
Proceedings Paper:
Song, X., Gaurav, A., Pillai, P.B. et al. (5 more authors) (2023) Reservoir computing based on a solid electrolyte ZnO TFT: an attractive platform for flexible edge computing. In: 2023 IEEE International Flexible Electronics Technology Conference (IFETC). 5th IEEE International Flexible Electronics Technology Conference (IFETC) 2023, 13-16 Aug 2023, San Jose, USA. Institute of Electrical and Electronics Engineers , pp. 1-3. ISBN 979-8-3503-3209-4

https://doi.org/10.1109/ifetc57334.2023.10254868

Reservoir Computing based on a Solid Electrolyte ZnO TFT : An attractive platform for flexible edge computing

Xiaoyao Song, Ankit Gaurav[1], Premlal B Pillai, Ashwani Kumar, Sanjeev Manhas[1], Aditya Gilra[2], E. Vasilaki[2] and M. M. De Souza
EEE department, University of Sheffield, Sheffield, UK
[1]ECE Department, IIT Roorkee, Roorkee India
[2]Department of Computer Science, University of Sheffield, Sheffield UK.
e-mail: m.desouza@sheffield.ac.uk;

***Abstract*—Implementation of accurate neural network models in edge applications such as wearables is limited by the hardware platform due to constraints of power/area. We highlight novel concepts in reservoir computing that rely on a volatile three terminal solid electrolyte thin film synaptic transistor, whose conductance can be controlled by the gate and drain voltages to enhance the richness of the reservoir and operate in the off-state. The proposed approach achieves an accuracy of 94% in image processing, significantly higher than equivalent applications of reservoir computing based on two-terminal memristors, primarily because we avoid down-sampling by training the readout after every pulse.**

Keywords—reservoir computing, Solid electrolyte FET, ZnO/Ta_2O_5)

I. INTRODUCTION

Fully integrated skin-inspired sensor systems based on sensor arrays, signal processing, and wireless transmission are of significant interest in healthcare diagnostics, prosthetics, motion tracking, soft robotics and AR/VR based human-machine interfaces. To realize such all-in-one integrated flexible sensor systems, signal collection, processing, and wireless transmission modules should ideally be incorporated with the sensing components on the same platform [1]. Nevertheless, present day capability favours hybrid flexible electronics, where soft and flexible sensor components are integrated with ultra-low power silicon-based CMOS chips on flexible PCBs [2], mainly because oxide electronics remains hampered by the lack of a high performing p-type transistor counterpart. Such systems can benefit from machine learning algorithms not just to minimise the amount of data stored on the sensor or transmitted to the cloud but also to potentially minimise the number of sensors required at the system level [3].

In comparison to non-volatile memory candidates such as RERAM, STT-MRAM or PCM, we highlight our progress with a synaptic three terminal ZnO thin film transistor with a Ta_2O_3 gate insulator, the Solid Electrolyte FET (SEFET) [4]. The device volatility makes it eminently suitable for Reservoir computing (RC), a branch of AI that offers a highly efficient framework for processing temporal inputs at a low training cost compared to conventional Recurrent Neural Networks (RNNs).

II. METHODOLOGY

A. Experimental Fabrication and Device Mechanism

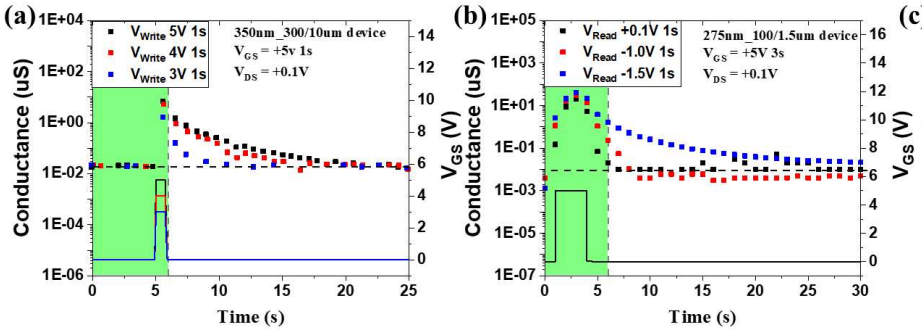
Bottom gated TFTs were fabricated on glass using conducting Indium Tin Oxide as the gate, 275 nm Ta_2O_5 as gate insulator and 40 nm ZnO as channel via Radio Frequency sputtering. The devices were subjected to thermal annealing at 80^0 C for 24 hrs. Patterning of the source/drain regions and deposition of Al metal S/D contacts were achieved using standard photolithography, thermal evaporation and finally lift off in organic solvents. The electrical characteristics were measured using a Keysight B2902A SMU unit with a Desert Cryogenic probe station.

The conventional ReRAM is a non-volatile and promising two terminal technology that is attractive because it is (i) simple to integrate on top of CMOS, and (ii) capable of high switching speed (~1ns). It relies on a filamentary process that typically requires a high forming voltage. The read and write operations occur at the same terminal resulting in typically high power consumption in $\mu W - mW$, although exceptions are reported in [5]. It is challenging to reliably scale a RERAM at < 0.3V, resulting in a trade-off between speed and operating voltage (common to all memory devices). The STT MRAM is also ideally not suited for edge computation due to high power for write, as well as high write disturb. A three-terminal non-volatile Ferroelectric RAM (FERAM) on the other hand offers advantages of low power by operating the device in the off state. However, its non-volatility implies that external RC elements need to be incorporated alongside the device in order to control the volatility in applications of reservoir computing [6].

The SEFET is a thin film transistor, based on a non-filamentary process. Application of a gate voltage induces motion of oxygen vacancies in a relatively thick Ta_2O_5 insulator, that leads to an internal electric field in the insulator, opposite in direction to the applied field. This is a necessary condition for negative capacitance, with a subthreshold slope of upto 26 mV/decade achieved practically in our transistor during the reverse sweep of the gate voltage [7]. The gate current characteristics show a distinctive negative differential resistance in its characteristics [8], making it an example of an active memristor. The change in drain conductance, upon application of a gate voltage can be directly measured between the source and drain (S/D) terminals, but with the added benefit of operation when the device is off [9]). The use of a separate control terminal makes the implementation of a reservoir efficient and straightforward compared to a two-terminal-based artificial neural network.

The unique ionic migration-based switching mechanism of the SEFET also presents new opportunities for multiple device conductance states and retention time that leads to enhanced richness of the reservoir. The maximum conductance after potentiation shown in Fig.1 and the subsequent retention time exhibit strong linearity with the write voltage, which is crucial in neuromorphic applications. Since ions migrate by drift diffusion, the conductance and retention are determined by the

Present address: Premlal B Pillai: Department of Physics, University of Manchester. Ashwani Kumar: Focal Agent Ltd, 124 City Road, London, EC1V 2NX; Aditya Gilra: Centrum Wiskunde & Informatica, Science Park 123, 1098 XG Amsterdam, Netherlands.



(c)

Action table	Pulse amplitude (V)		
Input \ Bit	Xxx	xXx	xxX
0	0.8	0.8	0.8
1	3.5	3.5	3.5
2	4.5	4.5	4.5
3	5.5	5.5	5.5

Fig. 1. The conductance and voltage scheme of the SE-FET as the function of the device **(a)** write and **(b)** read voltage, and **(c)** Action table mapping the gate voltage to multiple pixel values.

relative potential difference between the drain/source and the gate. Consequently, adjusting the read voltage can also modify these parameters. Applying a positive potential results in a higher potential on the drain/source compared to the gate, promoting faster decay. Conversely, using a negative read voltage leads to a relatively higher potential at the gate than the drain/source, thus enhancing the conductance and retention time. Fig 1 (c) shows a quaternary assignment of positive gate voltages to pixel values, but this can be significantly enhanced by further optimisation of the reading scheme.

B. Reservoir Computing

Our framework of the SE-FET based reservoir system is described in Fig. 2. In the first instance, we repeatedly use characteristics from a single SEFET fed by a temporal input pulse stream $u_n(t)$ rather than multiple FETs, because of excellent reproducibility as highlighted in Fig 3. An actual hardware implementation of a reservoir with multiple SE-FETs (for enhanced speed), would not require interconnection, as in conventional software-based RC systems because of the inherent short-term memory of our device [10] . A small read voltage of -1V after *each* pulse applied to the drain terminal, when the gate is off, is used to read the drain conductance. This process is repeated for all input sequences, and after each input sequence, a reset pulse is applied to set the device back to the initial state. We have shown that reading the conductance after every pulse, rather than down-sampling, increases the richness of the reservoir, leading to much higher learning efficiency than has been reported in reservoir computing by any comparable memristors to date [11]. The output response of 3 different SE-FETs, when subjected to a subset of all 64 unique sequences, corresponding to quaternary rather than binary digitsis shown, resulting in a pool of 192 (3× 64) reservoir states in Fig. 3, where each sequence has 3 different recorded reservoir states due to device-to-device variation, which adds to the richness of the reservoir. Although variation can be seen in terms of current magnitude from device to device, all devices show the same trend. The reservoir states are spread non-linearly throughout the range of readout current ~ (0- 60 µA). This shows the ability of nonlinear transformation of the input signal when subjected to all possible input scenarios.

We use the measured and recorded read current values for training and testing the readout network offline. Only the weight matrix W_{out} connecting the reservoir states to the output were trained using logistic regression with the liblinear Solver (Library for Large Linear Classification) from Python's scikit-learn library (scikit-learn), which uses a gradient descent algorithm.

We have explored image and voice recognition as examples of applications. For image processing, we trained our readout network using 60000 images from the MNIST database. For testing, a separate 10000 image sample set was fed into the reservoir. For voice recognition the NIST TI46 database consisting of 500 samples of spoken digits (0-9) by 5 different female speakers was used. Lyon's passive ear model was used

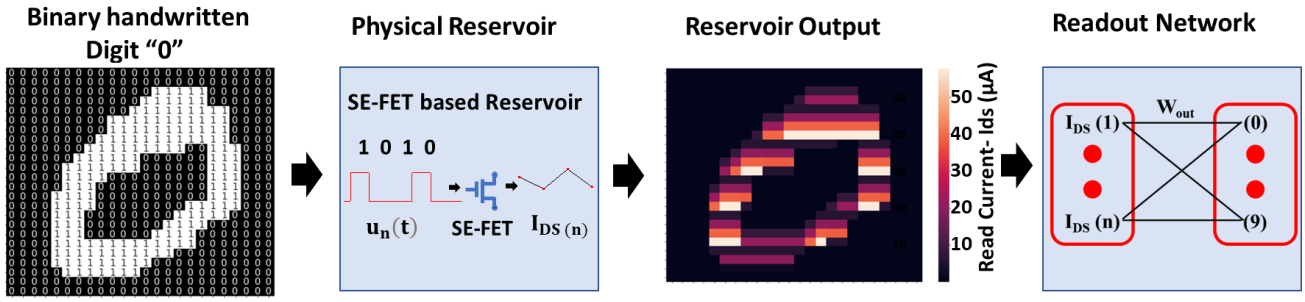


Fig. 2. Framework and process flow of a dynamic SE-FET-based reservoir system. The pre-processed binary MNIST handwritten digit is fed as a sequence of 4-bit binary numbers into the SE-FET device acting as a reservoir. The role of the reservoir here is to map simple binary features into much richer feature sets. The output of the reservoir shows different magnitudes of the read current adding more level of information than just 0 and 1. The subsequent read current is measured and recorded after each pulse and used to train the readout network using logistic regression. This simple approach in reservoir computing can be used to convert any input into richer features sets before feeding into the read out network for high accuracy.

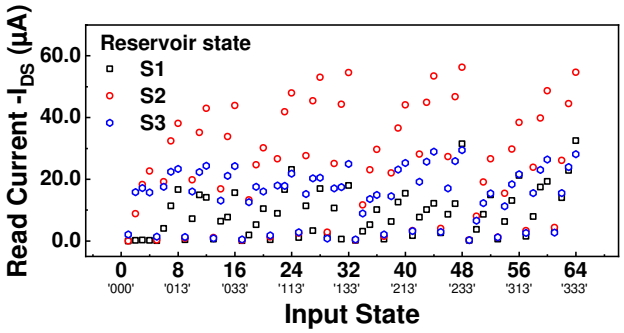


Fig. 3. Example response of 3 different SE-FET device (S1, S2, and S3) subjected to 64 temporal input combinations showing similar trends as well as device to device variability.

to pre-process the speech before feeding into the reservoir. The model includes filters to divide the input into frequency channels, half wave rectifiers to identify the actual information from the filtered signal, and automated gain control to compress it.

III. RESULTS

A. *Performance of the Reservoir*

Each row of the image was divided into 6 sub-sections, each containing 4 pixels, to allow better separation of the inputs. After each input sequence, a small reset pulse of -3V was applied. The image was fed into the reservoir in 4-pixel sub-sections as input voltage pulse streams, and the output response recorded after each input pulse. The output corresponding to MNIST digits 5 and 2 is shown in Fig. 4. The difference in read current at each sequence number results in better classification of the digits.

In the case of audio applications, the pre-processed spoken digit 0 using Lyon's passive ear model is shown in Fig. 5 (a), where the lower channel number captures the higher frequency components and vice-versa. This allows the capture of

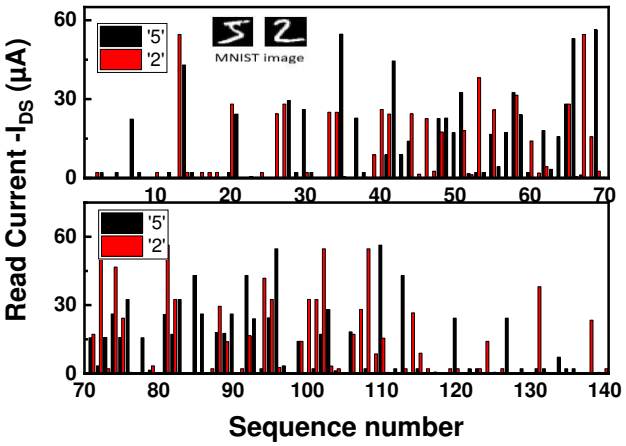


Fig. 4. Reservoir output corresponding to the MNIST image of digit "5" and "2" in the inset, showing the difference in read current value at each sequence number. Each row of the image is fed into the reservoir by converting it into a pulse stream (of 0's and 1's) divided into a sequence of 4 pixels each which are numbered from 1-140.

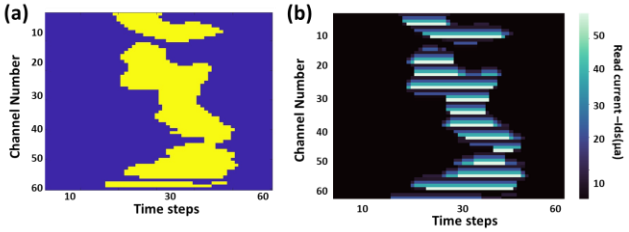


Fig. 5. (a) The pre-processed digitized spoken digit 0 by Lyon's passive ear model. **(b)** The complete response of the reservoir to spoken digit 0 adding more levels of information than just 0 and 1.

TABLE I. Comparison of the SE-FET based RC systems with reported literature.

Ref.	Device	MNIST Accuracy	NIST TI46 Accuracy
[13-14]	Memristor (WO_x)	88.1%	99.2%
[15]	Memristor (SiO2:Ag)	83%	-
[16]	Memristor (TiO_x/TaO_y)	-	99.6%
[17]	Spintronic	-	99.6%
[18]	Photonic	~97%	-
[11-12]	SE-FET (ZnO/Ta_2O_5)	94.44%	99.4%

frequency and amplitude of audio as essential features for speech classification. The complete response of the SE-FET reservoir corresponding to pre-process spoken digit is shown in Fig.5 (b). This is then used for training and testing of the readout network.

A comparison of our work with others reported in the literature is indicated in Table I. Audio leads to higher accuracy mainly because of a limited dataset for training and testing. Other than photonic neural networks, our approach highlights higher learning accuracy for image processing.

IV. CONCLUSIONS

This work highlights new approaches to harness the variability and volatility of a TFT to generate high dimensionality reservoirs by read and write operations that vary with voltage, as well as by increasing the sampling frequency. The transistors are readily integrable at the BEOL and amenable to Compute in Memory for the implementation of data processing.

ACKNOWLEDGMENT

We acknowledge funds from UKIERI and MHRD-SPARC under grant code No. P436 between University of Sheffield and IIT Roorkee

REFERENCES

[1] E. Ozer *et al.*, *Nat. Electron.* 3, 419 (2020); [2] W. Gao *et al.*, *Nature*, 529, 509,(2016); [3] K. K. Kim *et al.*, *Nat. Commun.*, 11, 2149, (2020). [4] P. B Pillai et al, *ACS Appl. Mater. Interfaces*, 9, 1609 (2017). [5] H.-S. P. Wong, et al: https://nano.stanford.edu/stanford-memory-trends. [6] Y. H. Jang *et al.*, *Nat. Commun.*, 12, 5727, (2021). [7] A. Kumar, et al, *ACS Appl. Mater. Interfaces*, 10, 19812 (2018). [8] P. B. Pillai, et al, *ACS Appl. Mater. Interfaces*, 10, 9782 (2018). [9] X. Song et al, *IEEE J. Elect Devices Soc.*, 7, 1232 (2019). [10] G. Tanaka *et al.*, *Neural Networks*, 115, 100, (2019). [11] A. Gaurav *et al.*, *Front. Electron.*, doi: 10.3389/felec.2022.869013 (2022) [12] A.Gaurav, et al Proc. of EDTM, 2023.[13] C. Du, et al *Nat. Commun.*, 8, 1 (2017). [13] J. Moon *et al.*, *Nat. Electron.*, 2, 480 (2019). [14] R. Midya *et al.*, *Adv. Intell. Syst.*, 1,. 1900084, (2019). [15] Y. Zhong, et al, ," *Nat. Commun.*, 12,. 1 (2021). [16] J. Torrejon *et al.*, *Nature*, 547, 428 (2017). [17] R. M. Nguimdo, et al, *Opt. Express*, 28, 27989 (2020). [18]R. M. Nguimdo et al Opt. Express, vol. 28, no. 19, p. 27989, 2020.