Dynamic Memristor-based Reservoir Computing for High-Efficiency Spatiotemporal Signal Processing

Yanan Zhong
Tsinghua University
Jianshi Tang
Tsinghua University https://orcid.org/0000-0001-8369-0067
Xinyi Li
Institute of Microelectronics, Tsinghua University
Bin Gao
Tsinghua University https://orcid.org/0000-0002-2417-983X
He Qian
Institute of Microelectronics, Tsinghua University
Huaqiang Wu (wuhq@tsinghua.edu.cn)
Tsinghua University https://orcid.org/0000-0001-8359-7997

Article

Keywords: dynamic memristor, reservoir computing, waveform classification, spoken-digit recognition, handwritten-digit recognition

Posted Date: July 16th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-40717/v1

Version of Record: A version of this preprint was published at Nature Communications on January 18th, 2021. See the published version at https://doi.org/10.1038/s41467-020-20692-1.

Abstract

Reservoir computing (RC) is a highly efficient network for processing spatiotemporal signals due to its low training cost compared to standard recurrent neural networks. The design of different reservoir states plays a very important role in the hardware implementation of RC system. Recent studies have used the device-to-device variation to generate different reservoir states; however, this method is not well controllable and reproducible. To solve this problem, we report a dynamic memristor-based RC system. By applying a controllable mask process, we reveal that even a single dynamic memristor can generate rich reservoir states and realize the complete reservoir function. We further build a parallel RC system that can efficiently handle spatiotemporal tasks including spoken-digit and handwritten-digit recognitions, in which high classification accuracies of 99.6% and 97.6% have been achieved, respectively. The performance of dynamic memristor-based RC system is almost equivalent to the software-based one. Besides, our RC system does not require additional read operations, which can make full use of the device nonlinearity and further improve the system efficiency. Our work could pave the road towards high-efficiency memristor-based RC systems to handle more complex spatiotemporal tasks in the future.

Introduction

In recent years, artificial neural networks (ANNs) have developed rapidly and played an important role in many different fields such as object detection,[1,2] natural language processing,[3] autonomous driving,[4] security,[5] *etc*. Generally, ANNs can be loosely divided into two main categories depending on the network structure. One is feedforward neural networks in which the neurons are separated into layers and the signal only goes forward. There are many kinds of feedforward neural networks, including the well-known convolutional neural network (CNN),[6] which are widely used to process static spatial patterns such as image recognition and object detection. However, this type of network may not be suitable for processing temporal signals because of the feedforward structure. The other kind of ANNs is recurrent neural network (RNN)[7,8] in which the neurons have recurrent connections. As a result, input signal can stay in the network for a while so that short-term memory can be realized in this way. Therefore, RNN is capable of dealing with temporal tasks. Unlike feedforward neural networks, the training of RNN is usually very difficult and requires extensive computational power, mainly due to the problem of exploding or vanishing gradients in recurrent structures. In order to solve this problem, the concept of reservoir computing (RC) was proposed.[9,10] The main difference between RC and RNN is that in a RC network only the weights connected to the output layer need to be trained and the rest of the network remain fixed. As a result, the training process becomes linear and many simple training algorithms can be used, such as linear regression. At the software level, it has been shown that RC systems can achieve satisfactory performance in speech recognition,[11] adaptive filtering,[12] time series prediction[13,14] and many other fields.[15] For high system efficiency, many new materials and devices, such as spintronic oscillators,[16,17] photonic modules,[18–20] or memristors,[21–23] have been studied for the hardware implementation of RC systems. Among them, remarkable progress has been made on memristors for the implementation of [illegible]esistive switching properties.[24–29] Meanwhile, the inherent

dynamic properties and nonlinear behaviour of memristors also make them very suitable for the implementation of RC systems.[30,31] In a RC system, the richness of the reservoir states is an important factor that largely determines the system performance. In previous works, different reservoir states were generated by using device-to-device variations.[21,22] Although this method can generate many reservoir states, it is difficult to control the variation between devices and hence it is not well reproducible. Besides, in these demonstrations, the memristor conductance was regarded as the reservoir state,[21–23] so after each input signal, a read signal must be followed to read out the device conductance. This additional read operation would limit the speed of such RC systems.

In this report, we demonstrate a dynamic memristor-based RC system that using a controllable mask process to generate rich reservoir states. Besides, we directly use the memristor response to the input signal as the reservoir state, which can take advantage of the device nonlinearity and does not require additional read operations. By controlling the condition of mask process, the implemented RC system can process spatiotemporal signal efficiently. Different temporal classification tasks of waveform classification and spoken-digit recognition are demonstrated in our RC system, where an extremely low normalized root mean square error (NRMSE) of 0.14 and word error rate of 0.4% are achieved respectively. Meanwhile, the spatial classification task of handwritten-digit recognition is also performed in our system, and a high classification accuracy of 97.6% is obtained, which is close to the value obtained with a software-based RC system.

Results

Dynamic Memristor-based RC System. The dynamic memristor used in this work has a vertically stacked cross-point structure of Ti/TiO_x/TaO_y/Pt (50 nm/16 nm/30 nm/50 nm), as schematically illustrated in Fig. 1a. The cross-sectional transmission electron microscope (TEM) image of the device is shown in Fig. 1b, and the corresponding elements distribution profile from energy-dispersive spectroscopy is shown in Fig. 1c. The details of device fabrication are described in the Experimental Section. The standard memristive I–V hysteresis curves over multiple cycles are shown in Fig. 1d. The repeatable I–V loops indicate a high stability and reliability of the device. Also, the I–V curve is highly asymmetric under positive and negative voltage sweeps, which can be attributed to the Schottky barrier at the TaO_y/Pt interface.[32] Such a strong nonlinearity of the dynamic memristor can be directly used to realize the activation function commonly used in ANNs. The dynamic characteristics of the device are also explored as shown in Fig. 1e. A write voltage pulse (amplitude of 3.0 V and pulse width of 1 ms) followed by several read voltage pulses (1.9 V, 10 μs) is applied on the device and the responding current is recorded for subsequent analysis. It can be seen from Fig. 1e that the current is integrated under the large write pulse and then decays under the small read pulses, as the migration and diffusion of oxygen ions modulates the barrier height at the electrode/oxide interfaces.[32] The behaviour of current decay over time is further analyzed in Fig. 1f, where a simple exponential relationship is used to fit the curve and the characteristic time t_0 obtained by fitting is about 400 μs. These experimental results imply that the output

dent on the current input but also relies on the history of the

input signal.[33,34] Such short-term memory of the dynamic memristor gives it the ability to equivalently implement the neural network with recursive connections.[35] Combining the I–V nonlinearity and short-term memory of the device, we realized a dynamic memristor-based RC system. As a comparison, Fig. 2a shows a conventional RC system which consists of three parts: input layer, reservoir and output layer. The reservoir is the core of the RC system, which produces a large number of reservoir states that are very important for classification. Traditional approaches of making a reservoir use a network consisting of random connections of nonlinear neuron nodes. The interactions among neurons can remember the history information of the input signals and produce rich reservoir states. However, such RC architecture needs the random connections between multiple devices, which is very difficult for hardware implementation. In order to solve this problem, we incorporate the concept of time multiplexing and use a mask process to generate virtual nodes in time domain.[35] Through the dynamic and nonlinear response of the memristor, these virtual nodes are nonlinearly coupled to each other (see Fig. S1 in Supplementary Information). Figure 2b shows the schematic diagram of a dynamic memristor-based RC system based on this new architecture. Firstly, the input signal is pre-processed through a time multiplexing procedure during which the input signal is multiplied by a mask matrix and then converted to a train of voltage pulses through a signal generation system. Every frame of the input signal can generate a pulse train of total length τ and pulse width δ. Secondly, the pre-processed input is fed to the reservoir which consists of a memristor connected in series with a load resistor of R_L = 4.7 kΩ. The R_L is used to convert the memristor output current to a voltage signal which is then sampled as the reservoir states (that are the output of virtual nodes as shown in Fig. 1d). Finally, the output vector is a linear combination of the reservoir states and the weights are trained through linear regression. The details of the measurement setup are described in the Methods Section.

Waveform Classification. In the above discussion, we proposed that a simple system connecting a dynamic memristor with a resistor can be regarded as a reservoir, which can generate a large amount of reservoir states for subsequent signal processing. In order to improve the system performance in practice, several single memristor-based reservoirs are connected in parallel to build a large parallel RC system as shown in Fig. 2c. A simple waveform classification task is used to test the temporal signal processing capability of our RC system.[36,37] As shown in Fig. 2d, the input sequence is a random combination of sine and square waveforms, and the desired output is the binary sequence that consists of 0 and 1 representing sine and square waveforms respectively. To achieve the optimal classification results, we use 10 reservoirs in parallel, where the mask (a one-dimensional sequence with a length of 4 in this case) is different from each other. At the same time, the I–V nonlinearity of dynamic memristor is directly used as the activation function as shown in Fig. S4. In every time interval τ, the output of RC system is the linear combination of all the reservoir states, where the weights are trained through simple linear regression method. NRMSE is used to measure the classification error,[38] which is described as:

$$\mathrm{NRMSE} = \sqrt{\frac{\langle \| y(t) - y_{target}(t) \|^2 \rangle}{\langle \| y_{target}(t) - \langle y_{target}(t) \rangle \|^2 \rangle}}$$

where y(t) is the output of RC system, $y_{target}(t)$ is the desired output, ||·|| denotes the Euclidean norm, and <·> denotes the empirical mean. During the test, the lowest NRMSE we obtained is 0.14 and a typical result is also shown in the Fig. 2d. In addition, we find the length of the mask sequence has a critical influence on the performance of the RC system. As shown in Fig. 2e, the NRMSE of classification changes with the mask length M when keeping the reservoir size the same $M \times N = 40$ (N is the number of reservoirs in parallel). We can see that NRMSE becomes very large when the mask length is either too long or too short and reaches the minimum value as the mask length is about 4. To explain such dependence, let us consider two extreme cases with mask lengths of 40 and 1. When the mask length is as long as 40, the overall change of memristor conductance over duration τ is very large, which means the reservoir states would easily reach the upper or lower limit, thereby losing the ability to further process signals in subsequent durations. For this reason, the classification error is large when the mask length is long. On the other hand, when the mask length is as short as 1, the binary combination of the mask sequence would be very limited, which limits the type of the mask sequence. It means that in this case most of the reservoir states in the parallel RC system are the same and the effective reservoir states would not support successful classification, leading to a large classification error. So in order to achieve the best classification results, we find the appropriate mask length of 4 that yields the lowest NRMSE of 0.14. Further analysis about this and the test on cycle-to-cycle variation are discussed in Fig. S2 and Fig. S3 in Supplementary Information, respectively. Another point worth mentioning is that the RC system is based on a single memristor (i.e., $N = 1$) when the mask length is 40. It can be seen from the experimental results that the parallel RC system has a better performance than the single memristor-based RC system by adjusting the mask length (e.g., $N = 10$ when $M = 4$), which not only increases the system speed but also reduces the error rate.

Spoken-digit Recognition. To further evaluate the performance of dynamic memristor-based RC system on temporal classification tasks, the benchmark test of spoken-digit recognition is carried out using NIST TI-46 database. The input data are audio waveforms of isolated spoken digits (0 to 9 in English) pronounced by five different female speakers. The goal of spoken-digit recognition is to distinguish each digit independent of speakers. Therefore, feature extraction of audio signals is very important. Figure 3a–c illustrate the procedure of feature extraction of digit 9 based on the RC method. According to a standard procedure in speech recognition, the original audio waveform (resampled at 8 kHz) in Fig. 3a (left panel) is firstly filtered into a spectrum with 64 frequency channels per frame by using Lyon's passive ear model.[39] The channel values that represent the amplitude of the corresponding frequency for each frame are then transferred to the time domain with a duration of τ as shown in Fig. 3a (right panel). Figure 3b shows the pre-processed input signal after the mask process. Different from the previous waveform classification task, the mask here is a two-dimensional matrix composed of randomly assigned binary values (-1 and 1). In each interval of duration τ, the spectrum signal is multiplied by a $64 \times M$ mask matrix to generate the input voltage sequence with a time step δ equal to $1/M$ of τ, where M is the mask length. The pre-processed input signal is then applied to the dynamic memristor, and the corresponding current is

firstly converted to a voltage signal through the series resistor R_L and then amplified and collected by the amplifier and ADC. The recorded memristor response is shown in Fig. 3c and the number of sampling points is set to be equal to M per interval τ. The time step is chosen as δ = 120 μs which must be shorter than the relaxation time t_0 (400 μs) of dynamic memristor. The mask and recording processes are repeated N times with different mask matrices in order to mimic N-parallel RC system. After that, the N times memristor response in each duration τ is combined into the reservoir states for subsequent classification.

The classification process contains two steps: training and testing. The 500 audio samples from TI-46 database are divided into two groups: 450 randomly selected samples for training and the rest 50 samples for testing. We use a 10-dimensional vector (target vector) to represent the classification result for the ten digits. For example, if the target digit is 9, the tenth number in the target vector will be 1 while the others should be 0. After feature extraction, the spoken digits are transformed into the reservoir states in each time interval τ. The classification procedure is performed once at each interval and the final classification result is obtained from majority voting of the results at all intervals of one digit.[11,16] In an ideal situation, a correct classification can be given at each interval. We assume a weight matrix (W_{out}) that can transform the reservoir states, which can be treated as an ($M \times N$)-dimensional vector, in each interval τ to the target vector. Therefore, the goal of the training process is to find a proper W_{out} for all the training samples to generate output vectors close to the corresponding target vectors. Here the linear regression method is used to calculate W_{out}. We generate a target matrix Y_{target} by combining the target vectors at all the time intervals used for training. In the same way, we can also generate a response matrix X by combining the response vectors at all of the time intervals used for training. Subsequently, the weight matrix W_{out} is given by $W_{out} = Y_{target}X^T(XX^T)^\dagger$,[40] where the symbol † represents Moore–Penrose pseudo-inverse.

During the testing process, the output vectors at all intervals of one digit are summed up. To obtain the final classification result, the element with the maximum value in the summed output vector predicts the corresponding digit (a winner-take-all method).[35] To evaluate the accuracy, the recognition rate is defined as the percentage of correctly identified digits in all the testing digits. Furthermore, a 10-fold cross validation is used to ensure the reliability of the obtained recognition rate. To do that, the training and testing processes are repeated 10 times and the data are randomly selected for training and testing for each time. The final recognition rate is the average of all the test results during 10-fold cross validation. Figure 3d shows the predicted digits obtained from the memristor-based RC system versus the correct digits, where the color depth is proportional to the number of correctly classified digits. The word error rate is as low as 0.4% (i.e., recognition rate of 99.6%), when M and N are set to be 10 and 40 respectively. In Fig. 3e, the dependence of the word error rate on the mask length is investigated, where the total reservoir size (M × N) remains constant at 400. Similar to the previous waveform classification task, the word error rate increases when the mask length is too long or too short. It can be seen from the experimental data that the lowest average word error rate is achieved when the mask length is about 10. the RC system has also been studied, and the experimental

result is shown in Fig. S5 of Supplementary Information. It is found that the word error rate decreases with the reservoir size, because a larger reservoir can create more reservoir states and hence retain more features of the input signals.

Handwritten-digit Recognition. In addition to the temporal signal processing in the above spoken-digit recognition, we also perform the handwritten-digit recognition to demonstrate the classification of spatial signals. We randomly select 20,000 images from MNIST database as the input data, in which 18,000 images are used for training and the rest 2,000 images for testing. In order to use our memristor-based RC system to realize handwritten-digit recognition, we first convert the spatial pattern of a handwritten digit image into a temporal signal.[41] As an illustration, the preprocessing of digit 9 is shown in Fig. 4a. At the beginning, the original 28 × 28 image is resized into 15 × 15 and it is reproduced in three copies. Then the resized images are rotated by three different angles (0°, 30°, 90°). This step is proved to be necessary, because it can increase the number of features when we transform the image into a temporal signal.[42] Subsequently, the three rotated images are combined and chopped vertically. After that, the obtained 45 × 15 image is divided into 15 sub-images and each of these has a dimension of 45 × 1. Finally, every piece of the chopped image is transferred into a temporal signal with a duration τ and the amplitude corresponds to the pixel grey-scale values. The mask process is similar as the one used in the previous spoken-digit recognition. During each time interval τ, the pre-processed signal is multiplied by a 45 × M mask matrix to generate the input voltage sequence with a time step δ (δ = 120 μs), where M is equal to 4. N-parallel RC system is realized by using different mask matrices. Here N is 300 to generate a total reservoir size of 1,200. The training and testing processes are similar to the spoken-digit recognition task and the only difference is that the feature vectors used for classification are composed of reservoir states selected within three time intervals of τ.[42] Fig. 4b shows the predicted digits versus the correct digits, where an overall recognition rate of 97.6% is achieved by the memristor-based RC system. Furthermore, the comparison of the recognition accuracy between software-based RC and memristor-based RC is shown in Fig. 4c, where the accuracy loss for our RC system is only 0.4% compared with standard software-based RC system. It is therefore demonstrated that the dynamic memristor-based RC system can also process spatial signal efficiently.

Discussion

In summary, a high-performance parallel RC system has been realized using a novel Ti/TiO_x/TaO_y/Pt dynamic memristor. By applying a simple mask process, we show that even a single dynamic memristor can be treated as a reservoir which is subsequently used to build a parallel RC system. By choosing the appropriate mask length and the number of reservoirs, our RC system can process both spatial and temporal signals efficiently. Low NRMSE and word error rate of 0.14 and 0.4% have been achieved for the waveform classification and spoken-digit recognition, respectively, and meanwhile a high recognition rats of 97.6% has been achieved for handwritten-digit recognition with just 0.4% accuracy loss compared to software-based RC system. The parallel RC system in this work is implemented on a single memristor running in serial mode, which is very compact and efficient, proving the feasibility and high efficiency of

memristor-based RC system. To further enable parallel processing of input signals and increase the complexity of the RC system, a more sophisticated RC system based on multiple memristors with inner connections will be constructed in the future.

Methods

Device Fabrication. The dynamic memristor device was fabricated as a cross-point structure on a silicon substrate with 200 nm thermally grown silicon oxide on it. Firstly, inert metal Pt was deposited and patterned on the substrate as the bottom electrode. The thickness and width of the bottom electrode are 50 nm and 10 μm respectively. Then, the functional 30 nm-thick TaO_y and 16 nm-thick TiO_x oxide layers were deposited by the reactive sputtering method with Ar and O_2 mixed atmosphere.[43] Finally, the top electrode Ti was deposited and patterned with the same thickness and width as the bottom electrode.

Measurement Setup. The basic electrical behaviors of the dynamic memristor were characterized at room temperature in a probe station connecting to a semiconductor parameter analyzer (Agilent B1500). The thickness of each layer of the device was verified by transmission electron microscope (TEM). The experimental RC system is realized with the cooperation of personal computer (PC), microcontroller unit (MCU) with peripheral circuits and memristor device. The PC is used to run the basic loop of RC algorithm which is realized by MATLAB code. The MCU used in our experiment is STM32 with 12-bit DAC and ADC modules. The peripheral circuits consist of input amplifiers and output amplifiers. The function of STM32 and amplifier is to connect the PC with the memristor device. Take the spoken-digit recognition task for example. The PC preprocesses the spoken signal into a discrete sequence of real numbers between − 1 and 1. This data sequence is transferred to the buffer of STM32 through UART communication. The DAC module of STM32 then generates voltage pulses with pulse width of 120 μs and amplitude (0 ~ 3.3 V) corresponding to data values. The input amplifier resizes the amplitude of voltage pulse between − 3 to 3 V and applies it to the memristor device. The constant R_L in series with the memristor is used to convert the response current into a voltage signal. The value of 4.7 kΩ is selected according to the resistance range of our memristor and the magnification of the amplifier. The output amplifier transforms the small current signal of memristor into a large voltage signal (0 ~ 3.3 V) which is then sampled by the ADC module. Finally, the ADC data is transferred from STM32 back to the PC for post-processing. The software-based RC is implemented in MATLAB using appropriate parameters.

Declarations

Acknowledgements

This work was supported in part by China key research and development program (2019YFB2205403), Natural Science Foundation of China (61974081, 91964104, 61851404).

References

1. Ren S., He K., Girshick R. & Sun J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **39**, 1137–1149 (2017).
2. Redmon J., Divvala S. K., Girshick R. & Farhadi A., in 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 779–788 (IEEE, 2016).
3. Deng L. *et al.*, in 2013 *IEEE International Conference on Acoustics, Speech and Signal Processing*, 8604–8608 (IEEE, 2013).
4. Chen C., Seff A., Kornhauser A. L. & Xiao J., in 2015 *IEEE International Conference on Computer Vision (ICCV)*, 2722–2730 (IEEE, 2015).
5. Kang M. & Kang J. Intrusion Detection System Using Deep Neural Network for In-Vehicle Network Security. *PLoS One* **11**, (2016).
6. LeCun Y., Bengio Y. & Hinton G. Deep Learning. *Nature* **521**, 436–444 (2015).
7. Hopfield J. J. Neural Networks and Physical Systems with Emergent Collective Computational Abilities. *Proc. Natl Acad. Sci. USA* **79**, 2554–2558 (1982).
8. Hochreiter S. & Schmidhuber J. Long Short-term Memory. *Neural Comput.* **9**, 1735–1780 (1997).
9. Maass W., Natschläger T. & Markram H. Real-Time Computing Without Stable States: A New Framework for Neural Computation Based on Perturbations. *Neural Comput.* **14**, 2531–2560 (2002).
10. Jaeger H. The "echo state" Approach to Analysing and Training Recurrent Neural Networks-with an Erratum Note'. *Bonn, Germany: German National Research Center for Information Technology GMD Technical Report* **148**, (2001).
11. Verstraeten D., Schrauwen B. & Stroobandt D., in *The 2006 IEEE International Joint Conference on Neural Network Proceedings*, 1050–1053 (IEEE, 2006).
12. Jaeger H. & Haas H. Harnessing Nonlinearity: Predicting Chaotic Systems and Saving Energy in Wireless Communication. *Science* **304**, 78–80 (2004).
13. Jaeger H., in *Proceedings of the 15th International Conference on Neural Information Processing Systems*, 609–616 (MIT Press, 2002).
14. Pathak J. *et al.* Model-Free Prediction of Large Spatiotemporally Chaotic Systems from Data: A Reservoir Computing Approach. *Phys. Rev. Lett.* **120**, 024102 (2018).
15. Tanaka G. *et al.* Recent Advances in Physical Reservoir Computing: A review. *Neural Netw.* **115**, 100–123 (2019).
16. Torrejon J. *et al.* Neuromorphic Computing with Nanoscale Spintronic Oscillators. *Nature* **547**, 428–431 (2017).
17. Nakane R., Tanaka G. & Hirose A. Reservoir Computing With Spin Waves Excited in a Garnet Film. *IEEE Access* **6**, 4462–4469 (2018).
18. Martinenghi R. *et al.* Photonic Nonlinear Transient Computing with Multiple-delay Wavelength Dynamics. *Phys. Rev. Lett.* **108**, 244101 (2012).
19. Vandoorne K. *et al.* Experimental Demonstration of Reservoir Computing on a Silicon Photonics Chip.

20. Antonik P. *et al.* Online Training of an Opto-Electronic Reservoir Computer Applied to Real-Time Channel Equalization. *IEEE Trans. Neural Netw.* **28**, 2686–2698 (2017).
21. Du C. *et al.* Reservoir Computing Using Dynamic Memristors for Temporal Information Processing. *Nat. Commun.* **8**, 2204 (2017).
22. Moon J. *et al.* Temporal Data Classification and Forecasting Using a Memristor-based Reservoir Computing System. *Nat. Electron.* **2**, 480–487 (2019).
23. Midya R. *et al.* Reservoir Computing Using Diffusive Memristors. *Adv. Intell. Syst.* **1**, 1900084 (2019).
24. Yao P. *et al.* Fully Hardware-implemented Memristor Convolutional Neural Network. *Nature* **577**, 641–646 (2020).
25. Yao P. *et al.* Face Classification Using Electronic Synapses. *Nat. Commun.* **8**, 15199 (2017).
26. Hu M. *et al.* Memristor-Based Analog Computation and Neural Network Classification with a Dot Product Engine. *Adv. Mater.* **30**, 1705914 (2018).
27. Yang J. J., Strukov D. B. & Stewart D. R. Memristive Devices for Computing. *Nat. Nanotechnol.* **8**, 13–24 (2013).
28. Cai F. *et al.* A Fully Integrated Reprogrammable Memristor–CMOS System for Efficient Multiply–accumulate Operations. *Nat. Electron.* **2**, 290–299 (2019).
29. Tang J. *et al.* Bridging Biological and Artificial Neural Networks with Emerging Neuromorphic Devices: Fundamentals, Progress, and Challenges. *Adv. Mater.* **31**, 1902761 (2019).
30. Wang Z. *et al.* Memristors with Diffusive Dynamics as Synaptic Emulators for Neuromorphic Computing. *Nat. Mater.* **16**, 101–108 (2017).
31. Chang T., Jo S. H. & Lu W. Short-Term Memory to Long-Term Memory Transition in a Nanoscale Memristor. *ACS Nano* **5**, 7669–7676 (2011).
32. Li X. *et al.* Power-Efficient Neural Network with Artificial Dendrites. *Nat. Nanotechnol.* (2020) https://doi.org/10.1038/s41565-020-0722-5.
33. Chua L. Memristor-The missing circuit element. *IEEE Trans. Circuit Theory* **18**, 507–519 (1971).
34. Strukov D. B., Snider G. S., Stewart D. R. & Williams R. S. The Missing Memristor Found. *Nature* **453**, 80–83 (2008).
35. Appeltant L. *et al.* Information Processing Using a Single Dynamical Node as Complex System. *Nat. Commun.* **2**, 468 (2011).
36. Paquot Y. *et al.* Optoelectronic Reservoir Computing. *Sci. Rep.* **2**, 287–287 (2012).
37. Riou M. *et al.*, in *2017 IEEE International Electron Devices Meeting (IEDM)*, 36.33.31–36.33.34 (IEEE, 2017).
38. Rodan A. & Tino P. Minimum Complexity Echo State Network. *IEEE Trans. Neural Netw.* **22**, 131–144 (2011).
39. Lyon R. F., in *ICASSP '82. IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1282–1285 (IEEE, 1982).

40. Lukosevicius M. & Jaeger H. Survey: Reservoir computing approaches to recurrent neural network training. *Comput. Sci. Rev.* **3**, 127–149 (2009).
41. Jalalvand A., Van Wallendael G. & De Walle R. V., in 2015 *7th International Conference on Computational Intelligence, Communication Systems and Networks*, 146–151 (IEEE, 2015).
42. Schaetti N., Salomon M. & Couturier R., in 2016 *IEEE Intl Conference on Computational Science and Engineering (CSE) and IEEE Intl Conference on Embedded and Ubiquitous Computing (EUC) and 15th Intl Symposium on Distributed Computing and Applications for Business Engineering (DCABES)*, 484–491 (IEEE, 2016).
43. Li X. *et al.* Electrode-induced Digital-to-analog Resistive Switching in TaO_x-based RRAM Devices. *Nanotechnology* **27**, 305201 (2016).

Supplementary Figure Legends

Supplementary Figure S1. Investigation of the effect of the time step on the RC system. (a) The input voltage after mask process and the corresponding current response of the dynamic memristor, where the time step δ (50 ms) is much larger than the characteristic time t_0 (400 μs) of the device. In this case, the device rapidly saturates to a state that is independent of previous inputs. (b) As a result, the extracted virtual nodes are independent of each other, and each node is only coupled with itself at previous time step. (c) The input voltage after mask process and the corresponding current response of the dynamic memristor, where the time step δ (50 μs) is smaller than the characteristic time t_0 (400 μs) of the device. In this case, the device does not have enough time to reach a saturated state. (d) As a result, the extracted virtual nodes can be coupled with their neighbours efficiently, and hence a functional RC system can be implemented.

Supplementary Figure S2. (a) - (d) Waveform classification results when the mask length changes from 1 to 40. The first to fourth panels of each figure show the input waveform, the first reservoir state, all reservoir states, and the classification results, respectively. As we can see from the second panel of each figure, the first reservoir state extracted from the dynamic memristor response can transform the difference of waveforms into the change of amplitude, however this effect decreased with the increase of mask length. That is because the mask length can affect the overall change of memristor conductance over a duration τ, as we mentioned in the main text. In addition, From the third panel of each figure we can find that more and more reservoir states overlap as the mask length decreases. When the mask length is 1, only two reservoir states can be distinguished, leading to a large error rate. This result further confirms our conclusion in the main text.

Supplementary Figure S3. (a) Measured memristor response from the waveform classification task, repeated by 30 times. (b) The variation of NRMSE during repeated tests.

Supplementary Figure S4. (a) The nonlinear response region used for waveform classification task. Through a load resistor, the output current of memristor is converted into a voltage signal which is then

amplified by a factor of 3.5 times using an amplifier. In this way, the dynamic memristor response is mapped to a voltage range of 0 to 3.3 V for ADC sampling. (b) The nonlinear response region used for spoken-digit recognition and handwritten-digit recognition tasks. Here the memristor response is amplified by a factor of 371 while adding a voltage bias of 1 V. Because the values of input voltage mostly vary from − 2 to 2V when performing these two tasks, we choose the appropriate amplifier gain and bias so that both positive and negative output responses can be well sampled by the ADC.

Supplementary Figure S5. The tested word error rate changes with the reservoir size. In each test, the mask length remains constant at 10, and the total reservoir size is adjusted by changing the number of parallel reservoirs *N*. The error bar shows the variation between devices.

Figures

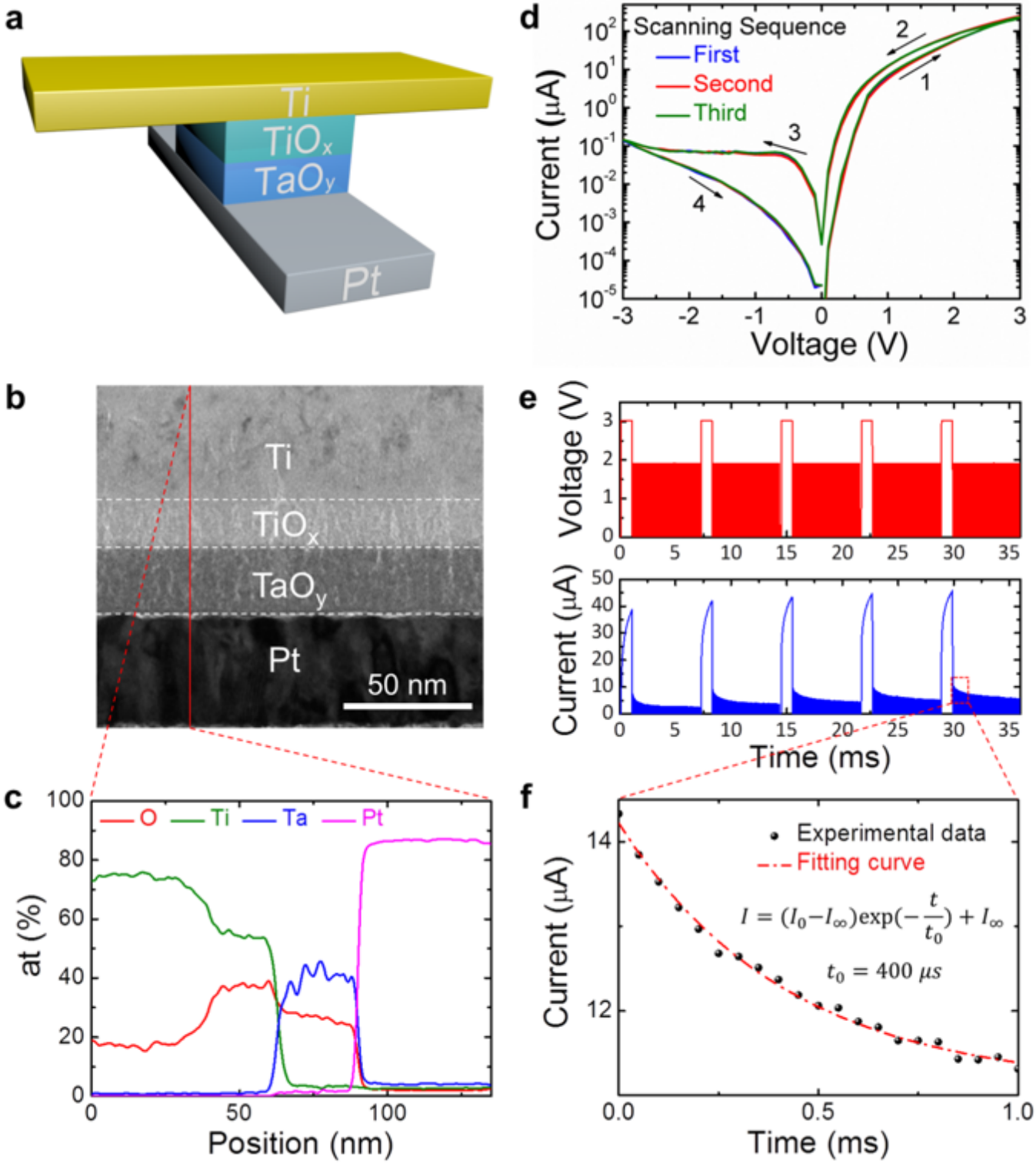

Figure 1

Device characteristics of dynamic memristor. (a) Device structure and (b) cross-sectional transmission electron microscope (TEM) image of the fabricated dynamic memristor, consisting of a vertically stacked structure of Ti/TiOx/TaOy/Pt (50 nm/16 nm/30 nm/50 nm). (c) Corresponding elements distribution profile from energy-dispersive spectroscopy. (d) Device I–V hysteresis curves. Three scans were repeated, [illegible] voltage scan. (e) The experiment exploring the dynamic

characteristics of device. Here the input sequence is a periodic signal composed of a write voltage pulse (3.0 V, 1 ms) followed by several read voltage pulses (1.9 V, 10 μs) in one period. The responding current is recorded for subsequent analysis. (f) The current decay with time follows a simple exponential relationship and the characteristic time t0 obtained by fitting is 400 μs.

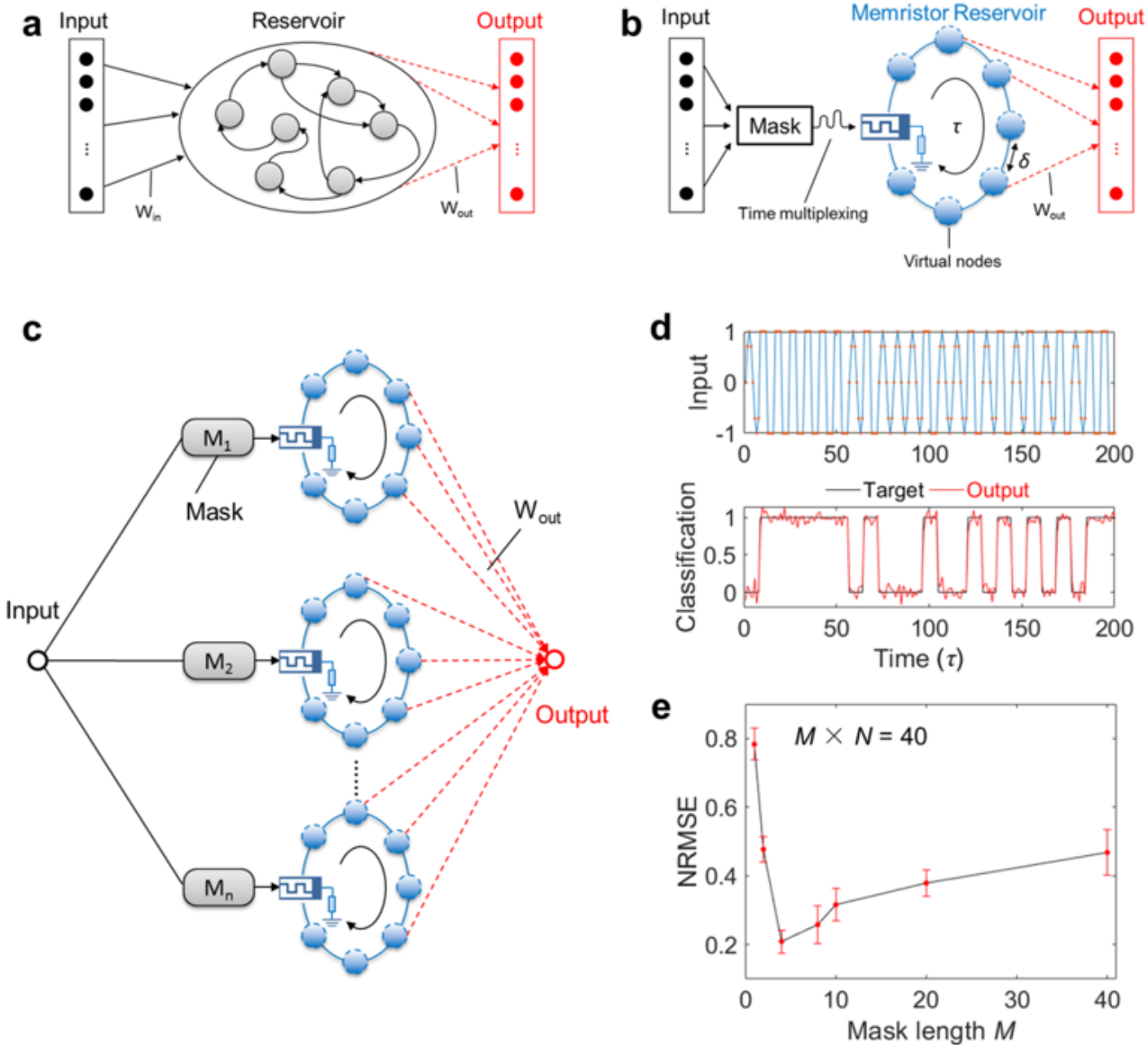

Figure 2

RC system architecture and waveform classification demonstration. (a) Schematic of a conventional RC system. The input is fed to a reservoir which is composed by a large number of nonlinear nodes. The internal connections among these nodes are random and fixed. The correct output learns from the states of nodes by training the output weights. (b) Schematic of the dynamic memristor-based RC system. For a given input, the input vector is transformed into a temporal signal through a mask (that is the time

multiplexing process), and then fed to the reservoir which consists of a dynamic memristor and a load resistor in series. The memristor responses within a duration time τ are selected as the virtual nodes with a fixed time step δ. The output vector is a linear combination of the values in the virtual nodes and the weights (Wout) can be trained through linear regression. (c) Schematic of a dynamic memristor-based parallel RC system, where the mask sequences are different for every single memristor RC unit. The output is the linear combination of all reservoir states. In our experiment, this parallel RC system is realized by testing single memristor in multiple cycles. (d) The input and classification result of sine and square waves. The input sequence is a random combination of sine and square waveforms, where the sampling points for each waveform are set to 8. The optimal classification results are achieved when the length of mask sequence and the number of reservoirs in parallel are set to 4 and 10 respectively and the lowest NRMSE we get is 0.14. (e) NRMSE changes with the mask length when keeping the reservoir size (that is the product of mask length M and number of reservoirs N) the same. Ten different devices are tested and the average of NRMSE reaches the minimum value as the mask length reaches 4. The error bar shows the variation between devices.

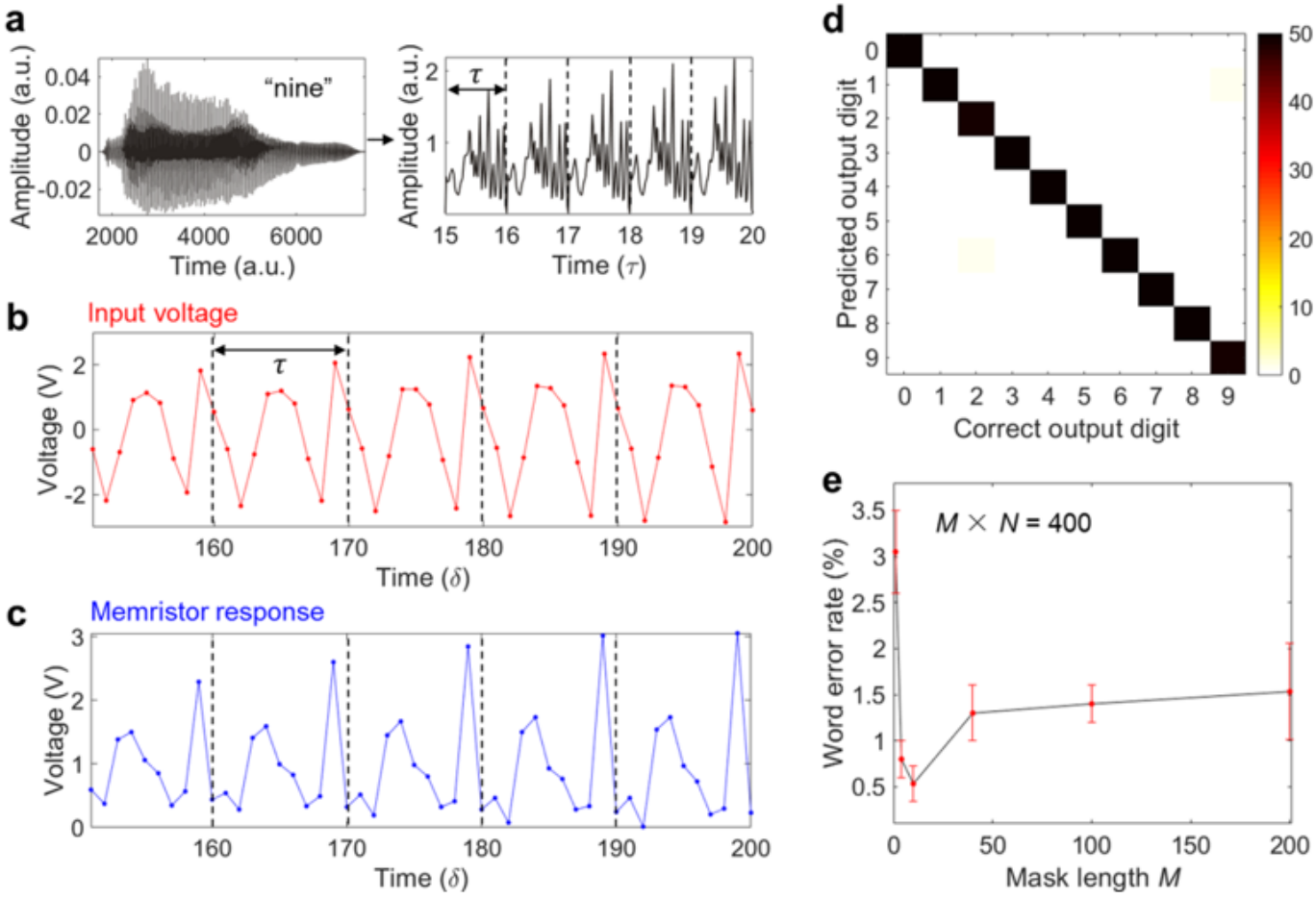

Figure 3

Spoken-digit recognition demonstration. (a) Left: typical audio waveform of digit 9 pronounced by a female speaker. Right: cochlear spectrum (64 channels per frame) of the corresponding audio waveform.

The channel values for each frame is transferred to the time domain with a duration of τ. (b) Time multiplexing process. In each interval of duration τ, the spectrum signal is multiplied by a mask matrix (64×M) containing randomly assigned binary values (-1 and 1) to generate the input voltage sequence with a fixed time step δ (δ = 120 μs) equal to 1/M of τ, where M is the mask length. The similar process repeats by N times with different mask matrices in order to mimic N-parallel RC system. (c) During each time duration, the dynamic memristor response is recorded. The device current is first converted into voltage through the load resistor, and then amplified and collected by the amplifier and ADC. After that, the N times memristor response in each duration of τ is combined into the reservoir states for subsequent classification. (d) Predicted results obtained from the memristor-based RC system versus the correct outputs, where the word error rate is as low as 0.4%. The two parameters M and N of the RC system are set to be 10 and 40 respectively. Colour bar represents the occurrence of each predicted result under the correct output. (e) Word error rate as a function of the mask length M, where the total reservoir size (M×N) remains constant at 400. Similar to the waveform classification task, the average of word error rate reaches the lowest value when M = 10. The error bar represents the variation between devices.

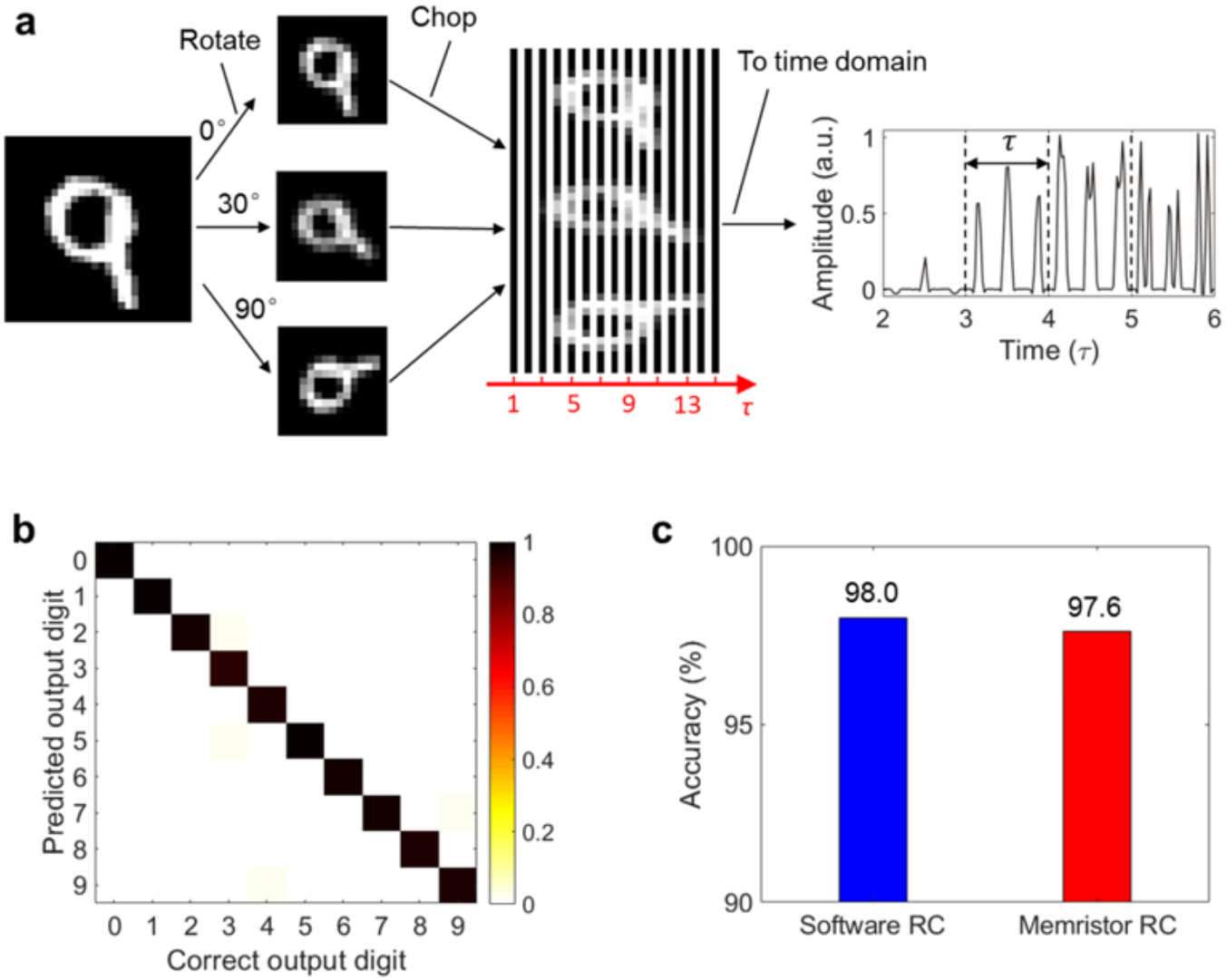

Figure 4

Handwritten-digit recognition demonstration. (a) Schematic of converting the spatial pattern of a handwritten digit image to a temporal signal. Firstly, the original digit 9 is rotated by three different angles (0°, 30°, 90°). Then, the three images are combined and chopped vertically. Finally, every piece of chopped image is transferred into a time-varying signal with a duration τ and the amplitude corresponds to the pixel grey-scale value. (b) Predicted results obtained from the memristor-based RC system versus the correct outputs and the overall recognition accuracy reaches 97.6%. The colour bar represents the normalized probability of each predicted result under the correct output. (c) The comparison of the recognition accuracy between software-based and dynamic memristor-based RC systems.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryInformation.docx