Voltage-Controlled Spintronic Stochastic Neuron Based on a Magnetic Tunnel Junction

Jialin Cai,[1,2] Bin Fang,[1] Like Zhang,[1,2] Wenxing Lv,[1,2] Baoshun Zhang,[1,2] Tiejun Zhou,[3]
Giovanni Finocchio,[4] and Zhongming Zeng[1,2,*]

[1] *Key Laboratory of Multifunctional Nanomaterials and Smart Systems, Suzhou Institute of Nano-Tech and Nano-Bionics, CAS, Suzhou, Jiangsu 215123, People's Republic of China*

[2] *School of Nano Technology and Nano Bionics, University of Science and Technology of China, Hefei, Anhui 230026, People's Republic of China*

[3] *College of Electronics and Information, Hangzhou Dianzi University, Hangzhou 310018, People's Republic of China*

[4] *Department of Mathematical and Computer Sciences, Physical Sciences and Earth Sciences, University of Messina, Messina 98166, Italy*

(Received 12 November 2018; revised manuscript received 19 January 2019; published 6 March 2019)

Stochastic units based on magnetic tunnel junctions have shown a high energy-efficient pathway to perform neuromorphic computing. We propose a voltage-controlled spintronic stochastic device based on a magnetic tunnel junction by introducing perpendicular anisotropy into the free layer. We observe a stochastic behavior at low bias and demonstrate that this behavior can be used to mimic the artificial neurons of an artificial neural network to recognize the handwritten digits in the Mixed National Institute of Standards and Technology (MNIST) database. Furthermore, the stochastic behavior can be modulated by a bias voltage owing to the voltage-controlled magnetic anisotropy effect. The voltage-controlled stochastic behavior is theoretically and experimentally studied, which indicates that it can be used to perform as an adaptive neuron. These results provide a way for building energy-efficient spintronic neuromorphic computing systems.

DOI: 10.1103/PhysRevApplied.11.034015

I. INTRODUCTION

While the current computing schemes face severe challenges of data processing speed and scalability, the human brain shows high efficiency and energy saving in performing cognitive and perception tasks. This ability relies on the protocol of learning and storage implemented by synapses and neurons in the brain. Recently, the utilization of hardware to mimic the functionalities involved in the neurons and synapses in the human brain has become an appealing subject, referred to as neuromorphic computing. Neuromorphic computing algorithms can be roughly divided into artificial neural networks (ANNs) and spiking neural networks (SNNs). Although the ANNs have already achieved great success at complex recognition problems [1,2], the computational costs are still much larger than that of the human brain. Many research efforts have been made to develop more biologically realistic SNN hardware systems to build up more powerful and energy-efficient neuromorphic computing systems [3–5]. For example, in the popular neuromorphic system, TrueNorth, built with one million neurons and 256 million synapses, the power density is 20 mW per cm^2, compared with 50–100 W per cm^2 of a typical central processing unit (CPU) [5]. However, it still consumes high energy and takes up a large area. In order to reduce the energy and area requirements, several post-CMOS technologies, which are able to offer a compact and energy-efficient implementation for neural and synaptic units, such as memristor [6–8], phase-change memory [9], and spintronic devices [10,11], have been proposed. Magnetic tunnel junctions (MTJs), the building blocks of spintronic devices [12–14], which possess the advantages of high working speeds, high endurance, low power consumption, and high integration capability, have recently been proved for wide applications in neuromorphic computing [15–18], such as mimicking multilevel synapses with a spin-torque memristor [17,19,20], emulating biological neurons with a spin-torque oscillator [18,21], and spiking neurons with spin-torque stochastic switching [15,22–28]. It is worth noting that the stochastic-switching MTJs have been able to act as artificial neurons in ANNs, which not only possess the high recognition performance of ANNs, but also consume low area and energy, as do SNNs [23,29].

Recently, superparamagnetic tunnel junctions have been proposed and show the capability of ultralow-energy

*jlcai2015@sinano.ac.cn

 © 2019 American Physical Society

neuromorphic computing [15,30]. The high energy efficiency results from the thermal fluctuations-induced stochastic magnetization switching at room temperature. Thus, no write operations are required and only low-energy readout operations are needed to output the random bits. Although the superparamagnetic tunnel junctions have the ability to compute with ultra-low-energy consumption, a highly efficient way to control the stochastic behavior is still lacking.

In this work, we report a voltage-controlled spintronic stochastic neuron based on MJT by introducing perpendicular magnetic anisotropy into the free layer through turning the thickness of the free layer to lower the energy barrier. We observe the stochastic switching under a low voltage bias and demonstrate that the switching probability in different magnetic fields of the MTJ can be used to mimic the activation function of an artificial neuron of ANNs. An ANN using the obtained activation function is constructed to recognize the handwritten digits in the Mixed National Institute of Standards and Technology (MNIST) database [31]. The activation function of the device shows controllability by the bias voltage, indicating a way to build up an all-spintronic neural network.

II. EXPERIMENTS

The MTJ multilayers are deposited by magnetron sputtering with a layered structure of bottom electrode/$Co_{20}Fe_{60}B_{20}$ (1.62 nm)/MgO (0.8 nm)/$Co_{40}Fe_{40}B_{20}$ (2.4 nm)/Ru (0.85 nm)/CoFe (2.3 nm)/PtMn (15 nm)/top electrode (thicknesses in nanometers) as illustrated in Fig. 1(a). The films are annealed for 2.0 h at a temperature of 300 °C in a magnetic field of 1 T. The interfacial perpendicular magnetic anisotropy (IPMA) is introduced by adjusting the thickness of the $Co_{40}Fe_{40}B_{20}$ free layer [12,32]. This IPMA has been successfully used for the design of high-efficiency spintronic devices such as decreasing the switching current of MTJs for storage applications [33], enlarging the emitted power of the spin-torque nano-oscillators (STNOs) [34], and increasing the sensibility of spin-torque diodes (STDs) [14,35]. Electron-beam lithography and ion milling are used to fabricate the pillar-shaped devices with an elliptical cross section. All data in the present work are from one 210×60 nm^2 device. The transport characterizations are performed with a current source (Keithley 2400), a nanovoltmeter (Keithley 2182), and an oscilloscope (Tektronix DPO73304D) at room temperature.

III. RESULTS AND DISCUSSION

First, we measure the magnetoresistance of the device. Figure 1(b) shows the resistance as a function of the in-plane magnetic field applied parallel to the magnetic easy axis of the nanopillar ($H_{||}$), at a bias current of $I_{dc} = 10\ \mu$A. As the field increases from −650 to +650 Oe, the resistance changes gradually from a low-resistance state to a high-resistance state, indicating the free-layer magnetization aligns antiparallel to the reference layer magnetization. The gradual change in the resistance confirms that the free layer is tilted out-of-plane from the easy axis. The resistance curve scan as a function of the out-of-plane field ($H_\perp$) [inset of Fig. 1(b)] also indicates a perpendicular free layer. At zero field, a small tilting angle (θ) from the out-of-plane configuration is estimated to be 9° through the resistance vs magnetic field relationship [35,36] due to the coupling between the free and reference layers.

Next, we study the low power stochastic behaviors of the device. The device investigated is a bistable system in which the two stable states are those characterized by low- and high-resistance states. As shown in Fig. 2(a), when we sweep the in-plane magnetic field with a small step $\Delta H = 1$ Oe at a dc voltage bias of 1 mV, telegraphic switching appears near zero field, and the switching probability can be tuned by an in-plane magnetic field. A similar phenomenon occurs under a perpendicular magnetic field, as shown in Fig. 2(b). Since the reference layer has the in-plane magnetization, and the applied in-plane magnetic field will change the resistance of the device, we choose to study the phenomenon that appears under the perpendicular magnetic field. If we fix the magnetic field and read the resistance of the device at 1 mV, we can obtain the

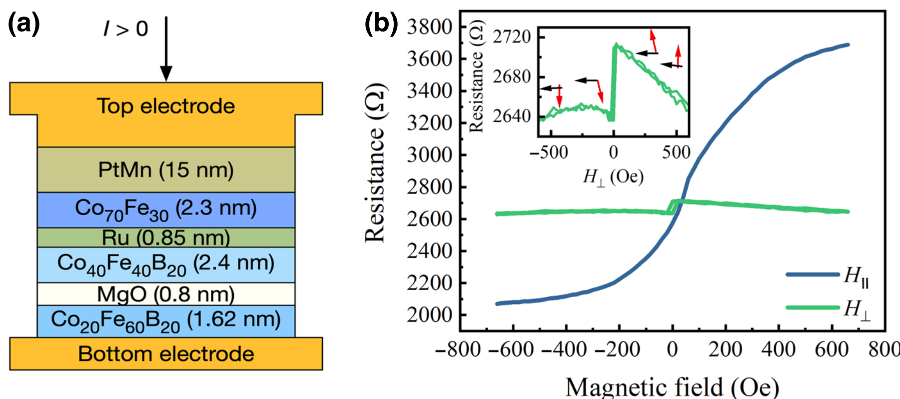

FIG. 1. (a) Schematic of the device structure. (b) The magnetoresistance curve of the MTJ device under in-plane magnetic field ($H_{||}$) and perpendicular magnetic field ($H_\perp$) for dc (I_{dc}) of 10 μA. The resistance scan as a function of the out-of-plane field (inset of b) clearly indicates the perpendicular free layer. The black (red) arrow denotes the magnetization direction of the reference (free) layer.

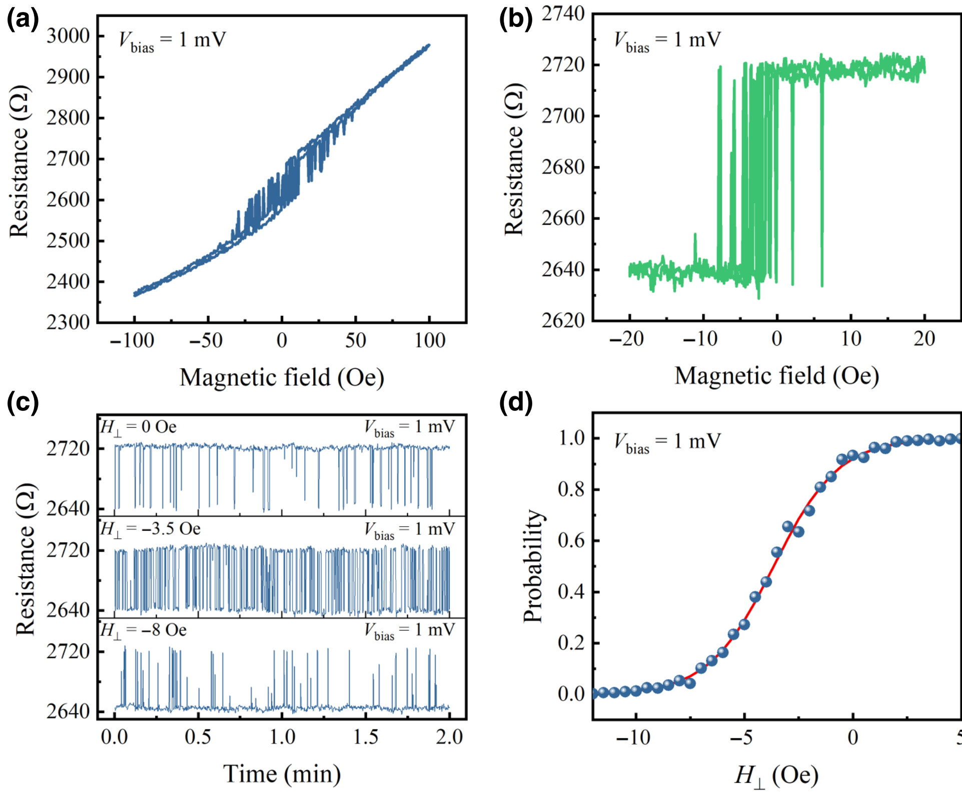

FIG. 2. Minor magnetoresistance curve of the MTJ device with step $\Delta H = 1$ Oe at voltage bias $V_{\text{bias}} = 1$ mV under (a) in-plane magnetic field ($H_{||}$) and (b) perpendicular magnetic field ($H_\perp$). (c) Time-domain stochastic switching characteristic of the MTJ device at $V_{\text{bias}} = 1$ mV in the absence of bias magnetic field. (d) Switching probability under various external perpendicular magnetic fields ($H_\perp$) at $V_{\text{bias}} = 1$ mV (blue dot). The data is fitted by a sigmoid probability density function (red line).

time-domain stochastic output of the device. Figure 2(c) shows the stochastic output performance of the device under zero magnetic field at 1 mV by sampling 1000 times in 2 min. These stochastic behaviors may occur because of the low energy barrier resulting from the introduction of the perpendicular magnetic anisotropy to the free layer. The energy barrier E_b of a tilted free layer can be expressed as [33,36,37]

$$\begin{aligned} E_b(V=0) &= K_{0,\text{eff}} V^* \\ &= (K_{\text{shape}} + K_\perp - K_d) V^*, \end{aligned} \tag{1}$$

where $K_{0,\text{eff}}$ is the effective energy density without voltage bias, K_{shape} is the in-plane shape-induced anisotropy energy density, $K_\perp$ is the perpendicular anisotropy energy density, and K_d is the out-of-plane demagnetizing energy density. According to Eq. (1), the energy barrier is related to the effective energy and the volume of the magnetic free layer, V^*. Thus, there are two ways to reduce the energy barrier. One is to reduce the volume of the device, which is unchangeable after the fabrication. The other is to lower the effective energy. Since $K_d \gg K_{\text{shape}}$, increasing the perpendicular anisotropy is able to reduce the energy barrier. Thus, with the introduction of an appropriate perpendicular anisotropy for the free layer that is able to compensate the out-of-plane demagnetization field, we can make the effective energy barrier as small as possible, leading to it behaving superparamagnetically without scaling down the device size.

If we define the low-resistance readout as the basic state, the switching probability can be easily calculated by $P = N_{\text{high}}/N_{\text{total}}$, where the N_{high} is the number of the high-resistance readout and N_{total} is the total sample number. By changing the perpendicular magnetic field and repeating the measurement, the switching probability of the device under different perpendicular magnetic fields can be obtained, as shown in Fig. 2(d). The switching probability can be described as [38]

$$P(\text{high}|H_\perp) = \frac{\tau_{\text{high}}}{\tau_{\text{low}} + \tau_{\text{high}}}, \tag{2}$$

where τ_{high} and τ_{low} are the dwell times from the low-to-high and high-to-low resistance states, respectively. Using the Neel-Brown formula, the dwell times can be expressed as [39]

$$\tau_{\text{high}} = \tau_0 \exp\left[\frac{E_b}{k_B T}\left(1 + \frac{H_\perp - H_s}{H_{k,\text{eff}}}\right)^2\right], \tag{3}$$

$$\tau_{\text{low}} = \tau_0 \exp\left[\frac{E_b}{k_B T}\left(1 - \frac{H_\perp - H_s}{H_{k,\text{eff}}}\right)^2\right], \tag{4}$$

where τ_0 is the inverse of the attempt frequency (assumed to be 1 ns) and $H_{k,\text{eff}}$ is the effective magnetic anisotropy

field. Substitution of Eqs. (3) and (4) into Eq. (2) leads to

$$P(\text{high}|H_\perp) = \frac{1}{1+\exp\left(-4\frac{E_b}{k_BT}\frac{H_\perp - H_s}{H_{k,\text{eff}}}\right)} = \frac{1}{1+\exp[-w(H_\perp - H_s)]}, \quad (5)$$

where w is the weight depending on the energy barrier, E_b. According to Eq. (5), the switching probability of the device can be fitted with the sigmoid function [$f(x) = (1/1+e^{-x})$], which is a normal kind of the activation function of the artificial neurons of ANNs, implying that our device can be used to mimic the artificial neurons of ANNs. When all signals input to a neuron, the neuron sums up all the input signals with their weights to decide whether or not to activate.

To explore the neuromorphic computing performance of our device, we build up a simple three-layer neural network with our devices as the stochastic spintronic neurons to solve the MNIST handwriting digit reorganization problem. The structure of the neural network is shown in Fig. 3(a), including one input layer, one hidden layer, and one output layer. All the images in the data set have 28 × 28 pixels, so the input layer has 784 neurons. The input signals are the perpendicular magnetic fields applied to each device, which is converted from the summed-up signals with weights by a linear transformation. The output voltage is converted to a current passing through a wire to produce an Oersted field that is applied to the next neurons. It can also apply the output voltage to a ferroelectric substrate under the MTJ, which can tune the effective field by modulating the strain from the ferroelectric substrate. All the images represent the handwriting digits from 0 to 9, and the output layer is set with 10 neurons. The hidden layer has 400 neurons. We use the back-propagation algorithm to train the network for 100 training epochs. Figure 3(b) shows the recognition rate vs training epochs of our stochastic SNN, together with the results from the developed ANN as a comparison. It can be seen that the spintronic neuron can work as well as the artificial sigmoid neuron [23,40] at a hardware level. At the beginning of the training process, the network with spintronic neurons is training faster. This may result from the linear transformation of the input signal, which is similar to the normalization of the input data that is a widely used technique to accelerate the network training [41]. The lower recognition rate at the final state of the spintronic ANN is given by the slight device variation in the switching probability for each neuron induced by the magnetic noise. The results indicate the great potential of spintronic stochastic neurons for neuromorphic computing. It is worth noting that the stochastic output of the device is under an ultralow bias ($V = 1$ mV), revealing the ultralow energy performance of the devices.

Furthermore, Eq. (5) indicates that stochastic behaviors of the device are tunable if we can modulate the energy barrier. According to Eq. (1), the perpendicular magnetic anisotropy energy can be controlled by voltage bias via the well-known voltage-controlled magnetic anisotropy (VCMA) effect along with the high-energy efficiency. By adding the linear dependence of the first-order anisotropy energy on voltage, Eq. (1) can be rewritten as

$$E_b = \left(K_{\text{eff}} - H_{\text{in}}M_s + \frac{M_s^2}{4K_{\text{eff}}}H_{\text{in}}^2\right)V^*,$$
$$K_{\text{eff}} = K_{0,\text{eff}} - \frac{\xi}{t_d}\frac{V}{t}, \quad (6)$$

where H_{in} is the in-plane exchange bias from the in-plane reference layer, t is the thickness of the free layer, t_d is the thickness of the MgO barrier, and ξ is the linear VCMA magnetoelectric coefficient. According to Eq. (6), we can tune the energy barrier of the device by the voltage bias to control the stochastic behavior. If there is no in-plane exchange bias or the nonlinear term is negligible when $M_s^2H_{\text{in}}^2/4K_{\text{eff}} \ll K_{\text{eff}}$, the energy barrier can be linearly tuned by the voltage bias. Thus, we study the stochastic behaviors of the device under different voltage biases. In Fig. 4(a), the voltage bias is swept from −200

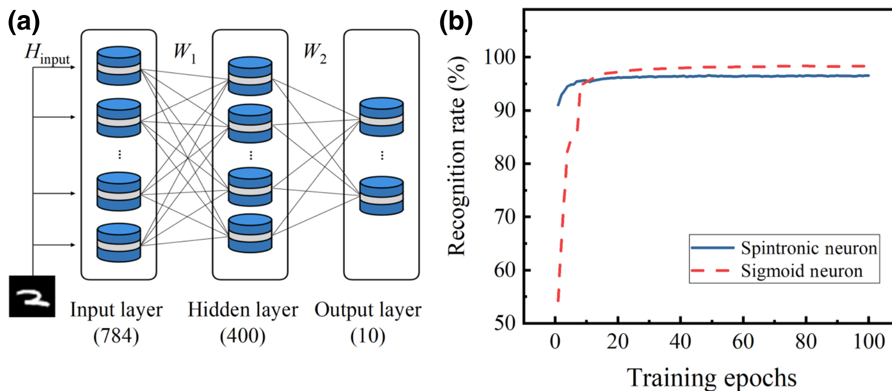

FIG. 3. (a) The structure of the ANN we construct to recognize the handwriting digits in the MNIST database (b) The transition in the recognition rate obtained from the neural network constructed with the MTJ (blue line) compared to a conventional neural network (red dash line).

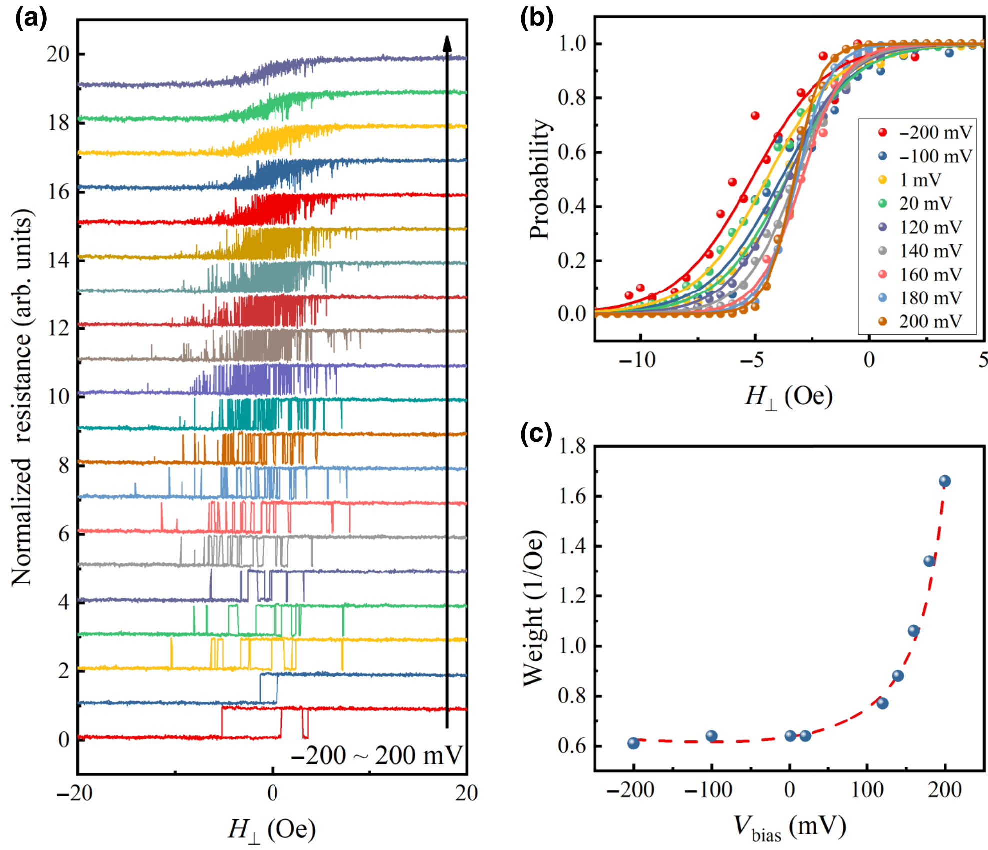

FIG. 4. Minor magnetoresistance curve of the MTJ with step $\Delta H = 1$ Oe at different voltage biases V_{bias} from −200 to 200 mV under the perpendicular magnetic field ($H_\perp$). (b) Switching probability under various external perpendicular magnetic fields ($H_\perp$) at different voltage biases (dot) fitted by the sigmoid probability density function (line). (c) The weight obtained from the fitting of the sigmoid probability density function (blue dot). The data can be fitted by the relation between the energy barrier and the voltage bias (red line).

to 200 mV under the perpendicular magnetic field. The stochastic telegraphic switching regime vs magnetic field is suppressed while increasing the voltage bias, and the switching rate of the device is increased along with the voltage bias. Similar to Fig. 2(d), we can obtain the weight at different voltage biases by fitting the data from Fig. 4(a) of Eq. (5), as shown in Fig. 4(b). The weight changing vs the voltage bias is shown in Fig. 4(c), which can be well fitted by Eq. (6), and shows a nonlinear relation between the energy barrier and the voltage bias. Since the sigmoid output characteristic of the device depends on the weight, we can tune the relation between input and output signals by the voltage bias. This performance indicates that we can use the voltage-controlled devices to perform as the adaptive neuron, which can further improve the neural network. With the voltage-controlled spintronic neurons, we can build up a spintronic hardware neural network, which can perform neuromorphic computing with higher energy efficiency and lower area consumption.

Finally, in order to verify the voltage-modulation stochastic behavior result from the VCMA effect, we perform the spin-torque ferromagnetic resonance (STFMR) measurement [42]. Figure 5(a) shows the STFMR spectra under different voltage biases. The resonant conditions can be expressed as follows [42]:

$$f_r = \frac{\mu_0 \gamma}{2\pi}\sqrt{f_1 f_2},$$
$$f_1 = H_{\text{ext}}\cos(\theta_H - \theta_M) + H_{K1}\cos^2(\theta_M) - H_{K2}\cos^4(\theta_M),$$
$$f_2 = H_{\text{ext}}\cos(\theta_H - \theta_M) + H_{K1}\cos(2\theta_M) - \frac{1}{2}H_{K2}[\cos(2\theta_M) + \cos(4\theta_M)], \quad (7)$$

where f_r is the resonant frequency, H_{K1} and H_{K2} are the first and second orders of the effective magnetic anisotropy field in the free layer, θ_M is the angle between the magnetization in the free layer and the normal to the film plane, and θ_H is the angle between the external magnetic field H_{ext} and the normal to the film plane. The first- and second-order anisotropy fields H_{K1} and H_{K2} as a function of V_{bias} can be obtained from the fitting of Eq. (7), shown in Fig. 5(b). The zero-voltage perpendicular anisotropy field is $\mu_0 H_{K1} = 56$ mT, consistent with previous reports [35].

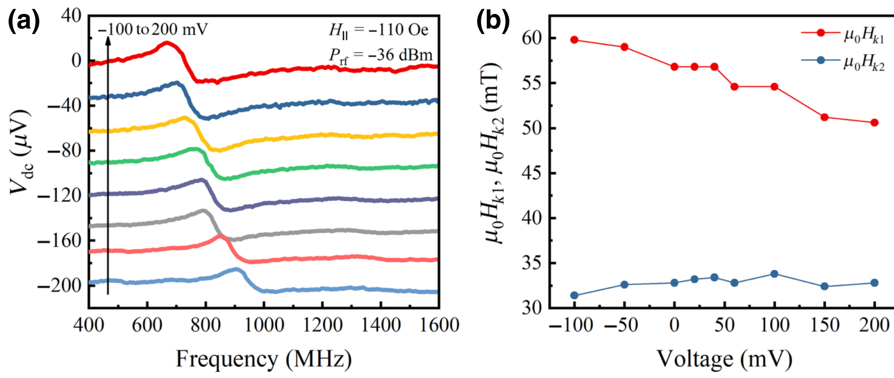

FIG. 5. (a) STFMR spectra under in-plane magnetic field $H_{||} = -110$ Oe at bias voltages V_{bias} from -100 to 200 mV. Each data set is offset by 35 μV. (b) First- and second-order anisotropy fields H_{K1} and H_{K2} as a function of V_{bias} obtained from the fitting of Eq. (7).

When we sweep the voltage bias from -100 to $+200$ mV, the first-order anisotropy field H_{K1} decreases while the second-order anisotropy field H_{K2} remains almost constant with the applied voltage bias. The modulation ratio of $\mu_0 H_{K1}$ is 34 mT/V, corresponding to a change in the magnetic anisotropy energy per unit area per applied electric field of $\xi = 26\,\text{fJ}\,\text{V}^{-1}\,\text{m}^{-1}$, in good agreement with previous reports for similar material structures [35,42,43]. According to Fig. 5(b), the is H_{K1} very small, which confirms the low barrier obtained by compensating for the out-of-plane demagnetization field and the existence of the VCMA effect is also proved.

IV. CONCLUSIONS

In this paper, we report a voltage-controlled spintronic stochastic neuron based on a MJT. By introducing the perpendicular anisotropy into the free layer of an in-plane MJT, the energy barrier is reduced and low-energy-cost stochastic behavior is obtained. The device shows the ability to mimic artificial neurons of an ANN to realize the recognition of the handwritten digits in the MNIST database, and the spintronic neuron can work as well as the artificial sigmoid neuron at a hardware level with ultralow energy consumption. Furthermore, the voltage-controlled stochastic behavior is proved theoretically and experimentally, which shows an adaptive neuron characteristic and finally confirms the result from the VCMA effect. Our results give a highly energy-efficient way to build up all spintronic neuromorphic computing systems.

ACKNOWLEDGMENTS

This work was supported by the Executive Programme of Scientific and Technological Cooperation between Italy and China for the years 2016–2018 (code CN16GR09, Grant No. 2016YFE0104100) titled “Nanoscale broadband spintransfer-torque microwave detector” funded by the Ministero degli Affari Esteri e della Cooperazione Internazionale. This work was supported in part by the National Science Foundation of China (Grants No. 51761145025, No. 11474311, and No. 11804370) and the National Postdoctoral Program for Innovative Talents (Grant No. BX201700275) and China Postdoctoral Science Foundation Funded Project (Grant No. 2017M621858).

[1] D. Hassabis, D. Kumaran, C. Summerfield, and M. Botvinick, Neuroscience-inspired artificial intelligence, Neuron **95**, 245 (2017).

[2] Y. LeCun, Y. Bengio, and G. Hinton, Deep learning, Nature **521**, 436 (2015).

[3] P. U. Diehl and M. Cook, Unsupervised learning of digit recognition using spike-timing-dependent plasticity, Front. Comput. Neurosci. **9**, 99 (2015).

[4] D. Kuzum, S. M. Yu, and H. S. P. Wong, Synaptic electronics: Materials, devices and applications, Nanotechnology **24**, 382001 (2013).

[5] P. A. Merolla, et al., A million spiking-neuron integrated circuit with a scalable communication network and interface, Science **345**, 668 (2014).

[6] B. Rajendran and F. Alibart, Neuromorphic computing based on emerging memory technologies, IEEE J. Emerg. Sel. Top. Circuits Syst. **6**, 198 (2016).

[7] A. Thomas, S. Niehorster, S. Fabretti, N. Shepheard, O. Kuschel, K. Kupper, J. Wollschlager, P. Kzysteczko, and E. Chicca, Tunnel junction based memristors as artificial synapses, Front. Neurosci. **9**, 241 (2015).

[8] D. Ielmini and H. S. P. Wong, In-memory computing with resistive switching devices, Nat. Electron. **1**, 333 (2018).

[9] D. Kuzum, R. G. D. Jeyasingh, B. Lee, and H. S. P. Wong, Nanoelectronic programmable synapses based on phase change materials for brain-inspired computing, Nano Lett. **12**, 2179 (2012).

[10] K. Y. Camsari, R. Faria, B. M. Sutton, and S. Datta, Stochastic p-bits for invertible logic, Phys. Rev. X **7**, 031014 (2017).

[11] P. Krzysteczko, J. Munchenberger, M. Schafers, G. Reiss, and A. Thomas, The memristive magnetic tunnel junction as a nanoscopic synapse-neuron system, Adv. Mater. **24**, 762 (2012).

[12] Z. Zeng, G. Finocchio, B. Zhang, P. K. Amiri, J. A. Katine, I. N. Krivorotov, Y. Huai, J. Langer, B. Azzerboni, K. L. Wang, and H. Jiang, Ultralow-current-density and bias-field-free spin-transfer nano-oscillator, Sci. Rep. **3**, 1426 (2013).

[13] Z. Zeng, G. Finocchio, and H. Jiang, Spin transfer nano-oscillators, Nanoscale **5**, 2219 (2013).

[14] L. Zhang, B. Fang, J. Cai, M. Carpentieri, V. Puliafito, F. Garescì, P. K. Amiri, G. Finocchio, and Z. Zeng, Ultrahigh detection sensitivity exceeding 105 v/w in spin-torque diode, Appl. Phys. Lett. **113**, 102401 (2018).

[15] A. Mizrahi, T. Hirtzlin, A. Fukushima, H. Kubota, S. Yuasa, J. Grollier, and D. Querlioz, Neural-like computing with populations of superparamagnetic basis functions, Nat. Commun. **9**, 1533 (2018).

[16] J. Grollier, D. Querlioz, and M. D. Stiles, Spintronic nanodevices for bioinspired computing, Proc. IEEE **104**, 2024 (2016).

[17] N. Locatelli, V. Cros, and J. Grollier, Spin-torque building blocks, Nat. Mater. **13**, 11 (2014).

[18] J. Torrejon, M. Riou, F. A. Araujo, S. Tsunegi, G. Khalsa, D. Querlioz, P. Bortolotti, V. Cros, K. Yakushiji, A. Fukushima, H. Kubota, S. Yuasa, M. D. Stiles, and J. Grollier, Neuromorphic computing with nanoscale spintronic oscillators, Nature **547**, 428 (2017).

[19] J. L. Cai, B. Fang, C. Wang, and Z. M. Zeng, Multilevel storage device based on domain-wall motion in a magnetic tunnel junction, Appl. Phys. Lett. **111**, 182410 (2017).

[20] W. A. Borders, H. Akima, S. Fukami, S. Moriya, S. Kurihara, Y. Horio, S. Sato, and H. Ohno, Analogue spin-orbit torque device for artificial-neural-network-based associative memory operation, Appl. Phys. Express **10**, 013007 (2017).

[21] H. Arai and H. Imamura, Neural-Network Computation Using Spin-Wave-Coupled Spin-Torque Oscillators, Phys. Rev. Appl. **10**, 024040 (2018).

[22] D. I. Suh, G. Y. Bae, H. S. Oh, and W. Park, Neural coding using telegraphic switching of magnetic tunnel junction, J. Appl. Phys. **117**, 17D714 (2015).

[23] A. Sengupta, M. Parsa, B. Han, and K. Roy, Probabilistic deep spiking neural systems enabled by magnetic tunnel junction, IEEE Trans. Electron Devices **63**, 2963 (2016).

[24] A. Sengupta and K. Roy, Short-Term Plasticity and Long-Term Potentiation in Magnetic Tunnel Junctions: Towards Volatile Synapses, Phys. Rev. Appl. **5**, 024012 (2016).

[25] G. Srinivasan, A. Sengupta, and K. Roy, Magnetic tunnel junction based long-term short-term stochastic synapse for a spiking neural network with on-chip STDP learning, Sci. Rep. **6**, 29545 (2016).

[26] N. V. Agudov, A. A. Dubkov, and B. Spagnolo, Escape from a metastable state with fluctuating barrier, Physica A **325**, 144 (2003).

[27] R. N. Mantegna, B. Spagnolo, L. Testa, and M. Trapanese, Stochastic resonance in magnetic systems described by Preisach hysteresis model, J. Appl. Phys. **97**, 1 (2005).

[28] N. V. Agudov, A. V. Krichigin, D. Valenti, and B. Spagnolo, Stochastic resonance in a trapping overdamped monostable system, Phys. Rev. E **81**, 051123 (2010).

[29] C. M. Liyanagedera, A. Sengupta, A. Jaiswal, and K. Roy, Stochastic Spiking Neural Networks Enabled by Magnetic Tunnel Junctions: From Nontelegraphic to Telegraphic Switching Regimes, Phys. Rev. Appl. **8**, 064017 (2017).

[30] D. Vodenicarevic, N. Locatelli, A. Mizrahi, J. S. Friedman, A. F. Vincent, M. Romera, A. Fukushima, K. Yakushiji, H. Kubota, S. Yuasa, S. Tiwari, J. Grollier, and D. Querlioz, Low-Energy Truly Random Number Generation with Superparamagnetic Tunnel Junctions for Unconventional Computing, Phys. Rev. Appl. **8**, 054045 (2017).

[31] C. C. Y. Lecun and C. J. C. Burges, The MNIST database of handwritten digits, http://yann.lecun.com/exdb/mnist/.

[32] S. Ikeda, K. Miura, H. Yamamoto, K. Mizunuma, H. D. Gan, M. Endo, S. Kanai, J. Hayakawa, F. Matsukura, and H. Ohno, A perpendicular-anisotropy CoFeB-MgO magnetic tunnel junction, Nat. Mater. **9**, 721 (2010).

[33] P. K. Amiri, Z. M. Zeng, J. Langer, H. Zhao, G. Rowlands, Y. J. Chen, I. N. Krivorotov, J. P. Wang, H. W. Jiang, J. A. Katine, Y. Huai, K. Galatsis, and K. L. Wang, Switching current reduction using perpendicular anisotropy in CoFeB-MgO magnetic tunnel junctions, Appl. Phys. Lett. **98**, 112507 (2011).

[34] Z. Zeng, P. K. Amiri, I. N. Krivorotov, H. Zhao, G. Finocchio, J.-P. Wang, J. A. Katine, Y. Huai, J. Langer, K. Galatsis, K. L. Wang, and H. Jiang, High-power coherent microwave emission from magnetic tunnel junction nano-oscillators with perpendicular anisotropy, ACS Nano **6**, 6115 (2012).

[35] B. Fang, M. Carpentieri, X. Hao, H. Jiang, J. A. Katine, I. N. Krivorotov, B. Ocker, J. Langer, K. L. Wang, B. Zhang, B. Azzerboni, P. K. Amiri, G. Finocchio, and Z. Zeng, Giant spin-torque diode sensitivity in the absence of bias magnetic field, Nat. Commun. **7**, 11259 (2016).

[36] C. Grezes, A. Rojas Rozas, F. Ebrahimi, J. G. Alzate, X. Cai, J. A. Katine, J. Langer, B. Ocker, P. Khalili Amiri, and K. L. Wang, In-plane magnetic field effect on switching voltage and thermal stability in electric-field-controlled perpendicular magnetic tunnel junctions, AIP Adv. **6**, 075014 (2016).

[37] R. Matsumoto, T. Nozaki, S. Yuasa, and H. Imamura, Voltage-Induced Precessional Switching at Zero-Bias Magnetic Field in a Conically Magnetized Free Layer, Phys. Rev. Appl. **9**, 014026 (2018).

[38] M. Bapna and S. A. Majetich, Current control of time-averaged magnetization in superparamagnetic tunnel junctions, Appl. Phys. Lett. **111**, 243107 (2017).

[39] W. Rippard, R. Heindl, M. Pufall, S. Russek, and A. Kos, Thermal relaxation rates of magnetic nanoparticles in the presence of magnetic fields and spin-transfer effects, Phys. Rev. B **84**, 064439 (2011).

[40] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, Gradient-based learning applied to document recognition, Proc. IEEE **86**, 2278 (1998).

[41] X. Glorot and Y. Bengio, Understanding the difficulty of training deep feedforward neural networks, in Proceedings of the thirteenth international conference on artificial intelligence and statistics (2010), p. 249.

[42] S. Kanai, M. Gajek, D. C. Worledge, F. Matsukura, and H. Ohno, Electric field-induced ferromagnetic resonance in a CoFeB/MgO magnetic tunnel junction under dc bias voltages, Appl. Phys. Lett. **105**, 242409 (2014).

[43] J. Zhu, J. A. Katine, G. E. Rowlands, Y. J. Chen, Z. Duan, J. G. Alzate, P. Upadhyaya, J. Langer, P. K. Amiri, K. L. Wang, and I. N. Krivorotov, Voltage-Induced Ferromagnetic Resonance in Magnetic Tunnel Junctions, Phys. Rev. Lett. **108**, 197203 (2012).