

Stochastic Synapses Made of Magnetic Domain Walls


Cen Wang,^{1,2,‡} Kuan Wang,^{1,2,3,‡} Xinyu Wen,^{1,2,3} Wei Luo,^{1,2} Shiheng Liang,⁴ Yue Zhang^{1,2,*} and Yuhui He^{1,2,3,†}

¹*School of Integrated Circuits, Huazhong University of Science and Technology, Wuhan 430074, People's Republic of China*

²*School of Optical and Electronic Information, Huazhong University of Science and Technology, Wuhan 430074, People's Republic of China*

³*Hubei Yangtze Memory Laboratories, Wuhan 430074, China*

⁴*Department of Physics, Hubei University, Wuhan 430062, China*

 (Received 29 July 2022; revised 6 October 2022; accepted 27 October 2022; published 6 December 2022)

The stochastic artificial neural network is widely studied nowadays, since it may address the critical issue of uncertainty quantification demanded in many application scenarios. At the hardware level, it calls for devices that can emulate stochastic weight-modification behaviors of synapses with high efficiency. In this work, we propose a stochastic spintronic synapse based on spin-torque-induced stochastic magnetic domain-wall (DW) dynamics, including stochastic DW generation by spin-transfer torque and deterministic DW motion induced by spin-orbit torque. The former plays the role of updating the synaptic weights stochastically, while the latter gives rise to multistates, namely, the analog property of the synapse. The proposed DW-based stochastic synapse requires few devices to implement sufficient resistance states. The neural-network-level simulation demonstrates that a spiking neural network based on these stochastic synapses is capable of classifying breast cancer data with a high accuracy of 95.7%.

DOI: [10.1103/PhysRevApplied.18.064014](https://doi.org/10.1103/PhysRevApplied.18.064014)

I. INTRODUCTION

Recent developments of artificial intelligence and machine learning have attracted significant interest owing to the successful application of the artificial neural network (ANN) and deep learning in numerous fields [1]. Although excellent performance is demonstrated for many benchmark tasks, ANN and deep-learning systems generally require very intensive computation in the implementation. Moreover, they are not quite capable of quantifying uncertainty, which plays a fundamental role in making predictions from the usually incomplete, inaccurate, and even contradictory input information of the real world [2]. For this aspect, probabilistic computing provides an effective adoption for uncertainty quantification in the imperfectly sensed and uncertain world [3]. By applying probability theory to evaluate uncertainty with a predictive model, probabilistic computing empowers us to infer unknown quantities, make predictions, and learn from data.

At the hardware level, a probabilistic ANN calls for neuromorphic devices that implement stochastic firing or synaptic weight update. However, utilizing silicon complementary metal oxide semiconductor circuits to

construct a probabilistic spiking neural network (SNN) always requires excessive energy and size. Therefore, exploring a stochastic SNN made of materials other than silicon has attracted wide attention, especially for stochastic artificial neurons. For example, Tuma *et al.* showed that stochastic phase-change neurons could achieve the temporal integration of postsynaptic potentials [4]. Wang *et al.* utilized stochastic neurons made of a CuS/Ge₄Se₆ conductive bridge to carry out Bayesian inference [5]. In addition to neurons, the synapse also plays a fundamental role in a SNN. A synapse provides adjustable weights for manipulating the connection between pre- and postneurons [6].

In recent years, spintronic devices have emerged to simulate neurons and synapses with high processing speed and low dissipation [7]. A typical example is mimicking the functions of neurons and synapses by exploiting resistance variation due to the motion of magnetic textures, such as skyrmions [8–13] and the domain wall (DW) [14–19]. Here, the DW is an ideal candidate for an artificial synapse thanks to continuous and nonvolatile resistance variation owing to DW motion. Especially, stochasticity can also be introduced into the DW-based synapse by exploiting the edge roughness and thermal fluctuation [15–19]. This stochasticity leads to the uncertainty in resistance measurement, and it appears to be challenging to control in experiments.

*yue-zhang@hust.edu.cn

†heyuhui@hust.edu.cn

‡C. Wang and K. Wang contributed equally to this work.

Here, we propose another stochastic magnetic DW-based synapse, where probabilistic weight updating is realized by utilizing the spin-torque-induced stochastic DW dynamics, including probabilistic DW generation by spin-transfer torque (STT) [20–22] and deterministic DW motion induced by spin-orbit torque (SOT) [23,24]. The probability for DW generation results from thermal fluctuation, and it can be controlled between 0 and 1 by modifying the current density for nucleating the DW. Not only do the multiple nonvolatile resistance states outline the DW-tuning devices as promising artificial synapses, but also the stochastic DW dynamics is exploited for constructing synapses with probabilistic learning obeying the spike-time-dependent-plasticity (STDP) rule [25]. The proposed stochastic synapse device is implemented in a SNN simulation to classify the breast cancer dataset under unsupervised learning, and the accuracy reaches 95.7%.

II. MODELS AND METHODS

We consider current-induced DW dynamics in a heavy-metal (HM)/ferromagnetic metal (FM) bilayer nanostrip [Fig. 1(a)]. Here, the FM layer exhibits perpendicular magnetic anisotropy. Under a perpendicularly injected spin-polarized current, a DW is first generated in the FM layer near the end of the nanostrip due to the STT effect [20]. This spin-polarized current can be generated by passing a current through the pinned FM layer of a spin valve or injecting a current from a nanoscale FM tip [21]. At 0 K, a DW deterministically forms when the current density is higher than a fixed critical value. At temperatures above 0 K, however, DW generation is probabilistic over a range of current density due to thermal fluctuation [22]. After DW formation, another longitudinal current is injected along the nanostrip to trigger deterministic DW motion [23]. This DW motion is based on the SOT effect due to the spin Hall effect in the HM layer or the Rashba effect at the HM/FM interface [24]. Different DW positions in the nanostrip correspond to different resistance values, which can be characterized by measuring the anomalous Hall resistance [26] or tunneling magnetoresistance [27]. The combination of probabilistic DW formation and deterministic DW motion is the basis for the stochastic weight adjustment of the DW-based synapse.

In the FM layer, magnetization initially points along the $-z$ direction, corresponding to minimal weight ω_{down} . In a long-term-potential (LTP) process, current J_1 is first applied along the $+z$ direction near the left end of the nanostrip ($l \times 100 \text{ nm}^2$) to generate a DW (up DW) with probability π_{up} at 300 K. Here, l is the length of the region for the injection of spin-polarized current. We find that a stable DW can be generated when l is longer than 10 nm, and the time for this DW generation is shorter than 1 ns, even when l is 50 nm. Therefore, l is fixed at 50 nm. After generation of the DW, current J_2 is applied

along the $+x$ direction. If there is an up DW generated in the previous stage, the up DW will be driven with a linear increase of the displacement with the pulse number [Fig. 1(c)]. When the up DW moves to the rightmost end of the nanostrip, the weight reaches its maximum value, ω_{up} . In a long-term-depression (LTD) procedure, current J_3 along the $-z$ direction and current J_4 along the $-x$ direction are applied to generate a DW (down DW) at the right end of the nanostrip with probability π_{down} and drive it back along the $-x$ direction. The procedure of weight updating mapped from the stochastic DW formation and deterministic DW motion is illustrated in Fig. 1(d).

We exploit the software “Object Oriented MicroMagnetic Framework” with a code of thermal fluctuation to simulate the stochastic DW dynamics. The simulation is based on numerically solving the augmented Landau-Lifshitz-Gilbert equation [28]:

$$\frac{d\vec{m}}{dt} = -\gamma\mu_0\vec{m} \times \vec{H}_{\text{eff}} + \alpha \left(\vec{m} \times \frac{d\vec{m}}{dt} \right) + \vec{\tau}_{\text{STT}} + \vec{\tau}_{\text{SOT}}. \quad (1)$$

Here, $\vec{m} = \vec{M}/M_S$ is the normalized magnetization vector and M_S is the saturation magnetization; γ is the gyromagnetic ratio of electrons; μ_0 is the vacuum permeability; α is the damping constant; and \vec{H}_{eff} is the effective field, including the exchange, demagnetization, anisotropy, Dzyaloshinskii-Moriya interaction (DMI), and thermal fluctuation. Here, thermal fluctuation behaves as white noise with an effective field with fluctuation following a Gaussian distribution. The average of the fluctuation field is zero, and the square of the standard error can be estimated by [29]

$$\sigma_{\text{thermal}}^2 = \frac{\alpha}{1 + \alpha} \frac{2k_B T}{\gamma\mu_0 M_S V}. \quad (2)$$

Here, T is temperature, k_B is the Boltzmann constant, and V is the volume of the FM layer. The dampinglike STT is given by [28]

$$\vec{\tau}_{\text{STT}} = \frac{\gamma\hbar JP}{2edM_s} \vec{m} \times (\vec{m} \times \vec{m}_\sigma). \quad (3)$$

Here, \hbar is the reduced Planck constant, J is the current density, P is the spin-polarization degree, e is the electron charge, d is the thickness of the FM layer, and \vec{m}_σ is the unit vector denoting the spin orientation of a spin-polarized current. The dampinglike SOT is described as [28]

$$\vec{\tau}_{\text{SOT}} = \frac{\gamma\hbar\theta_{\text{SH}}}{2edM_s} \vec{m} \times [\vec{m} \times (\vec{e}_z \times \vec{J}_{\text{HM}})]. \quad (4)$$

Here, θ_{SH} is the spin Hall angle. $\vec{J}_{\text{HM}} = (J_{\text{HM}}, 0, 0)$ with J_{HM} as the current density in the HM layer. \vec{e}_z is the

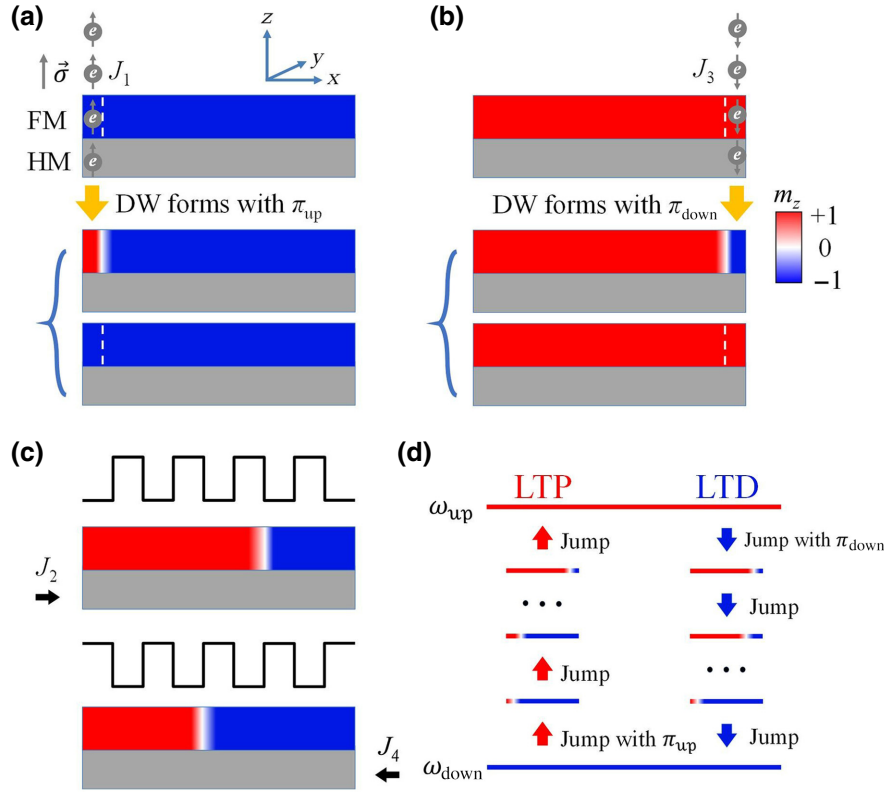


FIG. 1. Schematic of a stochastic DW memristor model. (a),(b) Probabilistic up- (down-) DW generation in a HM/FM heterostructure near the left (right) end of the track by STT, (c) DW motion under current pulses by SOT, (d) weight change mapped from resistance states. In the LTP (LTD) stage, weight increases (decreases) with probability π_{up} (π_{down}) owing to up- (down-) DW generation, and then it determinedly increases (decreases) under more of the same types of current stimuli.

unit vector along the thickness direction of the HM/FM bilayer.

The parameters for the simulation are as follows: the dimensions of the FM layer are $1024 \times 100 \times 1 \text{ nm}^3$ with a cell size of $4 \times 4 \times 1 \text{ nm}^3$. $T = 300 \text{ K}$ and $\alpha = 0.27$. $M_S = 9.1 \times 10^5 \text{ A m}^{-1}$ [30]. $\theta_{SH} = 0.4$ [31]. $P = 0.4$. The exchange stiffness constant (A) is $2 \times 10^{-11} \text{ J m}^{-1}$. The magnetic anisotropy constant (K) is $2 \times 10^6 \text{ J m}^{-3}$. The DMI constant (D) is 1.5 mJ m^{-2} .

III. RESULTS AND DISCUSSION

The electrical conductance of a DW memristor linearly changes with magnetization, which is proportional to the ratio between the red and blue areas, as indicated in Fig. 2(a). Figure 2(b) shows that the probability for DW formation can be controlled by changing J_1 or J_3 . Initially, magnetization points along the $-z$ ($+z$) direction, and the current pulse is perpendicularly injected at the left (right) end of the strip with a 1-ns width. The simulation is repeated 100 times. We calculate the probability for DW formation by dividing the time of DW formation by 100, and we find that at 300 K this probability as a function of J_1 or J_3 can be fitted by a sigmoidal function.

However, at 0 K, the DW is deterministically generated when the current density is higher than a critical value at $6.55 \times 10^{12} \text{ A m}^{-2}$ [Fig. 2(b)].

We consider a device with $m_z = -1$ as the initial state. After applying current J_1 , the DW may not form, so nothing will be driven, and the average m_z is kept constant under subsequent J_2 and J_4 [the lower line in Fig. 2(c)]. Once a DW is generated, the DW can be further driven. Here, we use four different J_2 and J_4 to drive a DW 10 times, and the average m_z and standard deviation (the error bar) are shown in Fig. 2(c). During the initial 0–3 ns, the current is along the $+x$ direction and the DW moves from left to right. During the subsequent 3–6 ns, the DW moves in the opposite direction with converse variation of m_z . The DW moves at a higher velocity under a larger current density with a smaller error bar. In addition to the stochasticity from thermal fluctuation, the edge roughness is another source of stochastic DW motion [18,19]. This edge roughness can be quantified by the random distribution of track width with a standard deviation σ_{SD} . Stochastic DW positions are seen at $\sigma_{SD} = 5 \text{ nm}$ in a 100-nm-wide track under a current density of $1.0 \times 10^{12} \text{ A m}^{-2}$. However, this stochastic DW position can be clearly depressed by moderately enhancing the current density to

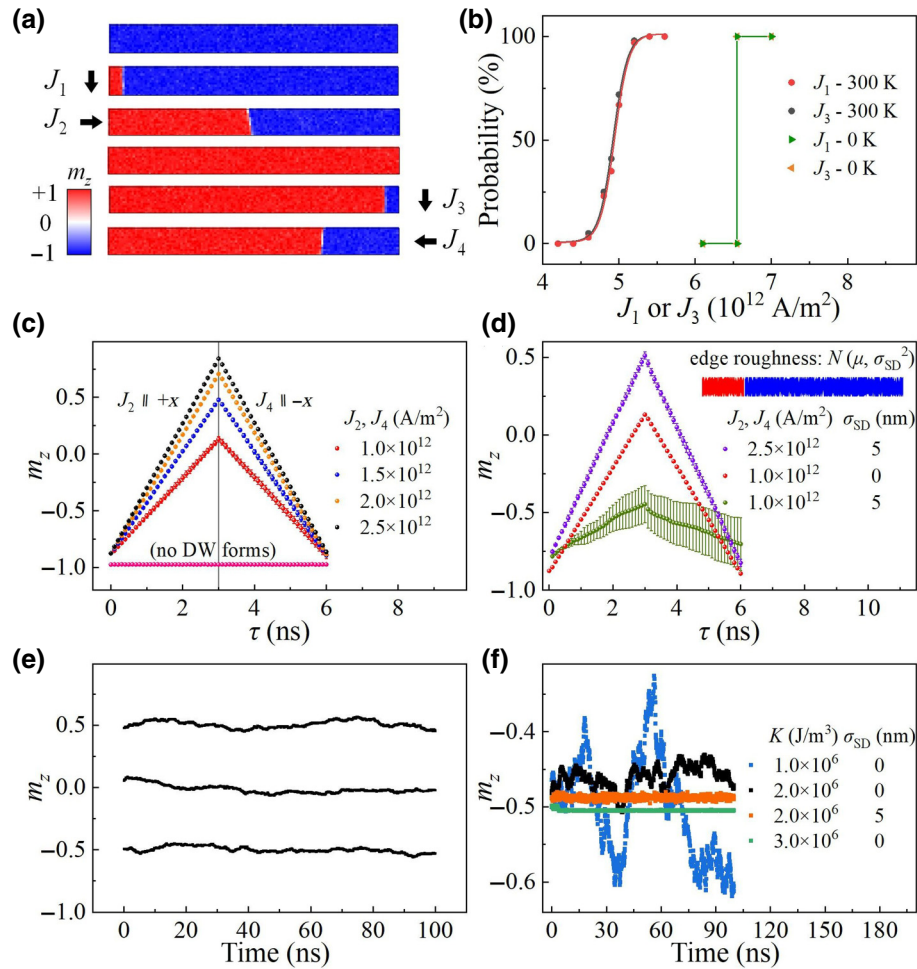


FIG. 2. Simulated stochastic DW generation and deterministic DW motion. (a) Snapshots of DW generation and motion under currents with various orientations. (b) Relationship between the probability of DW generation and current density J_1 or J_3 at 0 and 300 K. Probability of DW generation is estimated through repeated simulation 100 times at each current density, and that at 300 K can be well fitted by sigmoidal curves. (c),(d) Variation of m_z with respect to DW motion at different current densities and standard deviation (σ_{SD}) for edge roughness. (e),(f) Stability of the DW stopping at different positions, and the influence of anisotropy energy or edge roughness on DW stability.

2.5×10^{12} A m $^{-2}$ [Fig. 2(d)]. In a real application, the edge roughness can also be inhibited by exploiting an advanced fabrication technique like electron-beam lithography.

After driving a DW to different positions and removing the current, the DW stays for more than 100 ns, and the thermal effect causes the fluctuation of the DW position [Fig. 2(e)]. The stability of the DW position after removing the current is critical, since it determines the volatility and errors in the resistance manipulation. We show that this stability can be significantly improved by enhancing the anisotropy energy and edge roughness [Fig. 2(f)].

Based on the spin-torque-induced stochastic DW dynamics, we can construct a DW-based synapse with weight being stochastically adjusted. In a deterministic synapse, the weight is certain to be modulated when the strength of external stimulation exceeds a critical value. But if the weight adjustment is stochastic, the adjustable

states may be many more than that of a deterministic counterpart with the same device size. For example, for a deterministic binary memristor with two weights 0 and ω , a synapse made of N memristors has two states: 0 and $N \times \omega$. Nevertheless, if the weight adjustment is stochastic, there will be $N + 1$ possible weights: 0, ω , $2\omega \dots N \times \omega$ [32]. More generally, for a stochastic X -state memristor, the synapse made of N memristors can provide $N \times (X - 1) + 1$ weights.

We test the learning ability of a stochastic DW synapse by the simulation using the software “Python3.10.0.” The conductance, G , is mapped to the synaptic weight, ω , according to the equation $\omega = (G - G_{\min}) / (G_{\max} - G_{\min}) \times (\omega_{\text{up}} - \omega_{\text{dn}}) + \omega_{\text{dn}}$, where G_{\max} and G_{\min} mean the maximum and minimum conductance, respectively. Here, we set $\omega_{\text{down}} = 0$, $\omega_{\text{up}} = 1$, and $\pi_{\text{up}} = \pi_{\text{down}} = 0.5$. $n_{\text{up}} = 5$ ($n_{\text{down}} = 10$), which means the DW will reach the rightmost

(leftmost) end after 5 positive (10 negative) pulses in the LTP (LTD) procedure. During the test, 100 pulse pairs consisting of 80% LTP events and 20% LTD ones are first applied. Subsequently, another 100 pulse pairs made of 20% LTP events and 80% LTD ones are injected. The operation is repeated twice (green and pink lines in Fig. 3). One can see that, because of stochastic updating, the weight adjustments for the two training procedures can be different [Fig. 3(a)]. However, the twice-weight adjustments are the same for $\pi_{\text{up}} = \pi_{\text{down}} = 1$ [Fig. 3(b)].

Based on stochastic weight manipulation of the DW synapse, we further investigate its application in a SNN. Figure 4(a) shows the structure of a stochastic synapse composed of N parallel memristors between the input neuron, y_i , and the output one, z_k . The maximum weight is $\omega_{\text{max}} = N\omega_{\text{up}}$, and the minimum one is $\omega_{\text{min}} = N\omega_{\text{down}}$. Initially, no DW is generated, and the median weight is $(\omega_{\text{max}} + \omega_{\text{min}})/2$.

Current J_1 along the $+z$ direction is first applied at the left ends of all the memristors, and the up DW forms at probability π_{up} . Afterwards, current J_2 is applied to drive DWs along the $+x$ direction, which increases the total weight. The weight increase at each pulse can be adjusted by changing n_{up} . Finally, the magnetization becomes $+1$ and the weight reaches its maximum value, ω_{up} . In the LTD stage, current J_3 along the $-z$ direction is applied at the right end, while the down DW is generated only in the memristors with $m_z = 1$. After that, current J_4 triggers all DWs to move along the $-x$ direction. Owing to the stochastic DW generation and deterministic DW motion, the weight change includes not only the newly generated DWs, but also all the DWs formed in the previous stages.

Quantitatively, the stochastic weight updating under each current pulse is related to the number variation of the memristors in different states. In a LTP process, the state is expressed as $\omega_P = 0, \Delta\omega/n_{\text{up}}, 2 \times \Delta\omega/n_{\text{up}} \dots n_{\text{up}} \times \Delta\omega/n_{\text{up}}$, corresponding to states labeled by $j = 0, 1, 2, 3 \dots n_{\text{up}}$. Here, $\Delta\omega = \omega_{\text{up}} - \omega_{\text{down}}$, which indicates the weight difference between $m_z = 1$ and $m_z = -1$. We denote the number of memristors in the j th state as m_j .

The LTP weight updating for the $j = 1$ state is $\Delta W_1 = (m_0\pi_{\text{up}} - m_1)(\Delta\omega/n_{\text{up}})$. Here, the first term indicates the newly increased memristors at the $j = 1$ state from stochastic DW generation in m_0 memristors. The second term describes the weight adjustment from the m_1 memristors with the transition from the $j = 1$ state to the $j = 2$ one by deterministic DW motion. When $n_{\text{up}} > j > 1$, $\Delta W_j = (m_{j-1} - m_j)(j \Delta\omega/n_{\text{up}})$, which includes the transition from the $(j-1)$ th state to the j th one of the m_{j-1} memristors and that from the j th state to the $(j+1)$ th one of the m_j memristors based on deterministic DW motion. For the largest $j = n_{\text{up}}$, $\Delta W_{n_{\text{up}}} = m_{n_{\text{up}}}\Delta\omega$. The total weight adjustment in the LTP procedure is $\Delta W_{\text{LTP}} = \sum_{j=1}^{n_{\text{up}}} \Delta W_j$.

Similar weight modification in the LTD process can also be derived by considering a different initial state.

Based on the above discussion, the total weight adjustment satisfies the dynamics equation:

$$\frac{d}{dt} W_{ki} = s_k(t)[m_P T_P \omega_P y_i(t) + m_D T_D \omega_D (1 - y_i(t))]. \quad (5)$$

Here, $s_k(t)$ and $y_i(t)$ are the spiking signals from the post- and preneurons, respectively. The first term at the right-hand side of Eq. (5) depicts the LTP process with $m_P = [m_0 \dots m_j \dots m_{n_{\text{up}}}]$,

$$T_P = \begin{bmatrix} -\pi_{\text{up}} & \pi_{\text{up}} & 0 & 0 & \dots & 0 \\ 0 & -1 & 1 & 0 & \dots & 0 \\ 0 & 0 & -1 & 1 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 1 \\ 0 & 0 & \dots & \dots & \dots & 0 \end{bmatrix},$$

and

$$\omega_P = \begin{bmatrix} 0 \\ \Delta\omega/n_{\text{up}} \\ 2 \times \Delta\omega/n_{\text{up}} \\ \vdots \\ n_{\text{up}} \times \Delta\omega/n_{\text{up}} \end{bmatrix}.$$

The second term at the right-hand side of Eq. (5) describes the LTD process with $m_D = [m_0 \dots m_j \dots m_{n_{\text{down}}}]$,

$$T_D = \begin{bmatrix} 0 & 0 & 0 & 0 & \dots & 0 \\ 1 & -1 & 0 & 0 & \dots & 0 \\ 0 & 1 & -1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & -1 & 0 \\ 0 & 0 & \dots & \dots & \pi_{\text{down}} & -\pi_{\text{down}} \end{bmatrix},$$

and

$$\omega_D = \begin{bmatrix} 0 \\ \Delta\omega/n_{\text{down}} \\ 2 \times \Delta\omega/n_{\text{down}} \\ \vdots \\ n_{\text{down}} \times \Delta\omega/n_{\text{down}} \end{bmatrix}.$$

It is noticed that if $\pi_{\text{up}} = \pi_{\text{down}} = 0$, $m_P = [N/2 \ 0 \ \dots \ 0]$, and $m_P \times T_P = [0 \ \dots \ 0]$, so does the LTD term. Obviously, the weight cannot update. When $\pi = \pi_{\text{up}} = \pi_{\text{down}} = 1$, the DWs will be simultaneously generated in all the memristors. Under these circumstances, $m_j = N_j$ is the number of memristors for the j th state and

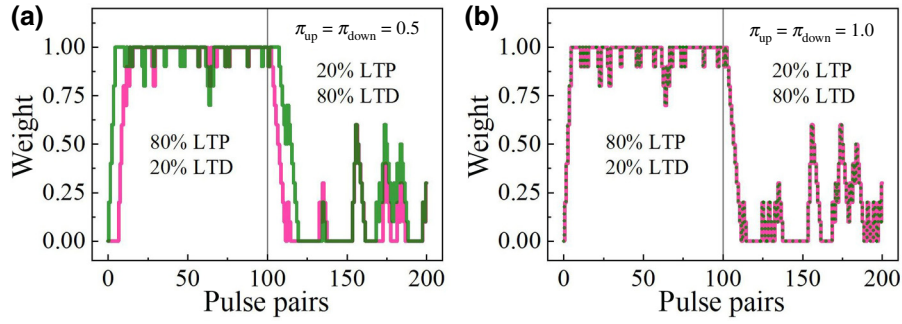


FIG. 3. Comparison of weight tuning between (a) a stochastic DW synapse with $\pi_{\text{up}} = \pi_{\text{down}} = 0.5$ and (b) a deterministic one. In a STPD pairing simulation, 200 programming pulses are applied to a single synapse, where during the first half 80% potentiative and 20% depressive pulses are imposed stochastically, while during the second half the ratio of the two types of pulses is reversed. In both (a),(b), the training procedures are repeated twice (green and pink lines).

0 for the other states. Equation (5) can be converted into

$$\frac{d}{dt}W_{ki} = N_j \Delta\omega S_k(t) \left[\frac{y_i(t)}{n_{\text{up}}} + \frac{[1 - y_i(t)]}{n_{\text{down}}} \right].$$

Compared with this deterministic circumstance, the introduction of probability significantly improves the flexibility of the weight manipulation for a synapse.

As an example in an application, we implement a stochastic DW synapse in a SNN to classify a breast tumor. We choose Wisconsin breast cancer data for unsupervised learning by the software Python3.10.0. This dataset consists of 699 instances with 9 features [33]: clump thickness, uniformity of cell size, uniformity of cell shape, marginal adhesion, single epithelial cell size, bare nuclei, bland chromatin, normal nucleoli, and mitoses. In our neuron network, each feature value (1–10) is encoded by 10 Poisson neurons y_i ($0 \leq i \leq 9$). Therefore, 90 input neurons are exploited. Among all the 699 instances, 399 are randomly selected for training and the rest are for testing.

Figure 4(b) shows the SNN model for this application. It consists of 2 layers: an input layer (90 neurons) and an output one (2 neurons with mutual inhibition). When a feature value is input, the corresponding input neurons fire at a spiking rate that is proportional to the feature value, while other input neurons stay silent. The leaky-integrate-fire model is adopted for the output neurons [34]. The input current through the synapse integrates as the membrane potential of the output neuron, and the accumulated potential continuously leaks a small amount. When the membrane potential reaches the threshold voltage, a postneuron fires, and the potential is then reset. The weight adjustment by a simplified STDP rule is also illustrated in Fig. 4(b). After preneuron y_i spiking, if postneuron z_k is spiked within the time window, potentiation occurs and the weight, W_{ki} , between y_i and z_k increases by $\Delta\omega/n_{\text{up}}$. Instead, if z_k is spiked out of the time window, W_{ki} decreases by $\Delta\omega/n_{\text{down}}$.

To effectively classify a tumor, a moderate probability between 0 and 1 is necessary for this application. If $\pi = 0$, since no DW forms, the network cannot work. When $\pi = 1$, the network fails to work either, because there will be no difference between two outputs for the same inputs and two identical deterministic synapses with the same initial weight (zero-DW state). Once an input signal accumulates to the threshold, both output neurons spike simultaneously, and both weights update without any difference. Thus, it cannot recognize data well.

We consider a SNN with seven ($N = 7$) HM/FM memristors in one synapse ($\omega_{\text{max}} = 0.04$ and $\omega_{\text{min}} = 0$). The weight of each device varies from 0 to ω_{max}/N . Among the seven devices, the initial m_z is +1 in four devices and -1 in the other three. $n_{\text{up}} = 15$; $n_{\text{down}} = 90$, and $\pi_{\text{up}} = \pi_{\text{down}} = 0.1\text{--}0.9$. After 3 epochs, the accuracy of the network converges to equilibrium. The average accuracy of the last 7 epochs can be tuned by changing the probability ($\pi = \pi_{\text{up}} = \pi_{\text{down}}$) [Fig. 4(c)]. The SNN of this synapse is able to classify the breast cancer dataset with an accuracy exceeding 90% when the probability is 0.1–0.9. The highest accuracy reaches 95.7% at $\pi = 0.7$. On the other hand, the difference between n_{up} and n_{down} also influences the accuracy. The accuracy varies around 90% by changing the ratio between n_{up} and n_{down} at $\pi = 0.5$ [Figs. 4(d) and 4(e)].

Finally, energy consumption (Joule heat) of the DW-based synapse system is estimated by $E_J = I^2 R t$, with I , R , and t as the current, resistance, and time, respectively. Therefore, energy consumption is influenced by current density, the size and resistivity of the device, and the velocity of DW motion. For SOT-induced DW motion, we calculate the energy consumption in a heavy metal layer, such as W, with a resistivity of $30 \times 10^{-8} \Omega\text{m}$ [35] (The shunting current in the ultrathin FM metal layer is generally neglected [19].) We find that in a W layer with dimensions of $200 \times 20 \times 5 \text{ nm}^3$, when the current density for driving the DW is around $3 \times 10^{11} \text{ A m}^{-2}$, the DW can move by 200 nm within 3 ns. This gives rise

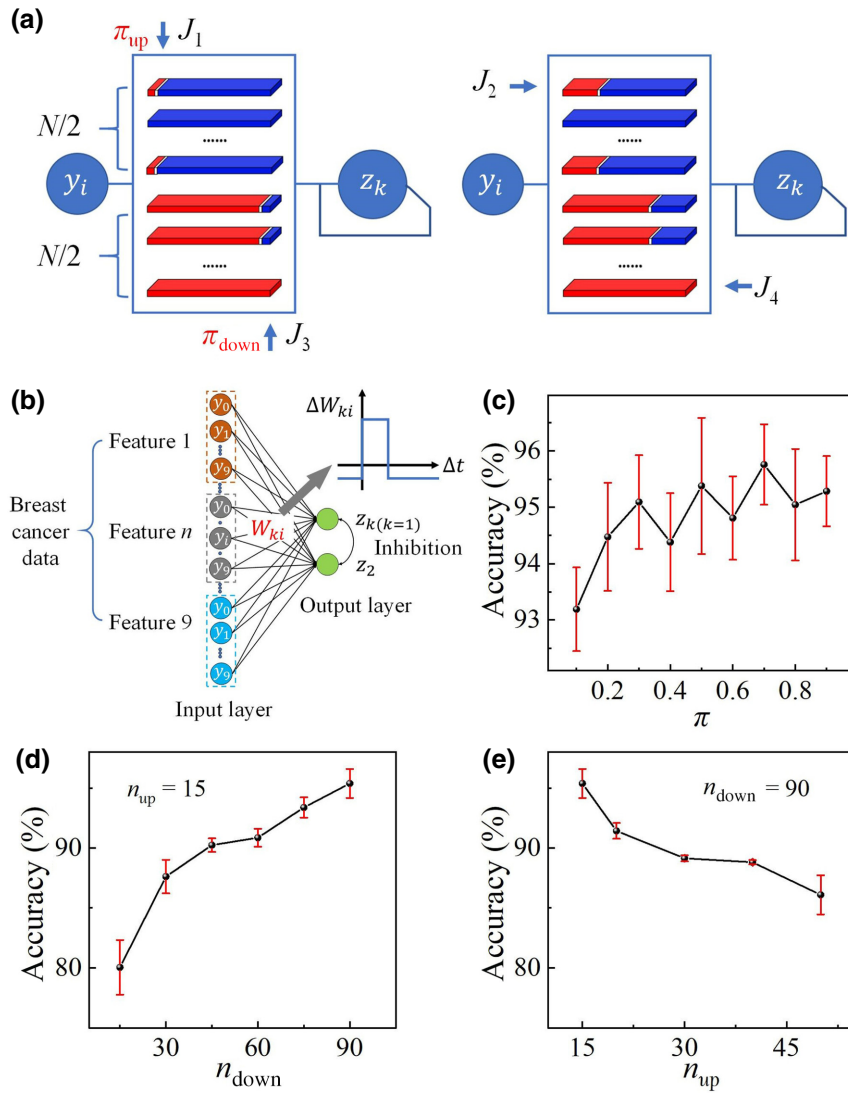


FIG. 4. SNN with stochastic DW synapses and its application for medical diagnosis. (a) Stochastic synapse composed of N nanostrips between input neuron y_i and output one z_k . (b) SNN constructed by the stochastic DW synapses connecting input neurons and output ones, and its application in the classification of the Wisconsin breast cancer data [33]. Inset shows the weight updates through the probabilistic STDP rule. (c) Accuracy of the classification after 10-epoch learning. Dots represent the average accuracy of the last 7 epochs, while red error bars denote standard deviations. (d) Accuracy of the classification as a function of n_{down} at a fixed n_{up} ($n_{\text{up}} = 15$) and (e) n_{down} at a fixed n_{up} ($n_{\text{down}} = 90$).

to energy consumption less than 2 fJ. In DW generation by STT, the current is injected in a region with a dimension between 200 and 2000 nm² with a current density around 5×10^{12} A m⁻², and the energy consumption is at a magnitude of tens of fJ. This is slightly higher than the voltage-assisted DW-based synapse [19] but still much smaller than traditional synapse devices at a magnitude of pJ [36].

IV. CONCLUSION

We propose a multistate nonvolatile artificial synapse based on spin-torque-induced probabilistic generation and deterministic motion of magnetic DWs. Because of stochastic DW generation at 300 K, the weight can be adjusted over a wide range with very few memristors. As a typical example of an application, this stochastic synapse model is used in a SNN to classify breast cancer, and the highest accuracy reaches 95.7%.

ACKNOWLEDGMENTS

The authors acknowledge financial support from the National Key Research and Development Program of China (Grant No. 2022YFE0103300) and the National Natural Science Foundation of China (Grants No. 51971098 and No. 61974051).

- [1] D. Silver, *et al.*, Mastering the game of Go with deep neural networks and tree search, *Nature* **529**, 484 (2016).
- [2] E. Begoli, T. Bhattacharya, and D. Kusnezov, The need for uncertainty quantification in machine-assisted medical decision making, *Nat. Mach. Intell.* **1**, 20 (2019).
- [3] Z. Ghahramani, Probabilistic machine learning and artificial intelligence, *Nature* **521**, 452 (2015).
- [4] T. Tuma, A. Pantazi, M. Le Gallo, A. Sebastian, and E. Eleftheriou, Stochastic phase-change neurons, *Nat. Nanotechnol.* **11**, 693 (2016).
- [5] K. Wang, Q. Hu, B. Gao, Q. Lin, F.-W. Zhuge, D.-Y. Zhang, L. Wang, Y.-H. He, R. H. Scheicher, H. Tong, and

- X.-S. Miao, Threshold switching memristor-based stochastic neurons for probabilistic computing, *Mater. Horiz.* **8**, 619 (2021).
- [6] J. Zhu, T. Zhang, Y. Yang, and R. Huang, A comprehensive review on emerging artificial neuromorphic devices, *Appl. Phys. Rev.* **7**, 011312 (2020).
- [7] J. Grollier, D. Querlioz, K. Y. Camsari, K. Everschor-Sitte, S. Fukami, and M. D. Stiles, Neuromorphic spintronics, *Nat. Electron.* **3**, 360 (2020).
- [8] X. Chen, W. Kang, D. Zhu, X. Zhang, N. Lei, Y. Zhang, Y. Zhou, and W. Zhao, A compact skyrmionics leaky-integrate-fire spiking neuron device, *Nanoscale* **10**, 6139 (2018).
- [9] X. Liang, X. Zhang, J. Xia, M. Ezawa, Y. Zhao, G. Zhao, and Y. Zhou, A spiking neuron constructed by the skyrmion-based spin torque nano-oscillator, *Appl. Phys. Lett.* **116**, 122402 (2020).
- [10] S. Li, W. Kang, Y. Huang, X. Zhang, Y. Zhou, and W. Zhao, Magnetic skyrmion-based artificial neuron device, *Nanotechnology* **28**, 31LT01 (2017).
- [11] Y. Huang, W. Kang, X. Zhang, Y. Zhou, and W. Zhao, Magnetic skyrmion-based synaptic devices, *Nanotechnology* **28**, 08LT02 (2017).
- [12] R. Chen, C. Li, Y. Li, J. J. Miles, G. Indiveri, S. Furber, V. F. Pavlidis, and C. Moutafis, Nanoscale Room-Temperature Multilayer Skyrmionic Synapse for Deep Spiking Neural Networks, *Phys. Rev. Appl.* **14**, 014096 (2020).
- [13] Z. Yu, M. Shen, Z. Zeng, S. Liang, Y. Liu, M. Chen, Z. Zhang, Z. Lu, L. You, X. Yang, Y. Zhang, and R. Xiong, Voltage-controlled skyrmion-based nanodevices for neuromorphic computing using a synthetic antiferromagnet, *Nanoscale* **2**, 1309 (2020).
- [14] S. Lequeux, J. Sampaio, V. Cros, K. Yakushiji, A. Fukushima, R. Matsumoto, H. Kubota, S. Yuasa, and J. Grollier, A magnetic synapse: Multilevel spin-torque memristor with perpendicular anisotropy, *Sci. Rep.* **6**, 31510 (2016).
- [15] S. Liu, T. P. Xiao, C. Cui, J. A. C. Incorvia, C. H. Bennett, and M. J. Marinella, A domain wall-magnetic tunnel junction artificial synapse with notched geometry for accurate and efficient training of deep neural networks, *Appl. Phys. Lett.* **118**, 202405 (2021).
- [16] W. A. Misba, M. Lozano, D. Querlioz, and J. Atulasimha, Energy efficient learning with low resolution stochastic domain wall synapse for deep neural networks, *IEEE Access* **10**, 84946 (2022).
- [17] S. A. Siddiqui, S. Dutta, A. Tang, L. Liu, C. A. Ross, and M. A. Baldo, Magnetic domain wall based synaptic and activation function generator for neuromorphic accelerators, *Nano. Lett.* **20**, 1033 (2019).
- [18] M. Sharad, C. Augustine, G. Panagopoulos, and K. Roy, Spin-based neuron model with domain-wall magnets as synapse, *IEEE Trans. Nanotechnol.* **11**, 843 (2012).
- [19] W. A. Misba, T. Kaisar, D. Bhattacharya, and J. Atulasimha, Voltage-controlled energy-efficient domain wall synapses with stochastic distribution of quantized weights in the presence of thermal noise and edge roughness, *IEEE Trans. Electron Devices* **69**, 1658 (2021).
- [20] S. S. P. Parkin, M. Hayashi, and L. Thomas, Magnetic domain-wall racetrack memory, *Science* **320**, 190 (2008).
- [21] J. C. Slonczewski, Current-driven excitation of magnetic multilayers, *J. Magn. Magn. Mater.* **159**, L1 (1996).
- [22] D. Bedau, H. Liu, J. Z. Sun, J. A. Katine, E. E. Fullerton, S. Mangin, and A. D. Kent, Spin-transfer pulse switching: From the dynamic to the thermally activated regime, *Appl. Phys. Lett.* **97**, 262502 (2010).
- [23] S. Emori, U. Bauer, S.-M. Ahn, E. Martinez, and G. S. D. Beach, Current-driven dynamics of chiral ferromagnetic domain walls, *Nat. Mater.* **12**, 611 (2013).
- [24] R. Ramaswamy, J.-M. Lee, K. Cai, and H. Yang, Recent advances in spin-orbit torques: Moving towards device applications, *Appl. Phys. Rev.* **5**, 031107 (2018).
- [25] W. Wang, G. Pedretti, V. Milo, R. Carboni, A. Calderoni, N. Ramaswamy, A. S. Spinelli, and D. Ielmini, Learning of spatiotemporal patterns in a spiking neural network with resistive switching synapses, *Sci. Adv.* **4**, eaat4752 (2018).
- [26] D. Chiba, M. Kawaguchi, S. Fukami, N. Ishiwata, K. Shimamura, K. Kobayashi, and T. Ono, Electric-field control of magnetic domain-wall velocity in ultrathin cobalt with perpendicular magnetization, *Nat. Commun.* **3**, 1 (2012).
- [27] K. Kondou, N. Ohshima, S. Kasai, Y. Nakatani, and T. Ono, Single shot detection of the magnetic domain wall motion by using tunnel magnetoresistance effect, *Appl. Phys. Express* **1**, 061302 (2008).
- [28] 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, and W. Zhao, Field-free switching of a perpendicular magnetic tunnel junction through the interplay of spin-orbit and spin-transfer torques, *Nat. Electron.* **1**, 582 (2018).
- [29] W. F. Brown Jr, Thermal fluctuations of a single-domain particle, *Phys. Rev.* **130**, 1677 (1963).
- [30] T. Herranen and L. Laurson, Barkhausen noise from precessional domain wall motion, *Phys. Rev. Lett.* **122**, 117205 (2019).
- [31] S. Mondal, S. Choudhury, N. Jha, A. Ganguly, J. Sinha, and A. Barman, All-optical detection of the spin Hall angle in W/CoFeB/SiO₂ heterostructures with varying thickness of the tungsten layer, *Phys. Rev. B* **96**, 054414 (2017).
- [32] J. Bill and R. Legenstein, A compound memristive synapse model for statistical learning through STDP in spiking neural networks, *Front. Neurosci.* **8**, 412 (2014).
- [33] W. H. Wolberg, W. N. Street, and O. L. Mangasarian, Machine learning techniques to diagnose breast cancer from image-processed nuclear features of fine needle aspirates, *Cancer Lett.* **77**, 163 (1994).
- [34] W. H. Brigner, N. Hassan, L. Jiang-Wei, X. Hu, D. Saha, C. H. Bennett, M. J. Marinella, J. Anne, C. Incorvia, F. Garcia-Sanchez, and J. S. Friedman, Shape-based magnetic domain wall drift for an artificial spintronic leaky integrate-and-fire neuron, *IEEE Trans. Electron Devices* **66**, 4970 (2019).
- [35] D. Choi, B. Wang, S. Chung, X. Liu, A. Darbal, A. Wise, N. T. Nuhfer, K. Barmak, A. P. Warren, K. R. Coffey, and M. F. Toney, Phase, grain structure, stress, and resistivity of sputter-deposited tungsten films, *J. Vac. Sci. Technol., A* **29**, 051512 (2011).
- [36] D. Kaushik, U. Singh, U. Sahu, I. Sreedevi, and D. Bhowmik, Comparing domain wall synapse with other non volatile memory devices for on-chip learning in analog hardware neural network, *AIP Adv.* **10**, 025111 (2020).