

ARTICLE

Open Access

Integrated neuromorphic computing networks by artificial spin synapses and spin neurons

Seungmo Yang¹, Jeonghun Shin², Taeyoon Kim³, Kyoung-Woong Moon¹, Jaewook Kim³, Gabriel Jang², Da Seul Hyeon², Jungyup Yang⁴, Chanyong Hwang¹, YeonJoo Jeong³ and Jin Pyo Hong²

Abstract

One long-standing goal in the emerging neuromorphic field is to create a reliable neural network hardware implementation that has low energy consumption, while providing massively parallel computation. Although diverse oxide-based devices have made significant progress as artificial synaptic and neuronal components, these devices still need further optimization regarding linearity, symmetry, and stability. Here, we present a proof-of-concept experiment for integrated neuromorphic computing networks by utilizing spintronics-based synapse (spin-S) and neuron (spin-N) devices, along with linear and symmetric weight responses for spin-S using a stripe domain and activation functions for spin-N. An integrated neural network of electrically connected spin-S and spin-N successfully proves the integration function for a simple pattern classification task. We simulate a spin-N network using the extracted device characteristics and demonstrate a high classification accuracy (over 93%) for the spin-S and spin-N optimization without the assistance of additional software or circuits required in previous reports. These experimental studies provide a new path toward establishing more compact and efficient neural network systems with optimized multifunctional spintronic devices.

Introduction

Advances in hardware technologies have resulted in hardware implementations of numerous neural network algorithms, including deep neural networks and convolutional neural networks, that use a feasible amount of computing resources. In turn, such implementations have fueled further algorithm developments^{1,2}. Intensive studies and commercialization are underway to accelerate execution of such algorithms, using mature complementary metal–oxide semiconductor (CMOS) technology; some examples are graphics processing units^{3,4}

and application-specific integrated circuits^{5–7}. However, this approach has the disadvantages of high power consumption and large-area coverage, which limits the use of recently developed algorithms despite significant breakthroughs in CMOS-based frameworks. To overcome these limitations, one promising approach is to apply the emerging in-memory analog computing concept in a crossbar array network, which is similar to an artificial synapse⁸. The simple Ohm's and Kirchhoff's laws for crossbar networks can enable massively parallel data processing that offer beneficial energy efficiency and performance^{9–11}. For example, previous studies using oxide-based resistive random access memory (RRAM) have successfully demonstrated in situ training of a simple perceptron algorithm in emerging device networks and validated the aforementioned concept using classification tasks^{12–14}. Nevertheless, hardware-based systems require further improvements. From an emerging device viewpoint, a linear and symmetric resistance change is important for achieving higher accuracy on a given task;

Correspondence: YeonJoo Jeong (jeongyeonjoo@kist.re.kr) or Jin Pyo Hong (jphong@hanyang.ac.kr)

¹Quantum Spin Team, Korea Research Institute of Standards and Science, Daejeon 34113, Republic of Korea

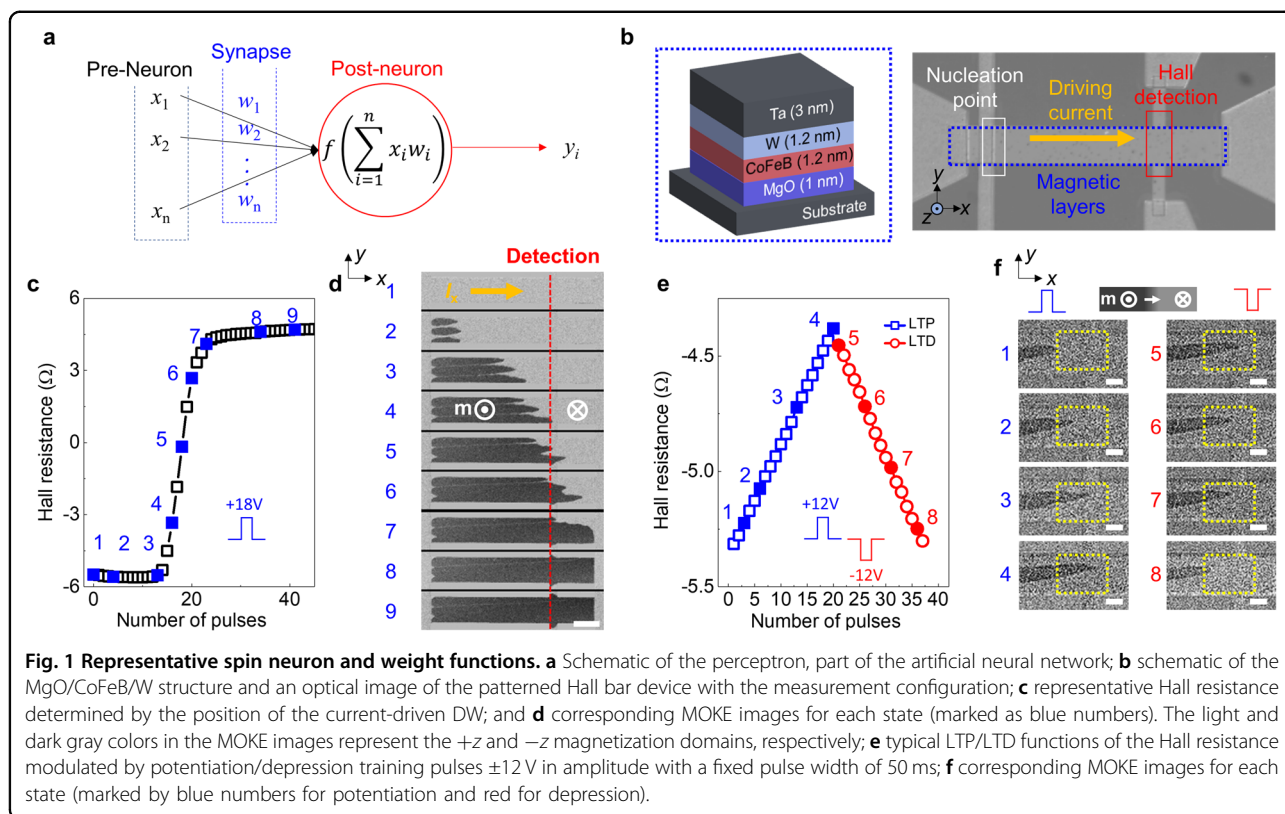
²Research Institute of Natural Science, Department of Physics and Division of Nanoscale Semiconductor engineering, Hanyang University, Seoul 04763, Republic of Korea

Full list of author information is available at the end of the article
These authors contributed equally: Seungmo Yang, Jeonghun Shin, Taeyoon Kim

© The Author(s) 2021



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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.



however, most RRAMs exhibit intrinsically imperfect linear and symmetric characteristics. Another essential component in an artificial neural network, a perceptron, and neuron, as shown in Fig. 1a, relies mainly on software or dedicated CMOS analog circuits. Thus, developing neuron components that can be monolithically integrated at a simple device level with the existing weight device crossbar array in a compatible fabrication process are necessary for widespread use in hardware-based neural networks.

In recent years, spintronics-based devices, such as those employing current-induced domain wall (DW) motion, have attracted considerable interest as basic building blocks for advanced neuromorphic component deployments. These devices offer low-power consumption, and highly stable and reproducible operation^{15–20}, according to the experimentally well-established model^{21,22}. Experimental demonstrations of DW-based neuromorphic components using magnetic tunnel junctions (MTJs) have been reported^{23,24}. MTJ-based resistance output devices generate a current output from the voltage input signal. Therefore, the current output signal from the weight device must be converted to the voltage signal to operate the activation function, which is achieved by an I - V converter with an operational amplifier (op-amp). Please note that I - V converter is required by all the

resistance-based neuromorphic computing components. However, if the weight devices were to generate an output signal as voltage, they could be connected directly to the activation generator to operate it. Recently²⁵, the performance of an artificial synapse was investigated by manipulating Hall voltage output signals conjugated by a novel spin texture, a magnetic skyrmion. However, previously reported output voltage-based artificial synapse devices still have the aforementioned issues (such as nonlinearity) that prevent practical implementations of hardware-based neural networks. In contrast, this work exhibits weight behavior linearity by employing a stripe domain motion that serves as a weight in the DNN algorithm and produces a spin-synapse (spin-S) voltage output signal instead of employing the conventional fully filled domain wall. Furthermore, by simply tuning the device operation principles, we also test a DW-based neuron (spin-N) in the same device geometry that functions as a sigmoidal activation function and has a voltage output signal. Together, these findings provide a new crossbar array configuration that employs voltage output (Hall voltage) weights and acts as a DNN accelerator. Furthermore, we propose the concept of how the voltage output signal of the spin-Ss can operate the spin-N at an array level, along with an experimental demonstration.

Basic behavior of domain wall-based device performance

Sample stacks of Si/SiO₂/1 nm MgO/1.2 nm Co₂₀Fe₆₀B₂₀/1.2 nm W/3 nm Ta with perpendicular magnetic anisotropy (PMA, see Fig. S1) were patterned into a Hall bar geometry with an asymmetric length, as shown in Fig. 1b. The distance between the nucleation region and the Hall detection region (red and white boxes, respectively, in Fig. 1b) was ~260 μm, and the channel width was 60 μm. Figure 1c, d shows the variation in the Hall resistance (R_H) under voltage pulses and the corresponding DW states observed from the contrast difference in polar magneto-optical Kerr effect (MOKE) microscopy images. Here, R_H is defined as the Hall voltage (V_H) detected in the Hall detection regime (red box) divided by the x -axis injected current (I_x). It is dominated by the z -component of the magnetization in the Hall detection region because the Hall voltage detected here is mostly dominated by the anomalous Hall effect. In its initial state, the entire magnetic layer is intentionally aligned with the $-z$ -axis (state 1 in Fig. 1c, d). Upon the application of a voltage pulse of +18 V and 50 ms with an external x -axis magnetic field (H_x) of 225 Oe, a DW is clearly formed at the nucleation point (details are presented in the “Method” section), and then the DW position shifts to state 2 or 3. However, R_H remains unaffected until the DW reaches the Hall detection region (red line). The R_H starts to increase in state 4, reflecting the arrival of the DW in the Hall detection region and becomes saturated after reaching state 7 upon consecutive voltage pulses. These DW behaviors, including the velocity, detection, and starting positions, can be precisely adjusted by changing diverse operation parameters and adjusting the device architecture, as described later.

Spin-synapse for highly linear and symmetric functions

Figure 1e shows the linear and symmetric variation in the R_H of spin-S versus the number of pulses during potentiation (blue)–depression (red) operation, along with the corresponding time-synchronized MOKE images (yellow box). These distinct weight features are mainly attributable to the uniform shifts of the DW caused by the consecutive pulses. Voltage pulses of ±12 V (duration of 50 ms) are introduced with a y -axis magnetic field of –80 Oe.

To gain insights into the linear and symmetric variation in R_H under identical voltage pulses, we adopt a creep scaling model for the DW motion, in which the velocity of the DW (v_{DM}) is explained by the Arrhenius form with an effective energy barrier height, $\alpha H_{z,eff}^{-\mu}(V)$:

$$\ln v_{DM} = \ln v_0 - \frac{\alpha H_{z,eff}^{-\mu}(V)}{k_B T}, \quad (1)$$

where v_0 , α , k_B , T , and $H_{z,eff}$ represent the characteristic velocity, scaling constant, Boltzmann constant, absolute

temperature, and z -component of the magnetic field, respectively, and the exponent μ is 0.25 (refs. 26,27). This model implies that v_{DM} depends only on the applied voltage and magnetic field; it does not depend on either the internal state or the DW position. Thus, the same v_{DM} is expected during the program and can be erased if identical pulses are used. In this regard, Fig. 2a shows a plot of the v_{DM} of spin-S versus the applied voltage under different magnetic fields, verifying the applicability of the model to our device. The experimental results are well-fitted with the creep model under the assumption of H_z , $SOT = \tau_{SOT} V$, clearly supporting the linearity and symmetry of the proposed spin-S (details on determining v_{DM} are presented in Fig. S2). However, in recent studies²⁸ using creep-based DW motion, similar performances for the weight behavior were not achieved, possibly owing to the depinning field (H_{dep}) arising at the Hall detection terminals²⁹. When a growing DW arrives at the Hall detection region, a change in the channel width causes modulation of H_{dep} , and the depinning effect varies as the DW grows out of the detection area. This results in a nonlinear variation in R_H for the conventional DW, as shown in Fig. 2c, where the detailed structure for the conventional DW experiment was Si/SiO₂/1 nm MgO/0.8 nm Co₂₀Fe₆₀B₂₀/1.2 nm W/3 nm Ta and intended to obtain a higher PMA induced by the thinner CoFeB layer. To avoid the possible depinning issue, a stripe domain (Fig. 2b) was adopted that does not touch the edge of the patterned devices during device operation. The stripe domain is controlled by adjusting the CoFeB thickness related to the variation in PMA values³⁰. The details for the conventional DWs were Si/SiO₂/1 nm MgO/0.8 nm Co₂₀Fe₆₀B₂₀/1.2 nm W/3 nm Ta to obtain a higher PMA induced by the thinner CoFeB layer. The proposed stripe domain can suppress the depinning effect owing to the unchanged environment when the DW passes through the Hall detection region, leading to a linear and symmetric R_H change (the red color in Fig. 2c) compared with the conventional DW case (the orange line in Fig. S3). To further achieve a precise weight performance by means of the stripe domains, the number of stripe domains associated with the ratio of the stripe domain width and device width must be determined. The stripe domain width can be controlled by material parameter engineering, including PMA, dipole energy, DMI, and pinning densities.

To further examine the foregoing observations, the effects of the total pulse numbers for potentiation and depression were evaluated, as shown in Fig. 2d. The linearity parameter (β) was quantitatively extracted from the curves using a quadratic model for the change in the Hall resistance under the assumption that the stripe DW has a trapezoidal shape, as shown in Fig. S4, where $\beta = 0$ indicates a completely linear function. The green line in

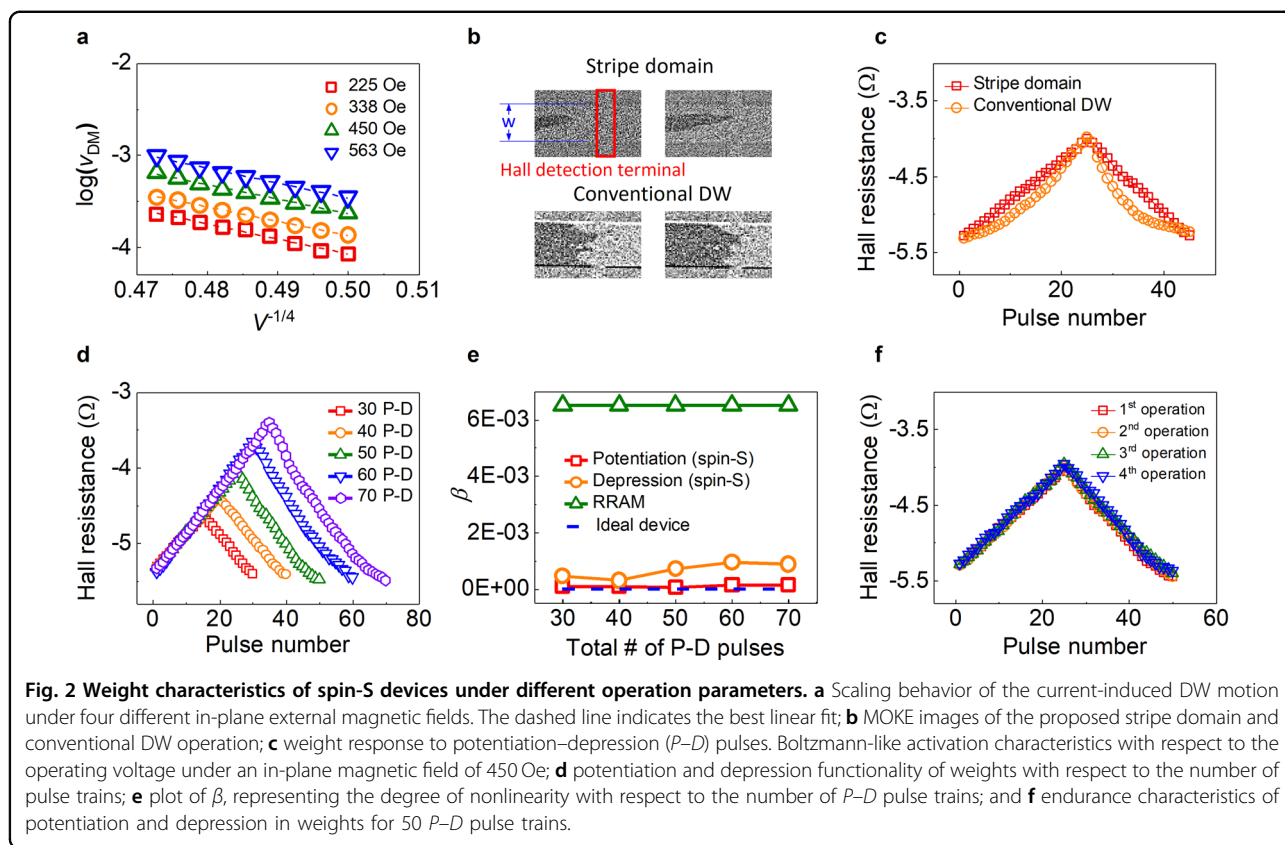


Fig. 2e represents the β of the representative oxide-based weight device³¹ exhibiting a higher β , and the red and orange lines reflect the β values taken from the spin-S for potentiation and depression, respectively. Both lines are relatively close to the ideal case (blue line). The slight difference in the β value for the potentiation (β_p) and depression (β_d) of spin-S can be explained by considering either the surface energy of the DW or a slight change in the shape of the stripe domain during operation³². Figure 2f shows a plot of the representative endurance features of R_H , which reflects the stability of the weight operation.

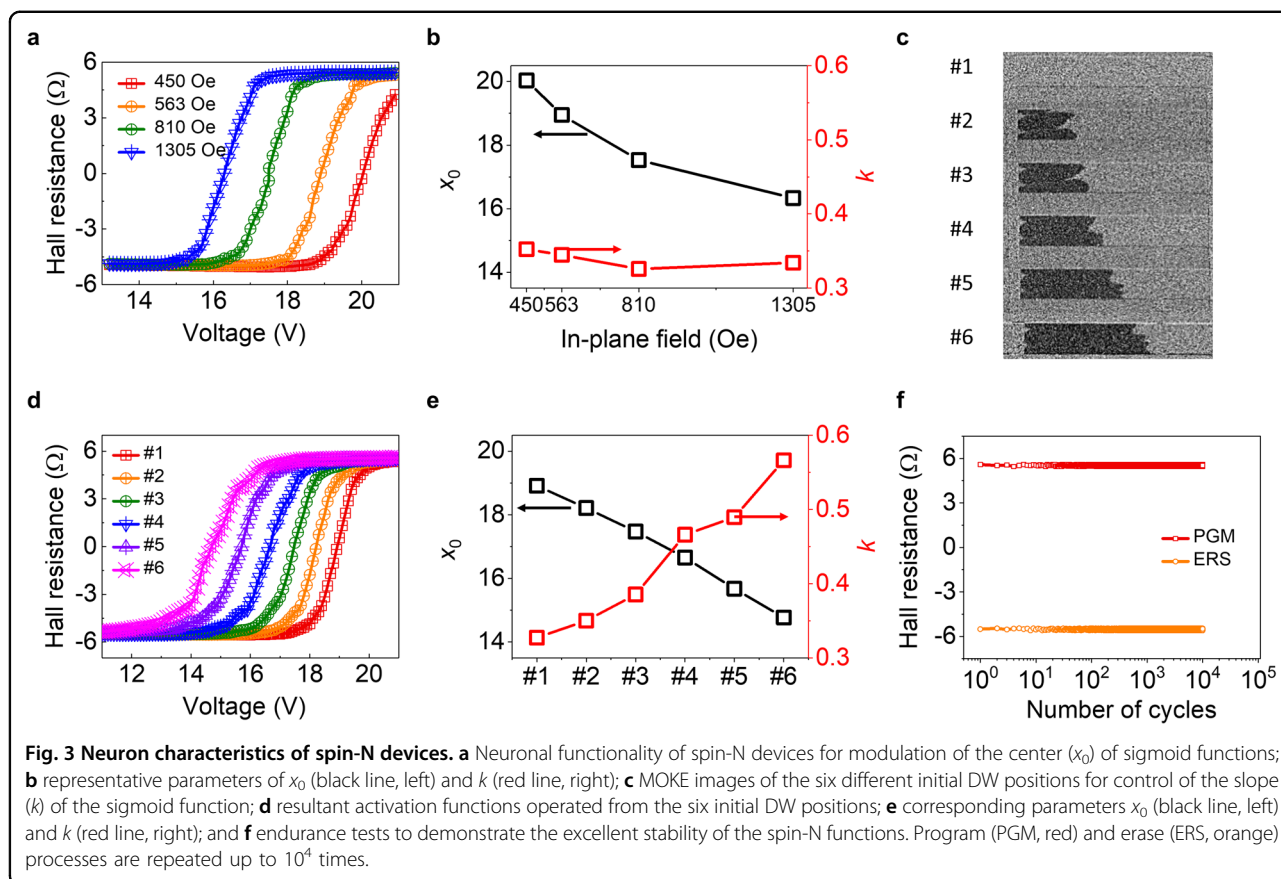
Spin-neuron for integration and activation functions

To satisfy the demand for compact artificial neuron–synapse integrated networks, a spin-N with a sigmoid function was prepared. The fabrication approach for the spin-N was identical to the spin-S described earlier to facilitate monolithic network integration. To achieve a basic sigmoid function, which is essential for solving nontrivial problems in multilayer networks³³, both the inactive states (1–3 in Fig. 1c) and the saturation states (7–9 in Fig. 1c) of the DW device were employed, as sources of nonlinearity in the spin-N. The plot in Fig. 3a shows the spin-N responses in R_H as a function of the voltage amplitude with a pulse width of 500 ms under

various x -axis magnetic fields (H_x). The sigmoid function can be fitted using the following equation:

$$y = \frac{A_1 - A_2}{1 + \exp\left(\frac{x-x_0}{k}\right)} + A_2, \tag{2}$$

where x_0 and k represent the rising point and slope parameters of the sigmoid function, respectively. Interestingly, both parameters can be tuned in the spin-N by varying the operation conditions or the device geometry. x_0 can be controlled by changing H_x ; that is, the shift in x_0 with an increase in H_x is a result of the higher initial DW velocity caused by the reduced DW energy (Fig. 2a). At a high H_x , the DW reaches the Hall detection region quickly, yielding a small value of x_0 . Figure 3b summarizes the relationship between x_0 and k at various H_x values, where k remains almost unaffected by H_x . However, because intentional control of k is one of the most important goals in the development of the sigmoid activation function, various k values were achieved, as shown in Fig. 3d, e, where the initial DW positions were intentionally selected by applying different erase pulse durations as indicated by the numbers in the corresponding MOKE images of Fig. 3c. When the distance from the initial DW position to the Hall detection region is shorter, the applied voltage when the DW passed the Hall



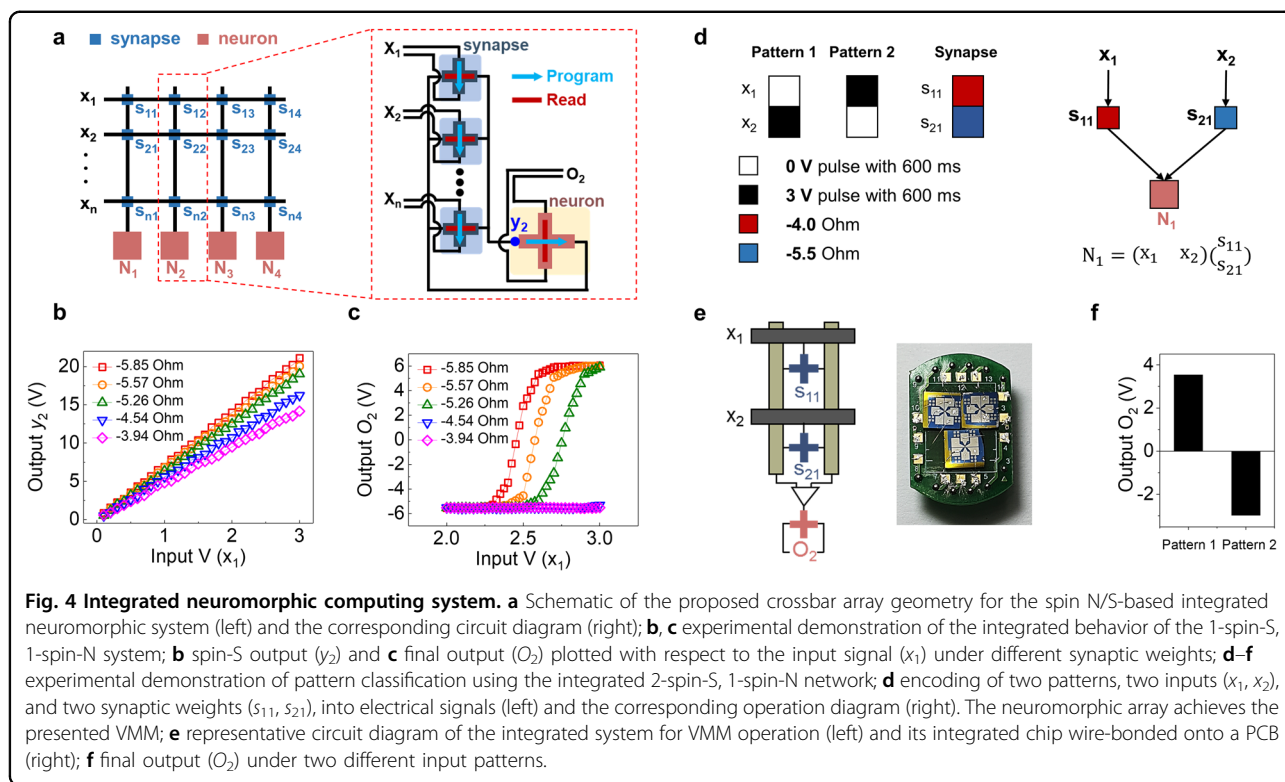
detection region is lower. Then, because k depends on how fast the DW passes the Hall detection region, x_0 shifts to the left and a change in k can be achieved. These tuning parameters can also be determined by varying the device geometry, such as the distance between the nucleation region and the Hall detection region. Finally, because the spin-N functions as a nonvolatile neuron, an erase operation is required after the activation level is read from the spin-N. Thus, the spin-N should also have high endurance. Figure 3f presents the endurance performance of spin-N in a cycling test (up to 10^4 cycles).

Integration of spin-S and spin-N for pattern classification

To facilitate massively parallel computation, we propose the incorporation of the obtained spin-S and spin-N functions into the crossbar array frame frequently adopted in conventional RRAM-based networks, as shown in Fig. 4a. The proposed spin-N/S devices are based on four-terminal electrodes, in which the programming path is separated from the read operation path and—more importantly—the input and output are both voltage signals. Thus, ideally, the two devices (spin-S and spin-N) can be directly connected to form a crossbar network. Although as simple proof-of-concept, we have connected

them via an op-amp, a proposed network array may not require an op-amp after further optimization, as described later. The working principle is as follows. The input voltages (x_n), which are proportional to the input amplitude, are introduced into the network, leading to the accumulation of charge carriers weighted by the synaptic weights (s_{nm}) at the Hall detection electrode. Then, the accumulated charges from all the weights in the same column gather and develop a Hall voltage. Hence, in the proposed network, all the information from all the connected weights is simply integrated in a convenient voltage form. The obtained total Hall voltage is appropriately adjusted through an op-amp to supply a suitable y_m to the next neuron stage. This operational principle follows vector matrix multiplication (VMM): $y_m = A \sum x_n s_{nm}$ where A represents the gain of the op-amp. The corresponding y_m drives the connected spin-N to operate in the same manner, producing an activation output (O_m) at each column.

To experimentally demonstrate the aforementioned operation principle, we plotted the response curves of y_m and O_m with respect to the input and weight values for integrated devices consisting of one spin-S and one spin-N, as shown in Fig. 4b, c. The output value at y_2 was clearly linearly dependent on the input and weight (Hall

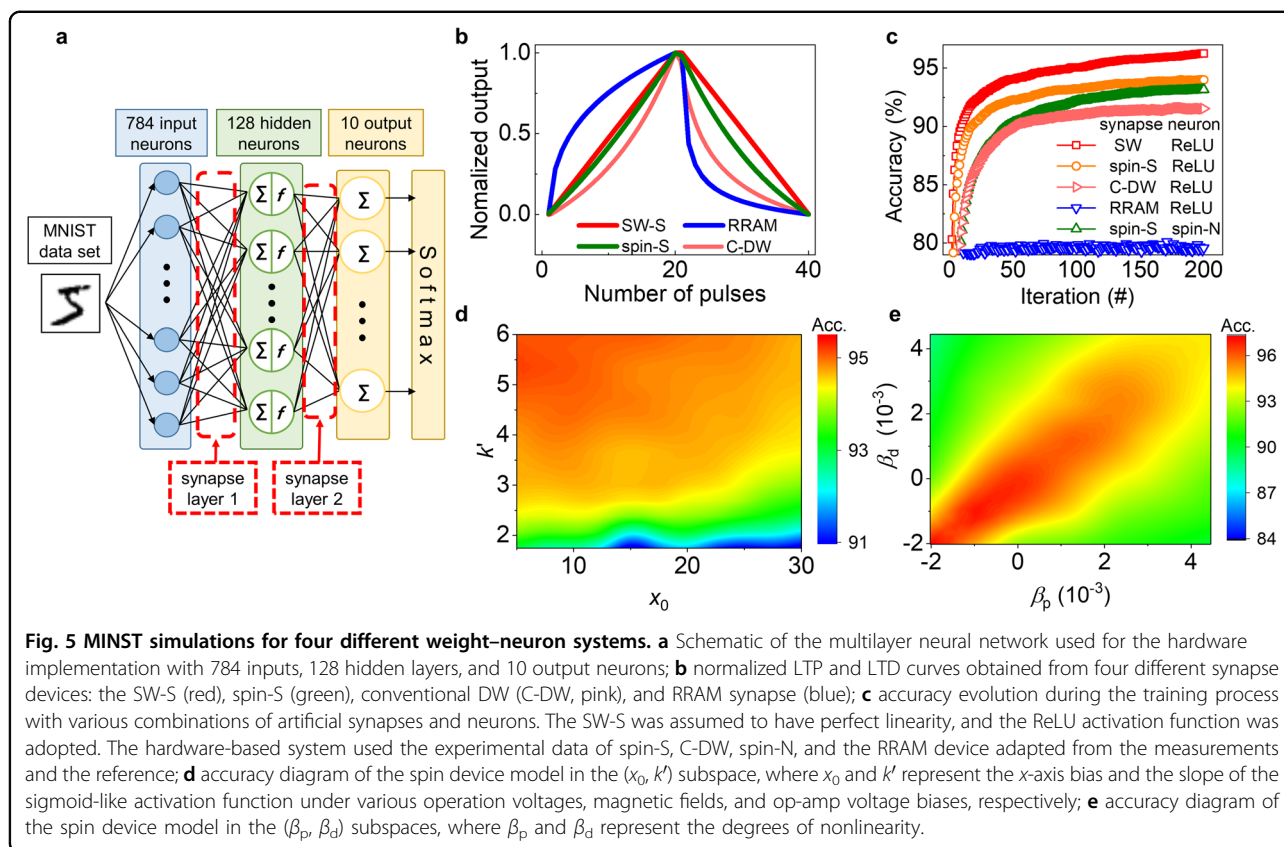


resistance). In addition, the final output of spin-N (O_2) generated representative activation curves depending on the y_2 level. These experimental observations demonstrate a successful VMM operation in the integrated spin-S and spin-N networks. We examined the concept further by considering a simple inference task of pattern classification utilizing the integrated frame from two spin-Ss and one spin-N ($2S + 1N$), where each device is electrically wire-connected in a printed circuit board (PCB), as shown in Fig. 4e. To conduct a proof-of-concept experiment for the integrated neuromorphic network behavior of Fig. 4, we installed a spin-N rotated by 90 degrees with respect to the spin-Ss under a single external magnetic field. Specifically, we used two manual patterns as inputs (Fig. 4d), where the weight column was already programmed to one of the patterns. With input pulses of 3 and 0 V (corresponding to the black and white pixels of the patterns, respectively), the network provides a higher activation value for the matched column. Figure 4f shows the two output levels corresponding to patterns 1 and 2 for a simple classification task performed by the small DW device array.

Simulation of a multilayer neural network with spin-S and spin-N

By exploiting the aforementioned observed spin-S and spin-N characteristics, we performed the pattern classification task shown in Fig. 5a by utilizing the Modified

National Institute of Standards and Technology³⁴ and Canadian Institute For Advanced Research 10 (CIFAR-10)³⁵ datasets (see the “Methods” section and Fig. S5). To evaluate the impact of the spin-N/S characteristics on the performance of the network, four different types of artificial synapse devices were selected: a software synapse (SW-S, red), the spin-S (stripe domain type, green), and the C-DW (conventional domain type, pink), as well as an oxide-based RRAM artificial synapse (blue) previously reported by another group²⁹. Figure 5b shows the representative long-term potentiation (LTP) and long-term depression (LTD) operational curves for the four different artificial synapses. The SW-S corresponded to the ideal condition for weight updating and error propagation, exhibiting high linearity and symmetry in the LTP and LTD curves. Thus, the SW-S can develop a precise weight update calculated via the gradient descent method during the entire training process, regardless of the current weight value. Figure 5c shows the results for the evolution of the classification accuracy under different combinations of neural components. As predicted, the SW-S/rectified linear unit (ReLU; red) combination exhibited the highest accuracy (>96%) after 200 iterations, whereas the oxide-based RRAM synapse device integrated with the ReLU (blue) exhibited the lowest accuracy (~79%). Surprisingly, due to its optimized characteristics, the spin-S/ReLU (orange) also generated a high accuracy (~94%) even without aid from a circuit or compensation



algorithm. In contrast, when the same ReLU activation function and the conventional DW type serving as weights (pink) were adopted, the accuracy reached only 91.5%. Moreover, the accuracy remained >93% even for the full hardware combination (i.e., the spin-S/spin-N case). These results indicate that the proposed spin-S and spin-N devices provide a novel solution for building a complete neuromorphic computing hardware implementation, while previously reported artificial synapse devices require a software-assisted ReLU activation function to play a neuron role. The outstanding abilities of the spin-S and spin-N components make them promising for use in fully operational artificial networks for high-performance systems with a simple artificial neural network design. These trends were confirmed in a simulation on the CIFAR-10 dataset, as depicted in Fig. S6. To establish the crucial relationship between the spin-S and spin-N device features and the fitting parameters, the simulation was conducted in an experimentally possible (x_0, k') subspace range, where two parameters— x_0 (start point) and k' (converted slope)—were selected for spin-N, as shown in Fig. 5d. As shown in Fig. 3e, k varied from 0.3 to 0.6; thus, the offset translation for the k value was established by adding an external resistor (Fig. S7) to employ the converted k (k') value in the simulation. The accuracy was increased to >95.18% by tuning the suitable

parameters of spin-N ($x_0 = 5.3$ and $k' = 5.5$) with the fixed weight characteristics of spin-S. Further performance enhancements could be achieved by adjusting the magnetic field, voltage range, or device design, as suggested in Fig. 3. In addition, the effects of the spin-S parameters, including the nonlinearity (β_p and β_d , defined in Fig. 2e), were simulated with the spin-N parameters fixed, as shown in Fig. 5e. Here, in addition to the nonlinearity, the symmetry between β_p and β_d is important to the accuracy. Our findings may provide guidelines for implementing novel hardware-based neural networks.

Potential advantages and issues of spin-N and spin-S for DNN accelerator applications

Our findings with the spin-N/S devices form an initial proof-of-concept experiment; such devices still have numerous limitations to the realization of real device applications. Thus, this section emphasizes the potential for employing spin-N and spin-S to construct DNN accelerators after further optimization. Possible approaches are discussed to resolve the current issues and to compare the potentials to those of other emerging devices.

The first issue is the operating speed. The operating speed of the current spin-N/S devices is governed by a DW motion using a creep model, as given in Fig. 2a.

While the creep regime of a DW motion is too slow to be used in a modern computing system. However, the device performance could be expanded to the flow regime of DW motion, providing a higher DW velocity of ~ 5700 m/s under ~ 1 ns pulses³⁶. This speed has also been demonstrated in a racetrack memory³⁷, supporting the future operation of spin-N/S devices in a GHz range.

The second issue involves scaling. This work employed relatively large ($60 \times 260 \mu\text{m}^2$) devices as an initial approach for spin N/S devices and to precisely detect domain state variation. However, the device size could be reduced to a sub-nm scale because the physically possible minimum size of spin-N/S can be estimated by the minimum stripe domain size. Note that when estimating a minimum stripe domain size, the stripe domain is a consequence of a trace of the half-skyrmion at the end. Thus, the minimum stripe width corresponds to the minimum half-skyrmion or minimum skyrmion size. To date, the theoretically and experimentally skyrmion sizes have been observed down to the sub-nm scale. Therefore, the stripe domain device (a half-skyrmion device) can be scaled to sub-nm sizes.

The third issue is high operating voltage. The operation voltage of the spin-N/S was ~ 20 V, which is incompatible with modern circuitry. We believe that the high operating voltage issue will be solvable in the future because spin-N/S device operation is based on current, not on voltage (electric field). In this paper, the driving force for the spin-N/S is spin-orbit torque (SOT), which is torque created by the spin Hall effect. Thus, the main parameter for spin-N/S devices is the operating current density, which is $\sim 10^{11}$ A/m² in our paper. However, to experimentally demonstrate a relation between DW states and electrical outputs, we fabricated spin-N/S devices at a relatively large size ($260 \times 60 \mu\text{m}$), which led to high resistance and is the reason why our prototype devices operated at a relatively high operating voltage (~ 20 V). At a real device design level, the ferromagnetic layer thickness, heavy-metal layer thickness, and the device width/length ratio could all be parameters for adjusting the device resistance. For example, one recent paper²⁵ utilized (Pt/CoFeGd/MgO) multilayers to operate skyrmion synapse devices with low resistance; the operation voltage was only a few millivolts with a subnanosecond pulse duration. It is worth noting that the operating nature of skyrmion synapse device paper was exactly the same as that of our work: SOT. Thus, further optimizing the device material parameters to lower device resistance may be a reliable approach for overcoming the high operating voltage issue in this work (more information is provided in Figs. S8 and S9).

The fourth issue is the required use of an additional external magnetic field during device operation. A possible alternative frame that would enable removal of the external magnetic field in future applications is as follows:

we describe the role of external parallel and perpendicular magnetic fields in the current direction briefly before explaining our alternative frame. First, the magnetic fields parallel to the current in spin-N have an effect on controlling the DW velocity, that is, the operating voltage of the spin neuron, as seen in Fig. 3a, b. Therefore, the role of the x -axis magnetic field in the spin neuron could easily be replaced by employing a suitable PMA value, which could be achieved by varying the CoFeB layer thickness or choosing a postannealing temperature. A more critical factor is the magnetic field perpendicular to the current in the spin-S device. This additional magnetic field serves to adjust the stripe domain elongation direction. The stripe domain with a sufficient DMI has a half-skyrmion at the end of the stripe domain, thereby inducing the skyrmion Hall effect. This effect implies that the skyrmion motion has a transverse component in velocity around the driving force (the driving current). As such, the stripe domain elongates along the direction deviating from the current, which consequently pushes the stripe domain toward the edges of the patterned device. As mentioned before, because the linearity of spin-S originates from stripe domain motion away from the edge, the stripe domain must elongate in the direction parallel to the current. Manipulation of the half-skyrmion Hall effect can be achieved by the in-magnetic field perpendicular to the driving force due to internal structure deformation. Among the various approaches for creating effective in-plane fields, one of the most compatible approaches is the insertion of the exchange bias layer. For example, a recent report³⁸ addressed magnetic field-free SOT switching by inserting an in-plane magnetized layer on a perpendicularly magnetized ferromagnet; that is, the stable external magnetic field can be replaced by inserting an exchange bias layer into the previous configuration. Thus, although our current work utilizes an external y -axis magnetic field to ensure the straight motion of the stripe domain, the insertion of an additional exchange bias layer may be crucial for removing the external magnetic field in the near future.

The possible advantages of spin-N/S devices over the existing emerging devices are as follows: first, spin-S has a higher linear weight variation due to the stripe domain motion. As described in Fig. 5e, the linearity of weight variation has a significant effect on the accuracy of the trained network. However, the emerging weight devices^{25,39,40} still possess nonlinearity features except for those reported in a few papers^{41,42} due to their resistance variation mechanism, and also possess relatively wide cycle-to-cycle distribution in device performance. However, the main information carriers of the spin-N/S (stripe domain or convention DW) are governed by the well-established physical model, thereby enabling the possibility of precise control of the information carriers. In

addition, the spin-N/S shares similar materials, structures, and operating schemes with the recently well-established SOT random access memories and racetrack memories, which have high endurance and retention features compared to those of RRAM and PCRAM devices. Table S1 summarizes the potential performance for diverse weight devices.

Conclusion

This study presented the first proof-of-concept demonstration of an integrated neuromorphic network using spintronics-based synapses (spin-S) and neurons (spin-N), both of which are prepared via the same fabrication process. We provided a crossbar array architecture for the Hall voltage output of the spin-S, not for conventional resistance output devices, and experimentally applied it to a simple pattern classification task using an electrically integrated two spin-Ss/one spin-N network, showing its possibility for constructing more compact neuromorphic computing networks. Simulations using experimentally determined parameters yielded a high accuracy (93%) in completely spin-N/S-based neural networks, thereby showing the possibility of developing compact and efficient spin-based neural networks. Nevertheless, further empirical observations and comparisons together with the elimination of the external magnetic field commonly required for device operation are needed to exploit a crucial device architecture.

Method

Sample fabrication

The films used in this study were deposited on 200-nm-thick thermally oxidized Si substrates via magnetron sputtering with a base pressure of $<7 \times 10^{-8}$ Torr at room temperature. To provide the PMA characteristics, a postannealing process was conducted at 350 °C for 30 min under vacuum conditions of $<1 \times 10^{-6}$ Torr with a 3-T perpendicular magnetic field. The asymmetric Hall bar geometry was obtained by utilizing photolithography and Ar ion milling, followed by an O₂ plasma ashing process for 2 min at 50-W radiofrequency power to completely remove the residual photoresist material hardened by the ion milling process. The electrodes for the Hall channel and the nucleation line were prepared as Ta (3 nm)/Pt (100 nm) layers.

MOKE microscopy and electrical measurement

A custom-built MOKE microscopy system with out-of-plane and in-plane electromagnets was employed to image the domains used in the spin-N and spin-S devices. As shown in Fig. S1, a stable PMA feature was observed. The +z and -z domains were clearly identified by the contrast difference in the MOKE microscopy images, as shown in Fig. 1d, f. To observe the current-induced DW

motion, four probes were incorporated in the MOKE system; two were connected to the voltage source path, and the other two were connected to the Hall voltage detection terminals. The Hall voltage was monitored using a Hewlett Packard 34401 A multimeter device. In addition, to synchronize the MOKE images with the Hall voltage signals, the MOKE images were programmed to be taken immediately after the injection of each voltage pulse.

Formation of DWs at nucleation sites and current-driven DW motion

With the Hall bar design, the driving current flowed mainly through the *W* layer along the *x*-direction (yellow line in the right-hand image in Fig. 1b). The magnetic CoFeB layer could easily be damaged in the nucleation region by the sputtering growth of the nucleation electrode, reflecting the presence of a significantly reduced PMA energy (K_{eff}) that was proportional to the energy required for magnetization reversal. Hence, the initially reversed magnetization drove the formation of the DW within the nucleation region; then, the DW spread out in the CoFeB layer along the driving current direction via the SOT phenomenon, where the Neel-type DW was stabilized by the finite DMI energy.

Network structure for simulation

In the neural network simulation, two-synapse-layer perceptron networks with 784 (28×28) input neurons, 128 hidden neurons, and 10 output neurons were used, as shown in Fig. 5a. ReLU and softmax³⁵ were adopted as the activation and loss functions, respectively. ReLU is highly popular due to its hardware-friendly implementation³⁶.

Acknowledgements

This work was supported in part by the Korea Institute of Science and Technology (KIST) Open Research Program (ORP) through Grant 2E30610 and in part by the National Research Foundation of Korea (Grant numbers NRF-2019M3F3A1A03079422, NRF-2016M3A7B4910249, and NRF-2019M3F3A1A02072478).

Author details

¹Quantum Spin Team, Korea Research Institute of Standards and Science, Daejeon 34113, Republic of Korea. ²Research Institute of Natural Science, Department of Physics and Division of Nanoscale Semiconductor engineering, Hanyang University, Seoul 04763, Republic of Korea. ³Center for Neuromorphic Engineering, Korea Institute of Science and Technology, Seoul 02792, Republic of Korea. ⁴Department of Physics, Kunsan National University, Gunsan 54150, Republic of Korea

Author contributions

J.P.H. directed and supported this project. S.Y. and J.S. designed the experiments, including the sample fabrication and analyses. T.K. and Y.J.J. conducted the simulations. S.Y., T.K., J.K. and Y.J.J. designed and proposed the circuit structure for the integrated neuromorphic computing networks. K.-W.M. helped with the domain wall dynamics analyses. G.J., D.H., J.Y.Y., and C.H. performed the sample design and etching. All the authors discussed the results and commented on the manuscript.

Data availability

All the relevant data are available from the authors upon reasonable request.

Conflict of interest

The authors declare that they have no conflict of interest.

Publisher's note

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

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41427-021-00282-3>.

Received: 31 March 2020 Revised: 9 December 2020 Accepted: 18 December 2020.

Published online: 29 January 2021

References

- LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
- Silver, D. et al. Mastering the game of Go without human knowledge. *Nature* **550**, 354–359 (2017).
- Nuvitadhi, E. et al. *Proceedings of the 2017 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays 5* (Monterey, California, USA, 2017).
- Raihan, M. A., Goli, N. & Aamodt, T. M. 2019 *IEEE International Symposium on Performance Analysis of Systems and Software 79–92* (ISPASS, Madison, Wisconsin, 2019).
- Jouppi, N. P. et al. 2017 *ACM/IEEE 44th Annual International Symposium on Computer Architecture 1–12* (ISCA, Toronto, ON).
- Jouppi, N., Young, C., Patil, N. & Patterson, D. Motivation for and evaluation of the first tensor processing unit. *IEEE Micro* **38**, 10–19 (2018).
- Chen, Y. et al. 2014 *47th Annual IEEE/ACM International Symposium on Microarchitecture*. 609–622 (IEEE Computer Society, 2014).
- Li, C., Wang, Z. & Rao, M. et al. Long short-term memory networks in memristor crossbar arrays. *Nat. Mach. Intell.* **1**, 49–57 (2019).
- Hu, M. et al. 2016 *53rd ACM/EDAC/IEEE Design Automation Conference (DAC)*, 1–6 (IEEE, Austin, TX).
- Sheridan, P. M. et al. Sparse coding with memristor networks. *Nat. Nanotechnol.* **12**, 784 (2017).
- Li, C. et al. Analogue signal and image processing with large memristor crossbars. *Nat. Electron.* **1**, 52–59 (2018).
- Prezioso, M. et al. Training and operation of an integrated neuromorphic network based on metal-oxide memristors. *Nature* **521**, 61 (2015).
- Bayat, F. M. et al. Implementation of multilayer perceptron network with highly uniform passive memristive crossbar circuits. *Nat. Commun.* **9**, 2331 (2018).
- Li, C. et al. Efficient and self-adaptive in-situ learning in multilayer memristor neural networks. *Nat. Commun.* **9**, 2385 (2018).
- Torreson, J. et al. Neuromorphic computing with nanoscale spintronic oscillators. *Nature* **547**, 428 (2017).
- Huang, Y., Kang, W., Zhang, X., Zhou, Y. & Zhao, W. Magnetic skyrmion-based synaptic devices. *Nanotechnology* **28**, 08LT02 (2017).
- Romera, M. et al. Vowel recognition with four coupled spin-torque nano-oscillators. *Nature* **563**, 230–234 (2018).
- Bhowmik, D. et al. On-chip learning for domain wall synapse based fully connected neural network. *J. Magn. Magn. Mater.* **489**, 165434 (2019).
- Yue, K., Liu, Y., Lake, R. K. & Parker, A. C. A brain-plausible neuromorphic on-the-fly learning system implemented with magnetic domain wall analog memristors. *Sci. Adv.* **5**, eaau8170 (2019).
- Sengupta, A., Shim, Y. & Roy, K. Proposal for an all-spin artificial neural network: emulating neural and synaptic functionalities through domain wall motion in ferromagnets. *IEEE Trans. Biomed.* **10**, 1152–1160 (2016).
- Lemerle, S. et al. Domain wall creep in an Ising ultrathin magnetic film. *Phys. Rev. Lett.* **80**, 849–852 (1998).
- Kim, K.-J. et al. Interdimensional universality of dynamic interfaces. *Nature* **458**, 740–742 (2009).
- Shibata, T. et al. Linear and symmetric conductance response of magnetic domain wall type spin-memristor for analog neuromorphic computing. *Appl. Phys. Express* **13**, 043004 (2020).
- Siddiqui, S. A. et al. Magnetic domain wall based synaptic and activation function generator for neuromorphic accelerators. *Nano Lett.* **20**, 1033–1040 (2020).
- Song, K. M. et al. Skyrmion-based artificial synapses for neuromorphic computing. *Nat. Electron.* **3**, 148–155 (2020).
- Lee, J.-C. et al. Universality classes of magnetic domain wall motion. *Phys. Rev. Lett.* **107**, 067201 (2011).
- Moon, K.-W. et al. Distinct universality classes of domain wall roughness in two-dimensional Pt/Co/Pt films. *Phys. Rev. Lett.* **110**, 107203 (2013).
- Zhang, S. et al. A spin-orbit-torque memristive device. *Adv. Electron.* **5**, 1800782 (2019).
- Jiang, W. et al. Direct observation of domain-wall surface tension by deflating or inflating a magnetic bubble. *Phys. Rev. Appl.* **9**, 024032 (2018).
- Jiang, W. et al. Blowing magnetic skyrmion bubbles. *Science* **349**, 283–286 (2015).
- Jang, J., Park, S., Burr, G. W., Hwang, H. & Jeong, Y. Optimization of conductance change in Pr_{1-x}Ca_xMnO₃-based synaptic devices for neuromorphic systems. *IEEE Electron Device Lett.* **36**, 457–459 (2015).
- Moon, K.-W. et al. Long-range domain wall tension in Pt/Co/Pt films with perpendicular magnetic anisotropy. *Appl. Phys. Express* **4**, 043004 (2011).
- Karlik, I. Jo. A. Ia. E. S. & Olgac, A. V. Performance analysis of various activation functions in generalized MLP architectures of neural networks. *Int. J. Artif. Intell. Expert Syst.* **1**, 111–122 (2011).
- Lecun, Y., Bottou, L., Bengio, Y. & Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **86**, 2278–2324 (1998).
- Krizhevsky, A. *Learning Multiple Layers of Features from Tiny Images*. Technical Report TR-2009 (University of Toronto, 2012).
- Cai, K. et al. Ultrafast and energy-efficient spin-orbit torque switching in compensated ferrimagnets. *Nat. Electron.* **3**, 37–42 (2020).
- Parkin, S. S. P., Hayashi, M. & Thomas, L. Magnetic domain-wall racetrack memory. *Science* **320**, 190–194 (2008).
- Lau, Y. C., Betto, D., Rode, K., Coey, J. M. D. & Stamenov, P. Spin-orbit torque switching without an external field using interlayer exchange coupling. *Nat. Nanotechnol.* **11**, 758–762 (2016).
- Choi, S. et al. A self-rectifying TaO_x/nanoporous TaO_x memristor synaptic array for learning and energy-efficient neuromorphic systems. *NPG Asia Mater.* **10**, 1097–1106 (2018).
- Lequeux, S. et al. A magnetic synapse: multilevel spin-torque memristor with perpendicular anisotropy. *Sci. Rep.* **6**, 31510 (2016).
- Fuller, E. J. et al. Parallel programming of an ionic floating-gate memory array for scalable neuromorphic computing. *Science* **364**, 570–574 (2019).
- Burgt, Y. et al. A non-volatile organic electrochemical device as a low-voltage artificial synapse for neuromorphic computing. *Nat. Mater.* **16**, 414–418 (2017).