UNIVERSITY OF SOUTHAMPTON

Neural signal processing with metal-oxide RRAM devices

by

Isha Gupta

A thesis submitted in partial fulfillment for the
degree of Doctor of Philosophy

in the
Faculty of Physical Sciences Engineering
School of Electronics and Computer Science

September 2018

Academic Thesis: Declaration of Authorship

I, Isha Gupta declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

Neural Signal Processing with metal-oxide RRAM devices

I confirm that:

This work was done wholly or mainly while in candidature for a research degree at this University;

Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;

Where I have consulted the published work of others, this is always clearly attributed;

Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;

I have acknowledged all main sources of help;

Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;

Parts of this work have been published as:

Gupta, I., Serb, A., Berdan, R., Khiat, A., Regoutz, A. and Prodromakis, T., 2015. 'A cell classifier for RRAM process development'. *IEEE Transactions on Circuits and Systems II: Express Briefs, 62(7)*, pp.676-680.

Gupta, I., Serb, A., Khiat, A., Zeitler, R., Vassanelli, S. and Prodromakis, T., 2016. 'Real-time encoding and compression of neuronal spikes by metal-oxide memristors'. *Nature communications, 7*.

Gupta, I., Serb, A., Khiat, A. and Prodromakis, T., 2016, May. 'Practical operation considerations for memristive integrating sensors'. *IEEE International Symposium on Circuits and Systems (ISCAS), 2016* (pp. 2322-2325).

Gupta, I., Serb, A., Khiat, A. and Prodromakis, T., 2017. 'Improving Detection Accuracy of Memristor-Based Bio-Signal Sensing Platform'. *IEEE Transactions on Biomedical Circuits and Systems, 11*(1), pp.203-211.

Gupta, I., Serb, A., Khiat, A. and Prodromakis, T., 2016. 'Towards a memristor-based spike-sorting platform'. *IEEE Biomedical Circuits and Systems*, pp.1-4.

Gupta, I., Serb, A., Berdan, R., Khiat, A. and Prodromakis, T., 2017. 'Volatility Characterization for RRAM Devices'. *IEEE Electron Device Letters, 38(1)*, pp.28-31.

Gupta, I., Serb, A., Khiat, A., Zeitler, R., Vassanelli, S. and Prodromakis, T., 2018. 'Sub 100nW volatile nano-metal-oxide memristor as synaptic-like encoder of neuronal spikes.' *IEEE Transactions on Biomedical Circuits and Systems.*

Gupta, I., Serb, A., Khiat, A., and Prodromakis, T., 'Mitigating noise effects in volatile nano-metal oxide devices, *IEEE International Symposium on Circuits and Systems (ISCAS), 2017.*

Gupta, I., Serb, A., Khiat, A., and Prodromakis, T., 'Spike Sorting using non-volatile metal-oxide devices, *Nature Scientific Reports, Under Review.*

Trapatseli, M., Carta, D., Regoutz, A., Khiat, A., Serb, A., Gupta, I. and Prodromakis, T., 2015. 'Conductive atomic force microscopy investigation of switching thresholds in titanium dioxide thin films'. *The Journal of Physical Chemistry C, 119(21)*, pp.11958-11964.

Regoutz, A., Gupta, I., Serb, A., Khiat, A., Borgatti, F., Lee, T.L., Schlueter, C., Torelli, P., Gobaut, B., Light, M. and Carta, D., 2016. 'Role and Optimization of the Active Oxide Layer in TiO_2 based RRAM'. *Advanced Functional Materials*, 26(4), pp.507-513.

Carta, D., Hitchcock, A.P., Guttmann, P., Regoutz, A., Khiat, A., Serb, A., Gupta, I. and Prodromakis, T., 2016. 'Spatially resolved TiO_x phases in switched RRAM devices using soft X-ray spectromicroscopy'. *Scientific reports, 6.*

Carta, D., Guttmann, P., Regoutz, A., Khiat, A., Serb, A., Gupta, I., Mehonic, A., Buckwell, M., Hudziak, S., Kenyon, A.J. and Prodromakis, T., 2016. 'X-ray spectromicroscopy investigation of soft and hard breakdown in RRAM devices'. *Nanotechnology, 27(34)*, p.345705.

Messaris, I., Nikolaidis, S., Serb, A., Stathopoulos, S., Gupta, I., Khiat, A., & Prodromakis, T. (2017, May), *A TiO_2 ReRAM parameter extraction method*, 2017 IEEE International Symposium on circuits and Systems (ISCAS) (pp. 1-4).

Messaris, I., Nikolaidis, S., Serb, A., Stathopoulos, S., Gupta, I., Khiat, A., & Prodromakis, T. (2017, May), *Live demonstartion: A TiO_2 ReRAM parameter extraction method*, 2017 IEEE International Symposium on circuits and Systems (ISCAS) (pp. 1-4).

Signature: Isha Gupta (March 2018)

Copyright Declaration

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF PHYSICAL SCIENCES AND ENGINEERING
SCHOOL OF ELECTRONICS AND COMPUTER SCIENCE

Doctor of Philosophy

by Isha Gupta

This thesis delves in the issue of processing huge volume of neural data recorded using *in-vitro* monitoring technologies. The real-time processing of neuronal signals imposes excessive strain on bandwidth, energy and computation capacity prohibiting scaling of present neural interfaces. This work offers a unique solution to the challenge of encoding electrophysiological neural bio-signal information in a power efficient way thus capable of impacting the field of neuroprosthetic applications or for instance, the emerging area of bio-electronic medicines.

The thesis mainly focussed on exploiting the intrinsic properties of nanoscale metal-oxide-metal devices commonly known as 'memristors' for demonstrating **spike detection** and **spike sorting** at the proof-of-concept level. Memristive devices were fed with extracellular neuronal activity and the thresholded integrating property of both **non-volatile** and **volatile** devices was used to differentiate between high-amplitude spiking events and low-amplitude background noise that forms the majority of the neuronal signals samples. The spiking events of supra-threshold strength are detected as memory state transitions thus compressing information on neuronal spikes in real-time. These experiments show a substantial improvement in the bandwidth required per sensing site (≈200:1), while concurrently offering more energy efficient paradigm, estimated at ≈100nW per channel as compared to the present state-of-the-art spike detection techniques. For all the experiments, for quantification of the obtained results, the spike detection performance was benchmarked against state-of-the-art template matching.

Importantly, the experimental work carried for demonstrating this application also involved developing electrical characterisation modules for carrying out en-masse testing of the fabricated devices and ensure process development. Furthermore, the same concept was used to demonstrate a much more computationally intensive task i.e. **'spike-sorting'**. It is a procedure of identifying the activity of individual neurons from the data collected through electrophysiological experiments. The experiments performed show how the intrinsic analogue programmability of memristive devices can be used to perform the task of spike sorting. This idea can thus potentially open new avenues for performing spike-detection and sorting both on-chip using miniaturised chips at minimal power costs, demonstrating the technology's potential to build scalable, yet energy efficient on-node processors for brain-chip interfaces.

Contents

List of Figures

List of Tables

Acknowledgements

THANK YOU FOR MAKING ME MORE THAN I AM

Coming together is a beginning. Keeping together is progress. Working together is success. While my name may be the only one you read on the front page of this thesis but by no means any work is solely an individual's work. I am forever indebted to a number of people who came together to walk with me through my transforming journey of PhD. First and foremost, I will like to duely acknowledge my extremely committed supervisors: *Professor Themis Prodromakis* and co-supervisors *Alexantrou Serb* and *Ali Khiat*. Thank you for offering me this wonderful opportunity. Thank you for believing in me. Your encouragement, patience and time has changed a great deal in me. I am sure I will have your continued advice and support just like family. I will also like to acknowledge *Professor Koushik Maharatna*, who was one my internal examiners. Thank you bringing valuable insight to my work.

Next, I will like to thank my colleagues and Nano-research group at Southampton. More specifically: *Radu Berdan*, your research made my research work easier, thank you for inventing a masterpiece. *Maria Trapatseli*, you definitely are a sister from another mother. Your endless supply of home-cooked food, delcious desserts, late night work rounds and conversations kept both of us going. *Arifa Ahmed*, thank you for the much needed shoulder and support. *Nikoalas Vasilakis*, without your teasing my days wouldn't be half the fun as they were. Keep the devilish spirit alive, it does kick some hearts. *Sumit Kalsi*, you know you mean a lot to me. Thank you being there. *Tasmiat Rahman*, thank you for your wise advices at important times, *Anushka Gangnayak* you have been a life saver many times, *Sally Chi and Sam Fishlock*, thank you for being great friends, *Simone Cortese*, obviously and of course, *Dear Angela and Laila*, thank you for being there! In the same vein, I would like to thank *Anna Regoutz* at *Imperial College London*, *Daniela Carta* at *University of Surrey, UK* and *Ioannis Messaris* for their support and commitment towards research work. You people made Southampton my home! I will also like to add due acknowledgements to the outreach team including: *Loukas Michalas, Jamie Reynolds, Stevan Stankovic, Daniel, Marilena, Rafida and many more (in advance apologies for not getting through all the names).*

I thank our collaborators on the RAMP project, *Prof. Stefano Vassenalli* at *University of Padova, Italy* and *Dr. Ralf Zeitler* at *Max Planck Institute, Germany* who extended all their resources for experiments during the course of this thesis. I admire your patience during two digit rounds of revision of manuscripts. This piece of research looks very different because of your input, feedback and expert knowledge.

Now, I will like to deeply acknowledge my family. *Family is not an important thing, it's EVERYTHING.* My two biggest source of inspiration, my paternal and maternal grandfather's,

Inder Prakash Gupta and *Dr. Chander Bhan Gupta*. Thank you for setting the bar way up high. I can only try and accomplish as much knowledge, fame and success as you people have in your respective lives. Thank you to the best momma and daddu, thank you for incredible, respectable, appreciable and impressive amount of understanding, support and encouragement. Your presence makes me feel blessed everyday. What me and my sister, Niki (*Dr. Bhavya Gupta*) are today is because of you both. Thank you for making us who we are. You told me when I was thirteen, an educated person can begin his/her life at any point in life and there has never been a bigger lesson. You said it all in few words, in one simple line.

Thank you my homie, *Varun Goel*, my brother, my best friend, my confidant, someone who knows me the best. You are my treasure and I love you so much. *Mr. Darshan Jindal*, *Mrs. Sheetal Jindal*, thank you for being a great elder brother and solving all my problems. I can sleep in peace because I know you guys are there. In one line, I simply want to thank my cousins (*Sonal Bansal, Animesh Bansal, Mohit Jindal, Monika Jindal, Uma Singhal, Vidushi Singhal, Yashovardhan Singhal, Surbhi Garg, Radhika Garg, Divya Garg, Sarthak Garg, Preeti Garg, Hiteshwar Jindal, Swati Jindal, Sakshi Jindal, Rohit Jindal, Khusi, Harshita, Jhanvi, Devash, Jai, Darsh*), elders, my masi's, my dadi, late nani, our family guru at Bhuma Niketan Ashram and the rest of the family. Its an honour to have people like you in my life and I miss you a lot!

Furthermore, I will dearly like to thank my beloved and patient friends: *Sahil Sankhla, Surbhi Jain, Anika Chakerwarti, Vanshika Bansal, Kamal Goel, Pushkar Arora, Saakshi Agarwal, Dhaval Vithlani, Rashmi Munot, Kunal Jain, Manini Gunjal, Rohit Nayak, Shreyans Daga, Subiksha Shekhar, Shrushti Prakash, Satyajit Das, Nayan Gupta, Medha Sethi and Nishant Agarwal*. You people have always been a call away for any help I have ever needed. Thanks for knowing exactly when to tell me what I want to hear and when I want to hear it the most. From selfies to parties, from heartbreak to loneliness, from roadtrips to skydiving and discovering the world beyond the borders, thanks for being there for me through it all. xoxo. I simply don't have words. I wish you all the best in your respective lives and you all are in my prayers. Crazy times and crazier yet to come...!!!

Lastly, I express my gratitude towards financial support of European Union (RAMP Project) and Innovation to Commercialisation of University Research (ICURe), especially *Alan Scarse and David Woolley* for providing me an enriching experience. I would also like to thank the fantastic team at *Galvani Bioelectronics/GSK, Stevenage* (especially *Tariqus Salam, Victor Pikov, Justin Perkins, Jason Miranda, Daniel Chew, Refet Firat, Jason Witherington, Cynthia Orme, Lynsay Evans, Clare de Vere, Matteo Donega and Kristoffer Famm*), for providing me an opportunity to do an internship and providing practical insight in the field of implantable devices. I look forward to gaining more experience in the coming year.

Knowledge is wealth in foreign land,
Intelligence is wealth in tough times,

Righteousness is wealth in other world,
Verily, good character is wealth everywhere and at all the times.

Love, Respect, Peace and Happiness,

Ms. Isha Gupta (2018)

You are what your deep, driving desire is
As your desire is, so is your will
As your will is, so is your deed
As your deed is, so is your destiny.
(Brihadaranyakopanishat 4.4.5)

To my ancestors, beloved family and friends....

Chapter 1

Introduction

1.1 Motivation

Reverse engineering the human brain and decoding the underlying information processes of biological systems requires integrated efforts from researchers with different scientific background [1]. Towards enabling this vision, advances in neural recording techniques [2–5] target the reliable acquisition of electrophysiological data from multiple neurons *in-vivo* and *in-vitro*. This has impacted our understanding of information processing by brain microcircuits [6] and brought new prospects for novel therapies based on adaptive neural stimulation [7]. To date, state-of-the-art implementations can record in-vivo [8] from up to thousand sites and from up to 30k [9] sites in-vitro using Complementary Metal Oxide Semiconductor (CMOS) based High Density Microelectrode Arrays (HDMEAs). Such advances in micro-sensors have been paralleled by considerable progress in neural processing microsystems [10, 11], which are capable of detecting neural spiking activity on-node [12, 13]. The relevant spike-detected information i.e. significant information differentiated from the background noise is then transmitted off-line wirelessly and techniques such as the Template Matching System (TMS) or Principle Component Analysis (PCA) [14] are then used off-line for 'spike-sorting' [15].

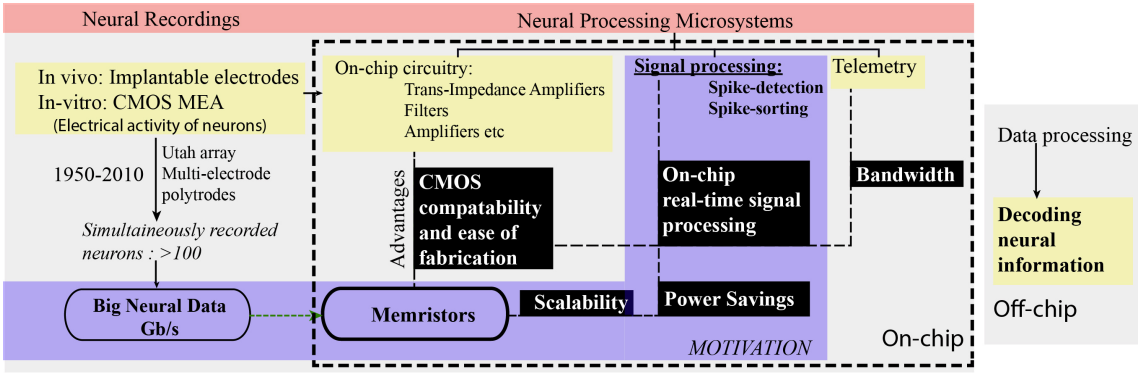


Figure 1.1: Motivation of this research thesis. The block in purple colour illustrates the scope of the experimental work.

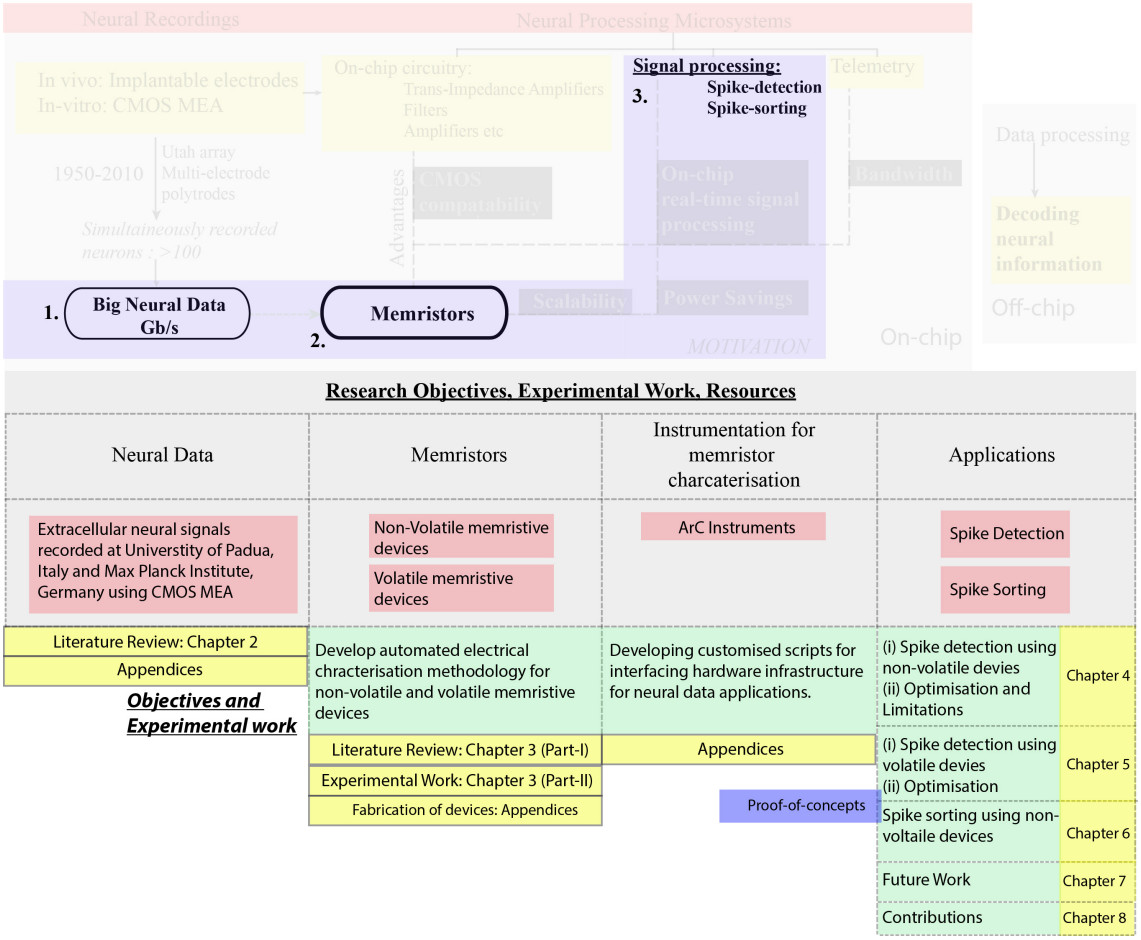


Figure 1.2: Research Objectives, Experimental work, Resources and Thesis Organisation

These methods by mapping the recorded electrical activity to the source neurons, offer insights in neural coding principles [16] and support novel neuroprosthetic applications [17–19]. However, further advances in the fast developing field of implantable neural interfaces [20] are hampered by key bottlenecks in the processing of neuronal spikes including: **a) computational power required to process the ever increasing volume of neural signals (Gb/s range presently) on-node and in real-time [21–24], b) bandwidth [25] and, c) scalability & deployment complexity**. Figure. 1.1 illustrates the motivation of this work using a block diagram.

To give a better perspective, at system architecture level of a neural interface, it comprises of, (a) an analogue front end comprising of a low noise amplifier, a high pass/low pass filter such as butterworth filter, an analogue-to-digital converter, and (b) back-end signal processing consisting of spike-detection and sorting circuits [26]. The power/area requirements for the front-end system is variable and depends upon a number of parameters such as gain, bandwidth, effective number of bits etc. For instance in [27], a 400 μm pitch neural spike recording channel, fabricated in a 130 nm standard CMOS technology, the front-end would consume 2.8 μW which goes up to 3.1 μW when the signal processing features are enabled. The area footprint is of the order of 1.6 mm^2 per channel. This thesis tackles the back-end

signal processing aspect, and aims at reducing the power/area requirements by the medium of memristive devices at a proof-of-concept level. For instance, the experimented memristive devices in this thesis are in the order of 200x200 nm^2 with the power dissipation of less than 100 nW (excluding full circuit implementations). Therefore, this solution in principle would offer several orders of magnitude of area/power savings over current state-of-the-art neural electronic systems.

Thus, memristive devices appear to be well-suited in providing a disruptive technological boost to this vision by performing the role of artificial synapses. Much akin to biological synapses, they possess the intrinsic ability to simultaneously carry out computational tasks and store information at aggressively downscaled volumes and power dissipation [28–30]. Intrinsic characteristics of TiOx memristors (used throughout this research work), such as input thresholding and analogue modulation of resistive state can be used in order to encode neuronal spiking activity while noise is inherently suppressed. A large part of the computational burden associated with spike processing can thus be relegated to highly scalable nanodevices that can be manufactured in the back-end-of-line part of Complementary Metal-Oxide Silicon (CMOS) electronics fabrication processes, thus costing potentially no chip layout area, and potentially lower power consumption. Such developments can help in paving the way towards advanced neuroprostheses or applications such as bioelectronic medicines where power dissipation remains as the major challenge.

1.2 Research Objectives

In this thesis, the author had access to (see Figure. 1.2):

a) fabricated solid-state memrisitve devices of many different stack structure's and materials, both packaged and directly on wafer fabricated in the cleanroom facilities established at University of Southampton[1],

b) commercial hardware infrastructure i.e. ArC Instruments used for automated electrical characterisation of the fabricated devices[2],

c) extracellular neural signals recorded from rat retinal ganglion cells from University of Padua, Italy[3] and Max Planck Institute, Germany[4].

Firstly, a new spike-detection approach based on metal-oxide resistive switching memory devices, also known as memristors was investigated. Fundamentally, memristive devices undergo non-volatile resitive state transitions as a function of the integral of the input voltage,

[1]Many thanks to Dr. Ali Khiat and Dr. Maria Trapatseli, University of Southampton.

[2]Many thanks to ArC Instruments Ltd for providing the instrument and Dr. Radu Berdan and Dr. Alexantrou Serb for helpful discussions to use this capability for testing of the devices.

[3]Many thanks to Prof. Stefano Vassenelli, University of Padua.

[4]Many thanks to Dr. Ralf Zeitler, Max Planck Institute/ Venneos Systems (www.venneos.com).

thus behaving as thresholded input integrators. Taking advantage of this property, TiO_x-based memrisitve devices were employed for spike-detection, as extracellular neural spikes recorded from retinal ganglion cells were encoded in gradual, non-volatile resistive state transitions, whereas the sub-threshold events (i.e. noise) were naturally filtered-off. This property made these devices suitable as noise-suppressing integrating sensors and are thus termed as Memristive Integrating Sensors (MIS). It should be noted that due to low yield of the fabricated wafers at the beginning of the project, it was essential to spend a large amount of time in devising a method for automatically electrically characterising the devices. Therefore, **secondly**, algorithms to characterise the non-volatile behaviour of devices were developed simultaneously. **Thirdly**, non-volatility of the devices was found to be limiting the spike detection performance due to saturation of the resistive state of the devices, therefore, the performance of the devices was optimised.

Fourthly, the approach of spike detection using non-volatile behaviour of the devices was advanced by exploiting an often overlooked crucial property of memristive devices that is 'volatility'. This approach recalls the way of operation of biological synapses that translate spiking frequency in gradual changes of postsynaptic conductance subject to a continuous self-resetting process. When used in the volatile operating region, memristive devices exhibit metastable memory state transitions following which they inherently relax to their initial resistive state range. It is demonstrated that volatility enables for naturally encoding spiking events into transient resistive changes. **Finally**, the experiments are performed to demonstrate the principle of 'spike-sorting' at proof-of-concept level.

1.3 Thesis Organisation

The organisation of the thesis is as follows:

The project involves in-depth characterisation of RRAM devices for spike detection and spike sorting in response to neural recordings obtained from extracellular environment. **Chapter. 2** presents a brief overview of the field of neural signal acquisition and processing.

Chapter. 3, Part I surveys the literature on RRAM/memristive devices and demonstrates the electrical characterisation of the employed devices. **Chapter. 3, Part II** details the automated algorithms developed for electrical characterisation of both non-volatile and volatile devices.

Chapter. 4 and **Chapter. 5** illustrates the application of spike detection using non-volatile and volatile devices respectively. In both the cases, the results have been benchmarked against a state-of-the art template matching system and the results have been optimised to improve the sensitivity of devices.

Chapter. 6 demonstrates the application of spike-sorting using non-volatile single devices at proof-of-concept level.

Chapter. 7 concludes the thesis and discusses the future work.

Chapter 2

Neural Signals: Acquisition and Processing

Neural recording and stimulation technologies have enabled scientists to design systems such as neuroprosthetics and brain-machine interfaces to understand, repair, enhance, or otherwise exploit properties of neural systems. This chapter sets the framework for the urgent need for devising intelligent solutions for handling the ever-increasing big neural data and provides a brief overview about neural signals and recordings, advances in micro-electrode technology, neural signal processing and the growth of neural data over the years.

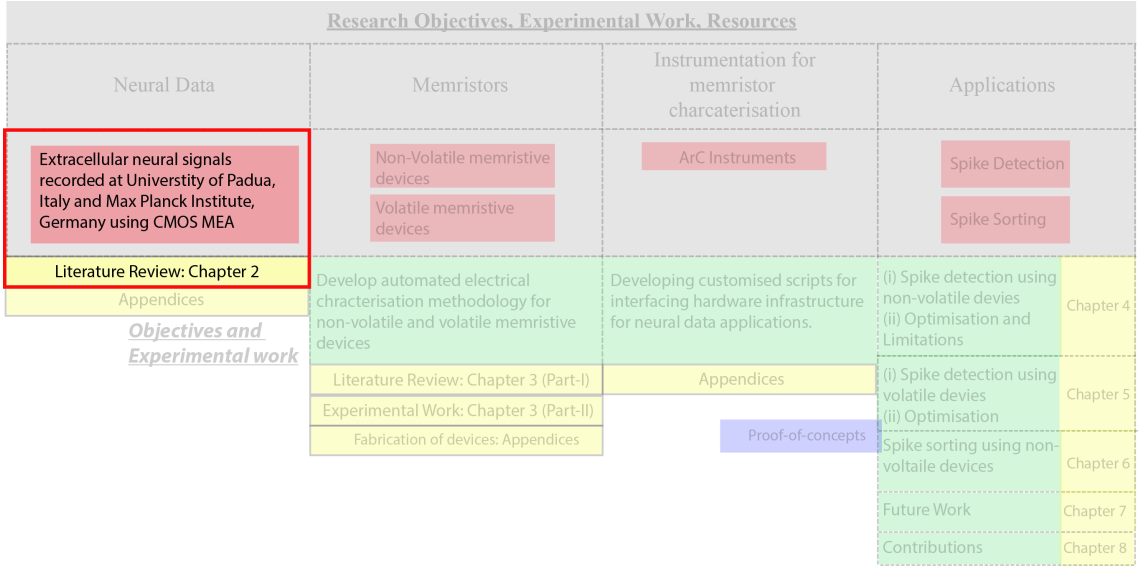


Figure 2.1: The research objective of this chapter.

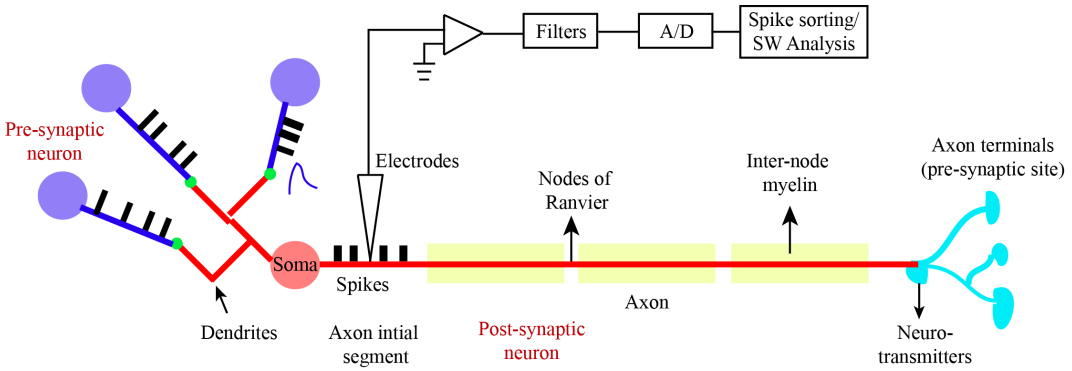


Figure 2.2: A neuron comprises of dendrites and axon branching out of soma. Axon is covered with a lipid myelin sheath which isolates it from the surrounding neurons and intermittently consists of small-gaps termed as 'Nodes of Ranvier'. The initial segment of the axon and uncovered small-gaps consists of ion-channels which helps in propagating the action potentials along the axon [35]. Axon and dendrites communicate via synapses (green). Neural activity can be acquired by placing microelectode in the region of interest.

2.1 Biology, Neural Signals and Recordings

2.1.1 Action potentials : source, nature and importance

Historically, the word 'neuron' was coined in 1891 [31]. Luigi Galvani in 1791 was the first person to develop a link between neural communication and electrical signal, where he showed contraction in a frog muscle by applying electric current [32]. Nervous system is built of separate elements i.e. dendrites and axons, which are receptive (input) devices and the sending (output) devices respectively, as shown in Fig. 2.2. These ingredients are physically disconnected (as can be seen under electron microscope) and communicate via synapses [33]. Action potentials (spikes) are fired from the axon branching out of soma leading to transmission of electrophysiological signals (electrical phenomena associated with nervous and other bodily activity). Action potentials usually last 2 ms, however, the rate of neuronal spikes is highly variable [34].

For centuries, scientists have been using electrophysiology to study the electrical properties of biological cells and tissues. Electrophysiological recordings can be made from within the cells (intracellular neural recordings) or from outside the cells (extracellular neural recordings). Neural activity in the brain gives rise to transmembrane currents that can be measured in the extracellular medium. Small diameter electrodes are placed in the brain tissue (or region of interest) to acquire action potentials in extracellular medium. A potential difference is applied between the ground and the tip of the microelectrode. The electrical changes measured at the tip of the electrode is the measure of the current, which flows in the extracellular medium. The major component of the acquired signal is generated by the action potential, however, this also includes current due to dendrites of surrounding neurons which are typically of low bandwidth. By bandpassing and filtering these signals, cellular action potentials of few neurons close enough to the electrode can be obtained [13]. For instance,

'spike trains' are high pass filtered (>300 Hz) extracellular action potentials of neuron and 'local-field potentials' measures low pass filtered (<300Hz) extracellular continuous signals in the vicinity of the recording electrode tip. Usually, these signals are digitised and then processed off-line. Since these signals are usually compounded from more than one neuron, it is essential to associate every signature of signal to a specific neuron. In other words, it is essential to distinguish single-unit activity from the multi-unit activity, which is also known as 'spike-sorting'.

Monitoring the activity of single neurons is invaluable to brain research at large. For instance, researchers believe that Brain-computer interfaces (BCI)[1] is a very promising technology and offers prospects for restoring motor ability in paralysed patients [36]. However, the vision is constrained from the limitations of trial-to-trial variability of brain function, the resolution and reliability of information detected, the complex electrical and spatial geometry of the brain and head [37].

2.1.2 Approaches to monitor the electrical activity of cells

To elucidate the dynamics and nature of neuronal code and neuronal memory, it is imperative to detect and monitor the electrical activity of large number of individual neurons at high-spatial temporal resolution for an extended period of time [11]. For this purpose, a large number and high density of electrical sensors are required which can monitor large neuronal assemblies (dissociated nerve cells/slices of brain tissue) at the resolution of single cell. Progression in past years to measure the electrical activity is reflected in the following two areas:

1. **Patch-clamp**: This is an invasive type of technique, which is used for making transmembrane measurements by accessing the cell interior with one of the electrodes termed as 'patch-pipette'. Although this technique has the capability to yield very accurate information on the electrophysiological properties of the cells (single ion-channels) but it can only monitor few cells in parallel (due to elaborate mechanical setup) and long-term monitoring is restricted. This approach however is mainly limited to fundamental discoveries and not appropriate for medical interventions as it causes irreversible damage to the cell it is recorded from. The action potentials measured across a cell membrane is typically of the order of 70 mV_{pp} [38].

2. **Micro-electrodes**: This is a non-invasive recording technique, which allows long term monitoring of the neural network. These are composite signals and the signal depends

[1] A brain computer interface (BCI), is a hardware and software communication system that enables human to interact with their surroundings, without the involvement of peripheral nerves and muscles, by using control signals generated for instance from electroencephalographic (EEG) activity. EEG measures electrical brain activity caused by the flow of electric currents during synaptic excitations of the dendrites in the neuron. These signals are recorded by placing the electrodes on the scalp. however, the quality of signal is poor because the signals will have to cross the scalp, scull, and many other layers.

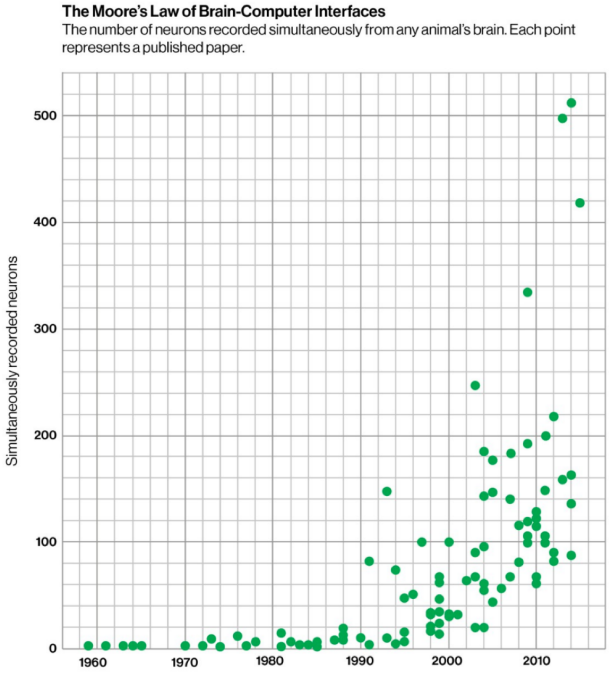


Figure 2.3: The numbers of neurons simultaneously recorded over the years.

on the distance from the electrode. Metal wire electrodes are used for extracellular recordings or more recently, the cells are cultured atop the sensors or electrodes (CMOS systems, discussed later). The amplitudes are as high as several hundred microvolts and frequency content extending up to perhaps 10 kHz [11, 39, 40].

2.1.3 Perspective: Growth of neural data

Neuroscience fundamentally aims at understanding the number of simultaneously recorded huge number of neurons, and for this it is essential to develop a perspective on how recording and analysis technique scale [41].

Simultaneously recorded neurons: There has been a exponential growth in the number of recorded neurons from 1950's [1, 22]. Statistically from the analysis from over 56 studies published between 1950-2010, the number has been doubling every 7 years [22] (see Fig. 2.3). It is important to note that this development has been accelerated by a number of innovations in parallel namely, advances in micro-electrode techniques, data transfer speeds, storage capacity, microelectronics and many more [11, 40, 42]. The pace of technological growth has been such that by the year 2010 one could record electrical activity from more than 500 neurons simultaneously. It would be right to say that such systems would have been impossible a few decades back. One of the biggest bottlenecks at the moment is the ever increasing volume of data being generated out of these systems.

Understanding massive neural data: Developing an understanding of how information about the external world is encoded in spike trains is the basis of recording 'neural data'.

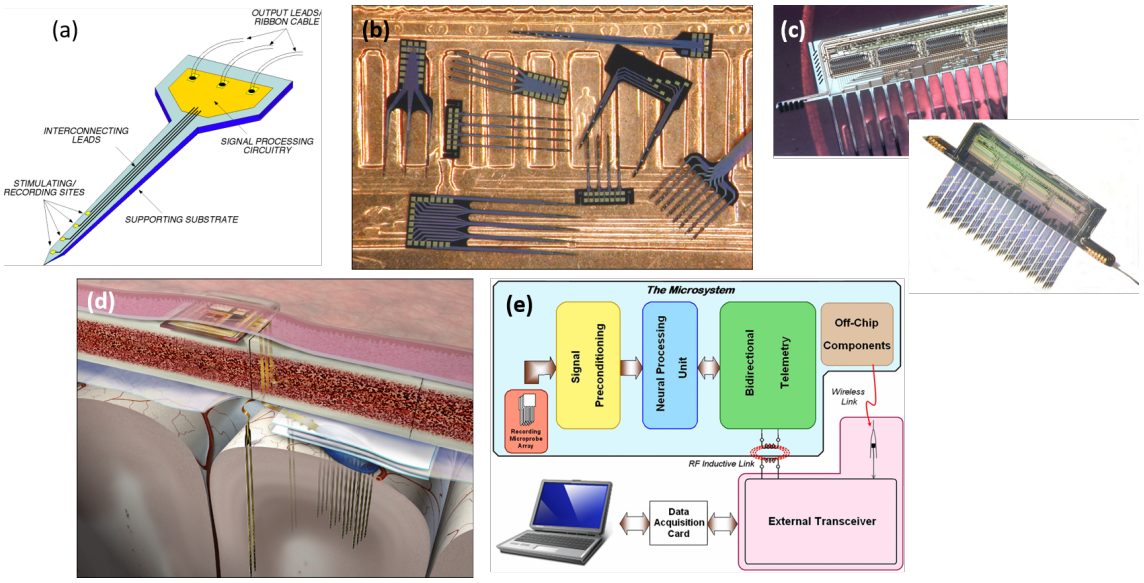


Figure 2.4: In-vivo interfacing. All figures replicated from [20]: (a) Early neural probes, b) several neural probe designs on the back of US penny, c) A 64-site eight-channel stimulating probes with sites on 400 μm centres with CMOS electronics for stimulus generation and recording, (d) and (e) A wireless implantable micro system and block diagram respectively.

The growth in the recording of the neuronal data comes with significant computational challenges. Modelling large number of neurons can certainly allow more accurate predictions of behaviour, however, the complex datasets compounded with dimensionality i.e. increasing numbers of features describing the dataset is a big impediment on computer runtime. Therefore, improving the efficiency of data analysis algorithms, finding order in rich multi-neuron data and developing hardware to accelerate them are currently active area of research.

2.2 Advances in Microelectrode technology

This section discusses the critical interface between the nervous system and microelectronics. The neural recordings can be recorded via *in-vivo* or *in-vitro* systems as discussed in the following paragraphs. It should be noted that both the fields are huge and vast in terms of research and commercialisation. It is impossible to cover an intensive review, therefore, the author tries to cover a brief summary on this field in regards to the importance of the research work in this thesis.

Interface between neurons and microelectronics (*in-vivo*):[2] the early neural probes

[2]Timeline:
(A) **Galvani** (1791): Dead frog's leg would twitch with the application of electrical stimulus.
(B) **Cajal**: Nobel Prize in Physiology (1906): His work on the anatomy of the nervous system showed that the nervous system had nerve cells or neurons and fibrillar structures in addition to blood vessels.
(C) **Edgar Adrian**: Nobel Prize in Physiology (1932): He showed it was possible to acquire electrical activity from neural tissues- a key breakthrough for neural recording technology.
(D) **Kenneth Cole and George Marmont** (1949): Voltage Clamp technique.
(E) **Hodgin and Huxley**: Nobel Prize in Physiology (1963): Explained the mechanism of action potentials

as shown in Figure. 2.4(a) were developed in Stanford University in 1960's. The structure consisted of a substrate with lithographically well defined conducting stimulating/recording sites interfaced to a signal-processing circuitry. Recording sites are usually formed of gold, aluminium or iridium oxide [20]. Over the years, a wide variety of designs for recording electrodes such as cuff electrodes, polytrodes, flexible electrodes etc [2, 43] have been explored. Biocompatible materials have been explored for use as the probe substrate such as silicon, metals (platinum, gold) etc, however, silicon substrates have remained the most widely used ones. One of the biggest reason for this is that the material is compatible with the realisation of probe structures that are very small (sub-micrometers) and on-chip circuitry for signal processing can be accommodated on the same substrate (for example, see Figure. 2.4(c)). Such recording electrodes facilitate neural signal acquisition for weeks and what prohibits the acquisition of reliable neural recording from such implantable electrodes for over months is the growth of few micrometers thick layer of glia. A number of approaches are being developed to increase the recording life such as coatings of materials that prevent protein adsorption [20]. Figure. 2.4(d) shows an example of fully implantable neural microsystem. This system in addition to micro-electrodes, needs circuitry for stimulus generation, signal amplification, spike detection and telemetry, which is also shown in detail in Figure. 2.4(e).

A list of commercially available electrodes from companies such as BlackRock Microsystems (Utah array), NeuroNexus, Cortec etc can be seen in Reference [43]. Recently there has been increasing research on the electrode designs and electrode technologies which can be integrated with the central and peripheral nervous system [44].

Interface between neurons and microelectronics (*in-vitro*): Late 1980s and 1990s saw advancement in CMOS technology, which lead to integration of electronics with the electrode structures. In addition with the capability of addressing large number of electrodes on chip, CMOS technology helps in improving the signal quality by conditioning the acquired signal on-site (with filters, amplifiers etc on-chip) and helps in sub-cellular investigation of the neurons [45], allowing for studying of entire neural network. This method can be used for studying biological preparations via two-dimensional Multi-electrode-arrays (MEA's) fabricated by an CMOS technology which helps in non-invasive, high resolution mapping of electrical activity [46]. Both kind of cell cultures can be basically studied with the CMOS systems i.e (a) **'slices'** - electrical activity is immediately recorded as soon as they are removed from the animal and (b) **'disassociated cell cultures'** - their relative in-vivo positions is no longer known.

As an example, Figure. 2.5 demonstrates the CMOS system used for acquiring the neural activity in this research work. Max Planck Institute of Biochemistry has developed a 128x128

to be 'electrical triggers' keeping the nervous system functioning.
(E) **Hubel and Wiesel**: Nobel Prize for Physiology (1981)- They used single neuron recordings from tungsten electrodes to map the visual cortex in rats. They were first to demonstrate that neural recordings essentially could be used to improve our understanding of neural structures.
(F) **Neher and Sakkmann:** Nobel Prize in Physiology (1991) - Valuable insight into the fundamental cell processes like the positive and negative ions flow in and out of the cells (nerve impulse conduction).

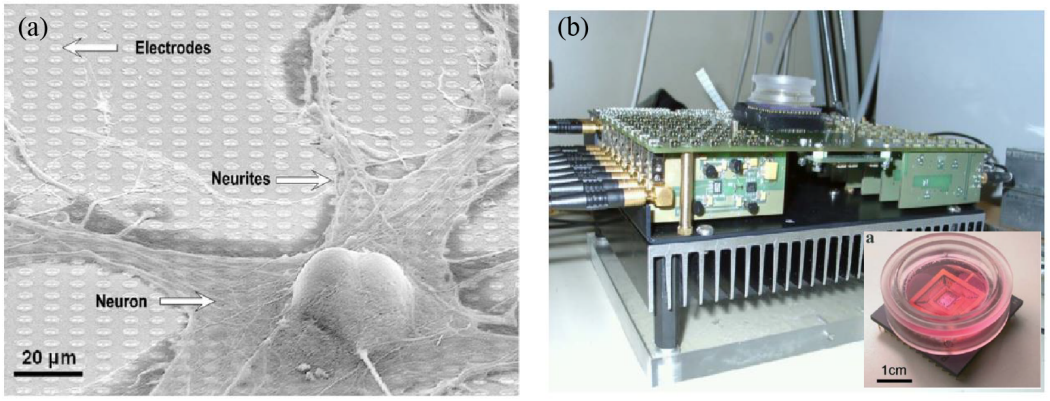


Figure 2.5: (a) Neurons on chip [47].(b) Image of measurement setup used to acquire neural recordings ([48]) and the inset shows he multi-transistor array neurochip fabricated on the basis of CMOS [39].

CMOS field-effect-based system for recording of neural activity [11] [49]. The chip size is 5.4 mm x 6.5 mm which comprises a sensor area of 1 mm x 1 mm on which 16,384 sensors are fabricated at a pitch of approximately 7.4 μm. The power consumption of the sensor chip is found to be 656 mW and all sensor spots are recorded with a frequency of 2 kHz. The signals detected are of the order of 100 μV to 5 mV peak-to-peak.

For this project (also detailed in Appendix. C) only *in-vitro* system was used, neural activity from the portions of mid-peripheral rabbit retina was recorded using MEA fabricated in CMOS technology as described above. The surface of CMOS multi-transistor array comprising of 128x128 sensor sites is insulated by thin, inert TiO_2/ZrO_2 layer. The oxide is connected to the gate of the field-effect transistor via metallic pathway. The current between the source and drain in the silicon-based field effect transistor is modulated by the application of local voltage changes within the affected neural tissue above the recording sites. Trans-Impedance amplifiers (TIA), fabricated on chip are then used to convert the signal into voltage and amplify the signals. In addition to this a number of systems based on CMOS with additional functionalities such as oxygen sensors and temperature sensors etc has been developed to study the neuronal activity covered extensively in review [11], which is out of scope of this thesis. However, many challenges arising out of processing of signals from these systems are discussed in subsequent sections.

2.3 Front-end electronic circuitry

Front end electronic circuitry comprising of neural amplifiers are indispensable part of neural micro systems. Neural signals acquired from the electrode interface are characterised by bandwidth i.e. the frequency content and their amplitude. The frequency of the signal can range from below 100 Hz to mid-range of 10's of kHz. The amplitude of these signals is usually low (μVs). Therefore, neural amplifiers with high gain, low-noise, low-power and

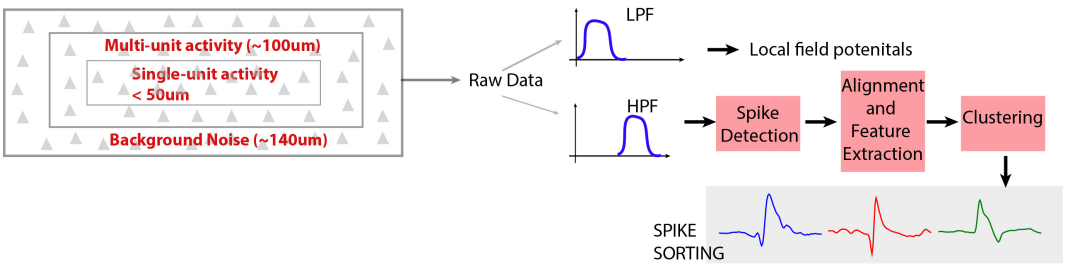


Figure 2.6: Conceptual diagram for spike detection and sorting. The single-unit activity i.e. activity of few neurons effectively in the vicinity of ≈50 μm from the recording electrode can be distinguished in individual components. The raw data is filtered to obtain local field potentials (LFP) and spike trains. Spike trains are then broken down in individual spike shapes by spike detection followed by feature extraction and clustering (triangles in grey represent neurons).

light weight characteristics is a pre-requisite, covered extensively in [50–52]. State-of-the-art amplifiers fit in the chip size of $9.6mm^2$, addressing a lower and upper cut-off frequency of 0.5-6 kHz respectively, with a power dissipation of 1.45 mW and ADC resolution of 10 bits [53]. There has been keen interest on fabricating the neural amplifiers close to the electrodes, also demonstrated in [53].

2.4 Neural Signal Processing

Signal processing is a discipline that encompasses the fundamental theory, applications, algorithms, and implementations of processing or transferring information contained in signals [54]. Neural signal processing is aimed to extract information from neural signals for the purpose of understanding how the brain represents and transmits information through neuronal ensembles. It is an emerging discipline that combines statistical signal processing, control and optimisation methods to process neuronal data. As the amount of neural data recorded through the micro-electrodes grow, it becomes an increasingly challenging demand on micro-electronics to transmit the recorded information. There are limitations on communication bandwidth of implantable devices, fundamentally constrained by the energy budget and the total amount of heat dissipation allowed to avoid tissue damage. These factors prohibit from transmitting such large volume of data wirelessly and therefore it is essential to find potential pathways to perform a part of signal processing on-chip with minimum power budget. Efforts in the area of hardware implantable methods which are capable of reducing the data rate while maintaining accuracy in spike sorting are being researched, which can be essentially based on data compression [55] or feature extraction [15, 56]. **However, it is important to note that improvements in field of micro-electrode technology have not been matched by parallel developments in the field of neural signal processing yet.**

Figure. 2.6 conceptually details the traditional process of spike detection and spike sorting. The raw data recorded from the neurons can be effectively filtered into local field potentials (it illustrates the dynamics of neural tissues surrounding the electrodes and are generated

by the input currents of the dendrites from the surrounding neurons) and 'spike trains' [57]. Filtering leads to visualisation of spikes on background noise. The bandpass/filtered data is then differentiated into 'spikes' and 'background noise' (called as spike detection) by using methods such as setting simple user-defined thresholds. Importantly, the activity of the neurons is **superimposed** and it is essential to distinguish the identity of each neuron individually by differentiating the spike shapes. The spike shape is affected by number of factors such as distance and orientation of the electrode from the recording neuron [58], however, to classify the shapes, several features such as spike amplitude and spike width are chosen and based on these features the spikes are classified into 'clusters' [15, 56]. Each cluster is then associated to a specific neuron and the clusters which cannot be distinguished are associated with 'multi-unit' activity (often characterised by lower amplitude) [12]. **In summary for spike sorting, the flow chart is alignment of detected spikes, feature extraction, clustering and finally association of clusters to original neurons known as classification.**

The following subsections cover a review on spike detection (subsection. 2.4.1) and spike sorting techniques (subsection. 2.4.2) along with the current challenges in the field.

2.4.1 Spike Detection Techniques: A Review

Spike detection is an important technique, where the spikes (the genuine neural activity/events) are distinguished from the background noise. The widely used method for this procedure is setting of a threshold. Setting of thresholds can be fixed or adaptive. In fixed-threshold, pre-set condition is used throughout the recording. On the other hand, adaptive threshold approach monitors the changes in the background and adjusts the threshold value accordingly. In any case, the threshold value has to be chosen carefully because if the threshold value is too small, noise fluctuations will lead to 'false positive' events and if it is too large, low amplitudes genuine spikes will be missed [13, 14].

Spike detection in early days [14] was performed by setting the threshold's manually by the user. Whenever the voltage signal crossed the threshold, a pulse would be generated to indicate a spike. In case, a spike waveform was needed, after every threshold certain user-specified points would be captured. Such methods are used till date and some of the approaches followed for setting of the thresholds for spike detection today are [59]: (a) **Absolute Value/Median:** The threshold can be applied to either the raw signal or to the absolute value of the waveform. In literature [59], absolute values of threshold are proven to be better than setting a simple threshold. Thresholds can also set using a certain multiple (3-5) of standard deviation of the noise, which would minimise the probability of noise exceeding the threshold. The standard deviation of the noise can be estimated using the entire signal including the spikes, but [60] showed that as the firing rate of the spikes increases, the thresholds becomes high. Therefore, they proposed another threshold estimate (Threshold $= 4\,\sigma$), where the estimation of σ is based on the calculation of median of the noise samples

[12]. (b) **Nonlinear Energy Operator (NEO)/ Teager Energy Operator (TEO):** This method is based on detecting the changes in the energy or amplitude of the signal without regarding the frequency. Since spikes are characterised by localised increased in energy, this method has shown many advantages. The threshold is generally set to a scaled version of the mean of estimation of NEO operator. The detailed equations/methods of estimation can be seen in [59, 61, 62]. Some other recent methods used for spike detection are **wavelet transform** [60] or setting of the thresholds using blocks of data to reduce the error rates and reduce effects due to noise distributions.

Alignment: Post spike detection, every detected spike is aligned to the set threshold, therefore, to reduce the effects of noise while performing the next step i.e. 'feature extraction', it is essential (beneficial) to align the spikes to a certain parameter such as maximum/ minimum amplitude values or to a maximum slope [63]. The misalignment of the spikes in time might also lead to spreading of the clusters while feature extraction, which in turn might further complicate the clustering procedure. Therefore, the procedure of alignment of spikes has proven to be useful.

2.4.2 Spike Sorting Techniques: A Review

Traditionally, spike sorting was performed using few features such as 'spike-amplitude' or 'spike-width' and techniques such as 'window discriminators' (time-amplitude windows) or 'template matching' [12, 14] were used. However, now a days these approaches are becoming impractical because large number of electrodes are used for recording the data. The other major impediment in using such techniques is the manual intervention by the user to compensate for the non-stationarity of the recordings and the constantly changing spikes shapes [15]. Broadly speaking, more number of discriminative features could have been used to distinguish the spike shapes, but till few decades back it was limited by the availability of the computer resources [13].

The field today is clearly moving towards **on-chip spike-sorting systems**. To put this in perspective, in conventional systems, neural data recorded through the implantable electrodes is transmitted to the recording/computing system via wires. Such tethered connections leads to increased chances of infection and restricts the mobility of the subject. Therefore, there has been an increased interest in wireless systems, which can permit the experiments/study in more ecological conditions. Examples showing wireless platforms capable of transmitting full broadband signal, whose performance was demonstrated similar to the wired systems [64] can be seen in literature. **However, wireless systems immediately lead to restriction on power, bandwidth and the amount of data that can be communicated through the wireless link.**

A more innovative approach to the problem is to perform most of the computation on-chip and transmit limited amount of information off-chip for further processing. People are

already following an hybrid approach to the problem, where initially spike shapes are sent off-line. The data processing is then performed and only few templates are sent back to the chip, so that subsequent spike-sorting can be performed using the computed templates [65].

Thus for the first stage of spike-sorting, i.e. feature extraction, the techniques can be loosely grouped into **off-chip and on-chip feature extraction methods** [66].

1. **Off-chip Feature Extraction Methods:**

   Principal Component Analysis (PCA) is one of the most trusted and widely implemented benchmark method for feature extraction and clustering. In this method, a set of principal components (PCs) are estimated, which estimates the directions in which most of the variance in the data is present and therefore helps in 'dimensionality reduction'. The dataset is decomposed in a set of eigen vectors and eigen values, which gives the information about the direction and value of variance respectively. To understand this intuitively, imagine for classifying a waveform with spikes and to distinguish that from noise, you use ten features, which equates to a 10-dimensional problem. PCA will decompose the whole data into ten eigen values and the highest eigen value would infer the maximum information about the dataset. For example, if the eigen value corresponding to spike amplitude had the highest number, that would mean the maximum information about the spikes is contained in the spike-amplitude and the rest can be ignored [14, 59, 67].

   In most of the analysis, in spike-sorting two or three principal components are used and the rest are ignored (see Figure. 2.7). The limitation to this method is that it can only be used offline after the acquisition of data and increasing number of dimensions leads to computational complexity making the whole process slower. Other method shown to work equally well for feature extraction is **Discrete Wavelet Transform** as introduced in [68]. This technique was introduced because it gave better resolution in time and frequency. It is defined as the convolution between the signal and the wavelet function. The details of this technique are out of the scope, however, just like PCA, in this technique a set of 'expansion coefficients' are estimated, which can then be clustered to achieve spike classification. This method also leads to high computational costs and therefore, just like PCA is difficult to implement for on-chip spike sorting.

   Importantly, **dimensionality reduction** is a crucial and significant step in spike sorting. The reduction of dimensions in the dataset leads to less burden on memory and computing power, which in turn would mean less area and power requirements for the spike-sorting hardware. Another important advantage is that this leads to improvement in accuracy of clustering. Using more dimensions to improve the spike classification is only beneficial to a certain point, after which it causes the cluster performance to degrade.

2. **Proposed On-chip Feature Extraction Methods:**

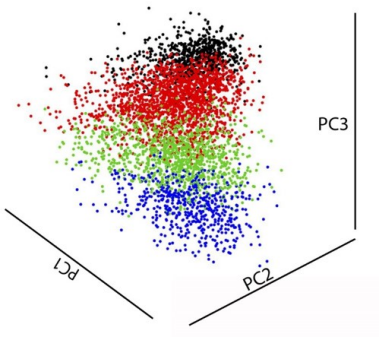


Figure 2.7: Concept: Feature extraction and clustering. PC: Principal Component

For simpler hardware implementations i.e for keeping the dimensionality low and reducing the computational burden, methods such as **Discrete Derivatives (DD)**, **First and Second Derivative** and **Integral Transform (IT)** have been proposed [59, 66]. DD are estimated by computing the slopes at each sample point of the spike over several time scales.

DD = s(n) - s(n-x); where s is the spike, n is the spike sample and x is an integer related to the time scale (chosen by the user). Similarly, IT is the method in which spikes are classified on the basis of areas under the negative and positive phases of the spikes (reducing the features to 2). In literature, DD in comparison to PCA, DWT and IT, has been found to be more accurate and less computationally complex [59].

The final step in spike sorting is **clustering** followed by **classification**. After feature extraction, extracted features are plotted on a scatter plot, as shown in Figure. 2.7. The cluster boundaries are then defined, and finally each cluster is associated to a neuron in the original recording known as 'classification'. In earlier days, clustering was performed by hand, which obviously lead to human biases, however now more automated methods are used for performing clustering such as **k-means** [12, 69]. It is a very simple and fast algorithm where each data point is assigned to the cluster with the nearest centroid. Despite being effective, this method can not be used in real-time for on-chip classification. Examples of other proposed methods are **super-paramagnetic clustering** [68], **Bayesian clustering** [13] and an overview of the published spike sorting algorithms detailing the detection, feature and classification algorithm is also covered in [69].

2.4.3 To sort, or not to sort?

A viable question to investigate for multi-unit activity is whether spike sorting is actually essential and useful, or simply the rate of spikes can be used [36]. The answer to this question lies in the area of application and the heterogeneity of the situation can be considered in the following points:

1. Studies such as [36], where the effects of different spike- sorting schemes on the quality of off-line reconstruction of arm trajectories is evaluated suggests that simple automatic spike-sorting methods can improve the efficiency of decoding from the motor cortex at minimal computational cost, compared to decoding without spike-sorting.

2. It can be seen in the literature, that several neural prosthesis implementations only require qualitative information. For example, in reference [7], the authors specify that especially in the very biomedically relevant cases, where cuff electrodes are employed, the signal it 'reads' arises from the multi-unit activity and identification of individual spikes is not necessary; rather a measure of overall activity over timescales longer than individual spikes is sufficient. Similarly, in the context of BMI applications, studies such as [70] suggest that information of sorted spikes is not essential .

3. Several other papers, for example [71], suggest that in few applications precise spike timing of detected spikes in not required. This can also be noted by the success of cochlear neuroprosthesis, which suggests that brain has remarkable capacity to make use of even the most limited amount of sensory stimulation and the plasticity of these neural circuits is such that the brain can interpret somewhat inappropriate but systematic stimulation of the sensory pathways, and can use this information to make useful judgements about the world.

   To put it in more simpler terms, we still don't have enough information or understanding about what is encoded in premotor and motor areas of the cortex to achieve higher orders of information transfer rates and more efforts thus need to be dedicated to understanding what kind of information is represented in the neuronal populations.

Therefore, it is entirely possible to make use of systems in applications where only set of detected spikes is known, or where the precise spike timing is not known or for that matter, where we have the whole set of spike sorted information.

2.4.4 Limitations and Challenges in Spike Detection and Spike Sorting

There are number of challenges in the field of spike detection and spike sorting, and there seems to be no optimal method that is capable of performing these techniques reliably and several issues remains in this domain. To mention a few, there is a huge amount of variability across signal and noise levels, the unknown signal-to-noise ratio, variability across channels further complicated by electrode drifts and variation with time. The electrochemical degradation of electrodes over time is also another variable parameter, and therefore it is essential to design the study carefully before the experiment starts.

Moving on to spike sorting, there is an issue with the validation of results due to lack of fundamental datasets (ground truths) containing spike shapes, which can be used for evaluating different classification algorithms [69]. Overlapping spikes, bursting neurons, electrode

drifts continue to remain challenging problems in the field of spike-sorting [12]. Another interesting aspect to keep in mind is the classification of clusters. Due to drift of signals over time and variable spike shapes emerging over time from the same neuron, the number of clusters fade and drift over time. There is a big room to develop algorithms which can combat this issue. Recently there are studies being performed to assess the minimum requirements for an on-chip spike sorting implementations that yields more accurate results in terms of computational cost and power consumption [65].

Several other are listed below.

1. **Design Constraints and Trade-Offs:** It is of great interest to the neuroscience community to develop spike sorting algorithms on a low power chip, to enable wireless transmission of the data and ensure ecological and secure conditions to perform experiments with animals. This research work [25], gives an excellent insight and highlights the trade-offs to be considered while designing a next generation neural interface. With the perspective of power budget of the entire system, it will be essential to carefully consider the application and scheme trade-off. For example, in the presented study [25], it is demonstrated that bandlimiting the signal of interest can help in achieving the same spike detection sensitivity with a small false detection increase, while reducing the front end amplifier power budget. To what level, this rates can be accepted is directly related to context of application under study.

2. One way of analysing the performance of spike sorting algorithms is to analyse the number of neurons it can detect. Studies such as [23] suggests a discrepancy in the number of neurons detected by the spike sorting algorithms and the number of neurons that were actually present in the original neural recording. In this case, 8-10 neurons were detected when 20 neurons were present in the original recording.

3. Spike sorting becomes an increasingly challenging task with increasing number of neurons [12]. The performance stays close to ideal up to 5-6 units [60], above which the performance degrades and there are misses especially with the low firing rates neurons.

Since their are no clear standards for benchmarking the performance of spike-sorting algorithms and clearly there is huge variation in the present ones, it is of urgent importance to develop models which accounts for the recordings of much higher complexity and can be used as benchmarking standards. The development of spike sorting methods have to go hand in hand with micro-electrode techniques.

2.5 Summary

The improvement in microelectrode fabrication technology and the corresponding integration with CMOS technology has led to simultaneousness neural recording from thousands of sites.

This has led to generation of huge volume of recorded data and the pivtol issues at present mainly centres around data sorting. Most of the present methods relating to spike-detection and sorting uses heavy and complex off-line processing. This thesis proposes one possible-method in this direction by using RRAM devices, which in future can also be integrated with CMOS system for on-node processing.

Chapter 3

Metal-Oxide RRAM

This chapter introduces the metal-oxide RRAM devices. In the present structure, section. 3.1 surveys the literature in the field and briefly demonstrates the spectrum of materials, mechanisms and switching behaviour of the metal-oxide devices. It also includes the results from electrical characterisation of devices (original contribution of the author). section. 3.6, is the experimental work of author detailing the automated characterisation protocols developed for en-masse testing of devices including the metal-oxide devices employed, device behaviour, methodology for electrical characterisation developed and resistive switching characteristics.

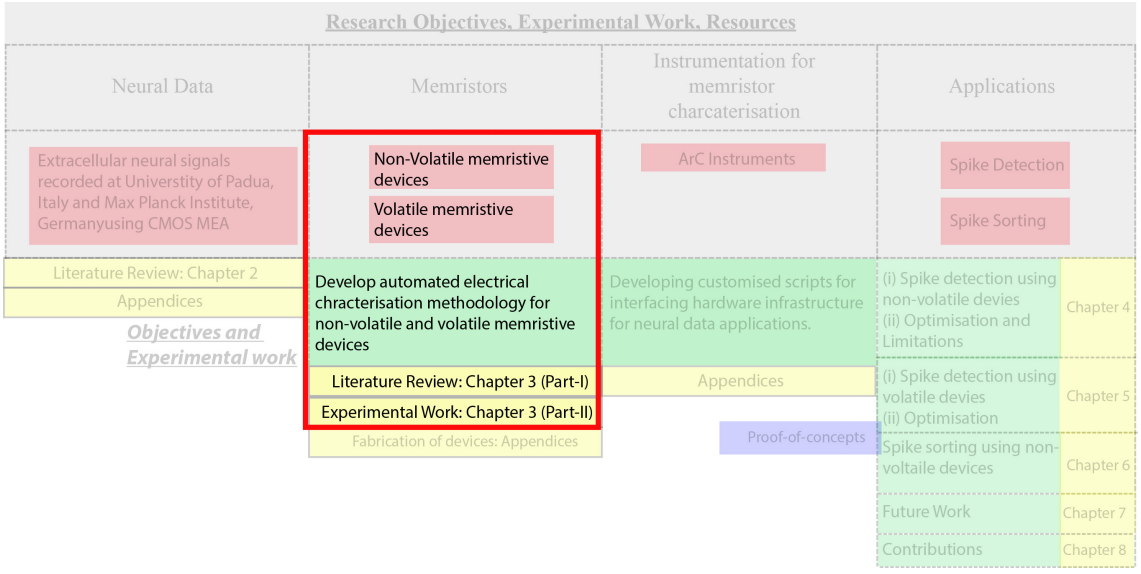


Figure 3.1: The research objective of this chapter.

3.1 RRAM: Background

Aggressive levels of miniaturisation has followed the Moore's law leading to shrinking and doubling of transistors on chip every two years [72]. This scaling paradigm, however, will

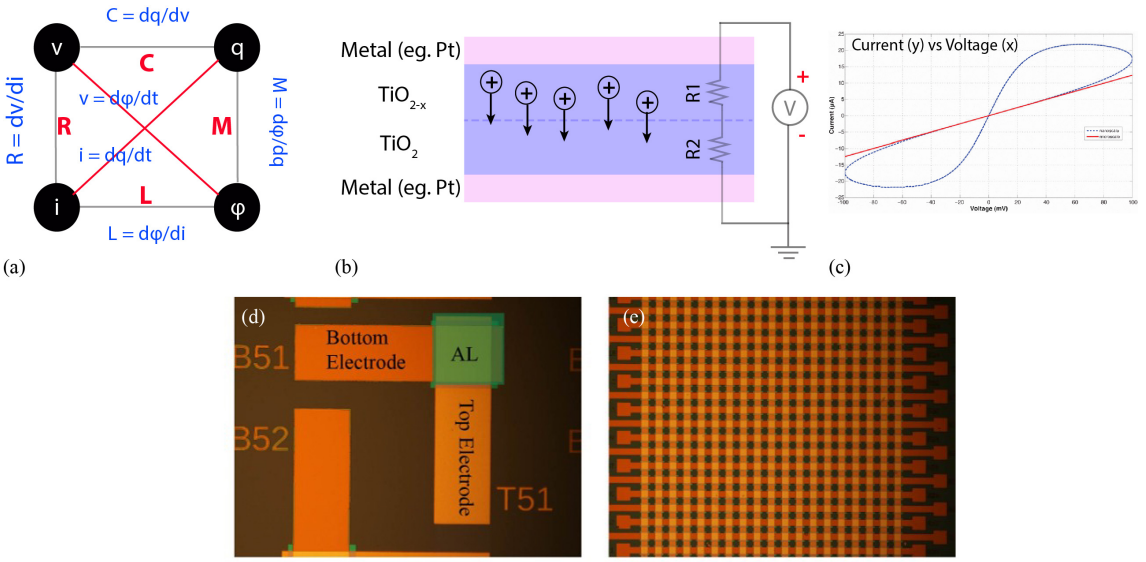


Figure 3.2: (a) Linking four circuit elements, (b) Physical structure of memristors and a basic electronic equivalent model of the device structure [75], (c) electrical characteristics of memristors i.e. non-linear relationship between current and voltage [78]. (d) Single devices, and (e) cross-bar architecture.

eventually come to an end and we will need devices which are not only infinitesimal but increasingly more capable [73]. The spectrum of emerging memories includes numerous potential candidates such as Magnetic Random Access Memory (MRAM), Phase-change Memory (PRAM), Ferroelectric Random Access Memories (FRAM) and Resistive Random Access Memory (RRAM) [74]. In 2008, Stanley Williams and his research team at Hewlett-Packard Labs demonstrated one potential candidate of such device, commonly known as 'Memristor' [75, 76]. Memristor are also referred as resistive RAM/RRAM/ReRAM which is a resistance based memory [1].

3.2 History and Science of Memristor

For more than a century, we know of three fundamental passive circuit elements i.e. capacitor (1745), inductor (1831) and resistor (1827) linking four circuit parameters namely current i, voltage v, charge q and flux ϕ via six different equations [75] (see Figure. 3.2(a)). In 1971, Leon Chua, a professor at Berkeley University, argued the existence of fourth fundamental circuit element called 'memristor' linking charge and flux as shown in Fig. 3.2(a) [77]. He proved that memristor behaviour could not be replicated by any other combination of three other elements, which is why he postulated 'memristor' to be the fourth fundamental circuit element.

The kind of memristor exploited within this thesis are two-terminal metal-insulator-metal devices (see Fig. 3.2(b)), whose resistance decreases depending on the magnitude, polarity

[1] MRAM, FRAM, PRAM are out of scope of this thesis. This research work studies only RRAM devices.

and the length of time the voltage is applied across its terminals. These devices can be fabricated either as single-devices or in a crossbar architecture (see Figure. 3.2(d) and (e)). If the charge flows in one direction of the circuit, the resistance of the component will increase. If the charge flows in opposite direction, the resistance will decrease. The change in the resistance of the device can be read by applying another small voltage. If the flow of charge is stopped by turning off the voltage, the component will 'remember' the last resistive state and when the voltage is applied again, the resistance of the device will be what the last resistive state was. In language of electronics, it can be thought of as two variable resistors connected in series, whose resistance can be varied by applying potential across the terminals. Therefore, it was named as 'memristor', combination of words 'memory' and 'resistor' meaning a device which remembers its history. Naturally, the 'remembering' property of the memristor makes these devices interesting for memory applications in computers.

Chua proved the way charge and flux are linked is similar to non-linear relationship between voltage and current [77] (see Figure. 3.2(c)). When electrically characterised, the devices demonstrate a 'pinched-hysteresis loop' and the resistance of the device varies according to the total amount of charge that has passed through the device. To understand it more clearly, it is not simple on and off, it illustrates a non-linear relationship and remembers the last state. It is essentially an integral of input voltage and this behaviour can also be qualitatively related to the way synapses function in brain.

Memristive devices can be classified in number of ways and here it is briefly reviewed on the basis of materials (section. 3.3), switching behaviour and mechanisms (section. 3.4).

3.3 Memristive materials

The author follows the classification followed in review article [30]. According to this classification, the switching materials are grouped into two categories i.e. anion devices and cation devices (see Table. 3.1).

3.4 Resistive switching behaviour, electrical characteristics and mechanisms

Resistive Switching behaviour - On the basis of electrical characteristics of the devices i.e. current-voltage characteristics, the device behaviour can be classified into unipolar and bipolar. In **unipolar** switching (also called as non-polar switching), the switching direction depends upon the amplitude and not on the polarity of the applied voltage. In **bipolar switching**, the switching is directional depending upon the polarity of the applied voltage and this behaviour can be linear or non-linear [80].

Resistive Switching mechanisms :

Table 3.1: Classification of resistive switching materials [30, 74, 79].

	Switching Materials	**Examples**	**Mobile species**	**Switching Mode**
Anion Devices (Valence Change memory)	Oxide insulators such as transition metal oxides, complex oxides, large bandgap dielectrics, and some non-oxide insulators such as nitride and chalcogenides.	MgO, TiO_x, ZrO_x, HfO_x, VO_x, NbO_x, TaO_x, CrO_x, MoO_x, WO_x, MnO_x, FeO_x, $SrTiO_3$, $SrZrO_3$, AlN, ZnTe.	Oxygen anion or equivalently the positive-charged oxygen vacancy.	Bipolar/ Unipolar
Cation devices (Electrochemical Metallization Memory)	One electrode is made from electrochemically active material such as Cu, Ag or alloy of these materials such as CuTe. Counter electrode is fabricated with electrochemically inert material such as W, Pt. Insulating materials: Ge_xS_x, As_2S_3, Cu_2S etc	Ta_2O_5, SiO_2, HfO_2, WO_3, MoO_x, ZrO_2, CuO_x etc.	Metallic Cation	Bipolar/ Unipolar

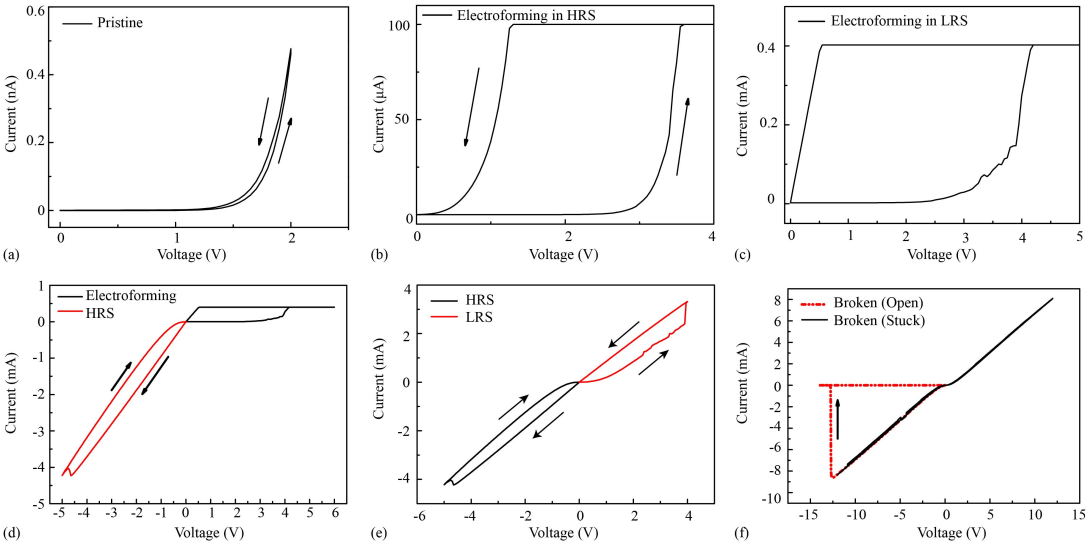


Figure 3.3: Electrical characteristics of the devices. (a) Pristine Device, (b) Electroforming in HRS, (c) Electroforming in LRS, (d) Device switched into HRS, (e) Device switched in LRS, (f) Broken device (Open and Stuck).

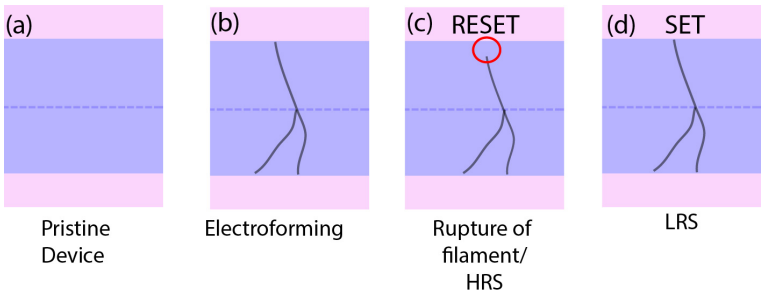


Figure 3.4: Filamentary switching mechanism. (a) Pristine Device, (b) Electroforming procedure, (c) Rupture of filament (HRS), and (d) SET process, LRS.

1. **Filamentry Mechanism:**[2]In Anion/VCM cells [81, 82], the mode of switching behaviour can be unipolar or bipolar. The resistive switching mechanism in different materials varies depending upon number of parameters such as structure of the device, electrical protocol applied, the effect of device fabrication (symmetric or asymmetric), electrode material etc. However, in the bigger landscape, the first crucial step in understanding the switching mechanism is the 'electroforming' procedure. In most of the cases it has been found that due to 'electroforming', the electrical (or thermal effect) leads to creation of both, negatively charged oxygen ions and positively charged metal ions. This results in a chemical change in the oxide leading to valence state change and creation of a new phase in the active layer, which in turn leads to 'resistive switching'. Simply put, 'electroforming' is an irreversible soft dielectric breakdown of the active layer performed with a DC sweep with the current limit enabled. For instance, this is an observation from TiO_x memrisitve devices employed during this research work (see Figure. 3.3(a)) [3]. The difference between the 'soft' and 'hard' breakdown can be seen in Figure. 3.3(b) and (c) respectively. The compliance current protects the device from forming in ohmic region. Notably, due to electroforming resistance of the device decreases from several GΩ's to MΩ's or lower [4].

   According to mainstream literature, electroforming leads to creation of conductive filamentary pathways whose diameter can be tens and hundreds of nanometers [79, 81, 83, 84] (see Figure. 3.4(b)). These filaments have been understood to be 'switching channels', whose rupture and formation repeatedly leads to switching between a high resistive state (HRS) and low resistive state (LRS), as shown in Fig. 3.3(d), (e) respectively and Fig. 3.4(c),(d). The generally accepted explanation for this phenomena is as follows: as the voltage applied across the terminals is increased, the electric field between the terminal grows. When it grows sufficiently, the conductive filament inside the materials starts growing from one electrode to another electrode controlled by the

[2]The switching mechanism has only been described from the perspective of electrical characterisation of devices.

[3]Electrical characterisation results from single-devices.

[4]Electrical characteristics of the devices were measured using 'Keithley 4200' semiconductor characterisation suite.

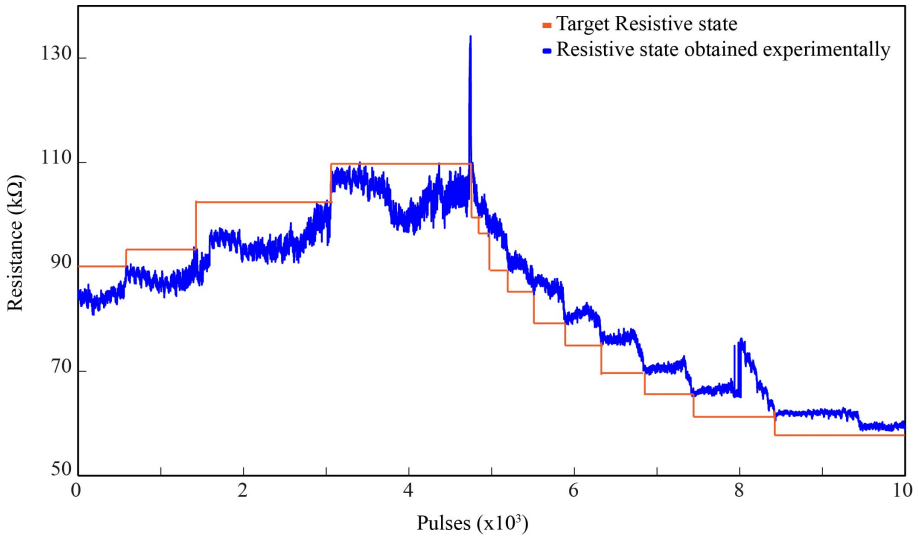


Figure 3.5: Capability of the device to switch to multi-resistive state.

application of compliance current. The application of voltage during the 'electroforming' procedure is usually higher than what is required during the switching procedure. It is observed that application of voltages under a fixed range, leads to repeated cycling of devices between HRS and LRS. However, using invasive voltages such as $\approx \pm 15\,V$ can lead to breakdown/complete failure of the device (see Figure. 3.3(f)). Moreover, the switching between HRS and LRS can either be non-volatile in nature or volatile (discussed in section. 3.6).

Importantly, instead of using the traditional sweeps of current and voltage, an accurate analogue control of the device is possible by providing varying write and read pulses. The devices tend to show rich switching dynamics by showing the capability of the device to be programmed in multi-resistive states, as it can be seen in Figure. 3.5 [85]. This property can be exploited for various applications such as multi-bit memory [86], however, the electrical characterisation of devices tend to exhibit a range of variability, as it can be seen in in Figure. 3.6, 3.5, 3.7 and 3.8[5]. From the cumulative study of the experiments in this set, it was concluded that pulse-amplitude had a stronger impact on resistive state of the device (instead of pulse-width). Therefore it is justified and essential to device a method for standardising the electrical characterisation protocol (discussed in section. 3.6), primarily to reduce device-to-device variability.

In order to optimise the structure, the author has also experimented with different stack structures with modification in the top electrodes for instance, as shown in Table. 3.2 and Table. 3.3. The same behaviour has been demonstrated by various combinations but it is essential to engineer the devices and optimise one combination which can be best applied from the application perspective.

[5] Stack structure: (Ti (5nm) /Pt(10 nm) /TiO_2 (25 nm)/Pt (10 nm)) were biased using the hardware infrastructure as described in Appendix B. The processing of the data is carried out using MATLAB software. The devices exhibit bipolar switching after electroforming. The electroforming voltage for these devices was approximately 7 V.

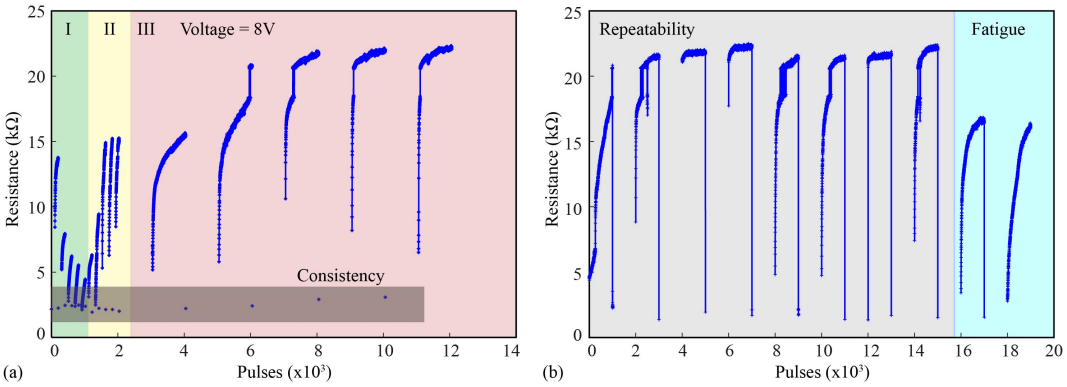


Figure 3.6: Biasing of packaged single devices using the hardware infrastructure (see Appendix. B). The applied pulse voltage is 8 V.(a) Resistive change in the state of DUT under pulse width 1ms (I), 10 ms (II) for 100 pulses, 10 ms for 1000 pulses.(b) Testing the repeatability of the same device which eventually tends to fatigue. The device switches to low resistive state with 8 V (10ms) and high resistive state with -5 V (1ms).

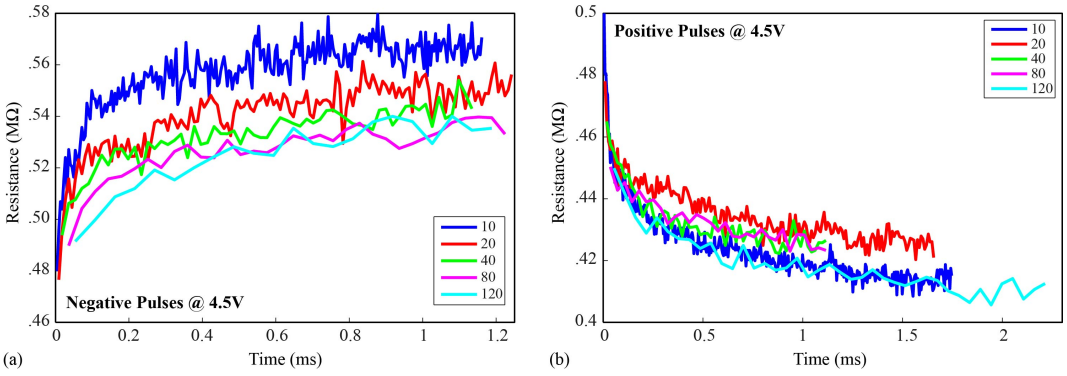


Figure 3.7: Resistive behaviour of the device with pulse width of 10μs (blue), 20μs (red), 40μs (green), 80μs (pink) and 120μs (cyan) for pulse amplitude (a) -4.5 V and (b) +4.5 V.

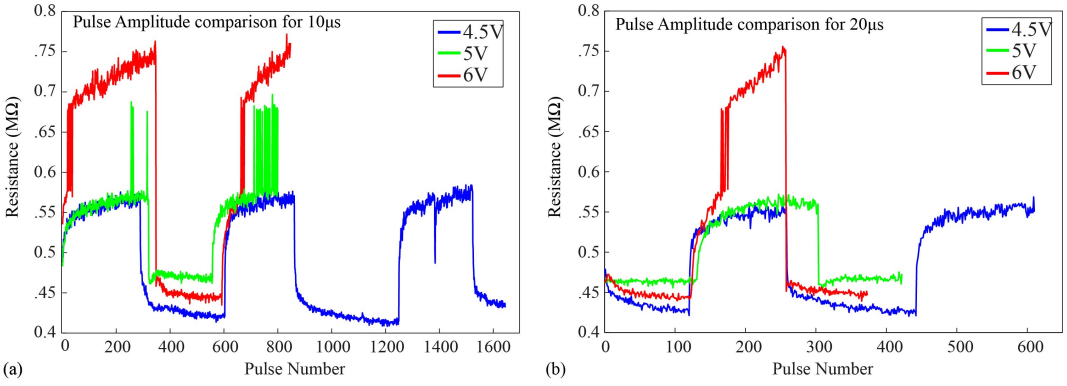


Figure 3.8: Resistive behaviour of the device with pulse amplitude of 4.5,5 and 6 V with pulse width of (a) 10μs and (b) 20μs

Table 3.2: Results from the electrical characterisation of stack structure Ti (5nm)/Au(75nm)/TiO_x (25nm)/Top Electrode (75nm).

Device Structure: Ti (5nm)/Au(75nm)/TiO_x (25nm)/Top Electrode (75nm)		
Top Electrode	Results	Favourable
Tungsten (W)	Electroforming voltages were found to be 6-7 V. Devices seem to toggle between HRS and LRS for thousand of pulses.	Yes
Aluminium (Al)	Conductive	No
Gold (Au)	Conductive	No
Titanium Nitride (TiN)	Electroforming voltages were found to be 4V. Devices seem to toggle between HRS and LRS but with high voltage and pulse width.	Yes
Silver (Ag)	Conductive	No
Gold/Titanium Nitride (Au/TiN)	Conductive	No
Nickel (Ni)	After electroforming, devices seem to irreversibly switch below 500 Ω after few pulses.	No

Table 3.3: Results from the electrical characterisation of stack structure Ti (5nm)/Pt (75nm)/TiO_x (25nm)/Top Electrode (75nm).

Device Structure: Ti (5nm)/Pt(75nm)/TiO_x (25nm)/Top Electrode (75nm)		
Top Electrode	Results	Favourable
Chromium (Cr)	Electroforming voltages were found to be 6-7 V. Devices seem to toggle between HRS and LRS for thousand of pulses.	Yes
Aluminium (Al)	Conductive	No
Gold (Au)	Conductive	No
Titanium Nitride (TiN)	Electroforming voltages were found to be 7-8 V (10-20ms pulse width). Devices seem to toggle between HRS and LRS but with high voltage and pulse width.	Yes
Silver (Ag)	Conductive	No
Titanium (Ti)	Conductive	No
Platinum (Pt)	Visible damage on all the electrodes.	No

2. **Resistive switching in Cation based devices**

   In cation based devices, one electrode is made from electrochemically active material such as Cu, Ag etc and the counter electrode is fabricated with electrochemically inert material such as W, Pt. The accepted explanation of the switching mechanism for such devices can be explained as follows: For example, in stack structure Ag/H_2O/Pt, a positive voltage on the Ag electrode oxidises the electrode atoms into Ag+ cations, which are dissolved in the electrolyte. These cations drift to the Pt electrode, are electrochemically reduced as Ag atoms at the cathode, which then grows to silver electrode and switches the device 'ON' (LRS). An opposite voltage would switch the

device 'OFF' (RESET) [30][6].

3.5 Potential Applications and Performance Parameters

It took many decades from postulation of a memrisitve device by Chua to the realisation of first memristive device at HP. Since then, the increasing scientific contribution demonstrating attributes such as low power consumption [87], high-speed [88], non-volatile behaviour [89], multi-state resistive switching memory [86, 90], scaling have made memristor's an ideal candidate for next-generation memory. Because memristors remember their last state, they can store data, using energy only when you toggle or read the state of a switch, unlike the capacitors in coventional DRAM [74], which will lose their stored charge if the power to the switch is turned off. Although, these devices have a very high potential to be used in high density memory and can replace FLASH but before that there are number of critical issues which needs to be addressed. **They tend to have a very high device-to-device variability and statistical study on the reliability, endurance [91] and retention [92] needs to be performed before it is used for future applications.**

Performance parameters for RRAM devices (in terms of device performances):

1. **Endurance:** It is the number of set/reset cycles which can be applied to the device, for example, TaO_x: 10^{12} switching cycles [93], WO_x: 10^7 cycles [94] and TiO_x: 10^6 cycles [95].
2. **Retention**: TiO_x - 10^4 seconds at 85 degree celsius [96]. HfO_x - 10^6 seconds at 200 degree celsius[97][7].
3. **Low Power:** Power consumption is estimated depending upon the amplitude and pulse-width of read and write pulses, for example, WO_x reset/set at 1.8 V/-1 V [94].
4. **OFF/ON Ratio:** Ratio of HRS to LRS. AlO_x - 10^5 [98].
5. **Switching Speed:** TaO_x [88, 93] and HfO_x [97] less than 10 ns.
6. **CMOS Compatibility:** A CMOS compatible process is where materials involved in the fabrication process can be easily incorporated in normal standardised semiconductor manufacturing processes (including the temperature range). TiO_x for instance is a CMOS compatible material [99, 100].

Lastly, memristive devices have also generated a lot of hope in the neuromorphic community. The classical von Neumann digital architecture which is based on the binary logic can not match the efficacy of the neural network in the mammalian brain. Developing a compact

[6] These devices have not been studied during this work.
[7] For memories it will be essential to have data retention of greater than 10 years

nanodevice which possess inherent learning and memory properties emulating the biological counterpart is a pre-requisite to realise more powerful machines which are more efficient in terms of speed and power consumption [101]. Mimicking the functions of brain, analog electronic circuits, hybrid circuits, containing many connected memristors and transistors could help research brain functions and disorders. Interdisciplinary research has led to an effective integration of two fields where the dynamics of memristors can be linked to synaptic connections between neurons [102, 103]. In principle, the connection between two neurons can be made stronger or weaker depending upon the polarity, strength and the length of electrical signal applied, which facilitates memory. Memristive devices can be thought of as a synapse/connection between the two neurons. Many groups are researching on this application and have demonstrated striking similarities between biological synapses and dynamics of memristors [101, 104–106].

Section. 3.6 of the chapter discusses the automated electrical characterisation protocols developed to carry out en-masse testing of the devices. Due to low yield of devices and high device-to-device variability, it was essential to device standardised approaches to carry out electrical characterisation of devices.

3.6 TiO_x RRAM Devices, Electrical Characterisation and Automated Characterisation Protocols

A major bottleneck towards further development of RRAM applications and their ultimate commercialisation is identifying which devices exhibit sufficiently good electrical characterisation for a specific application. Good cycing endurance [91], on/off resistive ratio [89] and retention figures [92] have been reported in literature but large scale monolithic arrays operating to such elevated specifications remains as yet unattained. Instrumentation for such mass testing has already been demonstrated [107], however, developing automated characterisation algorithms for characterising the devices in an automated fashion and determining suitable biasing parameters for obtaining resistive switching remains unattained. Or for that matter algorithms capable of characterising the metastable resistive state transitions i.e. volatility in the myriad RRAM configurations fabricated for achieving desired performance remains as a major bottleneck [108]. To ensure continued process development, similar protocols are imperative and during the course of this work two such algorithms were developed as discussed in the following sections (Section. 3.7.1 and Section. 3.7.2).

3.7 Characterisation Protocols

This protocol was developed for testing crossbar arrays.

3.7.1 Characterisation Protocol I

Automated cell classifier for resistive switching

The objective of this protocol is to build a statistical and highly automated algorithm that can identify functional devices within a crossbar array under test. The idea is to achieve RRAM process development by identifying the devices in a cross-bar array that can switch between high resistive state (HRS) and low resistive state (LRS) automatically. Importantly, a **'functional'** device or promising device in non-volatile region shows analogue **'waves'** when biased with pulses of opposite polarities. As shown in Fig. 3.9 the device toggles to **HRS** with negative pulses and **LRS** with positive pulses (after electroforming). No underlying assumption is made as to what levels of resistance should constitute a digital 1 or 0, making the proposed algorithm a particular effective process control tool, especially in immature technologies. In addition, testing of large number of devices can also help combat yield problems. In this respect efforts to develop fault analysis and mitigation algorithms and circuits specifically aimed at RRAM technology have already begun. However, much of the work in this field so far has been aimed at tackling memory faults such as stuck-at transition, and coupling faults within the context of conventional, CMOS-based memory arrays with simulated mature resistive switches as storage nodes [109, 110].

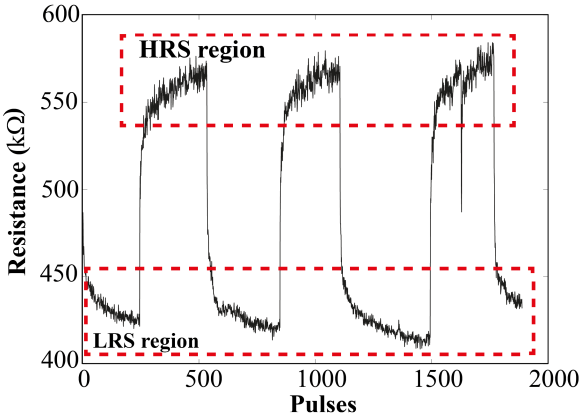


Figure 3.9: Resistive switching in TiO_x RRAM devices. A bipolar biasing protocol was used to introduce transitions between High Resistive State (HRS) and Low Resistive State (LRS).

3.7.1.1 Algorithm Description

Theoretical Background: The algorithm lies on the assumption that the functioning devices can toggle between HRS and LRS under a suitable biasing scheme. The test is designed to be applied to electroformed devices. The general concept of the testing algorithm relies on using standardised 'write pulse' train sets designed to drive each device-under-test (DUT) alternately to HRS and LRS. The device resistive state is assessed after each write pulse train via the application of multiple, non-invasive, low-voltage 'assessment pulses'. A two-means t-test is then performed (MATLAB function t-test2 (standard), two-tailed) on populations of readings taken after LRS-to-HRS pulse train sets (HRS set) vs. populations of readings taken after HRS-to-LRS events (LRS set).

The two-means t-test is described by (3.1):

$$t = \frac{\bar{\mu}_1 - \bar{\mu}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \tag{3.1}$$

where $\bar{\mu}_1$ and $\bar{\mu}_2$ are the means of the HRS and the LRS sets, σ_1 and σ_2 their respective standard deviations, n_1, n_2 denotes the number of readings in each set and t denotes the t-metric. The t-metric, n_1 and n_2 can be fed into statistical tables (standard) in order to compute corresponding p-values (p $\in$ [0,1]). The test directly outputs p-values which quantify statistical significance between differences in the means of the HRS and LRS sets. Functioning devices are expected to score low p-values whilst devices that fail to respond to biasing would sink towards p=1.

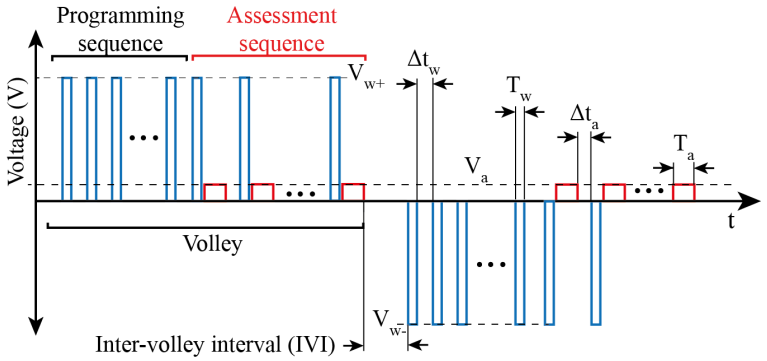


Figure 3.10: Typical biasing scheme applied to devices under test. Only two out of the four volleys are shown. The first volley toggles functioning devices into LRS whereas the second into HRS. Each assessment pulse results in an estimate of resistive state after the application of the 'write' pulse train.

3.7.1.2 Implementation and Testing

A. Devices and Hardware Infrastructure

The experiment was performed on 32x32 crossbar arrays of TiO_{2-x}. The devices comprise a vertical stack structure $Ti/Pt/TiO_{2-x}/Pt$ (5/10/25/10 nm). The fabrication procedure is identical to the one detailed in Appendix. A. The devices exhibit bipolar switching after electroforming [111], with electroforming voltages for this set to be approximately 6.5V. For these experiments, the hardware infrastructure used is described in Appendix. B.

B. Biasing Protocol

Each device was individually subjected to a biasing protocol organised in 'volleys' consisting of programming (subscript 'w') and assessment (subscript 'a') sequences where successive volleys employ pulses of opposite polarities. Programming sequences consist of a succession of square wave write pulses with controlled pulse widths (T_w), inter-pulse intervals (Δt_w) and amplitudes (V_{w+}, V_{w-}). Chosen control parameters for the device batch fabricated according to Appendix. A and the specifications are summarised in Table 3.4 (see Fig. 3.10). Determining the biasing protocol control parameters is a critical stage of the overall process control procedure, which was performed manually. Typically, a medium number (50-100) of standalone and not a part of crossbar devices were subjected to a variety of ad-hoc input waveforms by a human operator. This yields a rough estimate of pulse control parameter ranges (amplitude, duration, and interpulse interval) within which the DUTs show a response and outside which DUTs stop reacting or experience irreversible breakdown.

Assessment sequences consist of series of read pulses paired with a preceding write pulse. The write pulse width and amplitude remains unchanged between programming and assessment sequences whilst read pulses always occur at a standard voltage of 0.5V. Read pulse duration,

Parameter	Description	Value
Δt_w	Write pulse inter-event interval	$100\,\mu s$
T_w	Write pulse width	$100\,\mu s$
V_{w+}	Positive Write Pulse Voltage	$4.5\,V$
V_{w-}	Negative Write Pulse Voltage	$-4.5\,V$
V_a	Read pulse voltage	$0.5\,V$
Δt_a	Read seequence inter-event interval	Automatic
T_a	Read pulse width	Automatic
IVI	Inter-volley interval	$250\,\mu s$

Table 3.4: Biasing protocol parameters use to identify working devices. Parameters are illustrated in Fig. 3.10

and assessment sequence inter-event interval parameters (T_a, ΔT_a) are automatically determined by the system. The choice of forcing 'read' pulse to be preceded by a 'write' pulse was a measure for mitigating the possible effects of device volatility. The testing algorithm seeks devices that can toggle between different resistive states regardless of whether this change is nonvolatile or transient. Interspersing 'read' and 'write' pulses means the target device resistive state is maintained close to HRS or LRS longer; thus, longer assessment sequences can be utilised. Four volleys were used in order to ensure that at least three resistive switching events are captured from functioning devices irrespective of their initial states. The larger the number of volleys, the more switching events may be captured, but also more likely that weaker devices fatigue and become inoperable.

Each volley consisted of ten write pulses and five assessment pulses. The number of write pulses was set to ten because it was found that some devices reacted to each individual pulses weakly (small change in resistive state). Repeated pulsing allowed such weak devices to toggle memory states with a reduced risk of comprising the functionality of more easily switchable devices as would have been the case if higher bias voltage had been employed. Multiple assessment pulses are used to allow an evaluation of the uncertainty involved in computing the resistive state of the DUT after the application of the write pulse train. The larger the number of assessment pulses, the more accurate the estimation of mean resistive switching and associated measurement uncertainty and the more likely that it includes drift due to effects such as volatility [112] or natural degradation of the device with switching.

The Inter-volley interval (IVI) was kept as short as possible in order to minimise measurement time, but the possibility that the choice of IVI may affect the results must be investigated further (notably in regard to heat accumulation, particularly in encapsulated packaged samples). Finally, by controlling the number of volleys, write and assessment pulses can gear the test toward more qualitative switchability testing in immature processes or determine specific performance metrics in more mature processes.

Low p-values can be the result of either genuine switching or cross-talk. In response, a cross-talk filtering stage was added to the testing routine. Cross-talk may arise as a result

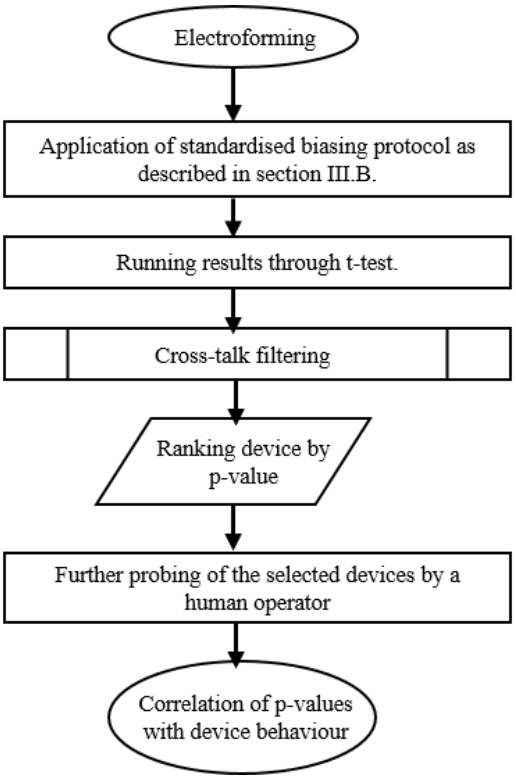


Figure 3.11: Typical testing procedure for the RRAM devices

of the implementation of the biasing scheme during the write operation [113]. During prior experiments with hardware infrastructure it was found that devices with resistive states above 5 MΩ were particularly susceptible to such cross-talk effects. This was found to be an inherent limitation of the testing-hardware, where imperfect wordline and bitline grounding, crossbar line resistances, and other imperfections impose an upper limit to cross-point impedance that can be reliably measured. Therefore, to mitigate this known effect, devices whose computed mean LRS resistance was measured as > 1 MΩ after application of the biasing protocol were eliminated from further data processing.

C. Typical Testing Procedure

A total of 20 crossbar arrays, either packaged or on-wafer, were tested according to the procedure described in Fig. 3.11 and as follows.

1. Electroforming
2. Application of standardised biasing protocol as described in the Implementation and testing section.
3. Running through T-test.
4. Cross-talk filtering.
5. Ranking Devices by p-value.
6. Further probing of the device by human operator.

p-value bins	Crossbar I		Crossbar II	
	n	S/T	n	S/T
[0.0-0.1]	35	10/10	34	7/25
(0.1-0.2]	21	2/10	12	2/10
(0.2-0.3]	7	3/7	7	2/7
(0.3-0.4]	5	0/5	12	4/10
(0.4-0.5]	16	2/10	7	1/7
(0.5-0.6]	16	2/10	13	2/10
(0.6-0.7]	17	0/10	19	1/10
(0.7-0.8]	23	0/10	23	0/10
(0.8-0.9]	36	1/10	82	0/10
(0.9-1.0]	793	1/10	798	0/10

Table 3.5: Second round measurement results: Data from 2 crossbars. [n = Total number of devices found in each bin; S/T = Switchable devices/Total number of devices probed further]

7. Correlation of the p-value with device behaviour.

The final two steps in the testing procedure were performed to visualise the difference in the typical behaviour of devices that score radically different p-values.

3.7.1.3 Results and Discussion

For each tested crossbar array a histogram of devices binned by p-value was constructed. Two typical examples are shown in Fig. 3.12. Initial tests were performed on two crossbar arrays, one accessed via probe card (see hardware infrastructure Appendix. B, Fig. B.1(b)) and the other in a package (see Fig. B.1(a)). 25 devices from the probe card sample and 23 devices from the packaged sample fell into the [0, 0.1] p-value bin. When probed further as described above (implementation and testing) it was found that 8 and 18 devices respectively, carried on behaving consistently with the results of the t-test, i.e. showed clearly identifiable HRS and LRS. None of the devices tested from the p-value bin (0.9, 1.0] were found to be working. Thus the test effectively relegates the non-functional devices to the bottom of the list.

In order to gain better understanding of the behaviour of devices residing in different bins, the experiment was repeated on a new set of crossbar arrays. In this second round of measurements the arrays were comprehensively measured with devices from each of the p-value bins individually tested. Results are shown in Table. 3.5 and are consistent with the preliminary measurements: devices which show distinctive HRS and LRS when probed are generally those that have scored low p-values during testing. These well-behaved devices were manually cycled hundreds of times to ensure that they continue to operate well after the conclusion of the test. Rigorous endurance testing to help determine the effects of pulsing parameter choice on device endurance remains as the future work.

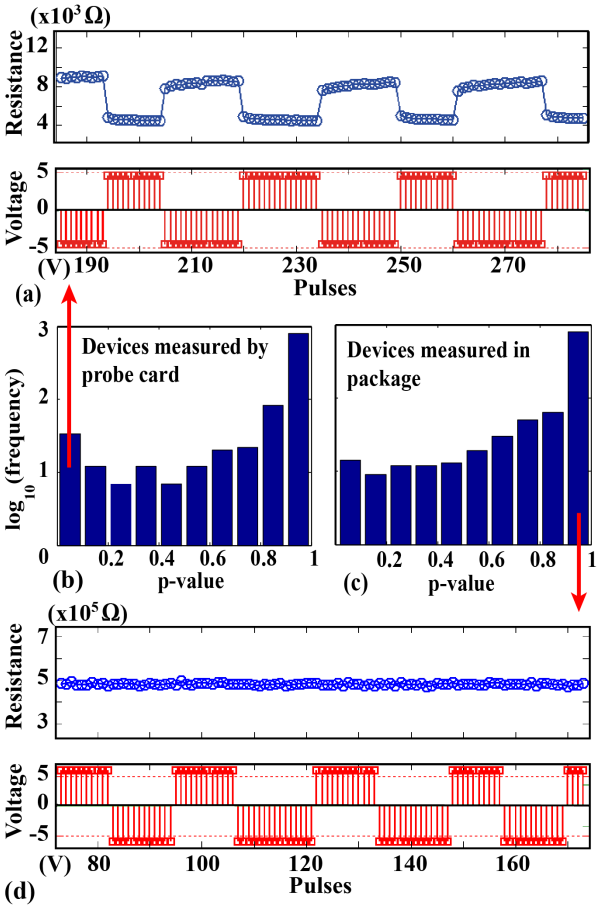


Figure 3.12: First measurement round results: (b) and (c) Typical examples of histograms showing number of devices versus p-value taken from probe-card-accessed and packaged crossbar arrays, respectively. Typical behaviour of devices that scored (a) low p-values and (d) high p-values.

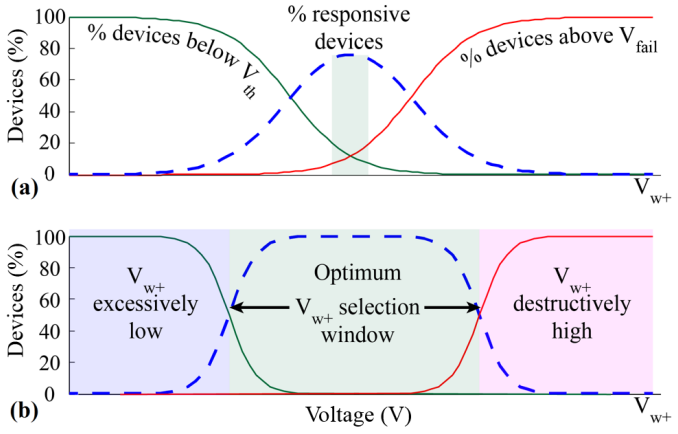


Figure 3.13: Conceptual example of effect of positive write pulse voltage (V_{w+}) selection on the outcome of the proposed testing algorithm. (a) Device stack exhibiting high switching threshold (V_{th}) an failure threshold (V_{fail}) variability. (b) Optimised device stack exhibiting clear gap between (V_{th}) and (V_{fail}) within which switching parameters can be set freely.

It is important to note that this test is intended to be applied to the electroformed devices only. This is because if the device electroforms during the test, the values of the mean of HRS and LRS would become highly skewed resulting in low p-values. All the results in these experiments were taken from fully formed arrays. Devices that scored low p-value but failed to show satisfactory switching were found to be cases of damage during testing and crosstalk that slipped through crosstalk filtering step. Checks revealed no incidences of 'in-test electroforming' in the results here.

Finally, the selection of pulsing parameters (in Table. 3.4) is of crucial significance. Using excessively invasive pulses will damage large number of devices, whereas overly gently pulses may fail to elicit resistive switching which might be the case in mid p-value bins but needs further experimental work. Fig. 3.13 b shows a conceptual example of how the selection of V_{w+} may affect the behaviour of devices during the test. In the case of the mature optimised device stack (see Fig. 3.13 a), there is no gap between the distributions of device switching thresholds (V_{th}: minimum switching threshold) and destruction thresholds (V_{fail}: maximum operating voltage) within which switching parameters can be set freely with relatively minor effects on the final outcome. In stacks where the distributions of V_{th} and V_{fail} overlap strongly (see Fig. 3.13 a), an optimum value must be found.

The protocol is summarised in section. 3.8.

3.7.2 Characterisation Protocol II

Volatility characterisation for metal-oxide TiO_x RRAM devices

Introduction: Contrary to the conventional use as solid-state memory, memristors can be used as synaptic mimics offering promising aspects for fabricating brain-like neuromorphic systems [103]. Typical electrical characterisation of two-terminal and highly scalable memristive devices initially involves an electroforming process and subsequently, application of bipolar pulses resulting in tunable resistive switching leading to thermodynamically stable 'non-volatile' resistive state transitions (as seen in Characterisation Protocol I). However, this can be contrasted with another crucial aspect of the memristive devices 'post-electroforming' i.e. the resistive state 'volatility' which demonstrates **the short-term dynamics of the DUT**. The devices tend to exhibit metastable memory state transitions following which the devices relax to its equilibrium resistive state in a definite time window [105, 108, 114, 115]. This volatility property can be extremely useful for emulating biological functions such as STP/STD (short term potentiation/depression) [101] in memristors and thus merits further in-depth methodical study. The developed characterisation protocol aims at ensuring a methodical process development for automated characterisation of RRAM arrays for operating them in volatile region.

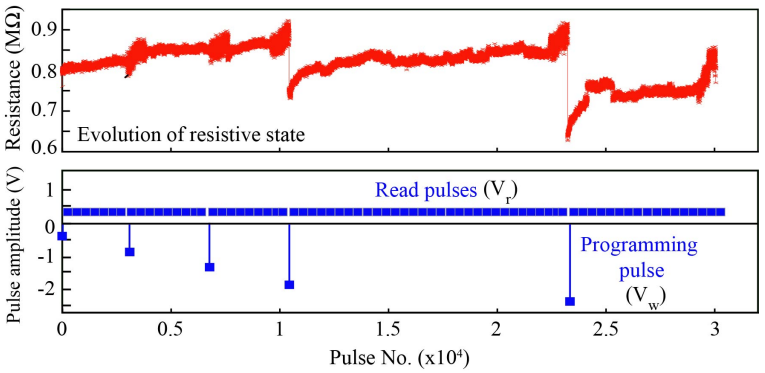


Figure 3.14: Typical volatile characteristics of the solid-state TiO_x RRAM devices. Programming pulse and read pulse is denoted by V_w and V_r respectively.

3.7.2.1 Volatility Characterisation Protocol

An automated testing routine capable of characterising the volatility observed in the TiO_x devices based on the standard two mean t-test is developed. The volatility characterisation protocol operates on the principle of applying a series of progressively more invasive voltage pulses and monitoring the resistive state of the DUT after each individual pulse at fixed read-out voltage over time, as shown in Fig. 3.14 lower and upper panel respectively. The module is split into two parts: Stage 1 measures the resistive state decay over time and halts when no significant resistive state change is recorded in a programmable time window; and Stage 2 which validates the equilibrium state by repeated measurement over 60 seconds. The output of the module determines the time the t-test lasted for and importantly the voltage range under which the DUT can be safely operated in purely volatile region.

RRAM devices and Hardware Infrastructure: For these experiments, solid-state 60 μm x 60 μm and 200 nm x 200 nm TiO_x devices with vertical stack structure of Ti/Pt/TiO_x/Pt (5/10/25/10 nm) and (5/10/10/10 nm) were utilised [116][8]. For electrical characterisation, custom made in-house fabricated hardware described in Appendix. B. In this experiments devices were accessed directly on-wafer via a probe card.

3.7.2.2 Algorithm Description and Implementation

The volatility algorithm is designed in two stages. In stage I, the 't-test' stage, the resistive state of the DUT is assessed and then the system applies a user-defined programming pulse of suitable amplitude and width (V_w, T_w) as illustrated in Fig. 3.15 a. Immediately afterwards, the DUT is subjected to a single, low-voltage read pulse (V_r, T_r), during which DUT resistive state is assessed multiple times. The system reads the state of the DUT in batches (B) of n measurements, where n is user-defined. Parameter description is also given in Table. 3.6. Data recorded in each batch is then used to estimate whether the DUT has reached an equilibrium/steady-state condition i.e. if the device has relaxed to a stable resistive state

[8]The device fabrication procedure for micro- and nano-devices is described in Appendix. A

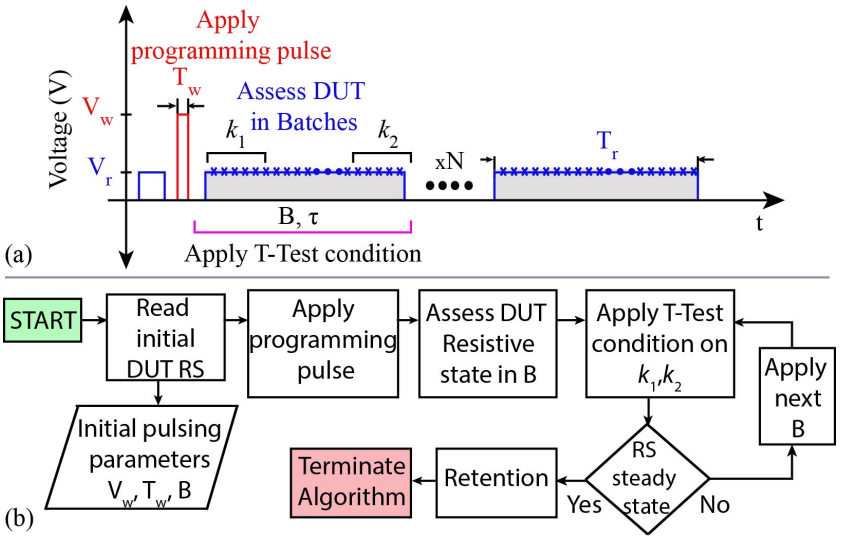


Figure 3.15: Volatility module. (a) Schematic for the volatility module. The red and blue color illustrates the programming pulse and read pulses respectively. Grey color indicates one batch of read pulses. (b) Flow chart for the volatility algorithm implementing standard t-test condition.

Table 3.6: Volatility Algorithm parameter list

Parameter	Description	Units	Value
V_w	Write pulse voltage	V	0 to -3 V
V_r	Read pulse voltage	V	0.2 V
T_w	Write pulse width	μs	100 μs
T_r	Read pulse width	μs	Automatic
B	One batch	-	-
n	No. of reads in a single batch	-	1000
τ	Time for one batch	-	$\approx$1s
N	Number of Batches	-	-
k_1, k_2	No. of read pulses used in the beginning and end of each batch	-	10,10

range. This is done using a standard two mean t-test to compare the means of the first and last k-values within each batch against measured variance as depicted in Equation. 3.2. Importantly, the system makes no assumption on what the equilibrium state should be.

The two-means t-test is described by (3.2):

$$t = \frac{\bar{\mu}_1 - \bar{\mu}_2}{\sqrt{\frac{\sigma_1^2}{k_1} + \frac{\sigma_2^2}{k_2}}} \tag{3.2}$$

where k_1 and k_2 are the number of pulses used for estimation of mean (μ_1, μ_2) and standard deviation (σ_1, σ_2) at the beginning and end of each batch. If the computed absolute value is less than the set threshold (experimentally optimised at '1' for this work) the t-test phase is terminated otherwise the next batch of read pulses is applied and the same procedure is

repeated until the DUT reaches a steady-state (Fig. 3.15 b). The choice of k value and t-metric threshold determines the confidence level at which the system decides that the device has reached steady state. Moreover, a user-defined ultimate stop condition is implemented in the algorithm in case the t-metric proves unable to drop below the t-test threshold.

Once Stage I of the algorithm detects that resistive state is stable, it proceeds to Stage 2, which further checks the 'retention' of this new equilibrium state. In this phase, the resistive state of the DUT is read for one minute following which the module is concluded. The output of the volatile module determines two things. Firstly, the module records the time-stamps at the start and end of each batch so that the system can determine how long the t-test lasted for. Secondly, by comparing the volatile and non-volatile changes against each other as in Fig. 3.16 b,c we can easily determine the voltage range under which the devices can be safely operated in purely volatile mode.

The volatility algorithm was applied on stand-alone RRAM arrays of electroformed devices. The employed devices are electroformed using a ramp of pulses of negative polarity and the electroforming voltages of the utilised TiO_x devices for these experiments is $\approx$ -4 V. The devices are formed in the resistive state range of 500 kΩ to 2 MΩ range. Subsequently, the devices are characterised using the volatility module and the batch size is fixed at 1000. V_w varies from -0.2 V up to -3 V and the T_w is chosen to be 100μs. The V_r is fixed at 0.2 V and the T_w is automatically determined by the system. One batch of thousand read points lasts for $\approx$ 1s and k is fixed at 10.

3.7.2.3 Results and Discussion

This section presents results obtained from volatility algorithm implemented on solid-state TiO_x devices as illustrated in Fig. 3.16. Figure. 3.16 a shows the pulsing history of a 60x60 μm^2 device as progressively higher voltages are applied in steps of 0.5 V. Four different instances of resistive state measurement for each step of voltage are represented i.e. I, II, III and IV. Point I and II marks the assessment of the DUT before and after the stimulus is applied whilst III and IV indicates the resistive state measurement after the completion of T-test and retention phase respectively. Table. 3.7 indicates the read values of the first and last batch as recorded at points II and III for -2V respectively. For lower programming voltages i.e from -0.5 V upto -1.5 V there is no significant change in the resistive state of the DUT. However, for input stimuli of -2 V and -2.5 V a metastable resistive state change of $\approx$ 100 kΩ can be noted following which the DUT stabilises to a different resistive state. For example, for -2 V steady-state (IV) is $\approx$ 0.54 MΩ which is slightly different in comparison to its initial resistive state (I) of $\approx$ 0.59 MΩ. Notably, the similar behaviour is evident across the entire family of employed TiO_x devices.

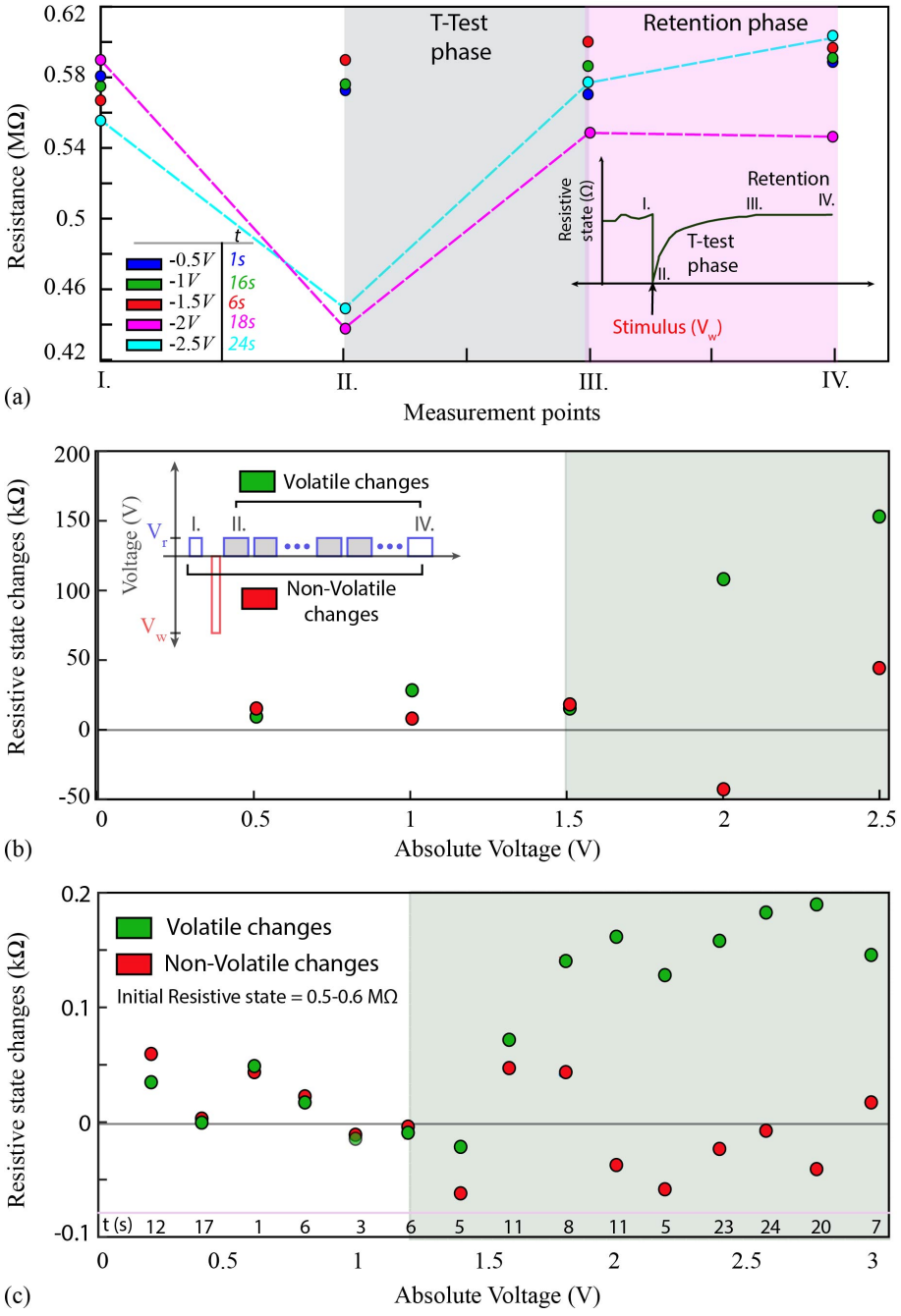


Figure 3.16: Volatility characterisation. (a) Resistive state measurements for the DUT. The grey and pink band indicates the t-test and retention phase respectively. The presented time (t) in the inset is the duration the DUT takes to exit the t-test phase. (b) Volatile (green colour) and non-volatile (red colour) resistive state changes for the same DUT as in (a). (c) Same as (b) for a $200\,nm^2$ memristive device. Grey band in (b) and (c) indicates the threshold of the DUT.

Table 3.7: Read values for round -2V in Fig. 3.16 a. All values are in kΩ.

	Initial batch, Point II		Last batch, Point III	
	k_1	k_2	k_1	k_2
1.	430.430	520.741	541.548	539.134
2.	430.779	517.044	541.713	543.060
3.	430.065	514.365	543.308	545.988
4.	429.221	515.162	542.657	547.596
5.	431.048	515.226	543.320	541.090
6.	430.563	514.206	545.474	540.972
7.	424.894	516.916	542.598	540.937
8.	427.879	514.152	542.468	545.330
9.	429.672	517.625	541.007	539.426
10.	430.802	511.324	541.043	538.272
$\|t - testvalue\|$	94.5		0.365	

For Fig. 3.16 b, non-volatile and volatile resistive state changes for the same device are estimated. For this, resistive state changes between the steady state attained after the completion of the retention phase i.e. point IV and resistive state measurements before and after the application of stimulus i.e. point I (non-volatile change) and II (volatile resistive state change) respectively are calculated. Evidently -1.5 V marks the threshold for the volatility region below which there are no changes in the resistive state of the DUT and above which clear volatile changes can be noted. Similarly, volatility was characterised for 200 nm^2 memristive devices where voltages with smaller resolution of 0.2 V was applied and the results are illustrated in Fig. 3.16 c. The threshold of the DUT in this case was found to be in the region of -1.2 V and can be understood as a soft-threshold. For input stimulus, -1.4 V and -1.6 V the resistive state changes are small in comparison to the higher voltages where the changes are evident. The resistive state changes in the sub-threshold region are considered as insignificant and are mainly attributed to background fluctuations caused due to the measurement noise. In contrast, for voltages higher than -3 V for 100 μs pulse width, the devices tend to destabilise and drift to higher resistive state range resulting in abrupt resistive state transitions or in some cases the devices switch and change their baseline operating range to less than 100 kΩ range thus eventually operating in the non-volatile region.

Importantly, in an ideal case the supra-threshold non-volatile resistive state changes i.e. the red circles in Fig. 3.16 b,c should approximately be zero. However, in reality this is not the case for example in Fig. 3.16 c at 2.2 V, 2.8 V, the DUT demonstrates an over-convergence whereas at points 1.6 V, 1.8 V the DUT seems to have under-converged. However, from the variance in the distribution of the non-volatile changes presented in Fig. 3.16 b,c it is sound to hypothesize that the illustrated non-volatile changes are centered around zero and thus the devices can be purely operated in the volatile region. From Fig. 3.16 for the used parameters, it can be concluded that the safe region of operation of employed TiO_x devices to obtain volatility is ≈-1.5V upto ≈-3V, where the time measurements for the t-test phase are indicated in Fig. 3.16 a,c. Thus through this protocol, the voltage range for pure volatile operation in memristors (TiOx-based) can be determined in an automated manner. This is

crucial for engineering the volatility property of memristive devices towards building biophysically realistic neural processing systems regardless of specific memristor implementation.

3.8 Summary

This chapter introduced the RRAM devices along with the developed automated characterisation protocols essential to identify functional non-volatile and volatile devices. In the protocol for non-volatile devices, a testing algorithm based on a two mean t-test is proposed. The test uses custom-made hardware to apply a stimuli which drive each device of a RRAM crossbar array in distinct resistive states (HRS and LRS) and determines switchability on the basis of statistical significance between resistive state readings obtained in the two states. Further DUT examination indicated that functional devices typically exhibit low p-values (< 0.1). This method effectively helped overcome device characterisation testing bottlenecks faced during RRAM process development, thereby accelerating both process development cycles and the pace of understanding of established and emerging memrisitve devices and materials alike [117]. In the second protocol for volatile devices, an algorithm capable of characterising volatility i.e. the metastable resistive state changes in RRAM- TiO_x devices in an automated manner using the standard t-test is proposed. The output of the module determines the operating range of devices to exploit volatility and the time required for the devices to relax to its equilibrium resistive state under the chosen set of characterisation parameters. Moreover, the obtained data can be extrapolated to obtain the non-volatile and volatile curves for the DUT.

Chapter 4

Spike detection using metal-oxide devices in non-volatile regime

4.1 Introduction

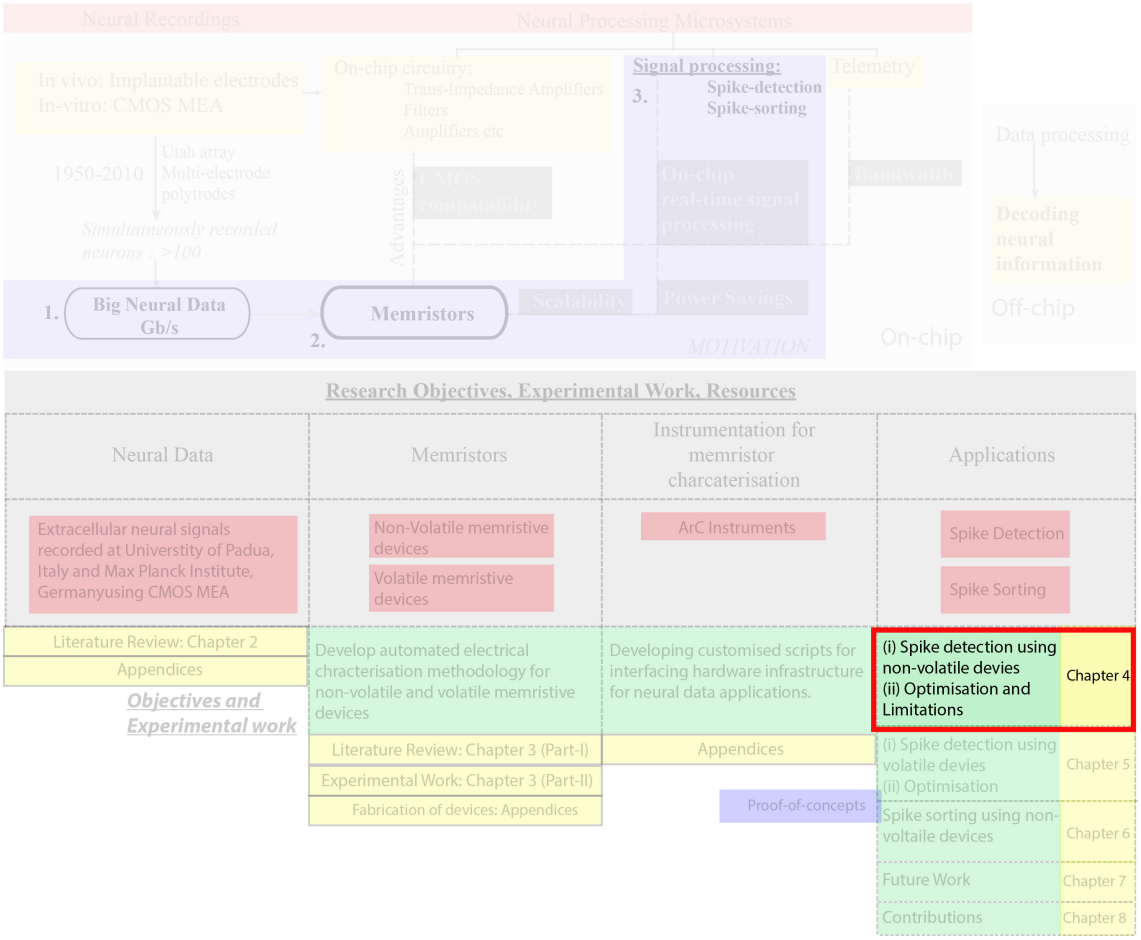


Figure 4.1: The research objective of this experimental chapter.

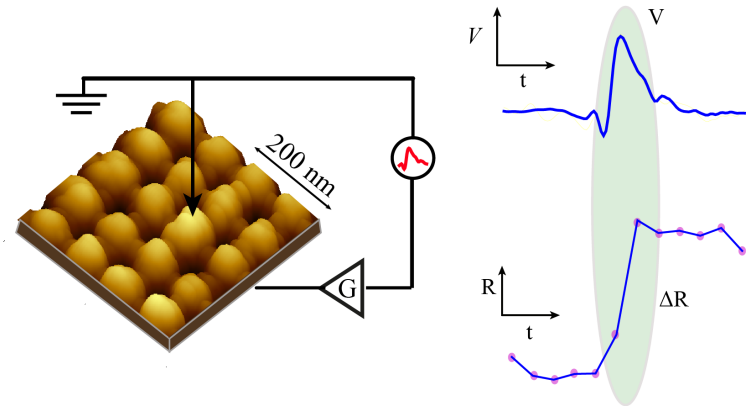


Figure 4.2: Concept figure: The basic concept of memristive integrating sensors revolves around transforming the action potentials, or signals into resistive state changes of the device-under-test. Importantly, only the signals above the threshold of the device would cause a change in resistive state of the device.

This chapter presents the proof-of-concept of spike detection using non-volatile metal oxide devices. Spikes from a large number of neurons have lead to the issue of 'big data' that require on-line processing under most stringent conditions, such as minimal power dissipation and on-chip space occupancy [12]. In this chapter, it is shown how emerging memristive devices capable of resistive switching appear to be well suited for 'spike-detection'. Intrinsic properties of metal-oxide TiO_x memristors, such as their analogue memory capacity occurring above certain voltage thresholds are found to be useful for encoding and compressing neural spiking activity recorded by MEA's.

The chapter is organised as follows: Section 4.2 demonstrates the behaviour and electrical characteristics of memristors suitable for spike-detection, Section 4.3 shows the memristor-based spike-detection platform and the employed signal processing strategy throughout the experiments, Section 4.4 illustrates benchmarking of memristor-based spike-detection platform against state-of-the-art template matching system. Section 4.5 discusses the concept of scaling the introduced spike-detection platform to a multi-channel array level and integrating memristors in high density neural recording platforms. Section 4.6 investigates the approaches for improving the detection accuracy of proposed spike-detection platform. Section 4.7 discusses the power consumption, performance limiting factors and possible measures for further improving the detection accuracy, and finally section. 4.8 summarises the chapter.

4.2 Memristive Integrating Sensors (MIS)

As originally proposed by Chua, memristors are capable of changing their resistive state as a function of the integral of their input voltage; a phenomenon known as 'resistive switching' [77]. Solid-state TiO_x memristors with a metal-insulator-metal architecture, as shown in Fig. 4.3(a) inset, were fabricated on a Si/SiO_2 substrate; detailed process parameters appear

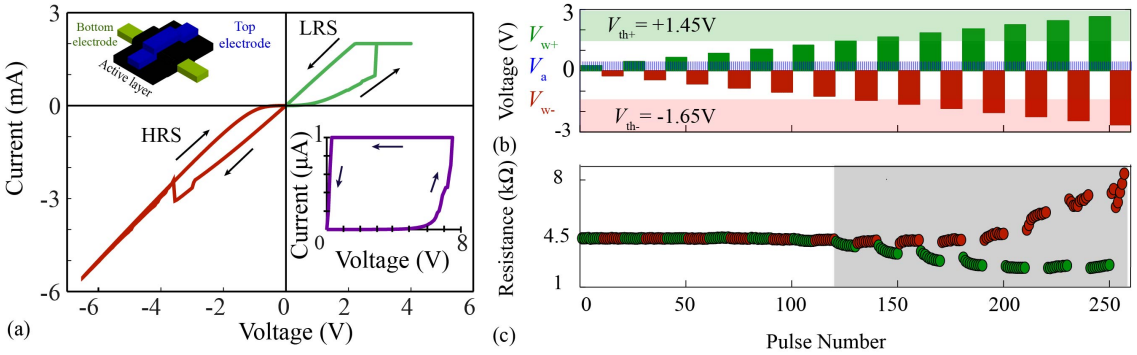


Figure 4.3: Device architecture and electrical characterisation of solid-state TiO_x RRAM devices.(a) Inset in the first quadrant represents a schematic illustration of a TiO_x RRAM memristor. The inset in the fourth quadrant shows the 'electroforming step' where the device forms to 'ON' state under positive bias, shown in purple colour. After electroforming, the device exhibits bipolar switching with low resistive state (LRS, illustrated in green) to high resistive state (HRS, illustrated in red) transitions at $\approx \pm 3$V. RESET occurs under negative bias whilst the complementary SET transitions occur under positive bias. (b,c) Pulse based characterisation of the DUT. Resistive state changes (bottom trace) in response to the input pulses above threshold, V_{w+} and V_{w-} indicated in green and red circles respectively. Read pulses of amplitude V_a are indicated in blue. In bipolar devices two inherent thresholds exist, one for each voltage polarity. For this device, we obtained V_{th+} = +1.45 V and V_{th-} = -1.65 V as indicated by the green and pink shadowed areas of the plot respectively.

under Appendix A [118][1]. Fig. 4.3(a) illustrates DC characterisation of the electroformed device (inset), demonstrating a typical pinched-hysteresis memristive device signature. For electrical characterisation voltages are applied to the top electrode (TE) whilst the bottom electrode (BE) is grounded for all the measurements. As shown in the figure, in the used TiO_x memristors, the devices 'post-electroforming' demonstrate reversible switching where the devices switch to 'low resistive state (LRS/SET)' with positive polarity and 'high resistive state (HRS/RESET)' with negative polarity[2].

Interestingly, in the normal operating regime where the device supports reversible bipolar resistive switching, intrinsic voltage threshold (denoted as V_{th-}/V_{th+}) accounts for the response to voltage pulsing events. This memristive behaviour and fundamental property of the device is apparent in Fig. 4.3(b,c). Subjecting the Device-under-Test (DUT) to a train of input programming pulses in alternating polarities gives rise to gradual resistive state transitions, provided the pulse amplitude exceeds the devices inherent bipolar switching thresholds. For instance for the DUT in Fig. 4.3(b,c), the device switches to HRS/RESET and LRS/SET with negative and positive polarity respectively, only after the voltage of the stimulus exceeds the inherent thresholds of the DUT, here identified as $V_{th+} = 1.45$ V and $V_{th-} =$ -1.65 V respectively. In Fig. 4.3(b,c), the train of input programming pulses was applied at 100 μs (write operation) and the device memory state was read after each programming pulse at $\approx$0.5V (read operation).

[1]The devices used for the experimental work were not fabricated by the author of this thesis.

[2]Electroforming typically occurs at $\approx$+6.5 V and bipolar switching is usually observed below $\pm$3 V as shown in Fig. 4.3

This capability for gradual switching can be exploited to encode multiple significant spiking events as small changes in a devices resistive state. This assumption is first explored deterministically, by employing known pulse events. Fig. 4.4 shows the response of a typical DUT to trains of 200 identical square-wave events of negative and positive polarities respectively, as illustrated in the insets. Each writing pulse has a fixed 100 μs duration and suitable amplitude to induce a resistive state change. It is followed by a reading pulse of fixed 0.5V amplitude and an automatically determined duration, t_a [119]. Notably, the pulse amplitude required to elicit a resistive state change of similar strength but in the opposite direction could differ, indicating an inherent asymmetry in the devices characteristics. This bidirectional, gradual (analogue), saturating switching, could be fitted by second order exponential functions of input voltage integral thus defining the input-output relation of an integrating sensor for distinct stimulation protocol. For the DUT, curve-fitting was carried out using standard curve-fitting tool in MATLAB. The data from the resistive state of the devices for negative (Fig 4.3(c)) and positive (Fig 4.3(c)) pulses were separately fitted to second-order exponential function, that is f($\int$Vdt) = A$e^{\beta \int V dt}$ + B$e^{\gamma \int V dt}$, where V is the fixed pulse voltage indicating non-volatile resistive states transitions. The goodness of fit (R-squre) for Fig. 4.4(a) and Fig. 4.4(b) was found to be 91.37% and 97.01% respectively.

Notably, as TiO_x prototype act as thresholded integrator, hereafter the device is named as 'Memristive Integrating Sensor' (MIS) and in the following sections it is shown how the thresholded-integrator attribute can be particularly useful for compressing information and suppressing noise in signals with low signal-to-noise ratios (SNR), such as data recorded from the activity of neurons/cells. This approach allows for significant, supra-threshold events to be registered as measurable changes in the device memory state, whilst sub-threshold events are inherently suppressed thus exploiting metal-oxide based resistive switches as neuronal spike integrators. It is important to note that the thresholded integrator behaviour is preserved only within a given range of input voltages (typically $\pm 3\,V$), above which the fabricated set of TiO_x devices tend to show behavioural changes indicative of physical damage (abrupt changes in resistive state and voltage operating ranges).

4.3 Memristors as encoders of neuronal spikes

The natural ability of memristors to integrate significant events provides an efficient way of encoding and compressing information on neuronal firing in real-time, as recorded by neuronal probes. In this section, the basic concept of the memristor-based spiking platform is exemplified in sub-section 4.3.1, signal-processing strategy is described in sub-section 4.3.2 and finally the method of estimating the spikes-detected by the memristor-based platform is reported in sub-section 4.3.3.

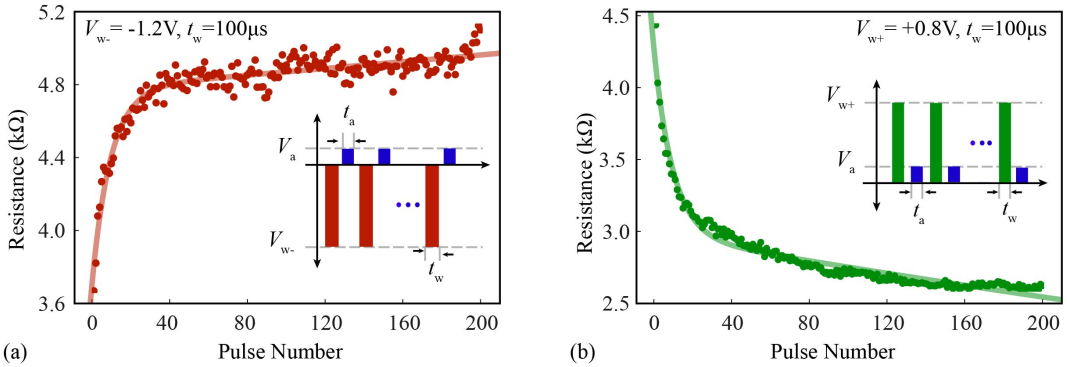


Figure 4.4: (a), (b) Gradual resistive switching under a pulse train stimulation (200 pulses per train) as illustrated in the respective insets. The DUT response is fitted with a second order exponential function (continuous line). Typical biasing scheme parameters (insets): negative write pulse voltage V_{w-} = -1.2 V, positive write pulse voltage V_{w+} = +0.8 V, read pulse voltage V_a = 0.5 V, write pulse width t_w = 100 μs and read pulse width t_a automatically determined by the measurement system.

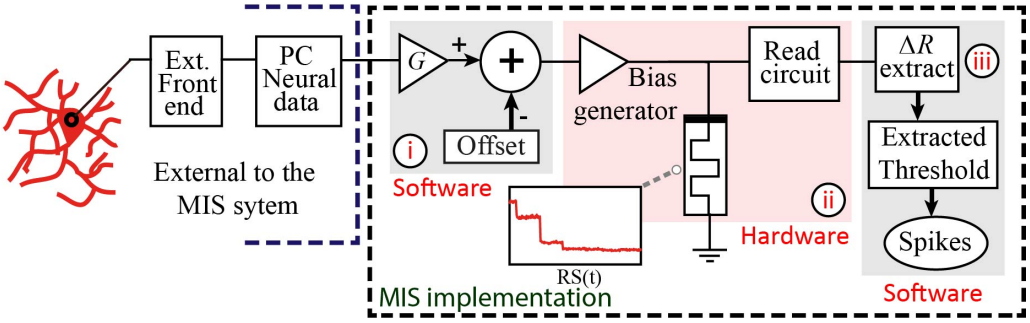


Figure 4.5: Memristive Integrating Sensors (MIS) concept and operation. Block diagram of the signal processing in the proposed spike-detection system. An external front-end (a CMOS Multi-Electrode-Array (MEA) system) located **'externally'** to the MIS platform records extracellular neuronal signals and amplifies them. The pre-amplified, acquired neural recordings are then fed into the characterisation instrument, suitably gain-boosted (G) and offset (V_{off}) to render them compatible with the memristors voltage operating regimes (i). The conditioned waveform is fed into a memristor and its resistive state is then periodically assessed (ii). Changes in resistive state caused by spiking events are extracted offline (iii).

4.3.1 Neural spiking compression with memristors

An overall block-diagram of the system used in the neural signal processing through memristive-integrating sensors is illustrated in Fig. 4.5. An 'external' front-end platform[3] which is fed into the MIS system as a series of voltage-time samples. The signal-chain in MIS platform begins by pre-amplifying the incoming signal to voltage levels suitable for operating the memristor sitting at the core of the MIS and then proceeding to apply the pre-amplified signals to the memristor in real-time. Periodically, the memristor's resistive state is assessed and when a significant change in comparison to the previous state is detected, the system registers a spiking event. This is more clearly explained as follows:

[3]The details of the front-end system used for recording the electrical activity from retinal ganglion cells are detailed in Appendix C, *Neural Data Acquisition System*.

External Front-end System

1. Fig. 4.5 (purple dashed line) - MIS system implementation was experimentally validated on spiking activity of retinal ganglion cells pre-recorded by an external Multi Electrode Array (MEA) front-end system [47](see Appendix C). The MEA employed records the raw bio-signals, which lie in the 0.1 mV-1 mV range, and then uses its own in-built amplifiers to boost them to the 10 mV-500 mV range. The resulting recordings are then stored off-line as voltage-time series.
2. The stored recordings are used as inputs to the MIS platform in isolation from the front-end, that is the front-end has not been connected to the MIS platform in real-time.

MIS Implementation

3. Fig. 4.5(i) - The processing of neural signals through MIS platform begins when the stored voltage-time series are subjected to amplification (G) and offset (V_{off}) in software on the PC that runs the MIS platform. The parameters are chosen such that only the most significant events (large amplitude extracellular spikes) would exceed the inherent threshold (V_{th-}/V_{th+}) and noise would be automatically suppressed.
4. Fig. 4.5(ii) - The resulting, pre-conditioned waveform is then transmitted from the PC to the memristor testing and operation instrument (see Appendix B), which physically implements the MIS system. The instrument, in turn, plays back the waveform to a target memristive device. In order to assess the distinct resistive state changes during the streaming of recorded neural activity, the DUT is periodically disconnected from the neural signal feed, and connected to a read-out circuit that captures the devices resistive state, digitises it and subsequently sends the resistive state reading back to the PC. Importantly, only a limited amount of data is returned to the PC when compared to the full voltage time series in conventional systems.
5. Fig. 4.5(iii) - A software converts the incoming series of resistive state readings into a series of resistive state changes, subsequently keeping only the largest ones that mark significant events in the neural signal. These significant resistive state changes are estimated as the 'spikes' detected by the MIS platform. Noteworthy this filtering process, based on an assessment of resistive state changes in absentia of an input signal may be engineered in order to fine-tune SNR on neuronal activity.

4.3.2 Signal-conditioning methodology

Each neural recording was fed to an in-house developed memristor characterisation instrument, as described in Fig. 4.5(ii). The customised hardware handled the software-implemented linear gain and offset conditioning operations, electrically interfaced test memristors and carried out the DUT resistive state assessment procedures. Neural signal voltage time-traces ($\approx$63k samples) as illustrated in Fig. 4.6(a), blue trace were fed into the target device in batches of 1000 data-points as depicted in the yellow block. Each data sample

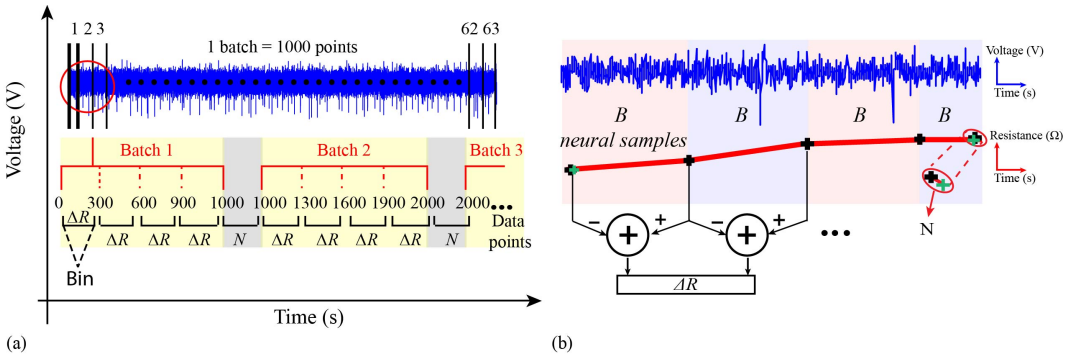


Figure 4.6: Standard read-out scheme for evaluating the time evolution of the resistive state of test devices subjected to input neuronal signal. (a) A single neural recording (blue trace) contains exactly 63016 data-points. Input data is fed to the devices in batches of 1000 data points with resistive state of the device being assessed at the beginning of each batch, then every 300 samples (termed as bin, B) and finally at the end of each batch. A single neural recording consists of 63 batches. Changes in resistive state (ΔR) are extracted from consecutive resistive state assessments. Resistive state changes occurring between the last measurement of each batch and the first measurement of the next batch, i.e. with no interceding pulse biasing, (N) can be used to determine the 'noise band'. (b) Conceptual read-out scheme for evaluating the time evolution of the resistive state of test devices subjected to input stimulation for one batch. Assessment points are marked by crosses.

lasts 82 μs and the sampling frequency is 12.2 kHz. Within each batch resistive state was assessed at the beginning of each batch, then every 300 samples and at the end of the batch (standard scheme: assess initial resistive state and after application of the 300th, 600th, 900th and 1000th data-points). Subsequently, changes in DUT resistive state (ΔR) can be extracted from pairs of consecutive resistive state readings, whilst resistive state changes occurring between the last measurement of each batch and the first measurement of the next batch, that is with no interceding pulse biasing, (N, grey block) provide an estimate of measurement uncertainty thus generating the noise band. The conceptual read-out scheme for a single batch is more clearly presented in Fig. 4.6(b). A flow diagram has also been added as illustrated in Fig. 4.7.

Notably, since the events are transduced as non-volatile resistive state transitions one can afford smaller sampling rates that benefits further time-resolution data-rate. Thus, by following the described signal processing strategy for a single neural recording we obtain 316 ΔR values, corresponding to 252 ΔR bins and 64 noise level samples.

4.3.3 Spike detected in MIS-based platform

The obtained ΔR bins and noise level measurements are post-processed to filter out 'significant' resistive state changes from the 'insignificant' resistive state changes. To realise this concept consider Fig. 4.8. An input signals where neuronal spikes span both negative and positive voltages was tested including a repeatability check. In the case of *in-vivo* recordings, spikes often span both negative and positive voltage polarities, depending on experimental

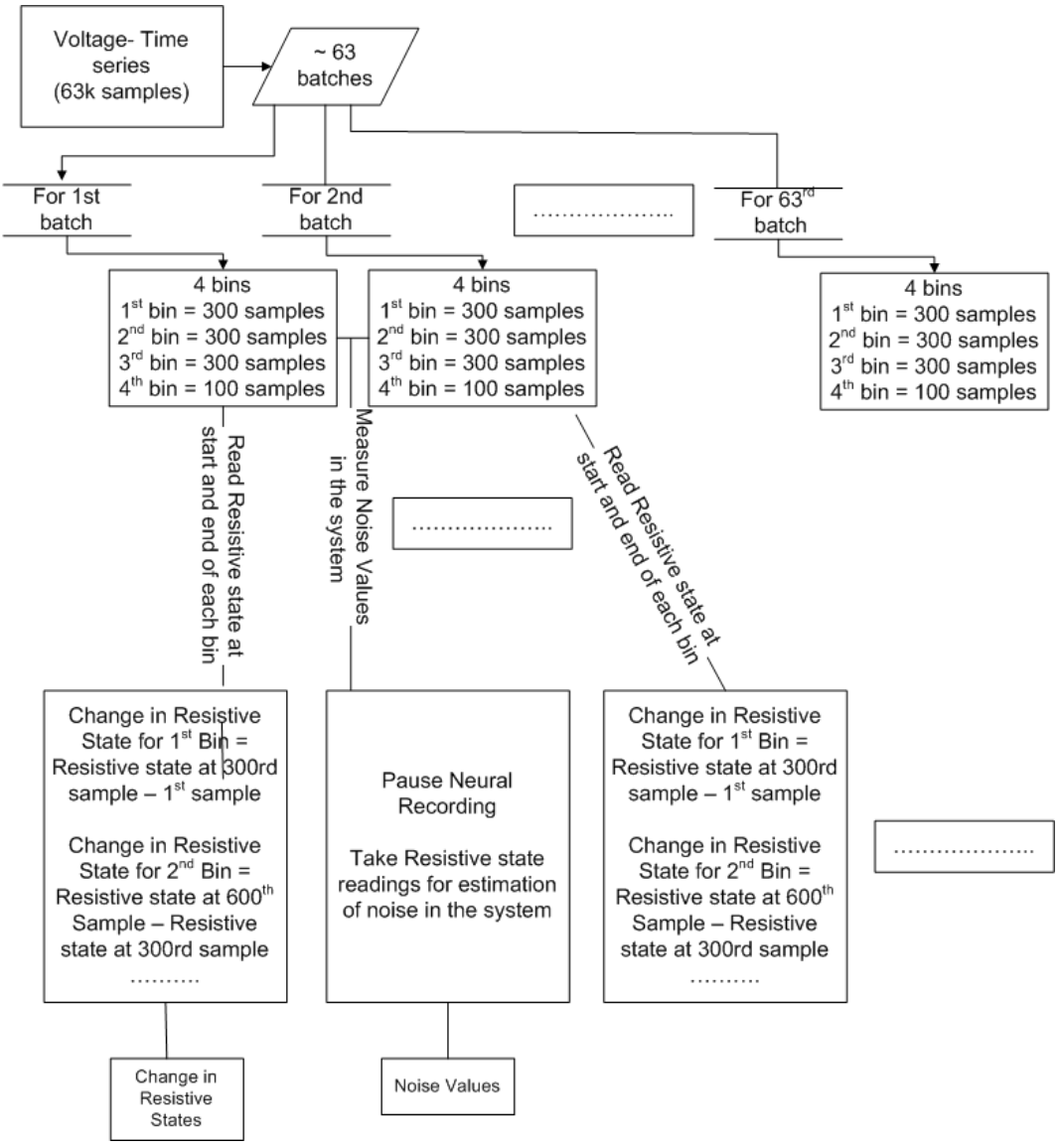


Figure 4.7: Flow chart for signal processing methodology

conditions, for example the position of the recording electrode relative to neuronal compartments and their associated ionic conductances [120]. It is thus relevant to demonstrate the MIS operation for both signal polarities, as explored here at a proof-of-concept level.

Figure. 4.8(a) depicts a waveform consisting of four, concatenated copies of the same retinal recording. Each copy was subjected to appropriate scaling and offsetting, and two of the copies were polarity-inverted. The corresponding resistive state transient response throughout this test is shown in Fig. 4.8(b). Significant changes in resistive state correspond to clear, supra-threshold events. It can be noted that spike detection successfully occurs at both polarities, achieving qualitatively similar modulation over two signals. Consequently the normalised resistive state changes between consecutive reads is plotted as a function of the maximum voltage magnitude of interceding events Fig. 4.8(c). The resistive state changes at x=0 within the grey band correspond to noise measurements. All the resistive state changes

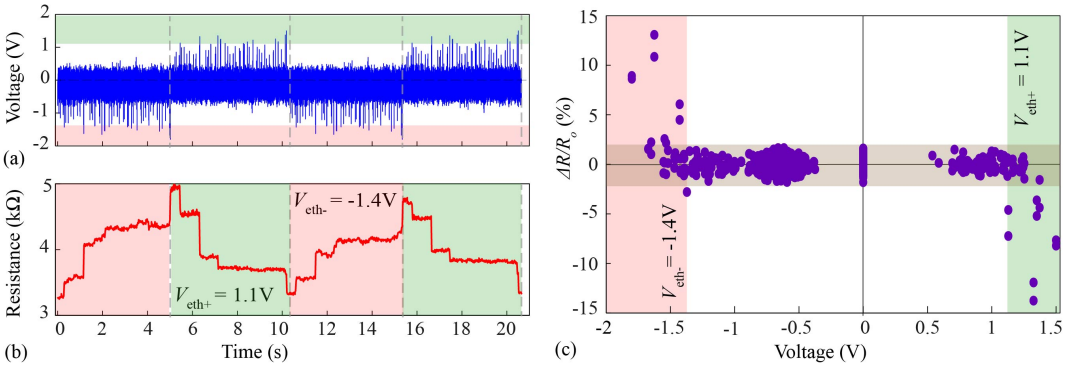


Figure 4.8: (a),(b) (Upper Trace) An arbitrary input waveform consisting of four concatenated copies of the same retinal cell neural recording and artificially inverted to produce spike trains with alternating polarities is used. (Bottom Trace) The response of DUT in response to the waveform shown in (a). The collated recording copies in (a) have been subjected to appropriate scaling and offsetting in order to accommodate the devices asymmetric threshold voltages, resulting in balanced resistive state SET and RESET. (c) Fractional resistive state modulation ($\Delta R/R_o$) extrapolated from (b) showcasing significant resistive state modulation occurring only above V_{eth+} and below V_{eth-} while intermediate bias values (noise, N) leads to no significant change. The extracted threshold voltages are identified here as, V_{eth+} = 1.1 V and V_{eth-} = -1.4 V represented in the green and pink band respectively.

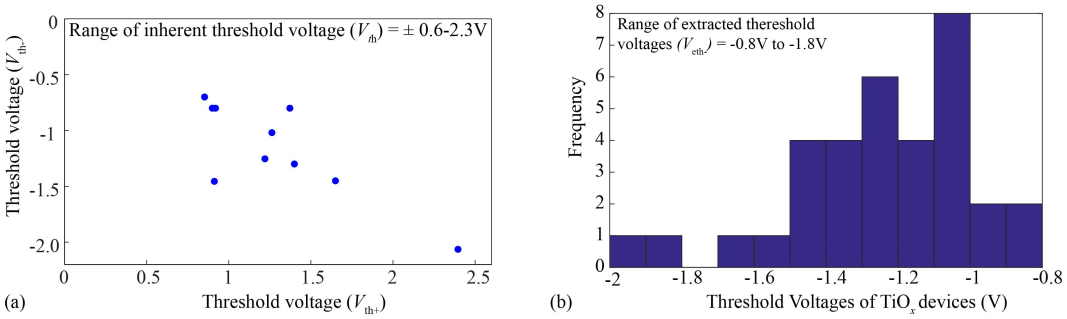


Figure 4.9: Inherent and effective threshold of the memristive devices. (a) Range of the 'inherent threshold voltage (V_{th-}/V_{th+}) is approximately ± 0.6-2.3V. (b) Distribution of effective threshold voltage (V_{eth-}/V_{eth+}) of the TiOx devices used. Range of the effective threshold voltages used for the experiment is -0.8V to -1.8V.

within this band are discarded and the remaining points are used to define the memristor's effective operating threshold voltages (V_{eth+}/V_{eth-}) which partition the plot into three distinct areas: two of them correspond to significant resistive state modulation (larger than V_{eth-} and less than V_{eth+}) and the last one ($[V_{eth+}, V_{eth-}]$) containing resistive state changes that are indistinguishable from the estimated background noise.

Importantly, whilst the inherent threshold of the device performs a coarse filtering action, the effective threshold ultimately determines SNR. The range of inherent threshold voltage is approximately ± 0.6-2.3 V as determined by the sample of data obtained from ten devices (Fig. 4.9(a)). This value is used as an approximation for the setting of gain and offset parameters for the processing of neural recordings before it is used for biasing the memristive device. Range of effective threshold voltages for the TiO_x prototypes employed throughout this study was 0.8V to -1.8V. The histogram in Fig. 4.9(b) is plotted for a sample of 34

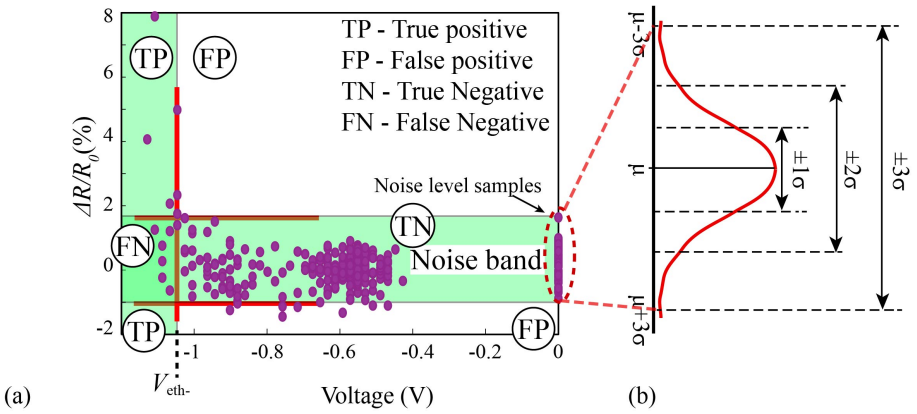


Figure 4.10: Concept of estimation of effective threshold voltage (V_{eth-}) and quantification of data obtained from the MIS platform. (a) Normalised plot of resistive state changes ($\Delta R/R_o$) values in each bin is plotted as a function of maximum voltage magnitude of interceding events. (b) Illustration of 6 sigma method used for the setting of the 'noise band limits'. Quantification parameters: True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN).

devices. Moreover, since the MIS system detects normalized changes in the resistive state rather than absolute values, the device variability is heavily compensated for such that MIS operation is routinely available.

Furthermore, the limits of the noise band in (Fig. 4.8(c)) are set using the 6 σ method that is, mean (μ) $\pm$ 3*standard deviations (σ) of noise level samples (Fig. 4.10(b)). The noise band consists of noise samples and insignificant resistive state changes arising as a result of weak amplitude neural signals which cannot be differentiated from the resistive state changes arising due to noise samples and hence discarded. Everything outside the noise band is considered as a significant resistive state change.

Notably, the estimation of V_{eth-} is determined by the noise band limits which classifies the data in four groups as conceptually illustrated in Fig. 4.10(a). In a simple thresholded integrator, everything above the threshold is assumed to be a spike and everything below is suppressed as noise. However, due to existence of noise band in the MIS platform, the data in MIS platform is quantified as follows: Outside the noise band and below V_{eth-}: True Positives (TP), outside the noise band and above V_{eth-}: False Positives (FP), inside the noise band and below V_{eth-}: False Negatives (FN), inside the noise band and above V_{eth-}: True Negatives (TN), shown in Fig. 4.10(a).

4.4 Benchmarking

The performance of the proposed MIS concept was benchmarked against a state-of-the-art template matching system [40] (details in [4, 121]). The front-end system for recording neural data from retinal ganglion cells (cell culture) remains the same for both the systems. The first resulting performance comparison between the two approaches is presented in Fig. 4.11.

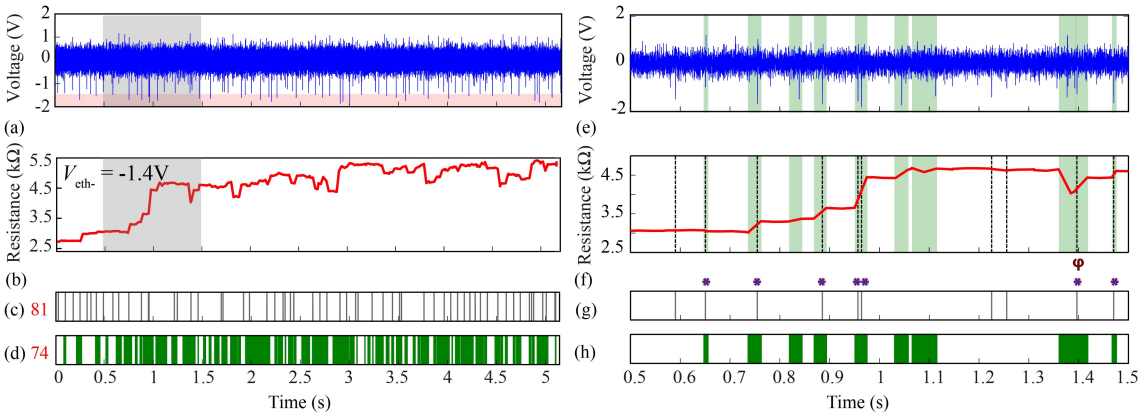


Figure 4.11: Benchmarking memristor-based system against state-of-art template matching system. (a) A pre-conditioned neural recording trace with gain and offset value of 2.8 and 0 respectively, causes the resistive state time evolution shown in (b). (c) 81 spikes were detected by the template matching system, with grey lines indicating spike positions. (d) Green shading indicates time intervals within which one or more spikes were detected through the MIS; total of 74. (e) and (f) are close-ups of the neural recording and resistive state evolution shaded grey in (a) and (b) respectively. Time intervals where the MIS detects spikes are shaded green whilst the locations of spikes detected by the template matching system are indicated by grey vertical dashed lines. (g) and (h) Comparison of the detection of spikes by the two systems. The asterisk mark (*) indicates agreement between the two systems and the ϕ symbol indicates the resistive state drop associated to the occurrence of a large-amplitude positive event.

In this case, an offset ($V_{off} = 0$) and amplification (G = 2.8) on the recording was used as shown in Fig. 4.11(a) and the devices resistive state was assessed as per the standard signal processing scheme as described in section 4.3.2. Figure. 4.11(b) illustrates the resistive state evolution of the tested MIS device in response to the input signal shown in Fig. 4.11(a). One can observe clear changes in the devices resistance corresponding to spiking events whose magnitude exceeds V_{eth-} (spiking events in pink band in Fig. 4.11(a)).

In this particular example, the incoming spikes mainly occur in negative polarity hence there is an overall increase in resistance, from approximately 2.5 kΩ to 5.5 kΩ. However, the presence of a few events in opposite polarity that exceed V_{eth+}, cause occasional resistive state drops. A clear example indicated by ϕ in Fig. 4.11(g,h) can be observed at ≈1.4s in Fig. 4.11(a,b) and Fig. 4.11(e,f) where the resistive state reduces from ≈4.5 kΩ to 4 kΩ indicated by 'ϕ'. However, optimising the value of V_{off} provides additional flexibility for compensating for this effect. Another point to note is as the MIS is capturing and storing significant events as non-volatile resistive changes, one can afford to use relatively low sampling rates for minimizing the overall requirements in data storage/handling. Along this line, the output of this system is quantified at discrete time bins containing one or more detected events (see for example Fig. 4.11(e, f-h) at approximately 0.96s). For this recording, MIS system identified 74 bins containing significant events denoted in Fig. 4.11(d), whilst the template-matching-based system detected 81 significant events shown in Fig. 4.11(c).

Comparing the MIS and template-matching approaches within a representative time-window of 1 s duration in Fig. 4.11 (g and h), indicates a similar performance in spike detection at

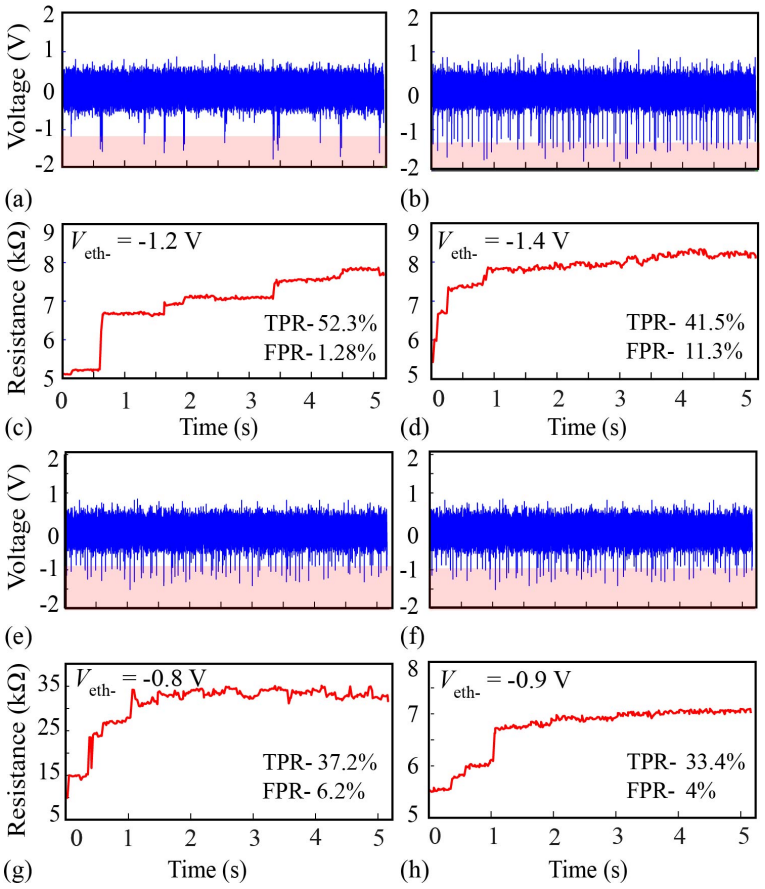


Figure 4.12: Robustness of memristive devices. (a), (b), (c), (d) Response of a single memristor to two blocks of neural recording data containing significantly different patterns of activity shown in (a) and (b). The pink band indicates the extracted threshold of the device-under-test (V_{eth-}). The respective figures indicate quantification parameters, that is rate of true positives (TPR) and false positives (FPR) respectively. (e), (f), (g), (h) Response of two different devices to a common reference block of neural recording data shown in (e) and (f). Initial resistance of device was set to approximately 5 kΩ. In all these experiments, signal conditioning parameters, that is software added gain and offset remained fixed at G = 2.2, $V_{off} = 0$ respectively.

majority of instances, as noted via asterisk symbol (*) marks. It is interesting to note that MIS approach results into registering apparent events (for example at 0.83s, 1.05s, 1.1s) that are missed by the template-matching method while it fails recognising other possible events (for example at 0.59s, 1.22s, 1.25s), presumably due to a conservative selection of signal conditioning gain.

Furthermore, the robustness of the observed behaviour and its potential for data compression via improvement of SNR is further demonstrated by experimenting: a) the response of a single MIS to blocks of neural recording data containing significantly different patterns of activity (Fig. 4.12 (a,b,c,d)) and b) the response of different devices to a common neural recording obtained from MEA as illustrated in Fig. 4.12 (e,f,g,h). In the former one device, many recordings case it can be observed how intense activity leads to a larger overall resistive state modulation and how particularly strong events tend to cause distinct non-volatile changes in memory states. Thus, resistive state traces compress information on both the firing rate and

Table 4.1: Spike count and quantification parameters for the neural recordings in Fig. 4.12 in comparison to the state-of-the-art template matching system (TMS). Spikes detected by Memristive Integrating Sensor (MIS) and Template Matching System (TMS). True Positives: TP, False Positives: FP, True Negatives: TN, False Negatives: FN.

S.No.	TMS	MIS	TP	FP	TN	FN	Rate of TP (%)	Rate of FP (%)
One device-to two neural recordings (Fig. 4.12 a,b,c,d)								
Recording I	20	14	11	3	230	3	52.3	1.28
Recording II	77	52	32	20	156	45	41.5	11.3
Many devices-one neural recording (Fig. 4.12 e,f,g,h)								
Device I III	78	40	29	11	164	49	37.2	6.2
Device 2 IV	78	33	26	7	168	52	33.4	4

spikes amplitude. Instead, in the latter one recording, many devices case we observe that despite the quantitative variability in device behaviour, most of the marked resistive state transitions tend to concur in time with significant events present in the input waveform.

To evaluate the quantification parameters and understand the efficiency of the MIS system the rates of True Positives (TPs) and False Positives (FP) are estimated [4]. For this, initially spikes detected by the template matching system are collapsed into time bins similar to MIS system. For instance the output of TMS system in Fig. 4.11(c) is 81 spikes. This collapses into 78 bins. Subsequently, each time bin is examined in order to determine whether the TMS and MIS system yields the same results. For the purpose of benchmarking it is assumed that TMS is a perfect spike detector and therefore TP, TN (True Negatives), FP and FN (False Negatives) are redefined. TP now means that both TMS and MIS system detect activity in a given bin. Similarly, TN: systems agree there is no activity in a given bin, FP: MIS registers activity whilst TMS does not, FN: MIS registers no activity while TMS does.

$$\text{Rate of } TP = \frac{TP}{TP + FN} \tag{4.1}$$

$$\text{Rate of } FP = \frac{FP}{FP + TN} \tag{4.2}$$

Quantification parameters for Fig. 4.12a-h are estimated using equations 4.1 and 4.2 and are indicated in Table. 4.1.

[4] It is important to note that in these experiments, 'Template Matching' based approach is used as a ground truth for determining the rate of true and false events.

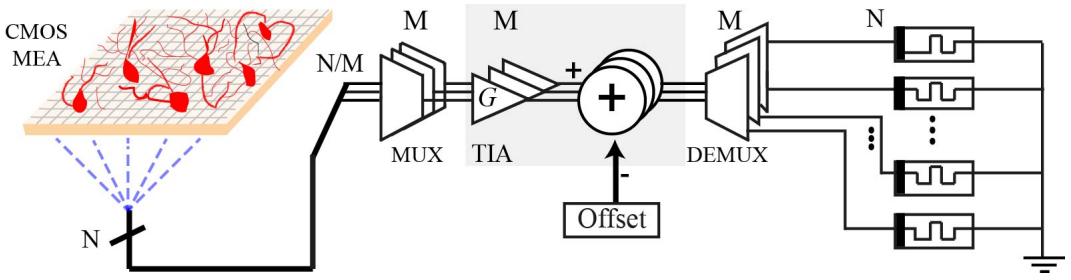


Figure 4.13: Towards array level integration. (a) Conceptual diagram indicating conditioning of data from N pixels through multiple gain and offset cascade (M). CMOS MEA: Multi-transistor Array Block, manufactured in standard, commercially available CMOS technology. TIA: Trans-impedance amplifier converting current to voltage with appropriate amplification.

4.5 Conceptual implementation at multi-channel array level

Efforts to reduce data bandwidth echo current research in on-chip spike-sorting [12, 19], where the proposed memristor-based approach can prove to be disruptive in exploiting the inherent data-compression capability of highly scalable, low power nanodevices that could potentially extend the scaling and processing capacity of neural recording platforms substantially. In essence this approach replicates the strategy adopted by natural synapses for signal compression where information on spike number and firing rate is stored into gradual changes of the post-synaptic membrane conductance. In contrast, present state-of-the-art neural activity monitoring platforms [48], like the MEA-based system described in Appendix. C, rely entirely on front-end circuitry for detecting and transmitting all data off-line for processing.

4.5.1 Concept

The concept introduced in Fig. 4.5, when directly interfaced with front-end-circuitry, can be exploited for advancing the present state-of-art in high density neural recording platforms [122]. The presented concept is amenable for scaling to a multi-channel array level, as illustrated in Fig. 4.13, for capturing the activity of neural networks in real-time. In this hybrid system [5], data from each of the 'N' pixels in the array arrives as an analogue current from the MEA and is multiplexed onto one of the 'M' on-chip trans-impedance amplifier (TIA) blocks, which are followed by on-chip gain and offset stages. Thus, a small number

[5]For the experimental validation of MIS system, neural activity was recorded from slices of dissected mid-peripheral rabbit retinal ganglion cells using CMOS technology (Fig.C.1) [123, 124]. The CMOS based multi-transistor array (MEA) consists of 128x128 sensor sites (pitch 7.8μm [39] , chip size of 5.4 mm x 6.5 mm with an active area of 1 mm x 1 mm), which records the data at a sampling rate of 12.2 kHz and outputs a current time series containing approximately 63k samples. The sensor sites of the CMOS-MEA are insulated by an inert TiO_2/ZrO_2 layer and a thin metal layer beneath the oxide layer is connected to the gate of the field-effect transistor. The voltage changes due to the interfaced neural tissue/cells above the recording sites are used for modulating the source-drain current in the MOSFET. Trans-Impedance Amplifiers (TIA) fabricated on-chip convert the signal into voltage and amplify the signal from a 0.1 mV- 1 mV range up to 10-500 mV in range. This amplified signal is then used as an input for our platform. It should be clearly noted that the CMOS MEA was kept external to the memristor-based spike detection platform and in this entire work the CMOS MEA is termed as the 'front-end' system. The data has not been made publicly available

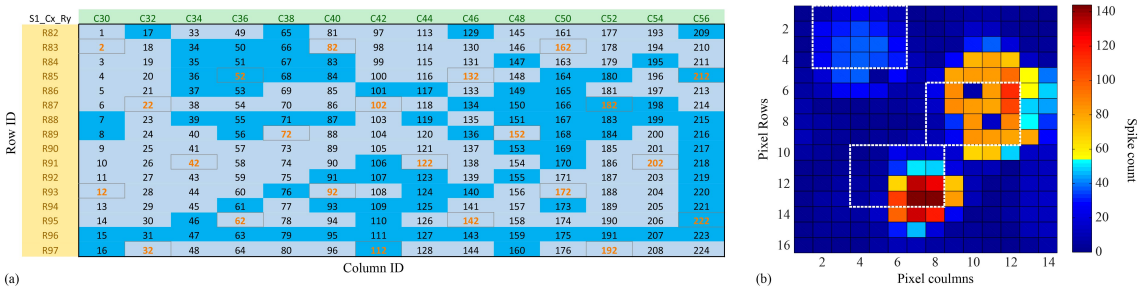


Figure 4.14: (a) Overview of the MEA 16x14 sub-array recordings. Green cells: MEA column coordinates of recordings employed. Yellow cells: MEA row coordinates of recordings employed. Blue cells: unique pixel identifier numbers. Dark blue cells indicate recordings that contain at least 1 data sample with magnitude above 0.5 V; this includes a number of recordings that barely pass the 0.5V mark. The map therefore represents the presence of spikes as benchmarked by simple threshold method. The 23 recordings illustrated in orange colour were chosen as a representative sample of the entire array and were then used to estimate gain (G) and offset values (V_{off}) for the main experiments. (b) Spike count pixel map yielded by the conventional template matching system for the MEA sub-array recordings used to obtain 4.11 (c). Three major regions of activity can be clearly seen.

of both gain and offset stages are time-shared by every pixel in the array. The conditioned recording data points are then de-multiplexed to a memristor bank, that can be fabricated into the back-end of the chip, in good proximity to the MEA recording sites. MIS output is then generated by sequentially measuring the resistive states of each memristor in the bank. The low frequency at which memristor read-outs are generated for example 200 times lower data rate vis-a-vis input stream arriving from the MEA if a standard scheme is used as described in Section. 4.3.2 allows the MIS system to carry out all measurements through a single or few, time-shared TIA feeding into Analogue-to-Digital Converter (ADC). The digitised results are then sent off-chip. It can be understood that a practical implementation of a monolithically integrated system will involve addressing the challenges associated with the integration of a MIS array with CMOS-based front-end circuitry, while the required MIS control can be accommodated as peripheral circuitry with sneak-path issues existing in dense RRAM crossbar configuration mitigated via selector topologies [125]. Importantly, in an ideal situation every channel should also be autonomous without the need of tuning between 100s and 1000s of channels.

4.5.2 Working Principle

In this experiment, the introduced concept was validated via a hybrid approach that is capable of processing 224 distinct recording traces stemming from a 16x14 pixel subset of the employed MEA system (see Appendix. C) atop which retinal ganglion cells were cultured. The sub-array was found to cover three cells after processing all recordings with a state-of-art array-level template matching system (TMS), using an extended principal component analysis method (PCA), as shown in Fig. 4.14 (a,b). An initial MIS calibration was performed in order to set suitable values for G and V_{off}. This entailed selecting a spatially sparse subset

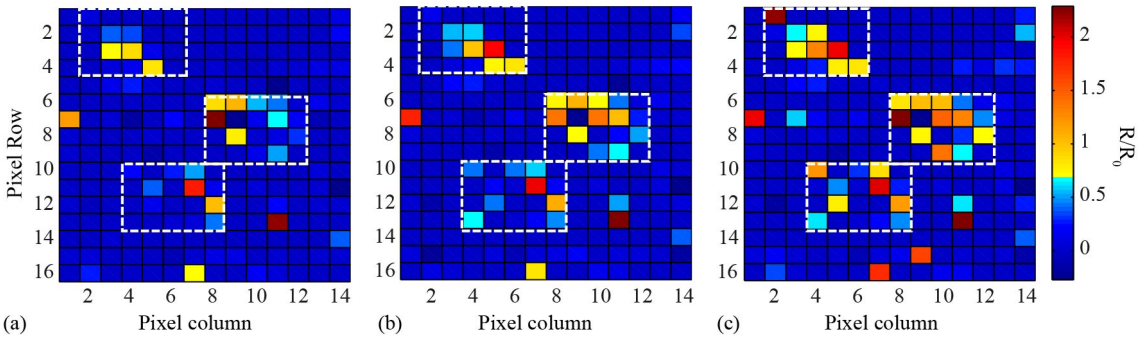


Figure 4.15: (a), (b) and (c) Time evolution of normalised resistive state throughout 16 x 14 test sub-array at $t_1 = 1.63$s, $t_2 = 3.27$s and $t_3 = 5.16$s respectively. The gain and offset values for the neural recordings was fixed at 2.8 and 0 respectively. Three clusters of activity can be discriminated which roughly correspond to the three main regions of activity detected by the template matching system (marked by white, dashed lines) in 4.14 (b). Colour bar mapping has been chosen to correspond to the one employed in Fig. 4.14.

of 23 pixels (Fig. 4.14 (a), cells marked in orange), and examining their neural recording waveforms to gauge average maximum/minimum voltage amplitudes as well as the typical levels of background activity. In this experiment, to ensure a suitable signal-to-noise ratio, gain and offset was set to 2.8 and 0 for all memristive devices. Importantly, these parameters were kept fixed for all recordings, they were not changed for accommodating individual memristive device behavioural variations, or distinct features of the employed recordings. Every utilised memristive device in this experiment was initialised to a common low resistive state in the range of 2-4 kΩ that for the given parameters yielded a useful MIS operating range up to the set 15 kΩ high resistive state.

These pre-processed 224 neural recordings were then fed to sixteen different memristive devices and spatio-temporal changes in the arrays memory state were monitored, snapshots of which are shown in Figs. 4.15 (a,b,c) for distinct time instances: $t_1 = 1.63$s, $t_2 = 3.27$s and $t_3 = 5.16$s respectively. Since the neuronal activity is encoded as non-volatile resistive state changes, one can observe an accumulation of activity clustered around three major centres: at pixel (row, column) locations (3,4), (7,10) and (11,7). Particularly the final array state, shown in Fig. 4.15 (c), qualitatively resembles the activity extracted by the conventional template matching method to the same neural recording data-set as shown in Fig. 4.14 (b). Thus MIS array results into a change in the resistive state that is strongly correlated to the strength of the individual spiking events. This attribute allows an added advantage of preserving information on both event amplitude and polarity, which in principle improves the data compression rate.

Furthermore, a few pixels exhibiting strong resistive state changes despite not appearing to belong to any well-defined cluster of activity can be distinguished in Fig. 4.16 (a). This discrepancy follows the argument presented previously for Figs. 4.11 (g,h), hinting that single, exceedingly strong events may lead to resistive state changes comparable to those arising as a result of accumulated activity. Thus by extending the idea to the array level and experimentally demonstrating that a memristor-based neural activity sensor could operate

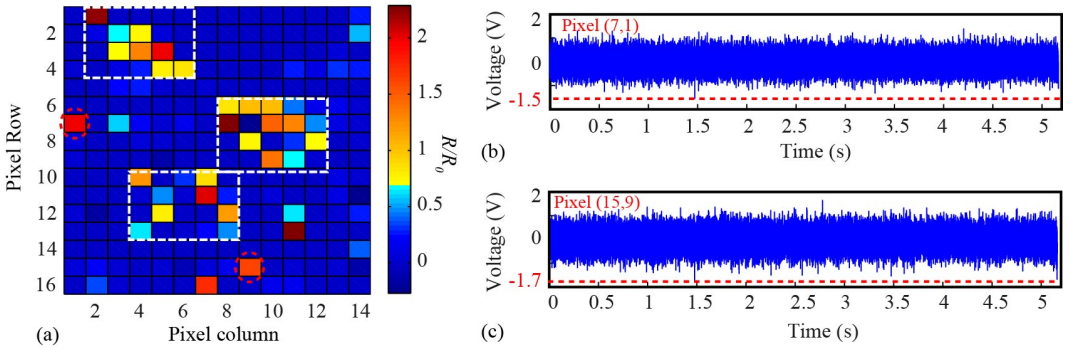


Figure 4.16: The origin of 'stray' pixels in the final output state of memristive integrating sensor as shown in 4.15 (c). (a) Memristive device array final state pixel map. It shows the DUT final/initial resistive state ratio. The three main regions of activity are enclosed in white, dashed line boxes, as are two of the four 'stray' pixels circled in red that show remarkable change in state without belonging to any of the three main clusters. (b) and (c) Investigation of the neural recordings corresponding to these stray pixels and it was found that they all contain lone, remarkably high amplitude events i.e. -1.5V for the former and -1.7V for the latter.

as a neuronal spike encoder by compressing information on the spike amplitude and firing rate, it can be seen that this approach can be potentially used for monitoring the activity of multiple cells at large-scales. In neuron-to-neuron communication, information carried by spikes is effectively compressed in changes of synaptic strength and in a similar fashion, information of spikes recorded by neural implants can be stored, in a compressed form and with potentially minimal power consumption (discussed in section 4.8), via single memristive devices.

4.6 Approaches for Improving the detection accuracy of MIS platform

Initial experiments confirmed the efficacy of TiO_x memristors to be operated as MIS devices by extracting significant neural activity events from pre-recorded in-vitro retinal cell activity. However, it was found out that one of the main performance-limiting factor is the programming saturation of the memristive devices [126]. For instance, in Fig. 4.11 (f) at $\approx 1.2\,s$ the DUT fails to detect any significant neuronal spike which is due to the saturation of the resistive state of the device at $\approx 4.5\,k\Omega$. Immediately afterwards, the presence of a strong event in the opposite direction causes a resistive state drop which renders the device capable of detecting further changes in the resistive state. Therefore, in this section it is investigated how key system parameters such as input signal pre-amplification, offset levels or noise band settings may impact the performance of the devices and can possibly be exploited to improve the detection accuracy of the devices for spike-detection.

More specifically, subsection 4.6.1 demonstrates the results for optimisation of operational parameters i.e. gain and offset values. Subsection 4.6.2, shows the effects of sampling rate and subsection 4.6.3, 4.6.4, 4.6.5 analyses the resistive state data in response to the neural

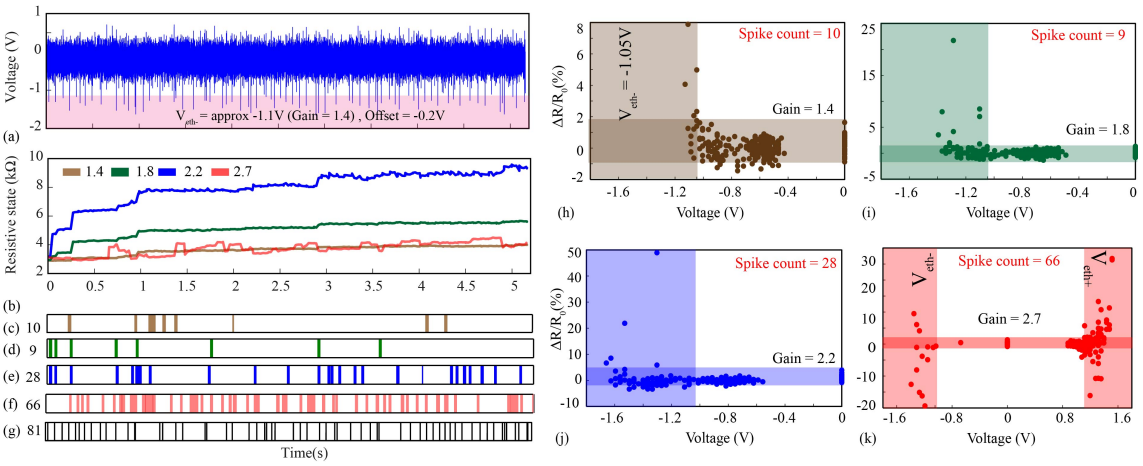


Figure 4.17: Comparison of different gain values for one neural recording and the spike count detected by the MIS. (a) Neural recording employed for the test. (b) The time evolution of DUT resistive state with gain 1.4, 1.8, 2.2 and 2.7. The offset for the first three gains is -0.2 V and +0.4 V for the last. (c),(d),(e) and (f) Spike locations and count as detected by MIS system when the gain is set to 1.4, 1.8, 2.2 and 2.7 respectively. (g) Spike locations and count as calculated by the state-of-the-art template matching system. (h), (i), (j) and (k). Normalised changes in resistive state for each bin plotted as a function of maximum voltage magnitude within bin for gain 1.4, 1.8, 2.2 and 2.7, respectively. Time axis in (g) applies to entire column.

recording in depth to gauge an optimised setting of the noise band. Finally, post-analysing the optimisation strategies, subsection 4.6.6 illustrates an improved measure for spike detection through non-volatile devices.

4.6.1 Optimisation of gain and offset

In these experiments, a MIS DUT was tested with a reference neural recording subjected to four different gain and two different offset parameter sets. The DUT was always initialised to a low initial resistive state value of approximately $R_{on} = 3\,k\Omega$ between runs as normal spiking activity causes increases in resistive state for the employed set-up. Figure. 4.17 shows the resulting ΔR traces and Table 4.2 summarises key operating parameters and results.

For gains 1.4, 1.8 and 2.2, the offset was set to -0.2 V and the resistive state at the end of the run R_{off} is found to be 4, 5.6 and 9.4 $k\Omega$, respectively. It can be noticed that at low gain (1.4) the MIS is filtering out too much activity, thus resulting in failure to detect the first two spikes in the recording, in contrast to the higher gain cases (Fig. 4.17 (a,b-e)). At high gain (2.2) we see an initial phase (up to $\approx 1.2\,s$) where activity detection results are in fairly good agreement with supra-threshold events in the neural recording (Fig. 4.17(a)) and also coincide with the case where gain = 1.8. However, after that period (after 1.2 s) it can be observed that the device tends to saturate towards its operational resistive state ceiling (maximum resistive state attained within safe operating voltage range). This precipitates a decreased sensitivity to spiking activity coupled with increased sensitivity to voltages with polarity opposite to the majority of neural spikes (modelled and experimentally shown in [127, 128]),

Table 4.2: Comparison of key parameters for different gain and offset values. (N = Noise, R_{init} = Intial state of the device and R_{end} = Final state of the device.

Gain	**Offset (V)**	**Spike count**	**R_{init} (kΩ)**	**R_{end} (kΩ)**	**Max N (%)**	**Min N (%)**
1.4	-0.2	10	2.9	4.0	1.6	-0.88
1.8	-0.2	9	2.9	5.6	1.34	-0.83
2.2	-0.2	28	3.2	9.4	3.9	-0.81
2.7	+0.4	66	3	4	1.3	-0.77

as evident from Fig. 4.17(b). These results show that increased sensitivity to spiking activity (higher gain) needs to be addressed with more frequent device reset operations.

Next, for each gain and offset value, the normalised changes in DUT resistive state are plotted as a function of maximum voltage magnitude recorded in each bin as shown in Fig. 4.17 (h,i,j,k). For comparison, the DUT thresholds ($V_{eth-,eth+} =\approx \pm 1.1\,V$) are also marked. Though this method of visualising data implicitly assigns DUT ΔR's to the highest magnitude data points in each bin, it is still a useful tool given the proven sensitivity of employed devices to voltage peaks and the fact that the neural recording spikes are the most prominent input signal feature. Notably, as gain increases, an increasing number of data points populating the false negative quadrant of each plot can be observed. This can be interpreted as the effect of gradual device saturation. Simultaneously, it can be observed that in the case of gain = 2.2, a large number of data points are in false negative territory, but at negative ΔR values caused due to events in positive polarity (once the device tends towards saturation, the device is more sensitive to events in positive polarity), which may account for a lot of the detected spikes in Fig. 4.17 (e).

Finally, a further experiment was ran with a different offset of $+0.4\,V$ (higher gain = 2.7 to compensate for effect of offset on spikes) with the aim to examine whether noise and spurious activity can be exploited as a constant resetting mechanism for used MIS device during normal operation. Results (Fig. 4.17 (b,f,k)) confirm that the principle indeed works; the DUT always remains within a tight resistive operational range and a significantly higher number of spiking events leading to positive ΔR is detected throughout the recording. The cost of this self-resetting mechanism via noise/spurious activity is that by deliberately allowing events of the non-desired polarity to affect DUT resistive state we are masking genuine neural activity in bins that feature reset activity (negative ΔR). Circuit design techniques for e.g. utilising redundant devices may potentially compensate for this effect.

For comparison purposes, spike count results from the template matching system are also shown (Fig. 4.17). Whilst the MIS system detects significantly fewer spikes than the template matcher, it shows good agreement with it at the start of the recording. This may indicate that with more frequent device resetting and optimised noise band limit definition significant improvements in performance are achievable. Thus efficient operation of MIS devices requires careful choice of gain and offset parameters.

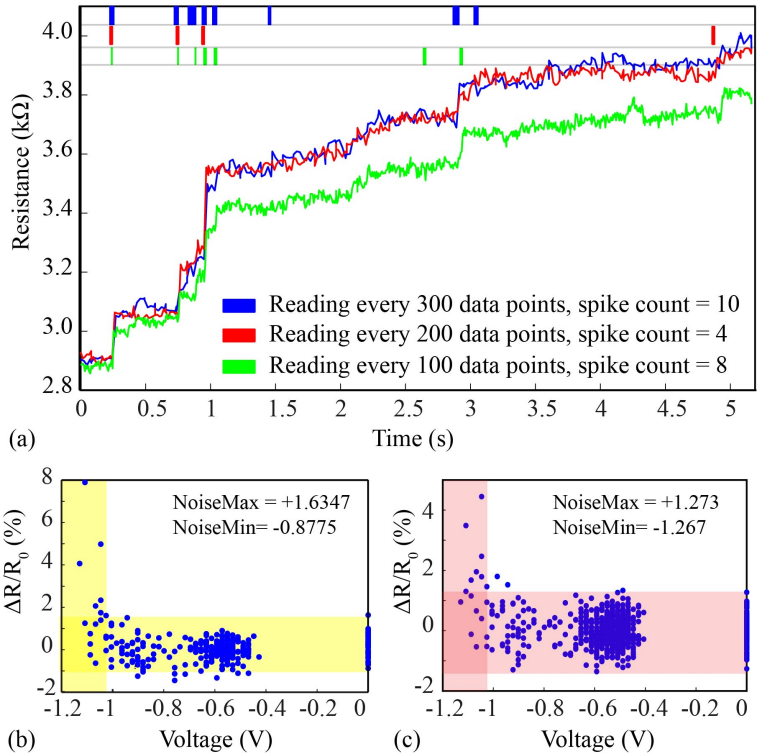


Figure 4.18: Comparison of MIS data for different sampling rates. (a) The response of the DUT when subjected to neural waveform shown in Fig. 4.17 (a) with gain and offset value of 1.4 and -0.2, respectively. The device was biased thrice and evaluated for three sampling rates (read after every 300, 200 and 100 data points - blue, red and green traces respectively). Corresponding time bin raster plots also shown (inset). (b) and (c) Normalised changes in the resistive state of the device for sampling periods of 300 and 100 pts. plotted as a function of max. bin voltage including noise band estimation.

4.6.2 Effect of sampling rate

The standard data read-out scheme as described in section 4.3.2 employs relatively low sampling rates (i.e. reads the state of the device after every 300 data points) at the cost of timing resolution. To understand the effects of higher sampling rates, the DUT was subjected to the neural recording from Fig. 4.17 (a) with different periods of resistive state assessment. Qualitatively consistent results were obtained for sampling at every 300, 200 and 100 points as shown in Fig. 4.18 (a) indicating flexible operation along the time-resolution/data-rate trade-off curve. The device was intialised to R_{on} of 2.8 kΩ after each round of measurements with R_{end} reaching $\approx 3.8\,k\Omega$ in every run. All runs showed three spikes consistently up to approx. the $1\,s$ mark. The spike count for the sampling rate 300 and 100 data points agree for three other instances (upto $\approx 3\,s$), with rest being false positives. On plotting the normailsed changes in ΔR vs Vmax. bin voltage (Fig. 4.18 (b and c)), it is found that sum of the discrepancies between different runs can be attributed to difference in the noise band settings (current noise band settings i.e. 6σ method include outliers). These results strongly indicated towards possibility of optimising the noise band limits and improving the detection accuracy of the MIS platform.

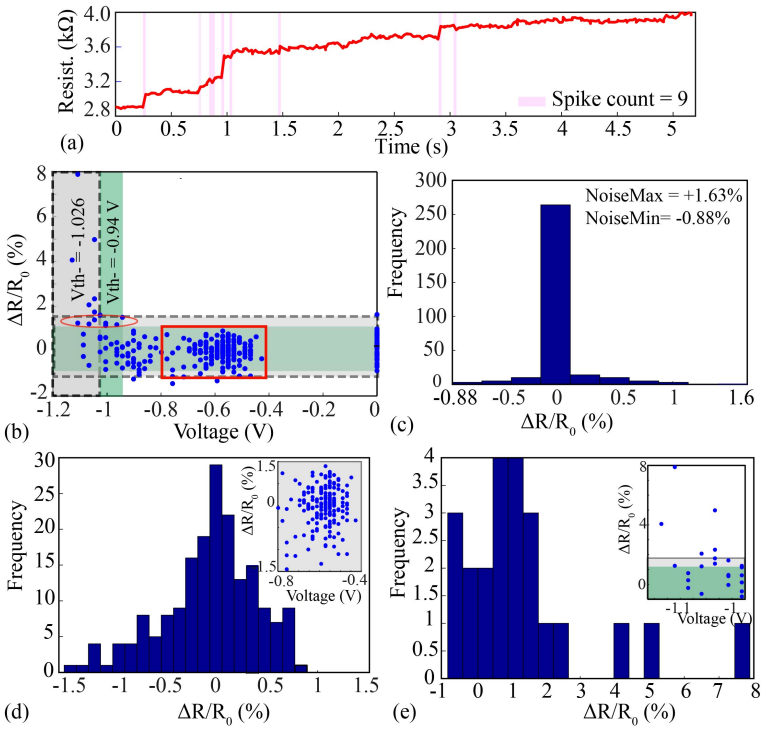


Figure 4.19: Analysis of $\Delta R/R_0$ plots. (a) Resistive state evolution of the DUT in response to the neural recording in Fig. 4.17 (a). The estimated spike count of the MIS system is 9 and the detected significant activity is illustrated as pink bands. (b) Normalised changes in the resistive state of the DUT for the entire neural recording shown in Fig. 4.17 (a) plotted as a function of max. bin voltage. The grey and green band indicates the difference in the estimation of threshold voltage when the outlier is included and eliminated respectively. (c) Histogram of the noise band estimations (with outlier). Histograms of the $\Delta R/R_0$ plots for voltages (d) -0.4 V to -0.8 V and (e) -1 V to -1.2 V. The inset indicates the corresponding voltage sensitivity plots.

4.6.3 Data Analysis: $\Delta R/R$ distributions vs $|V|_{max}$

Examining the $\Delta R/R_0(|V|_{max})$ plot (voltage sensitivity plot) can reveal important information including the effects of using different methods for estimating the noise band limits [129]. Figure. 4.19 (a) represents the response of a DUT in response to the neural recording shown in Fig. 4.17 (a) illustrating the output of the spike-count (9) from the MIS system. Figure. 4.19 (b) represents the voltage sensitivity plot for the corresponding response of the DUT. The noise band limits as computed by the 6σ method are outlined by the black, dashed lines. The normalised resistive state change histogram for data-set 'N' is displayed in Fig. 4.19 (c). This demonstrates a major concentration of the data around zero on the x-axis with few outliers on the extreme ends. In case of a live neuronal feed, the spread of ΔR data from bins where the maximum input magnitude $|V|$ is safely below the extracted $|V_{th-}|$ (red rectangle in Fig 4.19 (b) - 'sub-threshold' data-set), can be considered as a more realistic representation of the noise band under normal operation. The histogram plot spans from -0.4 V to approximately -0.8 V (Fig 4.19 (d)) which is expected to be safely under the extracted threshold voltage magnitude and shows that resistive state changes are centred around zero as is the case of data-set 'N' (Fig 4.19 (c)), but the overall distribution of points

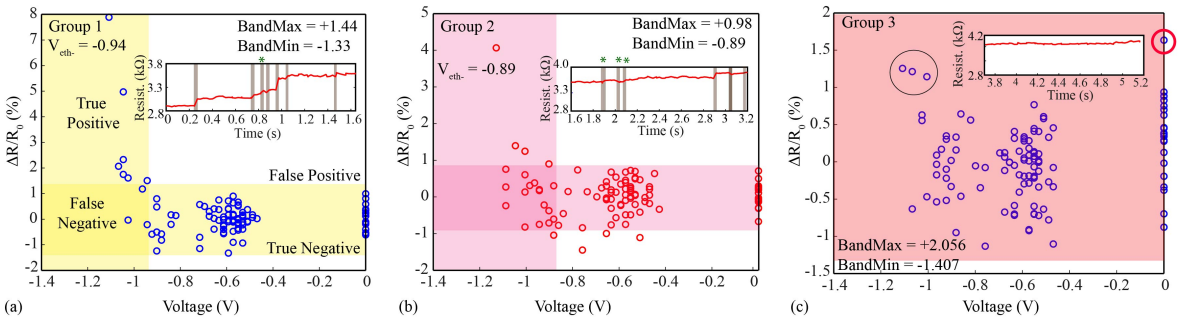


Figure 4.20: Analysis of Noise bands. In MIS system, one neural recording of ≈63 k data points (5.2 s) yields 316 ΔR values. Over here the data is analysed in three groups i.e Group 1 (1-100 ΔR values), Group 2 (101-200 ΔR values) and Group 3 (201-316 ΔR values). (a), (b) and (c) Normalised changes in the resistive state of the device for groups plotted as a function of max. bin voltage including noise band estimation for the respective groups. Noise band(s) are estimated using the 6σ approach (bandmax/bandmin) which is different from the max. and min. noise band limits. The inset shows the time-evolution of the resistive state of the device with the brown bands indicating the spikes detected in each group. '*' indicates the spikes not recorded in the output of the MIS system when the entire data for the neural recording is analysed (Fig. 4.19 (a)). The red and black circle in (c) indicates one of the outliers and potential significant ΔR changes respectively.

is substantially different. For the region where $|V| > |V_{th-}|$ i.e. $V < -1\,V$ (Fig 4.19 (e)), a bulk of data lies close to zero possibly due to saturation of the DUT whilst simultaneously substantial resistive state changes are clearly distinguishable as can also be seen from the distribution in this scenario.

Notably, comparing the 'sub-threshold' and 'N' data-sets we observe that the noise-band as determined by applying the 6σ method on 'N' seems excessively conservative: many data points that occur at high $|V|_{max}$ and feature a resistive state change are clearly distinguishable from what we observe in the sub-threshold dataset (circled in red) which are considered as insignificant. Cutting the extent of the noise band to the green shaded area in Fig. 4.19 (b) is expected to lead to better results. In this case this was achieved by using a 4σ approach instead of 6σ, which changed the extracted V_{th-} value to -0.94 V.

4.6.4 Data Analysis: Activity detection over time

A neural waveform of approximately 63 k samples on using 300 as the sampling rate (200 as the data compression rate) following the standard signal processing scheme, folds into 316 ΔR values and a significant resistive state change is indicative of supra-threshold events in the time bin. Resistive state read-outs for the DUT biased with the neural recording as shown in Fig. 4.17 (a) are analysed in smaller groups in order to study the effects of evolution of resistive state of the DUT with time on the noise band. Initially, in this section, a conservative approach i.e 6σ process for the estimation of noise band is followed. The voltage sensitivity plots for 316 ΔR values are binned in three groups i.e 1-100 points (Fig 4.20 (a)), 101-200 points (Fig 4.20 (b)) and 201-316 points (Fig 4.20 (c)). Estimation of noise bands limits in

Table 4.3: Comparison of Noise band values with and without outlier (Fig. 4.20). G1: Group1, G2: Group 2, G3: Group 3, V_{th-}: Threshold voltage

	V_{eth} (V) (with/ without outlier)	Noise Max/ Min (%) (with outlier)	Noise Max/ Min (%) (without outlier)	Noise band (6 σ)
G1	-0.94/-0.94	+0.99/-0.61	+0.99/-0.61	+1.44/-1.33
G2	-0.89/-0.89	+0.71/-0.67	+0.71/0.67	+0.98/-0.89
G3	-/-1	+1.63/-0.83	+1/-0.83	+2.05/-1.41

the illustrated figures is made on the basis of each individual group which will be different from the limits of the 'whole recording' noise band (Fig. 4.19 (b), grey band).

Majority of the ΔR changes are recorded in the first 100 ΔR values (Fig 4.20 (a)) as the device gradually reaches its operational ceiling (high resistive state); clearly illustrated in the inset of the figure. The extracted threshold of the device (V_{eth-}) is approximately -0.9 V whereas the max and min values for the noise band are +0.99% and -0.61%, respectively (Table. 4.3). Most of the spikes detected (outside +1.44/-1.29 % estimated using 6σ) in this group are also recorded in the 'whole recording' experiment of the MIS system; only one is missed (marked by '*').

For the next 100 ΔR values (Fig 4.20 (b)), the resistive state of the DUT is beginning to saturate with few ΔR changes detected. The total number of spikes detected in this group is equal to 7 however the ones indicated with '*' are excluded in the 'whole recording' spike count (as can be seen in Fig. 4.19 (a)). The values for the limits of noise band and extracted V_{eth-} are similar to the previous case. However, the false negative (FN) quadrant is significantly crowded in comparison to Fig. 4.20 (a), possibly as an effect of DUT resistive state saturation.

Analysing the data from the last set of ΔR values(Fig 4.20 (c)), it is found that the resistive state of the DUT seems completely saturated. There is a substantial increase in the max. noise band to +1.63% due to the introduction of an 'outlier' (marked in red-circle). This masks the possible significant ΔR changes indicated by the black-circle (Fig 4.20 (c)). The broadening of the noise band limits in this group affects the overall estimation of the noise band for the full neural recording, which then covers significant ΔR changes detected in the initial two groups. This leads to a reduction in the spike-count of the system statistically revealing less activity in comparison to the detected. However, if the outlier is not taken in consideration the max. noise band shoots down to 1% resulting in a V_{eth-} of approx. -1 V offering much fairer estimation of the detected neural activity. This illustrates the limitations of the approach used until now in evaluating the noise band (6σ) and countering the discrepancy in the significant events detected and finally recorded as MIS output.

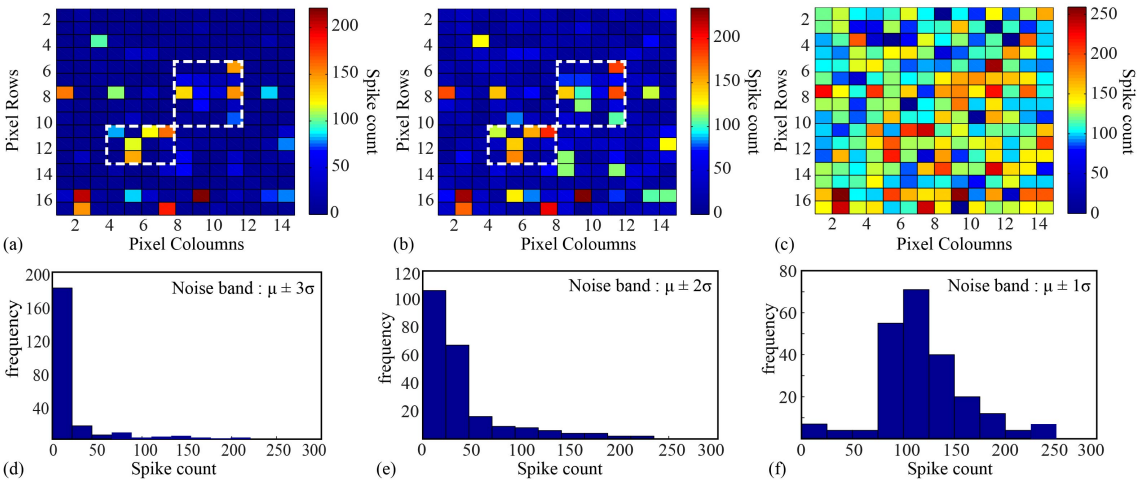


Figure 4.21: Surface plot and histograms of the spike counted estimated by the MIS for 16 x 14 neural recordings when the noise bands are set to be (a), (d) $\mu \pm 3\sigma$, (b), (e) $\mu \pm 2\sigma$ and (c), (f) $\mu \pm 1\sigma$, respectively.

4.6.5 Data Analysis: Noise band settings comparison

In this section, an in-depth study of the noise-band settings is performed. Three different settings of noise band constituting 2σ $(\mu \pm 1\sigma)$,4σ $(\mu \pm 2\sigma)$ and 6σ $(\mu \pm 3\sigma)$ is done and the most appropriate method is chosen for performing the benchmarking of our approach.

4.6.5.1 Approach comparison on 16x14 MEA sub-array

The conservative 6σ approach for determining noise band limits is not necessarily the best option. If a 4σ approach $(\mu \pm 2\sigma)$ is considered we would expect 95% of the noise band data to be still included under Gaussian distribution assumptions and the outlier to be eliminated. This would lead to a noise band as given by the green shaded area in Fig 4.19 (b). Using this method, the system is capable of picking up many more significant resistive state changes as indicated by the red encircled events in Fig. 4.19 (b). On the other hand, adopting an even more liberal $\mu \pm 1\sigma$ approach is expected to include excessive amounts of unwanted background activity in spike estimation. In this section, experimental results from the different approaches for statistical estimation of noise band to reach an optimised approach to calculate the output of MIS system are presented and benchmarked against the template matching system. For testing the methods of implementation of noise bands, 224 distinct neural recordings extracted from a 16×14 pixel sub-array of an MTA system as illustrated in Fig. 4.15 were used.

To set gain (G) and offset (V_{off}), a random subset of 23 neural recordings was chosen from the available set to firstly estimate the maximum/minimum voltage amplitudes and secondly to examine the typical levels of noise as mentioned in section 4.5.2, Fig. 4.14. Any recording that contained at least one data sample with magnitude above 0.5 V was considered to contain genuinely significant neural activity. For this experiment, G, V_{off} and sampling

Table 4.4: Approach comparison for a whole neural recording using $\mu \pm 3\sigma$ and $\mu \pm 2\sigma$ method (Fig. 4.22). Spikes estimated for template matchng system (TMS), Spike bins estimated for the template matching system (TMSBIN). Parameters used: G= 2.2, $V_{off} = 0$.

	TMS/TMSBIN	$\mu \pm 3\sigma$	$\mu \pm 2\sigma$
Spike count	82/78	19	36
TP	-	19	28
FP	-	0	8
TN	-	174	166
FN	-	59	50
Rate of TP (%)	-	24.3	35.8
Rate of FP	-	0	4.59

rate parameters were set to 2.8, 0 and 300 (standard read-out scheme), respectively, which were kept constant for all the recordings. The values were not customised to each recording channel/pixel to accommodate individual device behaviour or protect devices against the presence of excessively high amplitude spikes in the neural waveform. All the electroformed TiO_x based devices used for this experiment were initialised to a low resistive state (2-4 kΩ). Upon biasing with neural recordings, the devices showed an operating range extending up to 15 $k\,\Omega$ (high resistive state).

The resulting 224 time evolution of the resistive state of biased memristive devices were analysed to compare the spike count of the MIS system when the noise band limitations were set to: (i) $\mu \pm 3\sigma$, (ii) $\mu \pm 2\sigma$ and (iii) $\mu \pm 1\sigma$. Figure. 4.21 (a),(b) and (c) shows the final spike-count status of the array for three different methods. For the former method (i), two clusters of neural activity is detected with respect to spike count which is improved by adopting method (ii). However, method (iii) introduces a lot of unwanted background activity i.e. noise in the spike count which is clearly evident from the spike count map and the corresponding frequency in the histogram plot. Again, an important distinction between the MIS and the template matching system is that the output of the MIS system encodes the strength of the signal in the changes in the resistive state in contrast to the template matching system which only outputs spike count estimated above a certain set threshold.

4.6.5.2 Approach comparison on a single device

A comparison of the output of the MIS system for two statistical methods i.e $\mu \pm 3\sigma$ (i) and $\mu \pm 2\sigma$ (ii) in response to one of the neural recordings on 16x14 sub-array (a reference recording) is depicted in Fig. 4.22. The total number of spikes obtained in the former case is 19, displayed as green bands in Fig 4.22 (b). By using the second method, the spikes count increases to 36 (Fig 4.22 (c)), with additional spikes illustrated as pink bands. Both sets of MIS results are then benchmarked against the state-of-the-art TMS. To that end the spikes detected by the TMS system are initially collapsed into time bins (TMSBIN) similar to the MIS system and shown as a raster plot in Fig. 4.22 (d). Total number of bins: 252

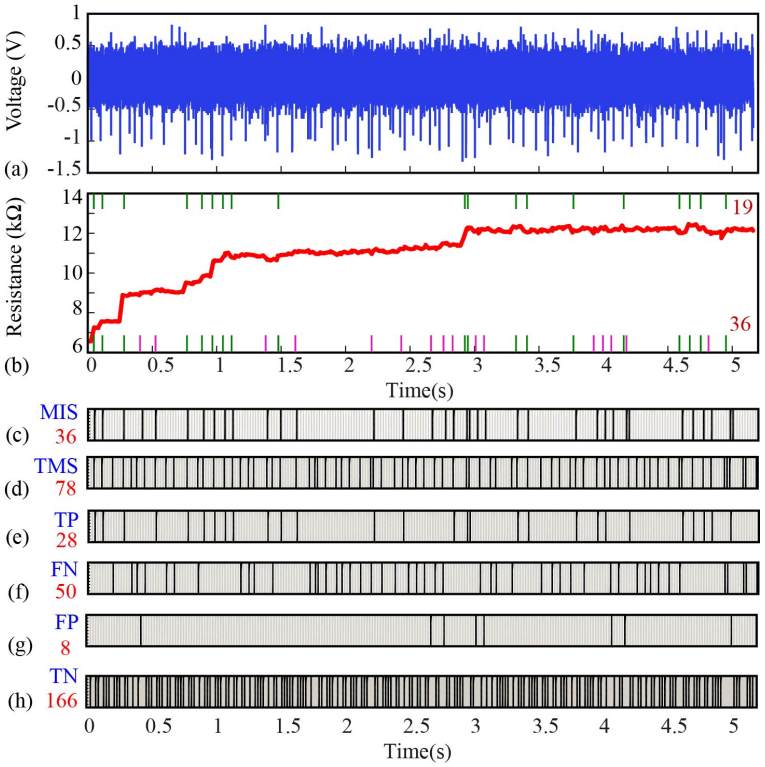


Figure 4.22: Comparison of the spike count estimated by the MIS system and benchmarking against state-of-the-art template matching system. (a) Neural recording employed to bias the TiO_x based device. (b) Gradual change in the resistive state of the device in response to the supra-threshold events in the neural recording. Upper panel (green) and lower band (green and pink) indicates the spike estimated with the noise band set to $\mu \pm 3\sigma$ and $\mu \pm 2\sigma$, respectively. Spike count adds up to 19 for the former and (c) 36 for the latter. (d) Spike bins estimated for the template matching system i.e. TMSBIN (78). (e), (f), (g) and (h) Raster plots for TP (28), FN (50), FP (8) and TN (166) estimated by the MIS system. The total number of bin columns are 252.

i.e. excluding the noise bins. The output of the TMS system for this recording is 82 spikes, which collapse onto 78 bins. Subsequently, each time bin is examined in order to determine whether the MIS and TMS systems yield the same results. For the purposes of benchmarking it is implicitly assumed that the TMS system is a perfect spike detector and the rates of TPs and FPs, are calculated using equations (4.1 and 4.2).

Results from the reference recording yield a TP of 28 (Fig. 4.22 (e)) FN, FP and TN stand at 50, 8 and 166 respectively as shown in Fig. 4.22 (f-h) for $\mu \pm 2\sigma$ method. The rate for TP is 35.8% which has improved in comparison to $\mu \pm 3\sigma$ which is estimated to be 24.3%. Computed FP rate are 4.59%.

4.6.6 Frequent resets strategy with optimised gain, offset and noise band settings

Based on the data analysis of neural activity detection of the MIS device as seen in section 4.6.4, relatively low true positive rate may be at least partially ascribed to resistive

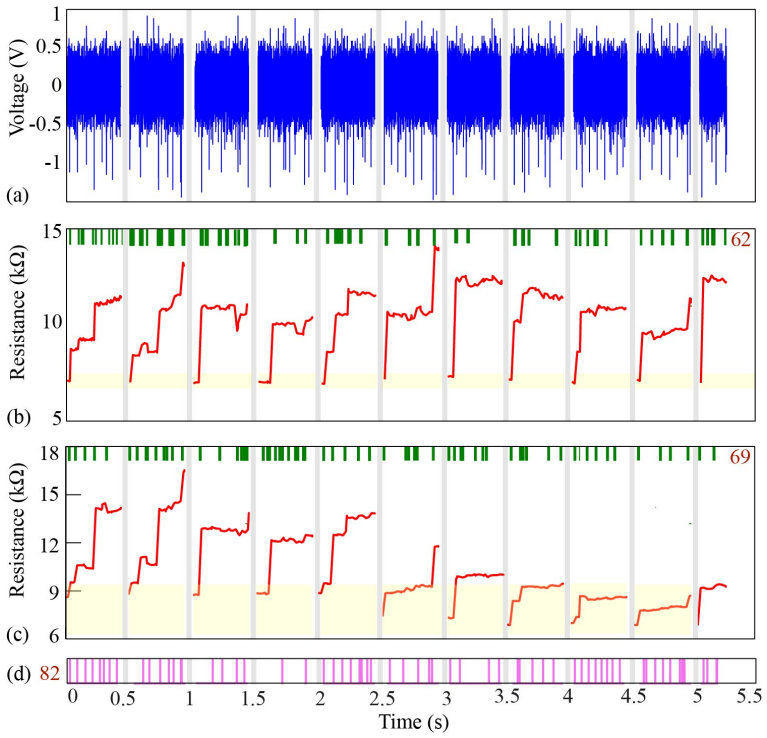


Figure 4.23: Frequent resetting of the memristive device in order to improve the spike count of the system. (a) Slicing of the neural recording used in Fig. 4.22 in eleven parts each lasting for 0.5 s separated by grey bands. The device is manually reset to its initial state after each run (yellow bands in (b) and (c)). R_{on} is approx. equal to 6-8 kΩ. (b) and (c) Output of the device when biased with neural recording in (a) with G = 2.2, V_{off} = 0 and G = 2, V_{off} = -0.2. Green bands indicates the bins in which neural activity is detected by the MIS system when the noise band is calculated using $\mu \pm 2\sigma$ method. (d) Spikes indicated by the TMS, which is equal to 82.

Table 4.5: Spike count and quantification parameters when the noise band is set to $\mu \pm 2\sigma$ (Fig 4.23 b and c respectively) in comparison to the state-of-the-art template matching system. Gain and offset parameters used in Fig. 4.23 b: G = 2.2 and V_{off} = 0. Gain and offset parameters used in Fig. 4.23 c: G = 2.2 and V_{off} = -0.2 . TP: True Positves, FP: False Positives, FN: False Negatives, TN: True Negatives, TMS: Spikes estimated by template matchng system (TMS), TMSBIN: Spike bins estimated for the template matching system.

Neural Recording	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**Total**
TMS spikes	7	8	8	7	9	7	6	8	8	8	6	82
TMS bins	7	7	8	7	8	7	6	7	8	8	3	76
$\mu \pm 2^*\sigma$, Fig. 4.23 b	11	9	7	4	5	3	2	4	6	6	5	62
TP	4	6	6	3	2	1	2	3	3	5	3	35
FP	7	3	1	1	3	2	0	1	3	1	2	24
TN	46	14	15	16	13	15	18	16	13	15	6	187
FN	3	1	2	4	6	6	4	4	5	3	2	40
Rate of TP (%)	57.14	85.17	75	42.8	25	14.28	33.33	42.8	37.5	62.5	60	≈47
Rate of FP (%)	13.20	17.64	6.25	5.88	18.75	11.76	0	5.88	18.75	6.25	25	≈11
$\mu \pm 2^*\sigma$, Fig. 4.23(c)	5	9	7	11	6	5	8	6	6	4	2	69
TP	5	7	5	5	4	2	3	5	5	4	2	47
FP	0	2	2	6	2	3	5	1	1	0	0	22
TN	17	15	14	11	14	14	13	16	15	16	8	153
FN	2	0	3	2	4	5	3	2	3	4	3	31
Rate of TP (%)	71.4	100	62.5	71.4	50	28.5	50	71.4	62.5	50	40	≈60
Rate of FP (%)	0	11.7	12.5	35.2	12.5	17.6	27.7	5.88	6.25	0	0	≈13

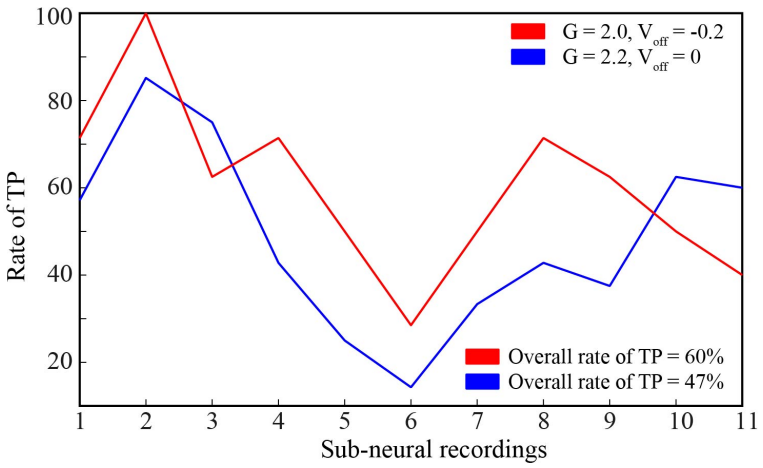


Figure 4.24: Rate of True positives detected for two cases illustrated in Fig. 4.23 b and c with different gain and offset settings for the same device.

state saturation and one approach to improve performance is employing frequent resets. Frequent resetting of the DUT is expected to limit the time the device spends at its resistive state operating ceiling i.e. maximum high resistive state the device attains with the employed characterisation parameters and thus improve its ability to detect significant activity in the neural waveform. To test this hypothesis, the same reference recording containing ≈ 63 k data points, shown in Fig. 4.22 (a), was sliced in eleven parts of approximately 6 k data points; each containing multiple significant spikes as shown in Fig. 4.23 (a). The DUT was initialised to 6-8 kΩ 'manually' before each sub-recording was played to the memristive device. To reset the device to its initial resistive state, positive pulses of 100 μs were used. The resistive state time-evolution of the device in response to the neural data is illustrated in Fig. 4.23 (b). Gain and offset parameters were left unchanged at G = 2.2, $V_{off} = 0$ and the quantification is illustrated in Table 4.5.

Increased number of resistive state changes in the device adds to the capability of the MIS system to detect significant neural activity. The spike-count for each recording slice as estimated by the MIS system when the noise band is set to $\mu \pm 2\sigma$ is also shown in Fig. 4.23 (b). Significant improvement in the system is detected in comparison to the whole neural recording experiment in section 4.6.5.2. Estimated spike count with frequent resets has increased to 62 with overall rate of TP improving to 47%. An interesting feature of the resistive state time evolution traces in the sliced recording experiment is the presence of a number of noticeable DUT resistive state changes towards lower resistive state. This contrasts with the overall trend towards higher resistive state. These resistive state drops may have happened by positive polarity activity in the neural recording, a hypothesis that is supported by the visible correlation between positive 'spikes' and resistive state drops in e.g. traces 2 and 3 from Fig. 4.23 (a,b). Whether these 'spikes' represent significant neuronal activity or not is uncertain, but in the case they are not we would hypothesise that their presence could influence ΔR and consequently spoil results in their respective bins by masking resistive state changes induced by clearly significant, 'negative' polarity events.

Table 4.6: Overall comparison of MIS performance using the three different operating methodologies from section 4.6.6, Fig. 4.23, Table. 4.5.

S.No	Approach with 4σ noise-band	Rate of TP (%)
1	Whole recording	4.26
2	Frequent resets	7
3	Frequent resets with optimised gain/offset	60

In order to test this hypothesis, the sliced recording experiment was repeated on the same DUT but this time with G = 2 and V_{off} = -0.2, where the offset guarantees that despite increased gain the maximum amplitude of positive polarity events is reduced to $\approx +0.5\,V$. Results are shown in Fig. 4.23 (c). The total number of spikes detected in this case increases to 69 ($4\,\sigma$ method) with significantly improved quantification parameters as noted in Table 4.5 and Fig. 4.24. TMS results corroborate our hypothesis. The overall rate of TP is found to be approx 60% with individual slices featuring rates as high as 100%. This can be seen in sub-recording 2 (Fig. 4.23 (c)). The improved results may be a result of the combination between higher spike/noise contrast brought about by increased gain and the weakening of positive polarity events by application of appropriate offset. Table. 4.6 depicts a summary of the overall improvement in the performance of our system according to the approaches discussed in the preceding sections.

4.6.7 Discussion

Narrowing the noise band estimates from the conservative $\mu \pm 3\sigma$ of the 'N' data-set to the $\mu \pm 2\sigma$ leads to a significant improvement in the rate of true positives for a relatively small increase in the rate of false positives. This reflects the fact that even though the total number of 'true' detections remains roughly constant it becomes much more likely to detect a true positive with the $\mu \pm 2\sigma$ method. Naturally, the tolerance of each specific application to the occurrence of false positives and false negatives will ultimately determine whether the proposed optimisation is acceptable or not [130]. It can be well understood that by utilising relatively simple techniques the amount of information that can be extracted from a seemingly information-poor series of sparsely sampled memristor resistive state readings can be substantially increased. Benchmarking the obtained results against widely used state-of-the-art template matching system utilising accurate principal component analysis spike-sorting algorithm confirms the validity of the introduced approach [59].

Frequently resetting memristors have a highly beneficial effect on significant activity detection by preventing resistive state saturation at a relatively low cost. The theoretical upper bound for the resetting frequency for MIS platform would be determined by the number of bins/unit time but in the demonstrated TiO_x memristors case resetting operation is carried out very infrequently ($\approx 1/6.4k$pts). Resets are only necessary when the device saturates, in operando

in the presented case, the benefits from increasing reset frequency should rapidly diminish beyond a fairly low frequency, far from the maximum allowed. Under normal operation, the reset should be reducible to a single pulse of appropriate amplitude, in contrast to the intricate operation used in these experiments which was performed in order to maintain initial test run resistive state within a relatively narrow range. Importantly, the differential nature of the MIS system (outputs determined by ΔR, not resistive state) would imply that the exact initial state after each reset is of limited importance so long as it is guaranteed to be 'sufficiently far' from the operational resistive state ceiling of the device which results in saturation of the DUT.

The mitigation of the effects of positive polarity events on bin accuracy is subject to the interplay between two effects: on the one hand positive polarity events can act as a continuous (albeit poorly controlled) reset mechanism [127] that may potentially help memristors avoid resistive state saturation for extended periods of time, but on the other hand the time bins during which these resets occur may mask significant neuronal activity affecting the false negative quadrant. Furthermore, memristors exhibiting asymmetric switching thresholds for the two polarities, as is the case for employed devices, implies that careful engineering of the offset and gain settings may be required before the balance between neural activity-induced resistive sate changes and resets is tuned sufficiently well. Optimisation of gain/offset settings proves advantageous in order to restrict significant events for one-polairty leading to well-controlled and engineered MIS operation.

4.7 Power estimations

The proposed memristor-based neural activity sensor can potentially open a new path towards energy- and area-effcient bio-signal acquistion and processing. It is therefore important to consider how it may perform against more traditional, fully CMOS-based approaches at the circuit level. For this, power estimation for the operation of DUT according to the standard signal-processing schematic for one batch is considered i.e. for every batch of thousand data points, the state of the device is read five times. The following calculations have been performed assuming the resistive state of the device to be $10\,\mathrm{k\Omega}$ and series resistance (compliance) to be $1\,\mathrm{k\Omega}$ (as seen in the experimental results). The read voltage of the device is set as $+0.5\,\mathrm{V}$, write voltage is conservatively assumed to be $+5\,\mathrm{V}$, sampling frequency is $82\,\mu\mathrm{s}$ and the pulse width used for the experiment is $100\,\mu\mathrm{s}$ (see Appendix. D for more details).

Under these assumptions, the write operation i.e. biasing of the device with neural recording for one batch of 1000 samples would cost $250\,\mu\mathrm{J}$ and the read operation (reading the state of the device five times in a batch) would equal to $11.5\,\mathrm{nJ}$. The energy dissipated in resetting operation with a single pulse of $100\,\mu\mathrm{s}$ will be equal to $250\,\mathrm{nJ}$ giving the average power consumption for the sampling frequency to be $\approx 3\mathrm{mW}$. However, experimental evidence suggests that good memrisitve sensor operation can be obtained if we employ 'voltage-time' trade off

and map the highest amplitude onto ≈5V pulses at <100ns [131]. This was also found to apply when the devices are operated in the 100kΩ resistive state range. Under these conditions, biasing of a memristor with an amplified neural sample (write operation) would cost a max. of 250μW$\times$100ns $= 0.025$ nJ and the read-out operation would cost 0.8μW$\times100\mu s$ $= 0.08$ nJ for an average power dissipation per channel of ∼300 nW. These projected power-dissipation figures are already significantly less than the present state-of-the-art continuous time spike detectors (∼720 nW-digital input) [24]. Clearly, the memristor read-out and biasing circuitry will require an additional power. Most importantly, the proposed technology is demonstrated here at a proof of concept-level via large prototype devices ($60 \times 60\mu m^2$) and clearly the presented power/density considerations are not a reflection of the technologys full potential. A full circuit implementation power estimations would also include considerations of 100s and 1000s of channels with precise schematic of peripheral circuitry. Substantial improvements in power consumption can be achieved by further downscaling and/or operating memristors at even higher resistive state ranges, for example operating the device in 1 MΩ region can further reduce the power dissipation remarkably by two orders of magnitude. *in-silico* simulations can be performed in order to develop an initial idea, however, this has not been performed during the course of this thesis and remains as the future work. In addition, the bandwidth required to assess the resistive state should be rather low as compared to the raw input data-rate and further power efficiency gains can be expected by integrating the MIS elements atop state-of-art CMOS.

4.8 Summary

The chapter presented a novel recording system exploiting the intrinsic synapse-like attributes of metal-oxide memristive devices to compress information on neuronal firing. Experimental results show that single devices are capable of identifying significant spiking events while suppressing noise. Memristive integrating sensor encodes the presence of events in non-volatile resistive state changes, allowing the flexibility to trade off sampling rates for timing resolution. This is particularly useful when information is rate- or spike-count-coded and where only a measure of overall activity within given time bins is requested. Sensitivity requirements and DUT resistive state saturation affects significant spiking event capture quality and hence key operating parameters such as DUT reset intervals and gain/offset settings should be chosen carefully. The principle of exploiting the neural recording waveform itself to continuously reset the device in-operando was also studied. Besides, sampling at different time bin resolutions and the method for optimisation of noise bands was also explored. These BEOL integrable MIS elements can complement current state-of-the-art neural recording platforms and can be potentially integrated directly atop front-end circuitry, however, the precise details of the integration system and circuitry remains as the future work. Moreover, as the memristor resistive state changes are linked to amplitude and polarity of the input

waveform (signal envelope), this information is preserved in the magnitude of resistive state modulation potentially paving the way towards ‘spike-sorting’.

Chapter 5

Spike detection using metal-oxide devices in volatile regime

5.1 Introduction

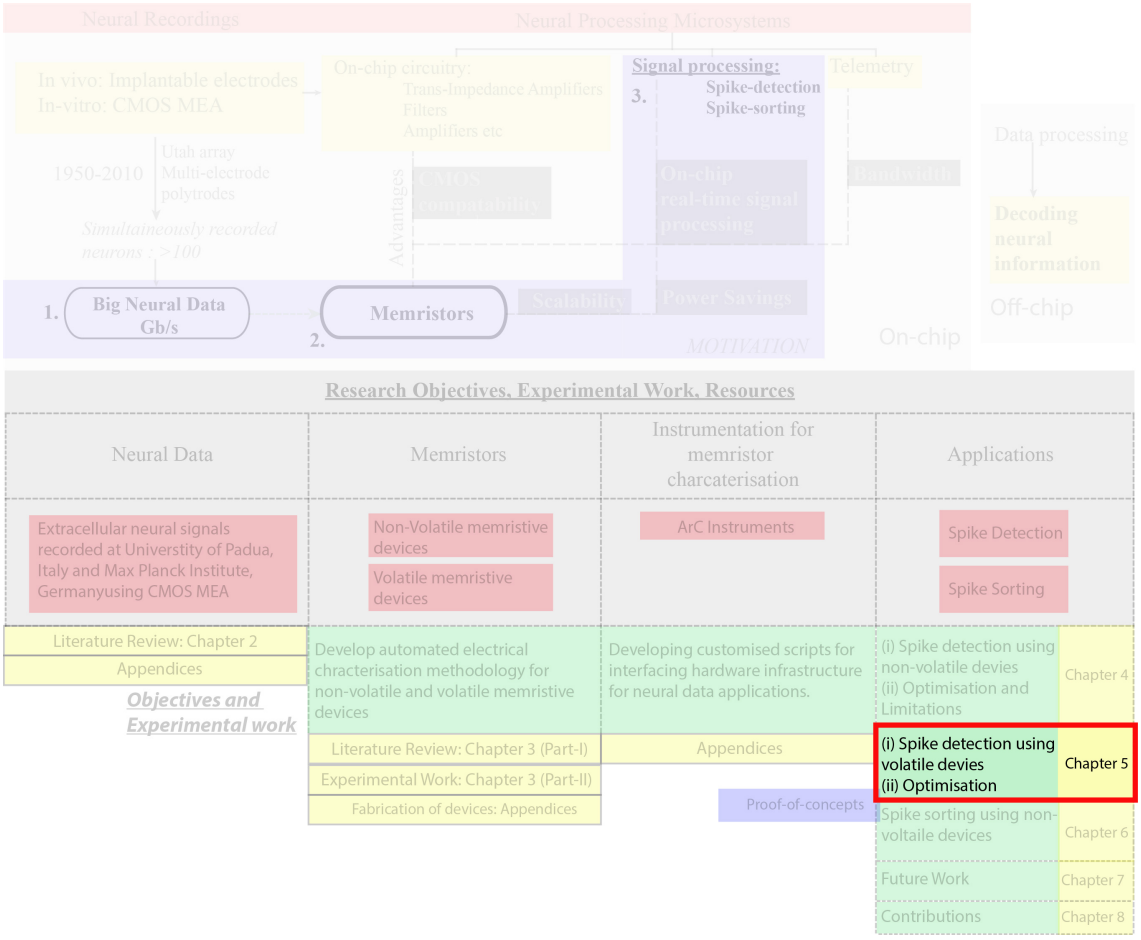


Figure 5.1: The research objective of this experimental chapter.

In the last chapter, a new spike-detection approach based on metal-oxide resistive switching memory devices was introduced [118, 126, 129]. Fundamentally, memristive devices undergo non-volatile resistive state transitions as a function of the integral of the input voltage, behaving as thresholded input integrators. Taking advantage of this property, it was demonstrated that TiO_x-based memristive devices can be employed for spike-detection, as extracellular neural spikes recorded from retinal ganglion cells were encoded in gradual, non-volatile resistive state transitions, whereas the sub-threshold events (i.e. noise) were naturally filtered-off [118]. This property makes these devices suitable as noise-suppressing integrating sensors and are thus termed as Memristive Integrating Sensors (MIS). Non-volatility, however, was strongly limiting detection performance, as after saturation of the resistive state the devices, it failed to register any significant neural activity [126]. Consequently, performance was optimised by manual operation through frequent resets to the initial devices resistive state, which however impacts negatively on the overall power consumption [129].

In this chapter, previous findings are advanced by exploiting another crucial property of memristive devices that is 'volatility' [105, 108, 114, 115]. This approach recalls the way of operation of biological synapses that translate spiking frequency in gradual changes of postsynaptic conductance subject to a continuous inherent self-resetting process [132]. When used in the volatile operating region, memristive devices exhibit metastable memory state transitions following which they inherently relax to their initial resistive state range [133]. It is demonstrated that volatility enables for naturally encoding spiking events into transient resistive changes. The self-resetting mechanism of used devices ensures that these operate far from their resistive saturation region, which overall enhances the attained spike-detection accuracy. More clearly, similarly to biological synapse, information on spike amplitude and frequency is transduced in metastable resistive state transitions of the nano-scale memristive devices (200x200 nm^2) as oppose to non-volatile resistive state transitions, which are capable of self-resetting and of continuous encoding of spiking activity. Moreover, the fact that volatile phenomena are more pronounced at higher resistive states [108] reduces the overall power dissipation to less than 100 nW, setting a new state-of-art in spike detectors [24].

In this chapter, section. 5.2 demonstrates the operation of memristive devices as volatile cells, section. 5.3, 5.4 illustrates the experimental results for nanoscale memristive devices as volatile spike encoders and section. 5.5, 5.6 implements the strategies for optimising the performance of the employed devices. Finally, section. 5.7 discusses the obtained results with power estimation and section. 5.8 summarises the chapter.

5.2 Operation of memristive devices as volatile cells

TiO_x metal-oxide memristive devices (micro- and nano- scale) with metal-insulator-metal architecture, as shown in Fig. 5.2 (a), were fabricated on a Si/SiO_2 substrate as detailed in

Appendix A [134]. A preliminary characterisation of the devices is needed to identify appropriate operation conditions in the volatile region [133]. A custom hardware infrastructure (Appendix B) was used for the electrical characterisation. The devices were electroformed using the procedure and characterisation protocol described in Chapter. 3 [133]. In brief, the devices are subjected to a high voltage stress until there is a sudden non-volatile change in their resistive state, known as electroforming [135][1]. The devices can then be operated in either non-volatile (described in Chapter 4, [118])[2] or volatile manner depending upon the polarity of the voltage stress applied during the electroforming procedure and the strength of stimuli [105]. Importantly, the devices feature inherent threshold levels, below which there is no change in the resistive state and above which metastable resistive state transitions are observed.

When operated in the volatile region, devices undergo metastable resistive transitions within a high resistive state range and are capable of inherently re-attaining their approximate initial resistive state. To enable this study, a volatility characterisation algorithm as described in Chapter 3 was developed [133]. To recapitulate, the algorithm applies a series of progressively more invasive voltage pulses and then monitors the resistive state of the DUT and its retention time-scale (Fig. 5.2 (b and c)). The module uses a standard two mean t-test to analyse the resistive state decay over time, terminating when the devices relax back to an equilibrium condition. Notably, the algorithm makes no assumption on what the equilibrium resistive state should be. The equilibrium condition eventually corresponding to a non-volatile residual change of resistance of the DUT is determined through a retention test, which is implemented throughout a user-defined time window (for instance 60s in Fig. 5.2 (c)). At the end, the output of the algorithm determines the time elapsing to achieve the equilibrium condition and the voltage ranges under which the devices can be safely operated in the volatile region, as estimated by comparing non-volatile with volatile changes. As shown in Fig. 5.2 (b and c), the operating resistive state region of the DUT was approximately 700 kΩ - 1.4 MΩ, using negative as the dominant stimulus polarity with 1 μs pulses (see Figure 5.3 for 100 μs pulse width).

The output of the volatility module is estimate of resistive state changes between the steady state and measurements taken immediately before and after the applied voltage stimulus[3]. This results in discrimination between non-volatile and volatile resistive changes for a given voltage stimulus as exemplified in Fig. 5.2 (d), where a -1.8 V threshold voltage marks the

[1] **Electroforming procedure for volatile devices**: For Electroforming stage: For prototype devices with μm scale active areas, negative polarity pulses (-6 V to -8V) were used. In the case of devices with nm scale (200nm x 200nm) active core areas, positive polarity pulses (+4 V to +6 V) were employed.

[2] **Device Electrical Characterisation in non-volatile regime**: In the non-volatile regime, the devices typicaly electroform in the range of +6 to +8 V and the device is then considered to be in low resistive state that is 'ON' state. The resistive state of the device decreases from 10's of MΩ to less than 50kΩ. Thereafter, a train of input programming pulses in alternating polarities is applied at a fixed duration of 100μs which leads to reversible resistive switching. The device is switched to low resistive state (ON/SET state) and high resistive state (OFF/RESET state) with positive and negative polarity respectively after the applied stimulus exceeds the DUTs inherent threshold voltage. Importantly, when operated in their non-volatile region, devices gradually switch within a 2 kΩ to 15 kΩ range as oppose to much higher resistive state range in volatile region.

[3] More specific details are presented in Chapter 3, Volatility Characterisation Protocol

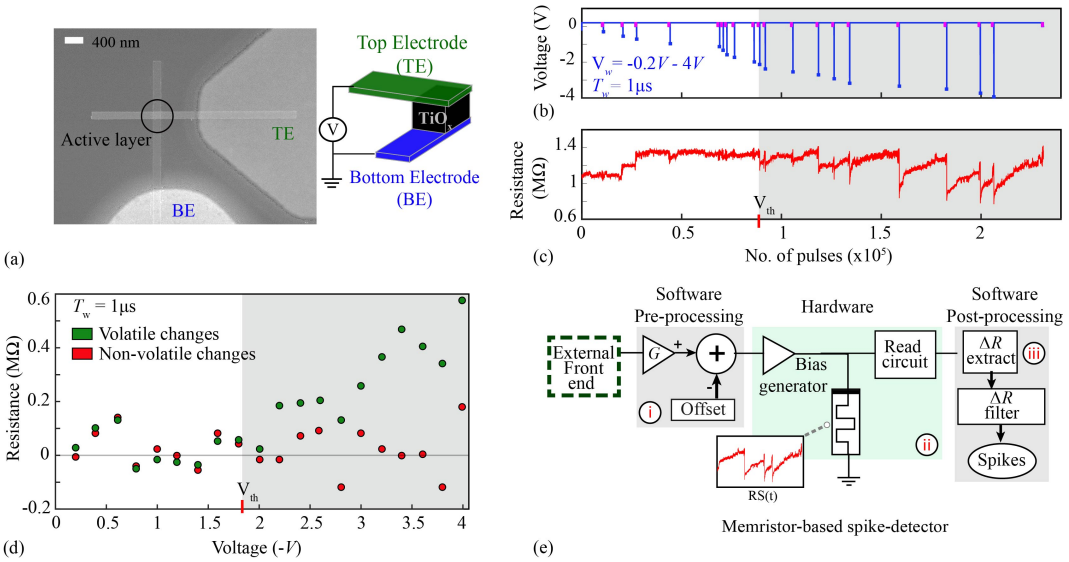


Figure 5.2: Device architecture and electrical characterisation of nano-scale TiO_x memristive devices in the volatile region. (a) Schematic (right) and Scanning Electron Microscopic (SEM) image (left) of the employed 200 nm x 200 nm memristive device. (b,c) Electrical characterisation of memristive device exhibiting volatility using an automatic volatility characterisation module [133] and the resistive state evolution of the DUT in response to the applied stimuli is illustrated in (c). The device operates in the 600 kΩ - 1.3 MΩ range. After a first electroforming stage, the algorithm applies a series of progressively increasing input pulses (blue bars) of amplitude (V_w) and width (T_w) to the target device, following which its resistive state is monitored. In the example, the write pulses (V_w) were applied in the range of -0.2 V to -4 V in steps of 0.2 V. T_w was 1 μs. The module operates on a standard statistical two mean t-test condition and terminates when the device reaches a steady state. Thus, equilibrium retention is verified by repeated readings over 60s (marked as pink rectangles), following which a new stimulus is delivered. (d) Determination of the operating voltage range of the DUT in volatile region. For every step of input stimulus applied, volatile (green circles) and residual non-volatile (red circles) resistive state changes were measured. The grey band indicates the voltage region for ensuring operation of the DUT in volatile conditions, with approximately -1.8 V being the inherent threshold voltage of the DUT (V_{th}). (e) Schematic for the implementation of memristor-based spike-detection platform [118].

transition of the DUT to a prevalent volatility state. The resistive state changes in the sub-threshold region are considered as insignificant and are mainly attributed to the background fluctuations caused due to the measurement noise. The range of identified inherent threshold voltages for the employed TiO_x-based memristive devices varied from approximately -0.6 V to -2.5 V (see Fig. 5.4). In conclusion, through the volatility module the range in which the devices could be safely operated in the volatile region and the rough estimate of relaxation times of the device to equilibrium state can be determined. The estimated volatility parameters were subsequently used when pre-processing the neural recordings in the spike-encoder platform. The platforms schematic is illustrated in Fig. 5.2 (e) and an overall picture of the spike-encoding system including the front-end system is presented in Chapter 4. Same neural neural recordings as used in the last chapter were used. These were obtained from slices of mid-peripheral rabbit retinal ganglion cells placed above an external front-end CMOS based MEA (see Appendix. C). Importantly, in these experiments there was no modification performed on the front-end system, it was kept completely external to the experimental platform.

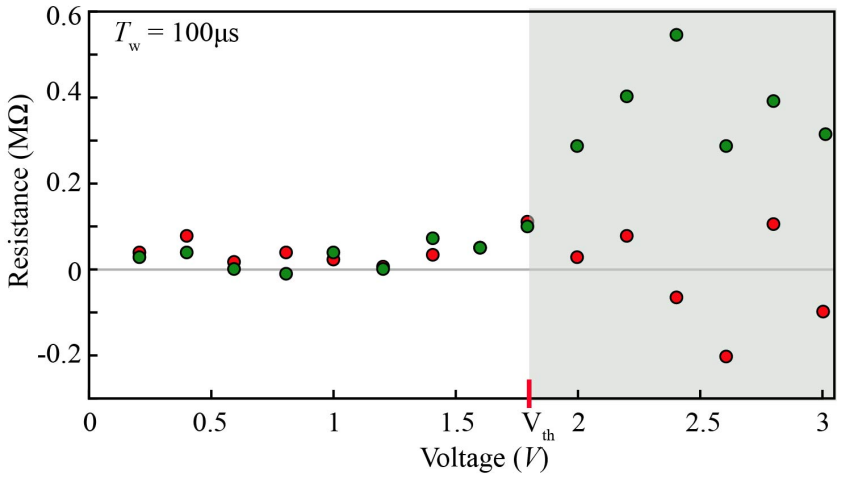


Figure 5.3: Volatility characterisation of the TiO_x nano-devices (200 nm x 200 nm). Input stimulus was applied to the DUT in range of 0 to -3 V in steps of 0.2 V. The pulse width, T_W, of the applied voltage was fixed, in this case, to 100 μs. Red and green circles indicate the non-volatile and volatile changes respectively. For the DUT the threshold of the device was found to be approximately -1.8 V. The grey band indicates the region where the device can be operated in the volatile region.

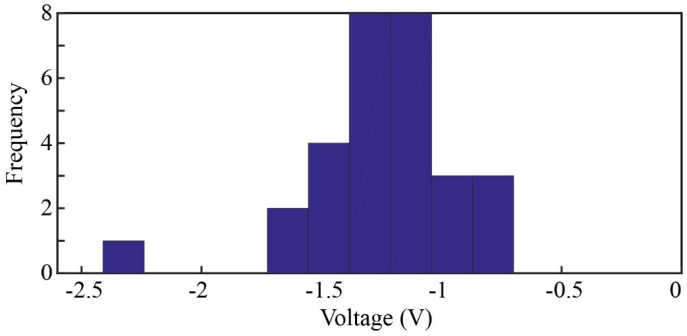


Figure 5.4: Range of inherent threshold voltages (V_{th-}) for the memristive devices in volatile region as extracted from 29 different devices.

Neural recordings and System Schematic[4]: arrive from the front-end as voltage-time data series in the range of approximately $\pm$0.5V. For the experiments in this work blocks of neural recordings containing approximately 63k samples recorded at a sampling rate of 12.2 kHz were used.

System Schematic: To re-iterate, in the first stage ('i- Fig. 5.2 (e)) the signals are pre-processed using a suitable Gain (G) and Offset (V_{off}) value. The neural signals are amplified such that the spikes but not the noise are above the volatile threshold of the DUT. Pre-processed neural recordings are then passed through the memristive devices in batches ('ii- Fig. 5.2 (e)) and the resistive state of the DUT is read periodically in real-time. In this experimental studies, a standardised signal processing protocol is followed (as described in chapter 4): For each thousand data points batch, the resistive state of the DUT is recorded at the beginning, then every 300 points and finally at the end of the batch compressing

[4]The setup in Fig.4.4 and Fig. 5.2(e) is identical. In the former methodology, memristive devices in non-volatile regime were used, and in the latter memristive devices in volatile regime.

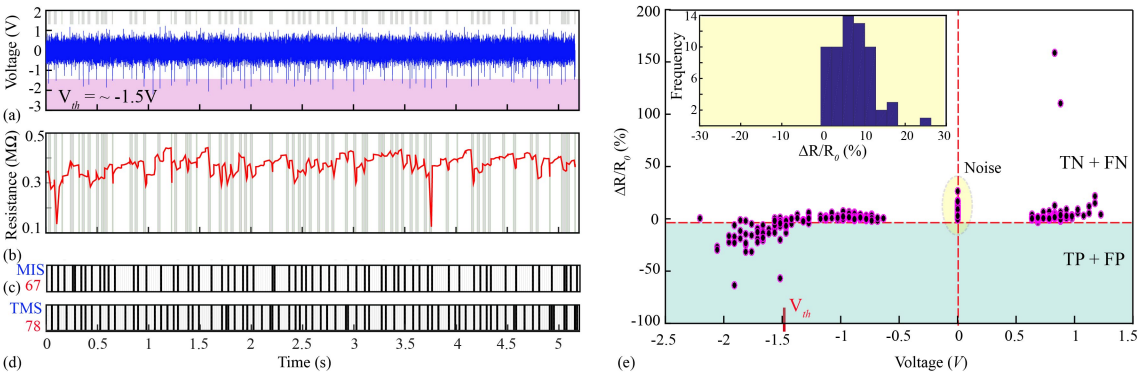


Figure 5.5: Spike-detection via volatile metal-oxide memristive device and benchmarking of the results against the TMS. (a) Neural recording used for biasing the DUT. Gain (G) and offset (V_{off}) values in this experiment were fixed at 3.2 and 0, respectively. The pink band indicates the inherent threshold voltage of the DUT (V_{th}) which is approximately equal to -1.5 V. (b) Resistive state evolution of the DUT with time in response to the neural recording in (a). Time intervals where our spike-detector identifies events are indicated in grey. (c) total number of spikes detected by volatile spike detector and (d) via the TMS. (e) Normalised changes in the resistive state of the device ($\Delta R/R_0$) in each bin are plotted as a function of highest voltage magnitude in each bin. The resistive state measurements in the yellow eclipse represents the noise measurements made at the end of each batch and beginning of the next batch with no neural feed. The inset represents the histogram for the noise measurements. The red dashed line on the horizontal axis represents the boundaries of the noise band estimated using the 4σ method that is $\mu \pm 2\sigma$ whilst the green band indicates the significant resistive state changes detected. TP: True Positives, FP: False Positives, TN: True Negatives and FN: False Negatives.

the data by a factor of 200. This leads to segmentation of the neural data into smaller bins. Pairs of consecutive measurements in each batch are used to estimate the resistive state changes, whilst the pair of resistive state measurements taken at the end of each batch and the beginning of the following batch is used to estimate noise or reference data values. Finally, resistive state changes compressed in this way are processed off-line and compared to noise ('iii-Fig. 5.2 (e)), with significant changes being estimated as spikes. The concept for real-time implementation at this stage is similar to the one proposed in section.4.5.1.

5.3 Memristive devices as volatile spike-detectors

Employing TiO_x memristive devices as volatile spike-detectors requires pre-processing any neural recording input to match the volatile operating region of devices as determined using the volatility characterisation algorithm. To illustrate the concept, a neural recording that contains a dense spiking pattern as illustrated in Fig. 5.5(a) was chosen. The operational parameters i.e. gain and offset values used for pre-processing the input neural recording were for this case fixed at 3.2 and 0, respectively. On biasing the target device with the neural recording the intrinsic reset capability of the DUT can be clearly noted in Fig. 5.5(b). For instance, the initial resistive state of the DUT is approximately 350 kΩ following which the device demonstrates metastable resistive state transitions towards a low resistive state in response to supra-threshold spiking events. In the case where no subsequent supra-threshold

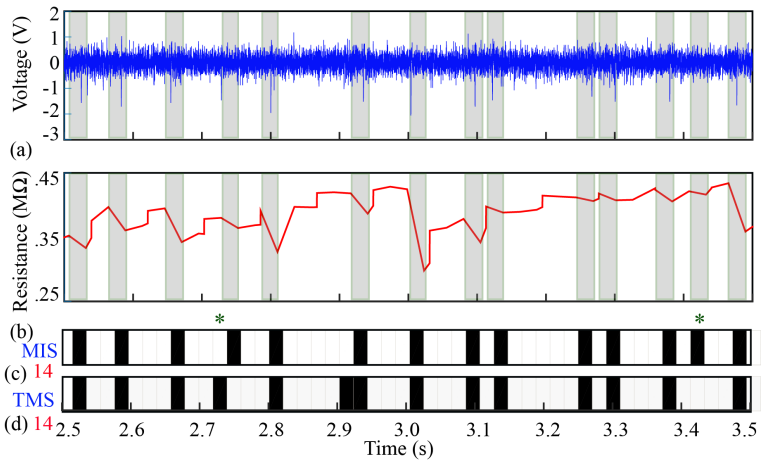


Figure 5.6: Biasing of 60 μm x 60 μm TiO_x devices using a neural recording in the volatile region. Close-up of the neural recording employed in Fig. 5.5(a) and the resistive state evolution in (b) for time window 2.5s - 3.5s, respectively. (c) and (d) Spikes detected by the volatile spike detection platform and the TMS system, respectively. In the close-up window the two systems agree for majority of instances except the point marked by *. The grey band indicates the bins with significant resistive state changes estimated as spikes. The quantification parameters for this specific recording are as follows: TP, FP, TN and FN are 58, 9, 166 and 20 respectively. The rate of TP (TPR) and FP (FPR) for this specific recording is estimated to be 74.3 and 5.14 respectively.

events occur, the state of the device relaxes back towards the initial device state, demonstrating an inherent reset. The close-ups of the Fig. 5.5(a, b) in window 2.5s - 3.5s are presented in Fig. 5.6 to illustrate this point more clearly.

The number and approximate timing of spikes are estimated after post-processing of the resistive state measurements obtained using the standard schematic as described in section 5.2. As shown in Fig. 5.5(e), resistive state change ($\Delta R/R_0$) in each bin is plotted as a function of highest voltage magnitude in each bin. The resistive state change magnitude distribution of noise measurements is used for identifying significant resistive state modulation that corresponds to spiking events. For the sake of clarity Fig. 5.7 demonstrates the difference of setting of the noise band in the volatile and non-volatile region. The inset of Fig. 5.5(e) represents the histogram for the noise measurements indicating an excessive inclination towards the positive polarity. Since the dominant stimulus polarity is negative and as a result neuronal activity-induced resistive changes are in the negative direction, the noise measurements in the positive direction are completely discarded and only the measurements in the negative direction are used to estimate meaningful noise band boundaries. This filters out the intrinsic reset transitions, which occur exclusively in the positive direction. Thus, noise band boundaries are estimated using only negative polarity noise measurements with a 4σ method (assuming Gaussian distribution), as indicated by the horizontal dashed line in Fig. 5.5(e). Everything outside this band in the negative region is considered as significant modulation corresponding to a spiking event whilst all events registered within this band are disregarded as these do not correspond to state modulations due to spiking events. Following this methodology, the total number of spikes detected, as shown in Fig. 5.5(c), is equal to 67.

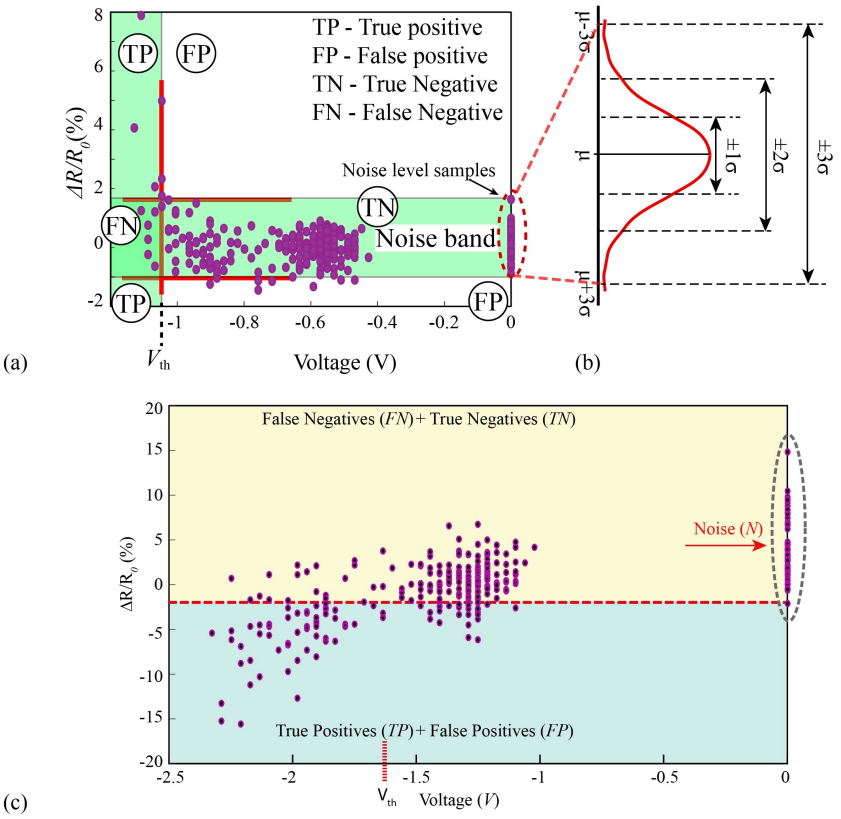


Figure 5.7: Comparison of the Noise band diagrams in the **volatile** and **non-volatile** region of operation [129]. (a) Noise-band diagram settings in the **non-volatile region**. Normalised plot of resistive state changes ($\Delta R/R_0$) in each bin is plotted as a function of maximum voltage magnitude in each bin. The noise-band (horizontal band marked in green colour) is estimated using the noise band measurements made at the end of each batch and the beginning of next batch with no interceding neural data points. (b) Assuming Gaussian distribution, mean and the standard deviation are estimated. 6 sigma method is chosen to set the boundaries of the noise-band. (c) Noise-band diagram in the **volatile region**. For the experiments with neural recordings in the volatile region most of the significant spikes are present in the negative polarity. Noise measurements are marked using a grey dashed eclipse indicating the spread of measurements which notably is inclined in the positive polarity. For the noise band boundary settings in the volatile region, the noise band measurements in the positive polarity are discarded and only the measurements in the negative polarity are used. The boundary is set using the 4-sigma method which is indicated by the horizontal red dashed line dividing the resistive state changes in two quadrants i.e. TP + FP (spikes detected) and FN + TN.

The performance of volatile memristor spike-detector is then benchmarked against the established state-of-the-art template matching system (TMS, [4, 121])[5]. As depicted in Fig. 5.5(d), the total number of spikes determined by the TMS is equal to 78. In Fig. 5.5(e), approximately -1.5 V represents the inherent threshold voltage of the DUT. The negative quadrant (green) represents the spikes detected i.e. sum total of True Positives (TP) and False Positives (FP) whilst the positive quadrant indicates the False Negatives (FN) and True

[5]The results are benchmarked against the established template matching system and the quantification parameters are redefined. TP (TN) indicates when the two system agree (disagree) for the presence of spike. FP indicates a spike detected by spike-detection platform and not by the TMS system. FN indicates a spike indicated by the TMS system and not spike-detection platform. The rate of TP (TPR) and FP (FPR) are estimated using the following equations: Rate of TP = TP/(TP+FN), Rate of TP = FP/(FP+TN)

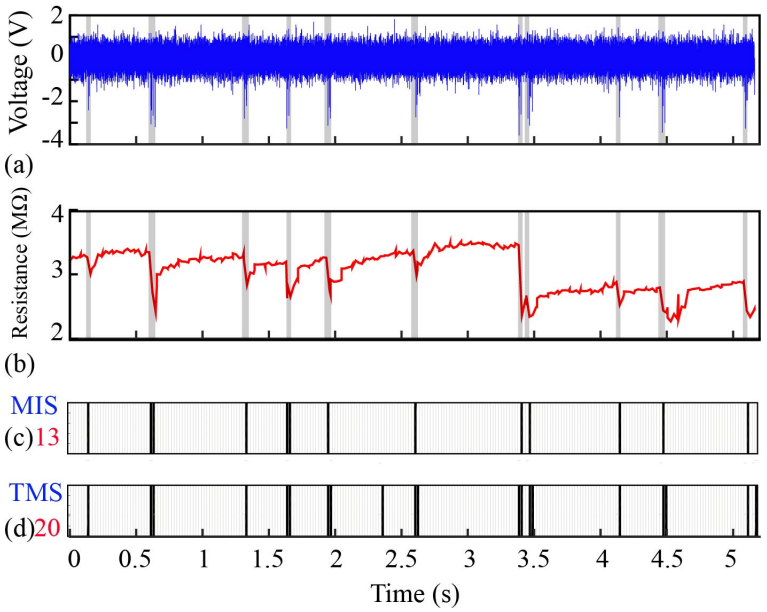


Figure 5.8: Robustness of TiO_x devices. Biasing of 60 μm x 60 μm devices using a different neural recording in comparison to the one presented in Fig. 5.5 in the volatile region. (a) Neural recording used for the biasing of the DUT. (b) Resistive state evolution of the DUT with time. (c) and (d) Spikes detected by volatile spike-detection platform (13) and TMS (20) respectively. The grey band indicates the bins with significant resistive state changes estimated as spikes in the output of our platform. The quantification parameters for this neural recording are as follows: TP, FP, TN and FN are 13, 0, 233 and 7 respectively. The rate of TP (TPR) and FP (FPR) is found to be 65 and 0 respectively.

Negatives (TN). These quantification parameters are used to evaluate the sensitivity of the volatile system through the true positive rate (TPR) and false positive rate (FPR) of detection. Assuming TMS to be a perfect spike detector, the two values are estimated to be 74.35% and 5.14% respectively. Both values represent a significant improvement in performance compared to when the devices are operated in the non-volatile region. For example, for the same neural recording in Fig. 4.23, TPR and FPR after introducing optimised manual frequent resets is equal to 60% and 30% respectively.

The concept of volatile spike-detection via memristive devices was initially validated in Fig. 5.5 using a memristive device of dimensions 60 μm x 60 μm. The robustness of the devices for the same dimensions is illustrated in Fig. 5.8. A neural recording with significantly different spiking pattern in comparison to Fig. 5.5(a) was used to bias a memristive device. TPR and FPR in this case were equal to 65% and 0%, respectively. Besides, similar results from fifteen different devices are tabulated in Table. 5.1, where the highest TPR and FPR obtained is equal to 88.4% and 13% respectively.

Table 5.1: Robustness of TiO_x memristive devices. For this experiment, devices with different dimensions i.e. 60 μm x 60 μm and 200 nm x 200 nm and different neural recordings with significantly different spiking pattern were used. For the pre-processing of the neural recording, the operational parameters that is G and V_{off} were varied. The quantification parameters are indicated in the table with the estimated rate of true positives and false positives. The rows in yellow indicate the performance of different memristive devices on the same input data. VMS: Volatile Spike Detection platform, TMS: Template matching system, TP: True Positives, FP: False Positives, TN: True Negatives, FN: False Negatives.

Device Dimensions	Vth-(V)	Gain	Offset	VMS	TMS	TP	FP	TN	FN	Rate of TP (%)	Rate of FP (%)
μm^2											
60 x 60	-1.4	3.2	0	67	78	58	9	166	20	74.35	5.14
60 x 60	-1.4	2.6	0	39	20	13	26	207	7	65	11.15
60 x 60	-1.17	2.4	-0.6	22	20	11	11	222	9	55	4.7
60 x 60	-1.7	2.4	-0.4	15	20	13	2	231	7	65	0.85
60 x 60	-1.1	2	0	15	20	13	2	231	7	65	0.85
60 x 60	-0.7	2.2	0	74	78	47	27	148	31	60.25	15.4
60 x 60	-0.7	2.2	-0.2	52	78	42	10	165	36	53.8	5.71
60 x 60	-1.34	2.2	0	12	20	12	0	233	8	60	0
60 x 60	-1	4.4	0	56	78	25	31	144	53	32	17.7
60 x 60	-1	4.4	0	57	78	40	17	158	38	51.3	9.7
60 x 60	-1	4.8	0	76	78	53	23	152	25	70	14
60 x 60	-1.2	4.8	-0.2	102	78	69	33	142	9	88.46	18.8
60 x 60	-1.2	2.9	0	62	78	59	3	172	19	75.6	1.71
60 x 60	-1.2	2.6	0	63	78	60	3	172	18	77	1.71
60 x 60	-2.41	4.4	0	13	20	13	0	233	7	65	0
60 x 60	-0.7	2.6	-0.4	92	78	69	23	152	9	88.4	13
nm^2											
200x200	-1.1	2.8	-0.4	54	78	31	24	151	47	40	13.7
200x200	-1.25	2.6	-0.6	46	78	35	11	164	43	44.8	6.28
200x200	-1.3	2.6	-0.6	30	78	21	9	166	57	27	5
200x200	-1.63	2.8	-0.4	10	78	10	0	175	68	12.8	0
200x200	-1.25	2.6	-0.6	42	78	32	10	165	46	41	5.74
200x200	-1.25	2.6	-0.6	78	78	54	24	151	24	70	13.7
200x200	-1.3	Max= 0	Min= -2.2	44	78	22	22	153	56	28.2	12.5
200x200	-1.13	2.6	-0.6	47	78	27	20	155	51	34.6	11.4
200x200	-1.2	3	-0.6	47	78	27	20	155	51	34.6	11.4
200x200	-1.3	3	-0.6	42	78	27	15	160	51	34.6	8.5
200x200	-1.3	2.6	-0.6	74	78	43	31	144	35	55.12	17.71
200x200	-1.4	2.6	-0.6	72	78	37	35	140	41	47.43	20
200x200	-1.45	2.6	-0.6	43	78	27	16	159	51	34.61	9.14

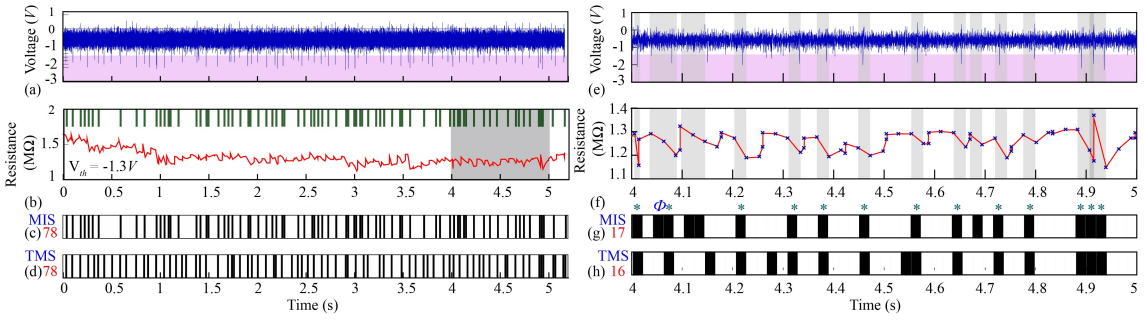


Figure 5.9: Scalability of the TiO_x devices to the nanoscale dimensions and operation of nano-devices as volatile spike detectors. (a) The employed neural recording pre-processed using gain and offset value of 2.6 and -0.6 respectively. The pink band indicates the inherent threshold voltage of the DUT (V_{th}). (b) Resistive state changes in the target DUT in response to the neural data in (a). The active core area of the nano-devices used for this experiment is 200nm x 200nm. The threshold voltage of the DUT is -1.3 V. (c), (d) Total number of spikes detected by both volatile spike detection platform and the TMS system is equal to 78. The black bins indicate the spike positions. (e) and (f) are close-ups for the neural recording and the resistive state evolution shaded grey in (a) and (b) respectively. 'X' cross mark in red indicates the positions where the resistive state measurements are taken. Time intervals where an event is detected are represented in grey. (g) and (h) spikes detected by both the system. The asterisk '*' indicates the positions where the two systems agree and 'ϕ symbol discusses a specific case of discrepancy between the two systems.

5.4 Nanoscale-metal oxide devices as volatile spike encoders

Memristive device technologies offer huge advantages in terms of scalability and can be accommodated in Back-End-Of-Line (BEOL) of CMOS technologies, which can potentially benefit future implantable neuroprosthetic platforms. The presented approach was further supported with downscaled memristive devices of 200nm x 200nm dimensions. The devices were prototyped using the fabrication procedure described in the Appendix A. In this case, the same neural recording as shown in Fig. 5.5(a) was employed and similarly the obtained results were benchmarked against the TMS. For this case the neural recording was amplified using G and V_{off} values of 2.6 and -0.6 respectively, as illustrated in Fig. 5.9.

The transient response of the target device resistance is illustrated in Fig. 5.9(b) along with the corresponding neural recording time series. Compared to the previous case, this device was operated at a higher resistive state region of 1-1.5 MΩ. The spikes detected from both this and the TMS system was calculated to be 78, as shown in Fig. 5.9(c) and (d) respectively. The response of the target device can be more closely scrutinized in Fig. 5.9(e) and (f), which illustrates the neural recording and resistive state response of the DUT during the 4-5sec window in Fig. 5.9(a) and (b) respectively. Marker 'x' is used to indicate the resistive state measurements, following the standard schematic. Importantly, between each pair of measurements the system is blind to the behaviour of the devices, however the bin size is a user-defined design parameter. Sampling rate (detection accuracy) and power consumption can be traded against each other.

The memristive devices undergo a resistive drop in response to the supra-threshold events. For instance, in the first bin shown in Fig. 5.9(f), the state of the device drops from approximately 1.3 MΩ to 1.15 MΩ. The asterisk '*' in Figs. 5.9(g, h) further confirms concurrence between the two systems and that is true for the majority of the instances. More specifically, the two systems agree for 13 over 17 instances detected, as shown in Fig. 5.9(g). The symbol 'ϕ' indicates an instance of mismatch between the two systems. Interestingly, at this instance the volatile spike detection system detects a neural event that closely resembles a spike, while the TMS fails. On further careful examination of the neural recording it was observed that the TMS also fails to detect apparent spiking events at approximately 1.1s, 1.6s, 1.9s, 2.2s, 2.5s, 2.7s, 3.3s, 3.6s, 3.9s, some of which are detected by the proposed system [136]. On the other hand, in Fig. 5.9(g), the volatile spike-detection system fails to detect spikes occurring at 4.55s. From these observations it can be safely concluded that, although TMS is assumed to be a perfect spike-detector for these experiments, however, in practical operation this is not the case. Benchmarking of the detected spikes revealed a TPR and FPR of 70% and 13.7% respectively. Spike detection results carried out by thirteen different nano-devices with pre-processed neural recordings are reported in Table. 5.1.

5.5 Optimising operation parameters for volatile devices

A Receiver Operating Characteristic sensitivity curve, that is defined as the rate of TP vs FP, is illustrated in Fig. 5.10(a) for devices of different dimensions. The details for the quantification parameters are presented in Table. 5.1. It can be seen that as the dimensions of the devices are reduced from 60 μm x 60 μm down to 200nm x 200nm the detection accuracy of the system is reduced. One way for optimising the performance of the downscaled devices is to tune the gain and offset parameters to an optimum level. Hence additional test were performed with three different gain settings 2.2, 2.4 and 2.6 and a fixed offset setting at -0.6 V, as illustrated in Fig. 5.10(b). For every round of gain, the experiment was repeated for five times and the details of the quantification parameters are illustrated in Table. 5.2. The asterisk symbol, '*' represents the average of the quantification parameters for every round of gain. Thus, an improvement in the gain resulted into an increase in rate of TP from 23.1% to 46.7%, and these numbers can be plausibly further improved in the future by estimating the optimum bin size for practical implementations of the volatile spike-detector and exploring different stacks of memristive devices with engineered volatile characteristics might also be investigated for optimising performance.

5.6 Mitigating Noise-effects in volatile metal-oxide devices

These section further examines the effect of sub-threshold events and operational parameters such as signal amplification and offset on the response of memristive devices in volatile region

Table 5.2: Optimisation of 200nm x 200nm TiO_x memristive devices. Gain and offset parameters for one of the 200nm x 200nm device was optimised. Three different values of gain i.e. 2.2, 2.4 and 2.6 were used with constant offset values i.e. -0.6. For each round of gain the experiment was repeated five times. The quantification parameters for each round when benchmarked against the state-of-the-art template matching system are indicated in the illustrated table. VMS: Spikes detected by our platform, TMS: Template matching system, TP: True Positives, FP: False Positives, TN: True Negatives, FN: False Negatives.

Device Dimensions (nm^2)	Gain	Offset	VMS	TMS	TP	FP	TN	FN	Rate of TP (%)	Rate of FP (%)
200x200	2.2	-0.6	66	78	32	34	141	46	41.02	19.42
200x200	2.2	-0.6	25	78	17	8	167	61	21.79	4.57
200x200	2.2	-0.6	23	78	13	10	165	65	16.67	5.71
200x200	2.2	-0.6	34	78	19	15	160	59	24.35	8.57
200x200	2.2	-0.6	11	78	9	2	173	69	11.53	1.14
									23.07	7.88
200x200	2.4	-0.6	32	78	25	7	168	53	32.05	4
200x200	2.4	-0.6	47	78	29	18	157	49	37.18	10.28
200x200	2.4	-0.6	40	78	26	14	161	52	33.34	8
200x200	2.4	-0.6	53	78	32	21	154	46	41.02	12
200x200	2.4	-0.6	35	78	27	8	167	51	34.61	4.57
									35.64	7.77
200x200	2.6	-0.6	62	78	38	24	151	40	48.71	13.71
200x200	2.6	-0.6	50	78	35	15	160	43	44.87	8.57
200x200	2.6	-0.6	60	78	40	20	155	38	51.28	11.42
200x200	2.6	-0.6	34	78	26	8	167	52	33.34	4.57
200x200	2.6	-0.6	74	78	43	31	144	35	55.128	17.71
									46.66	11.19

[137]. In these experiments, the effects of sub-threshold, unipolar and bipolar neural events on a TiO_x nano memristive device are analysed. It is found that inclusion of events in positive polarity leads to subsequent increase in the number of false events when benchmarked against state-of-the-art spike detector (template matching system). The performance of the system is thereafter optimised by determining optimum amplification settings and employing an offset such that positive polarity events in the input signal are minimised.

5.6.1 Effects of unipolar and bipolar neural events

For this experiment raw neural recording i.e. prior to pre-processing was modified. To generate events in one polarity as shown in Fig. 5.11(a), the threshold was set to -0.5 V and all the events above were discarded to study only the effects of spiking events in one polarity. The remaining events were then amplified to [0, -3 V] range. Similarly for Fig. 5.11(b),

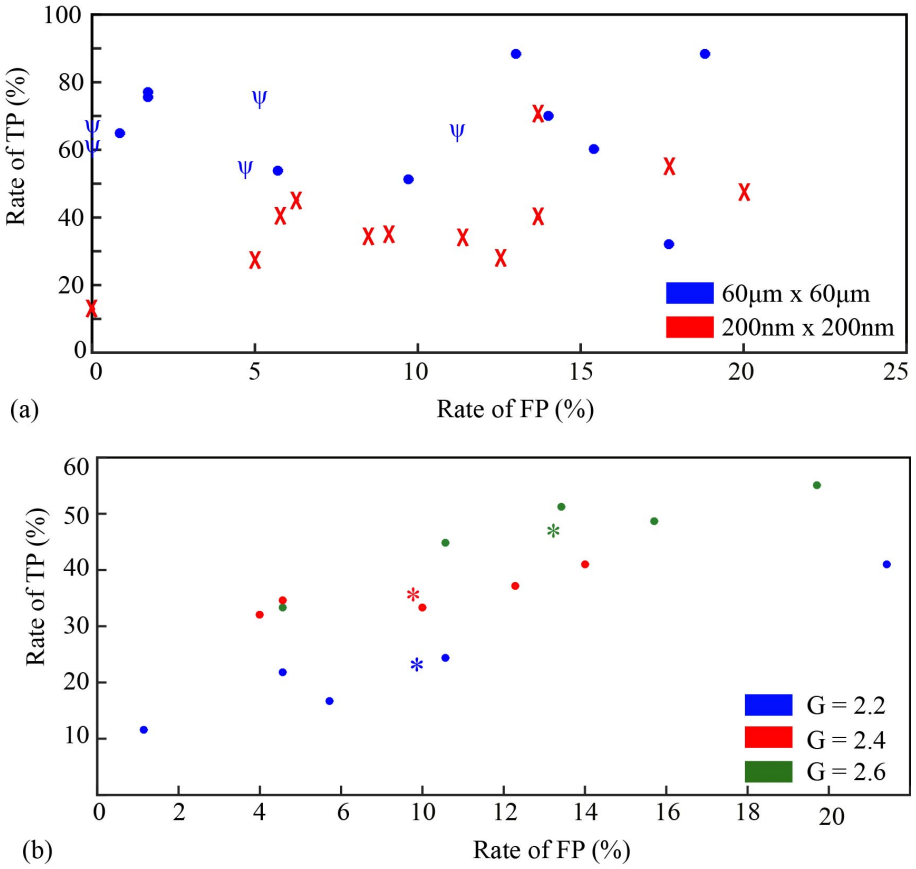


Figure 5.10: Comparison of Receiver Operating Characteristics (ROC) that is rate of true positives vs rate of false positives for memristors of micro- and nano- dimensions. (a) The blue and red colour indicates the ROC curves for 60 μm x 60 μm and 200nm x 200 nm devices respectively. Symbol psi 'ψ' and circle in blue colour indicates different neural recordings with different spiking pattern used for the DUT. (b) Optimisation of spike detection capability for one of the 200nm x 200nm device. Blue, red and green colour indicates the three different gain parameters i.e. 2.2, 2.4 and 2.6 respectively chosen for the experiment where the offset was kept constant at -0.6. For every gain the experiment was repeated five times. Asterisk symbol '*' indicates the average of the quantification parameters for each round of gain.

two threshold's were set at +0.3 V and -0.5 V and the remaining events above and below respectively were amplified to [+0.5 V, -3 V] range to the study the relation of spiking events in one polarity to events in different polarity. The resistive state transitions of a DUT in response to the pre-processed signals are illustrated in Fig. 5.11(a) and (b) respectively. The grey band are spikes detected by volatile spike detection platform.

The effect of positive polarity events for instance at $\approx$ 0.3 s, 1.6 s and 2.4 s as illustrated by green bands in Fig. 5.11(b) can be noted. The quantification parameters are presented in Table. 5.3 and it can be noted that the inclusion of positive polarity events increases the TPR from 44.8% to 58.9%. However, the FPR correspondingly increases from 1.14% to 6.28%. Moreover, the resistive state change within a bin (300 data points) is attributed to the strongest pulse, therefore at number of bipolar instances such as 0.4 s, 1.5 s, 2 s, 3.4 s, 4.4 s etc the change in resistive state is assigned to negative polarity events.

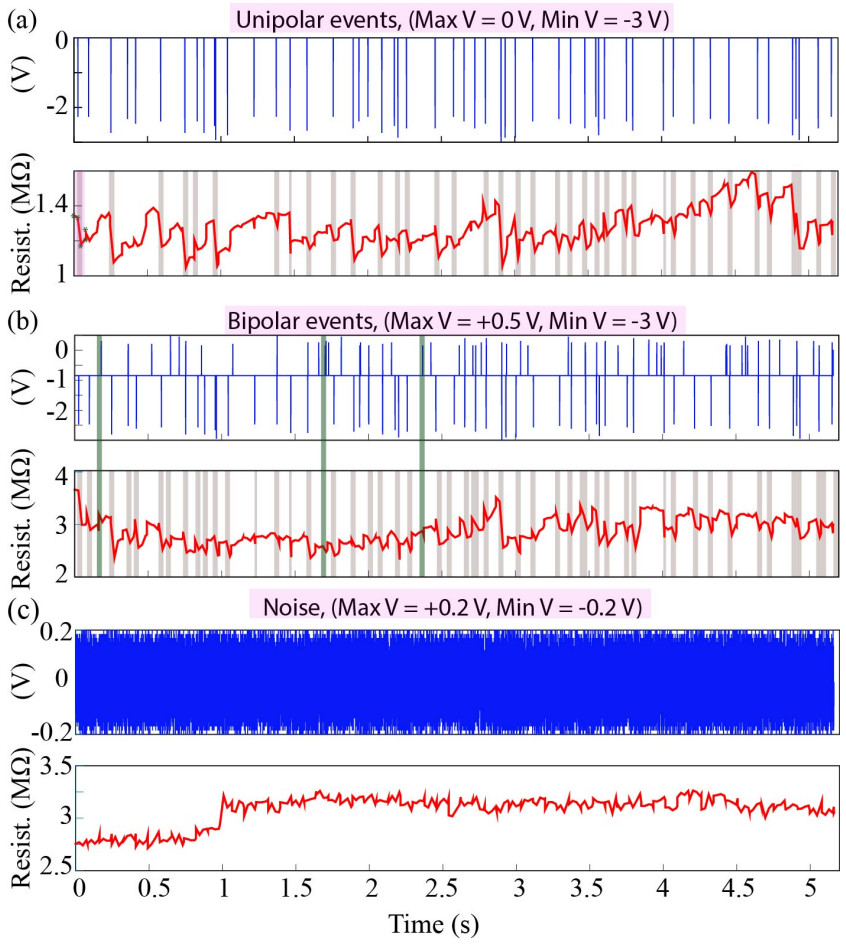


Figure 5.11: Analysing the effect of events on nano volatile memristive device. (a), (b) (c) (Upper panel's) Biasing a device-under-test (DUT) with unipolar, bipolar and noise events, pre-processed using different settings. (a), (b), (c) (Lower panel's) Response of the DUT in response to the biased events. The grey bands indicates the spikes detected by our platform. The green bands indicates few instances of spikes detected due to positive polarity events.

Table 5.3: Comparison of quantification parameters for the unipolar and bipolar events in 5.11(a) and (b). VMS: Spikes detected volatile spike-detection platform, TMS: Spikes detected by Template matching system, True positives (TP), False Positives (FP), True Negatives (TN), False Negatives (FN), True Positive Rate (TPR) and False Positive Rate (FPR).

	Max/Min (V)	VMS	TMS	TP	FP	TN	FN	TPR (%)	FPR (%)
1.	0/-3	37	78	35	2	173	43	44.8	1.14
2.	+0.5/-3	57	78	46	11	164	32	58.9	6.28

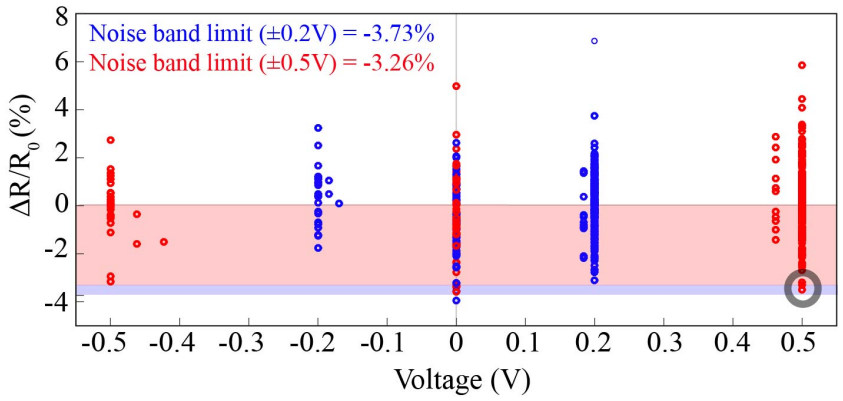


Figure 5.12: Comparison of noise band diagrams i.e. $\Delta R/R_0$ vs voltage plots for sub-threshold events. The blue and red colour illustrates the response of the DUT in response to the neural events in Fig. 5.11(c) when mapped to approximately ± 0.2 V and ± 0.5 V respectively.

5.6.2 Effect of sub-threshold events

The inherent volatile threshold V_{th-} of the employed devices are in the range of ≈ 0.5 V to 2.5 V. To understand the effect of sub-threshold events on the volatile behavior of the devices, the following experiment was carried out. Similarly to Fig. 5.11(a), two thresholds were set to +0.3 V and -0.5 V and the activity within was amplified to [+0.2, -0.2] V range deliberately excluding the spikes, as shown in Fig. 5.11(c). Notably, a drift in the resistive state of the DUT from ≈ 2.8 Ω to ≈ 3.2 MΩ in response to the sub-thresholds can be noted.

For better understanding, the normalised changes in the resistive state of the DUT in each bin ($\Delta R/R_0$) is plotted as a function of the highest voltage magnitude in each bin (as shown in blue colour, Fig. 5.12) for [+0.2, -0.2] V and [+0.5, -0.5] V. The noise measurements are illustrated on x=0, the measurements in positive polarity are discarded and the noise band limit is set using 4σ method using negative noise measurements (optimised for this system)[6]. The noise band limit is estimated to be -3.73%. It can be observed that all resistive state changes at $\approx \pm 0.2$ V fall within the estimated noise band limits. Since, we account for normalised changes in resistive state in each bin, slight drift in the resistive state of the DUT is naturally filtered off and therefore these voltages doesn't affect the output of the proposed neural detector.

However, on amplifying the neural events in Fig. 5.11(c) to ± 0.5 V, $\Delta R/R_0$ vs V plot in red colour in Fig. 5.12 is obtained. The noise band limit is estimated to be -3.26 % and the resistive state changes in black circle at $+0.5$ V fall outside the noise band thus registered as spikes confirming the effect of positive polarity events. Following these observations positive polarity events were modulated using an optimum offset value such that strength of events in positive polarity were minimised, which also justifies the importance of this key operational parameter.

[6]It should be noted that resistive state changes in positive polarity are discarded. Since, the devices relaxes back in positive polarity following a resistive state transition, inclusion of noise events in positive polarity doesn't indicate fair estimation of the noise band limits.

Table 5.4: Comparison of different gains with constant offset values for a single neural recording

	Gain	Offset	VMS	TMS	TP	FP	TN	FN	Rate of TP (%)	Rate of FP (%)	Noise Max (%)	Noise Min(%)	Noise band limit (%)
1.	2.6	-0.6	28	78	23	5	170	55	29.48	2.8	+7.52	-3.15	-4.48
3.	3.2	-0.6	56	78	47	9	166	31	60.25	5.14	+23.51	-2.99	-4.57
4.	3.6	-0.6	63	78	46	17	158	32	58.9	9.71	+21.15	-3.18	-4.196

5.6.3 Optimum settings for volatile devices

Considering the observations made in Fig. 5.11 and Fig. 5.12, the operational parameters for pre-processing the neural signal were reconsidered and optimised. The same neural recording (as shown in Fig. 5.11) was amplified using different gain values whilst the offset value was kept constant. Figure. 5.13 (a), (b) and (c) illustrates the $(\Delta R/R_0)$ vs V plots for the response of a DUT when biased with same neural recording with G value set as 2.6, 3.2 and 3.6 respectively. The offset was kept constant at -0.6 V. The number of spikes detected by the volatile spike detection platform was benchmarked against the TMS and the quantification parameters are demonstrated in Table. 5.4.

Visually in Fig. 5.13(a,b and c), the grey area indicates the noise measurements that have been discarded. The horizontal blue, red and green bar indicates the noise band with black dashed line indicating the noise band limit. Everything outside the noise band in the negative quadrant indicates the number of detected spikes. This can be also seen in histogram plots presented in Fig. 5.13(c, d and e), where the yellow bands indicate the spikes detected by the system. It was observed that as the gain was increased from 2.6 to 3.2, the rate of TP almost doubled i.e from 29.48% to 60.25%. At the same time the corresponding increase in the rate of FP was from 2.8% to 5.14% (see Table. 5.4). More clearly, on comparing Fig. 5.13 (a),(b) the pink quadrants in the latter indicates a significant increase in the detected spikes thus indicating the positive effects of increase in gain. The noise band limit in both the cases is $\approx$ -4.5%. Similar observations were made for G = 2.8, where the rate of TP was found to be 38.48 %.

Interestingly, as the gain was further increased from 3.2 to 3.6, the rate of TP remained approximately the same, however, the rate of FP almost doubled to 9.71% where the noise band limit was estimated to be -4.2% (see Table. 5.4). Gain (3.6) and V_{off} (-0.6 V) for this DUT in essence illustrates the optimum point of operation. This can be generalised to an understanding where most significant supra-threshold events are above the inherent threshold of the device and insignificant events/noise are below the threshold. An increase in gain beyond an optimum point leads to inclusion of insignificant amplified noise events above the threshold of the DUT and thus effects the sensitivity of the neural detector by increasing the FP.

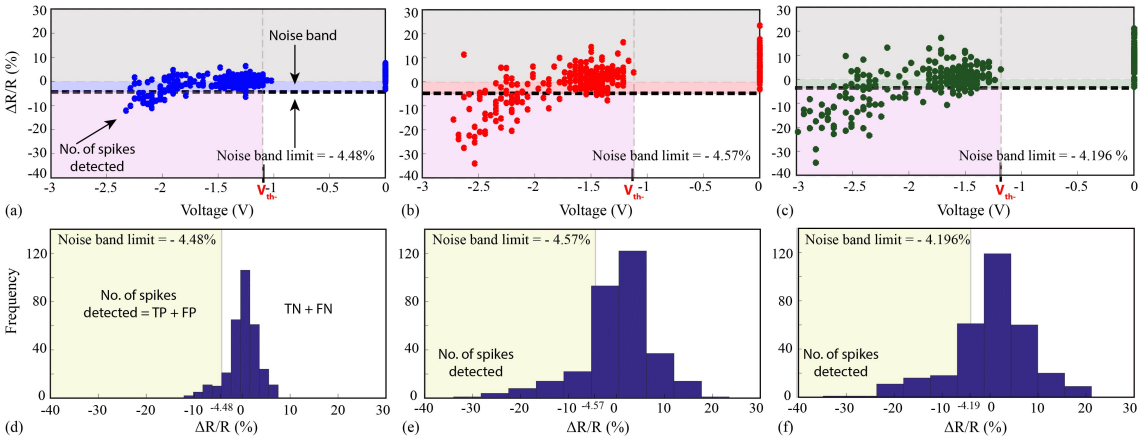


Figure 5.13: A single DUT was biased with the neural recording in Fig. 5.11. (a), (b), (c) The plots for $(\Delta R/R_0)$ with respect to highest voltage magnitude in each bin for gain equal to 2.6, 3.2 and 3.6 respectively. The offset was kept constant at -0.6 V. (d), (e), (f) Corresponding histograms for $(\Delta R/R_0)$ for the three cases respectively. The yellow bands indicate the number of spikes detected. V_{th-} indicates the inherent threshold voltage of the DUT ($\approx$ -1.2 V).

5.7 Discussion

Power Savings: A major prerequisite in the implementation of fully implantable neural interfaces is to develop electronic platforms capable of executing low-power and on-site signal processing of the acquired neural activity. These experiments demonstrate that metal-oxide memristive nanodevices can be operated in their volatile region as scalable spike-detection elements. The optimised performance is primarily due to the intrinsic resetting capability of such elements that eliminates the need for manually resetting the device to avoid its saturation. The major advantage of the these experiment lies in the prospects for dramatically reducing the power consumption per channel of implantable neural spike-detection circuits. Consider the example illustrated in Chapter 4, Fig. 4.23: for a neural recording consisting of approximately 63k data-points the target device is manually reset eleven times using a pulse of positive polarity of 100 μs pulse width. Realistically assuming +3V as the operating voltage and 10kΩ as the resistive state of the device, the amount of power dissipated including the reset operation by this system will be approximately equal to 3 mW per channel. The detailed power estimation methodology is presented in Appendix. D. The reset operation consumes about 250 nJ of energy and 11 such resets for one neural recording would consume 2.7 μJ of energy which will be conserved in the volatile region.

Additional power savings also stem from the fact that the volatile regime exploited in this case occurs at relatively high resistive state regions (e.g., consider the electrical characterisation results of nano-devices in Fig. 5.2(c and d). Assuming the operating resistive state and pulse width of the DUT to be 1MΩ and 1 μs respectively, the amount of energy dissipated per channel can be estimated using the standard batch processing schematic. The read operation with standard read out voltage of +0.2V will be equal to 0.2 pJ (0.04 pJ multiplied by 5 reads per batch) and the write operation at 3V will cost 9 nJ (9pJ multiplied by 1000 samples per

batch). The average power dissipated per channel at 12.2 kHz as the sampling frequency can thus be estimated to be approximately 100 nW (see Appendix. D). Voltage-time trade-off [131] can further assist in reducing the power dissipation by one order of magnitude, for instance, operating the same with 100 ns pulse widths will further reduce the power dissipated to 10 nW per channel per device. These measured results are already significantly lower than current state-of-the-art spike detectors projected at approximately 700 nW [24]. Naturally, a full system application would include other power overheads required by the memristor read-out and biasing circuitry, which are not considered in the present work as such circuitry will be cited in the periphery and would be shared (multiplexed) by each memristive device.

Device Variability: - One can also argue that the usability of the memristive devices could be limited due to the recurrently observed device-to-device variation. Nonetheless, this approach exploits normalised changes in each bin instead of absolute changes, therefore minimising any performance compromises due to device variability. Besides, the introduced volatility characterisation module leads to an automatic en-masse characterisation of memristive devices and determines the safe region for operation of devices in the volatile region. Clearly, if higher and more invasive voltages were applied, the devices may switch to the non-volatile region and change their baseline operating region [133].

The focal point of this work is to push the scaling limits for the employed devices towards deep-submicron arrays [116], along with optimised spike-detection capability. For future, the use of alternative or engineered materials as the devices active cores could allow tuning the volatile characteristics of devices [115] and in turn the self-reset achieved by the spike-detector. On the other hand, these parameters will also crucially depend upon the specific application under study, accounting for the sensitivity required for a particular application [7, 71, 138]. It should be further noted that the resistive state modulation encountered by these prototype devices is related to the amplitude and polarity of the input neural signal and the richness of the signal is in principle preserved [139]. One can possibly open a new avenue for scalable and power efficient on-node spike-sorting; a prerequisite for fulfilling the electroceuticals and more broadly the bioelectronics vision [138].

5.8 Summary

In summary, taking the cue from how biological synapses compress spiking information in post-synaptic conductance changes, a novel concept for neural-spike detection and encoding using intrinsic volatile behaviour of nanoscale metal-oxide memristive devices is demonstrated. Presented results prove that single nanoscale volatile devices are capable of identifying significant spiking activity in the input neural waveform in a highly power efficient manner. The effects of sub-threshold, bipolar and unipolar neural events on the volatile behavior of the device are studied. By mitigating the noise effects and setting an optimum point of key operational parameters, the performance of the neural detector is enhanced.

Chapter 6

Spike Sorting using metal-oxide devices in non-volatile regime

6.1 Introduction

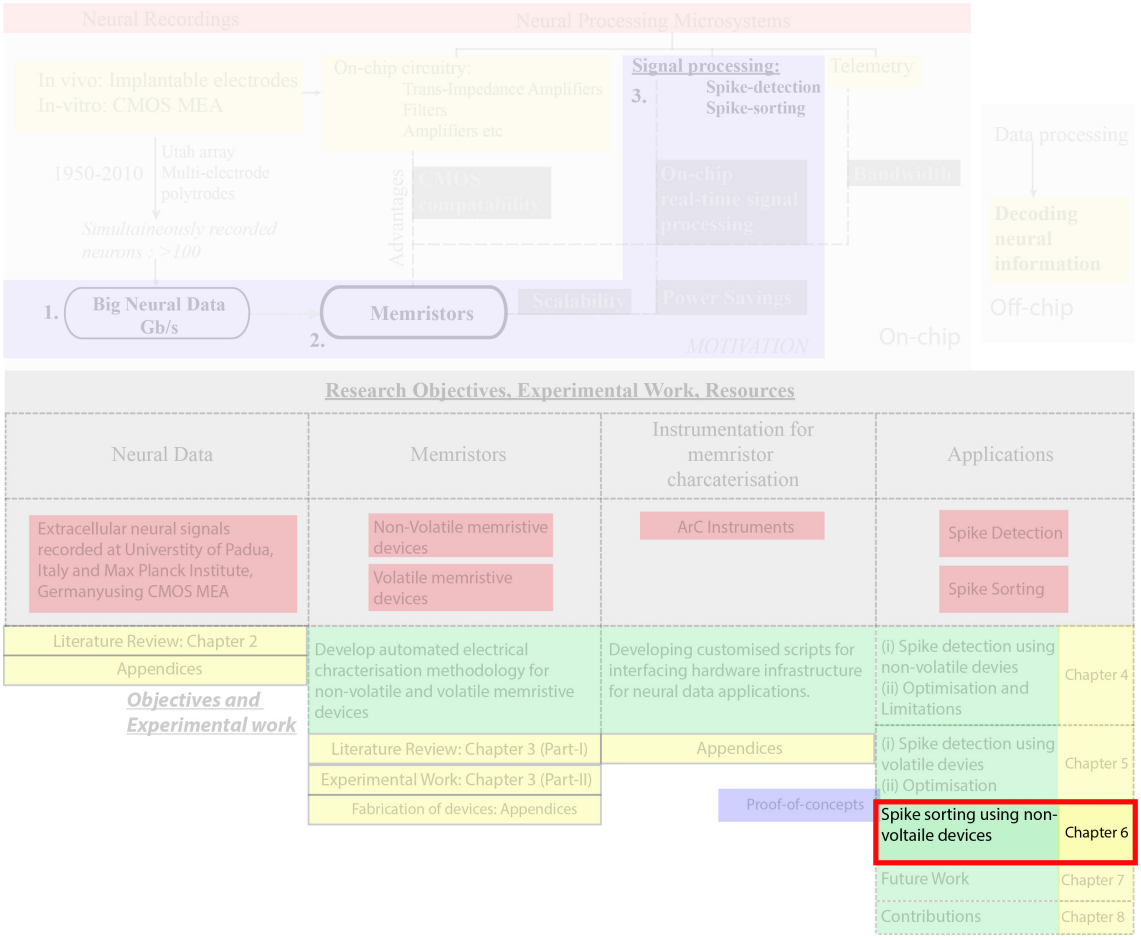


Figure 6.1: The research objective of this experimental chapter.

In the last chapters, it was demonstrated how memristive devices can perform the task of 'spike detection' by relying on metal-oxide memristors. When presented with an input voltage waveform the devices accumulate changes in resistive state linked to the instantaneous signal magnitude and polarity, so long as this exceeds the device threshold. However, another major question in neuroscience research is to analyse how much information is processed from the large ensemble of neurons, and this is majorly limited by the computational resources required for 'sorting' the acquired data [34].

Spike sorting is the procedure of identifying the activity of individual neurons from data collected through electrophysiological experiments [12, 60, 140, 141]. Typically this involves processing raw neuronal data by first detecting the presence of action potential (spiking) activity, then extracting appropriately chosen features and finally, clustering the results; each cluster corresponding to an individual neuron. Importantly, memristive devices perform the function of 'detection' and 'feature extraction'. Classification and Clustering is performed in post-processing stage by the investigator. The complexity of the spike-sorting algorithms remains as one of the major bottlenecks in emerging neuroprosthetic applications [12, 13, 19, 67].

In this chapter, it is shown how the intrinsic analogue programmability of memristive devices can be exploited to perform 'spike-sorting'. Since the neural waveform is transduced in the memory state changes of the memristive devices, the number of features that can be exploited for spike-sorting is limited. However, the richness of the input waveform is preserved in the resistive state changes of the DUT, which are sensitive to both 'amplitude' and 'polarity' of the biased event. Thus, the effect of these two parameters on the resistive state changes of the DUT are analysed (section. 6.2).

Section. 6.3 demonstrates the spike-sorting results from publicly available simulated neural signals constructed based on electrophysiological recordings where 'ground truth' is known. Finally, section. 6.4 concludes the chapter.

6.2 Spike sorting using non-volatile metal oxide devices

When presented with an input voltage waveform the devices accumulate changes in resistive state linked to the instantaneous signal magnitude and polarity, so long as this exceeds the device threshold. This section demonstrates how intrinsic analogue programmability of memristive devices can be exploited to perform 'spike-sorting'.

6.2.1 Proof-of-concept/Initial Results

For these experiments, 60x60 μm^2 solid-state TiO_x devices with a vertical stack structure of Ti/Pt/TiO_x/Pt (5/10/25/10 nm) fabricated on Si/SiO_2 substrate were used (see Appendix. A)[139]. Before use, all the devices were electroformed by employing a ramp of

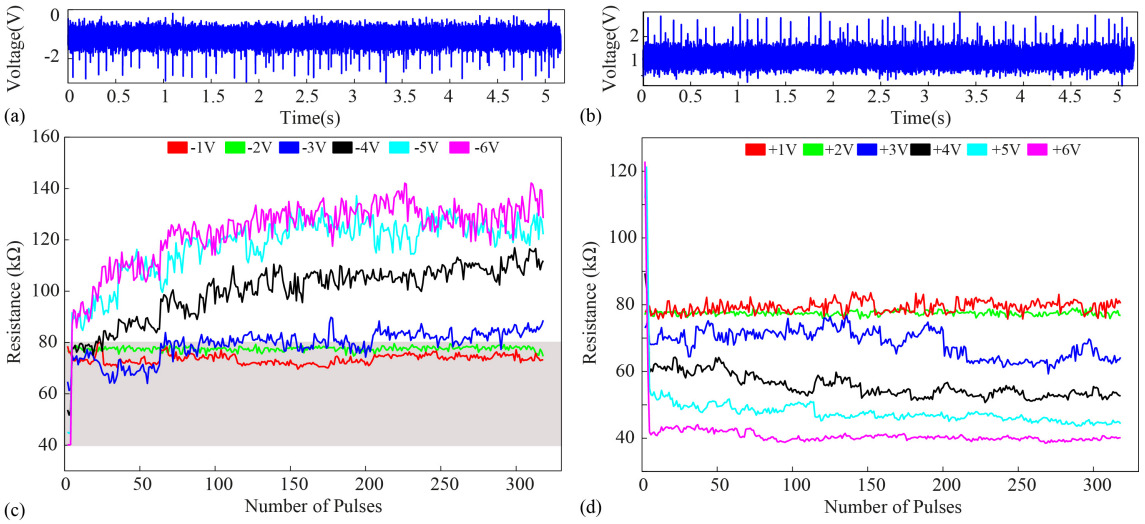


Figure 6.2: A DUT subjected to a reference neural recording amplified with voltages from -1V to -6V in steps of 1V. (a, b) The reference neural recording used for this experiment with significant events in the negative and positive polarity respectively. (c,d) The resistive state output of the device as captured by the MIS platform after every round of experiment. Notably, in these devices, with negative (positive) polarity there is an overall increase (decrease) in the resistive state of the device. Sub threshold region: 1-2 V, Supra-threshold region: 3-6 V.

positive voltages (forming voltage of the devices is $\approx +6.5$ V). Next, a 100 μs pulse-based stimulation protocol is used to switch the resistive state of the DUT in both the polarities [135]. The range of internal threshold (V_{th-}) of TiO_x devices in this case was $\pm$ 0.6-2.2 V.

For this study, a reference neural waveform (voltage-time series) was amplified from -1 V to -6 V in steps of 1 V, as illustrated in Fig. 6.2(a). The resulting resistive state of the target device in response to the biased neural recording as captured by the MIS platform is illustrated in Fig. 6.2(b). The DUT was manually reset to an initial resistive state (R_i) after every round of experiments using a positive polarity pulse lasting for 100 μs (R_i range being 40-80 kΩ). No resistive state changes can be observed for lower voltages i.e. -1 V and -2 V ('sub-threshold' region) whilst distinct resistive state changes can be observed for higher voltages i.e. -3 V - -6 V ('supra-threshold' region); -3 V being the V_{th-} of the DUT. Similar experiment was performed for the events in the opposite polarity, with the V_{th+} of the device being +3 V. The 'sub-threshold' region promotes the filtering of noise whereas the 'supra-threshold' helps in spike-detection (extrapolated from the significant resistive state changes in each bin). These results firmly establishes the fact that output of the MIS devices is sensitive to: a) amplitude, and b) polarity of the input signal. Moreover, since the output of memristive device is function of integral of input voltage, the resistive state changes in principal could be mapped to the spike-strength.

Furthermore, Fig. 6.2 also confirms that the continuous operation of the device for a specific amplitude leads to saturation of the resistive state of the DUT; failing to detect any further significant activity in the input waveform. For instance at -4 V, the resistive state of the device continuously increases from 80 kΩ to 100 kΩ for $\approx$100 pulses, following which the DUT

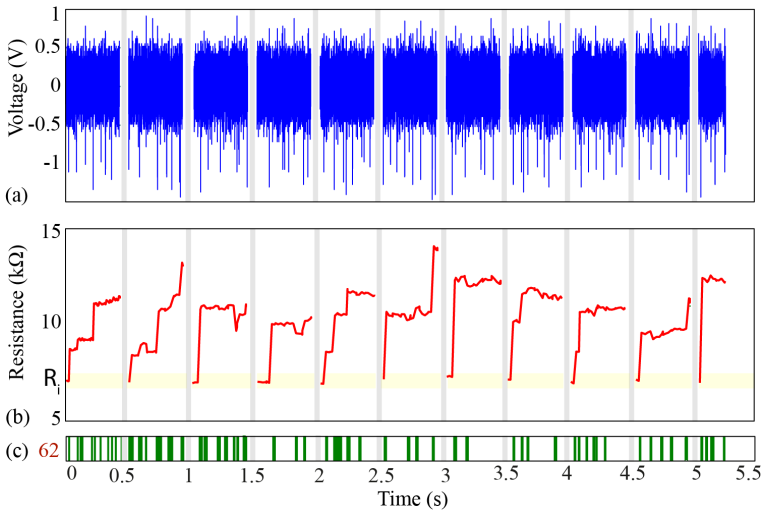


Figure 6.3: Manual frequent resets in the initial resistive state (R_i/R_{on}) of the device upon saturation. (a) Slicing of a neural recording in eleven parts. Each sub-neural recording lasts approximately 0.5 s separated by grey bands. (b) Transient response of the DUT in response to the biased sub neural recording with gain = 0.2 and offset = 0. (c) The number of spikes detected corresponding to each slice. The yellow band indicates the range of the initial resistive state (R_i/R_{on}) which falls in the range of 6-8 kΩ.

saturates. One step further increase in the voltage (-5 V) causes the device to saturate at a higher resistive state of ≈130 kΩ, however, an additional increase of voltage (i.e. -6 V) infers no change in the resistive state of the DUT. This crucial observation suggests two important things: (a) as the device reaches its operational ceiling (high resistive state), the device needs to be 'reset' to its initial state in order to obtain continuous changes in the resistive state and thus avoid saturation (as validated in Chapter 4, see section. 4.6.6), and (b) to map the resistive state changes to distinct neural events with specific amplitude the R_i of the DUT must be in an appropriately smaller range and should not drift considerably. Notably, the large amplitude V_{th-} obtained in this experiment are countered by using optimised TiOx devices for successive experiments.

Saturation of the resistive state of the devices was tackled using 'regular resets', as discussed in section. 4.6.6. The resistive state changes are sensitive to both amplitude and polarity of the input neural signal. Therefore, the information on spike amplitude/duration is preserved in the magnitude of the resistive state modulation to a certain extent, which can be used to locate the address of source event in the biased neural waveform. To examine this hypothesis, for these experiments related to spike sorting in this section, the same neural recording with 63 kS as illustrated in Fig. 6.3(a) was sliced in eleven subsets. Each sub-neural recording lasted for approximately 0.5 s and contained 6 k data points. The gain and offset parameters for this specific experiment was fixed at 2.2 and 0, respectively. The resistive state time-evolution of the target device in response to the biased sub-neural recordings is shown in Fig. 6.3(b), separated by grey bands. The DUT was manually initialised to an R_i of ≈6-8 kΩ after feeding each subset of neural data, represented by yellow bands in Fig. 6.3(b).

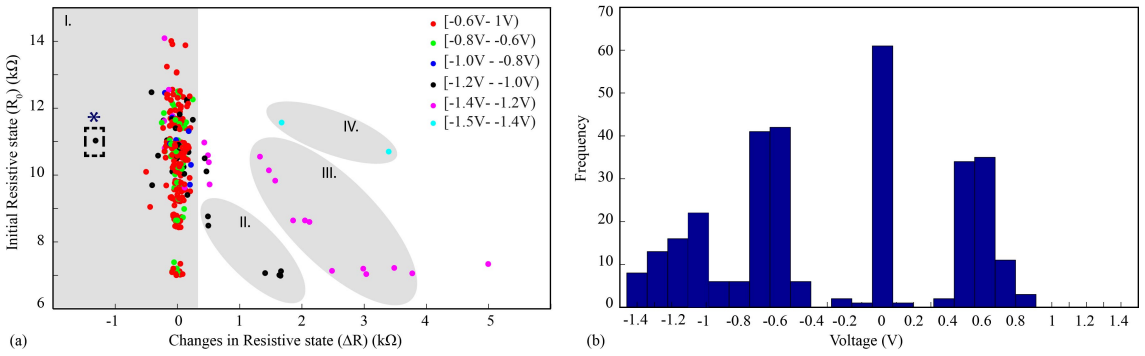


Figure 6.4: Towards spike sorting. (a) Scatter plots for the initial resistive state (R_o) vs changes in resistive state (Δ R) in each bin for every sub-neural recording employed in Fig. 6.3. '*' symbol denotes the outlier event in the group, [-1.2 V - -1.0 V) (b) Frequency histogram for the amplified spiking events used for this experiment.

In section. 4.6.6, we analysed the normalised resistive state changes in each bin and estimated the spike-count for each recording slice as presented in Fig. 6.3(c). Over here the data is analysed in a different facet from the objective of mapping the resistive state changes to the corresponding spike amplitude. Since the initial resistive state (R_i) for all the sub-neural recordings is in relatively smaller range, the resistive state changes (Δ R) are now plotted as a function of initial resistive state (R_0) in each bin, as illustrated in Fig. 6.4(a). These super-imposed scatter plots are analysed in reference to the maximum assigned amplitude of the neural (spiking) event in each bin. Figure. 6.4(b) represents the frequency distribution of the maximum spike amplitude values obtained from all the bins.

To sort the resistive state changes with respect to the spike-amplitude, the entire range of the spike-amplitude is divided into smaller groups. For this recording, the obtained range is (-1.5 V to 1 V) as seen in Fig. 6.4(b) is further sub-divided into six groups, as summarised in Table. 6.1. All the resistive state changes in Fig. 6.4(a) are analysed in reference to the assumed six groups of spike-amplitude. The choice of the first group was done on the basis that throughout the experimental measurements, the V_{th-} of the TiO_x devices is found to be greater than $\pm$ 0.6 V, hence the 'sub-threshold region' was assumed as one group. As for the remaining groups, the range is divided in intervals of 0.2 V, to introduce no experimental bias. In addition, Table. 6.1 indicates the maximum and minimum amplitude of the spiking-events present in each sub-group.

The first, second and third group i.e. [-0.6 V, +1 V), [-0.8 V, -0.6 V) and [-1.0 V, -0.8 V), mostly constitutes of 'noise' and does not affect the resistive state of the DUT. Clearly, most of the resistive state changes are centred around zero and fall under the 'sub-threshold' region of the device. These resistive state changes indicate the first cluster presented in Fig. 6.4(a), marked in grey (I). For the fourth sub-group i.e. [-1.2 V, -1.0 V), a ΔR of up to $\approx$2 kΩ is noted. Importantly, an event present in the beginning of the recording will lead to a stronger overall change in the resistive state of the device in comparison to the same event present in the latter part of the recording. This observation is in line with the saturation behaviour of the resistive state of the MIS devices obtained for different amplitudes, as explained in

Table 6.1: Distribution of the range of the amplitude of spiking events in sub-groups. Maximum and minimum value of spiking events in each sub-group is also indicated.

Bins	Max. Voltage (V)	Min. Voltage
[-0.6V,1V)	0.9084	-0.5840
[-0.8V,-0.6V)	-0.6164	-0.7786
[-1.0V,-0.8V)	-0.8110	-0.9733
[-1.2V,-1.0V)	-1.0057	-1.1679
[-1.4V,-1.2V)	-1.2003	-1.395
[-1.5V,-1.4V)	-1.4274	-1.4599

Fig. 6.3(b) and section. 4.6.6. This explains the smaller change in resistive state with events with higher R_0 (for instance at $\approx$ 8-9 kΩ). Therefore, these resistive state changes can be grouped in one cluster.

From the same fourth group, the resistive state change marked by asterisk symbol ('*'), as shown in Fig. 6.4(a) is an outlier and can be explained. This specific point belongs to the third sub-neural recording in Fig. 6.3(a), containing a strong positive polarity event which leads to a resistive state drop. However, in the operation of MIS platform, the resistive state measurements are assigned to the max. amplitude event in each bin. Therefore, this resistive state change has been assigned to the higher negative polarity event i.e. -1.1 V, present in the same bin, preceding the positive polarity event although the change in the state happens due to the positve polarity event.

Similarly, for the fifth group containing stronger amplitude events i.e. [-1.4 V,-1.2 V), a higher ΔR of upto 4 kΩ is obtained which points to the III cluster. Finally, for the last group i.e. [-1.5 V,-1.4 V), the IV cluster is obtained where ΔR changes are similar to the III cluster. This can again be explained by the fact that the spiking events are present in the latter part of the recording, which can be understood by a higher initial resistive state i.e. R_o (as illustrated), in turn causing relatively smaller changes in resistive state. The three distinct clusters containing significant resistive state changes (clusters II, III and IV) signifies the presence of at least three distinct spikes in the considered groups for spike amplitude.

Importantly, these results was the preliminary approach to decipher the structure present in the resistive state changes in relation to the spike amplitude. They are entirely based on the assumption that by following this methodology, a) spikes with similar amplitude cannot be distinguished, and b) the initial resistive state (R_i) of the employed device should be approximately in a similar range i.e. should not drift drastically. The major significance of these results lies in the fact that previous experiments limited the operation of MIS platform with 'spike-detection' function only, however, the presented approach was the first insight towards 'spike-sorting' capability from the same platform, given the mentioned assumptions.

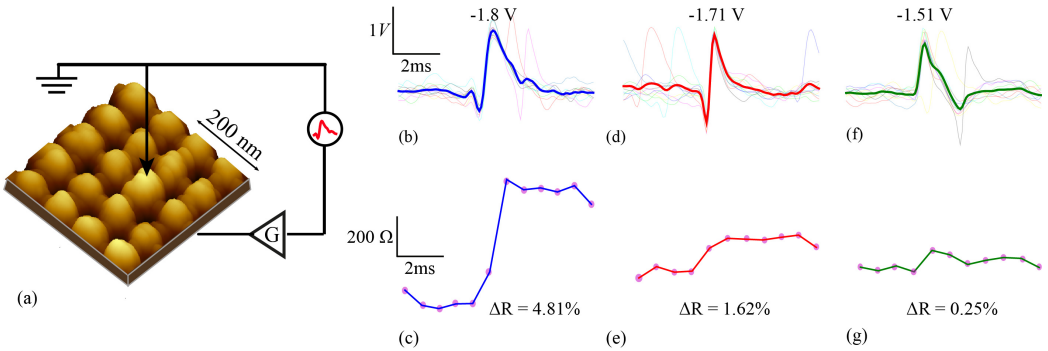


Figure 6.5: Memristive devices and neural signals. (a) Simplified schematic of experimental set-up. The top electrode of each test device was grounded whilst suitably amplified neural signal data samples were applied to the bottom electrode. Inset: Atomic Force Microscopy image memristive devices. (b,d,f) Neural signal data used as input to the memristive spike sorter. Each panel corresponds to neural spikes generated by different neurons and consequently featuring different signature waveforms. Thick traces: average spike waveform for each class (average of 10 instances). Thin traces: 10 different individual instances of spikes (c,e,g). Data includes amplification as shown in (a). Response of memristive devices to inputs from (b,d,f) respectively. Memristor resistive state jumps are observed in tight correlation to input signal voltage peaks. Pink dots: measured resistive state values. Blue: spike type/class I. Red: type II. Green: type III.

6.3 Memristor-based spike sorting platform

In this section, we use publicly available simulated neural signals constructed based on electrophysiological recordings from the cortex and basal ganglia [60] as the input data for the memristive spike sorter (see Appendix. E, Figure. E.1) [142]. The data contained three distinct single-unit activity waveform prototypes overlaid on a noise background as shown in Fig. 6.5 and Appendix. E, Table. E.1 and E.2.

Fig. 6.5(a) depicts a simplified diagram of a single memristor-based spike sorter channel. Analogue input neural data is subjected to suitable amplification using the instrument (see Appendix. B) and is then fed into the memristive device as described in Chapter 4 (section. 4.3.2). The amplification step ensures that action potential contributions to the input waveform exceed the devices thresholds, whilst background noise does not [118].

Fig. 6.5(b-g), illustrates the resistive state of the device assessed periodically and the differences between consecutive readings. The prominent peaks in the three example neural signals cause distinctive changes in the resistive state of the DUT. In order to avoid accumulating resistive state changes to the point of saturation, the devices are regularly reset to a suitable resistive state baseline [129] (section. 4.6.6).

6.3.1 Repeatability

For the first experiment, averages of each of the three single-unit waveforms were obtained by pooling ten random instances from each class. These were then arranged in a spike triplet

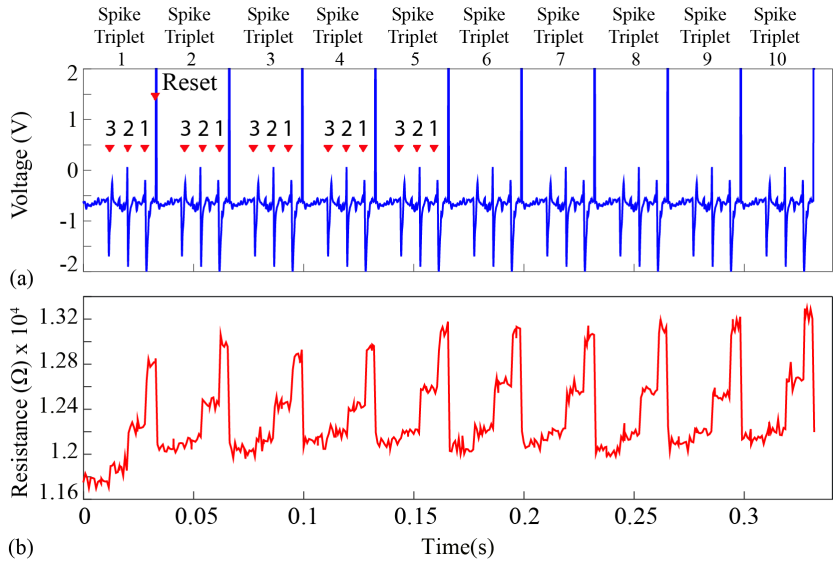


Figure 6.6: (a) Input signal used. Ten copies of a spike triplet were concatenated and fed to the DUT. The triplet was composed of the three average spike prototypes shown in Fig. 6.5, arranged in increasing order of amplitude i.e. order 3, 2, 1. Reset pulses followed every spike triplet. This stimulus was chosen in order to experimentally demonstrate the response repeatability of the DUT. (b) Resistive state evolution of the device in response to the neural recording in (a).

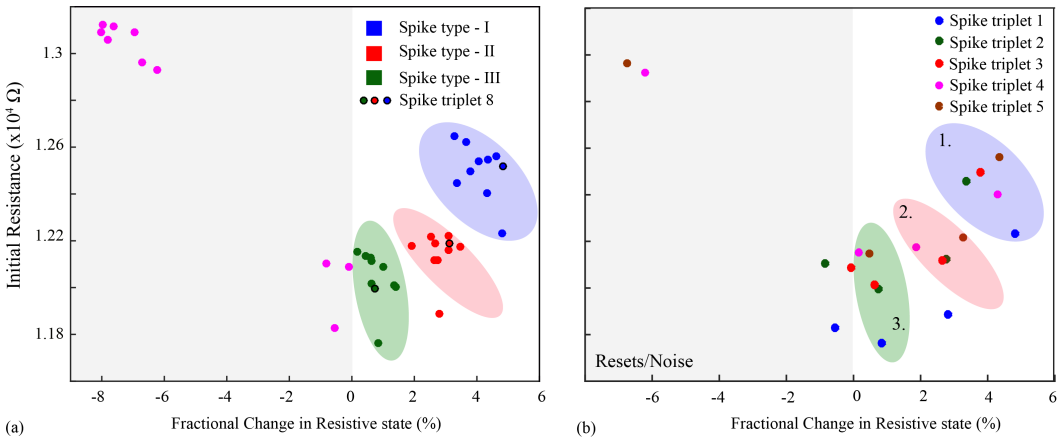


Figure 6.7: Resistive state change between every pair of consecutive measurements plotted as a function of the first resistive state measurement for the same pair. (a) The emergence of three distinct clusters of data-points corresponding to the three input spike classes is observed. Highlighted data-points (black outline): points gathered while applying the 8th triplet as input. (b) Data points are colour-coded for first five spike triplets.

Table 6.2: Raw resistive state measurements taken during the first spike triplet shown in Fig. 6.6. Each resistive state read-out is indexed with a unique identifier ('B1, B2...'). Each batch contains 100 input data points and each bin contains 10 input data points. This means that resistive state measurements were taken at the beginning of each batch of size 100 and then after every bin of size 10. One final measurement was taken at the end of batch whilst the neural feed was paused (noise estimation). Important: (a) For estimating the fractional change in resistive states for each spike prototype in the spike triplet, resistive states highlighted in yellow and green were used as the initial and final resistive states respectively. (b) For estimating fractional changes in resistive states due to background noise/resets, resistive states indicated in yellow and red colours (S.No. 11/12) were used as the initial and final resistive states respectively. The averages of these values were plotted in Figure. 6.7 as a function of the initial resistive state.

S.No.	Resistive State Reads (RSR's - ohms)	Batch 1	RSR's	Batch 2	RSR's	Batch 3	RSR's	Batch 4
	Background/Resets		Spike III		Spike II		Spike I	
1.	11757.99	B1	11782.02	B13	11898.06	B25	12311.09	B37
2.	11828.52	B2	11763.65	B14	11886.4	B26	12232.41	B38
3.	11790.73	B3	11757.92	B15	11809.14	B27	12235.02	B39
4.	11699.34	B4	11760.06	B16	11983.17	B28	12241.67	B40
5.	11790.92	B5	11706.08	B17	11855.17	B29	12267.59	B41
6.	11790.03	B6	11894.95	B18	12191.47	B30	12653.33	B42
7.	11783.03	B7	11899.9	B19	12244.21	B31	12843.73	B43
8.	11697.71	B8	11924.52	B20	12283.07	B32	12776.11	B44
9.	11755.09	B9	11834.5	B21	12238.37	B33	12796.08	B45
10.	11766.31	B10	11841.8	B22	12179.92	B34	12816.69	B46
11.	11739.83	B11	11889.99	B23	12279.27	B35	12813.46	B47
12.	11784.82	B12	11855.77	B24	12202.14	B36	12834.55	B48
Fractional Changes in Resistive state - igoring data point at S.No. = 1								
S.No. (10-2)	-0.526		0.664		2.469		4.776	
S.No. (11-2)	-0.749		1.073		3.305		4.750	
S.No. (12-2)	-0.369		0.783		2.656		4.922	
Averages	-0.548		0.840		2.810		4.816	
Fractional Changes in Resistive state - including data point at S.No. = 1								
S.No. (10-1)	0.071		0.507		2.369		4.107	
S.No. (11-1)	-0.154		0.916		3.204		4.080	
S.No. (12-1)	0.228		0.625		2.556		4.251	
Averages	0.0481		0.683		2.709		4.146	

as shown in Fig. 6.6, at the end of which a reset pulse was appended. Ten spike triplets were sequentially fed into the test memristor and its resistive state was regularly assessed 3 times during each triplet, producing results similar to Fig. 6.5(b-g) (amplification gain: -1.3 offset: -0.63 V). Plotting the change in resistive state between every pair of consecutive measurements vs. the resistive state in the first measurement of the same pair results in Fig. 6.7. These results capture the impact of each spikes strength on setting the memory state of the device with respect to its initial state and their corresponding R(ΔR) points in Fig. 6.7. An example of the raw resistive state measurement taken during the first spike triplet and the methodology for processing the fractional change in the resistive states from raw resistive states is presented in Table. 6.2.

Figure. 6.7(a,b) represents the clear clustering of data-points associated with the three distinct spike waveform prototypes. Importantly, in this figure the same data is represented in two different ways to visualise a) three distinct clusters in response to three different single-unit spike waveforms and (Fig. 6.7(a, b)) colour coded data points for first five spike triplets (Fig. 6.7(b)). This demonstrates that the memristive device is capable of intrinsically

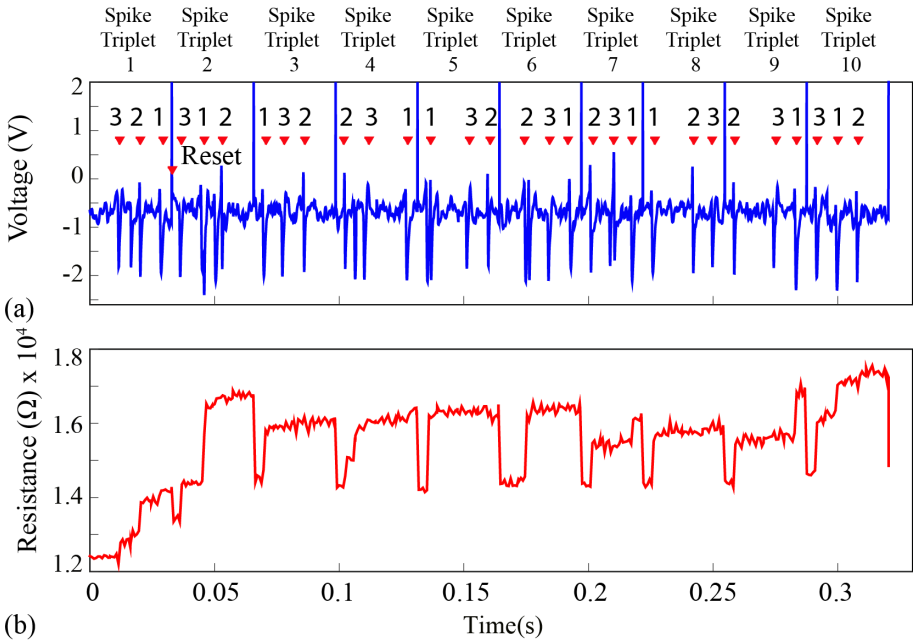


Figure 6.8: (a) Input neural waveform used. Randomly chosen spike waveform instances, one belonging to each prototype, were arranged in random order as shown in the figure in order to form a spike triplet. Reset pulses followed every spike triplet. Ten such randomly generated triplets were concatenated and fed to the device-under-test. (b) Resistive state evolution of the device in response to the neural recording in (a).

performing spike sorting consistently. As the waveform input to the device is identical for each triplet, any variation in R(ΔR) response arises mainly from device variability.

6.3.2 Randomised instances

For the second experiment, the triplets were constructed from individual spike instances (not averaged prototypes) and their order was randomised as shown in Fig. 6.8. This setup accounts for background noise-induced spike shape variability [60]. Results are shown in Fig. 6.9, where the intrinsic variability in device behaviour is compounded by the variability in the spike waveforms. Notably, despite the fact that clustering is no longer as clear as in Fig. 6.7, it is still possible to linearly separate the majority of events. The misclassified spike that caused strong change in resistive state (enclosed in a box) was the result of the event instance containing two spikes in close succession (main spike plus a stray spike). Moreover, measured resistive states and fractional changes in resistive state experienced by the memristor as a result of applying the ten spike triplets from Figure. 6.8 are presented in Table. 6.3.

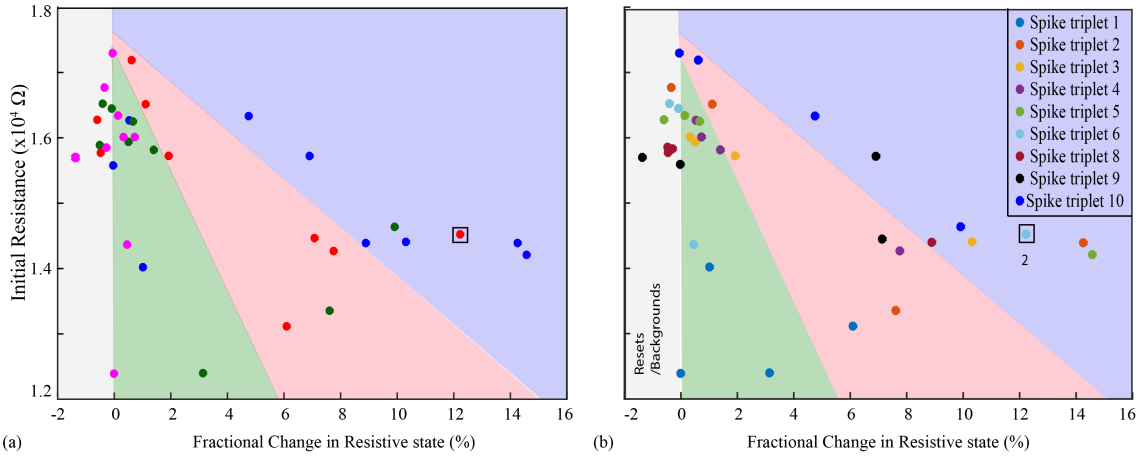


Figure 6.9: Resistive state change between every pair of consecutive measurements plotted as a function of the first resistive state measurement for the same pair. (a) The emergence of three distinct clusters of data-points corresponding to the three input spike classes is observed. (b) Data points are colour-coded for ten spike triplets.

Table 6.3: Measured resistive states and fractional changes in resistive state experienced by the memristor as a result of applying the ten spike triplets from Figure. 6.8. In each row, the effects of the different spike components of each triplet are shown separately.

	RSR (% change)	Initial RS ()	RSR (% change)	Initial RS ()	RSR (% change)	Initial RS ()	RSR (% change)	Initial RS ()
Spike Triplet (Order fed)	Background,Resets							
Spike Triplet 1 (3,2,1)	-0.009	12392.02	3.13	12400.28	6.08	13111.9	1.01	14022.04
Spike Triplet 2 (3,1,2)	-0.348	16769.94	7.61	13351.12	1.10	16514.83	14.25	14387.25
Spike Triplet 3 (1,3,2)	0.331	16012.95	0.493	15939.68	1.92	15721.62	10.304	14407.54
Spike Triplet 4 (2,3,1)	0.524	16265.24	1.40	15812.35	7.75	14267.21	0.72	16012.79
Spike Triplet 5 (1,3,2)	-0.607	16277.72	0.134	16337.895	0.663	16246.95	14.57	0.455
Spike Triplet 6 (2,3,1)	-0.091	16451.23	-0.413	16518.9	0.456	14365.22	12.221	14525.54
Spike Triplet 8 (1,2,3)	-0.472	15875.08	-0.451	15775.74	-0.298	15827.51	8.881	14395.29
Spike Triplet 9 (2,3,1)	-0.025	15591.67	-1.375	15698.12	7.125	14452.56	6.900	15719.95
Spike Triplet 10 (3,1,2)	-0.053	17298.12	9.90	14640.5	0.61	17190.34	4.75	16335.51

6.4 Discussion and Summary

The chapter shows the first approach for mapping the memory state changes of 'memristive integrating sensor' to the spiking events in the original neural waveform. The resetting operation saves the memristive device from undergoing saturation and permits continuous detection of spikes in the biased neural waveform. The obtained resistive state transitions are then studied in reference to the spike amplitude aggregated in different groups. Clusters of resistive state changes can be differentiated in response to the assumed groups of spike amplitudes, thus confirming the potential of MIS platform for 'spike-sorting.' This work used simple features such as 'amplitude' and 'polarity' for identification of individual spikes. In future, one could also explore features such as 'area under spike i.e integral transform' and compare this to the method adopted in this work.

It is also important to stress that this work has been performed with the objective of demonstrating how single devices can perform both spike detection and spike sorting at no additional power cost. The development of this work into full circuit implementations with how much support/auxiliary circuits/hardware would be needed remains as the future work. At this stage, this work confirms the capability of single memristive devices in non-volatile regime to repeatedly identify the spiking events and grouping them in clusters. Again how this performs against other standard methods of spike-sorting is yet to be determined. Additionally, the performance of memristive devices in volatile regime for spike sorting has not been fully determined and would require further experimental studies.

Chapter 7

Conclusions and Future Work

7.1 Conclusions/Summary

This thesis is the first scientific evidence demonstrating the employment of memrisitve devices for the purpose of spike detection and spike-sorting paving the way towards advanced neuroprostheses or applications such as bioelectronics medicines, where the power dissipation remains as the major challenge. As a first piece of the puzzle, this thesis accomplishes the task of process development of RRAM devices. It studied the behaviour and electrical characteristics of TiO_x devices in-depth for both single devices and crossbars deciphering the effects of parameters such as pulse-width and amplitude. Automated electrical characterisation protocols covering both volatile and non-volatile behaviour of the devices were developed to carry out en-masse testing of the fabricated devices. The development of such protocols helped in determination of the operating range of the functional devices consistently saving huge amount of manual effort and time. More importantly it ensured reliable & repeated testing of devices leading to accelerated process development cycles[1].

Development of the characterisation protocol connected the dots to engineer the fabricated devices for the application of spike-detection which remains as a major challenge in the fast growing field of neural interfaces. Therefore in the second stage, the intrinsic synapse-like attributes of single TiO_x devices were exploited and the proof-of-concept for spike detection was demonstrated. This was initially studied using non-volatile properties of the devices where the supra-threshold events were encoded in the non-volatile resistive transitions of the devices, whereas the noise was naturally suppressed. The results were benchmarked against state-of-the-art template matching system and the performance of the system was optimised to obtain sensitivity comparable to the benchmarking standard. This was done by experimenting with the operational parameters i.e. gain, offset and noise band settings. As a result, the experiments demonstrated a substantial improvement in the bandwidth with

[1] From the electrical characterisation perspective. This thesis did not evaluate the switching mechanisms of the memristive devices

minimal power budget of ≈300nW per channel. The subsequent in-depth analysis revealed saturation of the devices in non-volatile mode to be strongly performance limiting. This was due to saturation of the memory state of the devices. As a counter measure frequent resets in the memory states were employed to improve the performance of the devices, however, this impacted negatively on the overall power consumption.

The downside of non-volatility pushed the experimental work towards exploring different property of memristive devices i.e. 'volatility'. The same application of neural spike-detection was then explored using meta-stable resistive transitions of nanoscale metal-oxide memristive devices. The self-resetting property of the devices addressed the issue of saturation and in this scenario, power dissipation was found to be ≈100 nW with the footprint of the used devices to be 200x200 nm^2. It is of crucial importance to note that the estimated power dissipation is only for spike-detection per batch. The power estimations for full circuit implementations would include additional calculations for instance power dissipated by read and write circuit, which was out of scope of this thesis.

This work also took initial steps to explore and address another major challenge i.e. 'spike sorting'. Conceptually, the resistive state changes of the memristive devices were mapped to the spiking events in the original neural waveform. The approach for 'spike-sorting' was demonstrated using non-volatile memristive devices at no extra power cost in addition to spike detection.

To conclude, the author of this thesis experimentally found the memristive devices in volatile regime to be better suited for application of 'spike detection' over memristive devices in 'non-volatile' region. This is due to their self-resetting capabilities in addition to reduced power consumption. However, for 'spike-sorting' the initial results of memristive devices in volatile regime are negative and only the ones in non-volatile regime succeeded (as discussed in Chapter 6). This is an open ended question at this stage and would require additional experiments as detailed in the next section. Another important point in this thesis to remember is that all the work has been demonstrated at 'single devices' level and the power consumption has also been estimated on similar basis also limited by a big issue of 'device-to-device' variability. How these devices will transform to 'crossbar level' and how they would perform when integrated at hardware level with additional circuitry is yet to determined and would require additional simulations and experiments.

On the basis of these conclusions and summary, the author recommends this work to diversify in the following future work.

7.2 Author's viewpoint: Recommendations for future work

The research work in this thesis bring new application prospects for memristive devices, diversifying from conventional digital memory applications towards enabling active neural

interfacing technologies that are very much needed for realising the electroceuticals vision [138]. In the view of author, one could build upon the presented work in the following ways.

1. **Process development of memrisitve devices:** In order to use RRAM devices reliably for the application of spike detection and sorting, it is of paramount importance to research on alternative or engineered materials as devices' active core. It is essential to innovate the device stack structures which could potentially allow for achieving smaller range of variation in threshold voltages, tuning the volatile characteristics of devices and in turn the self-reset achieved by the spike-detector reducing the overall power consumption. On the other hand, in the case of non-volatile materials, finding substitute materials which would require less frequent resets, it would be beneficial to develop an understanding about switching mechanisms which remains debatable. Subsequently, the natural next questions to answer would be if the same reliable process for signal processing could be applied to crossbar arrays and if that can be optimised down to nano-scale. A far-fetched thought would be if these devices could potentially be integrated with micro-electrode arrays on flexible substrates on-site to permit more localised neural signal on-chip processing. These devices do open the doors to some intriguing scientific questions but how far they will perform remains to be seen.

2. **Further development of automated protocols:** The art of reliably detecting functioning RRAM devices in silicon is constantly evolving. Further work for developing the automated protocols would include: a) an investigation of device behaviour in mid-range p-value bins, currently suspected to be a mixture of devices that require more invasive pulsing for successful switching and devices that experience mild levels of cross-talk, b) development of a more powerful version of algorithm capable of quantifying cross-talk, c) assessment of thermal impact of biasing protocol, d) applying the proposed test strategy to different device stacks fabricated for RRAM technology (for both volatile and non-volatile), and e) perhaps retention property could be analysed in better depth.

3. **Full circuit implementations:** The thesis estimates the power dissipation at a single-device level. The precise details of integration of the neural circuitry and full circuit implementations remains as the future work. Clearly, the memristor read-out and biasing circuitry will require an additional power and therefore the presented considerations are not a reflection of technology's full potential. Substantial improvements in the power consumptions has already been demonstrated by using volatile devices by operating them in higher resistive state regions but how power efficient will they be in a full circuit implementation, still needs to be analysed and performing simulations would potentially be the first step in this direction.

4. **Improving sensitivity/detection accuracy rates:** Variable sensitivity, detection and accuracy rates can be accepted depending on the application under study. Therefore, it is imminent to study these devices from a specific application perspective and devise strategies to improve the accuracy rates at nano-scale.

5. **Spike detection and sorting with nano-scale, non-volatile devices:** In this work, large prototype non-volatile devices were used for demonstrating the concept of spike-detection and sorting. Optimising this concept towards nano-scale, non-volatile devices with less intrinsic threshold and device-to-device variability could be one possible next step. Again statistical analysis would be fundamental to prove this point. Moreover, many significant questions such as how will the devices be integrated at system level or how devices will perform in cases were action potentials with different shapes but same amplitude exist remains as the future work. To a certain extent, the nature of spike sorting also depends upon number of questions such as the application/disease under study, the nature of the recording i.e. intracellular or extracellular, assessing if the recording contains multi-unit activity or single-unit activity or the compound action potentials. Therefore, to assess the performance of devices in all such scenarios more closely, this subject needs to be studied in-depth in future.

6. **Spike sorting with nano-scale volatile devices:** An immediate next question that can be answered post this experimental work would be if the volatile devices could perform spike- sorting. From the presented results in case of volatile devices, there is no clear evidence if these devices could perform spike sorting demonstrating clear clustering.

Appendix A

Fabrication of Solid-state Metal-Oxide Devices

Dr. Ali Khiat provided the necessary support with the fabrication of devices.
Dr. Khiat is a senior experimental Officer at Southampton Nanofabrication Centre, University of Southampton. His current main research interests are micro-/nano-fabrication, optimisation, metrology and characterization of memristive devices.

In this experimental work, two sets of devices i.e. nano- and micro- structure devices were experimentally used. The fabrication procedure for the two sets is described as follows:

All the micrometer sized devices exploited in this experimental work were fabricated according to the following procedure. 200 nm of insulating SiO_2 was thermally grown on 6-inch silicon wafer. The three main patterning steps were processed, each contains optical lithography, film deposition and lift-off process. In the first step, 5 nm Titanium (Ti) and 10 nm Platinum (Pt) films were deposited via electron-beam evaporation technology to serve as bottom electrodes, Ti was used for adhesion purposes. In the second step, magnetron reactive sputtering system was used to deposit the near-stoichiometric TiO_{2-x} (x=0.06) active core from Ti metal target. Two plasma sources were used to ensure near stoichiometric film. 25 nm thick TiO_{2-x} was deposited. In the final step, 10nm Pt top electrodes were deposited using electron-beam evaporation system. At the end of processing, the wafer was diced into 9x9 mm^2 chips, which were then wire-bonded in standard packages for measurements. The final stack is constituted of Ti/Pt/TiO_{2-x}/Pt (5/10/25/10 nm) (Fig. A.1).

All nano-devices (Fig. A.2), that is, Ti/Pt/TiO_{2-x}/TiN (5/10/10/40 nm) were fabricated as follows: 6-inch wafer was thermally oxidised to grow 200 nm SiO_2, which serves as an insulating layer. Then, direct write e-beam lithography method was adopted, using JEOL JBX 9300FS tool, to pattern the bottom electrodes (BEs) nanowires. BEs constituting 5 nm adhesive Ti layer and 10 nm Pt film were deposited using e-beam evaporation. Bottom

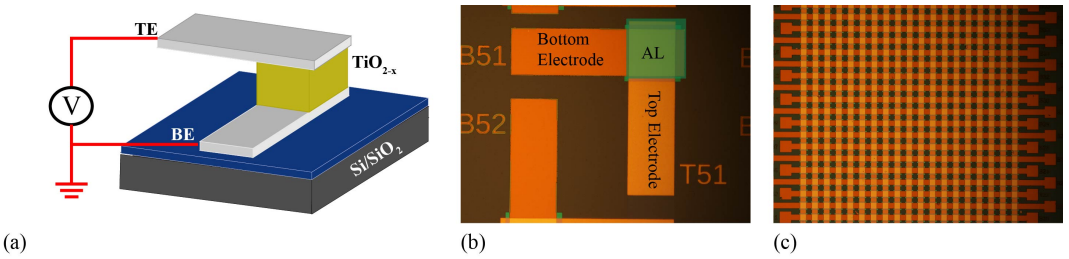


Figure A.1: (a) Vertical stack structure for RRAM devices. The fabricated devices have two types of RRAM architectures i.e. (a) Single devices with L-shape structure and (b) 32x32 Crossbar structures. 'T' and 'B' denotes the top and bottom electrode respectively. AL in green colour denotes the active layer.

access-electrodes (large features) were then defined via conventional photolithography patterning, e-beam evaporation of Ti/Au (5 nm/25 nm) and lift-off process. Access-electrodes connect the pads to the nanowires. To pattern the active layer, optical lithography, reactive sputtering and lift-off process were also used. 10 nm near-stoichiometric TiO_{2-x} active layer was sputtered with Leybold Helios Pro XL Sputterer from a Ti metal target. Next, 40 nm thick TiN top electrode (TEs) nanowires and 25 nm thick Au top access-electrodes were obtained in a similar manner to BEs and to the bottom access-electrodes, respectively, with the TiN films deposited via a Leybold Helios Pro XL Sputterer.

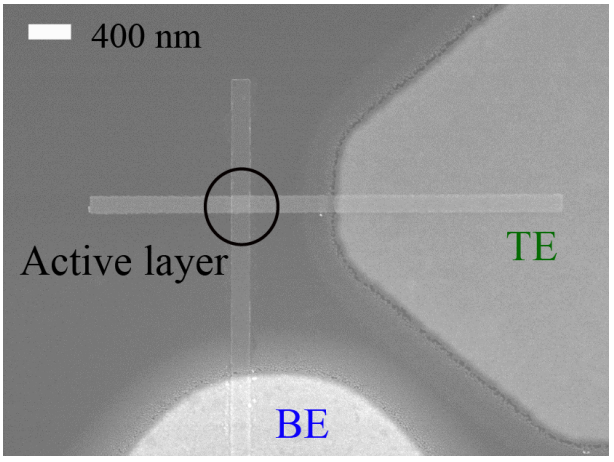


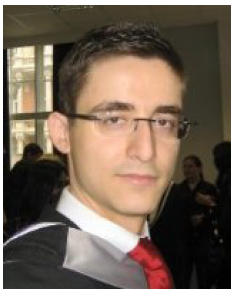
Figure A.2: Device architecture of nanoscale TiOx memristive devices. Scanning Electron Microscopic (SEM) image of the employed 200nm x 200nm memrisitve device.

Appendix B

Characterisation Tool for RRAM devices: ArC Instruments

Dr. Radu Berdan and Dr. Alexantrou Serb provided the necessary support with the instrumentation.

Dr. Radu Berdan recently completed his PhD in 'Applications of Memristors in Conventional Analouge Electronics' from Imperial College London, UK. Presently, he is the CEO of ArC Instruments, Product Lead and oversees the development of the full range of array control instruments.



Dr. Serb is a research fellow at the Electronics and Computer Science (ECS) dept., University of Southampton, UK. His research interests are: instrumentation, algorithms and applications for RRAM testing, and neuro-inspired engineering.

The group at Southampton has developed a microcontroller-based PCB mounted infrastructure to implement pulsing protocol on the devices. The instrument is capable of addressing devices embedded in crossbar arrays of up to 1kb in size (32x32, (Fig. B.1c)). The system has the capability of either testing packaged arrays (Fig. B.1a) or communicating to a multi-channel probe card for direct testing on wafer (Fig. B.1b) [107].

The hardware is supported by custom-made software as shown in Fig. B.2 that permits exhaustive, device-by-device testing of the entire crossbar array or an array of individual devices in one, fully automated round of measurements. The high end user friendly Graphical User Interface (GUI) enables programming of the device-under-test. It offers the flexibility of

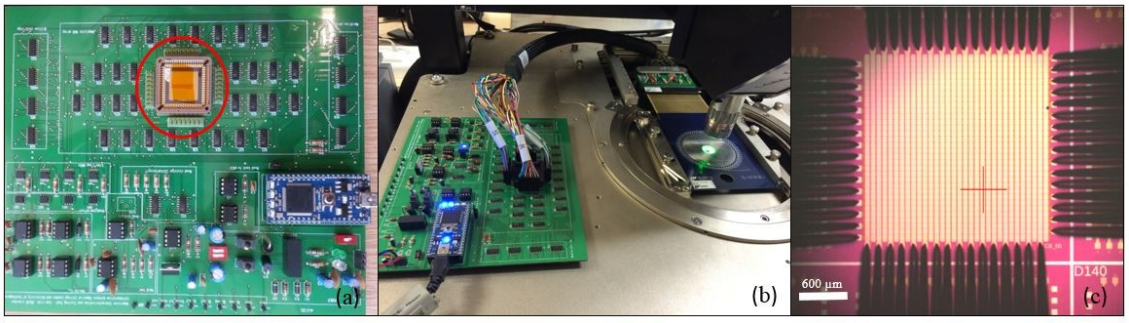


Figure B.1: Hardware Infrastructure. (a) PCB-mounted system overview with single devices, (b) PCB-mounted system overview with packaged devices, (c) System connected to multi-channel probe card, (d) Microscopic Image showing probe card needles touching down on a 32x32 crossbar array.

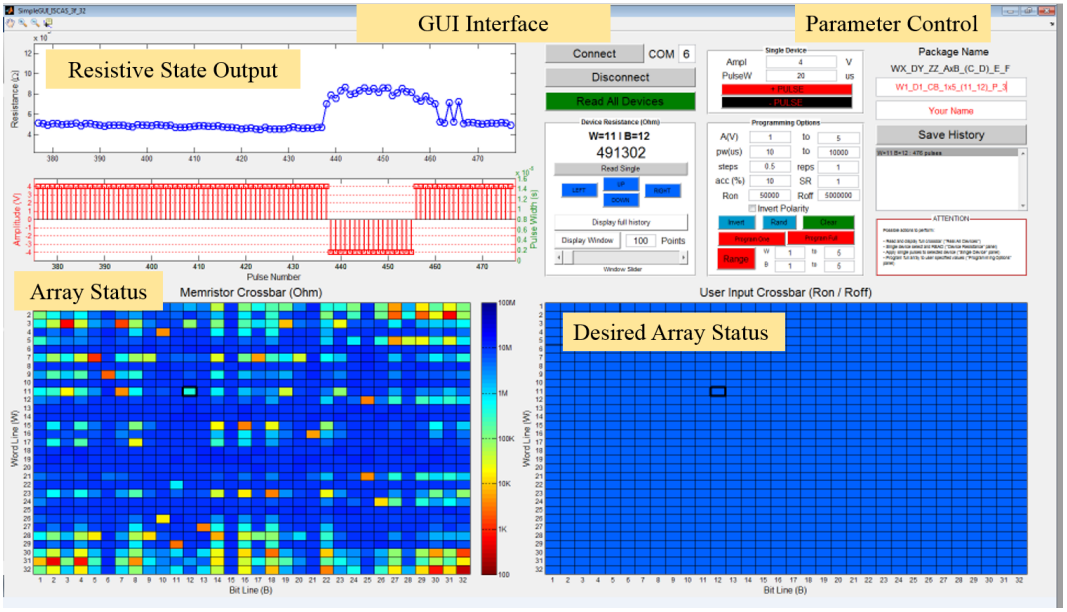


Figure B.2: GUI Interface for programming the devices under test.

multiple programming parameters for instance pulse-width, pulse-amplitude, steps, accuracy, low resistive state or high resistive state.

Appendix C

Neural Data Acquisition System

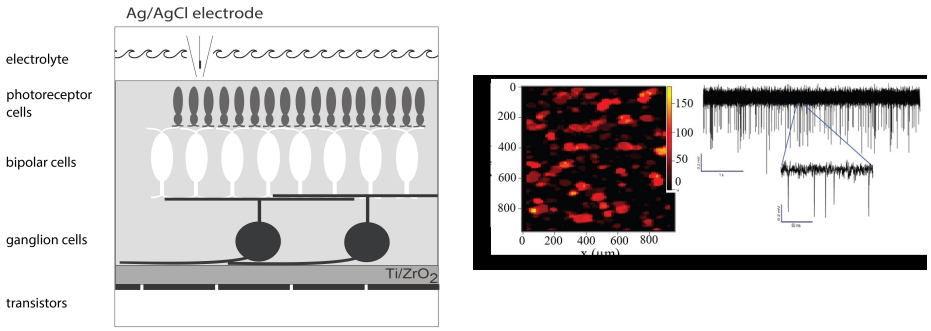
Front-end neural recording system (CMOS Multi Electrode Array)

Prof. Stefano Vassanelli and Dr. Ralf Zeitler provided the neural recordings used during this research work.

Prof. Stefano Vassanelli received a medical degree from the University of Padova (Italy) in 1992 and the PhD in Molecular and Cellular biology and Pathology from the same university in 1999. Currently, he is Professor of Physiology and teaching both for the Engineering and Medical faculities. He has been working on the development of neurotechnologies for recording, stimulation and processing of signals generated by neuronal networks.

Dr. Ralf Zeitler studied Physics at the Technical University of Munich (Germany), graduating with Diploma in Physics in 2004. In 2009 he received his doctor's degree in physics from the university for the research of the noise processes in the cell-semiconductor interface at Max Planck Institute for Biochemistry. From 2012-2015 he was research group leader at Max Planck Institute for Intelligent Systems working on a CMOS based sensor platform for living cells, which led to spin-off Venneos GmbH in 2014, where he currently the technical managing director.

For the experimental validation of MIS system, neural activity was recorded from slices of dissected mid-peripheral rabbit retinal ganglion cells using CMOS technology (Fig.C.1) [123, 124]. The CMOS based multi-transistor array (MEA) consists of 128x128 sensor sites, which records the data at a sampling rate of 12.2 kHz and outputs a current time series containing approximately 63k samples. The sensor sites of the CMOS-MEA are insulated by an inert TiO_2/ZrO_2 layer and a thin metal layer beneath the oxide layer is connected to the gate of the field-effect transistor. The voltage changes due to the interfaced neural tissue/cells above the recording sites are used for modulating the source-drain current in



(a) Illustration of the retina/chip configuration depicting the multi-electrode arrays upon which tissue slices from retinal ganglion cells are placed and measured directly.(b) (Left) Colour coded activity map of electrical activity from tissue slices of rabbit ganglion cells places atop the MEA. (Right) Traces of pre-amplified data (voltage-time series) obtained from the front end set-up in the range of 100's of mV.

the MOSFET. Trans-Impedance Amplifiers (TIA) fabricated on-chip convert the signal into voltage and amplify the signal from a 0.1 mV- 1 mV range up to 10-500 mV in range. This amplified signal is then used as an input for our platform. It should be clearly noted that the CMOS MEA was kept external to the memristor-based spike detection platform and in this entire work the CMOS MEA is termed as the 'front-end' system.

Appendix D

Power Estimations

D.1 Estimation of power dissipated per channel using manual frequent resetting method (Non-volatile region)

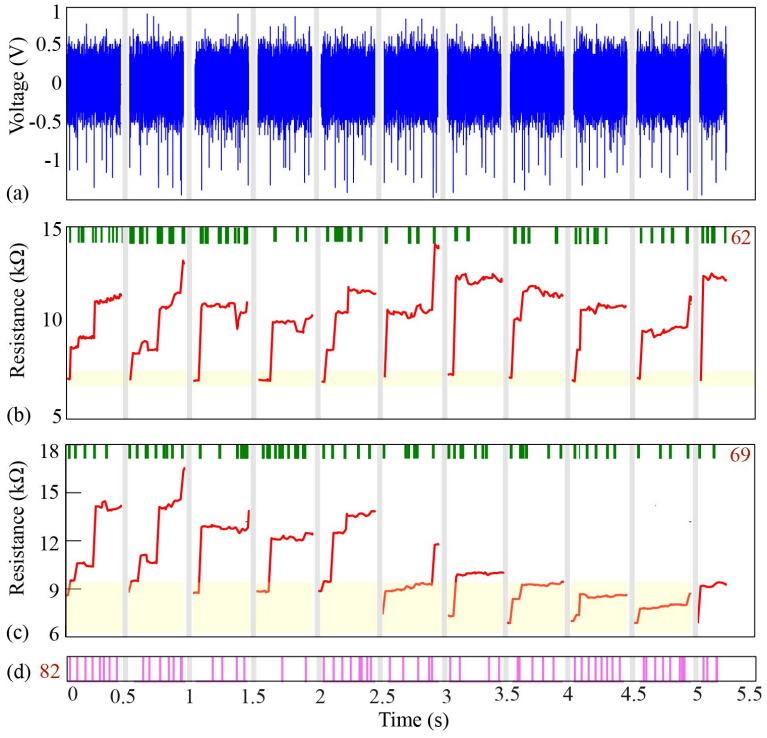


Figure D.1: Frequent resetting of the memristive device in order to improve the spike count of the system. (a) Slicing of the neural recording used in Fig. 4.22 in eleven parts each lasting for 0.5 s separated by grey bands. The device is manually reset to its initial state after each run (yellow bands in (b) and (c)). R_{on} is approx. equal to 6-8 kΩ. (b) and (c) Output of the device when biased with neural recording in (a) with G = 2.2, V_{off} = 0 and G = 2, V_{off} = -0.2. Green bands indicates the bins in which neural activity is detected by the MIS system when the noise band is calculated using $\mu \pm 2\sigma$ method. (d) Spikes indicated by the TMS, which is equal to 82.

Assumptions:

1. The following calculations have been done assuming the resistive state of the device to be 10 kΩ and series resistance (compliance) to be 1 kΩ.
2. The read voltage of the device is +0.5V and the write voltage is conservatively assumed to be +5V.
3. The pulse width used for the experiment is 100 μs.
4. For every batch of thousand data points, the state of the device is read five times.
5. P: Power, R: Resistance, V: Voltage, E: Energy, t: time (employed pulse width).
6. Sampling frequency is 12.2kHz.

Energy consumed for read operation:

P = V^2/R= $(0.5)^2$/(1+10) kΩ = 0.023 mW

E = P x t = 0.023 mW x 100 μs = 2.3 nJ

E dissipated for 5 read operations = 2.3 nJ x 5 = 11.5 nJ

Energy consumed for write operation:

P = V^2/R = $(5)^2$/(10) kΩ = 2.5 mW

E = P x t = 2.5 mW x 100 μs = 250 nJ

For 1000 samples, 1000x 250 nJ = 250 μJ

Resetting operation using one single pulse:

E = 2.5 mW x100 μs = 250 nJ

Average Power Consumption:

$\approx 3mW$

D.2 Estimation of power dissipated per channel in volatile MIS platform

Assumptions:

1. The following calculations have been done assuming the resistive state of the device to be 1 MΩ and series resistance (compliance) to be 100 kΩ.
2. The read voltage of the device is +0.2V and the write voltage is conservatively assumed to be +3V.

3. The pulse width used for the experiment is 1 μs.
4. For every batch of thousand data points, the state of the device is read five times.
5. P: Power, R: Resistance, V: Voltage, E: Energy, t: time (employed pulse width).
6. Sampling frequency is 12.2kHz.

Energy consumed for read operation:

P = V^2/R = $(0.2)^2/(1\,\text{M}\Omega + 100\,\text{k}\Omega) = 0.04\,\mu\text{W}$

E = P x t = 0.04 μW x 1 μs = 0.04 pJ

E dissipated for 5 read operations = 0.04 pJ x 5 = 0.2 pJ

Energy consumed for write operation:

P = V^2/R = $(3)^2/(1)$ MΩ = 9 μW

E = P x t = 9 μW x 1 μs = 9 pJ

For 1000 samples (per batch), 1000 x 9 pJ = 9 nJ

Average Power Consumption:

$\approx$100 nW.

Importantly, if the same experiment is performed with 100 ns pulses, the power consumption would be approximately 10 nW per batch.

Appendix E

Neural Data for Spike sorting

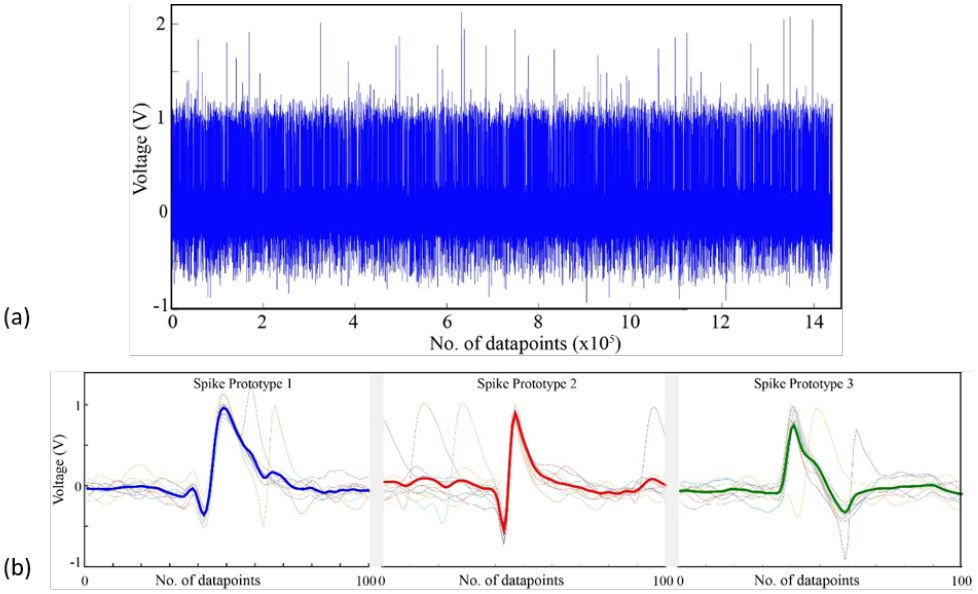


Figure E.1: (a) Raw neural recording data used for the experiments. The data is publicly available from University of Leicester (http://www2.le.ac.uk/centres/csn/software, **Dataset 2 - Simulated Extra-cellular recordings**. The neural recording contains three distinct single-unit spike waveforms superposed on background noise. Detailed description on how this neural recording data was synthesised is presented in [60, 143]. (b) For the experiments, ten different instances of each spike prototype were randomly extracted, as illustrated in the three insets. The spike timings of the instances are documented in Table. E.1. Each instance of every spike waveform contains 100 data points i.e. 19 points before and 80 points after the spike registration timestamp. Thick blue, green and red traces show the averages of ten different instances of each spike waveform. The maximum and minimum voltage values in each average spike prototype are further presented in Table. E.2.

Details on Generation of multi-unit activity:

Detailed description on how this neural recording data was synthesised is presented in [60, 143]. Research paper [143] shows how realistic simulations of extracellular recordings can be obtained. Research paper [60], page 1670 summarises the 'Data Simulation' procedure. To reiterate, the simulations were created using a database with 594 different averaged spike

shapes, compiled from recordings in the monkey neocortex and basal ganglia. Multi-unit activity i.e. spikes that can be detected but cannot be clustered into different single units due to their small amplitude is largely produced by neurons located around 100-150 μm from electrode tip. In this particular scenario, multi-unit activity was created by mixing the activity of database of 594 spike shapes using amplitudes uniformly distributed between 0.5 and 1.5 times the level of detection threshold. The detection threshold was set at 4 times the estimation of the standard deviation of the noise. Simulations of 2 min each with a sampling rate of 24 kHz were generated. The noise activity was obtained by superimposing a large number of spikes, simulating the contribution of far away neurons from the electrode tip (superimposed at random times and amplitudes).

Table E.1: Spike timings for ten randomly chosen instances of three distinct spike prototypes (as shown in Fig. E.1). Units: data point indices.

S.No.	Spike Waveform 1 Spike timings	Spike Waveform 2 Spike timings	Spike Waveform 3 Spike timings
1.	2785	20173	45450
2.	54871	41370	465326
3.	127642	107486	606816
4.	207821	194634	854280
5.	395160	300366	939685
6.	708914	395161	1091743
7.	903698	631599	1208267
8.	1011435	886880	1314658
9.	1273906	1082916	1436188
10.	1420564	1330703	204139

Table E.2: Maximum and minimum voltage values in the three average spike prototype waveforms as illustrated in Fig. E.1.

S.No.	Average Spike Prototypes	Max. (V)	Min. (V)
1.	Spike Prototype 1	1.042	-0.33
2.	Spike Prototype 2	0.968	-0.52
3.	Spike Prototype 3	0.850	-0.30

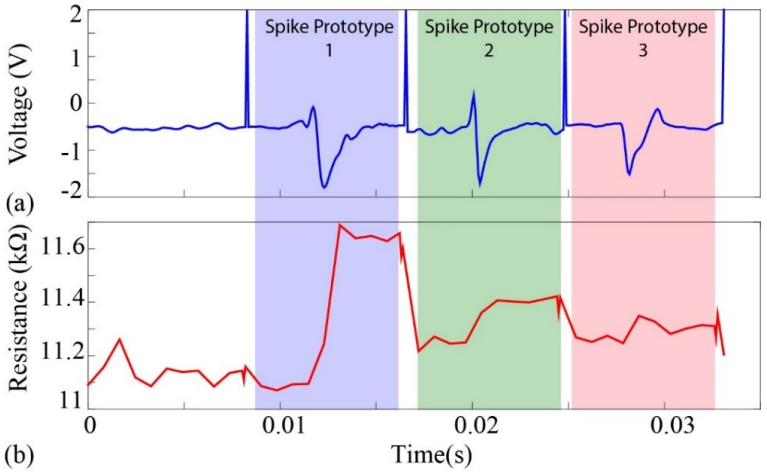


Figure E.2: (a) Pre-processed recording used for Figure. 6.5(b-g) in the main manuscript. Reset pulses of +2V were used to maintain the device functional within the operational resistive state region. (b) Resistive state changes for the device in response to the recording in (a) over time. Blue, green and red shadings correspond to the three distinct average prototypes 1, 2 and 3 respectively.

Appendix F

Contributions

F.1 Research publications specific to the proposed thesis

1. Gupta, I., Serb, A., Berdan, R., Khiat, A., Regoutz, A. and Prodromakis, T., 2015. 'A cell classifier for RRAM process development'. *IEEE Transactions on Circuits and Systems II: Express Briefs, 62(7)*, pp.676-680.
2. Gupta, I., Serb, A., Khiat, A., Zeitler, R., Vassanelli, S. and Prodromakis, T., 2016. 'Real-time encoding and compression of neuronal spikes by metal-oxide memristors'. *Nature communications, 7*.
3. Gupta, I., Serb, A., Khiat, A. and Prodromakis, T., 2016, May. 'Practical operation considerations for memristive integrating sensors'. *IEEE International Symposium on Circuits and Systems (ISCAS), 2016* (pp. 2322-2325). (Best paper award)
4. Gupta, I., Serb, A., Khiat, A. and Prodromakis, T., 2017. 'Improving Detection Accuracy of Memristor-Based Bio-Signal Sensing Platform'. *IEEE Transactions on Biomedical Circuits and Systems, 11*(1), pp.203-211.
5. Gupta, I., Serb, A., Khiat, A. and Prodromakis, T., 2016. 'Towards a memristor-based spike-sorting platform'. *IEEE Biomedical Circuits and Systems*, pp.1-4.
6. Gupta, I., Serb, A., Berdan, R., Khiat, A. and Prodromakis, T., 2017. 'Volatility Characterization for RRAM Devices'. *IEEE Electron Device Letters, 38(1)*, pp.28-31.
7. Gupta, I., Serb, A., Khiat, A., Zeitler, R., Vassanelli, S. and Prodromakis, T., 2018. 'Sub 100nW volatile nano-metal-oxide memristor as synaptic-like encoder of neuronal spikes.' *IEEE Transactions on Biomedical Circuits and Systems.*
8. Gupta, I., Serb, A., Khiat, A., and Prodromakis, T., 'Mitigating noise effects in volatile nano-metal oxide devices, *IEEE International Symposium on Circuits and Systems (ISCAS), 2017.*
9. Gupta, I., Serb, A., Khiat, A., and Prodromakis, T., 'Spike Sorting using non-volatile metal-oxide devices, *Nature Scientific Reports, Under Review.*

F.2 Contributions to group's research

1. Trapatseli, M., Carta, D., Regoutz, A., Khiat, A., Serb, A., Gupta, I. and Prodromakis, T., 2015. 'Conductive atomic force microscopy investigation of switching thresholds in titanium dioxide thin films'. *The Journal of Physical Chemistry C, 119(21)*, pp.11958-11964.
2. Regoutz, A., Gupta, I., Serb, A., Khiat, A., Borgatti, F., Lee, T.L., Schlueter, C., Torelli, P., Gobaut, B., Light, M. and Carta, D., 2016. 'Role and Optimization of the Active Oxide Layer in TiO_2 based RRAM'. *Advanced Functional Materials*, 26(4), pp.507-513.
3. Carta, D., Hitchcock, A.P., Guttmann, P., Regoutz, A., Khiat, A., Serb, A., Gupta, I. and Prodromakis, T., 2016. 'Spatially resolved TiO_x phases in switched RRAM devices using soft X-ray spectromicroscopy'. *Scientific reports, 6*.
4. Carta, D., Guttmann, P., Regoutz, A., Khiat, A., Serb, A., Gupta, I., Mehonic, A., Buckwell, M., Hudziak, S., Kenyon, A.J. and Prodromakis, T., 2016. 'X-ray spectromicroscopy investigation of soft and hard breakdown in RRAM devices'. *Nanotechnology, 27(34)*, p.345705.
5. Messaris, I., Nikolaidis, S., Serb, A., Stathopoulos, S., Gupta, I., Khiat, A., & Prodromakis, T. (2017, May), *A TiO_2 ReRAM parameter extraction method*, 2017 IEEE International Symposium on circuits and Systems (ISCAS) (pp. 1-4).

F.3 Innovations

1. Method and system for processing data from a sensor, filed on 18th October, 2016. (Now PCT level)
2. Method and system for detecting events in an input signal, filed on 11th November, 2016. (Now PCT level)

One who reads, writes, sees, inquires, lives in the company of learned, his intellect expands as the lotus petals expands due to the rays of sun. (Subhashitani)

Bibliography

[1] György Buzsáki. Large-scale recording of neuronal ensembles. *Nature neuroscience*, 7 (5):446–51, may 2004. ISSN 1097-6256. doi: 10.1038/nn1233. URL .

[2] Timothy J Blanche, Martin A Spacek, Jamille F Hetke, Nicholas V Swindale, and J Timothy. Polytrodes : High-Density Silicon Electrode Arrays for Large-Scale Multiunit Recording. *Journal of neural engineering*, 93.5:2987–3000, 2005. doi: 10.1152/jn.01023.2004.

[3] Felix Franke, David Jäckel, Jelena Dragas, Jan Müller, Milos Radivojevic, Douglas Bakkum, and Andreas Hierlemann. High-density microelectrode array recordings and real-time spike sorting for closed-loop experiments : an emerging technology to study neural plasticity. 6(December):1–7, 2012. doi: 10.3389/fncir.2012.00105.

[4] R. Lambacher, A., Vitzthum, V., Zeitler and P. Eickenscheidt, M., Eversmann, B., Thewes, R., Fromherz. Identifying firing mammalian neurons in networks with high-resolution multi-transistor array (MTA). *Applied Physics A*, 102(1):1–11, sep 2010. ISSN 0947-8396. doi: 10.1007/s00339-010-6046-9. URL .

[5] S Frey, U, Egert, U, Heer, F, Hafizovic and A Hierlemann. Microelectronic system for high-resolution mapping of extracellular electric fields applied to brain slices. *Biosensors & bioelectronics*, 24(7):2191–8, mar 2009. ISSN 1873-4235. doi: 10.1016/j.bios.2008.11.028. URL .

[6] Zoltán Berényi, Antal, Somogyvári, Eran Nagy, Anett J, Roux, Lisa, Long, John D, Fujisawa, Shigeyoshi, Stark, and György Leonardo, Anthony, Harris, Timothy D, Buzsáki. Large-scale, high-density (up to 512 channels) recording of local circuits in behaving animals. *Journal of neurophysiology*, 111(5):1132–49, mar 2014. ISSN 1522-1598. doi: 10.1152/jn.00785.2013. URL .

[7] X Navarro, T B Krueger, N Lago, S Micera, T Stieglitz, and P Dario. A Critical Review of Interfaces with the Peripheral Nervous System for the Control of Neuroprotheses and Hybrid Bionic Systems. *Journal of Peripheral Nervous System*, 10:229–258, 2005. ISSN 1085-9489. doi: 10.1111/j.1085-9489.2005.10303.x.

[8] Carolina Mora Lopez, Srinjoy Mitra, Jan Putzeys, Bogdan Raducanu, Marco Ballini, Alexandru Andrei, Simone Severi, Marleen Welkenhuysen, Chris Van Hoof, Silke Musa,

and Refet Firat Yazicioglu. A 966-electrode neural probe with 384 configurable channels in 0.13um SOI CMOS. *2016 IEEE International Solid-State Circuits Conference (ISSCC)*, pages 392–393, 2016. doi: 10.1109/ISSCC.2016.7418072. URL .

[9] Roland Eversmann, Bjorn, Lambacher, Armin, Gerling, Thomas, Kunze, Alexander, Fromherz, Peter, Thewes. A neural tissue interfacing chip for in-vitro applications with 32k recording / stimulation channels on an active area of 2.6 mm2. In *2011 Proceedings of the ESSCIRC (ESSCIRC)*, pages 211–214. IEEE, sep 2011. ISBN 978-1-4577-0703-2. doi: 10.1109/ESSCIRC.2011.6044902.

[10] Benoit Gosselin. Recent advances in neural recording microsystems. *Sensors*, 11(5): 4572–4597, 2011. ISSN 14248220. doi: 10.3390/s110504572.

[11] A Hierlemann, U Frey, S Hafizovic, and F Heer. Growing Cells Atop Microelectronic Chips: Interfacing Electrogenic Cells In Vitro With CMOS-Based Microelectrode Arrays. *Proceedings of the IEEE*, 99(2):252–284, feb 2011. ISSN 0018-9219. doi: 10.1109/JPROC.2010.2066532.

[12] Hernan Gonzalo Rey, Carlos Pedreira, and Rodrigo Quian Quiroga. Past, present and future of spike sorting techniques. *Brain research bulletin*, 119:106–117, apr 2015. ISSN 1873-2747. doi: 10.1016/j.brainresbull.2015.04.007. URL .

[13] Michael S Lewicki. A review of methods for spike sorting : the detection and classification of neural action potentials classification of neural action potentials. *Network: Computation in Neural systems*, 9.4(January):R53–R78, 1998. doi: 10.1088/0954-898X.

[14] Edward M Schmidt. Computer seperation of multi-unit neuroelectric data: a review. *Journal of Neuroscience Methods*, 12:95–111, 1984.

[15] Rodrigo Quian Quiroga. Spike sorting. *Current biology : CB*, 22(2):R45–6, jan 2012. ISSN 1879-0445. doi: 10.1016/j.cub.2011.11.005. URL .

[16] Miguel A L Nicolelis and Mikhail A Lebedev. Principles of neural ensemble physiology underlying the operation of brain-machine interfaces. *Nat Rev Neurosci*, 10(7):530–540, jul 2009. ISSN 1471-003X. URL .

[17] Nicholas G. Hatsopouplos and John P. Donoghue. The Science of Neural Interface Systems. *Annu Rev Neurosci*, 32:249–266, 2009. doi: 10.1146/annurev.neuro.051508.135241.The.

[18] Mark L Homer, Arto V Nurmikko, John P Donoghue, and Leigh R Hochberg. Sensors and decoding for intracortical brain computer interfaces. *Annual review of biomedical engineering*, 15:383–405, 2013.

[19] Luis Fernando Nicolas-Alonso and Jaime Gomez-Gil. Brain computer interfaces, a review. *Sensors (Basel, Switzerland)*, 12(2):1211–79, jan 2012. ISSN 1424-8220. doi: 10.3390/s120201211. URL .

[20] By Kensall D Wise, Amir M Sodagar, Ying Yao, Mayurachat Ning Gulari, Gayatri E Perlin, and Khalil Najafi. Microelectrodes, Microelectronics, and Implantable Neural Microsystems. *Proceedings of the IEEE*, 96(7):1184–1202, 2008.

[21] Gytis Baranauskas. What limits the performance of current invasive brain machine interfaces? *Frontiers in systems neuroscience*, 8(April):68, jan 2014. ISSN 1662-5137. doi: 10.3389/fnsys.2014.00068. URL .

[22] Konrad P Stevenson, Ian HKording. How advances in neural recording affect data analysis. *Nature neuroscience*, 14(2):139–42, feb 2011. ISSN 1546-1726. doi: 10.1038/nn.2731. URL .

[23] Carlos Pedreira, Juan Martinez, Matias J. Ison, and Rodrigo Quian Quiroga. How many neurons can we see with current spike sorting algorithms? *Journal of Neuroscience Methods*, 211:58–65, 2012. ISSN 01650270. doi: 10.1016/j.jneumeth.2012.07.010.

[24] Sivylla E. Paraskevopoulou and Timothy G. Constandinou. A sub-1uW neural spike-peak detection and spike-count rate encoding circuit. In *2011 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, pages 29–32. IEEE, nov 2011. ISBN 978-1-4577-1470-2. doi: 10.1109/BioCAS.2011.6107719.

[25] Amir Eftekhar, Sivylla E Paraskevopoulou, and Timothy G Constandinou. Towards a Next Generation Neural Interface : Optimizing Power , Bandwidth and Data Quality. pages 122–125, 2010.

[26] Deren Y Barsakcioglu, Yan Liu, Pooja Bhunjun, Joaquin Navajas, Amir Eftekhar, Andrew Jackson, Rodrigo Quian Quiroga, and Timothy G Constandinou. An analogue front-end model for developing neural spike sorting systems. *IEEE transactions on biomedical circuits and systems*, 8(2):216–227, 2014.

[27] Alberto Rodriguez-Perez, Jesús Ruiz-Amaya, Manuel Delgado-Restituto, and Angel Rodriguez-Vazquez. A low-power programmable neural spike detection channel with embedded calibration and data compression. *IEEE transactions on biomedical circuits and systems*, 6(2):87–100, 2012.

[28] Leon Chua and Sung Mo Kang. Memristive Devices and Systems. *Proceedings of the IEEE*, 64(2):209–223, 1976.

[29] Themistoklis Prodromakis, Christofer Toumazou, and Leon Chua. Two centuries of memristors. *Nature materials*, 11(6):478–81, jun 2012. ISSN 1476-1122. doi: 10.1038/nmat3338. URL .

[30] J Joshua Yang, Dmitri B Strukov, and Duncan R Stewart. Memristive devices for computing. *Nature nanotechnology*, 8(1):13–24, jan 2013. ISSN 1748-3395. doi: 10.1038/nnano.2012.240.

[31] Andreas Winkelmann. Wilhelm von waldeyer-hartz (1836–1921): An anatomist who left his mark. *Clinical Anatomy*, 20(3):231–234, 2007.

[32] Marco Piccolino. Luigi galvani and animal electricity: two centuries after the foundation of electrophysiology. *Trends in neurosciences*, 20(10):443–448, 1997.

[33] Edward George Tandy Liddell. Charles scott sherrington 1857-1952, 1952.

[34] Mikhail G Marblestone, Adam H, Zamft, Bradley M, Maguire, Yael G, Shapiro, David A Cybulski, Thaddeus R, Glaser, Joshua I, Amodei, Dario, Stranges, P, Benjamin, Kalhor, Reza, Dalrymple, Dongjin Seo, Elad Alon, Michel M Maharbiz, Jose M Carmena, Jan M Rabaey, Edward S Boyden, George M Church, and Konrad P Kording. Physical principles for scalable neural recording. *Frontiers in computational neuroscience*, 7:137, jan 2013. ISSN 1662-5188. doi: 10.3389/fncom.2013.00137.

[35] John G Nicholls, A Robert Martin, Bruce G Wallace, and Paul A Fuchs. *From neuron to brain*, volume 271. Sinauer Associates Sunderland, MA, 2001.

[36] Sonia Todorova, Patrick Sadtler, Aaron Batista, Steven Chase, and Valérie Ventura. To sort or not to sort: the impact of spike-sorting on neural decoding performance. *Journal of neural engineering*, 11(5):056005, oct 2014. ISSN 1741-2552. doi: 10.1088/1741-2560/11/5/056005. URL .

[37] Jonathan R Wolpaw, Niels Birbaumer, Dennis J McFarland, Gert Pfurtscheller, and Theresa M Vaughan. Brain–computer interfaces for communication and control. *Clinical neurophysiology*, 113(6):767–791, 2002.

[38] Areles Molleman. *Patch clamping: an introductory guide to patch clamp electrophysiology*. John Wiley & Sons, 2003.

[39] Electrical imaging of neuronal activity by multi-transistor-array (MTA) recording at 7.8 um resolution. 1611:1607–1611, 2004. doi: 10.1007/s00339-004-2991-5.

[40] Patrick Lichtsteiner, Christoph Posch, and Tobi Delbruck. A 128x128 120dB 15μs Latency Asynchronous Temporal Contrast Vision Sensor. *IEEE Journal of Solid-State Circuits*, 43(2):566–576, feb 2008. ISSN 0018-9200. doi: 10.1109/JSSC.2007.914337.

[41] Stefano Vassanelli and Mufti Mahmud. Trends and Challenges in Neuroengineering: Toward Intelligent Neuroprostheses through Brain-Brain Inspired Systems Communication. *Frontiers in Neuroscience*, 10(September):438, 2016. ISSN 1662-453X. doi: 10.3389/fnins.2016.00438. URL .

[42] Kilian Imfeld, Simon Neukom, Alessandro Maccione, Yannick Bornat, Sergio Martinoia, Pierre-André Farine, Milena Koudelka-Hep, and Luca Berdondini. Large-scale, high-resolution data acquisition system for extracellular recording of electrophysiological activity. *IEEE Transactions on biomedical engineering*, 55(8):2064–2073, 2008.

[43] Anoop C. Patil and Nitish V. Thakor. Implantable neurotechnologies: a review of micro- and nanoelectrodes for neural recording. *Medical and Biological Engineering and Computing*, 54(1):1–22, 2016. ISSN 17410444. doi: 10.1007/s11517-015-1430-4.

[44] Dominique M Durand, Maysam Ghovanloo, and Elliot Krames. Time to address the problems at the neural interface. *Journal of neural engineering*, 11(2):020201, 2014.

[45] U Frey, U Egert, D Jackel, J Sedivy, M Ballini, P Livi, F Faraci, F Heer, S Hafizovic, B Roscic, et al. Depth recording capabilities of planar high-density microelectrode arrays. In *Neural Engineering, 2009. NER'09. 4th International IEEE/EMBS Conference on*, pages 207–210. IEEE, 2009.

[46] Peter Fromherz. Microelectronics meets molecular and neurobiology. In *Electron Devices Meeting, 2001. IEDM'01. Technical Digest. International*, pages 16–1. IEEE, 2001.

[47] Christian Eversmann, Björn, Jenkner, Martin, Hofmann, Franz, Paulus, Birgit Brederlow, Ralf, Holzapfl, Matthias Fromherz, Peter, Merz, and Roland Brenner, Markus, Schreiter, Matthias, Gabl, Reinhard, Plehnert, Kurt, Steinhauser, Michael, Eckstein, Gerald, Schmitt-landsiedel, Doris, Thewes. A 128x128 CMOS Biosensor Array for Extracellular Recording of Neural Activity. *IEEE Journal of Solid-State Circuits*, 38(12): 2306–2317, 2003.

[48] Doris Eversmann, Bjoern, Hofmann, Franz, Holzapfl, Birgit, Schmitt-landsiedel and Alexander Lambacher, Armin, Zeitler, Ralf, Merz, Matthias, Kunze. CMOS Sensor Array for Electrical Imaging of Neuronal Activity. *IEEE INTERNATIONAL SYMPOSIUM ON CIRCUITS AND SYSTEMS*, 4:3479–3482, 2005.

[49] B. Eversmann, M. Jenkner, F. Hofmann, C. Paulus, R. Brederlow, B. Holzapfl, P. Fromherz, M. Merz, M. Brenner, M. Schreiter, R. Gabl, K. Plehnert, M. Steinhauser, G. Eckstein, D. Schmitt-Landsiedel, and R. Thewes. A 128 x 128 cmos biosensor array for extracellular recording of neural activity. *IEEE Journal of Solid-State Circuits*, 38 (12):2306–2317, dec 2003. ISSN 0018-9200. doi: 10.1109/JSSC.2003.819174.

[50] Kian Ann Ng, Elliot Greenwald, Yong Ping Xu, and Nitish V. Thakor. Implantable neurotechnologies: a review of integrated circuit neural amplifiers. *Medical & Biological Engineering & Computing*, 54(1):45–62, 2016. ISSN 0140-0118. doi: 10.1007/s11517-015-1431-3. URL .

[51] Robert Rieger and John T. Taylor. Design strategies for multi-channel low-noise recording systems. *Analog Integrated Circuits and Signal Processing*, 58(2):123–133, 2009. ISSN 09251030. doi: 10.1007/s10470-008-9230-5.

[52] Eric Bharucha, Hassan Sepehrian, and Benoit Gosselin. A Survey of Neural Front End Amplifiers and Their Requirements toward Practical Neural Interfaces. *Journal*

of Low Power Electronics and Applications, 4:268–291, 2014. ISSN 2079-9268. doi: 10.3390/jlpea4040268. URL .

[53] Carolina Mora Lopez, Alexandru Andrei, Srinjoy Mitra, Marleen Welkenhuysen, Wolfgang Eberle, Carmen Bartic, Robert Puers, Refet Firat Yazicioglu, and Georges GE Gielen. An implantable 455-active-electrode 52-channel cmos neural probe. *IEEE Journal of Solid-State Circuits*, 49(1):248–261, 2014.

[54] Zhe Chen. A primer on neural signal processing. *IEEE Circuits and Systems Magazine*, 17(1):33–50, 2017.

[55] Reid R Harrison, Paul T Watkins, Ryan J Kier, Robert O Lovejoy, Daniel J Black, Bradley Greger, and Florian Solzbacher. A low-power integrated circuit for a wireless 100-electrode neural recording system. *IEEE Journal of Solid-State Circuits*, 42(1): 123–133, 2007.

[56] Stefano Quian Quiroga, Rodrigo, Panzeri. Extracting information from neuronal populations: information theory and decoding approaches. *Nat Rev Neurosci*, 10(3):173–185, mar 2009. ISSN 1471-003X. URL .

[57] Rajeev C Nongpiur, Dale J Shpak, and Andreas Antoniou. Improved design method for nearly linear-phase iir filters using constrained optimization. *IEEE Transactions on Signal Processing*, 61(4):895–906, 2013.

[58] Carl Gold, Darrell A Henze, Christof Koch, and Gyorgy Buzsaki. On the origin of the extracellular action potential waveform: a modeling study. *Journal of neurophysiology*, 95(5):3113–3128, 2006.

[59] Sarah Gibson, Jack W Judy, and Dejan Markovic. Comparison of spike-sorting algorithms for future hardware implementation. *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, 2008:5015–5020, 2008. ISSN 1557-170X. doi: 10.1109/IEMBS.2008.4650340.

[60] R Quian Quiroga, Z Nadasdy, and Y Ben-Shaul. Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. *Neural computation*, 16(8): 1661–1687, 2004. ISSN 0899-7667. doi: 10.1162/089976604774201631.

[61] Sudipta Mukhopadhyay and GC Ray. A new interpretation of nonlinear energy operator and its efficacy in spike detection. *IEEE Transactions on biomedical engineering*, 45 (2):180–187, 1998.

[62] Kyung Hwan Kim and Sung June Kim. Neural spike sorting under nearly 0-db signal-to-noise ratio using nonlinear energy operator and artificial neural-network classifier. *IEEE Transactions on Biomedical Engineering*, 47(10):1406–1411, 2000.

[63] Rishi Chandra and Lance M Optican. Detection, classification, and superposition resolution of action potentials in multiunit single-channel recordings by an on-line real-time neural network. *IEEE Transactions on Biomedical Engineering*, 44(5):403–412, 1997.

[64] Ming Yin, David A Borton, Jacob Komar, Naubahar Agha, Yao Lu, Hao Li, Jean Laurens, Yiran Lang, Qin Li, Christopher Bull, et al. Wireless neurosensor for full-spectrum electrophysiology recordings during free behavior. *Neuron*, 84(6):1170–1182, 2014.

[65] Joaquin Navajas, Deren Y. Barsakcioglu, Amir Eftekhar, Andrew Jackson, Timothy G. Constandinou, and Rodrigo Quian Quiroga. Minimum requirements for accurate and efficient real-time on-chip spike sorting. *Journal of Neuroscience Methods*, 230:51–64, 2014. ISSN 1872678X. doi: 10.1016/j.jneumeth.2014.04.018. URL .

[66] Majid Zamani and Andreas Demosthenous. Feature extraction using extrema sampling of discrete derivatives for spike sorting in implantable upper-limb neural prostheses. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 22(4):716–726, 2014.

[67] Emery N Brown, Robert E Kass, and Partha P Mitra. Multiple neural spike train data analysis: state-of-the-art and future challenges. *Nature neuroscience*, 7(5):456–61, may 2004. ISSN 1097-6256. doi: 10.1038/nn1228. URL .

[68] R Quian Quiroga, Z Nadasdy, and Y Ben-Shaul. Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. *Neural computation*, 16(8): 1661–1687, 2004. ISSN 0899-7667. doi: 10.1162/089976604774201631.

[69] Robert Bestel, Andreas W Daus, and Christiane Thielemann. A novel automated spike sorting algorithm with adaptable feature extraction. *Journal of neuroscience methods*, 211(1):168–178, 2012.

[70] Fabian Kloosterman, Stuart P Layton, Zhe Chen, and Matthew A Wilson. Bayesian decoding using unsorted spikes in the rat hippocampus. *Journal of neurophysiology*, 111(1):217–227, 2013.

[71] Richard A Normann. Technology insight: future neuroprosthetic therapies for disorders of the nervous system. *Nature clinical practice. Neurology*, 3(8):444–452, 2007. ISSN 1745-8358. doi: 10.1038/ncpneuro0556. URL .

[72] Gordon E Moore. Cramming more components onto integrated circuits. *Proceedings of the IEEE*, 86(1):82–85, 1998.

[73] R. Stanley Williams. How we found the Missing Memristor. *IEEE Spectrum*, 2008.

[74] A Jeong, Doo Seok, Thomas, Reji, Katiyar, R S, Scott, J F, Kohlstedt, H, Petraru and Cheol Seong Hwang. Emerging memories: resistive switching mechanisms and current

status. *Reports on Progress in Physics*, 75(7), jul 2012. doi: 10.1088/0034-4885/75/7/076502.

[75] Dmitri B Strukov, Gregory S Snider, Duncan R Stewart, and R Stanley Williams. The missing memristor found. 453(May):80–84, 2008. doi: 10.1038/nature06932.

[76] J Joshua Yang, Matthew D Pickett, Xuema Li, Douglas a a Ohlberg, Duncan R Stewart, and R Stanley Williams. Memristive switching mechanism for metal/oxide/metal nanodevices. *Nature nanotechnology*, 3(7):429–433, jul 2008. ISSN 1748-3395. doi: 10.1038/nnano.2008.160. URL .

[77] Leon Chua. Memristor- The Missing Circuit Element. *IEEE Transactions on Circuit theory*, 18(5)(September):507–19, 1971.

[78] T. Prodromakis and C. Toumazou. A review on memristive devices and applications. *2010 IEEE International Conference on Electronics, Circuits, and Systems, ICECS 2010 - Proceedings*, pages 934–937, 2010. doi: 10.1109/ICECS.2010.5724666.

[79] Akihito Sawa. Resistive switching in Rapid advances in information technology rely on high-speed and. *Review Literature And Arts Of The Americas*, 11(6):28–36, 2008. URL .

[80] J Joshua Yang, Dmitri B Strukov, and Duncan R Stewart. Memristive devices for computing. *Nature nanotechnology*, 8:13–24, 2013. ISSN 1748-3395. doi: 10.1038/nnano.2012.240.

[81] K Szot, M Rogala, W Speier, Z Klusek, A Besmehn, and R Waser. TiO2–a prototypical memristive material. *Nanotechnology*, 22(25):254001, jun 2011. ISSN 1361-6528. doi: 10.1088/0957-4484/22/25/254001. URL .

[82] Rainer Waser and Masakazu Aono. Nanoionics-based resistive switching memories. *Nature materials*, 6(11):833–840, 2007.

[83] Deok-Hwang Kwon, Kyung Min Kim, Jae Hyuck Jang, Jong Myeong Jeon, Min Hwan Lee, Gun Hwan Kim, Xiang-Shu Li, Gyeong-Su Park, Bora Lee, Seungwu Han, Miyoung Kim, and Cheol Seong Hwang. Atomic structure of conducting nanofilaments in TiO2 resistive switching memory. *Nature nanotechnology*, 5(2):148–53, feb 2010. ISSN 1748-3395. doi: 10.1038/nnano.2009.456. URL .

[84] Gyeong-Su Park, Young Bae Kim, Seong Yong Park, Xiang Shu Li, Sung Heo, Myoung-Jae Lee, Man Chang, Ji Hwan Kwon, M Kim, U-In Chung, Regina Dittmann, Rainer Waser, and Kinam Kim. In situ observation of filamentary conducting channels in an asymmetric TaO5-x/TaO2-x bilayer structure. *Nature communications*, 4:2382, jan 2013. ISSN 2041-1723. doi: 10.1038/ncomms3382. URL .

[85] Fabien Alibart, Ligang Gao, Brian D Hoskins, and Dmitri B Strukov. High precision tuning of state for memristive devices by adaptable variation-tolerant algorithm. *Nanotechnology*, 23(7):075201, feb 2012. ISSN 1361-6528. doi: 10.1088/0957-4484/23/7/075201. URL .

[86] Spyros Stathopoulos, Ali Khiat, Maria Trapatseli, Simone Cortese, Alexantrou Serb, Ilia Valov, and Themis Prodromakis. Multibit memory operation of metal-oxide bilayer memristors. *Scientific reports*, 7(1):17532, 2017.

[87] J Joshua Yang, John Paul Strachan, Qiangfei Xia, Douglas a a Ohlberg, Philip J Kuekes, Ronald D Kelley, William F Stickle, Duncan R Stewart, Gilberto Medeiros-Ribeiro, and R Stanley Williams. Diffusion of adhesion layer metals controls nanoscale memristive switching. *Advanced materials (Deerfield Beach, Fla.)*, 22(36):4034–8, sep 2010. ISSN 1521-4095. doi: 10.1002/adma.201000663. URL .

[88] Antonio C Torrezan, John Paul Strachan, Gilberto Medeiros-Ribeiro, and R Stanley Williams. Sub-nanosecond switching of a tantalum oxide memristor. *Nanotechnology*, 22(48):485203, dec 2011. ISSN 1361-6528. doi: 10.1088/0957-4484/22/48/485203. URL .

[89] Kuk-Hwan Kim, Sung Hyun Jo, Siddharth Gaba, and Wei Lu. Nanoscale resistive memory with intrinsic diode characteristics and long endurance. *Applied Physics Letters*, 96(5):053106, 2010. ISSN 00036951. doi: 10.1063/1.3294625. URL .

[90] Fabien Alibart, Ligang Gao, Brian D Hoskins, and Dmitri B Strukov. High precision tuning of state for memristive devices by adaptable variation-tolerant algorithm. *Nanotechnology*, 23(7):075201, feb 2012. ISSN 1361-6528. doi: 10.1088/0957-4484/23/7/075201. URL .

[91] J. Joshua Yang, M.-X. Zhang, John Paul Strachan, Feng Miao, Matthew D. Pickett, Ronald D. Kelley, G. Medeiros-Ribeiro, and R. Stanley Williams. High switching endurance in TaO[sub x] memristive devices. *Applied Physics Letters*, 97(23):232102, 2010. ISSN 00036951. doi: 10.1063/1.3524521. URL .

[92] Shimeng Yu, Hong-Yu Chen, Bin Gao, Jinfeng Kang, and H-S Philip Wong. Hfox-based vertical resistive switching random access memory suitable for bit-cost-effective three-dimensional cross-point architecture. *ACS nano*, 7(3):2320–2325, 2013.

[93] Myoung-Jae Lee, Chang Bum Lee, Dongsoo Lee, Seung Ryul Lee, Man Chang, Ji Hyun Hur, Young-Bae Kim, Chang-Jung Kim, David H. Seo, Sunae Seo, U-In Chung, In-Kyeong Yoo, and Kinam Kim. A fast, high-endurance and scalable non-volatile memory device made from asymmetric Ta2O5x/TaO2x bilayer structures. *Nature Materials*, 10 (8):625–630, 2011. ISSN 1476-1122. doi: 10.1038/nmat3070. URL .

[94] Seonghyun Kim, Kuyyadi P Biju, Minseok Jo, Seungjae Jung, Jubong Park, Joonmyoung Lee, Wootae Lee, Jungho Shin, Sangsu Park, and Hyunsang Hwang. Effect

of scaling wox-based rrams on their resistive switching characteristics. *IEEE Electron Device Letters*, 32(5):671–673, 2011.

[95] WC Chien, YC Chen, EK Lai, YD Yao, P Lin, SF Horng, J Gong, TH Chou, HM Lin, MN Chang, et al. Unipolar switching behaviors of rto wox rram. *IEEE Electron Device Letters*, 31(2):126–128, 2010.

[96] Jubong Park, Seungjae Jung, Joonmyoung Lee, Wootae Lee, Seonghyun Kim, Jungho Shin, and Hyunsang Hwang. Resistive switching characteristics of ultra-thin tiox. *Microelectronic Engineering*, 88(7):1136–1139, 2011.

[97] HY Lee, YS Chen, PS Chen, TY Wu, F Chen, CC Wang, PJ Tzeng, M-J Tsai, and C Lien. Low-power and nanosecond switching in robust hafnium oxide resistive memory with a thin ti cap. *IEEE Electron Device Letters*, 31(1):44–46, 2010.

[98] Ya-Ting Wu, Shyankay Jou, and Ping-Jung Yang. Resistance switching of thin alox and cu-doped-alox films. *Thin Solid Films*, 544:24–27, 2013.

[99] Peng Lin, Shuang Pi, and Qiangfei Xia. 3D integration of planar crossbar memristive devices with CMOS substrate. *Nanotechnology*, 25(40):405202, 2014. ISSN 1361-6528. doi: 10.1088/0957-4484/25/40/405202. URL .

[100] Shuang Pi, Peng Lin, Hao Jiang, Can Li, and Qiangfei Xia. Device engineering and CMOS integration of nanoscale memristors. *Proceedings - IEEE International Symposium on Circuits and Systems*, pages 425–427, 2014. ISSN 02714310. doi: 10.1109/ISCAS.2014.6865156.

[101] Rui Yang, Kazuya Terabe, Yiping Yao, Tohru Tsuruoka, Tsuyoshi Hasegawa, James K Gimzewski, and Masakazu Aono. Synaptic plasticity and memory functions achieved in a WO3-x-based nanoionics device by using the principle of atomic switch operation. *Nanotechnology*, 24(38):384003, sep 2013. ISSN 1361-6528. doi: 10.1088/0957-4484/24/38/384003.

[102] Themistoklis Prodromakis and Chris Toumazou. A review on memristive devices and applications. In *Electronics, Circuits, and Systems (ICECS), 2010 17th IEEE International Conference on*, pages 934–937. IEEE, 2010.

[103] Bernabé Linares-barranco and Teresa Serrano-gotarredona. Memristance can explain Spike-Time- Dependent-Plasticity in Neural Synapses. *Nature precedings*, (1), 2009. doi: 10101/npre.2009.3010.1.

[104] Yi Li, Yingpeng Zhong, Lei Xu, Jinjian Zhang, Xiaohua Xu, Huajun Sun, and Xiangshui Miao. Ultrafast synaptic events in a chalcogenide memristor. *Scientific reports*, 3: 1619, jan 2013. ISSN 2045-2322. doi: 10.1038/srep01619. URL .

[105] Radu Berdan, Eleni Vasilaki, Ali Khiat, Giacomo Indiveri, Alexandru Serb, and Themistoklis Prodromakis. Emulating short-term synaptic dynamics with memristive devices. *Scientific Reports*, 6, 2016. ISSN 2045-2322. doi: 10.1038/srep18639.

[106] Dmitri B Strukov. Nanotechnology: smart connections. *Nature*, 476(7361):403, 2011.

[107] Alexantrou Serb, Radu Berdan, Ali Khiat, Christos Papavassiliou, and Themistoklis Prodromakis. Live demonstration : A versatile , low-cost platform for testing large ReRAM cross-bar arrays . *IEEE INTERNATIONAL SYMPOSIUM ON CIRCUITS AND SYSTEMS*, 9(5):4799, 2014.

[108] I Valov, E Linn, S Tappertzhofen, S Schmelzer, J van den Hurk, F Lentz, and R Waser. Nanobatteries in redox-based resistive switches require extension of memristor theory. *Nature communications*, 4:1771, 2013. ISSN 2041-1723. doi: 10.1038/ncomms2784.

[109] Nor Zaidi Haron and Said Hamdioui. On Defect Oriented Testing for Hybrid CMOS/Memristor Memory. *2011 Asian Test Symposium*, pages 353–358, nov 2011. doi: 10.1109/ATS.2011.66. URL .

[110] Sachhidh Kannan, Jeyavijayan Rajendran, Ramesh Karri, and Ozgur Sinanoglu. Sneak-path testing of crossbar-based nonvolatile random access memories. *IEEE Transactions on Nanotechnology*, 12:413–426, 2013. ISSN 1536125X. doi: 10.1109/TNANO.2013.2253329.

[111] T. Prodromakis, K. Michelakis, and C. Toumazou. Switching mechanisms in microscale memristors. *Electronics Letters*, 46(1):63, 2010. ISSN 00135194. doi: 10.1049/el.2010.2716. URL .

[112] Radu Berdan, Chuan Lim, Ali Khiat, Christos Papavassiliou, and Themis Prodromakis. A memristor spice model accounting for volatile characteristics of practical reram. *IEEE Electron Device Letters*, 35(1):135–137, 2014.

[113] Dae Eun Seok, Jun Yeong, Song, Seul Ji, Yoon, Jung Ho, Yoon, Kyung Jean, Park, Tae Hyung, Kwon, Gun Hwan Lim, Hyungkwang, Kim, and Cheol Seong Jeong, Doo Seok, Hwang. A Review of Three-Dimensional Resistive Switching Cross-Bar Array Memories from the Integration and Materials Property Points of View. *Advanced Functional Materials*, 24(34):5316–5339, sep 2014. ISSN 1616301X. doi: 10.1002/adfm.201303520. URL .

[114] Dmitri B. Strukov and R. Stanley Williams. Exponential ionic drift: Fast switching and low volatility of thin-film memristors. *Applied Physics A: Materials Science and Processing*, 94(3):515–519, 2009. ISSN 09478396. doi: 10.1007/s00339-008-4975-3.

[115] Simone Cortese, Maria Trapatseli, Ali Khiat, and Themistoklis Prodromakis. On the origin of resistive switching volatility in Ni/TiO2/Ni stacks. *Journal of Applied Physics*, 120(6):065104, 2016. ISSN 0021-8979. doi: 10.1063/1.4960690.

[116] Ali Khiat, Peter Ayliffe, and Themistoklis Prodromakis. High Density Crossbar Arrays with Sub- 15 nm Single Cells via Liftoff Process Only. *Scientific reports*, 6:32614, 2016. ISSN 2045-2322. doi: 10.1038/srep32614.

[117] Ali Khiat, Iulia Salaoru, and Themistoklis Prodromakis. Resistive switching characteristics of indium-tin-oxide thin film devices. *Physica Status Solidi (a)*, 211(5):1194–1199, may 2014. ISSN 18626300. doi: 10.1002/pssa.201330646. URL .

[118] Isha Gupta, Alexantrou Serb, Ali Khiat, Ralf Zeitler, Stefano Vassanelli, and Themistoklis Prodromakis. Real-time encoding and compression of neuronal spikes by metal-oxide memristors. *Nature communications*, 7:12805, 2016.

[119] Radu Berdan, Alexander Serb, Ali Khiat, Anna Regoutz, Christos Papavassiliou, and Themis Prodromakis. A u-Controller-Based System for Interfacing Selectorless RRAM Crossbar Arrays. *IEEE Transactions on Electron Devices*, 62(7):2190–2196, jul 2015. ISSN 0018-9383. doi: 10.1109/TED.2015.2433676.

[120] S Vassanelli and P Fromherz. Transistor probes local potassium conductances in the adhesion region of cultured rat hippocampal neurons. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 19(16):6767–6773, 1999. ISSN 1529-2401.

[121] Ronen Segev, Joe Goodhouse, Jason Puchalla, and Michael J Berry II. Recording spikes from a large fraction of the ganglion cells in a retinal patch. *Nature neuroscience*, 7 (10):1155, 2004.

[122] Lieuwe B. Leene, Yan Liu, and Timothy G. Constandinou. A compact recording array for neural interfaces. *2013 IEEE Biomedical Circuits and Systems Conference, BioCAS 2013*, pages 97–100, 2013. doi: 10.1109/BioCAS.2013.6679648.

[123] Ralf Zeitler, Peter Fromherz, and Gunther Zeck. Extracellular voltage noise probes the interface between retina and silicon chip. *Applied Physics Letters*, 99(26):263702, 2011. ISSN 00036951. doi: 10.1063/1.3672224. URL .

[124] Günther Zeck, Armin Lambacher, and Peter Fromherz. Axonal transmission in the retina introduces a small dispersion of relative timing in the ganglion cell population response. *PloS one*, 6(6):e20810, jan 2011. ISSN 1932-6203. doi: 10.1371/journal.pone.0020810. URL .

[125] Alexander Serb, William Redman-White, Christos Papavassiliou, and Themistoklis Prodromakis. Practical Determination of Individual Element Resistive States in Selectorless RRAM Arrays. *IEEE Transactions on Circuits and Systems I: Regular Papers*, PP(99):1–9, 2015. doi: 10.1109/TCSI.2015.2476296.

[126] Isha Gupta, Alexantrou Serb, Ali Khiat, and Themistoklis Prodromakis. Practical operation considerations for memristive integrating sensors. In *Circuits and Systems (ISCAS), 2016 IEEE International Symposium on*, pages 2322–2325. IEEE, 2016.

[127] Themistoklis Prodromakis, Boon Pin Peh, Christos Papavassiliou, and Christofer Toumazou. A versatile memristor model with nonlinear dopant kinetics. *IEEE transactions on electron devices*, 58(9):3099–3105, 2011.

[128] Alexander Serb, Ali Khiat, and Themistoklis Prodromakis. An RRAM Biasing Parameter Optimizer. *IEEE Transactions on Electron Devices*, 62(11):3685 – 3691, 2015. ISSN 00189383. doi: 10.1109/TED.2015.2478491.

[129] Isha Gupta, Alexantrou Serb, Ali Khiat, and Themistokalis Prodromakis. Improving detection accuracy of memristor-based bio-signal sensing platform. *IEEE transactions on biomedical circuits and systems*, 11(1):203–211, 2017.

[130] Ilan N. Goodman and Don H. Johnson. Information theoretic bounds on neural prosthesis effectiveness: The importance of spike sorting. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, pages 5204–5207, 2008. ISSN 15206149. doi: 10.1109/ICASSP.2008.4518832.

[131] J Xing, A Serb, R Berdan, H Xui, and T Prodromakis. An FPGA-based instrument for en-masse RRAM characterisation with ns pulsing resolution. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 2015.

[132] Henry Markram and Misha Tsodyks. Redistribution of synaptic efficacy between neocortical pyramidal neurons. *Nature*, 382(6594):807, 1996.

[133] Isha Gupta, Alexantrou Serb, Radu Berdan, Ali Khiat, and Themistoklis Prodromakis. Volatility characterization for rram devices. *IEEE Electron Device Letters*, 38(1):28–31, 2017.

[134] Isha Gupta, Alexantrou Serb, Ali Khiat, Ralf Zeitler, Stefano Vassanelli, and Themistoklis Prodromakis. Sub 100nW volatile nano metal-oxide memristor as synaptic-like encoder of neuronal spikes. *ArXiv:1611.09671*.

[135] Isha Gupta, Alexantrou Serb, Radu Berdan, Ali Khiat, Ali Regoutz, and Themis Prodromakis. A Cell Classifier for RRAM Process Development. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 62(7):676–680, 2015. ISSN 15497747. doi: 10.1109/TCSII.2015.2415276.

[136] Isha Gupta, Alexantrou Serb, Ali Khiat, Ralf Zeitler, Stefano Vassanelli, and Themistoklis Prodromakis. Sub 100nw volatile nano-metal-oxide memristor as synaptic-like encoder of neuronal spikes. *arXiv preprint arXiv:1611.09671*, 2016.

[137] Isha Gupta, Alexantrou Serb, Ali Khiat, and Themistoklis Prodromakis. Mitigating noise effects in volatile nano-metal oxide neural detector. In *Circuits and Systems (ISCAS), 2017 IEEE International Symposium on*, pages 1–4. IEEE, 2017.

[138] Kristoffer Famm. A jump-start for electroceuticals. *Nature*, 496(7444):159–161, 2013.

[139] I Gupta, A Serb, A Khiat, and T Prodromakis. Towards a memristor-based spike-sorting platform. In *Biomedical Circuits and Systems Conference (BioCAS), 2016 IEEE*, pages 408–411. IEEE, 2016.

[140] Mufti Mahmud and Stefano Vassanelli. Processing and analysis of multichannel extracellular neuronal signals: State-of-the-art and challenges. *Frontiers in Neuroscience*, 10(JUN):1–12, 2016. ISSN 1662453X. doi: 10.3389/fnins.2016.00248.

[141] R. Quian Quiroga. What is the real shape of extracellular spikes? *Journal of Neuroscience Methods*, 177(1):194–198, 2009. ISSN 01650270. doi: 10.1016/j.jneumeth.2008.09.033.

[142] Isha Gupta, Alexantrou Serb, Ali Khiat, Maria Trapatseli, and Themistoklis Prodromakis. Spike sorting using non-volatile metal-oxide memristors. *arXiv preprint arXiv:1707.08772*, 2017.

[143] Juan Martinez, Carlos Pedreira, Matias J. Ison, and Rodrigo Quian Quiroga. Realistic simulation of extracellular recordings. *Journal of Neuroscience Methods*, 184(2):285–293, 2009. ISSN 01650270. doi: 10.1016/j.jneumeth.2009.08.017.