

An artificial synapse by superlattice-like phase-change material for low-power brain-inspired computing

Qing Hu(胡庆), Boyi Dong(董博义), Lun Wang(王伦), Enming Huang(黄恩铭), Hao Tong(童浩), Yuhui He(何毓辉), Ming Xu(徐明), Xiangshui Miao(缪向水)

Citation: Chin. Phys. B . 2020, 29(7): 070701 . **doi:** 10.1088/1674-1056/ab892a
Journal homepage: http://cpb.iphy.ac.cn; http://iopscience.iop.org/cpb

What follows is a list of articles you may be interested in

High-performance synaptic transistors for neuromorphic computing

Hai Zhong(钟海), Qin-Chao Sun(孙勤超), Guo Li(李果), Jian-Yu Du(杜剑宇), He-Yi Huang(黄河意), Er-Jia Guo(郭尔佳), Meng He(何萌), Can Wang(王灿), Guo-Zhen Yang(杨国桢), Chen Ge(葛琛), Kui-Juan Jin(金奎娟)

Chin. Phys. B . 2020, 29(4): 040703 . **doi:** 10.1088/1674-1056/ab7806

Preliminary abnormal electrocardiogram segment screening method for Holter data based on long short-term memory networks

Siying Chen(陈偲颖), Hongxing Liu(刘红星)

Chin. Phys. B . 2020, 29(4): 040701 . **doi:** 10.1088/1674-1056/ab6d52

Multiple Lagrange stability and Lyapunov asymptotical stability of delayed fractional-order Cohen-Grossberg neural networks

Yu-Jiao Huang(黄玉娇), Xiao-Yan Yuan(袁孝焰), Xu-Hua Yang(杨旭华), Hai-Xia Long(龙海霞), Jie Xiao(肖杰)

Chin. Phys. B . 2020, 29(2): 020703 . **doi:** 10.1088/1674-1056/ab6716

Coexistence and local Mittag-Leffler stability of fractional-order recurrent neural networks with discontinuous activation functions

Yu-Jiao Huang(黄玉娇), Shi-Jun Chen(陈时俊), Xu-Hua Yang(杨旭华), Jie Xiao(肖杰)

Chin. Phys. B . 2019, 28(4): 040701 . **doi:** 10.1088/1674-1056/28/4/040701

Synthesization of high-capacity auto-associative memories using complex-valued neural networks

Yu-Jiao Huang(黄玉娇), Xiao-Yan Wang(汪晓妍), Hai-Xia Long(龙海霞), Xu-Hua Yang(杨旭华)

Chin. Phys. B . 2016, 25(12): 120701 . **doi:** 10.1088/1674-1056/25/12/120701

SPECIAL TOPIC — Physics in neuromorphic devices

An artificial synapse by superlattice-like phase-change material for low-power brain-inspired computing*

Qing Hu(胡庆), Boyi Dong(董博义), Lun Wang(王伦), Enming Huang(黄恩铭), Hao Tong(童浩)†, Yuhui He(何毓辉), Ming Xu(徐明)‡, and Xiangshui Miao(缪向水)

Wuhan National Laboratory for Optoelectronics, School of Optical and Electronic Information, Huazhong University of Science and Technology, Wuhan 430074, *China*

(Received 27 February 2020; revised manuscript received 11 April 2020; accepted manuscript online 13 April 2020)

Phase-change material (PCM) is generating widespread interest as a new candidate for artificial synapses in bio-inspired computer systems. However, the amorphization process of PCM devices tends to be abrupt, unlike continuous synaptic depression. The relatively large power consumption and poor analog behavior of PCM devices greatly limit their applications. Here, we fabricate a GeTe/Sb_2Te_3 superlattice-like PCM device which allows a progressive RESET process. Our devices feature low-power consumption operation and potential high-density integration, which can effectively simulate biological synaptic characteristics. The programming energy can be further reduced by properly selecting the resistance range and operating method. The fabricated devices are implemented in both artificial neural networks (ANN) and convolutional neural network (CNN) simulations, demonstrating high accuracy in brain-like pattern recognition.

Keywords: superlattice-like, phase-change material, artificial synapse, low-power consumption

PACS: 07.05.Mh, 47.20.Hw, 83.10.Tv **DOI:** 10.1088/1674-1056/ab892a

1. Introduction

Artificial intelligence (AI) that allows machines to think and act like human brains has been the main focus of computer science in this century, and the application of AI has rapidly extended to various fields in the past few years, such as big data mining, large-scale visual/auditory recognition and classification, driverless cars, complex strategic games, etc.[1–4] Currently, these tasks and applications are based on computers with von Neumann architecture, using conventional central processing units and graphics processing units with off-chip memories, to implement large-scale neural network training which requires several kilowatts of power.[5] Although custom-designed neuromorphic hardware with complementary metal oxide semiconductor (CMOS) technologies could greatly reduce the energy, the power consumption is still a serious issue with the fast expansion of the network scale[6] due to high-frequency data exchange and transmission. It is found that the dynamic random off-chip weight storage access memory will consume 100 times more power than the on-chip memory.[7]

In the biological brain, memory and processing are highly combined to provide an efficient and low-power way of computation.[8] Hence, we can use new memory devices to mimic the way how our brain works. Recently, novel non-volatile devices which can store information in different resistance states and exhibit conductivity modulation based on the programming history, are promising for achieving synaptic dynamics in a compact and power-efficient manner.[9–17] Among these devices, the phase-change materials (PCMs) provide a simple way to be integrated on a large scale, showing a good application prospect. It has been demonstrated by recent impressive achievements using large numbers of synapses.[5,18–28] However, the traditional mushroom-shape PCM devices present relatively high-power consumption under the RESET operation[29] due to the melt-quench amorphization process.[30] During that process, the conductance of PCM devices changes abruptly which is undesirable for mimicking the synaptic events. The mainstream solution is to use two PCM devices in the form of differential pairs to achieve synaptic functionality (one device is used for potentiation and the other for depression).[5,31–33] But it requires long refresh operations, where all PCM devices are being reprogrammed. This not only increases the energy consumption of the system but also increases the complex of processing, thus limiting the largescale integration for phase change synapses. In recent years, superlattice-like (SLL) PCM devices have drawn many attentions in this field because they reduce the energy consumption and meanwhile still maintain good performance.[34] Although the working mechanism is still under intensive debates,[35–38] the performance of low-

*Project supported by the National Science and Technology Major Project of China (Grant No. 2017ZX02301007-002), the National Key R&D Plan of China (Grant No. 2017YFB0701701), and the National Natural Science Foundation of China (Grant Nos. 61774068 and 51772113). The authors acknowledge the support from Hubei Key Laboratory of Advanced Memories & Hubei Engineering Research Center on Microelectronics.

†Corresponding author. E-mail: tonghao@hust.edu.cn

‡Corresponding author. E-mail: mxu@hust.edu.cn

© 2020 Chinese Physical Society and IOP Publishing Ltd

power consumption and the outstanding phase transition behavior in SLL-PCM provides new opportunities to improve the performance of PCM devices for memory and computing applications.[39] However, there is no systematic study on how to use and operate SLL-PCM in neural networks or the accuracy of this type of device.

In this article, we fabricated a GeTe/Sb_2Te_3 SSL device to mimic the synaptic events. Through interface control, the analog behavior of this artificial synapse has been optimized. In the specific operation process, we have adopted two different schemes and improved them. The benchmark tasks for handwritten digits recognition are also simulated and verified with artificial neural networks (ANN) and convolutional neural networks (CNN), respectively.

2. Experiments

We fabricated the SSL GeTe/Sb_2Te_3 PCM cell with 250 nm diameter via-hole structure. First, 10 nm Ti, 100 nm Pt, and 100 nm SiO_2 isolation layer were deposited in sequence. Next, electron beam lithography (EBL) and inductive coupled plasma etching (ICPE) were performed to create a 250 nm hole and expose the bottom electrode (BE) contact pad. Photolithography was then employed to pattern the functional layer. After that, we deposited the functional layers through magnetron sputtering at room temperature. 2 nm Sb_2Te_3 was firstly deposited on the BE, and then 4 nm GeTe was deposited on Sb_2Te_3. After repeating this cycle 12 times, 2 nm Sb_2Te_3 was lastly deposited on the top. Using this method, we obtained GeTe/Sb_2Te_3 SSL functional layers with good heterogeneous structure. Finally, 100 nm Pt inert top electrode (TE) was deposited. The electrical properties measurements were carried out by Agilent B1500 A semiconductor analyzer. The top Pt electrode was biased and the bottom Pt electrode was grounded during the measurements.

3. Results and discussion

3.1. SLL PCM characteristics

The two phases have remarkable resistance contrast, as displayed by the *I*–*V* characteristics in Fig. 1(b): while the crystalline state shows a relatively low resistance, the amorphous state shows high resistance and a typical threshold switching behavior at the characteristic threshold voltage V_{th}. By controlling the applied pulse, we can achieve continuous conductance state in-between these two states, which is a key property for SLL PCM to emulate the synapse.[31] Another key property that enables brain-inspired computing is the accumulative behavior arising from the crystallization dynamics.[23] As shown in Fig. 2, one can induce progressive reduction in the size of the amorphous region, corresponding to the increasing conductance, by applying successive SET pulses with the gradually changing or unchanged amplitude. By applying SET voltage pulses of 3.2 V and the pulse width of 1 μs, the conductance of the device can continuously increase from 0.54 μS to 26.98 μS, with about 13 distinctive states (Fig. 2(a)). On the other hand, with the RESET voltage pulses of 4 V and the pulse width of 100 ns, the conductance of the device can be continuously reduced from 26.98 μS to 1.97 μS, with about 22 states (Fig. 2(b)). Because the interface significantly affects the thermal transport, the thermal conductivity of the SLL structure decreases sharply after the interface is introduced. The change in thermal conductivity can improve the heating efficiency of the SLL PCM device and improve the energy control ability. This allows the SLL device to have better analog characteristics and does not require additional design to compensate.

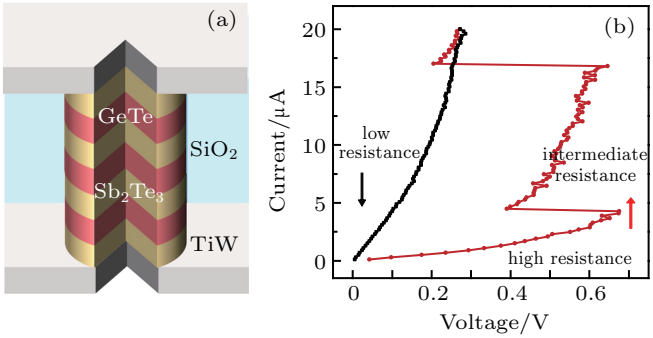


Fig. 1. Device fabrication and electrical characteristics. (a) Schematic of the via-hole GeTe/Sb_2Te_3 SLL PCM device. The total thickness of the intermediate functional layers is 74 nm. (b) A typical current–voltage curve of this formed PCM device, showing multi-level resistance states under certain operations.

By continuously adjusting the amplitude of the voltage pulses, more intermediate states can be obtained (Fig. 2(c)). Although applying the voltage pulse with the same amplitude will be easier and less expensive to implement in actual circuits, it may lead to less intermediate states, which directly affect the device performance as a synapse. From Fig. 2(c), we can see that although there are many conductance states near low conductance values, the window of difference between them is narrowed compared to high conductance values, which can be regarded as the same conductance. We have tried to increase the ratio of the high and low conductance of the device as much as possible, but the varying curve of the device conductance is not linear, and it is more concentrated near the low conductance. At the same time, operating a device near a high conductance will result in high-power consumption. Therefore, just by adjusting the amplitude of the pulse, the synaptic analog behavior becomes worse. But our further study found that the conductance change at low conductance shows better synaptic characteristics. Using the same amplitude variation operation scheme, while limiting the device conductance operating range, we can obtain the synaptic analog behavior of Fig. 2(d).

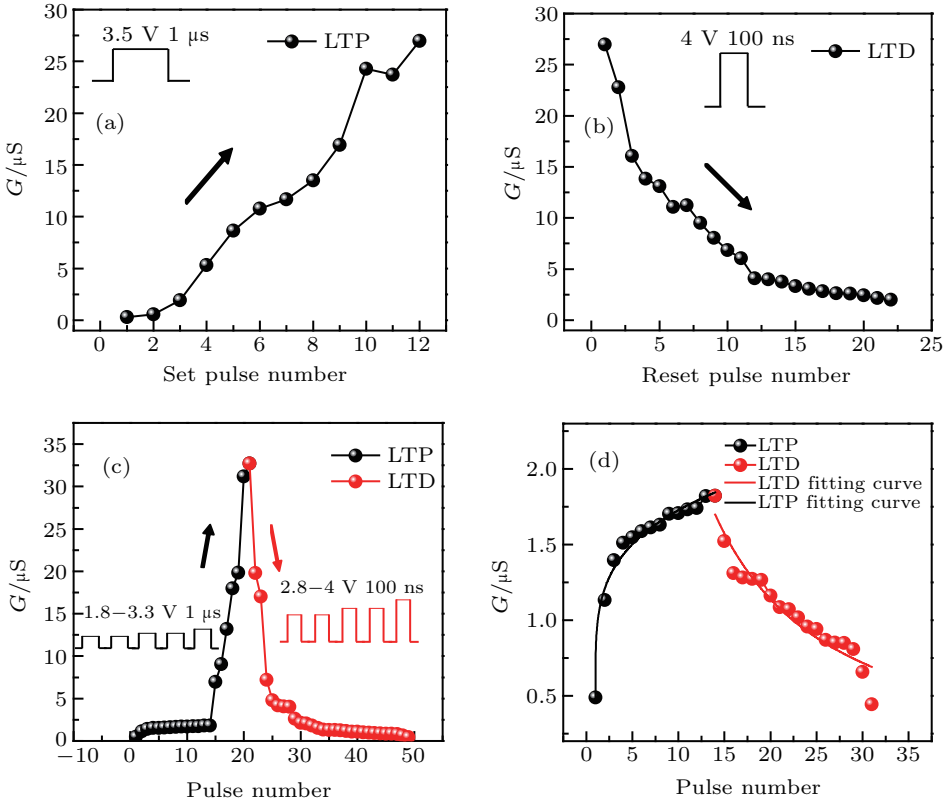


Fig. 2. Superlattice-like PCM synaptic characteristics. (a) Long term potentiation (LTP) properties of superlattice-like PCM synaptic with fixed SET pulse scheme. (b) Long term depression (LTD) properties of superlattice-like PCM synaptic with fixed RESET pulse scheme. (c) Gradual SET of the device is implemented by using pulses with increasing amplitude of 1.8–3.3 V range with 150 mV voltage steps. Gradual reset of the cell resistance is achieved by using stair-case pulses from 2.8–4 V range with 80 mV voltage steps. There are 2 pulses for each voltage step during the SET and RESET operations. (d) An excessively high maximum conductance state may result in a decrease in the effective number of intermediate states. We limit the working range of the conductance to maximize the effective number of intermediate states.

Table 1. Synaptic characteristics under different operations.

Synaptic properties	Figs. 2(a)–2(b)	Fig. 2(c)	Fig. 2(d)
LTP nonlinearity	1.718	4.959	3.755
LTD nonlinearity	4.328	7.081	1.346
Asymmetry	6.046	12.04	5.101
ON/OFF ratio	13.69	67.03	4.13
No. of conductance states	22	29	15

3.2. ANN MNIST task

Figure 3(a) shows the topological structure of the simulated three-layer feedforward ANN, where the classification of the modified National Institute of Standard and Technology (MNIST) handwritten digits is taken as an example. This is a fully connected network, with 784 input neurons (i), 250 hidden neurons (h), and 10 output neurons (o).

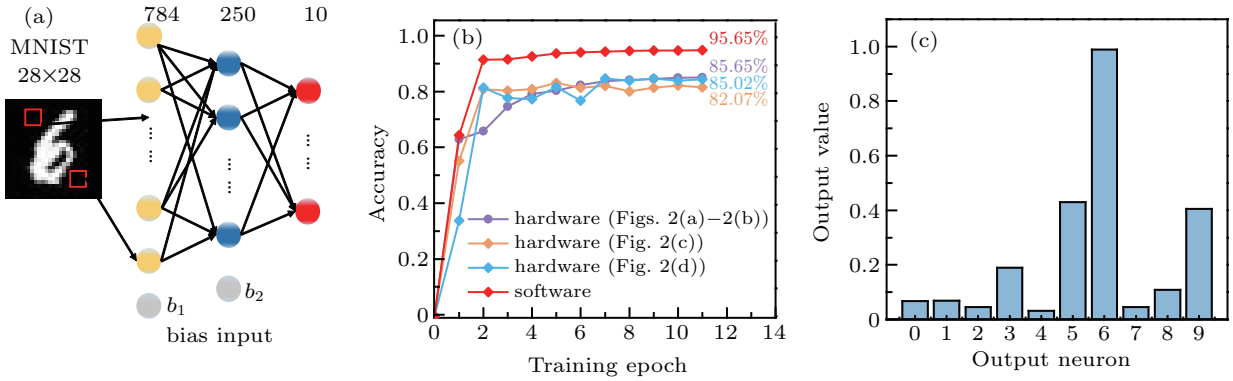


Fig. 3. Applications of superlattice-like PCM synapses in ANN. (a) Schematic illustration of the input digits and the three-layer artificial neuron network with bias input. (b) The final network recognition rate of the software synapse and synapse data in Fig. 2 using two different update methods. (c) The output results of 10 output neurons under digital picture 6 inputs. The second output neuron has the largest output value, so its corresponding number '6' is the recognition result of the network.

The analog weight update is used in the simulation based on superlattice-like PCM conductance modulation properties, such as the number of conductance states, G_{max}/G_{min}. Our network uses supervised learning method with traditional backpropagation algorithms.[40]

The programming process follows either one of the two learning rules below: an update scheme using unfixed voltage pulses and an update scheme with fixed voltage pulses, where η is the learning rate of the network,

$$\Delta w_{\mathrm{ih}} = -\eta \cdot \frac{\partial E}{\partial w_{\mathrm{ih}}}, \quad \Delta w_{\mathrm{ho}} = -\eta \cdot \frac{\partial E}{\partial w_{\mathrm{ho}}},$$
$$\Delta w_{\mathrm{ih}} = -\eta \cdot \mathrm{sgn}\frac{\partial E}{\partial w_{\mathrm{ih}}}, \quad \Delta w_{\mathrm{ho}} = -\eta \cdot \mathrm{sgn}\frac{\partial E}{\partial w_{\mathrm{ho}}}.$$

During each iteration, 5×10^4 pictures are used as the training process in random order, and another 1×10^4 pictures are used as the input to the test process. We use the batch learning method, updating the weights after each input of 100 pictures (Fig. 4(a)). The hardware implementation scheme of the two update methods is given by Figs. 4(b) and 4(c), respectively. Update method 1 does not focus on the size of the update, but only on the direction of adjustment. We can use the data in Figs. 2(a) and 2(b) to achieve this operation method, and apply the corresponding fixed amplitude pulses according to the sign of the weight change. On the other hand, to adjust the weight data from Figs. 2(c) or 2(d), we have to read the device state firstly and calculate the required pulses.

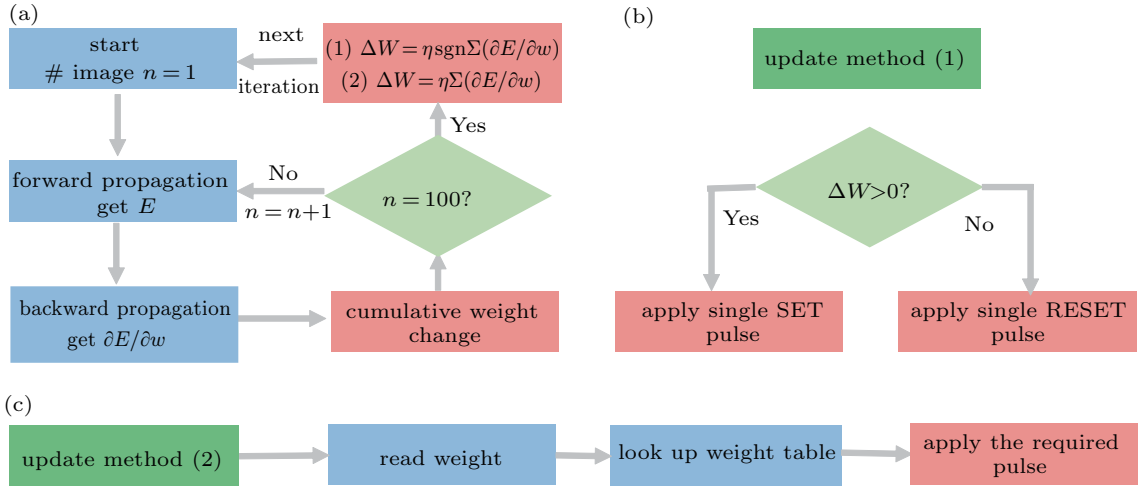


Fig. 4. Part of the training process flow chart and schematic diagram of two types of weight update methods. (a) The steps in the blue box mainly use the resistance characteristics of the device to reflect synaptic efficacy. Weight updates in the red box take advantage of the device's resistive state transition to reflect synaptic plasticity, where E is the loss function in the back-propagation algorithm. (b) The update method 1 corresponds to the weight data of Figs. 2(a) and 2(b). It does not require additional calculation of the pulse size to be applied, only the category. (c) The update method 2 corresponds to the weight data of Figs. 2(c) and 2(d). This method needs to obtain the current position of the weight and then calculate the pulse to be applied.

Using the three groups of synaptic weight data from Fig. 2, the final network recognition accuracy is shown in Fig. 3(b). When weight update method 1 is used, the network convergence speed is slightly slow (purple line). The final recognition rate of the three sets of data is slightly different, while there is a certain gap with the ideal recognition rate. Next, we explore the differences between them through a more complex network.

3.3. CNN MNIST task

Due to its parameter sharing mechanism and sparseness of connections, CNN has better recognition results than ordinary ANN.[41–44] The schematic diagram of our CNN network structure is shown in Fig. 5. We still use supervised learning, and the input and output encoding methods and network training methods are similar to the previous ANN.

Figure 5(b) reveals the results of preprocessing of the convolutional layer after training is completed. After two convolution pooling, the input is reduced in dimensions, and each small graph contains part of the information of the original graph, which is then used as the input of the fully connected layer. As the complexity of the network and the parameters increase, the network recognition rate in the ideal situation is greatly improved. Based on the highly non-linear synapse data in Fig. 2(c), the final network recognition rate drops sharply. Or from the effective conductance analysis, the actual adjustable weight value is too small, resulting in a reduction in classification accuracy. Although the on-off ratios of the devices in Figs. 2(a), 2(b), and 2(d) are different, the nonlinearity and the number of states of the devices are similar. We can see that the two purple and orange lines in Fig. 5(c), the two types of synapse data and operation methods, have similar final network recognition accuracy. As long as there is good discrimination between the conductance values and the number of conductance states is similar, $G_{\mathrm{max}}/G_{\mathrm{min}}$ has little effect on the results of the network. Therefore, in terms of the impact on the network results, it is a feasible solution to change the amplitude of the applied pulse and then limit the working range.

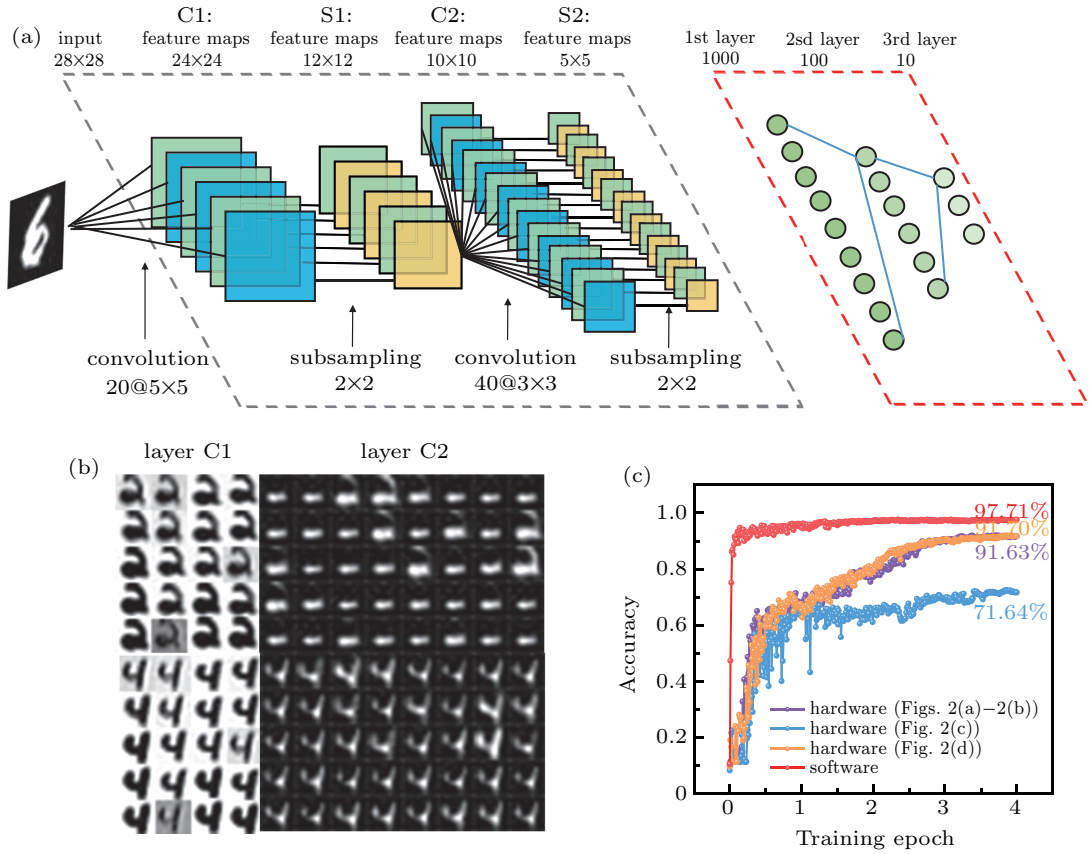


Fig. 5. Applications of superlattice-like PCM synapses in CNN. (a) Schematic diagram of a CNN comprising feature extraction and classification for a handwritten digit recognition task. Both the convolutional kernels in the feature extraction unit and the connections in the classification unit are simulated by superlattice-like PCM synapses. (b) Output results of the convolutional layer that performs feature extraction on the input picture. (c) The final network recognition rate of the software synapse and synapse data in Fig. 2 using two different update methods.

3.4. Energy efficiency

Low-power consumption of our SLL device is mainly due to interface scattering and phonon microstrip transport, which results in extremely low thermal conductivity, thus significantly improving heating efficiency.[45] We roughly estimate the power consumption of the device under pulse operation through equations below:

$$E_{\text{set}} \approx V_{\text{set}} \cdot I_{\text{set}} \cdot t_{\text{set}},$$
$$E_{\text{reset}} \approx V_{\text{reset}} \cdot I_{\text{reset}} \cdot t_{\text{reset}}.$$

Table 2. Energy statistics on average operating power consumption.

Data	E_{set}/pJ	E_{reset}/pJ
Figs. 2(a)–2(b)	104.42	7.97
Fig. 2(c)	59.67	2.90
Fig. 2(d)	8.24	1.39

The results are shown in Table 1. Although this calculated power consumption may be slightly different from the actual operating energy of the experimental devices, it can still reflect the effectiveness of the operating strategy. By limiting the operating range of the conductance, the power consumption of the device can be significantly reduced, which indicates that we can further reduce the power consumption by increasing the complexity of operations without affecting the accuracy of the network. Different operation methods can be tailored for different applications to further save the energy.

Table 3. Typical phase change synaptic power consumption.

Ref.	[46]	[47]	[48]	[31]	[32]
E_{set}/pJ	1.5–500	0.675	200–300	121	3.5
E_{reset}/pJ	50–90	50	–	1552	100

4. Conclusion

We have demonstrated that superlattice-like PCM devices can be used as artificial synapses for neuromorphic systems. The realization of gradual conductance depression improves the symmetry of synaptic weight update, which has greatly increased the accuracy of this neural network. Different strategies of voltage amplitude schemes have been selected according to different scenarios. By properly selecting the dynamic range of the device conductance, not only the power consumption, but also the performance of the neural synapses can be optimized. The simulations of this neural network based on the above device show great accuracy of 91.7% in recognizing the handwritten digits. Our results may facilitate the design of the neuromorphic hardware systems based on superlattice-like PCM devices.

References

[1] Yao P, Wu H Q, Gao B, Eryilmaz S B and Qian H 2017 *Nat. Commun.* **8** 15199
[2] Yang J J, Strukov D B and Stewart D R 2013 *Nat. Nanotechnol.* **8** 13
[3] Chi-Sang P and Kuan Z 2011 *Frontiers in Neuroscience* **5** 108
[4] Yu S M 2018 *Proc. IEEE* **106** 260
[5] Irem B, Manuel L G, R N S, Timoleon M, Thomas P, Tomas T, Bipin R, Yusuf L, Abu S and Evangelos E 2018 *Nat. Commun.* **9** 1
[6] Merolla P A, Arthur J V, Alvarez-Icaza R, Cassidy A S, Sawada J, Akopyan F, Jackson B L, Imam N, Guo C and Nakamura Y 2014 *Science* **345** 668
[7] Han S, Liu X Y, Mao H Z, Pu J, Pedram A, Horowitz M A and Dally W J 2016 *ACM SIGARCH Comput. Architecture News* **44** 243
[8] Sebastian A, Le Gallo M, Burr G W, Kim S, BrightSky M and Eleftheriou E 2018 *J. Appl. Phys.* **124** 111101
[9] Papandreou N, Pantazi A, Sebastian A, Eleftheriou E, Breitwisch M, Lam C and Pozidis H 2010 *Solid-State Electron.* **54** 991
[10] Wang Z R, Joshi S, Savel'ev S E, Jiang H, Midya R, Lin P, Hu M, Ge N, Strachan J P and Li Z Y 2017 *Nat. Mater.* **16** 101
[11] Lin C, Tian-Yu W, yawei d, Ming-Yang C, Hao Z, Qing-Qing S, Shi-Jin D, Peng Z, Leon C and Wei Z D 2018 *Nanoscale* **10** 15826
[12] Ambrogio S, Balatti S, Milo V, Carboni R, Wang Z, Calderoni A, Ramaswamy N and Ielmini D 2016 *IEEE Symposium on VLSI Technology*, June 14–16, 2016, Honolulu, HI, USA, pp. 1–2
[13] Yang Y C, Yin M H, Yu Z Z, Wang Z W, Zhang T, Cai Y M, Lu W D and Huang R 2017 *Adv. Electron. Mater.* **3** 1700032
[14] Jo S H, Chang T, Ebong I, Bhadviya B B, Mazumder P and Lu W 2010 *Nano Letters* **10** 1297
[15] Su S, Jian X C, Wang F, Han Y M, Tian Y X, Wang X Y, Zhang H Z and Zhang K L 2016 *Chin. Phys. B* **25** 107302
[16] Yao P, Wu H, Gao B, Tang J, Zhang Q, Zhang W, Yang J J and Qian H 2020 *Nature* **577** 641
[17] Wang S, He C, Tang J, Yang R, Shi D and Zhang G 2019 *Chin. Phys. B* **28** 017304
[18] Burr G, Narayanan P, Shelby R, Sidler S, Boybat I, di Nolfo C and Leblebici Y 2015 *IEEE International Electron Devices Meeting (IEDM)*, December 7–9, 2015, Washington, DC, USA, pp. 4.4.1–4.4.4
[19] Choi S, Tan S H, Li Z, Kim Y, Choi C, Chen P Y, Yeon H, Yu S and Kim J 2018 *Nat. Mater.* **17** 335
[20] Xie Y J, Kim W, Kim Y, Kim S, Gonsalves J, BrightSky M, Lam C, Zhu Y and Cha J J 2018 *Adv. Mater.* **30** 1705587
[21] Cheng Z G, Carlos R, Nathan Y, David W C, P P W H and Harish B 2018 *Adv. Mater.* **30** 1802435
[22] Tuma T, Pantazi A, Le Gallo M, Sebastian A and Eleftheriou E 2016 *Nat. Nanotechnol.* **11** 693
[23] Sebastian A, Le Gallo M and Krebs D 2014 *Nat. Commun.* **5** 4314
[24] Sidler S, Pantazi A, Wózniak S, Leblebici Y and Eleftheriou E 2017 *International Conference on Artificial Neural Networks (ICANN)*, September 11–14, 2017, Alghero, Sardinia, Italy, pp. 281–288
[25] Ambrogio S, Gallot M, Spoon K, Tsai H, Mackin C, Wesson M, Kariyappa S, Narayanan P, Liu C C and Kumar A 2019 *IEEE International Electron Devices Meeting (IEDM)*, December 9–11, 2019, San Francisco, USA, pp. 6.1.1–6.1.4
[26] Piveteau C, Le Gallo M, Khaddam-Aljameh R and Sebastian A 2019 *IEEE 11th International Memory Workshop (IMW)*, May 12–15, 2019, Monterey, CA, USA pp. 1–4
[27] Kim W, Bruce R, Masuda T, Fraczak G, Gong N, Adusumilli P, Ambrogio S, Tsai H, Bruley J and Han J P 2019 *Symposium on VLSI Technology*, June 9–14, 2019, Kyoto, Japan pp. T66–T67
[28] Tsai H, Ambrogio S, Mackin C, Narayanan P, Shelby R, Rocki K, Chen A and Burr G 2019 *Symposium on VLSI Technology*, June 9–14, 2019, Kyoto, Japan pp. T82–T83
[29] Zhou X L, Xia M J, Rao F, Wu L C and Zhang S B 2014 *Acs Appl. Mater Interfaces* **6** 14207
[30] Burr G W, Kurdi B N, Scott J C, Lam C H, Gopalakrishnan K and Shenoy R S 2008 *IBM J. Res. & Dev.* **52** 449
[31] Bichler O, Suri M, Querlioz D, Vuillaume D, DeSalvo B and Gamrat C 2012 *IEEE Trans. Electron. Devices* **59** 2206
[32] Burr G W, Shelby R M, Sidler S, Di Nolfo C, Jang J, Boybat I, Shenoy R S, Narayanan P, Virwani K and Giacometti E U 2015 *IEEE Trans. Electron. Devices* **62** 3498
[33] Stefano A, Pritish N, Hsinyu T, M S R, Irem B, Carmelo d N, Severin S, Massimo G, Martina B and P F N C 2018 *Nature* **558** 60
[34] Simpson R E, Fons P, Kolobov A V, Fukaya T, Krbal M, Yagi T and Tominaga J 2011 *Nat. Nanotechnol.* **6** 501
[35] Kalikka J, Zhou X, Dilcher E, Wall S, Li J and Simpson R E 2016 *Nat. Commun.* **7** 11983
[36] Zhou X L, Kalikka J, Ji X L, Wu L C, Song Z T and Simpson R E 2016 *Adv. Mater.* **28** 3007
[37] Momand J, Wang R, Boschker J E, Verheijen M A, Calarco R and Kooi B J 2015 *Nanoscale* **7** 19136
[38] Zhang W, Thiess A, Zalden P, Zeller R, Dederichs P H, Raty J Y, Wuttig M, Blügel S and Mazzarello R 2012 *Nat. Mater.* **11** 952
[39] Xian B L, Nian K C, Xue P W and Hong B S 2018 *Adv. Funct. Mater.* **28** 1803380
[40] Rummelhart D E, Hinton G E and Williams R J 1986 *Nature* **323** 533
[41] Krizhevsky A, Sutskever I and Hinton G 2012 *Advances in neural information processing systems* **25** 1097
[42] Cireşan D, Meier U and Schmidhuber J 2012 *IEEE conference on computer vision and pattern recognition*, June 16–21, 2012, Providence, RI, USA p. 3642
[43] Simard P, Steinkraus D and Platt J C 2003 *International Conference on Document Analysis and Recognition*, August 6–6, 2003, Edinburgh, UK, UK p. 3
[44] Ranzato M A, Poultney C, Chopra S and Lecun Y 2007 *Advances in Neural Information Processing Systems* (Boston: MIT Press) p. 1137
[45] Tong H, Miao X S, Cheng X M, Wang H, Zhang L, Sun J J, Tong F and Wang J H 2011 *Appl. Phys. Lett.* **98** 101904
[46] Kuzum D, Jeyasingh R G D, Yu S and Wong H S P 2012 *IEEE Trans. Electron. Devices* **59** 3489
[47] Kuzum D, Jeyasingh R G D, Lee B and Wong H S P 2012 *Nano Lett.* **12** 2179
[48] Kang D H, Jun H G, Ryoo K C, Jeong H and Sohn H 2015 *Neurocomputing* **155** 153