Proceedings Paper:
De Souza, M.M. orcid.org/0000-0002-7804-7154, Song, X., Gaurav, A. et al. (6 more authors) (2023) A delay system reservoir based on a nano-ionic Solid Electrolyte FET*. In: 2023 IEEE Nanotechnology Materials and Devices Conference (NMDC). 2023 IEEE Nanotechnology Materials and Devices Conference (NMDC), 22-25 Oct 2023, Paestum, Italy. Institute of Electrical and Electronics Engineers (IEEE) , pp. 139-140. ISBN 9798350335460

https://doi.org/10.1109/nmdc57951.2023.10344264

A delay system reservoir based on a nano-ionic Solid Electrolyte FET*

M. M. De Souza, X. Song, A. Gaurav[1], S. K. Manhas , P. B. Pillai, S. Sikdar, A. Kumar, A. Gilra[2], E. Vasilaki[2]
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
Invited Paper.

***Abstract*— Reservoir computing using delay systems took off when it was demonstrated that a single volatile memristor can be effectively utilized to reduce the burden of training recurrent neural networks, without the need for any interconnected reservoir nodes. Here we demonstrate that the maximum impact on the learning efficiency of a ZnO/Ta_2O_5 SEFET based reservoir is derived from (i) training the output after every pulse, (ii) device variability, and (iii) scanning the input image data horizontally and vertically.**

I. Introduction

Edge computing is becoming increasingly significant because of its potential to reduce the amount of data transferred to a "cloud" from any sensor platform. The main challenge facing edge computation is autonomy of operation in terms of decision-making as well as energy consumption. A well-known problem of memristor-based architectures for embedded intelligence lies in the accuracy of computation even in off-chip learning. For example, it is practically impossible for a memristor to match the 16-64bit accuracy of conventional Von Neumann processors. On-chip training incurs high overheads in terms of hardware, sub-optimal fidelity, long training times and chip specific training processes, whereas off-chip training, when it is undertaken in the cloud, causes inaccuracy, when translated on a sensor platform with different underlying technology. The issues affecting memristor based platforms may be attributed to variability, non-linearity of devices, as well as higher voltages during programming. Despite this, hybrid CMOS-memristor architectures are popular because two-terminal memristors can be co-integrated with ease on top of CMOS transistors (largely due to compatible materials/processes), as well as advances in 3D integration that suit the fabrication of 3D memristor arrays.

In applications such as voice or gesture recognition, the neural networks require the same time constants and dynamics as their input signals, ranging from the order of milliseconds-seconds. Such time scales are too slow for digital implementations, both in terms of hardware as well as AI algorithms. To address these, delay circuits have been demonstrated with external components in order to emulate a wide range of plasticity[1,2].

In this work we highlight a delay-system reservoir based on a three-terminal non-filamentary ZnO/Ta_2O_5 thin film transistor [3]. The memory of the device arises via a gate voltage dependent diffusion of vacancies in the gate insulator, resulting in long delay times (milliseconds-hours). The memory decay can further be adjusted via the applied gate and drain voltages, during read and reset [4]. Operating the device in the off-state, minimizes the power consumption[5]. In comparison to other reported works on memristor based reservoirs, we have demonstrated an improved learning efficiency of 3%, by training the readout network after every applied pulse, rather than training the weights at the end of the measured input sequence [6].

II. Methodology

A. *Experimental fabrication and mechanism of the SEFET*

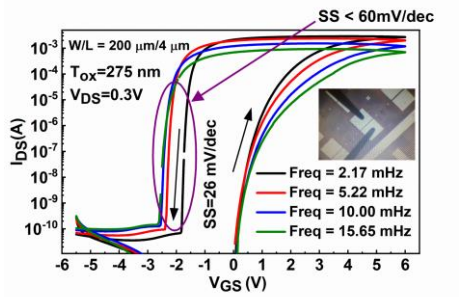

Fig. 1. I_D-V_G characteristics of the SEFET as a function of scan rate (Inset shows fabricated device).

Our methodology consists of using experimental data measured from a bottom gated TFT with an ITO gate, Ta_2O_5 gate insulator (120nm-350 nm) and 40 nm ZnO as channel deposited via Radio Frequency sputtering. The devices are subjected to thermal annealing at 80^0 C for 24 hrs. The measurements from either a single or up to 3 unconnected devices make up the delay system reservoir and are used to train the readout network[6].

We refer to the TFT as a Solid Electrolyte FET (SEFET), because of its inherent mechanism of memory which involves motion of ionized oxygen vacancies within the gate insulator. The device I_D-V_G characteristics show a counter-

Present address: Premlal B Pillai: Department of Physics, University of Manchester. Ashwani Kumar: Focal Agent Ltd, 124 City Road, London, EC1V 2NX; Aditya Gilra: Machine Learning Group, Centrum Wiskunde & Informatica, Science Park 123, 1098 XG Amsterdam, Netherlands.
Corresponding author: m.desouza@Sheffield.ac.uk; 44-114-2225167.

clockwise hysteresis with applied gate voltage and steep switching <60 mV/dec in the reverse sweep (Fig.1).

In the forward direction, vacancies in the insulator are driven towards the channel, resulting in an internal electric field that is opposite in direction to that of the applied gate voltage. This build-up of positive charge constitutes an additional electrolytic capacitance, whose value becomes negative[7], leading to steep switching without the involvement of any filamentary process unlike that in [8] or in ferroelectric FETs [9]. Inducing an electric field opposite in direction to that applied is a necessary condition for the body factor m<1.

A key characteristic of the device mechanism lies in a

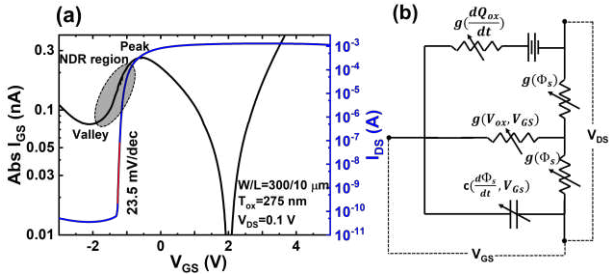

Fig. 2. (a) The gate current characteristics of the SEFET show a unique negative differential resistance. (b)The equivalent model of the gate current characteristics.

distinctive redox reaction in the insulator (Fig 2). Although we have established that the charge storage mechanism is more battery-controlled, rather than pseudo-capacitive [10], we have not yet established the origin nor the physical reaction. The unique negative differential resistance, can potentially be used to harness the device as a neuron[11].

Compared to a conventional model of a gate insulator, the equivalent circuit of an SEFET is represented by Figure 2b). It consists of an additional conductance $g(dQ_{OX}/dt)$ in series with a battery, and an additional oxide capacitance arising from the change of surface potential, $d\Psi_s/dt$ related to the migration of the ions in the insulator [12]. Both off-state

logic in memory as well as asynchronous learning of logic operations of the SEFET were demonstrated.

B. Principles of reservoir computing and delay systems.

In view of the unique volatility of the device, on similar

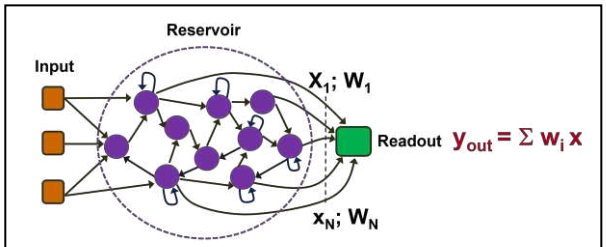

Fig. 3. A conventional reservoir consists of a number of connected nodes with fixed weights. Only the output weights are trained in the readout, easing the burden of training compared to a deep learning network.

timescales as biological applications, we have explored its application in reservoir computing. A classical reservoir consists of a recurrent network of non-linear nodes with feedback loops, connected by fixed weights, implying that the burden of training is shifted to the readout function. The transient dynamical response of the reservoir is linearly combined at the readout. Due to the projection of the low dimensional input data onto a high dimensional space, any non-linear output can be linearly decoded at the readout. The readout weights are trained by linear or logistic regression, thus learning is simpler whilst avoiding vanishing and exploding gradients in recurrent neural networks.

Three properties of the reservoir crucially influence the learning efficiency. (i) Separability: ie the non-linearity of the nodes results in a mapping of the inputs into uniquely separable outputs. (ii) Approximation: ie states which are slightly different map on to same targets (ie minimize noise) and (iii) Fading memory. This decides the length of time over which the reservoir can remember its previous state. Those that are in the far past can be forgotten whereas those nearer to the present, will affect the signal more strongly.

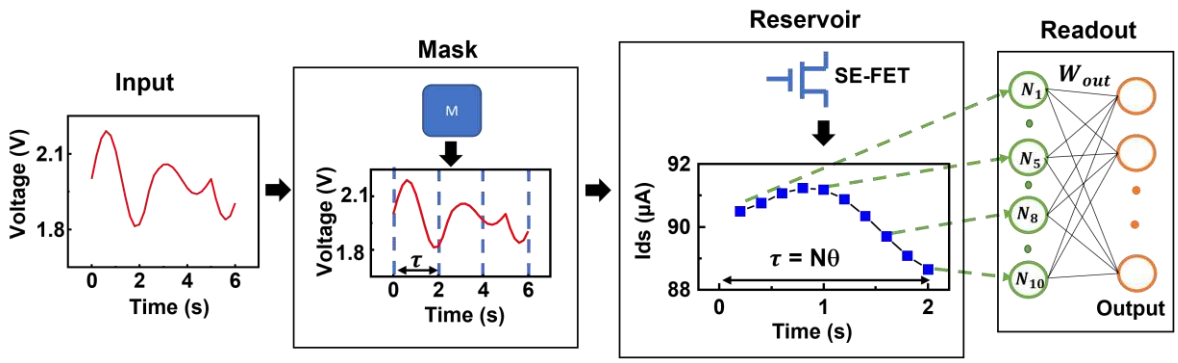

Fig.4. Framework of a dynamic SE-FET-based reservoir system. The input is transformed into a temporal signal with duration time τ via a mask and fed to the reservoir, consisting of one SE-FET device. The SE-FET response over time τ is selected as the reservoir output node (N) with a fixed time step θ. In the illustration, τ = 2 s is sampled every 200 ms thus 10 reservoir output nodes to be trained by the readout network with weights (W_{out}) using logistic regression.

Our framework of a delay system reservoir follows concepts from Appeltant et al [13], who first demonstrated that a single nonlinear node with a fading memory of timescale τ, could function as an efficient reservoir. In their conception, multiple connected spatial nodes are replaced by time multiplexing a single nonlinear node that processes the computing states sequentially. Hence, τ is divided into N sequential virtual nodes of time intervals θ that are delayed from the output hardware node. In their representation, a masking function for each virtual node is randomly varied and represents the weights of the classical reservoir.

Practical implementations of their approach and subsequent delay systems have been demonstrated with both CMOS based neurons[14], as well as memristors [15-17], when it was realized that the dynamic decay of volatile memristors is inherently an Appeltant et al delay system reservoir. In the case of a digitized input, the number of sampled bits before which the output is fed to a reservoir represents the mask length. When the time step is much larger than the decay time, the device rapidly saturates to a state that is independent of previous inputs [18]. As a result, the virtual nodes are only coupled to themselves and independent of other nodes. On the other hand, when the time interval θ is smaller than the decay time τ, the device does not have enough time to reach a saturated state. As a result, the virtual nodes can be coupled with their neighbours efficiently, leading to a functional "delay" RC system as shown by our example in (Fig 4).

In our implementation of the SEFET based reservoir, the entire input data is digitized and divided into sequences (of length 3 or 4 bits at a time). The length of the sequence represents the mask length. After each sequence, the device is reset by the application of a negative voltage of -5V on the gate to prevent saturation of the memory. The distinctiveness of our implementation is that a small read voltage is applied after *each* pulse at the drain terminal, when the gate is off, to read the drain conductance. Unlike previous works, we have shown that training the readout after every input pulse has the maximum impact to improve the learning efficiency which increases for a 4 bit sequence from 88% to 91% [6]. Off-state operation makes this approach possible when the input is digitized. However, even in the case of an analogue input, the three terminal device still offers the potential to operate continuously with its analogue output and adjustable magnitude of read voltage.

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.

III. Results

A. *Experimental characteristics of our reservoir*

There are multiple ways in which to enhance the richness of our reservoir, due to the availability of the third terminal in the SEFET: (i) pulse frequency (ii) Read voltage (including negative bias values, not possible in a conventional two terminal memristor) (iii) Gate voltage (iv) Reset voltages and last but not least reading out from multiple devices with device variability in parallel, all of which affect the learning accuracy.

Here we report the impacts of the write (Fig 5) and read

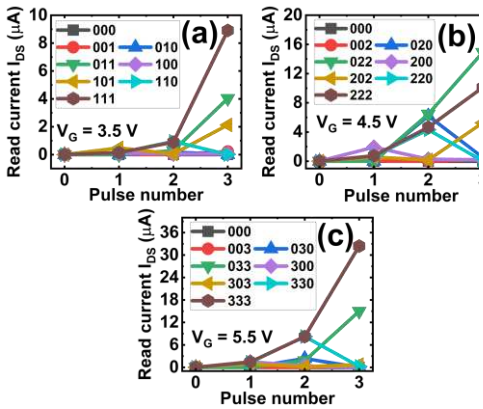

Fig. 5 (a-c). Impact of write voltage (V_G) on the read current.

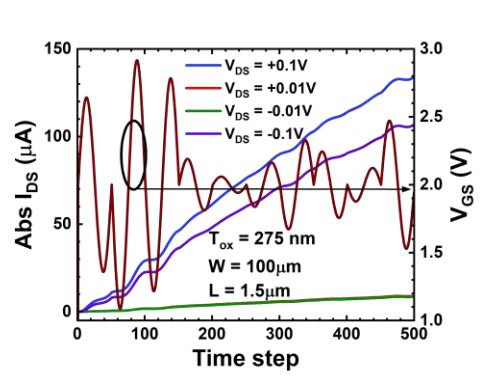

Fig. 6. Absolute drain current as a function of read voltage (V_{DS}) and gate voltage (V_{GS}).

voltages (Fig 6) on the read current. Fig 5 shows that higher gate voltages have the effect of increased separation that can be used as quaternary states to capture more information in

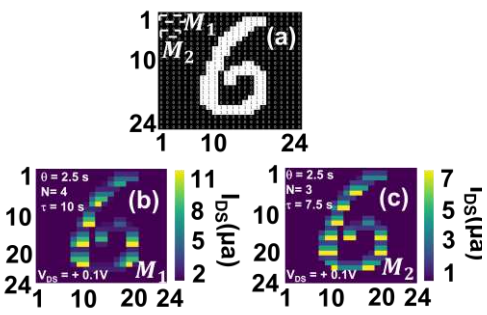

Fig. 7. (a) Pre-processed MNIST dataset image of digit 6 as an example; with image dimension of 24 × 24 pixels, showing two different masks M_1 and M_2 digitized in 4-bits and 3-bits respectively. (b-c) The complete recorded reservoir output response of digit 6 using masks M_1 and M_2.

the image, a subset of which is shown in the figure. Fig 6 shows that both positive and negative read currents are able to distinguish the outputs, with a better performance at a negative read voltage. This could be because of the nature of the built-up interfacial charge that constitutes the memory of the device. We show that read voltages as low as 10 mV are still capable of memory, though 100 mV shows a much better separability of states. Fig 6 also shows the output current of the device as a result of an arbitrary input analogue waveform.

B. *Performance of the Reservoir.*

We have explored both image and voice recognition[19] as examples of applications of the SEFET reservoir. Here we

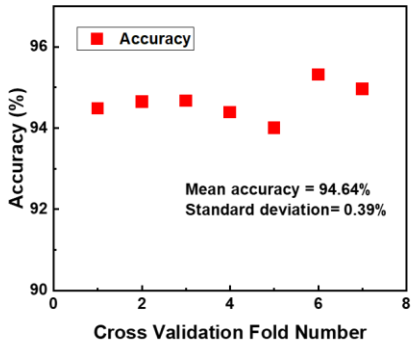

Fig 8. Accuracy for each fold in a 7-fold cross-validation test.

highlight the results with image processing. The SEFET response of the digitized image "6", over τ =10 and 7.5 seconds was sampled after each $\theta = 2.5$ seconds, resulting in N= 4 and 3 reservoir output nodes respectively per mask. The overall mean recognition of 90. 63% using mask M_1 and 91.54 % using mask M_2 is achieved across 7-fold cross-validation of the test set. This result shows that with shorter time duration of τ =7.5s the SEFET can capture the temporal information better resulting into higher accuracy.

Further improvement in accuracy is achieved by combining the complete recorded output response at three read voltages scanned both horizontally (row-wise) and vertically (column-wise). The increase of accuracy is due to local spatial correlations of the 2D image which we exploit by feeding sequentially in two orthogonal directions. With this representation of sample data, a maximum mean accuracy of 94.64% is demonstrated as shown in Fig8. The impact of the various input parameters in the optimisation of the learning accuracy is seen in Table 1. It is seen that varying the gate and drain voltages has less of an impact than the inherent device variability. An overall improvement of 3.5% is delivered by the reservoir leading to an overall efficiency of 94.64%, the highest reported to date, for a comparable task of image recognition using memristors as the memory element. We expect some of the limitations arise from digitisation of the signal.

C. *Benchmark performance Tables*

In comparing the performance of the SEFET with others reported in the literature in Table 2, it is observed that the accuracy of the voice recognition task is limited by the input dataset that only consists of 500 audio samples[20], in comparison to the MNIST database consisting of 60000 images samples. This means that learning accuracy can be properly tested as a function of the input variables, whereas in audio applications, learning efficiencies reported are > 99% [19,22,24,25]. We demonstrate the highest learning efficiency with our SEFET-based reservoir, and attribute it to a range of factors both at the device as well as system level implementation.

Approach	Mean Accuracy	Standard Deviation	Sequence Length
Benchmark without any reservoir (Input trained by logistic regression)	90.82%	0.61%	4-bit
1 SE-FET with input horizontally scanned	91.19%	0.54%	3-bit
3 SE-FETs with 3 gate voltages, input horizontally scanned	92.44%	0.57%	3-bit
3 SEFETs with same gate voltage, input horizontally scanned	92.97%	0.52%	3-bit
3 SE-FETs with 3 gate voltages, input horizontally + vertically scanned.	94.26%	0.46%	3-bit
3 SE-FETs with the same gate voltage; input horizontally + vertically scanned	94.44%	0.42%	3-bit
3 SE-FETs with the same gate voltage but 3 read voltages; input horizontally + vertically scanned (this work)	94.64%	0.39%	3-bit

Ref.	Device	MNIST Accuracy	NIST TI46 Accuracy
[21-22]	Memristor (WO_x)	88.1%	99.2%
[23]	Memristor (SiO2:Ag)	83%	-
[24]	Memristor ($TiOx/TaO_y$)	-	99.6%
[25]	Spintronic	-	99.6%
[26]	Photonic	~97%	-
[6, 19]	SE-FET (ZnO/Ta_2O_5)	94.44%	99.4%

IV. Conclusion

A careful study of the impact of various factors affecting the learning accuracy of a delay-system reservoir is presented. The richness of the reservoir can be further enhanced by using analogue input rather than digitization,

which results in loss of information.

ACKNOWLEDGMENT

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

REFERENCES

[1] M. Peyvand et al, "Dendritic Computation through Exploiting Resistive Memory as both Delays and Weights," arXiv preprint, arXiv:2305.06941, 2023.

[2] Yoon Ho Jang et al, "Time varying data processing with non-volatile memristor based temporal kernel", Nature Communications, vol. 12, pp. 5727, 2021.

[3] P.B. Pillai et al, "Nanoionics Based three terminal synaptic device using Zinc Oxide", ACS Applied Materials and Interfaces, vol. 9, pp 1609, 2017.

[4] Xiaoyao Song et al, "Reservoir Computing based on a Solid Electrolyte ZnO TFT : An attractive platform for flexible edge computing", Proc of IFETC 2023 (To be published).

[5] Xiaoyao Song et al "An ultra-low power 3-terminal memory device with write capability in the off-state", Proc. Of IEEE EDTM, 2019.

[6] A. Gaurav et al "Reservoir computing for temporal data classification using a dynamic Solid Electrolyte ZnO thin film transistor", Frontiers in Electronics, vol. 3, pp. 1-9, 2022.

[7] A. Kumar et al "Negative capacitance beyond ferroelectric switches", ACS materials and interfaces vol. 10, pp 19812, 2018.

[8] N. Shukla et al, "A Steep-Slope Transistor Based on Abrupt Electronic Phase Transition." Nature Communications, vol. 6, pp. 7812, 2015.

[9] S. Salahuddin et al, "Use of Negative Capacitance to Provide Voltage Amplification for Low Power Nanoscale Devices". Nano Letters, vol. 8, pp. 405, 2008.

[10] P. B. Pillai, et al, "Diffusion controlled Faradaic charge storage in high performance solid electrolyte gated Zinc Oxide thin film transistors" ACS Appl. Mater. Interfaces, vol. 10, pp. 9782, 2018.

[11] Wei Yi et al, "Biologically plausibility and stochasticity in scalable VO_2 active memristor neurons", Nature Communications vol. 9, pp. 4661, 2018.

[12] X. Song et al, "A Ta_2O_5/ZnO synaptic transistor for supervised learning in a crossbar", Proc. EDTM, 2021.

[13] L. Appeltant et al "Information processing using a single dynamical node as complex system", Nature Communications, vol. 2, pp. 468, 2011.

[14] Stefan Gerber et al "Neuromorphic implementation of ECG anomaly detection using delay chains". Proc. IEEE Biomedical and circuits conference, 2022.

[15] C. Du et al, "Reservoir computing using dynamic memristors for temporal information processing," Nature Communications, vol. 8, no. 1, pp. 1, 2017.

[16] J. Moon et al., "Temporal data classification and forecasting using a memristor-based reservoir computing system," Nature Electronics, vol. 2, no. 10, pp. 480, 2019.

[17] R. Midya et al., "Reservoir Computing Using Diffusive Memristors", Advanced Intelligent Systems, vol. 1, p. 1900084, 2019.

[18] Y. Zhong et al, "Dynamic memristor-based reservoir computing for high-efficiency temporal signal processing," Nature Communications, vol. 12, pp. 1, 2021.

[19] A Gaurav et al, "A Solid Electrolyte ZnO Thin Film Transistor for classification of spoken digits using Reservoir Computing", Proc of EDTM 2023.

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

[21] J. Torrejon et al., "Neuromorphic computing with nanoscale spintronic oscillators," Nature, vol. 547, pp. 428, 2017.

[22] R. M. Nguimdo et al, "Impact of optical coherence on the performance of large-scale spatiotemporal photonic reservoir computing systems," Optics Express, vol. 28, pp. 27989, 2020.