signal to noise ratio increases but after $T_{relax}/2$ the error increases again because of a loss of memory. This trend is also seen for an input amplitude of 400 mV. In that case, surprisingly the optimum seems to occur for a smaller θ of 50 ns. However the general result is always worse than for 500 mV input amplitude. Finally for an input amplitude of 300 mV, no clear trend is observed and the classification results are bad. For this particular voltage amplitude input, there does not seem to be any values of θ that have both sufficient memory and a sufficient signal-to-noise ratio.

4.2 Magnetic field and DC current dependence

4.2.1 Noise and amplitude ratio

Successful classification depends strongly of the operating point (Figure 7a). The magnetic field and the applied bias DC current determine the amplitude variation and the noise level of the oscillator signal. The amplitude of the measured signal and its asymmetry are measured through amplitude variations in both positive and negative directions, V_{up} and V_{dw} (Figure 7b upper graph), while the noise in the voltage amplitude is measured through the noise standard deviation ΔV (Figure 7b lower graph). We extract these parameters from the time traces of the voltage emitted from the oscillator at each bias point, and plot $V_{up}V_{dw}$ (Figure 7c) and $1/\Delta V$ (Figure 7d) as a function of the DC current I and field H .

Large oscillation amplitudes (red regions in Figure 7c) are obtained when the applied magnetic field is low, in such case the magnetization is weakly confined and when the applied DC current is high because it increases the spin torque on the

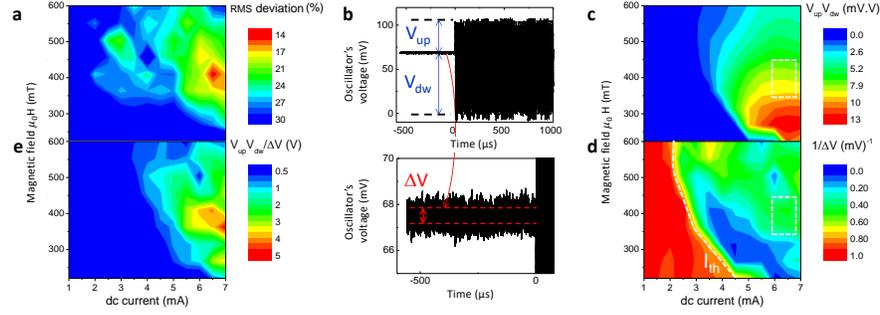


Fig. 7 a, Root-mean-square deviation of output-to-target deviations: map as a function of DC current I and magnetic field H . b, Amplitude variations in both positive and negative directions, V_{up} and V_{dw} (upper graph) and standard deviation of the noise in the voltage amplitude ΔV (lower graph). c, Maximal response of the oscillator to the input ($V_{up} V_{dw}$) in the H - I plane. d, Inverse of the noise level ($1/\Delta V$) in the H - I plane. The threshold current I_{th} is indicated in white dashed line. e, Map of the ratio maximal amplitude to noise $V_{up} V_{dw} / \Delta V$ in the H - I plane, showing that these parameters largely determine the performance of the oscillator (compare with a). Extracted from Torrejon et al [5].

magnetization. High noise regions (blue regions in Figure 7d) are obtained just above the threshold current for oscillation I_{th} . In such conditions, the oscillation amplitude increases rapidly as a function of the current and is sensitive to external fluctuations. The dotted white boxes in Figures 7c and d (currents of 6 mA to 7 mA and magnetic fields of 350 mT to 450 mT) highlight regions of high oscillation amplitude and low noise. Such bias conditions give root-mean-square deviations below 15 %, and there are no classification errors between sine and square waveforms. The similarity between the map of $V_{up} V_{dw} / \Delta V$ (Figure 7e) and the map of root-mean-square error (Figure 7a) confirms that the best conditions for the recognition are regimes with both high oscillation amplitude and low noise. We expect that the requirement, which we demonstrate here for a spin-torque oscillator, of a high signal-to-noise ratio to obtain good classification would apply to any type of nanoscale oscillator used for neuromorphic computing.

4.2.2 Amplitude level

Modifying the operating point modifies the oscillation amplitudes level as can be observed in Figure 8a. In the case of our sample, the highest amplitude are obtained for fields between 300 mT and 500 mT and for DC currents between 7 mA and 9 mA. During the reservoir computing experiments, we fix the magnetic field and we vary the current that the sample receives.

First the magnetic field should be selected so the input signal induces large and reproducible amplitude variation. By changing the the magnetic field, the threshold

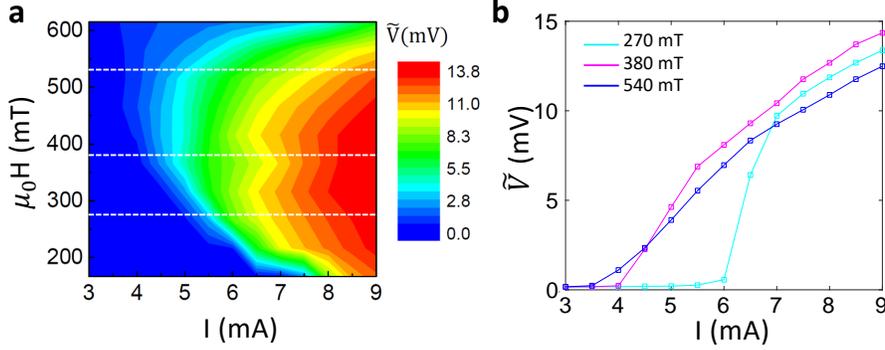


Fig. 8 a, Voltage amplitude \tilde{V} of the oscillator in the steady state: map in the I - H plane. b, Voltage amplitude \tilde{V} of the oscillator in the steady state as a function of I for $\mu_0 H = 273$ mT (cyan), $\mu_0 H = 379$ mT (magenta) and $\mu_0 H = 537$ mT (blue).

current and the derivative of \tilde{V} with the current are modified (Figure 8b). \tilde{V} is directly linked to the vortex orbit s . The oscillation amplitude is given theoretically by the following equation

$$\tilde{V}(t) = \lambda(H_{\perp}, I)s(t), \quad (1)$$

where s is the vortex radius and λ is a factor depending on the perpendicular magnetic field H_{\perp} and the DC current. s is a function of I/I_{th} and evolves approximately as $\sqrt{(I/I_{th}) - 1}$ [7]. Because of this dependency, for low magnetic field such that 297 mT, the threshold current I_{th} is high (5.9 mA) and \tilde{V} varies very rapidly with the input current I so the effect of a input current variation maybe not reproducible, particularly in the proximity of I_{th} . For higher magnetic field such as 537 mT, I_{th} is lower (4.0 mA) but the variation of \tilde{V} with I is slow, such as the input current variation results in small variation of \tilde{V} . Finally, optimal variations are obtained for intermediate values of field such as 379 mT, where the variations of input current induce significant change of \tilde{V} but the value of \tilde{V} is not too sensitive to noise.

Once the magnetic field is fixed, the DC current should be selected to obtained large variations of \tilde{V} and ensure that \tilde{V} does not evolve in a linear regime. The DC current influences both the amplitude and the asymmetry of the measured time traces as seen in Fig. 9 for three different DC currents, 4.5 mA, 6.5 mA and 9.0 mA. The left part of the figure shows the non-linear dependence of the oscillation amplitude on the current. The parts of this non-linear dependence explored for the chosen inputs are represented in the colored areas. An input of 300 mV induces a typical variation of DC current of roughly 4 mA. When the input is sent, the oscillator receives a current in a range ± 2 mA around the fixed DC current.

Changing the DC current explores different windows of the non-linear dependence of \tilde{V} on I , which is set by the magnetic field. This exploration window is represented by the vertical dashed lines. The central line is the DC current and the left and right lines are the extreme current values received by the oscillator. The level of the oscillation amplitude when no input signal is sent is the intersection of the central

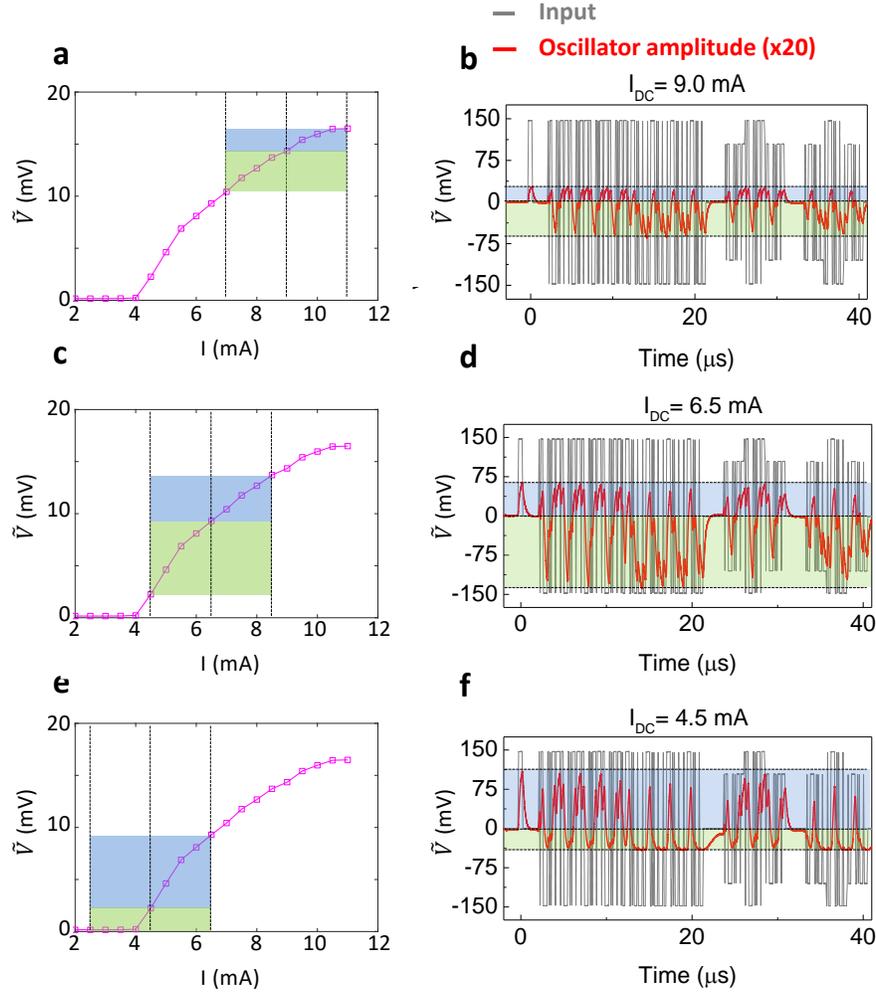


Fig. 9 Influence of the DC current on the oscillation amplitude variation: a,c,e \tilde{V} as a function of I . The blue (green) shaded area highlights the typical excursion in the voltage amplitude V_{up} (V_{dw}) that results when an input signal of $V_{in} = +150$ V ($V_{in} = -150$ mV) is injected. b,d,f The oscillator voltage amplitude, curves (in red) in response to the input waveform (in gray). Blue (green) shaded area represents V_{up} (V_{dw}). Curves are plotted for $\mu_0 H = 379$ mT and $I = 9.0$ mA (a,b), $I = 6.5$ mA (c,d) and $I = 4.5$ mA (e,f).

vertical line and the magenta curve $\tilde{V} = f(I)$. The higher oscillation amplitude value reached is the intersection of the right vertical line and the curve $\tilde{V} = f(I)$. V_{up} , the difference between the highest value reached by the oscillator amplitude when input is sent and the oscillation amplitude level due to the DC current, is represented in blue. Similarly the lowest amplitude reached when the input is sent is situated at the intersection between the left vertical line and the $\tilde{V} = f(I)$ curve. V_{dw} , the

difference between the oscillation amplitude due to the DC current and the lowest value reached by the oscillator amplitude when input is sent, is represented in green.

The right part of Fig. 9 shows the time traces corresponding to these different DC current conditions. V_{up} and V_{dw} are shown for the time traces with the blue and green areas. The apparent asymmetry of the measured time traces and the amplitude variation is completely determined by the portion of the non-linearity $\tilde{V} = f(I)$ that is explored. For 9.0 mA DC current value, the signal is small because the oscillation amplitude evolves in a part of the non-linearity $\tilde{V} = f(I)$ where variations are small (for high current the growth of \tilde{V} with I slows down). For 6.5 mA, the lowest current received by the oscillator is close but still greater than the threshold current (the oscillator receives currents between approximately 4.5 mA and 8.5 mA).

The region close to the threshold current is the region with the largest oscillation amplitude variations. The overall amplitude of the time traces for 6.5 mA is much larger than for 9.0 mA. Because the variation of the oscillation amplitude is stronger close to the threshold current, V_{dw} (green) is much larger than the V_{up} (blue). Decreasing the DC current to 4.5 mA gives saturation because the lowest DC current into the oscillator is then 2.5 mA, which is under the threshold current of 4 mA. Saturation, which maps a whole range of inputs to the same output, is detrimental for computation because these inputs can not be differentiated in later computation. In such cases the asymmetry of the time traces is reversed with V_{up} which is larger than V_{dw} .

An optimal oscillation amplitude variation requires an intermediate magnetic field (typically between 300 mT and 500 mT) so that the variation of the $\tilde{V} = f(I)$ non-linearity is not so abrupt such as for low field (because of the high I_{th} value) nor so slow as for high fields. Once the magnetic field is determined, the DC current should be chosen so that the lowest current sent to the oscillator from the AWG is close to the threshold current, giving the largest variations of the oscillation amplitude.

4.2.3 Noise and non-linearity trade-off

Good classification requires non-linearity in the reservoir. The non-linearity allows for separation of the different inputs by projecting the problem into a different space in the reservoir state. A large signal to noise ratio is also necessary, because otherwise similar inputs can be mapped onto very different reservoir states because of the noise so that the approximation property fails for the reservoir. Figure 10 shows that the non-linearity and the voltage noise vary considerably with DC current and magnetic field, suggesting that care is essential in selecting the working conditions.

To quantify the non-linearity, we computed the second derivative of the voltage with the current $\partial^2 \tilde{V} / \partial I^2$. Figure 10a shows that the largest non-linearities are close to the threshold current where the emission of the oscillator varies strongly. The noise level is evaluated by taking the standard deviation of the fluctuations of the voltage amplitude due to the noise (similarly to the noise evaluated in Sec. 3 Figure 4). Again, the noise also varies strongly close to the threshold current (Figure 10b).

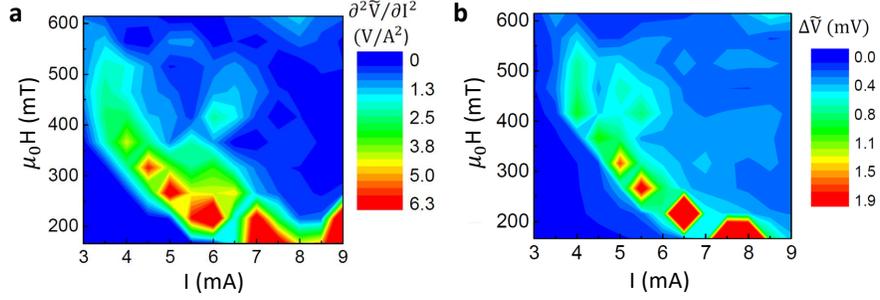


Fig. 10 a. The non-linearity $\partial^2 \tilde{V} / \partial I^2$ of the oscillator: map in the I - H plane. b. Amplitude noise $\Delta \tilde{V}$ of the oscillator in the steady state: map in the I - H plane.

Since spin-torque oscillators have a small magnetic volume, thermal noise affects the magnetization dynamics. The resulting voltage amplitude noise is large for large non-linearity, which quantifies the sensitivity of the system to perturbations. The correlation between voltage noise and non-linearity appears clearly in the comparison of Fig. 10a and b. Neuron non-linearity is a key ingredient for classification as it allows the separation of input data. On the other hand, noise in the response of neurons is detrimental for classification as it directly affects the output. Figure 7a shows the classification performance as a function of DC current and magnetic field. We find good performance by choosing a bias point with intermediate non-linearity and therefore intermediate noise, and where the neuron output changes strongly in response to the AC input. Such bias points allow enough non-linearity to classify while keeping a large enough signal to noise ratio to distinguish between outputs.

4.3 Conclusion

This section describes the experimental optimization of the θ time scale and the operating point for reservoir computing with a spin-torque oscillator. The θ time step plays both an important role in the memory of the reservoir (particularly when intrinsic memory is the only memory mechanism) and in the connectivity between the virtual neurons. A trade off leads to a value of θ short enough to assure memory and connectivity and long enough to assure a satisfactory amplitude variation. We find the best trade-off for $\theta = T_{relax}/2$ where T_{relax} is the relaxation time of the oscillator.

The operating point plays an important role in the optimization of these properties. The magnetic field should be between 300 mT and 500 mT so the variation of the non-linearity $\tilde{V} = f(I)$ is not too abrupt nor too smooth. At a given field, the DC current should be large enough to avoid reaching the threshold current and low enough to avoid a regime where \tilde{V} variation saturates. Finding the correct operating

point is a subtle balance, requiring both intermediate non-linearity and noise, while avoiding the threshold current.

5 General conclusion

In this chapter we demonstrate neuromorphic computing with a nanoscale hardware neuron. We use the non-linear transient dynamics of a single nanoscale spin-torque oscillator to emulate the response of an entire neural network, using the reservoir computing approach. The main physical properties used for this demonstration are the non-linear dependence of the emitted voltage amplitude on the applied current and the memory of the oscillator through the relaxation of the oscillation amplitude.

Using this non-linear dynamics, we classified sine and square waveforms, which requires both memory and non-linearity and we recognized successfully digits spoken by different speakers. For this last task we obtained a recognition rate of 99.6 % using cochlear decomposition prior to inputting the signal into the oscillator. With a spectrogram as time-to-frequency transformation, the final recognition rate is smaller (80 %) but the oscillator provides a higher gain.

The final recognition rate depends on the DC current and the magnetic field that is applied to the oscillator, because these bias conditions modify the regime where the oscillator operates determining both the signal-to-noise ratio and the non-linearity of the oscillator response. The noise and the non-linearity are correlated for spin-torque oscillators and optimal results are obtained for intermediate level of noise and non-linearity. The rate and the amplitude of the input which is sent to the oscillator also play an important role. The amplitude of the input influences the amplitude of the signal emitted by the oscillator and better results are obtained for larger input signals. The rate should ensure that the oscillator stays in a transient regime while keeping large amplitude variations of the emitted signal. Optimal variation time for the input is found for half of the oscillator relaxation time.

Using the reservoir computing approach, we show that spin-torque oscillators are stable enough and have both enough non-linearity and enough memory to perform neuromorphic computing. In this work we used a time-multiplexing approach for computation, allowing for the use of a single oscillator. This work combined with the ability of these oscillators to couple together opens the path to build large scale hardware neural networks with dense arrays of interconnected spin-torque oscillators.

Acknowledgements

This work was supported by the European Research Council ERC under Grant bioSPINspired 682955 and the French ministry of defense (DGA).

References

1. F. Rosenblatt, "The perceptron: a probabilistic model for information storage and organization in the brain.," *Psychological review*, vol. 65, no. 6, p. 386, 1958.
2. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, p. 436, 2015.
3. S. I. Kiselev, J. Sankey, I. Krivorotov, N. Emley, R. Schoelkopf, R. Buhrman, and D. Ralph, "Microwave oscillations of a nanomagnet driven by a spin-polarized current," *Nature*, vol. 425, no. 6956, p. 380, 2003.
4. W. H. Rippard, M. R. Pufall, S. Kaka, T. J. Silva, and S. E. Russek, "Current-driven microwave dynamics in magnetic point contacts as a function of applied field angle," *Physical Review B*, vol. 70, no. 10, p. 100406, 2004.
5. J. Torrejon, M. Riou, F. A. Araujo, S. Tsunegi, G. Khalsa, D. Querlioz, P. Bortolotti, V. Cros, K. Yakushiji, and A. Fukushima, "Neuromorphic computing with nanoscale spintronic oscillators," *Nature*, vol. 547, no. 7664, pp. 428–431, 2017.
6. M. Riou, F. A. Araujo, J. Torrejon, S. Tsunegi, G. Khalsa, D. Querlioz, P. Bortolotti, V. Cros, K. Yakushiji, A. Fukushima, *et al.*, "Neuromorphic computing through time-multiplexing with a spin-torque nano-oscillator," in *Electron Devices Meeting (IEDM), 2017 IEEE International*, pp. 36–3, IEEE, 2017.
7. E. Grimaldi, A. Dussaux, P. Bortolotti, J. Grollier, G. Pillet, A. Fukushima, H. Kubota, K. Yakushiji, S. Yuasa, and V. Cros, "Response to noise of a vortex based spin transfer nano-oscillator," *Physical Review B*, vol. 89, no. 10, p. 104404, 2014.
8. S. Tsunegi, H. Kubota, K. Yakushiji, M. Konoto, S. Tamaru, A. Fukushima, H. Arai, H. Imamura, E. Grimaldi, R. Lebrun, *et al.*, "High emission power and q factor in spin torque vortex oscillator consisting of feb free layer," *Applied Physics Express*, vol. 7, no. 6, p. 063009, 2014.
9. S. Tsunegi, K. Yakushiji, A. Fukushima, S. Yuasa, and H. Kubota, "Microwave emission power exceeding 10 μ W in spin torque vortex oscillator," *Applied Physics Letters*, vol. 109, no. 25, p. 252402, 2016.
10. L. Appeltant, M. C. Soriano, G. Van der Sande, J. Danckaert, S. Massar, J. Dambre, B. Schrauwen, C. R. Mirasso, and I. Fischer, "Information processing using a single dynamical node as complex system," *Nature communications*, vol. 2, p. 468, 2011.
11. A. Slavin and V. Tiberkevich, "Nonlinear auto-oscillator theory of microwave generation by spin-polarized current," *IEEE Transactions on Magnetism*, vol. 45, no. 4, pp. 1875–1918, 2009.
12. Y. Paquot, F. Duport, A. Smerieri, J. Dambre, B. Schrauwen, M. Haelterman, and S. Massar, "Optoelectronic reservoir computing," *Scientific reports*, vol. 2, p. 287, 2012.
13. L. Larger, M. C. Soriano, D. Brunner, L. Appeltant, J. M. Gutiérrez, L. Pesquera, C. R. Mirasso, and I. Fischer, "Photonic information processing beyond turing: an optoelectronic implementation of reservoir computing," *Optics express*, vol. 20, no. 3, pp. 3241–3249, 2012.
14. D. Brunner, M. C. Soriano, C. R. Mirasso, and I. Fischer, "Parallel photonic information processing at gigabyte per second data rates using transient states," *Nature communications*, vol. 4, p. 1364, 2013.
15. A. Dejonckheere, F. Duport, A. Smerieri, L. Fang, J.-L. Oudar, M. Haelterman, and S. Massar, "All-optical reservoir computer based on saturation of absorption," *Optics express*, vol. 22, no. 9, pp. 10868–10881, 2014.
16. Q. Vinckier, F. Duport, A. Smerieri, K. Vandoorne, P. Bienstman, M. Haelterman, and S. Massar, "High-performance photonic reservoir computer based on a coherently driven passive cavity," *Optica*, vol. 2, no. 5, pp. 438–446, 2015.
17. R. Lyon, "A computational model of filtering, detection, and compression in the cochlea," in *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'82.*, vol. 7, pp. 1282–1285, IEEE, 1982.
18. R. Penrose, "A generalized inverse for matrices," in *Mathematical proceedings of the Cambridge philosophical society*, vol. 51, pp. 406–413, Cambridge University Press, 1955.