Proceedings Paper:

Gaurav, A., Song, X., Manhas, S.K. et al. (2 more authors) (2023) A solid electrolyte ZnO thin film transistor for classification of spoken digits using Reservoir Computing. In: 2023 7th IEEE Electron Devices Technology & Manufacturing Conference (EDTM). 7th IEEE Electron Devices Technology & Manufacturing Conference (EDTM) 2023, 07-10 Mar 2023, Seoul, Korea. Institute of Electrical and Electronics Engineers , pp. 1-3. ISBN 9798350332520

https://doi.org/10.1109/edtm55494.2023.10103131

A Solid Electrolyte ZnO Thin Film Transistor for classification of spoken digits using Reservoir Computing

Ankit Gaurav[1], Xiaoyao Song[2], Sanjeev Kumar Manhas[1], Partha Pratim Roy[3]and Maria Merlyne De Souza[2]

[1] Departments of Electronics and Communication, IIT Roorkee, Roorkee, India, [2] Departments of Electronic and Electrical, University of Sheffield, Sheffield, U.K, [3]Departments of Computer science, IIT Roorkee, Roorkee, India

Contact: m.desouza@sheffield.ac.uk; sanjeev.manhas@ece.iitr.ac.in.

Abstract

The duration and interval of the input signal in two terminal memristor based reservoir systems is constrained by the decay time of a conventional memristor. Here, we demonstrate that the third terminal of a Solid Electrolyte ZnO/Ta_2O_5 Thin-film Transistor (SE-FET) can be used to control the decay time by using a variable read voltage, without any additional circuit elements. Using this approach, we have benchmarked the performance of our SE-FET based RC system for a task of recognition of spoken digits with a high accuracy of 99.4%.

(Keywords: Reservoir computing, Solid electrolyte FET, Spoken-digit classification, Lyons passive ear model)

Introduction

Physical Reservoir computing systems enable the classification of complex timing related tasks much more efficiently than recurrent neural networks (RNNs)[1]. RNNs require backward connections to retain information about the input, the previous as well as present states of the network. They are prone to exploding and vanishing gradients, which makes the training difficult and expensive. Physical dynamical reservoirs are more economical as they dispense with hidden layers and operate by mapping the input signal onto a higher-dimensional spatiotemporal feature space that is directly read by the output layer. The richness of reservoir states requires devices with a short-term memory (or fading memory) that can be mapped directly to time-varying input data. Nevertheless, the duration and interval of the input signal are limited in two terminal memristor based reservoir systems by time range in which sufficient memory decay occurs which is typically in the range of few milliseconds [1]–[3]. This severely limits their adaptability for different dynamical inputs that vary from 10^{-7} seconds for static situations such as image processing to a few seconds for detection of malignancy of lesions via ultrasound or the detection of arrhythmia via electrocardiogram. In these studies, the reservoirs were constructed based on diffusive memristors which are a class of volatile memristors whose switching is governed by fast diffusive species (e.g., Ag) [2]. Moreover, to solve this issue, a non-volatile memristor was combined with circuits elements such as a resistor and capacitor [4] which adds to the cost and complexity of the reservoir.

Recently, we have demonstrated improved learning by increasing the dimensionality of an RC system based on a three terminal solid electrolyte ZnO/Ta2O5 TFT (SE-FET) $(W \times L = 100\mu m \times 1.5\mu m)$ by sampling at lower rate (~1 Hz or lower) after each pulse rather than reading an entire sequence of the input [5].

In the current work, we evaluate its performance for classification of isolated spoken digits. We also demonstrate that the third terminal can be used to control the temporal dynamics (decay time) of the device by using a variable read voltage in the off state, which makes this approach more adaptable in biomimetic system than by using external elements such as resistors and capacitors.

Methodology

The framework of our SE-FET based reservoir system for spoken digits is described in Fig.1. Before feeding the isolated spoken-digit into the reservoir, it is preprocessed using Lyon's passive ear model [6] based on human cochlear channels. The model includes a number of filters to divide the input signal into frequency channels, half wave rectifiers (HWRs) to identify the actual information from the filtered signal, automated gain control (AGC) to compress a signal with a high dynamic range (which varies over twelve orders of magnitude) to a reasonable range (of about two orders of magnitude). The processed audio file is then converted into a pulse stream. The resultant pulse stream is applied to the SE-FET and the read current is subsequently measured and recorded after each pulse. The measured read current forming reservoir states were then used to train the readout network. We trained our readout network using logistic regression on 450 samples from the standard benchmark NIST TI46 database [7] of (0-9) digits spoken 10 times by 5 different female speakers, and testing was achieved on a separate sample set of 50, not used in training. To avoid the system from being overfitted to some selections of the training and testing data, 10-fold cross-validation was used.

Result and Discussion

The schematic of an SE-FET device, showing charge separation of oxygen ions and vacancies at respective opposite interfaces of the Ta_2O_5, upon application of positive gate voltage is shown in Fig. 2. The movement of oxygen vacancies (V^{2+}) in the gate insulator Ta_2O_5 caused by an electric field, drives the operation of this device [8]-[9]. The measured conductance values of the SE-FET with a pulse train of ±1 V/ 60 ms applied at the gate terminal shows the gradual change in conductance which results in multiple conductance values as shown in Fig. 3(a).

The conductance was measured in the off state [8] i.e. when no gate pulse is applied, as shown in the inset of Fig. 3(a). The controllable memory decay of the SE-FET is demonstrated in Fig. 3(b). This shows three different memory time decay constants of $\tau \approx 4s$, 6s, 25s achieved by using different read voltages applied at drain terminal. This temporal response of the SE- FET is attributed to the polarity-induced motion of oxygen vacancies in the gate insulator [9]. These results show that a negative read voltage can be used whose magnitude can control the retention time of the memory. Moreover, the processing of the same input using an SE-FET at different frequencies results in distinguishable output response as shown in Fig.4 (a). This shows that the SE-FET can be adapted to different temporal dynamics of the input data. Further, the response of an SE-FET to all possible 8 temporal combinations of a 3-bit input sequence is shown in Fig.4(b) which shows the uniqueness of the output when subjected to a unique combination at the inputs.

To demonstrate the SE-FET based RC system for spoken digit classification, we first digitized the pre-processed spoken digit by converting the Cochleagram (where each data point represents the firing probability of a hair cell sensitive to a certain frequency at a given time point) shown in Fig. 5 by setting a threshold (0.3) for the firing probability. We find that by setting a threshold of 0.3 preserves the richness of input while digitizing the Cochleagram. (This data point represents the firing probability). The digitized output is shown in Fig.6(a). Theoretically, there may be 2^{42} different input patterns if the entire channel were used as one single input pulse stream, which would be too challenging for one SE-FET to differentiate. Therefore, to enable better input separation and enhance the reservoir's dimensionality, we separated each channel into 14 sub-sections, each of which contained three data points. To prevent the influence from the previous input sequence, a small reset pulse of 3 V with a pulse width of 0.5 s is applied after each input sequence. With this consideration, the reservoir is fed with input voltage pulse streams, and the output response is recorded after each input pulse. Similarly, to extract features, each channel number (total 60 frequency channels each with 42-time steps) of the input is fed into the reservoir by converting it into a pulse stream (as exemplified in Fig.6(b) for channel number 53). As an example, all reservoir states corresponding to spoken digit '0' are shown in Fig.6(c). Through 10-fold cross-validation of the test set, an overall mean accuracy of 99% is attained for this implementation. Further improvement in accuracy is obtained by scanning the input time step-wise (vertically) rather than channel wise (horizontally). For this an overall mean accuracy of 99.4% was obtained, which is slightly better than the previously reported accuracy of 99.2% in memristor-based reservoir systems [3] in which a single read operation was performed at the end of the sequences. Our SE-FET-based RC system performs at par to earlier published work but do need the other physical devices to control decay time or different input processing techniques for example 99.6%, ~90% and 99.6% for spintronic[10], single magnetic domain wall[11] and memristor based RC system[12] respectively.

Conclusion

We experimentally demonstrated the SE-FET based reservoir system for spoken digit classification task. We demonstrate that the third terminal can be used to control the temporal dynamics decay time of the device by using variable read voltage in the off state. This is a significant advantage for applications that are suitable for real time monitoring of health conditions such as heart arrhythmia and detection of tumors via ultrasound.

Acknowledgments

The authors gratefully acknowledge that this project was funded by MHRD-SPARC under grant code No. P436 between University of Sheffield, UK and IIT-Roorkee India.

References

[1] C. Du, *et al.*, "Reservoir computing using dynamic memristors for temporal information processing," *Nat. Commun.*, 2017.

[2] R. Midya *et al.*, "Reservoir Computing Using Diffusive Memristors," *Adv. Intell. Syst.*, 2019.

[3] J. Moon *et al.*, "Temporal data classification and forecasting using a memristor-based reservoir computing system," *Nat. Electron.*, 2019.

[4] Y. H. Jang *et al.*, "Time-varying data processing with nonvolatile memristor-based TK," *Nat. Commun.*, 2021.

[5] A. Gaurav *et al.*, "Reservoir Computing for Temporal Data Classification Using a Dynamic Solid Electrolyte ZnO Thin Film Transistor," *Front. Electron.*,2022.

[6] R. F. Lyon, "A computational model of filtering, detection, and compression in the cochlea," *ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process.,* 1982.

[7] "Texas Instruments-Developed 46-Word Speaker-Dependent Isolated Word Corpus (TI46) NIST Speech Disc 7-1.1," 1991.

[8] X. Song *et al.*, "Off-State Operation of a Three Terminal Ionic FET for Logic-in-Memory," *IEEE J. EDS*, 2019.

[9] P. B. Pillai *et al.*, "Nanoionics-based three-terminal synaptic device using zinc oxide," *ACS Appl. Mater. Interfaces*, 2017.

[10] J. Torrejon *et al.*, "Neuromorphic computing with nanoscale spintronic oscillators," *Nature*,2017.

[11] R. V. Ababei *et al.*, "Neuromorphic computation with a single magnetic domain wall," *Sci. Rep.*, 2021.

[12] Y. Zhong *et al.*, "Dynamic memristor-based reservoir computing for high-efficiency temporal signal processing," *Nat. Commun.*, 2021.

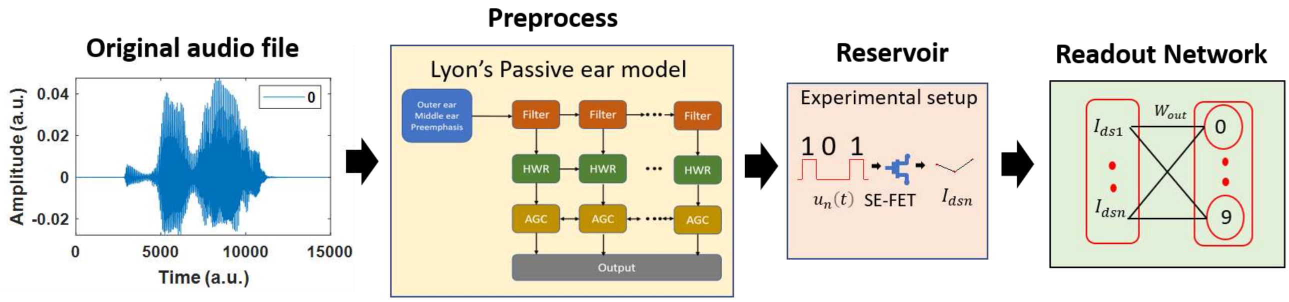

Fig.1. Framework and process flow of a dynamic SE-FET-based reservoir system for spoken digit classification. The original spoken digit is preprocessed by Lyon's passive ear model and digitized before feeding into the reservoir. The digitized input is then converted into a pulse stream and applied to the SE-FET. The subsequent read current is measured and recorded after each pulse and used to train the readout network using logistic regression.

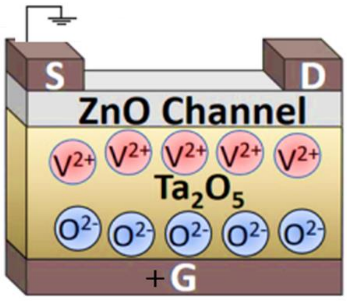

Fig.2. Schematic of a Ta_2O_5/ZnO SE-FET device, showing charge separation of oxygen ions and vacancies at respective opposite interfaces of the Ta_2O_5, upon application of an applied gate voltage.

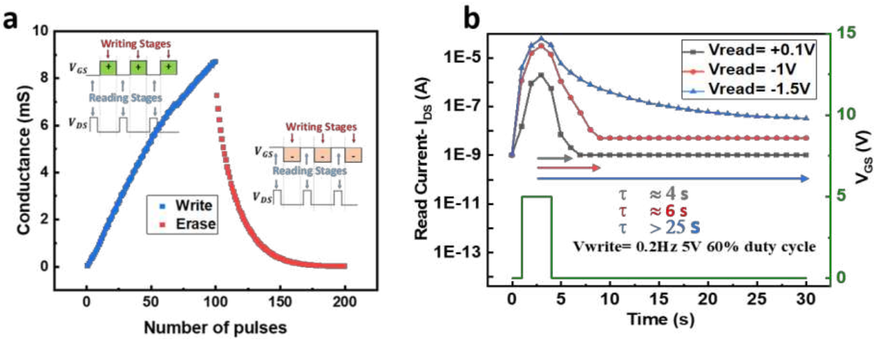

Fig.3. (a) Measured conductance of the SE-FET in the off-state showing gradual changes upon application of positive write pulses (+1 V, 60 ms) and negative erase pulses (−1 V, 60 ms). **(b)** Control of the temporal dynamics (fading memory) of the device is as simple as using a variable read voltage. The device is first programmed by single write pulses of 5 V at 0.42 Hz with 60% duty cycle and its subsequent read current is measured at 3 different read voltages. Showing three different memory time decay constants (τ).

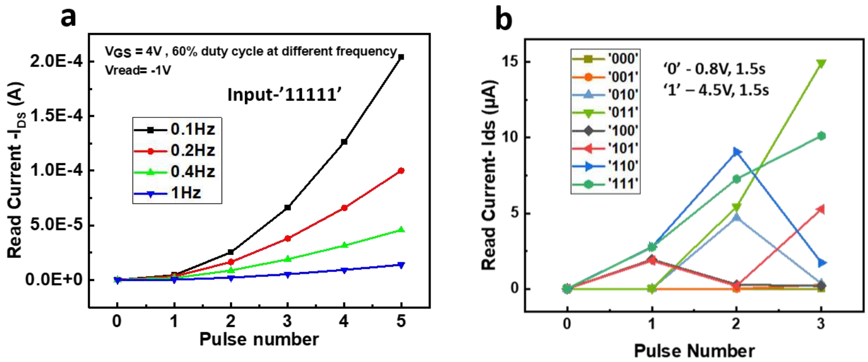

Fig.4. (a) Processing the same input using an SE-FET at different frequencies results in distinguishable output responses. **(b)** The SE-FET response when subjected to all possible eight combinations of temporal inputs, shows the uniqueness of the output.

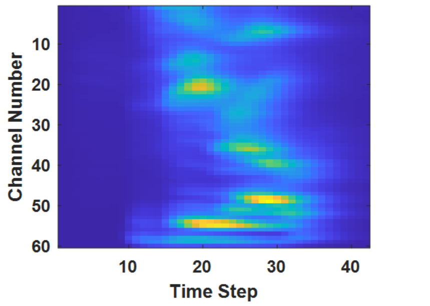

Fig.5. Cochleagram of the sample '0' after being processed by Lyon's passive ear model.

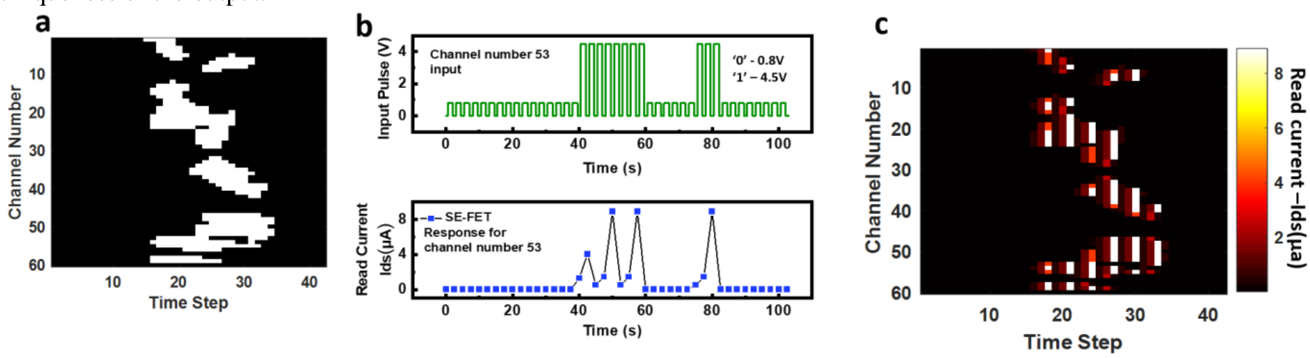

Fig.6. Results of a dynamic SE-FET-based RC system **(a)** Pre-processed digitized isolated spoken-digit '0'. **(b)** The temporal response of the SE-FET to an input pulse stream for channel number 53 is shown as an example. **(c)** The heatmap shows the complete recorded output response of spoken digit 0. The magnitude of the read current ranging from (0- ~10 μa).