


# Lightweight and highly robust memristor-based hybrid neural networks for electroencephalogram signal processing

Peiwen Tong(童霈文), Hui Xu(徐晖), Yi Sun(孙毅), Yongzhou Wang(汪泳州), Jie Peng(彭杰), Cen Liao(廖岑), Wei Wang(王伟), and Qingjiang Li(李清江)

**Citation:**Chin. Phys. B, 2023, 32 (**7**): 078505. DOI: 10.1088/1674-1056/ac9cbc
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 throughput *N*-modular redundancy for error correction design of memristive stateful logic

Xi Zhu(朱熙), Hui Xu(徐晖), Weiping Yang(杨为平), Zhiwei Li(李智炜), Haijun Liu(刘海军), Sen Liu(刘森), Yinan Wang(王义楠), and Hongchang Long(龙泓昌)

Chin. Phys. B, 2023, 32 (**1**): 018502. DOI: 10.1088/1674-1056/ac8f38

## Fabrication and investigation of ferroelectric memristors with various synaptic plasticities

Qi Qin(秦琦), Miaocheng Zhang(张缪城), Suhao Yao(姚苏昊), Xingyu Chen(陈星宇), Aoze Han(韩翱泽),Ziyang Chen(陈子洋), Chenxi Ma(马晨曦), Min Wang(王敏), Xintong Chen(陈昕彤), Yu Wang(王宇),Qiangqiang Zhang(张强强), Xiaoyan Liu(刘晓燕), Ertao Hu(胡二涛), Lei Wang(王磊), and Yi Tong(童祎)

Chin. Phys. B, 2022, 31 (**7**): 078502. DOI: 10.1088/1674-1056/ac3ece

## Universal memory based on phase-change materials: From phase-change random access memory to optoelectronic hybrid storage

Bo Liu(刘波), Tao Wei(魏涛), Jing Hu(胡敬), Wanfei Li(李宛飞), Yun Ling(凌云), Qianqian Liu(刘倩倩), Miao Cheng(程淼), and Zhitang Song(宋志棠)

Chin. Phys. B, 2021, 30 (**5**): 058504. DOI: 10.1088/1674-1056/abeedf

## Recent progress in optoelectronic neuromorphic devices

Yan-Bo Guo(郭延博), Li-Qiang Zhu(竺立强)

Chin. Phys. B, 2020, 29 (**7**): 078502. DOI: 10.1088/1674-1056/ab99b6

## A method of generating random bits by using electronic bipolar memristor

Bin-Bin Yang(杨彬彬), Nuo Xu(许诺), Er-Rui Zhou(周二瑞), Zhi-Wei Li(李智炜), Cheng Li(李成), Pin-Yun Yi(易品筠), Liang Fang(方粮)

Chin. Phys. B, 2020, 29 (**4**): 048505. DOI: 10.1088/1674-1056/ab77fd

---

# Lightweight and highly robust memristor-based hybrid neural networks for electroencephalogram signal processing

Peiwen Tong(童霈文), Hui Xu(徐晖), Yi Sun(孙毅), Yongzhou Wang(汪泳州), Jie Peng(彭杰), Cen Liao(廖岑), Wei Wang(王伟)[†], and Qingjiang Li(李清江)[‡]

*College of Electronic Science and Technology, National University of Defense Technology, Changsha* 410073, *China*

(Received 2 August 2022; revised manuscript received 22 September 2022; accepted manuscript online 21 October 2022)

Memristor-based neuromorphic computing shows great potential for high-speed and high-throughput signal processing applications, such as electroencephalogram (EEG) signal processing. Nonetheless, the size of one-transistor one-resistor (1T1R) memristor arrays is limited by the non-ideality of the devices, which prevents the hardware implementation of large and complex networks. In this work, we propose the depthwise separable convolution and bidirectional gate recurrent unit (DSC-BiGRU) network, a lightweight and highly robust hybrid neural network based on 1T1R arrays that enables efficient processing of EEG signals in the temporal, frequency and spatial domains by hybridizing DSC and BiGRU blocks. The network size is reduced and the network robustness is improved while ensuring the network classification accuracy. In the simulation, the measured non-idealities of the 1T1R array are brought into the network through statistical analysis. Compared with traditional convolutional networks, the network parameters are reduced by 95% and the network classification accuracy is improved by 21% at a 95% array yield rate and 5% tolerable error. This work demonstrates that lightweight and highly robust networks based on memristor arrays hold great promise for applications that rely on low consumption and high efficiency.

**Keywords:** memristor, lightweight, robust, hybrid neural networks, depthwise separable convolution, bidirectional gate recurrent unit (BiGRU), one-transistor one-resistor (1T1R) arrays

**PACS:** 85.35.–p, 84.37.+q, 07.05.Mh, 87.19.lv **DOI:** 10.1088/1674-1056/ac9cbc

## 1. Introduction

With the booming development of brain–computer synergy technology in recent years, brain electrode arrays have been expanded and the accuracy and speed of electroencephalogram (EEG) signal acquisition have been continuously improved. This has led to a dramatic increase in the amount of EEG data and has placed higher demands on the processing.[1–5] The traditional EEG signal processing system is affected by the high latency and power consumption resulting from the A/D conversion and separation architecture of the memory and computing unit architecture, which makes it more and more difficult to meet the increasing demand for high-speed, high-throughput EEG signal processing.[6,7]

The emergence of EEG signal processing systems based on memristors has been a boon for EEG signal processing.[8–10] With their good switching characteristics, excellent retention characteristics, static linear $I$–$V$ characteristics and biological likelihood characteristics, memristors can achieve fast computing with low power consumption and low latency in the analogue domain. However, the most advanced one-transistor one-resistor (1T1R) arrays are affected by the fabrication process and suffer from non-idealities such as array yield rate and device weight fluctuation.[11] These lead to higher limits of scale and demand robustness of the network. The typical end-to-end processing algorithms for EEG signals mainly use complex and massive multiply accumulate operations to pursue extreme performance.[12,13] This brings challenges to the implementation of networks in memristor arrays.

Here we propose the depthwise separable convolution and bidirectional gate recurrent unit (DSC-BiGRU) network, a lightweight and highly robust hybrid neural network based on 1T1R arrays, which extracts and learns the features of EEG signals from the temporal, frequency and spatial domains by using a hybrid of a convolutional neural network (CNN) and recurrent neural network (RNN). The novelty of the proposed network lies in improving the efficiency of network learning and reducing the utilization of resources. Within the network, the DSC block is used to reduce the network complexity and improve the robustness of the network[14] while the BiGRU block is used to improve the learning efficiency to reduce the size of the network. Simulated results show a 95% reduction in network parameter resources compared with the traditional convolutional networks DeepConvNet (DCN) and ShallowConvNet (SCN).[15] With a 95% array yield rate and 5% tolerance error, the network classification accuracy is improved by 21% to 85%. These results show that the DSC-BiGRU network can effectively achieve lightweight and highly robust EEG signal processing, thus providing a new solution for the application of memristor neuromorphic computing in brain–computer interfaces.

[†]Corresponding author. E-mail: wangwei_esss@nudt.edu.cn
[‡]Corresponding author. E-mail: qingjiangli@nudt.edu.cn

© 2023 Chinese Physical Society and IOP Publishing Ltd 

## 2. Application of EEG signal analysis system based on memristor arrays

### 2.1. EEG signal analysis system based on memristor arrays

Figure 1 illustrates an EEG signal analysis system based on memristor crossbar arrays. A complete EEG signal processing application (EEG communication, motor aids, metaverse, etc.) can be designed by integrating the system with neural probes for application scenarios. Multiply-accumulate operations are usually the most critical and computationally intensive part of EEG processing algorithms, and memristor arrays can achieve parallel multiply-accumulate calculations in the analog domain. This overcomes the limitations caused by the Von Neumann architecture and A/D conversion processing, providing an efficient hardware platform to perform various processing algorithms and helping to improve the latency, power consumption and scalability of applications. However, due to the limitation of present manufacturing and fabrication processes, the memristor array suffers from size and non-ideality problems.[11] Therefore, a high-performance neural network that can adapt to practical memristor arrays and still meet the high requirements of EEG signal processing is of significant importance.

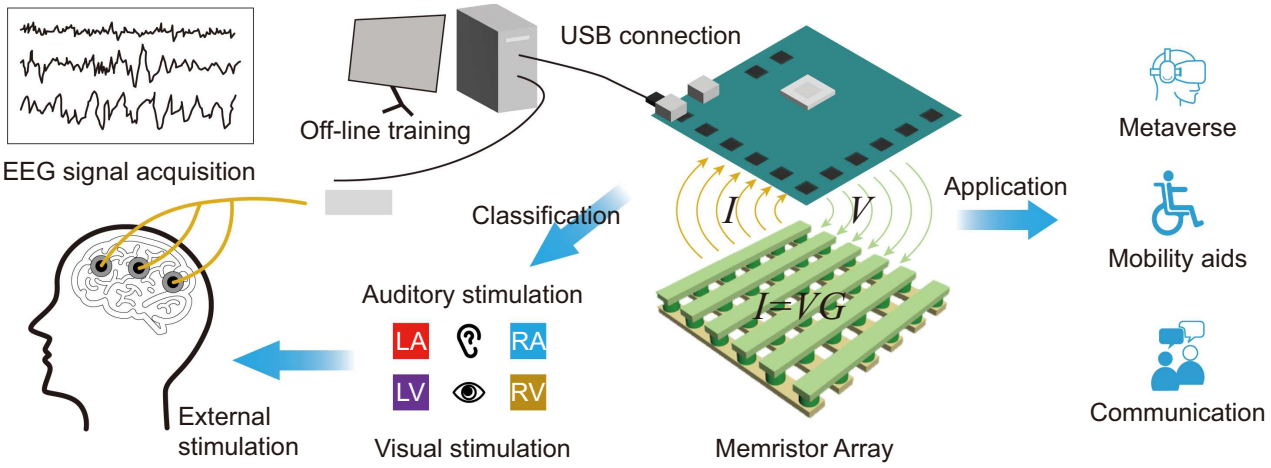



**Fig. 1.** Schematic diagram of brain–computer interface computing based on memristors. It consists of three parts: acquisition of the EEG signal by neural probes, recognition of the signal by a memristor-based processing system and application scenarios.

### 2.2. Description of the application task

An EEG signal classification task is used to verify the performance of the network. The application tasks are described in terms of dataset, network model training and network validation.

The performance of the network is validated by classifying event-related potential (ERP) EEG data in a four-class classification task using the sample dataset provided in the MNE package.[16,17] The data are acquired using the Neuromag Vectorview system at the MGH/HMS/MIT Athinoula A. Martinos Centre for Biomedical Imaging. Two hundred and eighty eight samples of EEG data are taken from a 60-channel electrode cap. During data acquisition, checkerboard patterns are presented in the subject's left and right visual fields and interspersed by tones in the left and right ears. The time interval between stimuli is 750 ms. A smiley face will incidentally appear in the center of the visual field. Subjects are asked to press a key with their right index finger soon after the appearance of the smiley face. The four classes used from this dataset are LA, left-ear auditory stimulation; RA, right-ear auditory stimulation; LV, left visual field stimulation; RV, right visual field stimulation.

In order to achieve the four-classification task, the network model is fitted by the Adam optimizer to minimize the classification cross-entropy loss function. The training process ends after 300 epochs and the model weights that produce the lowest validation set loss are saved. Dropout techniques are used to prevent overfitting during small sample training with a dropout rate of 0.5. The Keras API in TensorFlow on an NVIDIA GeForce RTX 2060 is used to train all models. It is worth noting that we omit the use of bias units in all convolutional layers for the purposes of hardware implementation.

Considering the limited size of the dataset we used, the simulation uses five-fold cross-validation to obtain network classification accuracy. This is achieved by dividing the dataset into five parts, with one part being taken for validation while the remaining four parts are used for training. In this way five different results are obtained, and the final result is obtained by averaging the five validation results. This extracts as much valid information as possible from the limited dataset to obtain more accurate results. In addition, the simulation is repeated 10 times to obtain more comprehensive results.

### 2.3. Device characterization and analysis

For reliable verification of the network performance and successful hardware implementation of the network, the char-

acteristics of practical memristors are measured. The writing error of the weight modulation is statistically analyzed and data for the weight fluctuation are provided to support the subsequent performance simulation.

The memristor measurements are based on our fabricated TiN/HfO$_x$/TaO$_x$/TiN 1T1R arrays. The electrical characteristics of the memristor are characterized using a Keithley 4200 SCS parameter analyzer and the board-level verification system equipped with a 1T1R array.

A DC voltage is applied to the gate of the target cell when we perform the modulation operation. During the SET process, the excitation is applied at the top electrode while the source electrode is grounded, as shown in the inset of Fig. 2(a). On the contrary, the excitation is applied at the source electrode and grounded at the top electrode during the RESET process. In Fig. 2(a), measured results show that our 1T1R device possesses excellent bidirectional analogue switching behavior, which allows the device conductance to be modulated continuously during SET and RESET. This will aid precise mapping of the network parameters. Importantly, the multi-conductance level exhibits good retention characteristics, as shown in Fig. 2(b). In order to reduce the total measurement time, 40 randomly selected devices are modulated to 40 different conductance states and the conductance values are measured every hour. These devices maintain a stable conductance state over a 24-h period. The results of the 1T1R cell measurements show that the device has good bipolar switching characteristics and retention characteristics to support the validation of the network in the array.

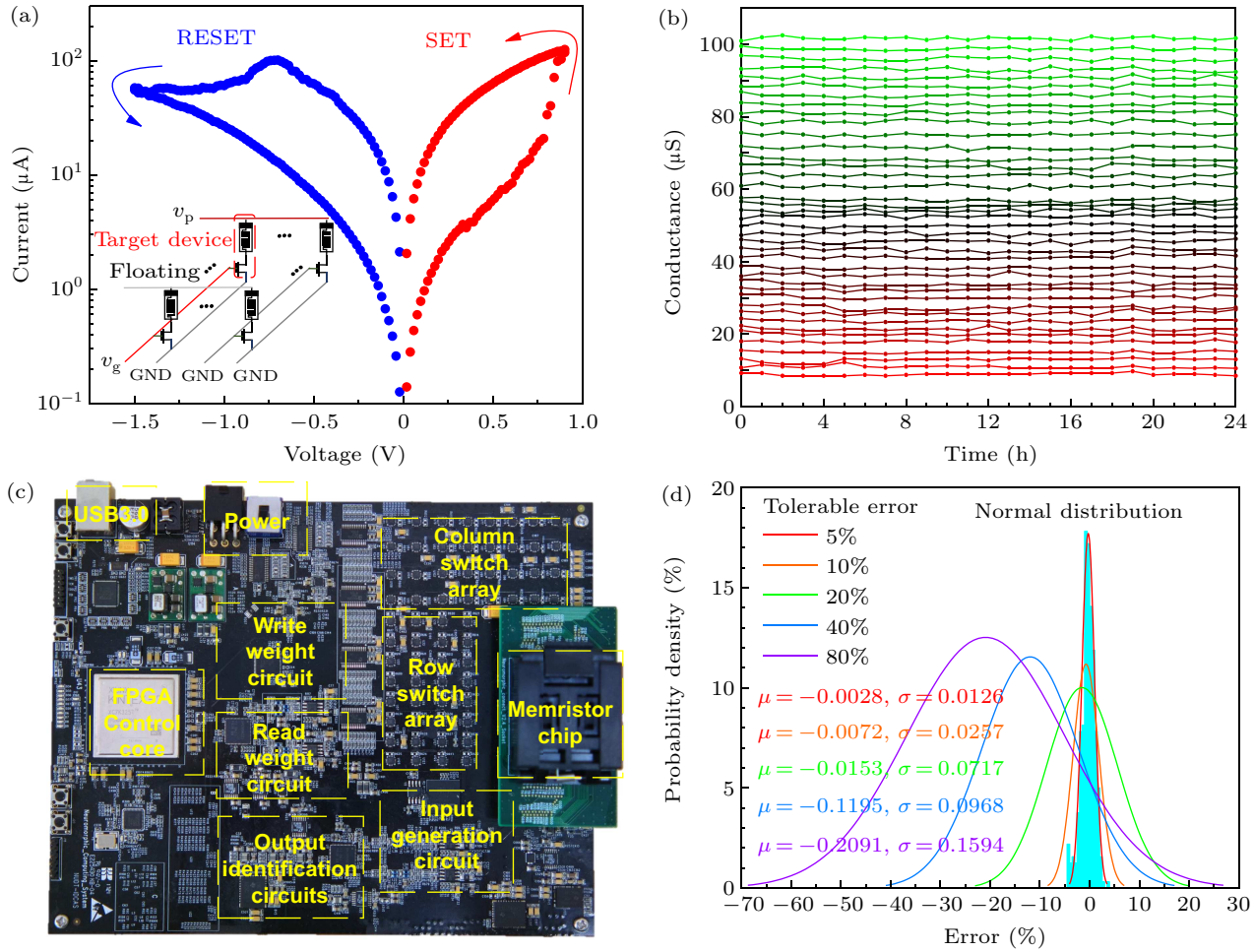



**Fig. 2.** (a) The 1T1R cell DC scan characteristics. (b) Retention characteristics of the 1T1R cell. (c) Composition of the 1T1R chip board-level verification system. (d) Statistical analysis of the weight modulation error under different tolerable errors. The tolerable error is the range of tolerable device conductance fluctuations at the end of modulation.

We explore a board-level verification system embedded with a 1T1R array to measure memristor characteristics and verify network parameter mapping, as shown in Fig. 2(c). The board has three circuits for writing weights, reading weights and identifying output currents, respectively. In this paper we use a variable gate voltage modulation method with rough and precise modulation performed in cooperation. This method is inspired by the variable gate voltage modulation method,[19] where a threshold value of ±20% from the target resistance value is set, with precise modulation within the threshold and rough modulation outside the threshold. Both rough and precise modulation increase the gate voltage gradually according

to the present conductance value. The difference between the two modulations is the different step size of the increase in gate voltage: rough modulation has a larger step size to facilitate fast convergence, while precise modulation has a smaller step size, which improves accuracy. In addition, different tolerable errors can be set to meet the requirements of different applications. For example, with a 5% tolerance error, the modulation is considered successful when the device resistance is modulated to within ±5% of the target resistance (5% near the ultimate in system accuracy).

The system and modulation method are verified for experiments. The 10 μS–100 μS range is divided equally into 32 conductance states with 10 devices in each conductance state. Measurements are performed at different tolerable errors. During error analysis, the practical error of the device meets the requirement of less than 5%, as shown in the cyan histogram in Fig. 2(d). Then we perform a statistical analysis and find it obeys a standard distribution ($\mu = -0.0028$, $\sigma = 0.012619$). Similarly, following the same approach, the statistical analysis is implemented for results with tolerable errors of 10%, 20%, 40% and 80%, as shown in Fig. 2(d). The results show that both ends of the fitted curves are basically included within the setting threshold (the mean values of the errors are −0.0072, −0.0153, −0.1195, and −0.2091, respectively), which proves that the modulation method can effectively achieve the tolerable error requirement. We find that starting each measurement with a low conductance state causes the center of the standard distribution to shift toward the negative axis. The statistical analysis of the weight modulation accuracy digitizes the random weight fluctuations to provide practical performance of the devices for the 1T1R array validation of the network, which makes the validation more reliable.

## 3. SCN-BiGRU network structure and performance simulation

### 3.1. SCN-BiGRU network structure

In order to complete the network implementation on the memristor, it is of primary importance to reduce the network's size. The proposed SCN-BiGRU network is an innovative small-scale network based on the SCN network. It is implemented by shrinking the class of convolutional kernels of the SCN network and adding recurrent neural networks. A visualization and full description of the network model is shown in Fig. 3(a), where the input is an EEG trial with $C$ channels (here 60) and $S$ time samples (here 151). The network consists of two blocks: the first is the SCN block, which uses CNN to extract and learn spatial and frequency features in the EEG signal. The second is the BiGRU block, which learns temporal feature information about the EEG signal through RNN learning and finally obtains classification results.

The SCN block extracts and learns spatial features by two sequentially performed convolutions. First, $N_1$ (here four) convolution filters of size $(1, L_1)$ are set, whose length is determined based on the sampling frequency of the data [here (1, 16)]. A convolution kernel of height 1 can be used as a filter, and the output contains EEG features with different band passes. This achieves feature extraction in the frequency domain. In addition, a convolution kernel with a height of 1 is equivalent to only handling the spatial features of the signal within the same moment. Then the signal space features are extracted using $N_2$ spatial filters of size $(C, 1, N_1)$, where $N_2 = 8$ and controls the number of spatial filters to be learned. In CNN image processing applications, deep directional convolution reduces the size of the parameters that need to be fitted and each convolutional kernel does not need to learn all the input data. Inspired by the filter-bank common spatial pattern (FBCSP)[18] algorithm for feature extraction, a separate spatial filter is given to the output of each temporal filter using depth directional convolution. Before applying exponential linear units (ELUs), batch normalization is applied along the dimensions of the features. Then, an average pooling layer of size (1, 7, 35) is used to further converge the features with a size of (1, 35) and a stride of 7.

In the BiGRU block, the obtained feature matrix is cut into blocks in temporal order [here, 15 (1,8) feature vectors] and sent to the GRU sequentially. This achieves learning features in the temporal domain. The GRU has two gates, a reset gate and an update gate. The reset gate determines how the new input information is combined with previous memories and the update gate defines how much of the previous memory is saved to the present time. A single layer of bi-directional GRUs is used which provides the complete past and future information for each cell in the input sequence. The parameter $U$ is set to determine the size of the output vector (here, 16). Then the output features are normalized and activated (ELUs). Finally, the information which condenses the temporal, frequency and spatial characteristics of the signal is sent into the dense layer and classified by the SoftMax function.

### 3.2. SCN-BiGRU network performance simulation

Hybrid neural networks of CNN and RNN are used to learn signal features in three dimensions (time, frequency and space). This improves the network learning efficiency and reduces the size of the network needed to accomplish the same work. In the comparison process, besides our proposed SCN-BiGRU hybrid neural network, three other common combinations are compared. The classification accuracy rates and scales of four hybrid networks (SG = SCN-GRU, SBG = SCN-BiGRU, SL = SCN-LSTM, SBL = SCN-BiLSTM) are compared with the traditional DCN and SCN, as shown in Fig. 3(b).

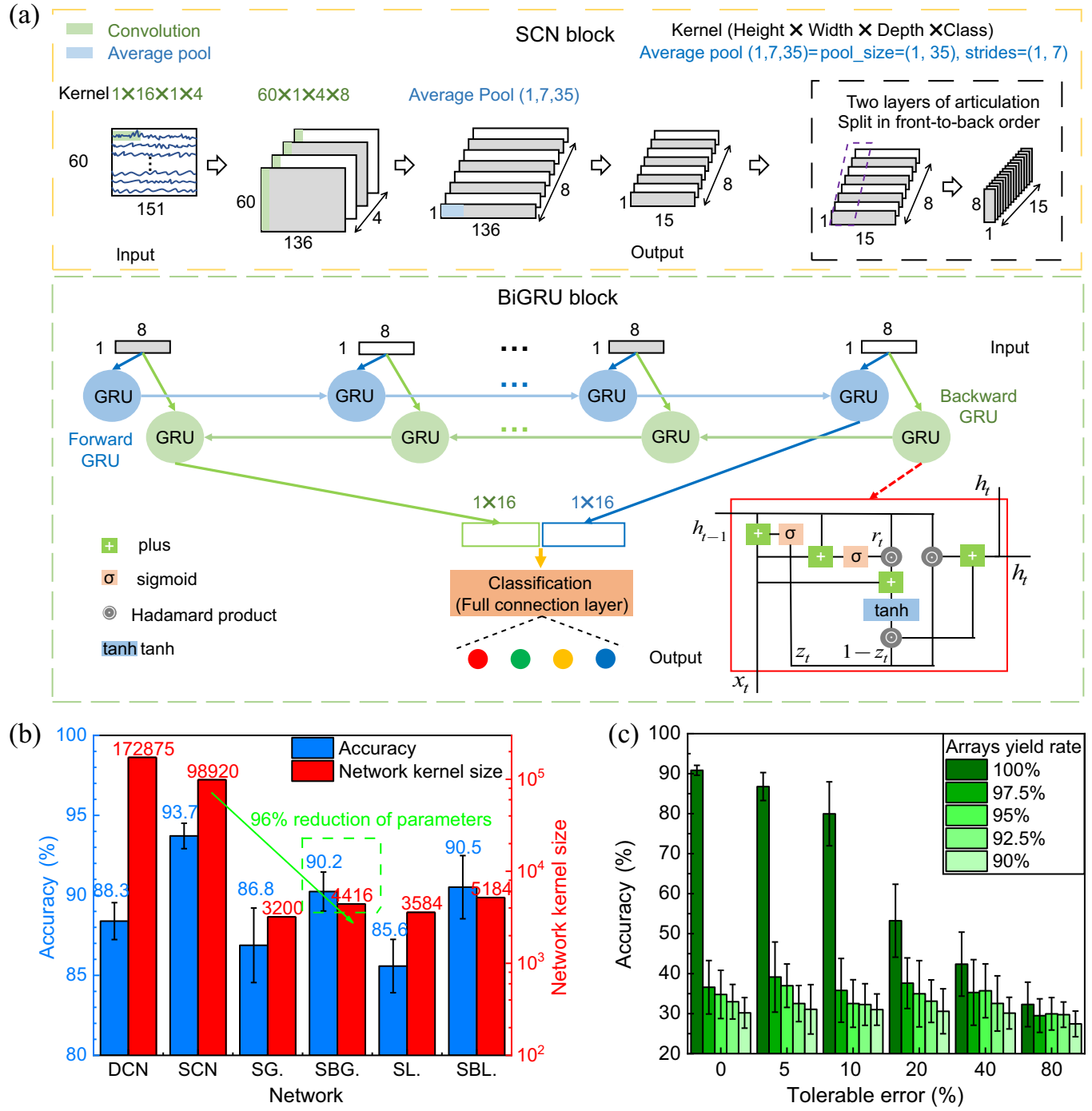



**Fig. 3.** (a) Overall visualization of the SCN-BiGRU network architecture. It mainly consists of a CNN block (SCN block) and a RNN block (BiGRU block). (b) Comparison of classification accuracy and scale of the hybrid neural network and conventional convolutional network. The SCN-BiGRU network obtains the most balanced performance. (c) Robustness analysis of SCN-BiGRU.

It can be seen from the simulated results that all four hybrid networks can effectively reduce the network size. The SCN-BiGRU network provides the most balanced network classification performance. The bi-directional recurrent network reveals the past and future information of the EEG signal at each time point, so higher accuracy is obtained. The SCN-BiGRU network classification accuracy rate is reduced by only 0.3%, with a 15% reduction in network parameters compared with SCN-BiLSTM. The GRU uses two gates to achieve the information storage and transfer that the LSTM achieves with three gates, which saves network resources and provides a more balanced performance with limited datasets. Compared with the superior SCN of the two convolutional networks, the network parameters are reduced by 96% while the classification accuracy is only reduced by 3.5%.

However, the network not only needs to meet the array size but must also adapt to the non-ideal characteristics of the memristors for the network to be implemented on an array of memristors. It mainly includes two aspects: on the one hand, the arrays have a limited yield rate with damaged de-

vices showing a high resistance state; on the other hand, there are writing errors in the weights, which fluctuate around the target value. Therefore, array yield rate and array errors are introduced to verify the robustness of the network. Some parameters are placed at 0, depending on the array yield rate, to observe the effect of different yield rates on network classification accuracy. The volatility of the weights is based on the statistical results of the practical measured weight writing errors under different tolerable errors, and the normal distribution of errors is applied to the parameters of the network. The results are shown in Fig. 3(c), and the robustness of SCN-BiGRU is very poor in meeting the needs of the memristor array.

Therefore, a lightweight and highly robust hybrid neural network is important in EEG signal processing based on 1T1R arrays.

## 4. DSC-BiGRU network structure and performance simulation

The proposed DSC-BiGRU network is a lightweight and highly robust hybrid neural network structure based on 1T1R arrays. The novelty of the network is that it can be implemented on a practical memristor array of limited size. A visualization and full description of the network model are shown in Fig. 4 and Table 1. The difference from the SCN-BiGRU network is the DSC block for the convolution part.

**Table 1.** The DSC-BiGRU architecture.

| Block | Layer | Input | Size | Kernel parameters | Output |
|---|---|---|---|---|---|
| 1 | Input | | | | $(C, S)$ |
| | Frequency Conv | $(C, S)$ | (1, 16, 1, 8) | $(1, L_1, 1, N_1)$ | $(C, S, N_1)$ |
| | Spatial Conv | $(C, S, N_1)$ | (60, 1, 8, 2) | $(C, 1, N_1, N_2)$ | $(1, S, N_3)$ |
| | Averagepool1 | | (1, 4) | | $(1, S/4, N_3)$ |
| | Depthwise Conv | $(1, S/4, N_3)$ | (1, 8, 16, 1) | $(1, L_2, N_3, 1)$ | $(1, S/4, N_3)$ |
| | Separable Conv | $(1, S/4, N_3)$ | (1, 1, 16, 16) | $(1, 1, N_3, N_3)$ | $(1, S/4, N_3)$ |
| | Averagepool2 | | (1, 8) | | $(1, S/32, N_3)$ |
| 2 | Forward GRU/backward GRU | $(1, S/32, N_3)$ | $(1, 16, 48, 2)/(1, 16, 48, 2)$ | $(1, N_3, 3U, 2)/(1, U, 3U, 2)$ | $(2U)$ |
| Classifier | Dense (SoftMax) | $(2U)$ | (32, 4) | $(2U, n)$ | $n$ |
| Note | | $n$ is the number of categories, $N_3 = N_1 N_2$ | | | |

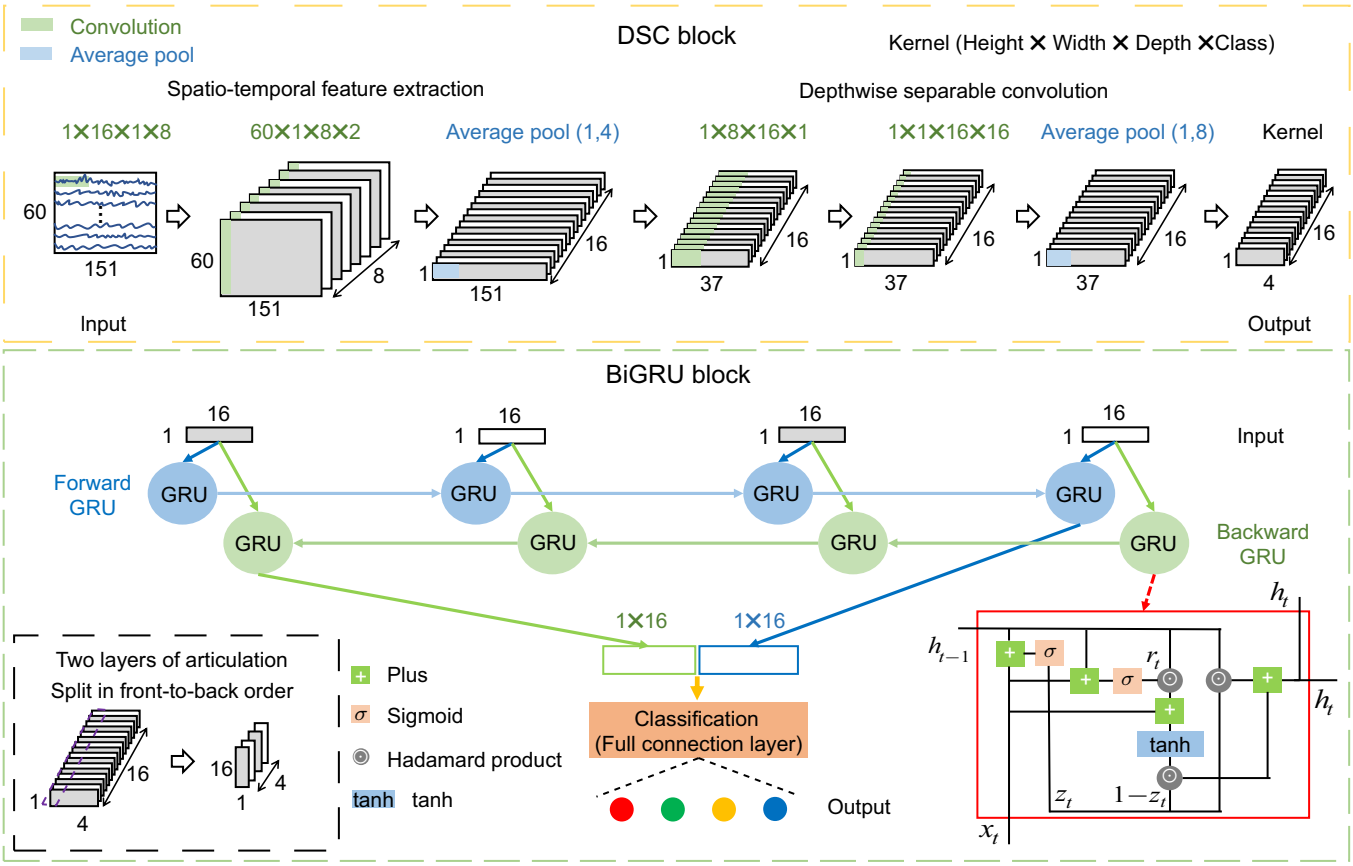



**Fig. 4.** Overall visualization of the DSC-BiGRU network architecture. It mainly consists of a CNN block (DSC block) and a RNN block (BiGRU block).

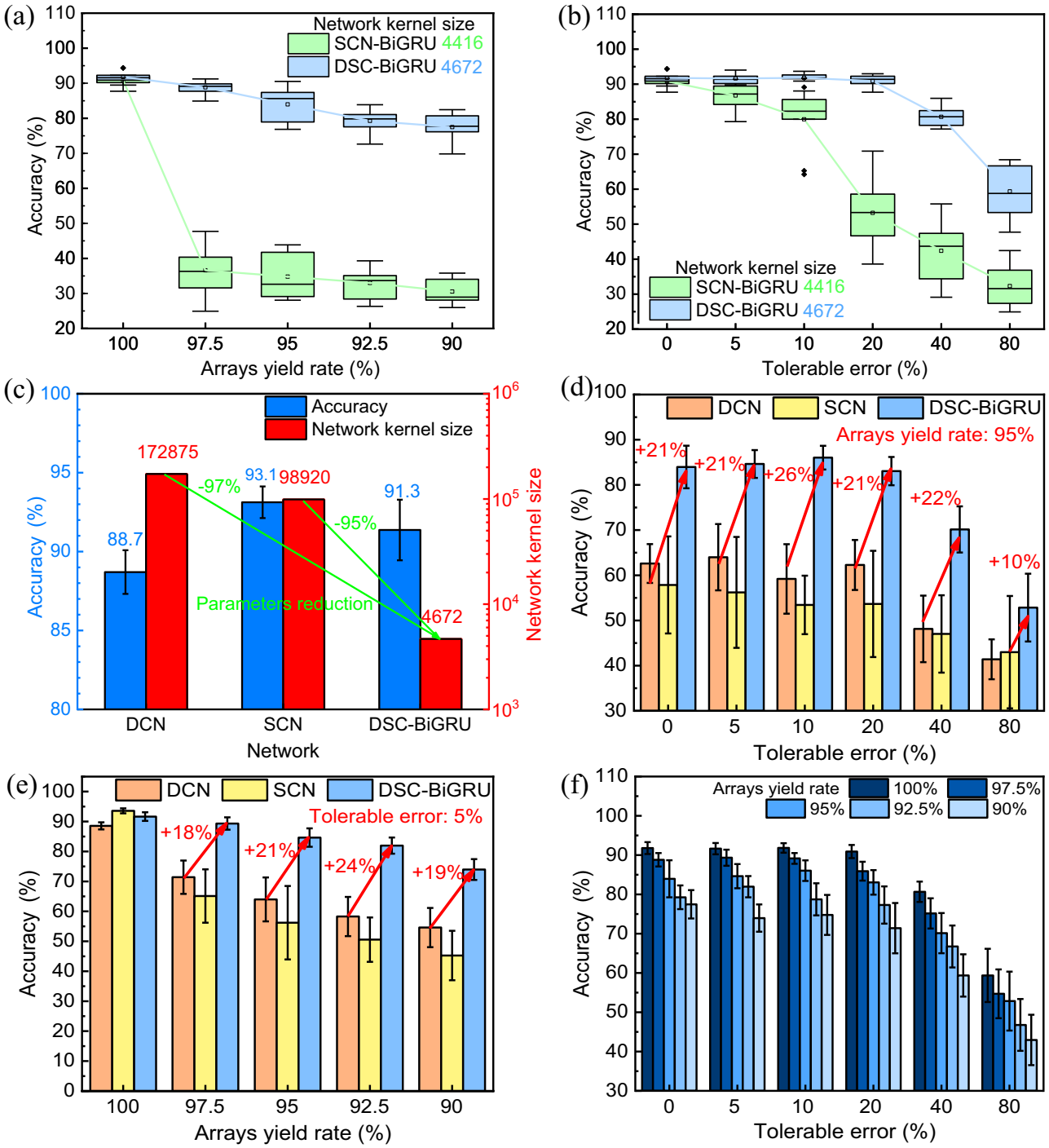



**Fig. 5.** (a) The effect of array yield rate on the accuracy of SCN-BiGRU and DSC-BiGRU networks. (b) The effect of array tolerance error on the accuracy of SCN-BiGRU and DSC-BiGRU networks. (c) Comparison of the classification accuracy and size of DSC-BiGRU and traditional convolutional networks (DCN, SCN). (d) The effect of array tolerance error on the accuracy of DSC-BiGRU and traditional convolutional networks (DCN, SCN). (e) The effect of array yield rate on the accuracy of DSC-BiGRU and traditional convolutional networks (DCN, SCN). (f) Robustness analysis of DSC-BiGRU.

The DSC block can be divided into two parts: spatial feature extraction and depthwise separable convolution. In spatial feature extraction, the two convolution steps are similar to the SCN-BiGRU network. First, eight convolution filters of size (1, 16) are set. Then two (60, 1, 8) spatial filters are used to extract the signal spatial features. Compared with the SCN-BiGRU network, the number of classes of convolutional kernels is compressed. Before applying ELUs, batch normalization is applied along the dimensions of the features. Then, an average pooling layer of size (1, 4) is used to further converge the features. Following the feature extraction process, features are learned using a depth-separable convolution. This is a depthwise convolution [here, of size (1, 8, 16)] followed by pointwise convolutions [here, of size (1, 1, 16, 16)] which reduces the convolution kernel parameters and improves network robustness compared with traditional CNN. The princi-

ple applies to multi-band feature signals such as EEG signals where the features in each feature map are first learned individually and then weighted and combined according to their importance. This allows for effective learning of features within different frequency bands as well as integrated learning of features across different frequency bands. Finally, an average pooling layer of size (1,8) is used for feature convergence.

In the BiGRU block, the obtained feature matrix is cut into blocks in temporal order [here, 4 (1,16) feature vectors] according to the same method as the SCN-BiGRU network and sent to the GRU sequentially. The information about the past and future of the EEG signal features is learned in BiGRU and the classification results are calculated in the full connection layer.

### 4.1. DSC-BiGRU network performance simulation

To verify the high accuracy of EEG signal classification achieved by DSC-BiGRU with limited scale and high robustness, we compare it with SCN-BiGRU and traditional convolutional networks (DCN, SCN). During the validation, the statistics for the practical weight modulation errors under different tolerable errors are brought into the network to test the robustness.

In Fig. 5(a), depending on the array yield rate some parameters are placed at 0 to observe the effect of different yield rates on network classification accuracy. As in Fig. 5(b), we investigated the effect of different weight fluctuations on network accuracy by adding the statistics of a normal distribution with different allowable errors to the network parameters. The results show clearly that the network size of DSC-BiGRU is comparable (4416 for SCN-BiGRU, 4672 for DSC-BiGRU) but the robustness is better than for SCN-BiGRU. This demonstrates that it is more suitable for memristor arrays than the SCN-BiGRU network.

DSC-BiGRU ensures similar network classification accuracy compared with traditional convolutional networks, with a 95% and 97% reduction in network parameters, respectively, as shown in Fig. 5(c). To compare the performance of the DCN, SCN and DSC-BiGRU networks in the case of a non-ideal memristor array, we reveal the effect of different tolerance errors (Fig. 5(d)) with a constant yield rate (95%) and the effect of different yield rates (Fig. 5(e)) with a constant tolerance error (5%), respectively. With the addition of the 1T1R array non-idealities, the performance superiority of the DSC-BiGRU network over traditional convolutional networks is even more significant. Specifically, with a 95% array yield rate and 5%, 10%, and 20% tolerance errors, the classification accuracy of the EEG signals of the DSC-BiGRU network is improved by 21%, 26%, and 22%, respectively. We also find that the impact of tolerance errors and yield rate on the network is different. As shown in Fig. 5(f), the classification accuracy of this network is stable in the error tolerance range of 0%–20%, which decreases as the array yield rate decreases. These results directly demonstrate that the impact of small deviations in all weights is less than the impact of missing some of the weights. It can be inferred that the reliability of the neural network performance in the implementation of the memristor array mainly depends on the overall mapping of the weights. This provides implications for future neural network hardware implementations in memristor arrays.

## 5. Conclusion and perspectives

In summary, we propose a lightweight and highly robust hybrid neural network (DSC-BiGRU) which utilizes the DSC and the BiGRU hybrid blocks to achieve classification of EEG signals by extracting and learning EEG signal feature information from three dimensions. The simulated experiments are performed by extracting the error distribution of the practical device. The simulated results demonstrate that the DSC-BiGRU network parameters are substantially reduced and the network classification accuracy is significantly improved compared with the DCN and SCN in the same non-ideal situation. This provides a new algorithm for implementing neural networks on larger memristor arrays in the future. Lightweight and highly robust hybrid neural networks based on memristors are of great importance for applications such as EEG signal processing.

## Acknowledgments

Project supported by the National Key Research and Development Program of China (Grant No. 2019YFB2205102) and the National Natural Science Foundation of China (Grant Nos. 61974164, 62074166, 61804181, 62004219, 62004220, and 62104256).

## References

[1] Shanechi M M 2019 *Nat. Neurosci.* **22** 1554
[2] Kao J C, Stavisky S D, Sussillo D, Nuyujukian P and Shenoy K V 2014 *Proc. IEEE* **102** 666
[3] Hong G and Lieber C M 2019 *Nat. Rev. Neurosci.* **20** 330
[4] Maharbiz M M, Muller R, Alon E, Rabaey J M and Carmena J M 2017 *Proc. IEEE* **105** 73
[5] Brandman D M, Cash S S and Hochberg L R 2017 *IEEE Trans. Neural Syst. Rehabil. Eng.* **25** 1687
[6] Liu X, Zhang M, Xiong T, Richardson A G, Lucas T H, Chin P S, Etienne-Cummings R, Tran T D and Van Der Spiegel J 2016 *IEEE Trans. Biomed. Circuits Syst.* **10** 874
[7] Cheng C H, Tsai P Y, Yang T Y, Cheng W H, Yen T Y, Luo Z, Qian X H, Chen Z X, Lin T H, Chen W H, Chen W M, Liang S F, Shaw F Z, Chang C S, Hsin Y L, Lee C Y, Ker M D and Wu C Y 2018 *IEEE J. Solid-State Circuits* **53** 3314
[8] Merkel C, Saleh Q, Donahue C and Kudithipudi D 2014 *Procedia Comput. Sci.* **41** 249
[9] Park S, Chu M, Kim J, Noh J, Jeon M, Hun Lee B, Hwang H, Lee B and Lee B G 2015 *Sci. Rep.* **5** 1
[10] Zhang J, Xu H, Long H, Hou K, Li Z, Yang Z, Wang Y and Li Q 2021 *IEEE 4th Int. Conf. Electron. Technol. ICET* **2021** 851

[11] Liu Z, Tang J, Hong B, Wu H and Gao B 2020 *Nat. Commun.* **11** 4234

[12] Liu J, Ye F and Xiong H 2021 *J. Neural Eng.* **18**

[13] Cai H 2018 *ACS Appl. Mater. Interfaces* **10** 22408

[14] Lawhern V J, Solon A J, Waytowich N R, Gordon S M, Hung C P and Lance B J 2018 *J. Neural Eng.* **15** 056013

[15] Schirrmeister R T, Springenberg J T, Fiederer L D J, Glasstetter M, Eggensperger K, Tangermann M, Hutter F, Burgard W and Ball T 2017 *Hum. Brain Mapp.* **38** 5391

[16] Ang K K, Chin Z Y, Zhang H and Guan C 2008 *Proc. Int. Jt. Conf. Neural Networks* 2390

[17] Gramfort A, Luessi M, Larson E, Engemann D A, Strohmeier D, Brodbeck C, Parkkonen L and Hämäläinen M S 2014 *Neuroimage* **86** 446

[18] Gramfort A, Luessi M, Larson E, Engemann D A, Strohmeier D, Brodbeck C, Goj R, Jas M, Brooks T, Parkkonen L and Hämäläinen M 2013 *Front. Neurosci.* **7**

[19] Chen J, Wu H, Gao B, Tang J, Zhang W and Qian H 2019 *IEEE Int. Work. Futur. Comput. IWOFC* **2019** 2019