

Nonvolatile Memristor-Based Reservoir Computing System with Reservoir Controllability

Yoonho Jang

Seoul National University

Ji Hun Kim

Seoul National University

Gyung Seok Woo

Seoul National University

Hyun Jae Lee

Seoul National University

Woohyun Kim

Seoul National University <https://orcid.org/0000-0001-7578-4813>

Jeong Woo Jeon

Seoul National University

Cheol Seong Hwang (✉ cheolsh@snu.ac.kr)

Seoul National University <https://orcid.org/0000-0002-6254-9758>

Article

Keywords: reservoir computing (RC), memristor

Posted Date: January 13th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-132081/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Nature Communications on September 30th, 2021. See the published version at <https://doi.org/10.1038/s41467-021-25925-5>.

Nonvolatile Memristor-Based Reservoir Computing System with Reservoir Controllability

Yoonho Jang, Jihun Kim, Gyung Seok Woo, Hyun Jae Lee, Woo Hyun Kim, Jeong Woo Jeon, and Cheol Seong Hwang*

Department of Materials Science and Engineering and Inter-university Semiconductor Research Center, College of Engineering, Seoul National University, Seoul, 08826, Republic of Korea

*Corresponding author (e-mail: cheolsh@snu.ac.kr)

Keywords: *Reservoir computing, Reservoir adaptation, Electronic switching memristor, MNIST recognition, Medical diagnosis*

Abstract

Recent advances in reservoir computing (RC) using memristors have made it possible to perform complicated timing-related recognition tasks using simple hardware. However, the fixed reservoir dynamics in previous studies have severely limited application fields. In this study, RC was implemented with a reservoir that consisted of a W/HfO₂/TiN memristor (M), a capacitor (C), and a resistor (R), in which the reservoir dynamics could be arbitrarily controlled by changing their parameters. After the capability of the RC to identify the static MNIST data set was proven, the system was adopted to recognize the sequential data set [ultrasound (malignancy of lesions) and electrocardiogram (arrhythmia)] that had a significantly different time constant (10^{-7} vs. 1 s). The suggested RC system feasibly performed the tasks by simply varying the C and R, while the M remained unvaried. These functionalities demonstrate the high adaptability of the present RC system compared to the previous ones.

Introduction

The convolutional neural network (CNN), based on the back-propagation learning rule, combines the convolutional layer and the fully connected layer.¹ It shows outstanding performance in static image processing (recognition and classification).^{2,3} However, when the temporal order of each input vector and the correlation between the input vectors are essential, such as for natural language recognition or translation, a method of processing the input over time is required, and CNN is not suitable for this purpose.⁴ Such an event sequence or time-dependent network operation can generally be represented by the relationship between the input (x_i) and the network state (H_i) as follows:

$$H_{i+1} = A(H_i, x_i), \tag{1}$$

where H_{i+1} represents the state of the network at time step $i+1$, and A , H_i , and x_i represent the activation function, the state, and the input, respectively, of the network at time step i . A typical network with such characteristics is a recurrent neural network (RNN) with the long-short-term memory learning rule,⁵ which mitigates the vanishing gradient descent problem of the classical RNN.⁶ Nonetheless, these artificial neural networks perform vast amounts of multiplication and accumulation operations during the learning and inference steps. If these calculations are performed using the conventional von Neumann computing system, even with the latest graphics processing unit, the cost of achieving the required processing speed and the energy consumption will be enormous.⁷

In this regard, the recent upsurge of studies on neural networks that use a memristor-based cross-bar array (CBA) based on Ohm's law and Kirchoff's law is notable.⁸⁻¹³ If the memristor used in such neural networks can process the event-sequence-related and temporal information, it can perform the RNN functionality. An even more desirable functionality is to extract the features of the input information (raw data vector) and feed them to the next classification layer, which is usually a fully connected network (FCN), as for the CNN. Reservoir computing (RC) is capable of these functions.^{14,15}

RC, which consists of a randomly generated RNN (reservoir) and an FCN (readout layer), extracts features at the reservoir and analyzes the result in the readout layer. The core part of the RC system is the reservoir, where the non-linear transformation of the input signal is performed, and the characteristics of the input signal are projected into a high-dimensional feature space, which is called the *reservoir state*.¹⁶ Afterwards, the reservoir state and the target output are matched during the training of the readout layer. To constitute a high-performance RC system, the reservoir devices (or system) must express not only the signal amplitude but also the time information. When a memristor is used as the reservoir, its resistance must be determined by the different input pulse signals with varying amplitudes and the intervals between such input signals. If the input signals to be discerned have simple and obviously distinguishable patterns, a memristor can sufficiently discern them by assigning different resistance values to them. However, for complicated and similar input patterns, high separability is required, which is usually challenging to achieve with a given type of memristor.^{17,18} Also, the input signals could have substantially different time constants, which further severely limits the memristor reservoir.^{19,20} In this case, a high-capacity of the reservoir adaptation must be achieved not only by modifying the memristor but also by incorporating additional circuit components. In a software-based RC system, several studies reported optimized RC systems that showed high performance through reservoir adaptation by modifying several parameters and connections between recurrent nodes.^{16,21}

Recently, various studies were conducted on hardware-based RC systems that use volatile memristors, in which a volatile memristor was used to process a time-varying input.¹⁷⁻²⁰ In those studies, the reservoirs were constructed based on ionic diffusion dynamics (diffusive memristors), in which the diffusive memristor was operated as follows. It was initially reset to the high-resistance state (HRS), and the input pulse later set it to a low-resistance state (LRS).¹⁷ However, such an LRS decays with time through the progressive rupture of the metallic (Ag) conducting filament (CF) with time, which was induced by the thermal energy of the ambient.

Therefore, it can respond not only to the number/strength of the input pulses but also to the timing between the input pulses and the time elapsed before the data read operation.

However, there are several limitations in using such reservoir dynamics. First, the duration and interval of the input signal are limited to the time range in which sufficient conductance decay occurs. For this reason, in the previous studies, it took 1 to 20 ms for one memristor to process 4-bit data, which is a speed that is significantly insufficient to process a large amount of data.^{17,18} Second, obtaining a reproducible reservoir state could be challenging. An Ag-filament-based diffusive memristor has a stochastic switching nature,¹⁷ so the variation of the reservoir state will be significantly large. Finally, reservoir adaptation could be difficult to achieve, given that the reservoir dynamics are totally determined by the material property, which renders the previous system useful for applications with a time scale similar to that of the specific memristor.¹⁸⁻²⁰

In this study, a device based on an electron trap/detrapping mechanism was used to solve the aforementioned issues.^{22,23} A W/HfO₂/TiN (WHT) memristor goes into an LRS when the trap is filled with electrons and shifts to a high-resistance state (HRS) when the trapped electrons are detrapped. Since the resistance switching is based on the electron trapping and not the ionic movement, reproducible results can be achieved.^{24,25} In addition, since the work functions between the top and bottom electrodes differ only slightly, there is limited built-in potential, so the device has good retention properties (Supplementary Fig. S1a).^{23,26} Nonetheless, the WHT memristor also has its own time constant of operation, so it is unlikely to achieve reservoir adaptability with a sufficiently large time constant range. This problem could be solved by combining the memristor with a capacitor (C) and a normal resistor (R). Under this circumstance, the R-C time constant of the circuit can be varied, and the memristor response to the temporal arrangement of the inputs can be controlled.

Figure 1a shows the RC system that can control the reservoir dynamics using a memristor, a normal resistor, and a capacitor (1M1R1C). In this RC system, the charging and discharging

of the capacitor transforms the signals applied to the device into various forms so that the conductance state of the memristor can be varied depending on the magnitude and sequential arrangement of the input signal (Supplementary Fig. S1b, c). Such a configuration of the RC system allows the arbitrary variation of the response dynamics by adjusting the sizes of the resistor, capacitor, and pulse width, etc. Therefore, the optimized RC system can be configured for tasks with vastly different time scales.

Device Analysis

Figure 1b shows the measured current-voltage (I-V) curve of the WHT device. During the electrical measurement, the W top electrode (TE) was biased, while the TiN bottom electrode (BE) was electrically grounded. The resistance of the device was changed from HRS to LRS by a positive bias (SET), and reverse switching was achieved by a negative bias (RESET). In both SET and RESET, gradual switching appeared, as shown in Fig. 1b and Supplementary Fig. S1b, c, which contributed to the high performance of the RC system. Supplementary Fig. S1d shows the cross-sectional scanning transmission electron microscopy (STEM) image of the WHT device, which revealed the W TE, the TiN BE, and the 4 nm-thick HfO₂ layer between the TE and BE. Supplementary Fig. S1e shows the X-ray photoelectron spectroscopy (XPS) analysis of the W/HfO₂ interface in the WHT device. Analysis of the W peak in the XPS data revealed the presence of tungsten sub-oxide (WO_x, x < 3) and a WO₃ layer. The energy-dispersive X-ray spectroscopy line scan result (Supplementary Fig. S1d, right portion) along the vertical line from TE to BE in the STEM image implies that a thin WO₃ was formed at the W/HfO₂ interface and WO_x (x << 3) was formed within the W bulk. Therefore, the WO_x may work as a voltage divider when the voltage is applied to the device, which will cause gradual SET and RESET performance.²⁷ This is a favorable characteristic, allowing the reservoir to have various states. Moreover, this WHT device does not have an electroforming step (Fig. 1b), which also contributed to the stable resistance switching operation

(Supplementary Figs. S2-4 and Supplementary Note 1). W and TiN have similar work functions of ~ 4.5 eV, which may render the energy band profile symmetric.^{28,29} The symmetric energy band profile is unfavorable for fluent electronic bipolar resistive switching (eBRS).^{23,26} However, the WO_3 layer formed at the W/HfO₂ interface can induce a Schottky barrier, whereas the HfO₂/TiN interface constitutes a quasi ohmic contact.^{27,30} Especially, the chemical interaction between the HfO₂ and TiN layers can produce defects within the HfO₂ layer, which provide the system with the electron traps that are necessary to induce the eBRS mechanism. With the application of the positive bias to the TE, the traps were filled with electrons that were injected from the TiN BE through the quasi ohmic contact, which switched the device to the LRS. Conversely, when the negative bias was applied, the device switched back to the HRS as the trapped electrons were detrapped, while the electron injection from the TE was blocked by the Schottky barrier at the W/HfO₂ interface.²⁶ Due to the presence of the WO_x layer, there was no need to set current compliance (CC) during the operation.

Reservoir Generation

The physical reservoir must satisfy the following four requirements: 1) high dimensionality, 2) non-linear transformation, 3) separability, and 4) dynamic adaptation. In this study, the reservoir was implemented by configuring the circuit, shown in Fig. 2a. Pulse streams were generated by a pulse generator (PG), where input signal '1' is converted to a high level, and '0' is converted to a low level. These pulse streams were delivered to channel 1 (CH1) and channel 2 (CH2) of an oscilloscope (OSC). A 50Ω resistor was assigned to CH1, which allowed monitoring of the input pulse shape. In CH2, a $1 \text{ M}\Omega$ resistor was connected to the device-under-test (DUT, the memristor) in series. From the estimated voltage from the CH2 resistor, the voltage applied to the DUT was inferred. Since the oscilloscope fixes the size of the CH2 resistor at $1 \text{ M}\Omega$, the overall series resistance to the memristor was adjusted by connecting a load resistor (R_L), as shown in the figure. Also, a capacitor was connected to the CH2 resistor

in parallel, which stored the charge supplied by the applied pulse voltage. In this specific experimental setup, its value was fixed at 180 pF. However, the dynamic time constant of the RC system was varied by changing R_L and the capacitance when it was necessary. The measurement consisted of two steps. In the first step, a pulse was generated at the PG, which caused SET switching in the memristor. In the second step, the conductance state of the memristor was read through the DC sweep.

Figure 2b shows the voltages transients over the memristor with a ‘0101+reference pulse’ (left) and a ‘1010+reference pulse’ (right), and Fig. 2c shows the corresponding voltage transients read at CH2. In these operations, 4 V, 200 μ s, and 0 V, 200 μ s pulses were programmed to represent ‘1’ and ‘0’, respectively. The role of the last reference pulse is explained as follows. The left panels of Figs. 2b and c show that since the first signal was ‘0’, no voltage appeared up to 0.2 ms. When the first ‘1’ signal was applied, the DUT showed a peak of up to \sim 3.5 V due to the involvement of the capacitive charging current, and it decayed to \sim 1.5 V after the capacitor charging was completed. At the same time, the CH2 voltage showed a corresponding gradual increase in the capacitor voltage, which was saturated at \sim 2.5 V. When the second ‘0’ signal came in, the capacitor was discharged, and the reverse current flowed into the DUT, which made its voltage negative, while the CH2 showed gradual decay of the capacitor voltage. It was noted from the CH2 voltage that the capacitor was not completely discharged during the 0.2 ms duration of ‘0’ signal, so when the subsequent ‘1’ signal came in, the capacitive charging current was not as high as in the previous ‘1’ signal case (where the DUT voltage peaked only up to \sim 2.5 V). Such an effect can be more evidently seen with the subsequent ‘1’ signal (the reference pulse), as there was almost no peak in the DUT. Therefore, in this case, the effective number of SET pulses applied to the DUT was only two (the first and second ‘1’ among the total three ‘1’s in the ‘01011’ sequence). After the entire pulse sequence was over, the memristor resistance was 2.82 M Ω .

In the case of the right panels in Figs. 2b and c, in contrast, each of the 1 signals is separated by 0 signals, and all the three '1's in the '10101' sequence switched the DUT to the SET state, which made its resistance $2.67\text{ M}\Omega$, despite the application of the same number of set pulses (three) in the two cases. It should be noted, however, that the last two peaks had a lower effect in decreasing the memristor resistance than the first one due to its lower peak height, which was induced by the incomplete discharging of the capacitor during the intervening '0' pulse cycle. This is not a demerit but actually a merit of this RC system, which allowed even higher separability and adaptability. Therefore, this RC system can recognize not only the different input pulse numbers but also their timing. Figures. 2b and c show several notable features. First, if the memristor has the same resistance for both the positive and negative current flows, the charging and discharging times of the capacitor, which can be estimated from CH2, should be identical. However, due to the built-in asymmetry of the band profile of the WHT memristor, the resistance at the positive bias of $\sim 2.5\text{ V}$ was ~ 100 times lower than that at the negative bias of $\sim 1.5\text{ V}$. Therefore, the charging was much faster than the discharging. This is the first factor that allows the RC system to have higher separability and adaptability. Second, the capacitance and R_L values can be arbitrarily taken to vary the charging and discharging times, which can eventually affect the effectiveness of the voltage pulse applied to the memristor. Third, the input voltage pulse height and duration are another knob that can further change the RC dynamics. These features rendered the RC system extremely flexible and adaptable to the various requirements, as shown in the next sections. Without the last reference pulse, such a systematic variation and examination of the memristor state control would have been improbable.

Modifying the Reservoir Dynamics

In this RC system with the given WHT memristor property and capacitance, R_L and the pulse height/duration were varied to examine the separability of the memristor. The capacitance

could also be varied, but it was fixed in this experiment section. Figure 3 shows several examples of the different degrees of separability of the RC system when these parameters were varied. The examples show the current value read at 0.5 V after the 16 different input patterns, from ‘0000’ to ‘1111’, were programmed to the PG, with the additional reference pulse added last. The x-axis numbers correspond to the different input patterns described in the inset table in Fig. 3e, and the different parameters, such as R_L , the input pulse, and the reference pulse, for each graph in Fig. 3 are summarized in Table I. It should be noted that in Fig. 3, the y-axis scales of each graph were varied to easily compare them. All the detailed pulse responses and analyses are included in Supplementary Figs. S5-9 and Supplementary Note 2. In Fig. 3a, wherein $R_L = 1 \text{ M}\Omega$, the signal pulse = 4 V, 100 μs , and the reference pulse = 4 V, 100 μs , the five patterns, ‘0000’, ‘0001’, ‘0011’, ‘0111’, and ‘1111’ are not clearly distinguished (an analysis of the separation of these inputs is shown in Supplementary Fig. S10). It was also noted that the ‘1000’ pattern resulted in the highest memristor conductance, although there were only two SET pulses (the first 1 and the reference pulse at the last SET pulse). This is because the reference pulse induced the highest peak voltage to the memristor because the interval between the two pulses, during which the capacitor was fully discharged, was the longest (the details are shown in Supplementary Fig. S11).

Of the six graphs in Fig. 3, Fig. 3c shows well the critical features of this RC system. The only difference of Fig. 3c from Fig. 3a is the pulse length [100 μs (in a) vs. 200 μs (in c)]. As the pulse width increases, the capacitor discharging during the 0 input increased, and the subsequent ‘1’ induced a higher peak voltage. The conductance levels in Fig. 3c can be clearly grouped into three levels, which are determined by the number of 1’s immediately after the ‘0’ (not the total number of ‘1’). For example, ‘0000’ has only one 1 after 0 (the reference pulse), so it induced the lowest conductance. Interestingly, ‘1111’ has the same low conductance even though it had five 1 inputs (including the reference pulse). This is because the only effective ‘1’

was the first one because all the other '1's do not have the preceding '0's, so they cannot produce peak voltage.

Another characteristic and most desirable setting could be seen in Fig. 3f, in which R_L was decreased to 10 k Ω , and the pulse width was decreased to 200 ns. This setting makes the capacitor charging per one voltage pulse ('1' signal) insufficient, and its discharging during the '0' signals faster. Overall, this makes the memristor conductance more linearly dependent on the total number of '1's, as shown in Fig. 3f (an example of insufficient charging and details of the effects are included in Supplementary Fig. S12). A short pulse length is also beneficial to rapidly process the input vectors.

Task Optimization: MNIST

To recognize the digit images in the Modified National Institute of Standards and Technology Database (MNIST), the reservoir dynamics were optimized to implement an RC system suitable for the task. To do this, the raw MNIST data set, composed of 784 pixels (28 x 28), had to be reconfigured to meet the requirement of this specific RC system, which is basically a binary system (0 and 1 inputs). According to Fig. 3f, 16 levels of the memristor conductance were readily discerned, so 4 bits were dealt with by one memristor. Therefore, the data in the 784 pixel images were binarized and chopped by 4 bits, which resulted in 196 4-bit input signals. To make the task analysis more efficient, the frequency of the appearance of inputs in the dataset was investigated, and it was confirmed that '0000' appeared most frequently, followed by '1111,' '1000,' '0011,' and '0001' (Supplementary Table 1). Therefore, in this task-optimized RC system, the task was performed effectively by setting the operation parameters so that the RC system could readily separate the responses to the inputs with a high frequency of appearance rather than separating the responses to all the 16 inputs. The data points indicated by the red circle in Fig. 3f correspond to these frequently appearing signal sets. Accordingly, the 196 4-bit input image data were converted to the 196 x 1 memristor

conductance vector (MCV), where the measurements were performed on a single 1M1R1C circuit, based on Fig. 3f. Using the 50,000 training images in the MNIST data set, 50,000 training MCVs were generated. These MCVs were used to train the 196 x 10 FCN (weights and biases), which were generated in a PyTorch simulation (Methods section). The trained RC system was used to infer the 10,000 MNIST test images, and the achieved accuracy was 90.1% (See Table II).

This reservoir took 1 μ s of time to process one pulse stream, which is $10^3 \sim 10^4$ times faster than in the previous studies.^{18,17} Table II also shows the comparison with other RC results using the diffusive memristors and the software-based single-layer FCN. This study focuses on the only memristive RC system that performs reservoir adaptation, and it showed the best performance in terms of accuracy and latency. In addition, in this study, higher accuracy was achieved using a 1/4-size FCN compared to a software-based FCN.¹ The hardware size of the RC system could be further decreased as the number of bits processed by the reservoir (nBPR) increases, for as long as the separability for the higher nBPR is guaranteed. Supplementary Fig. S13 shows the different read currents for the 3 to 6 bits (8 to 64 input patterns). Obviously, the separability decayed as the nBPR increased, but they were still be used to recognize the MNIST data set because not all the input patterns mattered equally. Table III shows the variation in the test accuracy of the MNIST data set using the same method as above, but with different nBPRs. As the nBPR increased from 3 to 6, which was accompanied by a decrease in the required memristor number from 252 to 112, the accuracy decreased from 90.7% to 86.3% (the confusion matrices are included in Supplementary Figure S14), which is not much lower than in the software-based FCN (784 x 10). The next section demonstrates the most crucial merit of this RC system by showing its capacity to process time-series data using medical diagnostic data.

Task Optimization: Medical Diagnosis

Medical diagnosis often requires analyzing time-varying data and making a quick diagnosis, but there are inevitable limitations such as high dependence on operators and high variability across different medical institutions. For a more accurate and objective medical diagnosis, a universal diagnosis system adaptable to various situations is essential. Automatic medical diagnosis using deep learning has considerable potential, and several studies have been conducted on it,³¹⁻³³ but most of them rely on the conventional image classification method, such as CNN. This means that the traditional medical diagnosis produces data images and analyzes them later, mostly *ex-situ*. This study suggests, however, a method for *in-situ* medical diagnosis in real-time using a 1M1R1C reservoir. The diagnostic application consists of two sections. The first section is a breast cancer diagnosis using ultrasound images, and the second section is an arrhythmia diagnosis based on electrocardiogram (ECG) results. These two applications have vastly different operating signal frequencies (MHz to Hz). In this study, a system for efficient medical diagnosis was implemented by optimizing the RC system for each task.

1) Diagnosis of malignancy in breast lesions

Ultrasound is used to diagnose and monitor breast cancer. In contrast to the conventional CNN, where the preprocessed images are identified, the proposed RC system in this study directly uses ultrasonic raw data without an imaging process, as shown in Fig. 4a. In the conventional ultrasound diagnosis, the ultrasound is transmitted to the piezoelectric material, where electrical signals are generated. The signal processor processes these signals to generate an ultrasound image, which the operator analyzes to diagnose the disease. However, if the reservoir can directly process the ultrasonic signal, the imaging process can be skipped, and an automatic diagnosis will be made at the readout layer. Therefore, this system makes real-time diagnosis simpler than in the existing ultrasound diagnosis.

The dataset used in the experiment consisted of an open-access database of raw ultrasound signals acquired from malignant and benign breast lesions.³⁴ Each sample consisted of 510 ultrasound (10 MHz) echo lines. After they were preprocessed for measurement convenience, they were converted into pulse streams and applied to the memristor (Methods section). Figure 4b shows the results of the voltage-time (V-t) measurement for one echo line of a benign sample (inset in Fig. 4b, and Methods section). The RC system performed learning/inference based on the training/test set that consisted of 14/6 benign cases and 14/6 malignant cases, and the results showed 100% accuracy. However, such 100% accuracy might originate from the insufficient number of available images and a small number of classes, so a larger dataset and subdivided classes are required for more accurate analysis, which is currently unavailable.

This method has two main advantages over the existing ultrasound diagnosis using CNN. First, diagnosis is performed using a much simpler system without a pre-imaging process. Second, one of the major difficulties in ultrasound analysis is the presence of artifacts.³¹ CNN may have difficulty in recognizing such artifacts because it performs learning and inference with the information on the artifacts. On the other hand, when using 1M1R1C, even with additional stimulation by artifacts, the capacitor only maintains the charging state. Therefore, the reservoir state is determined by the overall contour rather than by fine artifacts, and it can show higher performance.

2) Real-time arrhythmia diagnosis

Arrhythmia is a condition in which the heart has an irregular rhythm or an abnormal heart rate. Since malignant arrhythmia can cause sudden death due to a heart attack,³⁵ real-time ECG monitoring and diagnosis are required. The purpose of this experiment is to implement a system capable of real-time diagnosis of arrhythmia in response to an electrical signal caused by a heartbeat. For the experiment, a part of the MIT-BIH arrhythmia database³⁶ was used, and a task-optimized reservoir was utilized to distinguish between arrhythmia and normal cases. A

reservoir capable of responding to a signal with a frequency of 0.8 to 1.2 Hz was constructed using a 1 μ F capacitor parallel to CH2. In this case, a simple RC system composed of only one 1M1R1C reservoir could be used. Figure 4c shows a part of the ECG of a patient with arrhythmia. The electrical signal is generated at approximately 0.8-s intervals, and then arrhythmia occurs at 1.6 s (marked by a red arrow). When an electrical signal from a heartbeat is applied to the reservoir, the capacitor maintains a high charging level at a normal beat; but when an arrhythmia occurs, the capacitor is discharged at a longer interval than in the normal case, and SET switching occurs in the memristor by the next pulse (Supplementary Fig. S15). Since this reservoir responds only to arrhythmia, the reservoir state can reflect the pulse of the arrhythmia patient in real-time. Figure 4d shows the results of 5-minute reservoir monitoring based on ECG data of normal (case 1) and arrhythmic (cases 2 and 3) patients. In cases 2 and 3, 49, and 81 arrhythmias occurred, respectively. As a result, the conductance of the reservoir monitoring in case 3 was the highest, and the reservoir state was clearly distinguished according to the degree of arrhythmia. This single reservoir system was able to detect different arrhythmia conditions in real-time with low energy using a simple 1M1R1C circuit.

Conclusion

In this study, an RC system with high reservoir separability and dynamics controllability was demonstrated using a nonvolatile W/HfO₂/TiN memristor. A reservoir was generated by composing a 1M1R1C circuit. From asymmetric charging/discharging of the capacitor caused by the memristor, separability, which is the basic property of the reservoir, was achieved. In addition, the manner in which the reservoir reacted to the input signal was modified by changing various parameters such as the load resistor, capacitance, pulse width, and pulse height. Using these characteristics, the RC system was optimized to perform static data-based MNIST recognition application and sequential data-based medical diagnosis (ultrasound diagnosis and ECG-based diagnosis). For the MNIST recognition, a task-optimized system was used to

improve the separability of the inputs that frequently appeared in the dataset. Furthermore, the tradeoff between the reduction of the readout layer size and the performance was confirmed by increasing the nBPR. The RC-system-aided medical diagnosis was conducted for two situations with extremely contrasting input frequencies (1 Hz and 10 MHz). By implementing a reservoir configuration suitable for each task, the excellent performance was achieved without misdiagnosis. This is because this RC system has a circuit structure that can freely change the time constant according to various tasks. In particular, the most crucial point of this study is its demonstration that dynamic signals with vastly different time constants can be well distinguished by changing the resistor or capacitor added to the circuit using only one type of memristor.

The two types of hardware needed to implement the RC system proposed in this study are shown in Supplementary Fig. S16. Figure S16a shows that R_L can be changed by selecting different transistors at the end of the resistor metal line. In contrast, Fig. S16b shows that R_L can be varied by changing the normal memristor resistance. Therefore, it is expected that the fabrication of the hardware for the array configuration will be simple and that the reservoir dynamics can easily be changed even in the fabricated hardware.

Methods

Memristor fabrication. The array of cross-bar-type W/HfO₂/TiN memristors was fabricated. A 50 nm-thick TiN layer was sputtered (Endura, Applied Materials) on an SiO₂/Si substrate, and the TiN layer was patterned into a line shape to form a BE. The 2- to 10 μm-wide TiN BEs were patterned using conventional photolithography and the dry-etching system. After the patterning, the residual photoresist was removed with acetone and cleaned sequentially with deionized water. Then 4 nm HfO₂ was deposited using atomic layer deposition (ALD) at a 280 °C substrate temperature using a traveling-wave-type ALD reactor (CN-1 Co. Plus 200). A tetrakis-ethylmethyamido hafnium (TEMA-Hf) and O₃ were used as precursors for Hf and

oxygen, respectively. On the HfO₂ layer, 50-nm-thick W TEs were sputtered using the MHS-1500 sputtering system and patterned into 2- to 10 μm-wide lines using the conventional lift-off process. After the fabrication, the WHT device was analyzed using x-ray photoelectron spectroscopy (XPS, AXIS SUPRA, Kratos) and energy-dispersive x-ray spectroscopy (EDS, JEOL, JEM-ARM200F) to observe the formation of the tungsten oxide layer. Cross-sectional transmission electron microscope (TEM) images of the WHT memristor were observed using scanning transmission electron microscopy (STEM, JEOL, JEM-ARM200F).

Modified National Institute of Standards and Technology database. The data set, the Modified National Institute of Standards and Technology (MNIST) database,³⁷ is a large database of handwritten digit images. It is commonly used for training and testing of image processing systems such as artificial neural networks. The database was created by "remixing" the digit samples from NIST's original datasets.³⁸ This database consists of 60,000 training samples and 10,000 test samples.

Experimental set-up for the 1M1R1C reservoir computing. To compose the reservoir circuit, the WHT device with an area of 10 μm × 10 μm was connected to the pulse generator (PG, Agilent 81110A) and an oscilloscope (OSC). A 1M1R1C circuit was constructed by adding a load resistor to the circuit and setting the resistance values of CH1 and CH2 in the OSC to 50 Ω and 1 MΩ, respectively. A semiconductor parameter analyzer (SPA, Hewlett-Packard 4145B) was connected to the WHT device to perform the DC sweeps. To process the static and sequential data, the device states after the pulse streams were measured. After the measurement, the device was reset to the HRS state, and the process was repeated. The reservoir state was constructed based on the recorded device states, and the readout layer was trained based on it.

PyTorch simulation for the readout layer of the RC system. The logistic regression algorithm was used to train the readout layer for the MNIST recognition and breast lesion classification. The reservoir state (\mathbf{x}) in the form of an $n \times 1$ vector ($n = 196$ for the MNIST recognition and $n = 510$ for the breast lesion classification) was multiplied by the weight matrix (\mathbf{W}) of the readout layer to yield the weighted sum (\mathbf{z}).

$$\mathbf{z} = \mathbf{W}^T \cdot \mathbf{x} \quad (2)$$

The weighted sum was applied to the following softmax function to yield an output ($\hat{\mathbf{y}}$).

$$\hat{y}_j = \sigma(\mathbf{z})_j = \frac{e^{z_j}}{\sum_{k=1}^n e^{z_k}} \text{ for } j = 1, \dots, n. \quad (3)$$

The sum of the elements of the output vector became 1 and the output of the softmax function was perceived as a 'probability.' The cross-entropy loss was used for the loss function, which is defined as:

$$\text{loss} = -\frac{1}{N} \sum_{i=1}^N [\mathbf{y}_i \log(\hat{\mathbf{y}}_i) + (1 - \mathbf{y}_i) \log(1 - \hat{\mathbf{y}}_i)], \quad (4)$$

wherein N is the number of samples, and \mathbf{y}_i is the target output for input \mathbf{x}_i . To minimize the loss, a gradient-descent-based Adam optimizer³⁹ was used. Full-batch-type learning of the readout layer was performed in PyTorch.

RC system for medical diagnosis. 1) Ultrasound-based breast lesions diagnosis: Each sample in the database consisted of 510 ultrasound (10 MHz) echo lines, and the length of each echo line was different for each sample (100 ~ 300 μs). For measurement convenience, samples in which breast lesions appeared within 40 μs were used for learning and inference. The raw ultrasound data were binarized and converted into 100 ns pulses, which corresponded to a 10MHz frequency. Pulse streams that consisted of 400 100 ns-long pulses were applied to the memristor. The measurement setup was set at a 3.5 V pulse height, 4 V reference pulse height, and a 30 k Ω load resistance. The use of relatively large load resistance and a short pulse length made the reservoir sensitive to consecutive pulses, and the effect of the signal pulses remained

until the reference pulse (Fig. 4b). After the measurement, the 510×2 readout layer was trained based on the reservoir state that consisted of the 510 reservoir responses for each input. 2) ECG-based arrhythmia identification: The measurement was performed under the conditions of no R_L , a 2.5 V pulse height, a 200 ms length, and a 1 μ F capacitor.

Data availability. All the relevant data are available from the authors.

References

1. LeCun, Y., Bottou, L., Bengio, Y. & Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **86**, 2278–2323 (1998).
2. Krizhevsky, A., Sutskever, I. & Hinton, G. E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* (2017) doi:10.1145/3065386.
3. Dong, C., Loy, C. C., He, K. & Tang, X. Image Super-Resolution Using Deep Convolutional Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* (2016) doi:10.1109/TPAMI.2015.2439281.
4. Wang, Z., Yan, W. & Oates, T. Time series classification from scratch with deep neural networks: A strong baseline. in *Proceedings of the International Joint Conference on Neural Networks* (2017). doi:10.1109/IJCNN.2017.7966039.
5. Hochreiter, S. Long Short-Term Memory. **1780**, 1735–1780 (1997).
6. Hochreiter, S. The vanishing gradient problem during learning recurrent neural nets and problem solutions. *Int. J. Uncertainty, Fuzziness Knowledge-Based Syst.* **6**, 107–116 (1998).
7. Lecun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).

8. Kim, G. H. *et al.* 32×32 Crossbar Array Resistive Memory Composed of a Stacked Schottky Diode and Unipolar Resistive Memory. *Adv. Funct. Mater.* **23**, 1440–1449 (2013).
9. Jeong, D. S. & Hwang, C. S. Nonvolatile Memory Materials for Neuromorphic Intelligent Machines. *Adv. Mater.* **30**, 1–27 (2018).
10. Kim, K. M. *et al.* Low-Power, Self-Rectifying, and Forming-Free Memristor with an Asymmetric Programming Voltage for a High-Density Crossbar Application. *Nano Lett.* **16**, 6724–6732 (2016).
11. Li, C. *et al.* Analogue signal and image processing with large memristor crossbars. *Nat. Electron.* **1**, 52–59 (2018).
12. Lee, Y. K. *et al.* Matrix mapping on crossbar memory arrays with resistive interconnects and its use in in-memory compression of biosignals. *Micromachines* (2019) doi:10.3390/mi10050306.
13. Kim, Y. *et al.* Novel Selector-Induced Current-Limiting Effect through Asymmetry Control for High-Density One-Selector–One-Resistor Crossbar Arrays. *Adv. Electron. Mater.* **5**, 1–11 (2019).
14. Maass, W., Natschläger, T. & Markram, H. Real-time computing without stable states: A new framework for neural computation based on perturbations. *Neural Comput.* (2002) doi:10.1162/089976602760407955.
15. Jaeger, H. The ‘echo state’ approach to analysing and training recurrent neural networks. *GMD Rep.* (2001).
16. Lukoševičius, M. & Jaeger, H. Reservoir computing approaches to recurrent neural network training. *Comput. Sci. Rev.* **3**, 127–149 (2009).
17. Midya, R. *et al.* Reservoir Computing Using Diffusive Memristors. *Adv. Intell. Syst.* **1**, 1900084 (2019).

18. Du, C. *et al.* Reservoir computing using dynamic memristors for temporal information processing. *Nat. Commun.* **8**, 1–10 (2017).
19. Moon, J. *et al.* Temporal data classification and forecasting using a memristor-based reservoir computing system. *Nat. Electron.* **2**, 480–487 (2019).
20. Zhu, X., Wang, Q. & Lu, W. D. Memristor networks for real-time neural activity analysis. *Nat. Commun.* **11**, (2020).
21. Steil, J. J. Online reservoir adaptation by intrinsic plasticity for backpropagation-decorrelation and echo state learning. *Neural Networks* **20**, 353–364 (2007).
22. Kim, K. M. *et al.* A detailed understanding of the electronic bipolar resistance switching behavior in Pt/TiO₂/Pt structure. *Nanotechnology* (2011) doi:10.1088/0957-4484/22/25/254010.
23. Shao, X. L. *et al.* Electronic resistance switching in the Al/TiO_x/Al structure for forming-free and area-scalable memory. *Nanoscale* **7**, 11063–11074 (2015).
24. Lu, Y. *et al.* An electronic silicon-based memristor with a high switching uniformity. *Nat. Electron.* **2**, 66–74 (2019).
25. Kwon, S. *et al.* Structurally engineered nanoporous Ta₂O_{5-x} selector-less memristor for high uniformity and low power consumption. *ACS Appl. Mater. Interfaces* **9**, 34015–34023 (2017).
26. Kim, Y. *et al.* Nociceptive Memristor. *Adv. Mater.* **30**, 1–7 (2018).
27. Ryu, J. J. *et al.* Fully ‘Erase-free’ Multi-Bit Operation in HfO₂-Based Resistive Switching Device. *ACS Appl. Mater. Interfaces* **11**, 8234–8241 (2019).
28. Ang, S. S. Titanium nitride films with high oxygen concentration. *J. Electron. Mater.* **17**, 95–100 (1988).
29. Müller, E. W. Work function of tungsten single crystal planes measured by the field emission microscope. *J. Appl. Phys.* **26**, 732–737 (1955).

30. Yoon, J. H. *et al.* Highly uniform, electroforming-free, and self-rectifying resistive memory in the Pt/Ta₂O₅/HfO_{2-x}/TiN structure. *Adv. Funct. Mater.* **24**, 5086–5095 (2014).
31. Liu, S. *et al.* Deep Learning in Medical Ultrasound Analysis: A Review. *Engineering* **5**, 261–275 (2019).
32. Cui, R. & Liu, M. RNN-based longitudinal analysis for diagnosis of Alzheimer's disease. *Comput. Med. Imaging Graph.* **73**, 1–10 (2019).
33. Arena, P., Basile, A., Bucolo, M. & Fortuna, L. Image processing for medical diagnosis using CNN. *Nucl. Instruments Methods Phys. Res. Sect. A Accel. Spectrometers, Detect. Assoc. Equip.* **497**, 174–178 (2003).
34. Piotrkowska-Wróblewska, H., Dobruch-Sobczak, K., Byra, M. & Nowicki, A. Open access database of raw ultrasonic signals acquired from malignant and benign breast lesions. *Med. Phys.* **44**, 6105–6109 (2017).
35. To, U. E. *et al.* *Sddca.* **345**, 1473–1482 (2001).
36. Moody, G. B. & Mark, R. G. The impact of the MIT-BIH arrhythmia database. *IEEE Eng. Med. Biol. Mag.* **20**, 45–50 (2001).
37. Lecun Yann, Cortes Corinna & Burges Christopher. THE MNIST DATABASE of Handwritten Digits. *Courant Inst. Math. Sci.* (1998).
38. Grother, P. J. & Hanaoka, K. K. NIST Special Database 19 - Handprinted Forms and Characters Database. *Technical Report on Special Database 19* (2016).
39. Konur, O. Adam Optimizer. *Energy Education Science and Technology Part B: Social and Educational Studies* (2013).

Acknowledgment

This study was supported by the National Research Foundation of Korea (2020R1A3B2079882).

Author Contributions

Y.J. designed and fabricated the devices and performed the material analysis of the device based on various measurements. He also devised the circuit design for the reservoir and analyzed the characteristics of the reservoir through various electrical measurements. J.K. contributed to the implementation of the readout layer in the PyTorch simulation. K.S.W. contributed to device fabrication. H.J.L. and W.K. supported the data analysis. J.W.J supported the electrical measurements. C.S.H. directed the entire study and prepared the manuscript.

Competing Interests

The authors declare no conflict of interest.

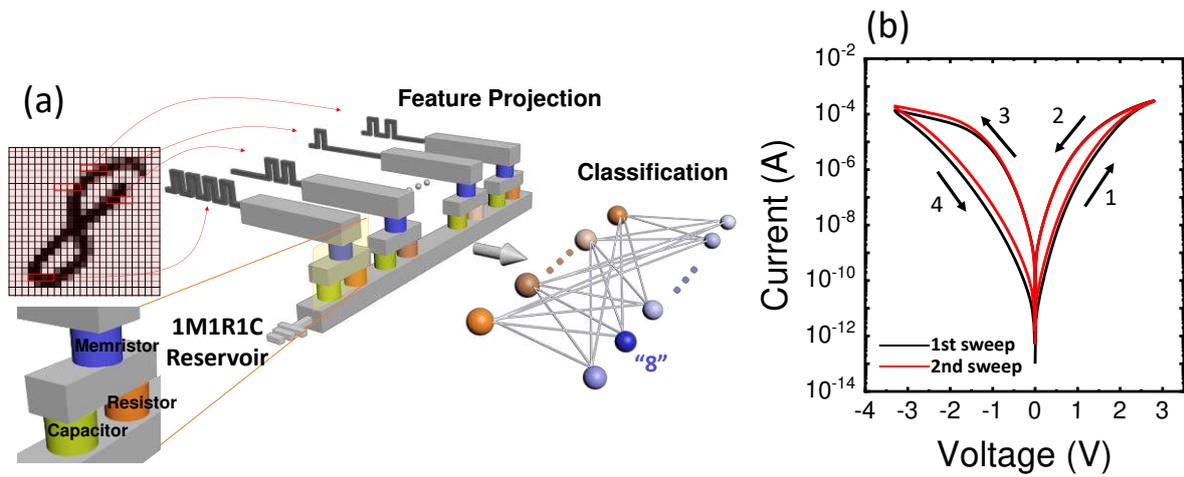


Figure 1: The structure of the 1M1R1C reservoir computing system and the I-V characteristics of the memristor used in the reservoir. **a)** The structure of the 1M1R1C reservoir computing system proposed in this study. The RC system can recognize images in the MNIST database through feature projection and classification. **b)** The I-V curve of the W/HfO₂/TiN memristor. The sweep order is marked in the figure. SET and RESET occurred in the positive bias and the negative bias, respectively, and gradual switching occurred in both switching conditions. Since the filament formation process is not required in this electronic switching device, no electroforming process is seen in the first sweep.

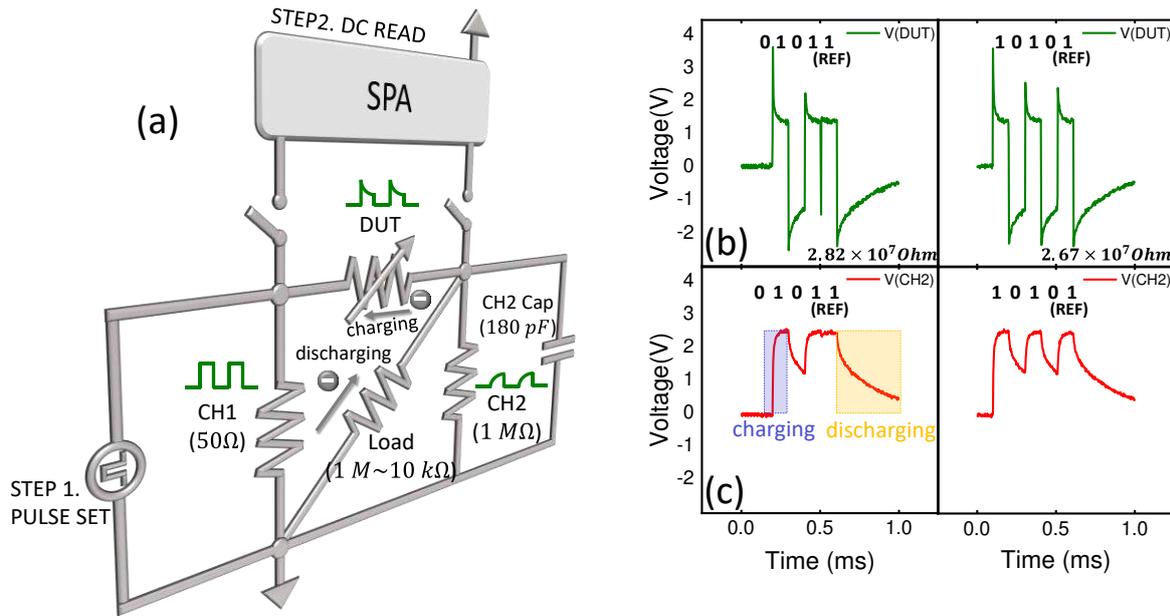


Figure 2: The circuit used as a reservoir in the experiment, and the V-t graphs were obtained from the DUT and CH2 of this circuit. **a)** A reservoir circuit composed of a memristor, resistors, and a capacitor. CH1 shows the shape of the input pulse stream, and CH2 shows the voltages applied to the 1M ohm resistor, from which the voltages across the DUT were obtained (green graph). **b)** The voltages applied to the memristor with a '0101+reference pulse' (left) and a '1010+reference pulse' (right). **c)** The voltages of the corresponding CH2, where the 4V and 0V voltage amplitudes represent '1' and '0,' respectively. The voltage across CH2 shows that the charging and discharging rates of the capacitor were asymmetric.

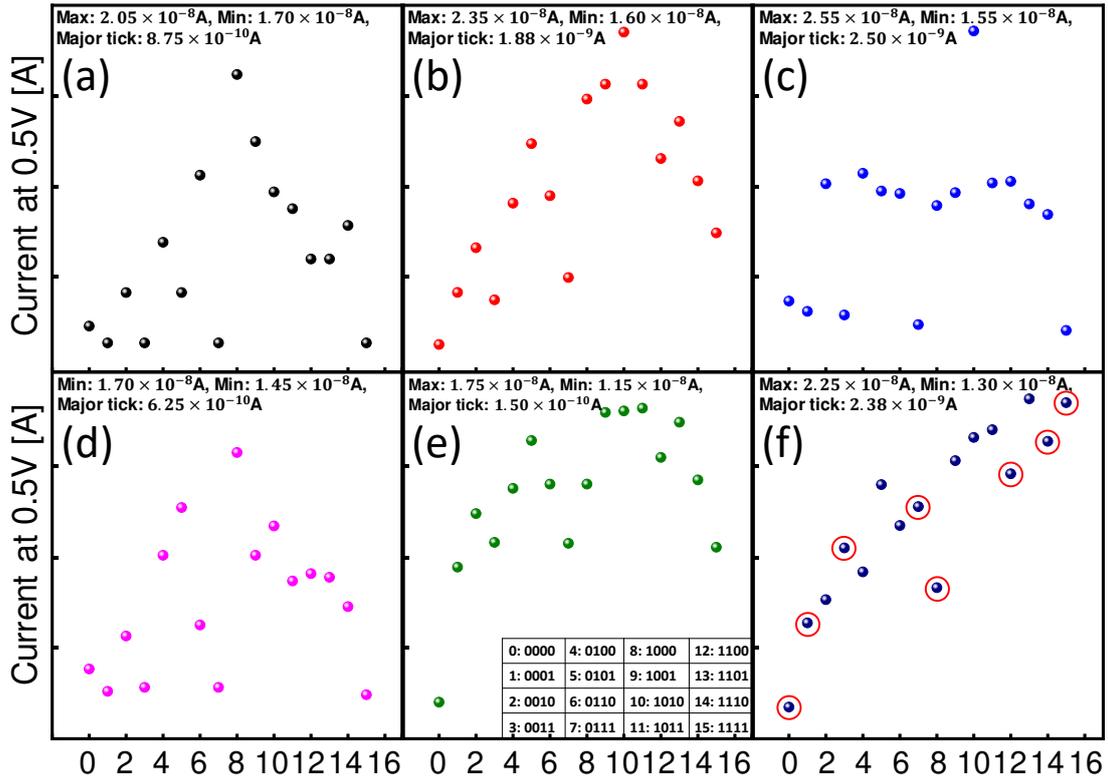


Figure 3: Experiment results to analyze the effect of changing parameters on the reservoir characteristics in the RC system. The read current at 0.5 V of the memristor for the pulse stream '0000'~'1111' corresponds to 0~15 in the inset table in **e**. **a**) The read current at 0.5V of the memristor for each input under the conditions of 1 M Ω R_L, 4 V signal pulse height, 100 μ s width, 4 V REF pulse height, and 100 μ s width. **b-e**) The read current at 0.5 V of the memristor for each input when R_L, pulse width, pulse height, and REF pulse height are changed respectively from the condition of **a**. The various parameter settings for each figure was summarized in Table I. The reservoir responses for each input of the reservoir optimized for the MNIST recognition are shown in **f**. Responses to inputs showing high prevalence in the dataset were well separated (marked by red circles).

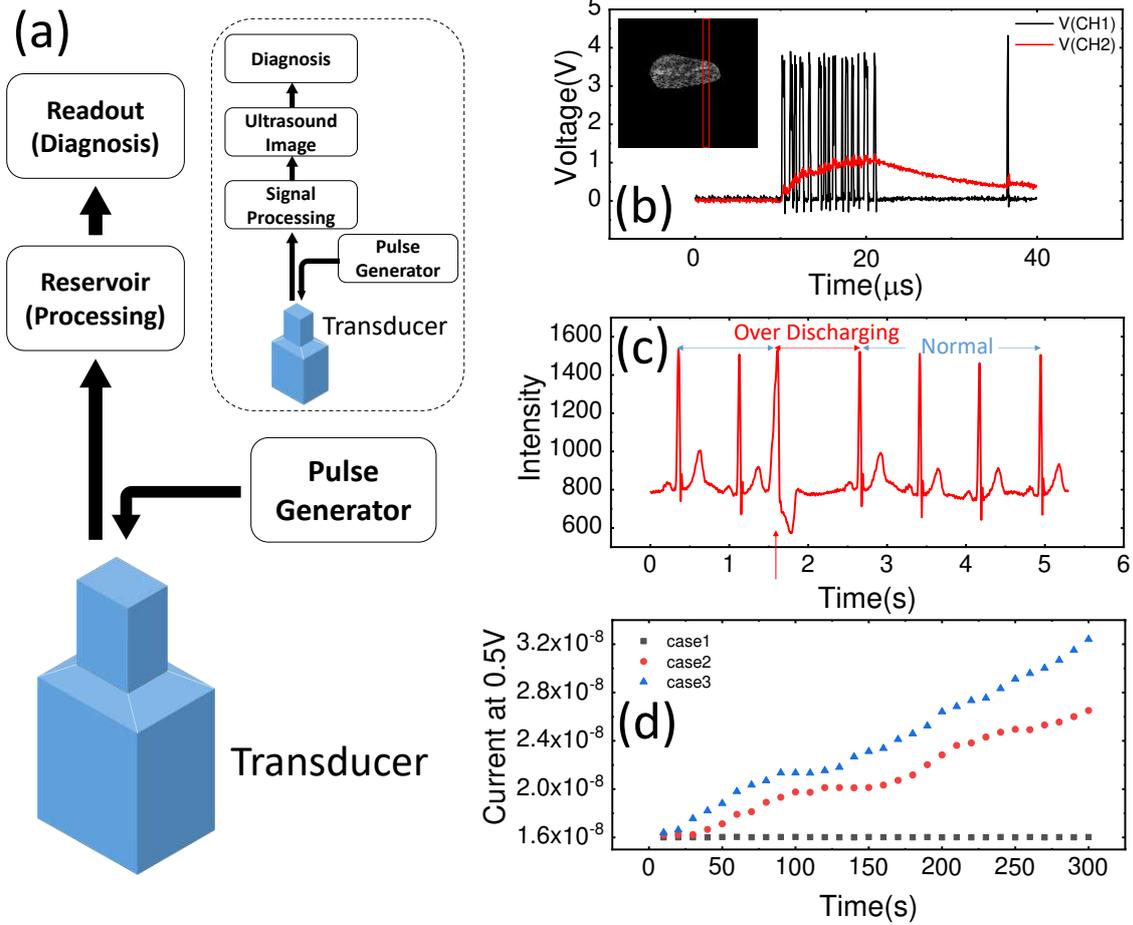


Figure 4: The automatic medical diagnosis system using the 1M1R1C reservoir and the experiment results in the two sections. **a)** A system for diagnosing the malignancy of breast lesions, which is much simpler than in the existing method (inset in **a**). In this system, ultrasonic signals are applied directly to the reservoir, so the imaging step is omitted. **b)** V-t graph for one echo line of a benign sample (inset in Figure 4b). **c)** A part of the electrocardiogram of a patient with arrhythmia. Long intervals caused by abnormal beats discharged the capacitor, and the conductance of the memristor increased in the next pulse. **d)** Five-minute reservoir monitoring based on the ECG of one normal patient (case 1) and two arrhythmic patients (cases 2 and 3). When arrhythmia occurred, the conductance of the memristor increased. Case 3, which had the most severe arrhythmia symptoms, showed the highest conductance.

Table I: The reservoir conditions (R_L , signal pulse, and REF pulse) used in Figure 3a-e

Reservoir Condition	R_L	Signal Pulse	REF Pulse
Fig. 3a	$1 M\Omega$	$4 V, 100 \mu s$	$4 V, 100 \mu s$
Fig. 3b	$120 k\Omega$	$4 V, 100 \mu s$	$4 V, 100 \mu s$
Fig. 3c	$1 M\Omega$	$4 V, 200 \mu s$	$4 V, 100 \mu s$
Fig. 3d	$1 M\Omega$	$3.5 V, 100 \mu s$	$3.5 V, 100 \mu s$
Fig. 3e	$1 M\Omega$	$4 V, 100 \mu s$	$3 V, 100 \mu s$
Fig. 3f	$10 k\Omega$	$3.5 V, 200 ns$	$3 V, 200 ns$

Table II: Comparison of the results of the MNIST recognition using memristive reservoir computing systems^{18,17} and a software-based system¹ (single-layer FCN), showing very fast processing and the highest accuracy in this study

Group	Accuracy	Latency in Reservoir	Reservoir Adaptation	Network Size	Image Size	Etc.
This Study	90%	$1 \mu s$	O	196×10	28×28	
Wei. D. Lu	85%	$10 ms$	X	88×10	22×20	14,000/2,000 Training/Test set
Joshua Yang	83%	$1 ms$	X	220×10	22×20	In situ training
Software (784x10 FCN)	88%	-	-	784×10	28×28	

Table III: Results of the MNIST recognition while increasing the number of bits processed in the reservoir, showing that as nBPR increased, both the size of the used readout layer and the recognition accuracy decreased

nBPR	Readout Layer Size	Accuracy
3	252x10	90.7%
4	196x10	90.1%
5	140x10	88.1%
6	112x10	86.3%

Figures

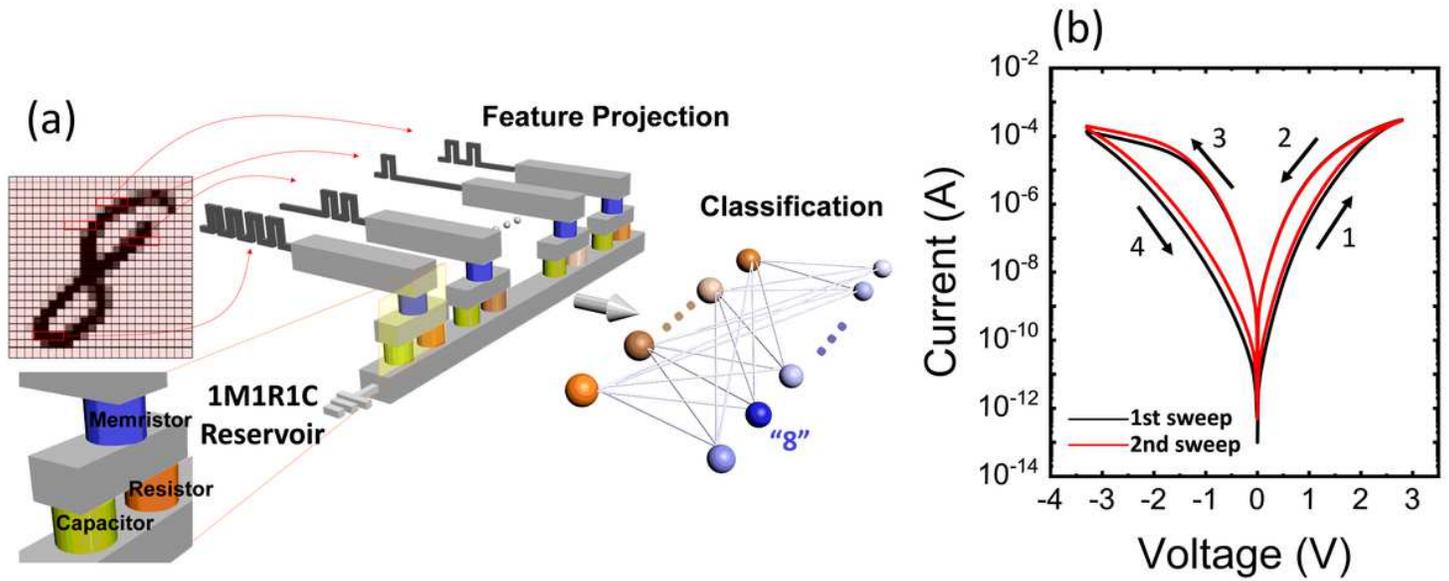


Figure 1

The structure of the 1M1R1C reservoir computing system and the I-V characteristics of the memristor used in the reservoir. a) The structure of the 1M1R1C reservoir computing system proposed in this study. The RC system can recognize images in the MNIST database through feature projection and classification. b) The I-V curve of the W/HfO₂/TiN memristor. The sweep order is marked in the figure. SET and RESET occurred in the positive bias and the negative bias, respectively, and gradual switching occurred in both switching conditions. Since the filament formation process is not required in this electronic switching device, no electroforming process is seen in the first sweep.

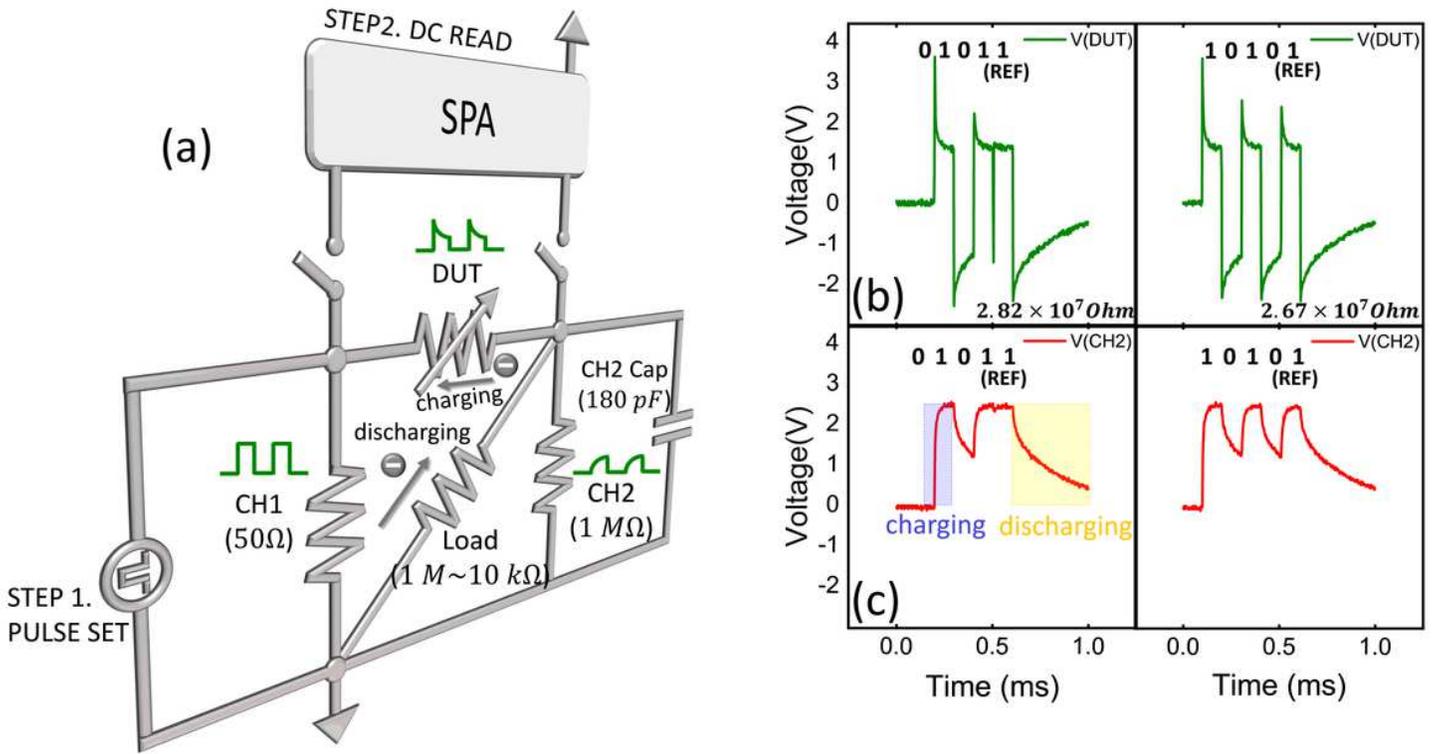


Figure 2

The circuit used as a reservoir in the experiment, and the V-t graphs were obtained from the DUT and CH2 of this circuit. a) A reservoir circuit composed of a memristor, resistors, and a capacitor. CH1 shows the shape of the input pulse stream, and CH2 shows the voltages applied to the 1M ohm resistor, from which the voltages across the DUT were obtained (green graph). b) The voltages applied to the memristor with a '0101+reference pulse' (left) and a '1010+reference pulse' (right). c) The voltages of the corresponding CH2, where the 4V and 0V voltage amplitudes represent '1' and '0,' respectively. The voltage across CH2 shows that the charging and discharging rates of the capacitor were asymmetric.

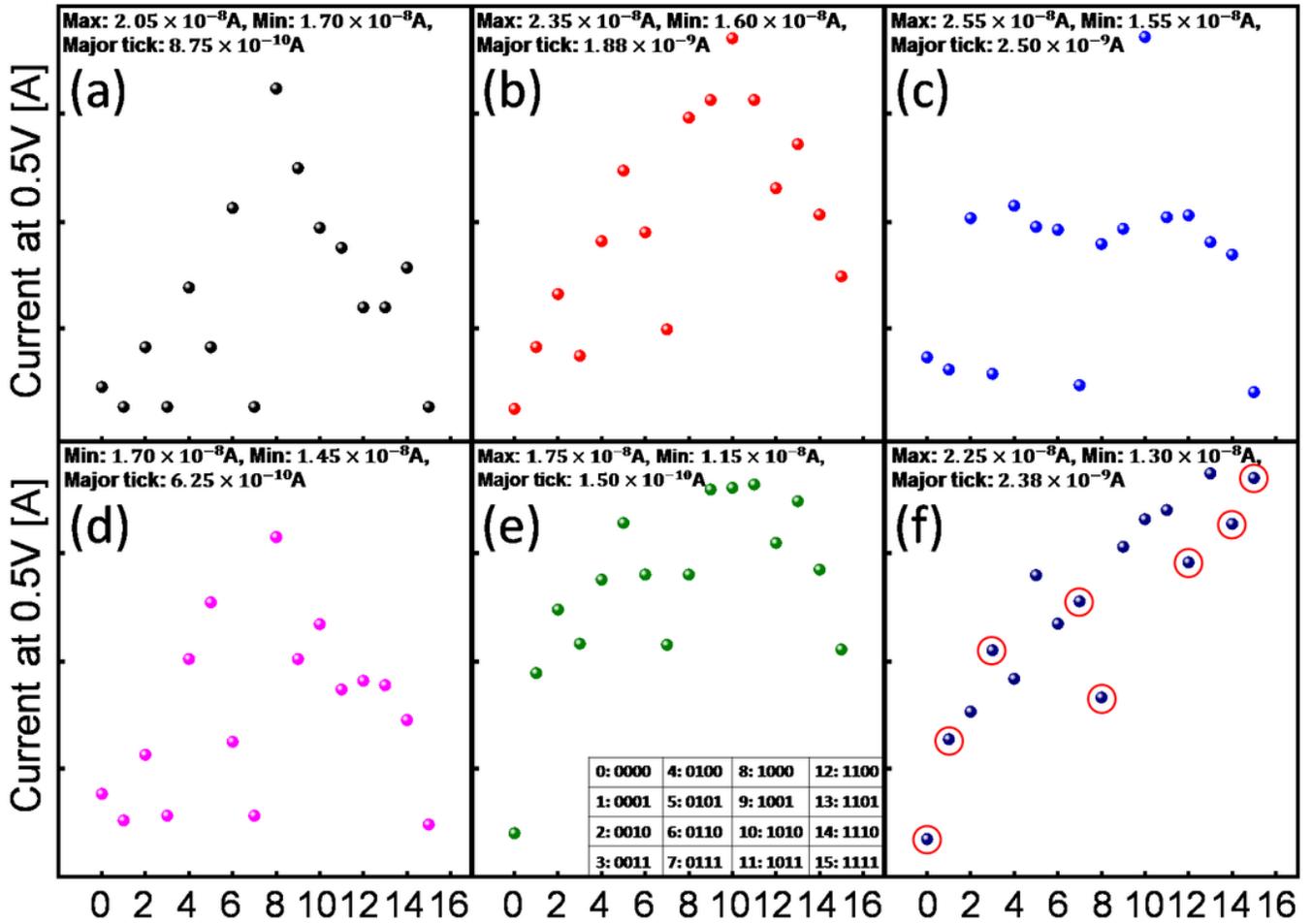


Figure 3

Experiment results to analyze the effect of changing parameters on the reservoir characteristics in the RC system. The read current at 0.5 V of the memristor for the pulse stream '0000'~'1111' corresponds to 0~15 in the inset table in e. a) The read current at 0.5V of the memristor for each input under the conditions of 1 M Ω RL, 4 V signal pulse height, 100 μ s width, 4 V REF pulse height, and 100 μ s width. b-e) The read current at 0.5 V of the memristor for each input when RL, pulse width, pulse height, and REF pulse height are changed respectively from the condition of a. The various parameter settings for each figure was summarized in Table I. The reservoir responses for each input of the reservoir optimized for the MNIST recognition are shown in f. Responses to inputs showing high prevalence in the dataset were well separated (marked by red circles).

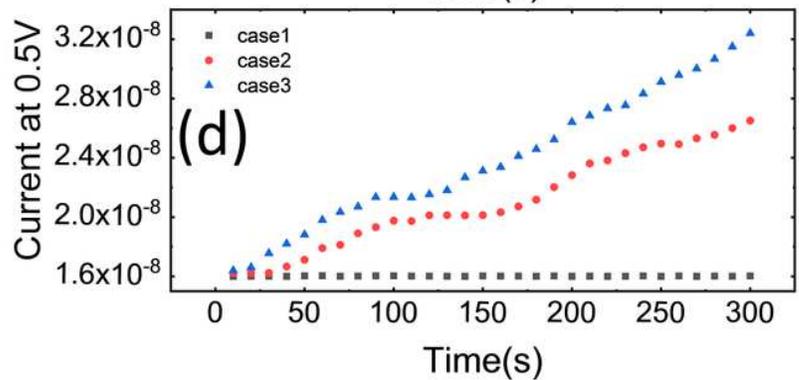
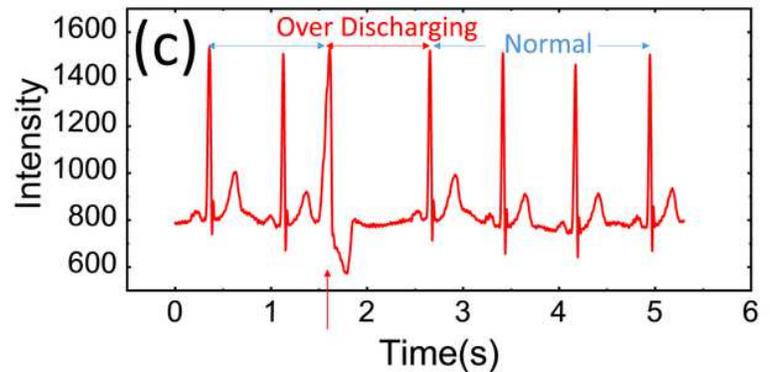
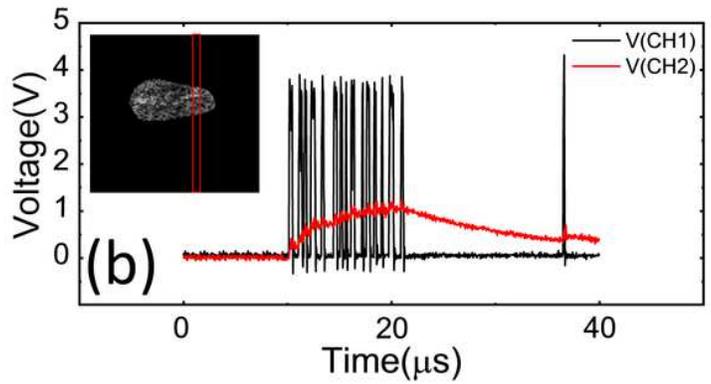
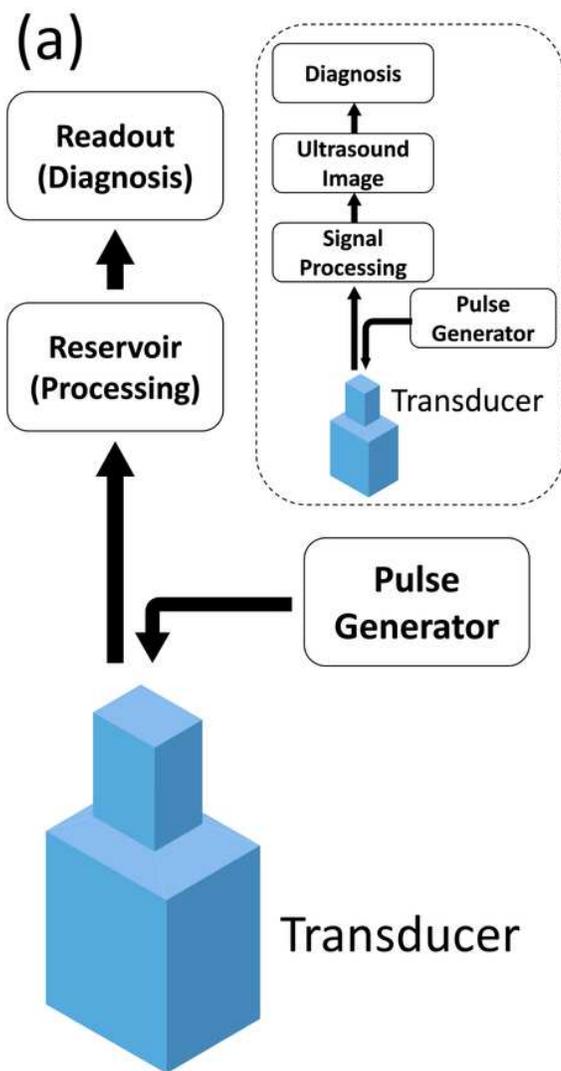


Figure 4

The automatic medical diagnosis system using the 1M1R1C reservoir and the experiment results in the two sections. a) A system for diagnosing the malignancy of breast lesions, which is much simpler than in the existing method (inset in a). In this system, ultrasonic signals are applied directly to the reservoir, so the imaging step is omitted. b) V-t graph for one echo line of a benign sample (inset in Figure 4b). c) A part of the electrocardiogram of a patient with arrhythmia. Long intervals caused by abnormal beats discharged the capacitor, and the conductance of the memristor increased in the next pulse. d) Five-minute reservoir monitoring based on the ECG of one normal patient (case 1) and two arrhythmic patients (cases 2 and 3). When arrhythmia occurred, the conductance of the memristor increased. Case 3, which had the most severe arrhythmia symptoms, showed the highest conductance.

Supplementary Files

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

- [RCManuscriptNatureCommunicationsSI.docx](#)