Neuromorphic computing using multilevel nonvolatile memories as synapses offers opportunities for future energy- and area-efficient artificial intelligence. Among these memories, artificial synapses based on current-induced magnetization switching driven by spin–orbit torques (SOTs) have attracted great attention recently. Herein, the gradient descent algorithm, a primary learning algorithm, implemented on a $2 \times 1$ SOT synaptic array is reported. Successful pattern classifications are experimentally realized through the tuning of cycle-to-cycle variation, linearity range, and linearity deviation of the multilevel SOT synapse. Also, a larger $m \times n$ SOT synaptic array with $m$ controlling transistors is proposed and it is found that the classification accuracies can be improved dramatically by decreasing the cycle-to-cycle variation. A way for the application of spin–orbit device arrays in neuromorphic computing is paved and the crucial importance of the cycle-to-cycle variation for a multilevel SOT synapse is suggested.

Inspired by the highly efficient human brain, where the information is stored in plastic synapses, neuromorphic computing architectures using nonvolatile memories\cite{1–3} as artificial synapses are expected to achieve ultralow power intelligent systems.\cite{4–7} Spin–orbit torque (SOT) devices\cite{8,9} with inherent nonvolatility and radiation resistance, subnanosecond switching dynamics, unlimited endurance, excellent stability, and verified complementary metal-oxide-semiconductor transistor (CMOS)-compatibility can also be implemented as the multilevel current-induced magnetization switching, which can be used to emulate the synaptic plasticity.\cite{10–18} Beyond the device level, artificial and spiking neural networks\cite{19,20} consisting of spin–orbit synapses have been applied for various purposes such as associative memories,\cite{21} deep belief learnings,\cite{22} on-the-fly learnings,\cite{23} etc. According to the studies on other emerging nonvolatile memories, especially the resistive switching random access memory (RRAM) and the phase change memory (PCM),\cite{24,25} the variation and linearity of the device are crucial parameters for neuromorphic applications.\cite{1,26–28} However, few studies have been conducted on these crucial parameters of multilevel spin–orbit devices, the study of which could be important for developing the neural networks based on SOT synaptic arrays.

Recently, the tunable multilevel SOT-induced out-of-plane magnetization switching was observed by adjusting the magnitude of an in-plane magnetic field\cite{10} which provides a powerful way to investigate the dependence of device properties on synaptic training behaviors. In this study, we demonstrate prototype pattern classifications based on tunable multilevel SOT synaptic arrays using the basic gradient descent algorithm, where an iterative algorithm optimizes the result by seeking the minimum value of a differentiable function (i.e., the cost function, whose gradient determines target values of updated synaptic weights). First, various indicators of the multilevel SOT synapse, including the cycle-to-cycle variation (the variation of the updated states between different operating cycles for identical
The linearity range, and the linearity deviation, were characterized as a function of the external in-plane magnetic field $H_x$. Then, the gradient descent algorithm was experimentally carried out on a $2 \times 1$ SOT synaptic array under varying $H_x$. The training results were analyzed using the aforementioned multilevel indicators. Finally, an extended $m \times n$ SOT synaptic array with $m$ current-controlled transistors are proposed, and the classification accuracy dependence on the cycle-to-cycle variation is simulated.

As schematically shown in Figure 1a, stacks of Si/SiO$_2$ substrate/Pt(3 nm)/Co(0.8 nm)/AlO$_x$(2 nm) were patterned into Hall bars with a cross region of $20 \times 20 \mu$m$^2$. In the presence of an external magnetic field $H_x$, current pulses $I_x$-induced out-of-plane magnetization ($m_z$) switching was characterized by detecting the anomalous Hall resistance ($R_{H}$) between each pulse with a measuring current of 100 $\mu$A. Consistent with the result of our previous work, multilevel SOT-induced $m_z$ switching and reductions of the maximum current-induced Hall resistances ($\pm R_{max}$) are obtained with increasing $H_x$, as shown in Figure 1b. The multilevel magnetization is attributed to the magnetic field-induced pinning potential difference, which causes the gradually enlarged reversed domain area as the current−pulse intensity increases.$^{[10]}$ To characterize cycle-to-cycle variations of the multilevel SOT device quantitatively, repeated experiments of single pulse-induced magnetization switching from their $-R_{max}$ states were investigated under various $H_x$. The measured $R_{H}$ in the linearity ranges of single pulse-induced switching at different $H_x$ are shown in Figure 1c, where the minimum and maximum Hall resistances of a fitting line are denoted as $R_a$ and $R_b$, respectively. In this article, synaptic trainings are conducted using such linear fitting lines instead of the original measured $R_1- I_x$ data. Therefore, four variables are introduced to describe the variation and linearity deviation of a SOT synapse. Primarily, the first two variables are the averaged absolute switching variation $\Delta R_{var}$ and the averaged absolute linearity deviation $\Delta R_{dev}$, where the first one describes the average height of error bars, and the second one describes the average maximum absolute difference between the measured value and the fitting line. Then, the averaged relative variation $\bar{\eta}_{var}$ and the averaged relative linearity deviation $\bar{\eta}_{dev}$ can be obtained as

$$\bar{\eta}_{var} = \frac{\Delta R_{var}}{2R_{max}} \quad (1)$$

and

$$\bar{\eta}_{dev} = \frac{\Delta R_{dev}}{R_b - R_a} \quad (2)$$

where $\bar{\eta}_{var}$ is the relative variation for the total current-induced switching range ($2R_{max}$), and $\bar{\eta}_{dev}$ is the relative linearity deviation for the linearity range ($R_b - R_a$).

![Figure 1](image-url)

Figure 1. Optimizing the linearity deviation of the multilevel SOT synapse by adjusting $H_x$. a) Schematic diagram of the SOT Hall bar device. b) Current pulse-induced $R_x-I_x$ loops at different external in-plane magnetic fields $H_x$. The duration of each pulse is 60 ms, and the anomalous Hall resistance $R_a$ was obtained by detecting the Hall voltage $V_H$ in y-direction with a low measuring current of 100 $\mu$A in x-direction after 2 s of each pulse. c) Resulted $R_a$, after various single positive current pulse $I_x$ under different $H_x$. Before the application of each single current pulse, $R_a$ is reset to a maximum current-induced negative value ($-R_{max}$) by applying a current pulse of $-12$ mA. Error bars are derived from ten repeated experimental trials, and each dashed line is the linear fitting of the middle values of all error bars of the same color. The starting and ending points of fitting lines are denoted as $R_a$ and $R_b$, respectively. d) Average absolute switching variation $\Delta R_{abs}$ and maximum current-induced anomalous Hall resistance change $2R_{max}$. e) Average absolute linearity deviation $\Delta R_{dev}$ and linearity range of anomalous Hall resistance ($R_b - R_a$). f) Average relative variation $\bar{\eta}_{var}$ and average relative linearity deviation $\bar{\eta}_{dev}$ under various $H_x$. 

© 2021 The Authors. Advanced Intelligent Systems published by Wiley-VCH GmbH
As shown in Figure 1d, both $2R_{\text{max}}$ and $|\Delta R_{\text{var}}|$ decrease with the increase in $H_x$. The decrease in $2R_{\text{max}}$ can be ascribed to the gradually tilted $m$. Meanwhile, the underlying mechanism of the decrease in $|\Delta R_{\text{var}}|$ are not clear at this stage, which can be attributed to the growing pinning effect. This pining effect may stabilize the current-driven domain wall motions and thereby enhance the cycle-to-cycle reproducibility. Unlike the monotonous decrease in $2R_{\text{max}}$, the $R_b-R_a$ as a function of $H_x$ initially exhibits an ascending trend for $H_x<2500$ Oe, followed by a descending trend when $H_x$ exceeds 2900 Oe, as shown in Figure 1e. The enlargement of $R_b-R_a$ from 2100 to 2500 Oe is consistent with the improved multilevel magnetization switching. Meanwhile, $|\Delta R_{\text{dev}}|$ decreases as $H_x$ increases in most cases, except for $H_x$ at 2500 and 2900 Oe. Due to the misalignment between the fitting line and measured datapoints, $|\Delta R_{\text{var}}|$ is unexpectedly large at 2500 and 2900 Oe, as shown in Note 1, Supporting Information. As a result, both evolutions of $\pi_{\max}$ and $\pi_{\text{var}}$ in Figure 1f show “U” shape trends when $H_x$ increases from 2100 to 3500 Oe, with bulges in $\pi_{\text{dev}}^{\max}$ and $\pi_{\text{dev}}$. The “U” shape trends enable us to tune the variation/deviation and optimize training results by controlling $H_x$, which will be discussed later.

As schematically shown in Figure 2a, a $2 \times 1$ synaptic array comprising two such tunable multilevel SOT devices was applied to carry out pattern classification using gradient descent algorithm, where the training and testing pattern samples ($X_1, X_2$) are shown in Figure 2b. The array output $Z$, the logistic function $Y$ (the activation function which controls the output based on $Z$ and mimics the step function output of human neurons using differentiable sigmoid function) as well as the cost function $J$ (the error between the label value and the actual output value) of pattern samples are obtained by measuring the $R_1$-related synaptic weights $W_i$ of the two SOT devices. Their relations can be described as the following equations.

$$Z = W_A X_1 + W_B X_2 + b$$

(3)

$$Y = \frac{1}{1 + e^{-Z}}$$

(4)

$$J = -y \ln Y - (1-y) \ln(1-Y)$$

(5)

where $b$ is a bias constant and $y$ is the label value of the pattern (integral number, ranges from 0 to 9). Then, the desired changes in $W_i$ are calculated in a gradient descent way as

$$\Delta W_i = -\alpha \frac{\partial J(W_i)}{\partial W_i}$$

(6)

where $\alpha$ is the learning rate. Finally, values of $\Delta W_i$ determine the direction and intensity of current pulses $I_i$ on the basis of fitting lines in Figure 1c. Note that the updated weight (represented by Hall resistance) also depends on the stochastic deviation $\Delta R_{\text{dev}}$, as discussed in Note 2, Supporting Information. Aforementioned training procedures are shown in Figure 2c. In brief, a successful pattern classification requires a lowest possible $J(-0)$, as well as significant differences between $Y$ values for the training ($Y \rightarrow 1$) and testing ($Y \rightarrow 0$) pattern samples.

According to this principle, training results for $\alpha = 0.5$ under various $H_z$, as shown in Figure 2d–g, can be qualitatively interpreted by $H_z$-dependent variations, linearity ranges, and linearity deviations shown in Figure 1d–f as followings: First, the cycle-to-cycle variation is very important to the gradient descent training, as trainings have failed for $\pi_{\text{var}} > 5.5\%$ at 2100 and 2900 Oe; Second, the linearity range is also an important factor, as the successful training at 2500 Oe shows similar $\pi_{\max}$ and $\pi_{\text{var}}$ ($\approx 5\%$) but nearly tripled $R_b-R_a$ ($\approx 1-2\,\Omega$) with the failed case at 3500 Oe. The effect of linearity range is further confirmed by the case at 3300 Oe. Although the synaptic device at 3300 Oe has small $\pi_{\max}$ and $\pi_{\text{var}}$ ($\approx 3.5\%$), it also has the slowest training process and the weakest identifiability (see cyan triangles in Figure 2d–g) among the successful cases from 2500 to 3300 Oe as the small $R_b-R_a$ ($\approx 0.7\,\Omega$); Third, for 2700 and 2900 Oe with both small $\pi_{\text{var}}$ ($<3.5\%$) and large $R_b-R_a$ (1.1–1.2\,Ω), the best training result is obtained only at 2700 Oe due to its smaller $\pi_{\max}$ ($\approx 4\%$), comparing with that of 2900 Oe ($>5\%$). Corresponding evolutions of $W_i$ and $R_{\max}$ in each SOT synapse are shown in Note 2, Supporting Information, and other two training results for larger learning rates of $\alpha = 1$ and $\alpha = 2$ are shown in Note 3, Supporting Information, where identical trends versus $H_x$ are observed. As a general understanding, we attribute the strong correlation between larger linearity ranges and better training results to the adjustable synaptic weights.

Furthermore, we have proposed a $m \times n$ SOT synaptic array for recognizing larger patterns. As shown in Figure 3a, $n$ columns of SOT synapses are parallelly connected and $m$ transistors are designed to switch the input current pulse $I_{\text{ext}}$ for certain synapse. For practical applications, $H_x$ can be replaced using SOT devices with optimized build-in plane magnetic fields, which can be achieved using either an static in-plane magnetic field or a “T-type” interlayer exchange coupling between an in-plane and an out-of-plane magnetized layer. Other magnetic field-free SOT switching strategies include the wedged layers, the asymmetric spin currents, the competing spin currents, etc. all of which have potential for the field-free multilevel magnetization switching in the future. The feasibility of this proposal is primitively verified by measuring the gate-voltage-controlled multilevel magnetization switching of a SOT device driven by a thin-film transistor (TFT). Moreover, the variation-dependence of gradient descent algorithm carried out on the proposed SOT synaptic array was investigated by examining the simulated classification accuracy of a one-hidden-layer backpropagation neural network, as shown in Figure 3b. The hidden and output layers export the same logistic functions as Equation (3), and synaptic weights are updated using a similar gradient descent method in Equation (6) with stochastic deviations of $\Delta W_i$ ranging from $-\pi_{\text{var}} \times W_{\text{range}}/2$ to $+\pi_{\text{var}} \times W_{\text{range}}/2$. Note that the $\pi_{\text{var}}$ here works equally with the $\Delta R_{\text{dev}}$, because we assume that those devices have the ideal linearity. The classification accuracy is defined as the percentage of successfully recognized testing pattern samples in all testing samples. The success of a recognition can be judged by comparing the label value of each sample with the number of the corresponding highest output $Y^{\text{out}}$ in Figure 3b. In this article, pattern classifications of two typical datasets, i.e., a small image version (8 × 8 pixels) of handwritten digits from the “Optical Recognition of Handwritten Digits” (ORHD) dataset and a large image version (28 × 28 pixels) of handwritten digits from the “Modified National Institute of Standards and Technology”
The (MNIST) dataset are simulated under various η var, as shown in Figure 3c,d respectively. More details of the simulations can be found in Note 4, Supporting Information. Obviously, the classification accuracies of both datasets drop substantially from 95% to 80% (ORHD ≈90%; MINST ≈80%) when the variation increases from 0% to 2%, and then even reduce to 50% when the variation reaches 8%, as shown in Figure 3e, suggesting the great importance of cycle-to-cycle variation for large-scale applications.

Figure 2. Gradient descent carrying out on a 2 × 1 SOT synaptic array. a,b) Schematic drawing of a) the 2 × 1 SOT synaptic array and b) the training pattern sample (1, 1) as well as two testing pattern samples (1, −1) and (−1, 1). c) Flow chart of the gradient descent algorithm used for pattern recognition in this work, where the step in the gray box is carried out experimentally on the 2 × 1 SOT synaptic array. d–g) Evolutions of d) cost function J and e–g) logistic function Y of e) the training pattern sample and f,g) two testing pattern samples as functions of epoch number N under various Hx. A learning rate of α = 0.5 is used here. Training results shown in (e–g) can be divided into three grades, i.e., 1) Failed trainings at 2100, 2300, and 3500 Oe, where Y < 0.5 is found for the training pattern sample; 2) Good trainings at 2500, 2900, 3100, and 3300 Oe, where there are relative low J and substantial differences in Y between the training and testing pattern samples; 3) Best training at 2700 Oe, where most efficient evolutions of J and Y with epoch number N are observed.
multilevel SOT synaptic arrays. To improve the variation and linearity of a multilevel SOT device, uniform pinning potential in the domain wall motion process is required. In addition, emerging technologies in skyrmion synapses may provide new strategies, where linear weight update is expected as it depends on the number of accumulated skyrmions.

In conclusion, we have carried out a pattern classification using the gradient descent algorithm on a $2 \times 1$ SOT synaptic array under various multilevel switching conditions. It is shown that training results are improved by decreasing cycle-to-cycle variation and linearity deviation as the linearity range increases. We also proposed a larger $m \times n$ SOT synaptic array with $m$ transistors, and the simulated classification accuracy shows a significant positive correlation with the cycle-to-cycle variation. Our findings are an important step toward the application of spin–orbit synaptic arrays in neuromorphic computing.
Experimental Section

Device Fabrication: First, Hall bars were patterned onto Si(0.5 mm)/SiO₂(190 nm) wafers by lithography. Then, Pt(3 nm)/Co(0.8 nm)/Al₂O₃(2 nm) stacks were deposited into the Hall bar patterns at room temperature by magnetron sputtering. Direct current (DC) sputtering was used to deposit the Pt and the Co layers, while radio-frequency (RF) sputtering was used to deposit the Al₂O₃ layer. The chamber base pressure was less than 2 × 10⁻⁶ Pa, and Ar gas was used for sputtering. The chamber pressures during the DC and the RF sputtering were 1.067 × 10⁻¹ Pa and 2.666 × 10⁻¹ Pa, respectively. No magnetic field was applied during the sputtering. The deposition rates for Pt, Co, and Al₂O₃ were 0.021, 0.013, and 0.0018 nm s⁻¹, respectively.

Device and Array Measurements: All the electrical measurements were carried out at room temperature using a Keithley 2602B as the current source and a Keithley 2182A as the nanovoltmeter. Note that there are two types of currents, i.e., the relatively large pulse currents Iₚ for driving SOT-induced magnetization switching, and the relatively small (100 μA) measuring DC for detecting the anomalous Hall voltage and thereby obtaining the anomalous Hall resistance for each device.

Neural Network Simulations: Details of the neural network simulations can be found in the Note 4, Supporting Information.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

X.L. and Y.C. contributed equally to this work. The authors acknowledge the support from National Key R&D Program of China (Grant No. 2017YFB080450700), National Natural Science Foundation of China (Grant No. 61774144 and 61922090), Chinese Academy of Sciences (Grant No. 61774144 and 61922090), National Natural Science Foundation of China (Grant No. Z190007).

Conflict of Interest

The authors declare no conflict of interest.

Keywords

cycle-to-cycle variations, gradient descents, linearity deviation, linearity ranges, multilevel magnetization switching, pattern classifications, spin–orbit torques

Received: July 30, 2020
Revised: November 25, 2020
Published online: February 8, 2021

[1] D. Ielmini, H. S. P. Wong, Nat. Electron. 2018, 1, 333.
[2] Z. Wang, H. Wu, G. W. Burri, C. S. Hwang, K. L. Wang, Q. Xia, J. J. Yang, Nat. Rev. Mater. 2020, 5, 173.
[3] Z. Wang, S. Joshi, S. E. Savelev, H. Jiang, R. Muida, P. Lin, M. Hu, N. Ge, J. P. Strachan, Z. Li, Q. Wu, M. Barnell, G.-L. Li, H. L. Xin, R. S. Williams, Q. Xia, J. J. Yang, Nat. Mater. 2017, 16, 101.
[4] J. Grollier, D. Querlioz, K. Y. Camsari, K. Everschor-Sitte, S. Fukami, M. D. Stiles, Nat. Electron. 2020, 3, 360.
[5] R. Guo, W. Lin, X. Yan, T. Venkatesan, J. Chen, Appl. Phys. Rev. 2020, 7, 011304.
[6] A. Sengupta, K. Roy, Appl. Phys. Express 2018, 11, 030101.
[7] Y. Zhang, Q. Zheng, X. Zhu, Z. Yuan, K. Xia, Sci. China: Phys., Mech. Astron. 2020, 63, 277531.
[8] I. M. Miran, K. Garelo, G. Gaudin, P.-J. Zermatten, M. V. Costache, S. Auffret, S. Bandiera, B. Rodmacq, A. Schuhl, P. Gambardella, Nature 2011, 476, 189.
[9] L. Liu, C.-F. Pai, Y. Li, H. W. Tseng, D. C. Ralph, R. A. Buhrman, Science 2012, 336, 555.
[10] Y. Cao, A. Rushforth, Y. Sheng, H. Zheng, K. Wang, Adv. Funct. Mater. 2019, 29, 1808104.
[11] S. Fukami, C. Zhang, S. DuttaGupta, A. Kurenkov, H. Ohno, Nat. Mater. 2016, 15, 535.
[12] A. Kurenkov, C. Zhang, S. DuttaGupta, S. Fukami, H. Ohno, Appl. Phys. Lett. 2017, 110, 092410.
[13] A. Kurenkov, S. DuttaGupta, C. Zhang, S. Fukami, Y. Horio, H. Ohno, Adv. Mater. 2019, 31, 1900636.
[14] X. H. Liu, K. W. Edmonds, Z. P. Zhou, K. Y. Wang, Phys. Rev. Appl. 2020, 13, 014059.
[15] J. Yun, Q. Bai, Z. Yan, M. Chang, J. Mao, Y. Zuo, D. Yang, L. Xi, D. Xue, Adv. Funct. Mater. 2020, 30, 1909092.
[16] P. Wadley, B. Howells, J. Železný, C. Andrews, V. Hills, R. P. Campion, V. Novák, K. Olejník, F. Macherczoo, S. S. Dhesi, S. Y. Martin, T. Wagner, J. Wunderlich, F. Freimuth, Y. Mokrousov, J. Kuněš, J. S. Chauhan, M. J. Grzybsowski, A. W. Rushforth, K. W. Edmonds, B. L. Gallagher, T. Jungwirth, Science 2016, 351, 587.
[17] K. Olejník, V. Schuler, X. Marti, V. Novák, Z. Žeďar, P. Wadley, R. P. Campion, K. W. Edmonds, B. L. Gallagher, J. Garces, M. Baumgartner, P. Gambardella, T. Jungwirth, Nat. Commun. 2017, 8, 15434.
[18] A. Sengupta, Z. Al Azim, X. Fong, K. Roy, Appl. Phys. Lett. 2015, 106, 091704.
[19] A. Sengupta, Y. Shim, K. Roy, IEEE Trans. Biomed. Circuits Syst. 2016, 10, 1152.
[20] A. Sengupta, A. Banerjee, K. Roy, Phys. Rev. Appl. 2016, 6, 064003.
[21] W. A. Borders, H. Akima, S. Fukami, S. Moriya, S. Kurihara, Y. Horio, S. Sato, H. Ohno, Appl. Phys. Express 2016, 10, 013007.
[22] V. Ostwal, R. Zand, R. DeMara, J. Appenzeller, IEEE J. Explor. Solid-State Comput. Devices Circuits 2019, 5, 182.
[23] K. Yue, Y. Liu, R. K. Lake, A. C. Parker, Sci. Adv. 2019, 5, eaau8170.
[24] S. Ambrogio, S. Balatti, A. Cubeta, A. Calderoni, N. Rasmawamy, D. Ielmini, IEEE Trans. Electron Devices 2014, 61, 2912.
[25] M. Rizzi, N. Ciocchini, A. Montefori, M. Fero, P. Fantini, A. L. Lacaia, D. Ielmini, IEEE Trans. Electron Devices 2015, 62, 2205.
[26] V. Milo, C. Malavaca, C. Monzio Compagnoni, D. Ielmini, Materials 2020, 13, 166.
[27] K. Moon, M. Kwak, J. Park, D. Lee, H. Hwang, IEEE Electron Device Lett. 2017, 38, 1023.
[28] A. Krishnaprasad, N. Choudhary, S. Das, D. Dev, H. Kalita, H.-S. Chung, O. Aina, Y. Jung, T. Roy, Appl. Phys. Lett. 2019, 115, 103104.
[29] C. Baeumner, R. Valenta, C. Schmitz, A. Locatelli, T. O. Mente, S. P. Rogers, A. Sala, N. Raab, S. Nemsak, M. Shim, ACS Nano 2017, 11, 6921.
[30] K. Garelo, F. Yasim, H. Hody, S. Courat, L. Souriau, S. H. Sharif, J. Swerts, R. Carpenter, S. Rao, W. Kim, J. Wu, K. K. V. Sethu, M. Pak, N. Jossart, D. Crotti, A. Furnémont, G. S. Kar, presented at Symp. on VLSI Circuits, Manufacturable 300mm platform solution for Field-Free Switching SOT-MRAM, Kyoto, Japan, June 2019.
[31] Y. Sheng, K. W. Edmonds, X. Ma, H. Zheng, K. Wang, Adv. Electron. Mater. 2018, 4, 1800224.
[32] X. Wang, C. Wan, W. Kong, X. Zhang, Y. Xing, C. Fang, B. Tao, W. Yang, L. Huang, H. Wu, M. Irfan, X. Han, Adv. Mater. 2018, 30, 1801318.
[33] Z. A. Bekele, X. Lan, K. Meng, X. Liu, J. Appl. Phys. 2020, 127, 113902.
[34] G. Yu, P. Upadhyaya, Y. Fan, J. G. Alzate, W. Jiang, K. L. Wong, S. Takei, S. A. Bender, L.-T. Chang, Y. Jiang, M. Lang, J. Tang, Y. Wang, Y. Tserkovnyak, P. K. Amiri, K. L. Wang, Nat. Nanotechnol. 2014, 9, 548.

[35] K. Cai, M. Yang, H. Ju, S. Wang, Y. Ji, B. Li, K. W. Edmonds, Y. Sheng, B. Zhang, N. Zhang, S. Liu, H. Zheng, K. Wang, Nat. Mater. 2017, 16, 712.

[36] Y. Cao, Y. Sheng, K. W. Edmonds, Y. Ji, H. Zheng, K. Wang, Adv. Mater. 2020, 32, 1907929.

[37] M. Wang, W. Cai, D. Zhu, Z. Wang, J. Kan, Z. Zhao, K. Cao, Z. Wang, Y. Zhang, T. Zhang, C. Park, J.-P. Wang, A. Fert, W. Zhao, Nat. Electron. 2018, 1, 582.

[38] N. Zhang, Y. Cao, Y. Li, A. W. Rushforth, Y. Ji, H. Zheng, K. Wang, Adv. Electron. Mater. 2020, 6, 2000296.

[39] M. Yang, Y. Deng, Z. Wu, K. Cai, K. W. Edmonds, Y. Li, Y. Sheng, S. Wang, Y. Cui, J. Luo, Y. Ji, H. Zheng, K. Wang, IEEE Electron Device Lett. 2019, 40, 1554.

[40] Q. Ma, Y. Li, D. B. Gopman, Y. P. Kabanov, R. D. Shull, C. L. Chien, Phys. Rev. Lett. 2018, 120, 117703.

[41] Y. Cao, G. Xing, H. Lin, N. Zhang, H. Zheng, K. Wang, iScience 2020, 23, 101614.

[42] D. Dua, Graff, Casey, (UCI) Machine Learning Repository, http://archive.ics.uci.edu/ml (accessed: April 2020).

[43] E. J. Fuller, F. E. Gabaly, F. Léonard, S. Agarwal, S. J. Plimpton, R. B. Jacobs-Gedrim, C. D. James, M. J. Marinella, A. A. Talin, Adv. Mater. 2017, 29, 1604310.

[44] C. C. Yann LeCun, Christopher J. C. Burges, THE MNIST DATABASE of handwritten digits, http://yann.lecun.com/exdb/mnist/ (accessed: April 2020).

[45] S. Choi, S. H. Tan, Z. Li, Y. Kim, C. Choi, P.-Y. Chen, H. Yeon, S. Yu, J. Kim, Nat. Mater. 2018, 17, 335.

[46] K. M. Song, J.-S. Jeong, B. Pan, X. Zhang, J. Xia, S. Cha, T.-E. Park, K. Kim, S. Finizio, J. Raabe, J. Chang, Y. Zhou, W. Zhao, W. Kang, H. Ju, S. Woo, Nat. Electron. 2020, 3, 148.