



OPEN

Classification tasks using input driven nonlinear magnetization dynamics in spin Hall oscillator

John Rex Mohan¹, Arun Jacob Mathew¹, Kazuma Nishimura¹, Ruoyan Feng¹, Rohit Medwal^{3,4}, Surbhi Gupta^{3,5}, Rajdeep Singh Rawat³ & Yasuhiro Fukuma^{1,2}✉

The inherent nonlinear magnetization dynamics in spintronic devices make them suitable candidates for neuromorphic hardware. Among spintronic devices, spin torque oscillators such as spin transfer torque oscillators and spin Hall oscillators have shown the capability to perform recognition tasks. In this paper, with the help of micromagnetic simulations, we model and demonstrate that the magnetization dynamics of a single spin Hall oscillator can be nonlinearly transformed by harnessing input pulse streams and can be utilized for classification tasks. The spin Hall oscillator utilizes the microwave spectral characteristics of its magnetization dynamics for processing a binary data input. The spectral change due to the nonlinear magnetization dynamics assists in real-time feature extraction and classification of 4-binary digit input patterns. The performance was tested for the classification of the standard MNIST handwritten digit data set and achieved an accuracy of 83.1% in a simple linear regression model. Our results suggest that modulating time-driven input data can generate diverse magnetization dynamics in the spin Hall oscillator that can be suitable for temporal or sequential information processing.

Machine learning with artificial neural networks (ANNs) has become an essential component of modern computing^{1,2}. ANNs perform classification and pattern recognition tasks by nonlinearly transforming input data into a higher-dimensional space feature map. The ANNs are hierarchically connected units, constructed between an input and an output layer (classifier layer), with a predetermined number of inner layers. The connection strength between each layer, called weights, are trained to perform specific tasks. The inner layers, commonly called “hidden layers”, perform convolution operations in Convolution Neural Networks (CNN) or weighted sum (perceptron) operations in Feed-forward Neural Networks (FNN), followed by a nonlinear activation of inputs^{3–5}. Typically, problems with pattern recognition involve Feature Extraction (FE) or Feature Selection (FS) methods for mapping. FE involves the transformation of the original data into feature vectors and their subsequent mapping to a feature space, whereas FS involves choosing the relevant feature among the available states without any functional mapping³.

Feature extraction is a computationally intensive process. It requires large matrix operations for weight optimization across multiple connected layers of the network, resulting in a complex learning process and a significant amount of data⁶. The FS method aims to reduce the data storage requirement by removing redundant data and selecting only relevant data via filters, wrappers, and embedded operations^{7,8}. However, because of frequent memory access and high memory requirement, traditional von Neumann architecture suffers from speed and efficiency issues. Furthermore, large processing operations hinder the implementation of ANN for real-time processing in memory-constrained devices^{4,9}. Neuromorphic computing processors and hardware capable of in-memory computing are being considered for performing various computational tasks^{10–15}. Various physical systems including photonics or spintronics based hardware are considered, which can combine processing and memory capabilities to carry out weight optimization, nonlinear activations and in-situ mathematical operations like matrix–vector multiplication^{2,15–17}. Despite these advancements, on-device inference and classification of multibit input data on the device itself are scarce^{18–20}. By having specialized inference or feature extraction computing units, the energy costs of feature mapping, which make up 90% of the costs in currently used ANNs, can

¹Department of Physics and Information Technology, Faculty of Computer Science and Systems Engineering, Kyushu Institute of Technology, Iizuka 820-8502, Japan. ²Research Center for Neuromorphic AI Hardware, Kyushu Institute of Technology, Kitakyushu 808-0196, Japan. ³Natural Sciences and Science Education, National Institute of Education, Nanyang Technological University, Singapore 637617, Singapore. ⁴Present address: Department of Physics, Indian Institute of Technology, Kanpur 208016, India. ⁵Present address: Department of Physics, Motilal Nehru National Institute of Technology Allahabad, Prayagraj 211004, India. ✉email: fukuma@phys.kyutech.ac.jp

be significantly decreased^{15,21}. However, implementing dedicated computing units require overcoming significant hurdles such as, (i) efficient signal processing to generate readily accessible signal parameter identification and computational outputs, (ii) integration with existing complementary metal–oxide–semiconductor (CMOS) circuits, and (iii) adaptability to the existing machine learning algorithms²². Therefore, it is important to identify and model physical devices that can perform computational tasks while consuming less power, especially for devices with limited memory.

Spintronic oscillators such as spin-torque oscillators (STOs) and spin Hall oscillators (SHOs) have been widely studied for their use as feature filters in computational tasks such as pattern recognition and classification tasks. In STOs, the spin transfer torque (STT) is induced by the electric current flowing through a multilayer structure consisting of a free magnetic layer and a fixed magnetic layer separated by a non-magnetic conductor or an insulator^{23,24}. The electric current passing through the fixed layer is spin-polarized in the direction of magnetization and exerts a torque on the free layer magnetization. The SHO consists of a ferromagnetic (FM) and heavy metal (HM) bilayer structure in which the generation of a pure spin current in the HM via the spin Hall effect (SHE) or the Rashba–Edelstein effect induces the spin orbit torque (SOT)^{25–28}. The SHOs have a few additional benefits despite STOs having a higher oscillatory output power. First, they are easier to fabricate due to their simpler bilayer structure^{29–32}. Second, the SOT which is caused by the pure spin current from the HM electrode, can be exerted over extended areas in SHOs, whereas the STT in STOs is a localized effect^{31,32}. SOT can be successfully used to excite and regulate a variety of magnetization oscillations^{33–36}. Furthermore, the development of charge-to-spin conversion efficiency through material engineering and the significant contribution of a field-like term to the SOT, caused by interfacial effects, provides an interesting opportunity to investigate SHO-based neuromorphic hardware^{37–40}.

In this article, we use micromagnetic simulations to model a SHO device in the framework of a single domain and take advantage of the microwave spectral properties of its magnetization dynamics to process binary 0 and 1 inputs. We show that the spectral filter technique can be used to directly classify multibit binary data. The present SHO model is a simplified model and based on physically accessible parameters, which makes them easily accessible for feature extraction. The main motivation behind this work is to reduce the dimensionality of feature maps in order to achieve fast and efficient information processing with low training cost. By controlling the magnetization dynamics of the SHO through the parameters of the input pulse pattern, we are able to classify the input sequences that contain 4-binary digit data. Next, we demonstrate recognition of handwritten digits of the Modified National Institute of Standards and Testing (MNIST) handwritten digit database. A SHO device with a modified input pulse pattern requires static readout and training of a linear network to classify handwritten digits.

Results and discussion

SHO device model. A conceptual schematization of the SHO-based feature filter is shown in Fig. 1a. It consists of a pulse input, a SHO, a frequency domain output and a filter. The modelled SHO is composed of platinum/permalloy bilayer (Pt/NiFe) with a lateral size of $100 \times 100 \text{ nm}^2$. Each layer has a thickness of 5 nm. The magnetization is aligned at an angle $\phi = 90^\circ$ (+Y), by applying an in-plane magnetic field, $\mu_0 H_{\text{ext}} = 100 \text{ mT}$. The SHE in Pt causes spin-dependent electron scattering to the top and bottom surfaces of Pt upon the flow of current I_c in the X direction. This leads to spin accumulation at the interface of Pt/NiFe and a subsequent transfer of spin angular momentum to NiFe results in a transverse flow (Z) of spin current (I_s). The ratio between charge current density (j_c) and spin current density (j_s) is characterized by the spin Hall conversion efficiency, θ_{SH} (spin Hall angle)^{25,26}. The j_s results in two SOTs known as field-like torque (FLT) and damping-like torque (DLT). FLT is the result of an interfacial effect due to the strength of spin accumulation at the interface, whereas DLT is the result of bulk phenomena of the SHE^{41,42}. Due to the minimal impact of FLT for the 5 nm thick Pt Layer, we have only taken into account the role of DLT on the dynamics of magnetization in the simulations. The natural NiFe damping can be controlled by DLT and fully compensated to produce auto-oscillations in the gigahertz frequency range by increasing j_c ^{29,33,34}. The “Methods” section contains information on the simulation technique and the material parameters. The input data are in the form of current pulses. The binary digits “1” and “0” are encoded as two distinct current values, I_1 and I_0 , respectively. The output channel is assumed to be of anisotropic magnetoresistance (AMR) based electrical read-out which depends on the magnetization component M_x ^{24,43}. To represent the collective behavior of the SHO for the given input signal, the fast Fourier transform (FFT) transforms the simulated time-dependent M_x into frequency spectra. The main benefit of frequency domain analysis is the reduced volume of output data for additional computations. The following section will discuss how a filtering mechanism can be used to perform feature extraction and a straightforward linear classification of input data.

Inherent magnetization dynamics excited by SOT. We begin by examining the magnetization dynamics of the modeled SHO device as a function of I_1 for a single current pulse with a pulse width (τ) of 3 ns and pulse rise and fall times of 1 ns. In Fig. 1b, the magnetization component M_x 's temporal evolution is depicted for $I_1 = 5.0, 5.5, 6.0,$ and 6.5 mA with $I_0 = 0 \text{ mA}$. SOT excitation of a ferromagnetic resonance mode causes magnetization oscillations at $I_1 = 5.0$ and 5.5 mA . They exhibit small angle precession, in which the cone angle increases with an increase of I_1 ^{44–46}. As shown by the colored circles in Fig. 1b, the initiation and relaxation times vary depending on the strength of I_1 . This indicates the manipulation of effective damping by the SOT. These excitations can be converted into self-sustaining auto-oscillations by gradually increasing the precession amplitude with τ until it saturates at the limit cycle of auto-oscillation⁴⁷. For $I_1 = 6.0$ and 6.5 mA , the oscillations correspond to the in-plane and out-of-plane auto-oscillation modes, respectively. The equilibrium energy, which regulates the auto-oscillation's limit cycle, is influenced by the device geometry, mode of excitation, and

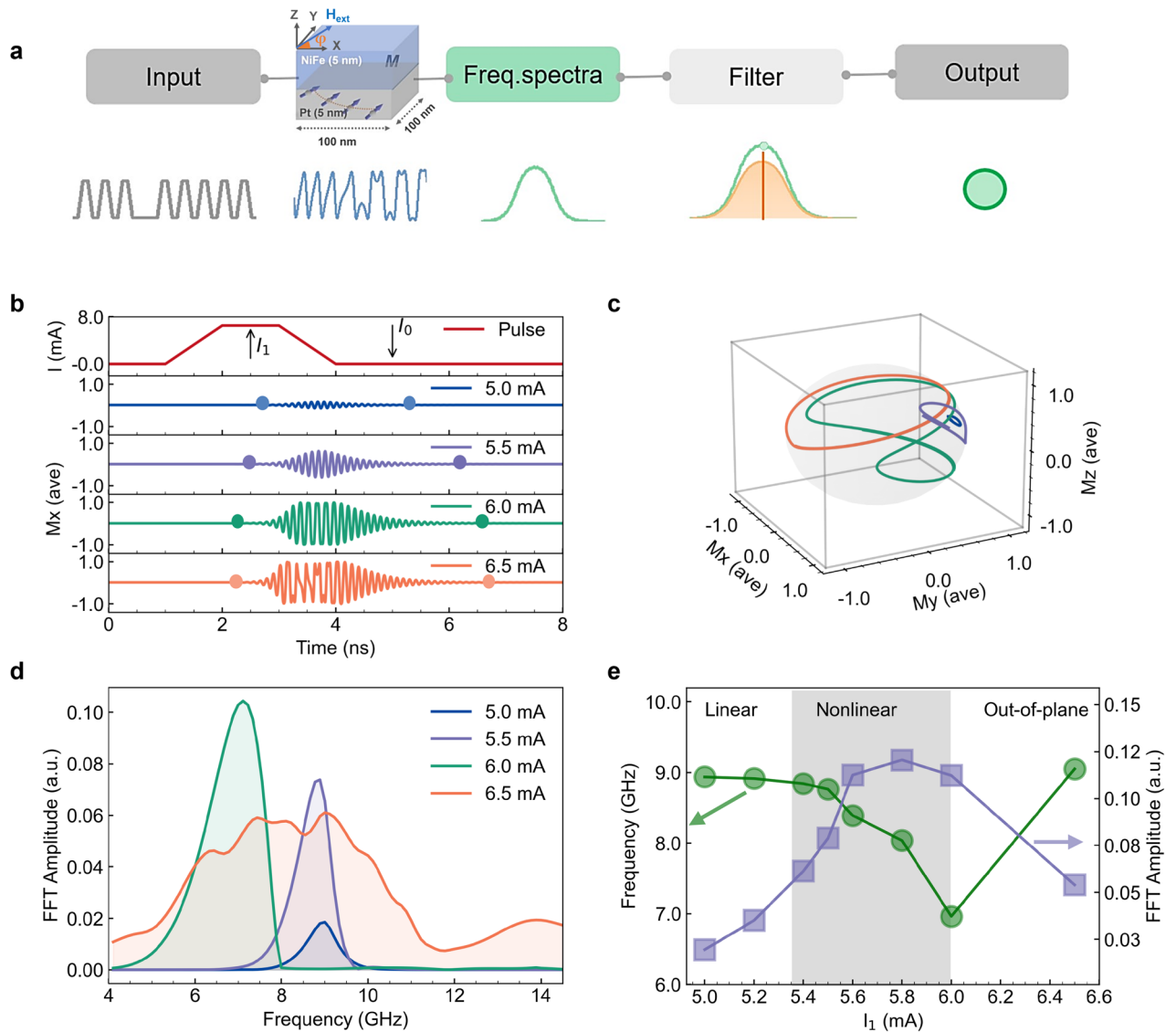


Figure 1. (a) Conceptual filtering process for the classification of inputs with use of a spin Hall oscillator. (b) Pulsed input current and magnetization dynamics of M_x component for $I_1 = 5.0\text{--}6.5$ mA, the initiation and relaxation time scales are indicated by colored circles. (c) Transition of small angle precession to large angle oscillation trajectories as a function of I_1 . (d) Fast Fourier transformed amplitude spectrum of time domain M_x component as a function of I_1 . (e) Frequency response and peak amplitude level of FFT spectrum as a function of I_1 .

direction of the effective field. The auto-oscillation orbit for the single magnetic domain model is circular in the out-of-plane direction and shaped like a clamshell in the in-plane direction⁴⁵. The single cycle trajectories for each oscillation mode are depicted in Fig. 1c.

In Fig. 1d, the FFT amplitude spectrum for the I_1 range (5 mA to 6.5 mA) is displayed. For $I_1 < 5.5$ mA, the FFT amplitude increases linearly with increasing I_1 , but the oscillation frequency, which can be seen from the peak position of the FFT amplitude in the spectra, is constant at 9.0 GHz. This constant frequency corresponds to the ferromagnetic resonance frequency (Supporting Fig. S1). For $5.5 \text{ mA} < I_1 < 6.0 \text{ mA}$, the frequency undergoes a red shift due to the large angle motion of magnetization components, as can be seen in Fig. 1c ($I_1 = 6.0$ mA), which reduces the effective demagnetization field in the NiFe layer. The magnetization component transverse to H_{ext} undergoes oscillations at twice the oscillation frequency in order to maintain a constant magnitude and thus reduces the frequency. As I_1 is increased further, the frequency decreases, reaching 7.1 GHz at $I_1 = 6.0$ mA. This frequency shift is attributed to the complex coupling of oscillatory amplitude and phase as predicted by the nonlinear auto-oscillator theory for STOs⁴⁸. For $I_1 = 6.5$ mA, the frequency increases to 9.1 GHz and the FFT amplitude reduces due to the out-of-plane oscillation, as can be seen in Fig. 1c ($I_1 = 6.5$ mA). As a result, the magnetization dynamics in the SHO can be divided into three regimes, as shown in Fig. 1e: the linear excitation regime for $I_1 < 5.5$ mA, where the frequency is constant with increasing precession amplitudes as a function of I_1 , the nonlinear excitation regime for $5.5 \text{ mA} < I_1 < 6.0$ mA, where the FFT amplitude is saturated and the

frequency decreases drastically with increasing I_1 and the out-of-plane oscillation regime for $I_1 > 6.5$ mA where the frequency increases and the FFT amplitude decreases. These nonlinear frequency amplitude relationships can be used to classify the inputs.

Binary inputs classification. Regular pulse scheme. After investigating magnetization dynamics, we look into the SHO's ability to classify n -binary input data. The pulse stream is represented by the encoded input signal $n - bi(t)$, which has current values I_0 and I_1 for "1" and "0", respectively. The pulse period (Δt) and width (τ), respectively, 4 ns and 3 ns. The pulse width τ , includes a rise time of 1 ns and a fall time of 1 ns. Figure 2 represents 4 - $bi(t)$ input pulse patterns, Mx responses, and FFT amplitude spectra (frequency) with input current values of $I_1 = 3.5$ mA and $I_0 = 0$ mA, which lies in the linear excitation regime. For the input pattern 1111 in Fig. 2a, the Mx response in Fig. 2b shows the magnetization oscillations with varying amplitude for each "1" input, and the corresponding FFT spectrum is shown in Fig. 2c. For the 1001 pattern in Fig. 2d, the magnetization relaxes to its initial state in the time between the two "1" inputs, as shown in Fig. 2e. The corresponding amplitudes in the FFT spectrum (Fig. 2c,f) for the input patterns of 1111 and 1001 allow one to clearly see the difference in magnetization dynamics.

In order to categorize the patterns of the input data and extract features from the outputs, we employ the classic signal processing tool of filtering. We apply the filter neuron concept, which is inspired by recurrent neural networks' use of it for feature extraction and handling time-varying outputs, to filter a specific feature in the output data^{49,50}. The linear excitation regime frequency, 9.0 GHz, is fixed as the filter and the FFT amplitude value at this frequency is set as the filter output characteristic feature in order to separate the input patterns in the output spectrum. The filtered amplitude values for the input patterns 1111 and 1001 at 9.0 GHz are 0.0042 and 0.00002, respectively, as shown in Fig. 2c,f. Note that this filtering strategy differs from the standard bit slicing techniques used in the computing paradigm. The bit slicing method maps n input elements to n output values and then performs additional computations for weight optimization⁹. But in this case, n input elements are mapped to a single output value using the quantization technique. This method is well suited for the reduction of output data and does not call for weight optimization for the classification task of input patterns, which can reduce the computation costs^{51,52}.

The relaxation of magnetization precession during the interval between two consecutive pulses poses a challenge to the classification task. For the 0101 input pattern in Fig. 2g, the Mx response is displayed in Fig. 2h. Since the magnetization has relaxed to its initial state prior to the second "1" pulse arrival, the individual "1" pulses exhibit the same oscillating amplitude. As can be seen in Fig. 2f,i, the resulting FFT spectra have the same amplitude as the 1001 pattern. In this case, it is not possible to distinguish between the SHO's output and any of the possible 4 - $bi(t)$ input patterns. Input pulse parameters I_0 , I_1 and τ can be varied to affect the dynamics of the magnetization, but in the linear excitation regime, patterns like 1000 and 0001 still produce the same FFT spectra. This is caused by the same magnetization dynamics, but in a different time frame, for each "1" pulse. We refer to this type of input pulse stream as the regular pulse scheme. Supporting Figures S2 and S3 provide an analysis of the magnetization dynamics for various I_1 values, as well as the resulting FFT spectra and the inability to separate 4 - $bi(t)$ input patterns.

Modified pulse scheme. To tackle the challenge faced in the regular pulse scheme, we resort to modifying the input driven magnetization dynamics rather than the internal structure of the device⁵³. Figure 3a depicts the pulse input of an excitatory pulse I_e , with a pulse width (t_1) of 7 ns. The Mx oscillation, as shown in Fig. 3b

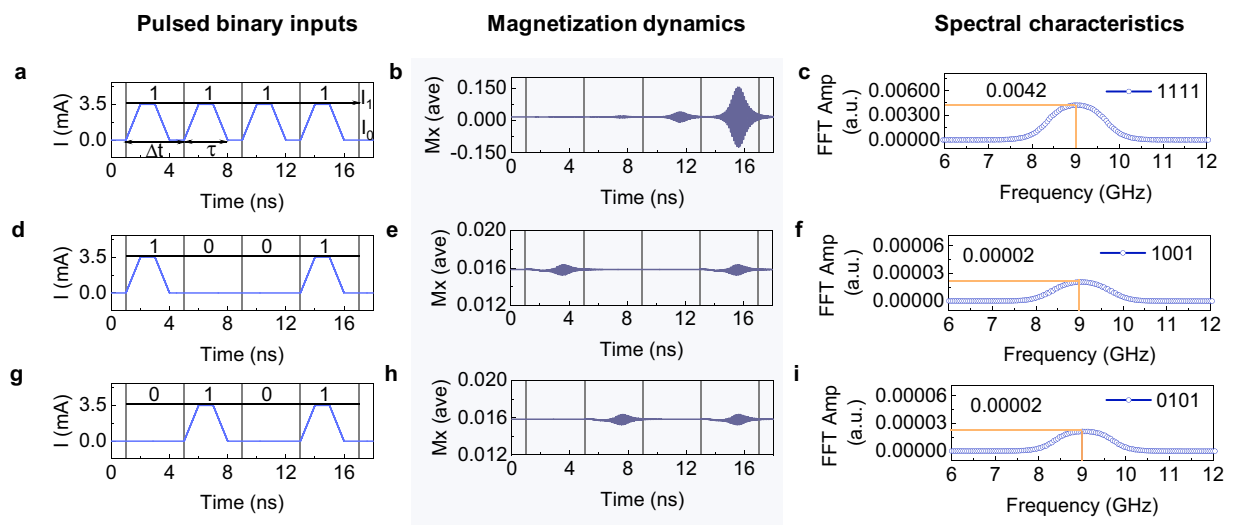


Figure 2. (a) Regular pulse scheme for 4-binary digit input pattern 1111 with pulse period (Δt) of 4 ns and pulse width (τ) of 3 ns. (b) Magnetization dynamics of Mx component and (c) corresponding spectral characteristics. Guide line in FFT shows the amplitude value at 9.0 GHz. Similar input pattern, magnetization dynamics and FFT in (d–f) for 1001 pulse pattern, and in (g–i) for 0101 pulse pattern.

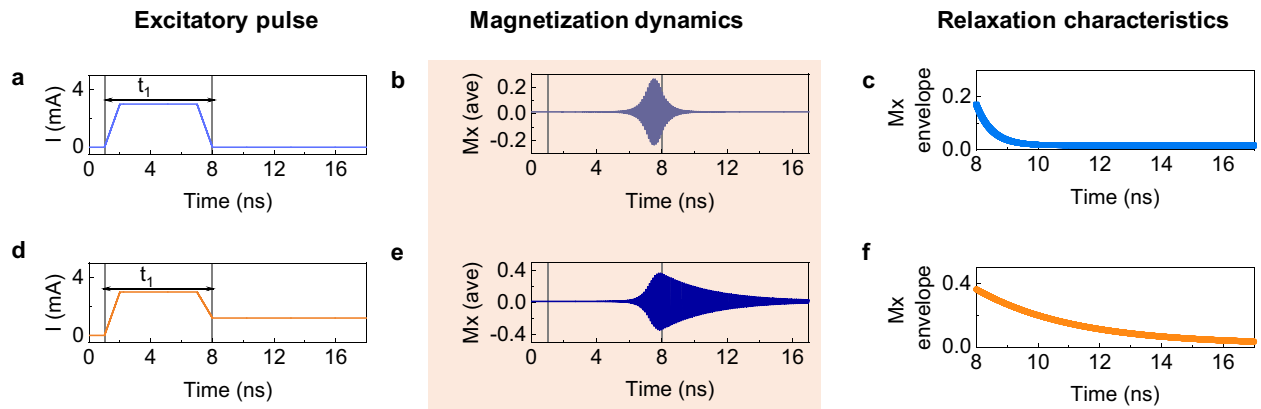


Figure 3. (a) Excitatory pulse ($I_e = 3.0$ mA) without offset current ($I_0 = 0$ mA). (b) Magnetization dynamics of Mx component and (c) relaxation characteristics plot from the upper envelope of the Mx time domain data after the pulse is off. (d) Excitatory pulse (I_e) with an offset current value ($I_0 = 1.2$ mA). (e) Magnetization dynamics of Mx component and (f) relaxation characteristics plot of the Mx time domain data.

reaches an amplitude of 0.18 and relaxes to the ground state within 2 ns when $I_e = 3.5$ mA and $I_0 = 0$ mA. This can be clearly seen from the upper envelope plot of Mx shown after the pulse is off at 8 ns in Fig. 3c. However, by introducing an offset value for I_0 , the relaxation time can be extended. Figure 3d shows an excitatory pulse with an offset value for I_0 ($I_e = 3.5$ mA, $I_0 = 1.2$ mA). The oscillation amplitude increases to 0.38 due to the increasing precessional amplitude with the offset current, as shown in Fig. 3e, and the relaxation period is extended by 8 ns, as shown in Fig. 3f. This allows us to modify the magnetization dynamics in the SHO during the inputs for different 4 – bi(t) patterns, as will be discussed below.

In order to extend the dynamics of magnetization relaxation for the duration of the input pulse pattern, a modified pulse scheme (I_{mod}) that includes the excitatory pulse (I_e) and a pulse gap (δ) prior to the introduction of the 4 – bi(t) is used. Hence, the SHO responds to a combination of two input signals, I_e and bi(t) given by,

$$I_{mod} = \begin{cases} I_e; & 0 < t < t_1 \\ n - bi(t); & t > t_1 + \delta \end{cases} \quad (1)$$

A modified pulse scheme with $I_e = 3.0$ mA, $t_1 = 7$ ns, $\delta = 5$ ns, $I_0 = 1.2$ mA, $I_1 = 2.4$ mA, $\Delta t = 4$ ns, $\tau = 3$ ns is used for illustration. For simplicity, we denote these input patterns with the above mentioned parameters as IP₁. Figure 4a shows the input pattern of 1010, the Mx response in Fig. 4b, and the FFT spectrum in Fig. 4c. Similarly, Fig. 4d shows the input pattern for 0101, the Mx response in Fig. 4e, and the FFT spectrum in Fig. 4f. As can be seen from the Mx responses in Fig. 4b,e, each “1” pulse exhibits a different oscillating amplitude as the magnetization dynamics is influenced by both the prior I_e and the corresponding pulses. Since I_e 's influence gradually diminishes over time, each output has a unique dynamic, and the degree of influence of previous inputs varies as a function of time. As anticipated, the variation in relaxation dynamics has a significant impact on oscillation amplitude. For the I_{mod} scheme, the FFTs were obtained from the Mx in the range of input pulse patterns $n - bi(t)$ (Supporting Fig. S4). Figure 4c,f show the variation in the FFT amplitude and frequency as well

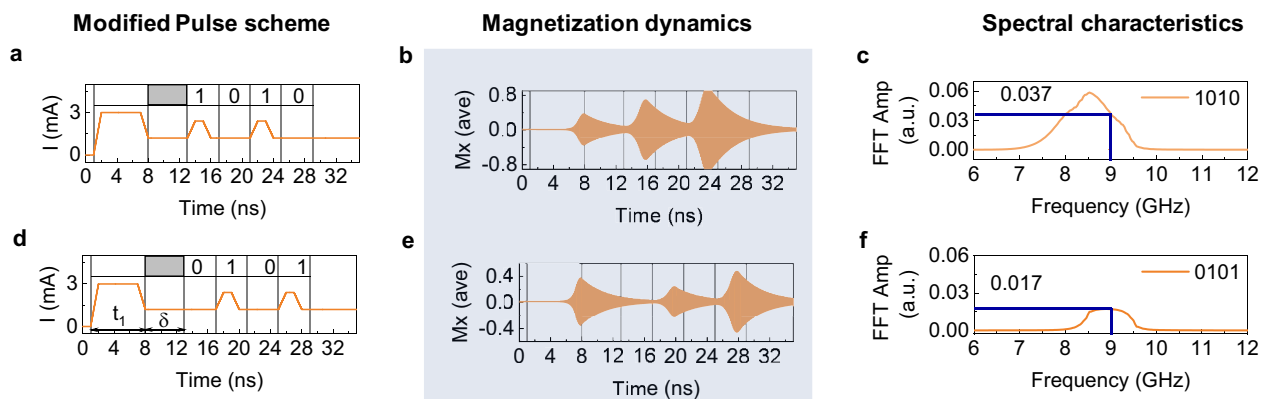


Figure 4. (a) Modified pulse scheme with an excitatory pulse $I_e = 3.0$ mA of pulse width $t_1 = 7$ ns, for 4-binary digit pattern 1010 with pulse period (Δt) of 4 ns and pulse width (τ) of 3 ns. The pulse gap of $\delta = 5$ ns between I_e and 4-binary digit patterns is shown with the grey box. (b) Magnetization dynamics of Mx component and (c) corresponding spectral characteristic FFT plot. (d–f) Similar plots of input pattern, magnetization dynamics and FFT for 0101 pulse pattern. Guide line in FFT plots show the amplitude value at 9.0 GHz.

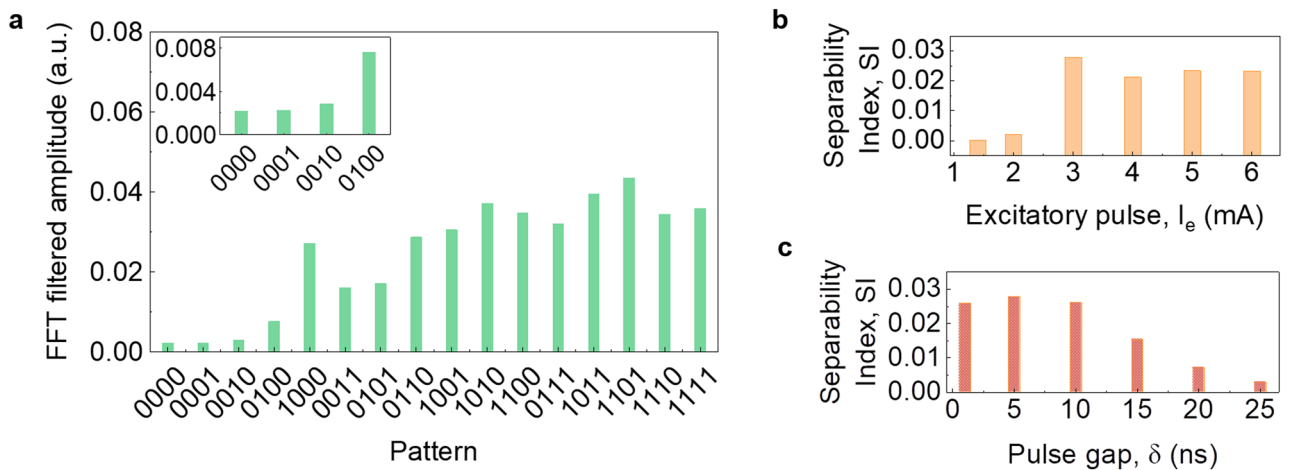


Figure 5. (a) Filtered amplitude at frequency of 9.0 GHz for all 16 4-binary digit inputs for the modified pulse scheme with input parameters $I_e = 3.0$ mA, $t_1 = 5$ ns, $\delta = 5$ ns, $I_0 = 1.2$ mA, $I_1 = 2.4$ mA, $\Delta t = 4$ ns, $\tau = 3$ ns. (b) Separability index for different excitatory pulse current values $I_e = 1.4$ –6.0 mA, with $\delta = 5$ ns. (c) Separability index for different pulse gap values $\delta = 1$ –25 ns with $I_e = 3.0$ mA.

as the filtered (9.0 GHz) amplitude values 0.037 and 0.017 for the patterns 1010 and 0101, respectively. Because of the input pattern-specific nonlinear magnetization dynamics, I_{mod} scheme can easily classify these two input patterns from the filtered amplitudes, whereas the previous regular pulse scheme could not. Similarly, the 1000 and 0001 patterns can also be distinguished using the filtered amplitude due to the variations in the relaxation rates. Supporting Figure S5 shows the 16 different $n - \text{bi}(t)$ input pattern combinations and their corresponding magnetization dynamics. Figure 5a displays the filtered FFT amplitude for the 16, 4 - $\text{bi}(t)$ input patterns using IP_1 . Due to the distinction in the filtered amplitude values, a 4 - $\text{bi}(t)$ pattern can be quantized and represented as a 1-dimensional analog output.

Separability index. We further examined how different input I_e and δ values in IP_1 affect the separation of 4- $\text{bi}(t)$ digit patterns. The results are compared using a commonly used metric in pattern recognition problems, called separability index, (SI) which is a measure of the average difference between the outputs, for different classes of input parameters^{54,55}. SI for IP_1 with $I_e = 1.4$ –6.0 mA and $\delta = 5$ ns is shown in Fig. 5b. For $I_e = 1.4$ and 2.0 mA, the magnetization precession excited by the first “1” pulse relaxes before the arrival of the next pulse. This results in the FFT amplitudes being very close to each other, leading to a low SI. On the other hand, for $I_e = 4.0$, 5.0, and 6.0 mA, the magnetization precession excited by the first “1” pulse shows a large amplitude of auto-oscillation, resulting in the consecutive $\text{bi}(t)$ pulses oscillating at the same amplitude. As a result, input patterns like 0111, 1011, 1101, 1110, and 1111 produce FFT amplitudes that are very close to one another. The close range of SI values for $I_e = 4.0$, 5.0, and 6.0 mA indicates that the excitatory pulse’s ability to distinguish between the various input patterns has been lost. We observed a maximum SI in the case of $I_e = 3.0$ mA, which indicates that a balance between the I_e pulse and $\text{bi}(t)$ pulses is needed.

Figure 5c shows SI for IP_1 with a range of pulse gaps, $\delta = 1$ –25 ns and $I_e = 3.0$ mA. The effect of I_e on $\text{bi}(t)$ pulses decreases as δ increases. Therefore, for input patterns like 0010 and 0001, the FFT amplitudes become similar to one another. I_e has a noticeable impact on $\text{bi}(t)$ for up to 20 ns before it is lost completely at $\delta = 25$ ns. For $\delta > 25$ ns, the magnetization dynamics are identical to the regular pulse scheme. Therefore, we can conclude that the separation of 4- $\text{bi}(t)$ inputs can be optimized to obtain the highest separation at the filtered FFT amplitude with proper I_e and δ . A comprehensive analysis of magnetization dynamics for various δ and I_e are summarized in the Supporting Figs. S6 and S7.

The MNIST handwritten digit classification. Finally, the SHO device with the modified pulse scheme is evaluated for the classification of handwritten digits of the Modified National Institute of Standards and Technology (MNIST) handwritten database⁵⁶. The database has 60,000 training images and 10,000 test images for the digits 0 to 9. Each image is 28×28 pixels in size and is displayed in grayscale, with pixel intensities ranging from 0 to 255. Figure 6a depicts the hypothetical model network, which contains an input layer, the SHO layer, and a classifier layer. The workflow of the network is assumed to have the following layer operations: the images are preprocessed in the input layer so that the original grayscale format is binarized with a threshold (pixel intensity $> 125 = 1$), where 1 stands for a white pixel and 0 for a black pixel. Following this, the images are divided into 4 pixel segments that move along rows and are then converted into input current pulses, creating a total of 196, 4 - $\text{bi}(t)$ input patterns. Each of the 4 - $\text{bi}(t)$ inputs are encoded as a pulse stream and fed to the SHO layer using the modified pulse scheme with IP_1 parameters. As was discussed in the previous section, the filtered FFT amplitude is collected as the output of the SHO layer for 4 - $\text{bi}(t)$ pattern at 9.0 GHz. For the results shown here, custom functions were created in Matlab and Python programming languages to replace 4 pixel values

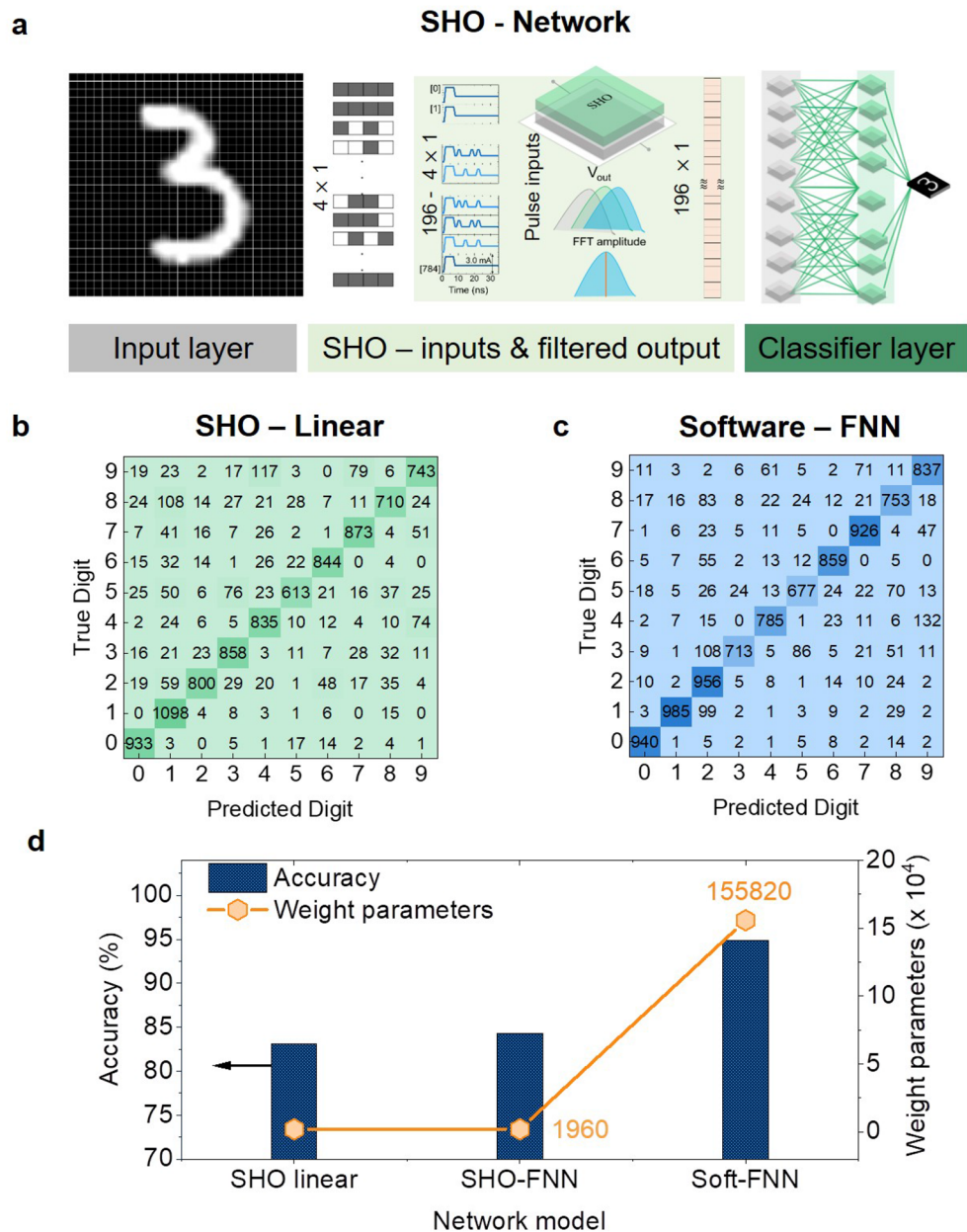


Figure 6. (a) Fully connected network model for classification of MNIST handwritten digit datasets with input layer, SHO layer and a classifier layer with 10 neurons for 10 digits. (b) Color map shows the confusion matrix classification results for 10,000 test images classified by SHO for a linear regression activation (SHO linear). (c) Software implemented feed forward neural network model's, (software-FNN) classification results color map. (d) Comparison of network model implemented with SHO and software as processing unit for filter operations along with classification accuracies (bar graph) and weight parameters (scatter-line plots).

with corresponding FFT amplitudes. The classifier layer has 10 nodes that are all fully connected and are used to categorize the 10 handwritten digits from the maximum entry.

The MNIST handwritten digit classification accuracy is evaluated as the ratio of the total number of correctly classified digits to the total number of digits. First, using simple linear regression the weight matrix was calculated using the training data and tested on the test set of 10,000 images. The accuracy obtained by this one-step calculation is 83.1%. The predicted vs. true digit confusion matrix, shown in Fig. 6b as a color map, displays the classification success for the IP₁ modified pulse scheme. The digit 5 is the least successful at being classified, which lowers the success rate as a whole. In addition, the SHO network model was evaluated in a supervised learning process carried out in Python using the Tensorflow machine learning module⁵⁷. We fixed the softmax activation function for the classifier layer and categorical cross entropy as loss function. The classification accuracy achieved with supervised learning with 20 epochs and 32 batch sizes was 86.6%. This indicates the overall

classification accuracy can be improved with classifier layer training. However, our aim is to reduce the previous layer computations for feature extraction.

To illustrate the reduced computations, we performed network training comparing with a software FNN. The software-FNN model consisted of 784 input neurons for each input pixel, a fully connected middle layer of 196 neurons for feature extraction, and ten neurons (the classifier layer) for classification. We used the rectified linear activation function (ReLU) for the feature extraction layers and the softmax activation function for the classifier layer. It is clear that the software-FNN requires 196 neurons and associated computations to generate 196 features for the classifier layer, similar to the SHO layer outputs. The confusion matrix of the software-FNN, shown in Fig. 6c, achieved an accuracy of 93.1%. When compared to the SHO network, the software-FNN improves accuracy by 7.5%. However, the computational cost of calculating the weight matrix is high. Figure 6d shows the accuracy and weight parameters required for the classification for the linear SHO (single linear weight computation), SHO-FNN (supervised learning), and software-FNN. It can be seen that with the use of the SHO layer, the weight computations required for a software network can be significantly reduced from 155,820 weight parameters to 1960 parameters. The ability of the SHO layer to directly infer distinct outputs for 4 – bi(t) patterns reduces the computation required. Although there is a trade-off between recognition accuracy and inference computations for the MNIST handwritten digit classification due to various factors like binarization of images, choice of classifier layer training methods, and loss functions, the SHO network is still favorable since it can achieve 83.1% with linear regression. Moreover, a sequence of 4 binary digits can be easily classified without any requirement of weight storage or computations, which can be beneficial for applications that need faster inference with reduced computations. Nevertheless, there is a huge opportunity to engage in the co-design of SHO as a dedicated feature mapping layer alongside already existing hardware and algorithms.

Conclusion

In this study, a single spin Hall oscillator capable of real-time classification of sixteen different binary 4 – bi(t) patterns (0000 to 1111) was demonstrated. Control and tuning of the intrinsic magnetization oscillations of the spin Hall oscillator were performed by modifying an input digit pulse pattern. The performance of the model was tested with the standard MNIST handwritten digit data set classification and achieved an accuracy of 83.1% with a linear training network and reduced output layer computations. Hence, these results are expected to provide an insight into the controlling and tuning of spintronic oscillators for real time and on-device neuromorphic computations.

Methods

Micromagnetic simulations. The micromagnetic simulations were carried out using the LLG micromagnetic simulator in the framework of a single domain model^{58–60}. The temporal dynamics of the ferromagnetic layer was solved by the Landau-Lifshitz-Gilbert (LLG) equation with the spin transfer torque term,

$$\frac{d\hat{\mathbf{m}}}{dt} = -\gamma\hat{\mathbf{m}} \times \mu_0\mathbf{H}_{\text{eff}} + \alpha\hat{\mathbf{m}} \times \frac{d\hat{\mathbf{m}}}{dt} - \gamma \frac{\hbar}{2|e|} \frac{\theta_{\text{SH}}|j_c|}{M_s t_{\text{FM}}} (\hat{\mathbf{m}} \times (\hat{\mathbf{m}} \times \hat{\boldsymbol{\sigma}})), \quad (2)$$

where $\hat{\mathbf{m}} = \frac{\mathbf{M}}{M_s}$ is the normalized magnetization vector, γ is the gyromagnetic ratio, α is the Gilbert damping parameter, μ_0 is the vacuum permeability, M_s is the saturation magnetization, \hbar is the reduced Planck constant, e is the electron charge, and t_{FM} is the thickness of the magnetic layer. The effective field \mathbf{H}_{eff} includes the external magnetic field, the magneto-crystalline anisotropy field, and the demagnetization field. θ_{SH} is the spin Hall angle, which characterizes the conversion efficiency of charge current density $\hat{\mathbf{j}}_c$ to spin current density $\hat{\mathbf{j}}_s$ in the heavy metal layer. $\hat{\boldsymbol{\sigma}} = -\text{sgn}\theta_{\text{SH}}(\hat{\mathbf{z}} \times \hat{\mathbf{j}}_c)$ is the orientation of spin injected into the ferromagnet, where $\hat{\mathbf{z}}$ and $\hat{\mathbf{j}}_c$ are the unit vectors in the direction of surface normal and the electrical current, respectively. The $-\text{sgn}$ factor changes with the position of ferromagnet, i.e., if the ferromagnet is atop the HM, $\hat{\mathbf{z}}$ would face into the HM, or if the ferromagnet is below the HM, $\hat{\mathbf{z}}$ would face into the ferromagnet but its sign would be opposite to the prior case. In accordance with experiments, the material parameters used in the simulations^{38,39}, are: $\mu_0 M_s = 1.0\text{T}$, $\alpha = 0.02$, $\theta_{\text{SH}} = 0.07$, resistivities $\rho_{\text{NiFe}} = 4.5 \times 10^{-7}\Omega\text{m}$ and $\rho_{\text{Pt}} = 2.0 \times 10^{-7}\Omega\text{m}$. The magnetic anisotropy is ignored, and an in-plane external magnetic field of strength $\mu_0 H_{\text{ext}} = 100\text{mT}$ is applied to saturate the magnetization along the +Y direction.

Separability index. The separability index (SI) were calculated between the N different outputs (a_1, \dots, a_n) of the spin Hall oscillator (filtered FFT amplitudes), corresponding to each of the N different input patterns via,

$$\text{SI} = \sum_{i=1}^N \sum_{j=1}^N \frac{a_i - a_j}{N^2}, \quad (3)$$

where i and j are the indices of summation and $N = 16$ (16 patterns). The SI value of a particular input current scheme indicates how well the different input patterns can be distinguished from the closeness of the output values. Thus, we can infer that the input current scheme having the highest SI value is the most suited for performing classification and recognition tasks.

Network training and testing. *Linear regression.* For the SHO based network training, all of the images in the MNIST dataset were binarized first, and with custom functions another dataset replacing the 4×1 or 2×2 pixels with FFT amplitudes was created. For the direct software based networks implementation, we used the inbuilt Matlab/Python functions for converting black and white images within the respective classification pro-

grams. For the linear regression readout layer, we used linear regression function executed in Matlab software. During the learning process, for a particular image, the output amplitudes were mapped into a one-dimensional feature vector of 196 elements. This process was repeated for all 60,000 training images, creating a $60,000 \times 196$ output matrix \mathbf{O} . Then, a 196×10 weight matrix \mathbf{W} (where each column corresponds to each digit from 0 to 9) was calculated using the output matrix \mathbf{O} and a label matrix \mathbf{L} ($60,000 \times 10$) containing the true labels for each training image. In each row of the label matrix, the $(l + 1)$ th column had a value of 1 and the remaining columns had a value of 0. l is the true digit ($l = 0, 1, \dots, 9$).

In linear regression, assuming a linear relationship between the output matrix \mathbf{O} and the label matrix \mathbf{L} ,

$$\mathbf{OW} = \mathbf{L}, \quad (4)$$

the weight matrix \mathbf{W} was calculated using pseudo inverse \mathbf{W} ,

$$\mathbf{W} = \mathbf{O}^\dagger \mathbf{L}. \quad (5)$$

Nonlinear activation functions. The SHO and software-FNN network models were evaluated in a supervised learning process performed in Python using the TensorFlow. All the networks were trained for 20 epochs for the weight optimization. The categorical cross-entropy error is reduced using the ADAM optimizer with default learning rate and a batch size of 32. The activation functions ReLU and softmax for the input (z) are given by,

$$\text{ReLU}(z) = \max(0, z), \quad (6)$$

$$\text{softmax}(z_i) = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}}. \quad (7)$$

The softmax z_i are the results of previous layer and K is the number of outputs, which corresponds to the number of classes to be distinguished ($K = 10$ for MNIST handwritten digits 0–9).

Data availability

The datasets generated and analysed for this study is available from Kyushu Institute of Technology repository at <http://hdl.handle.net/10228/00008940>. The MNIST dataset is available at <http://yann.lecun.com/exdb/mnist>.

Received: 2 March 2023; Accepted: 9 May 2023

Published online: 16 May 2023

References

- Goertzel, B. Artificial general intelligence: Concept, state of the art, and future prospects. *J. Artif. Gen. Intell.* **5**, 1–48. <https://doi.org/10.2478/jagi-2014-0001> (2014).
- Pei, J. *et al.* Towards artificial general intelligence with hybrid Tianjic chip architecture. *Nature* **572**, 106–111. <https://doi.org/10.1038/s41586-019-1424-8> (2019).
- Bishop, C. M. *et al.* *Neural Networks for Pattern Recognition* (Oxford University Press, 1995).
- LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444. <https://doi.org/10.1038/nature14539> (2015).
- Farabet, C., Couprie, C., Najman, L. & LeCun, Y. Learning hierarchical features for scene labeling. *IEEE Trans. Pattern Anal. Mach. Intell.* **35**, 1915–1929. <https://doi.org/10.1109/TPAMI.2012.231> (2013).
- Ripley, B. D. *Pattern Recognition and Neural Networks* (Cambridge University Press, 1996).
- Blum, A. L. & Langley, P. Selection of relevant features and examples in machine learning. *Artif. Intell.* **97**, 245–271. [https://doi.org/10.1016/s0004-3702\(97\)00063-5](https://doi.org/10.1016/s0004-3702(97)00063-5) (1997).
- Das, S. Filters, wrappers and a boosting-based hybrid for feature selection. In *Proc. Eighteenth International Conference on Machine Learning* vol. 1, 74–81 (Morgan Kaufmann Publishers Inc, 2001).
- Sebastian, A., Gallo, M. L., Khaddam-Aljameh, R. & Eleftheriou, E. Memory devices and applications for in-memory computing. *Nat. Nanotechnol.* **15**, 529–544. <https://doi.org/10.1038/s41565-020-0655-z> (2020).
- Zhang, W. *et al.* Neuro-inspired computing chips. *Nat. Electron.* **3**, 371–382. <https://doi.org/10.1038/s41928-020-0435-7> (2020).
- Zhou, J. & Chen, J. Prospect of spintronics in neuromorphic computing. *Adv. Electron. Mater.* **7**, 2100465. <https://doi.org/10.1002/aelm.202100465> (2021).
- Sangwan, V. K. & Hersam, M. C. Neuromorphic nanoelectronic materials. *Nat. Nanotechnol.* **15**, 517–528. <https://doi.org/10.1038/s41565-020-0647-z> (2020).
- Grollier, J. *et al.* Neuromorphic spintronics. *Nat. Electron.* **3**, 360–370. <https://doi.org/10.1038/s41928-019-0360-9> (2020).
- Schuman, C. D. *et al.* Opportunities for neuromorphic computing algorithms and applications. *Nat. Comput. Sci.* **2**, 10–19. <https://doi.org/10.1038/s43588-021-00184-y> (2022).
- Wright, L. G. *et al.* Deep physical neural networks trained with backpropagation. *Nature* **601**, 549–555. <https://doi.org/10.1038/s41586-021-04223-6> (2022).
- Ielmini, D. & Ambrogio, S. Emerging neuromorphic devices. *Nanotechnology* **31**, 92001. <https://doi.org/10.1088/1361-6528/ab554b> (2019).
- Choi, S., Yang, J. & Wang, G. Emerging memristive artificial synapses and neurons for energy-efficient neuromorphic computing. *Adv. Mater.* **32**, 2004659. <https://doi.org/10.1002/adma.202004659> (2020).
- Chicca, E., Stefanini, F., Bartolozzi, C. & Indiveri, G. Neuromorphic electronic circuits for building autonomous cognitive systems. *Proc. IEEE* **102**, 1367–1388. <https://doi.org/10.1109/JPROC.2014.2313954> (2014).
- Gembaczka, P. *et al.* Combination of sensor-embedded and secure server-distributed artificial intelligence for healthcare applications. *Curr. Dir. Biomed. Eng.* **5**, 29–32. <https://doi.org/10.1515/cdbme-2019-0008> (2019).
- Boriskov, P. IoT-oriented design of an associative memory based on impulsive hopfield neural network with rate coding of LIF oscillators. *Electronics* **9**, 1468. <https://doi.org/10.3390/electronics9091468> (2020).
- Patterson, D. *et al.* Carbon emissions and large neural network training. <https://doi.org/10.48550/ARXIV.2104.10350> (2021).
- Christensen, D. V. *et al.* 2022 roadmap on neuromorphic computing and engineering. *Neuromorph. Comput. Eng.* **2**, 022501. <https://doi.org/10.1088/2634-4386/ac4a83> (2022).

23. Slonczewski, J. C. Current-driven excitation of magnetic multilayers. *J. Magn. Magn. Mater.* **159**, L1–L7. [https://doi.org/10.1016/0304-8853\(96\)00062-5](https://doi.org/10.1016/0304-8853(96)00062-5) (1996).
24. Kiselev, S. I. *et al.* Microwave oscillations of a nanomagnet driven by a spin-polarized current. *Nature* **425**, 380–383. <https://doi.org/10.1038/nature01967> (2003).
25. Hirsch, J. E. Spin Hall effect. *Phys. Rev. Lett.* **83**, 1834–1837. <https://doi.org/10.1103/physrevlett.83.1834> (1999).
26. Sinova, J., Valenzuela, S. O., Wunderlich, J., Back, C. H. & Jungwirth, T. Spin Hall effects. *Rev. Mod. Phys.* **87**, 1213–1260. <https://doi.org/10.1103/RevModPhys.87.1213> (2015).
27. Miron, I. M. *et al.* Current-driven spin torque induced by the Rashba effect in a ferromagnetic metal layer. *Nat. Mater.* **9**, 230–234. <https://doi.org/10.1038/nmat2613> (2010).
28. Mellnik, A. R. *et al.* Spin-transfer torque generated by a topological insulator. *Nature* **511**, 449–451. <https://doi.org/10.1038/nature13534> (2014).
29. Demidov, V. E. *et al.* Magnetization oscillations and waves driven by pure spin currents. *Phys. Rep.* **673**, 1–31. <https://doi.org/10.1016/j.physrep.2017.01.001> (2017).
30. Chen, T. *et al.* Spin-torque and spin-hall nano-oscillators. *Proc. IEEE* **104**, 1919–1945. <https://doi.org/10.1109/JPROC.2016.2554518> (2016).
31. Duan, Z. *et al.* Nanowire spin torque oscillator driven by spin orbit torques. *Nat. Commun.* **5**, 6616. <https://doi.org/10.1038/ncomms6616> (2014).
32. Duan, Z. *et al.* Spin-wave modes in permalloy/platinum wires and tuning of the mode damping by spin Hall current. *Phys. Rev. B* **90**, 024427. <https://doi.org/10.1103/physrevb.90.024427> (2014).
33. Demidov, V. E. *et al.* Control of magnetic fluctuations by spin current. *Phys. Rev. Lett.* **107**, 107204 (2011), <https://doi.org/10.1103/physrevlett.107.107204>.
34. Demidov, V. E. *et al.* Magnetic nano-oscillator driven by pure spin current. *Nat. Mater.* **11**, 1028–1031. <https://doi.org/10.1038/nmat3459> (2012).
35. Fullara, H. *et al.* Spin-orbit torque-driven propagating spin waves. *Sci. Adv.* **5**, 8467. <https://doi.org/10.1126/sciadv.aax8467> (2019).
36. Dvornik, M., Awad, A. A. & Åkerman, J. Origin of Magnetization Auto-Oscillations in Constriction-Based Spin Hall Nano-Oscillators. *Phys. Rev. Appl.* **9**, 014017. <https://doi.org/10.1103/physrevapplied.9.014017> (2018).
37. Hoffmann, A. Spin Hall Effects in Metals. *IEEE Trans. Magn.* **49**, 5172–5193. <https://doi.org/10.1109/TMAG.2013.2262947> (2013).
38. Shashank, U. *et al.* Enhanced spin Hall effect in S-implanted Pt. *Adv. Quantum Technol.* **4**, 2000112. <https://doi.org/10.1002/qute.202000112> (2020).
39. Shashank, U. *et al.* Highly dose dependent damping-like spin-orbit torque efficiency in O-implanted Pt. *Appl. Phys. Lett.* **118**, 252406. <https://doi.org/10.1063/5.0054779> (2021).
40. Kim, K.-W., Lee, K.-J., Sinova, J., Lee, H.-W. & Stiles, M. D. Spin-orbit torques from interfacial spin-orbit coupling for various interfaces. *Phys. Rev. B* **96**, 104438. <https://doi.org/10.1103/physrevb.96.104438> (2017).
41. Ryu, J., Lee, S., Lee, K.-J. & Park, B.-G. Current-induced spin-orbit torques for spintronic applications. *Adv. Mater.* **32**, 1907148. <https://doi.org/10.1002/adma.201907148> (2020).
42. Nan, T. *et al.* Comparison of spin-orbit torques and spin pumping across NiFe/Pt and NiFe/Cu/Pt interfaces. *Phys. Rev. B* **91**, 214416. <https://doi.org/10.1103/PhysRevB.91.214416> (2015).
43. Liu, R. H., Lim, W. L. & Urazhdin, S. Spectral characteristics of the microwave emission by the spin Hall nano-oscillator. *Phys. Rev. Lett.* **110**, 147601. <https://doi.org/10.1103/physrevlett.110.147601> (2013).
44. Berkov, D. V. & Gorn, N. L. Magnetization precession due to a spin-polarized current in a thin nanoelement: Numerical simulation study. *Phys. Rev. B* **72**, 094401. <https://doi.org/10.1103/physrevb.72.094401> (2005).
45. Xiao, J., Zangwill, A. & Stiles, M. D. Macrospin models of spin transfer dynamics. *Phys. Rev. B* **72**, 014446. <https://doi.org/10.1103/physrevb.72.014446> (2005).
46. Slavini, A. N. & Kabos, P. Approximate theory of microwave generation in a current-driven magnetic nanocontact magnetized in an arbitrary direction. *IEEE Trans. Magn.* **41**, 1264–1273. <https://doi.org/10.1109/TMAG.2005.845915> (2005).
47. Kim, J.-V. Spin-torque oscillators. In *Solid State Physics* (ed. Kim, J.-V.) 217–294 (Elsevier, 2012).
48. Slavin, A. & Tiberkevich, V. Nonlinear auto-oscillator theory of microwave generation by spin-polarized current. *IEEE Trans. Magn.* **45**, 1875–1918. <https://doi.org/10.1109/TMAG.2008.2009935> (2009).
49. Holzmann, G. & Hauser, H. Echo state networks with filter neurons and a delay & sum readout. *Neural Netw.* **23**, 244–256. <https://doi.org/10.1016/j.neunet.2009.07.004> (2010).
50. Wyffels, F., Schrauwen, B., Verstraeten, D. & Stroobandt, D. Band-pass reservoir computing. In *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)* 3204–3209. <https://doi.org/10.1109/IJCNN.2008.4634252> (IEEE, 2008).
51. Hubara, I., Courbariaux, M., Soudry, D., El-Yaniv, R. & Bengio, Y. Quantized neural networks: Training neural networks with low precision weights and activations. *J. Mach. Learn. Res.* **18**, 1–30 (2018).
52. Pham, P., Abraham, J. A. & Chung, J. Training multi-bit quantized and binarized networks with a learnable symmetric quantizer. *IEEE Access* **9**, 47194–47203. <https://doi.org/10.1109/ACCESS.2021.3067889> (2021).
53. Akashi, N. *et al.* Input-driven bifurcations and information processing capacity in spintronic reservoirs. *Phys. Rev. Res.* **2**, 043303. <https://doi.org/10.1103/physrevresearch.2.043303> (2020).
54. Krishnan, S., Samudravijaya, K. & Rao, P. V. S. Feature selection for pattern classification with Gaussian mixture models: A new objective criterion. *Pattern Recognit. Lett.* **17**, 803–809. [https://doi.org/10.1016/0167-8655\(96\)00047-5](https://doi.org/10.1016/0167-8655(96)00047-5) (1996).
55. Gibbons, T. E. Unifying quality metrics for reservoir networks. In *The 2010 International Joint Conference on Neural Networks (IJCNN)* 1–7. <https://doi.org/10.1109/IJCNN.2010.5596307> (IEEE, 2010).
56. LeCun, Y., Cortes, C. & Burges, C. J. C. *The MNIST Database of Handwritten Digits*. <http://yann.lecun.com/exdb/mnist/> (Accessed 02 2021)
57. Abadi, M. *et al.* *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. <https://www.tensorflow.org/> (2015).
58. Deka, A., Tanaka, I., Mohan, J. R. & Fukuma, Y. Modulation of magnetization precession trajectories by perpendicular magnetic anisotropy in CoFeB thin films. *IEEE Trans. Magn.* **56**, 1–5. <https://doi.org/10.1109/TMAG.2020.2987288> (2020).
59. Deka, A., Sato, K., Tanaka, I. & Fukuma, Y. Simulations on the effect of magnetic anisotropy on switching of an easy cone magnetized free layer. *IEEE Trans. Magn.* **56**, 1–4. <https://doi.org/10.1109/TMAG.2019.2942339> (2020).
60. Scheinfein, M. R. & Price, E. A. LLG micromagnetic simulator. In *LLG Micromagnetic Simulator* (2015).

Acknowledgements

Financial support from JSPS Grant-in-Aid (KAKENHI No. 18H01862 and 19K21112) is gratefully acknowledged. This work is partly executed in response to support KIOXIA Corporation. R.M., S.G. and R.S.R. would like to acknowledge research Grants NRF-CRP21-2018-0003 from National Research Foundation Singapore and MOE2019-T2-1-058 from Ministry of Education Singapore.

Author contributions

Y.F. conceptualized and supervised the study. J.R.M., A.J.M., K.N. and R.F. did the micromagnetic simulations and classification tasks. J.R.M., R.M. and S.G. analysed data and wrote the manuscript with the help of Y.F. and R.S.R. All the authors discussed the results and commented to improve the quality of manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-34849-7>.

Correspondence and requests for materials should be addressed to Y.F.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023