

Biorealistic Learning on Memrisitive Network



Viet Cuong Vu

Principal Supervisor: Anthony J. Kenyon

Subsidiary Supervisor: Adnan Mehonic

Department of Electronics and Electrical Engineering
University College London

This dissertation is submitted in fulfilment for the degree of
Doctor of Philosophy

June 2025

Declaration

I, Viet Cuong Vu, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Viet Cuong Vu
June 2025

Abstract

latex.

Impact Statement

latex.

Contributions

List of Publications

- V. C. Vu, A. Kenyon, D. Joksas, A. Mehonic, D. J. Mannion, and W. H. Ng, “Spiking Neural Networks with Nonidealities from Memristive Silicon Oxide Devices,” in 2024 IEEE 24th International Conference on Nanotechnology (NANO), Jul. 2024, pp. 46–50.
- D. J. Mannion, V. C. Vu, W. H. Ng, A. Mehonic, and A. J. Kenyon, “Unipolar Potentiation and Depression in Memristive Devices Utilizing the Subthreshold Regime,” IEEE Transactions on Nanotechnology, vol. 22, pp. 313–320, 2023.

Conference Presentations

- "Unipolar Potentiation and Depression within Optically Active Memristive Devices Subthreshold Regime", EMRS Spring 2025.
- "Circuit-Based Modelling of Current Transients within the Memristive Devices Subthreshold Regime", MEMRISYS 2024.
- "A Compact SPICE Model for Current Transients within the Subthreshold Regime of Memristors", IEEE MetroXRAINE 2023.

Acknowledgements

latex.

Table of contents

List of figures	xiii
List of tables	xv
1 Introduction	1
1.1 Research Context	1
1.2 Motivations and Challenges	1
1.3 Thesis Outline	1
2 Theoretical Foundations	3
2.1 Overview	3
2.2 Neuroscience Primers	4
2.2.1 Neuron Anatomy and Electrophysiology	5
2.2.2 Spiking Neuron Dynamics	7
2.2.3 Synaptic Transmission and Plasticity	13
2.3 Foundations of Neuromorphic Computing	16
2.3.1 Memristor Fundamentals	17
2.3.2 In-memory Computing Paradigms	22
2.3.3 Encoding Plasticity in Memristors	27
2.4 Architectures and System-Level Integration	29
2.4.1 Hierarchical Modular Architectures	29
2.4.2 Hardware-Software Co-Design	32
2.4.3 Experimental Validations Strategy	34
2.5 Summary	38
3 Fabrication and Characterisation Methodologies	39
3.1 Fabrication Procedure	39
3.1.1 Device Properties	39
3.1.2 Manufacturing Steps	41

3.1.3	Experimental Setup	44
3.2	Electrical Characterisation	49
3.2.1	Unipolar Switching Mode	49
3.2.2	Bipolar Switching Mode	54
3.2.3	Alternate Operating Modes	57
3.3	Resistive Switching in Silicon Oxide	61
3.3.1	Conduction Mechanisms	62
3.3.2	Switching Model Analysis	67
3.4	Summary	67
References		69

List of figures

2.1	Labeled diagram of the neuron.	5
2.2	Spiking dynamics of a neuron.	7
2.3	Hodgkin-Huxley neuron model.	8
2.4	The Leaky Integrate-and-Fire neuron model.	12
2.5	Conceptual symmetries of resistor, capacitor, inductor, and memristor	18
2.6	Typical I-V characteristic of a memristor	19
2.7	The main RRAM types.	20
2.8	The memristor-based crossbar architecture.	23
2.9	The address-event representation.	30
3.1	Device Structure	40
3.2	Radio Frequency Magnetron Sputtering	42
3.3	Contact profilometer for thin film thickness measurements.	43
3.4	Two wire or Four wire (Kelvin) testing.	45
3.5	Transimpedance amplifier circuit.	47
3.6	Experimental setup of spike train measurements.	48
3.7	Initial Electroformation step for unipolar switching.	50
3.8	Non-volatile switching behaviour for unipolar device under current compliance.	51
3.9	Observation of Set and Reset process under the same sweep.	52
3.10	Observation of Set and Reset process under the same sweep.	53
3.11	Switching in unipolar devices across different electrode sizes.	53
3.12	Initial electroformation step for unipolar sample via a double sweep curve.	54
3.13	Observation of bipolar switching in asymmetric device with -2V Set and 2V Reset sweeps.	55
3.14	Cycling stress test for bipolar devices.	56
3.15	Switching in bipolar devices across different electrode sizes.	58
3.16	Multi-levels I-V characteristics in MIM devices.	58
3.17	Gradual increase (left) and decrease (right) in conductivity for bipolar device.	59

3.18 Current-time plots showing transitions between resistive states.	60
3.19 Volatile activities observed under different constant current inputs.	61
3.20 : Conductive regions for filamentary switching and interface switching. . .	62
3.21 Energy-band diagrams showing different conduction mechanisms.	63

List of tables

Chapter 1

Introduction

1.1 Research Context

1.2 Motivations and Challenges

1.3 Thesis Outline

Chapter 2

Theoretical Foundations

2.1 Overview

Neuromorphic computing aims to bridge the gap between neuroscience and artificial intelligence by emulating the structural and functional principles of biological neural systems. At the heart of this vision is a pressing research question: how can we implement biorealistic learning on memristive networks to develop energy-efficient, scalable and adaptive computing architectures? Addressing this question requires a multidisciplinary approach, combining insights from neurobiology, electronic materials science and computational modelling.

Traditional von Neumann architectures [69], characterised by their separation of memory and processing units, struggle to efficiently handle tasks that the human brain performs effortlessly, such as pattern recognition, sensory integration, and decision making. In contrast, the brain achieves these feats with remarkable energy efficiency and adaptability, in part due to the tightly coupled nature of computation and memory in its neural circuits. Mimicking these biological characteristics in silicon and emerging nanotechnologies has become the guiding principle of neuromorphic engineering [58].

Recent advances in memristive technologies have reignited interest in neuromorphic computing. Memristors, or memory resistors, are two-terminal non-volatile devices that can emulate synaptic plasticity by adjusting their conductance based on the history of voltage and current applied. This property lend itself to naturally support learning rules such as Hebbian learning [25] and spike-timing-dependent plasticity (STDP) [49]. When arranged in crossbar arrays, memristors offer a promising platform for in-memory computation, which can significantly reduce the power and latency associated with traditional data transfer bottlenecks.

This chapter presents a comprehensive discussion of biorealistic learning mechanisms and their physical realisation on memristive networks. By grounding the discussion in the neuroscientific principles that underlie learning and cognition, the chapter aims to elucidate how these biological processes can be abstracted and implemented in hardware.

The chapter begins with an overview of the biological basis of computation, providing an essential neuroscience primer. It then moves to device-level considerations, discussing the properties of memristive devices and their integration into neuromorphic architectures. Throughout, the emphasis is on aligning computational models with biological fidelity, while navigating the constraints and opportunities offered by emerging nanotechnologies.

2.2 Neuroscience Primers

Computational neuroscience employs a computational methodology to elucidate the mechanisms underlying brain function. This entails not only identifying the computations performed by the brain but also understanding the interactions between brain elements, such as neurons and synapses, that facilitate these computations.

The brain is capable of performing a vast array of computations, with the fundamental units of the brain generally considered to be neurons and synapses. In the context of the nervous system, a synapse is defined as a structure that is capable of facilitating the transfer of an electrical or chemical signal from a presynaptic neuron to a postsynaptic neuron.

This section provides a concise overview of the relevant biological details, in addition to the concepts and models from computational neuroscience that are employed or expanded upon in this study. These details provide invaluable preliminary information for accurately modelling the implementation of silicon oxide device-based neuromorphic hardware and bio-inspired computing.

In addition to the fundamental biological details that are pertinent to the subject under discussion, the section provides information both at the neural level and at the network level. It is important to acknowledge that the models outlined in this study are comparatively rudimentary when placed in contrast to the substantial corpus of evidence that has been amassed on the neural system. This extensive body of evidence [34], constitutes the preponderance of neural data concerning these domains. Consequently, this section presents only the most fundamental biological facts relevant to the present work.

2.2.1 Neuron Anatomy and Electrophysiology

A neuron is a specialised biological cell that processes and transmits information through electrical and chemical signals [54]. They represent only one of the numerous cell types within the brain, yet they are the most frequently discussed due to their status as the primary computational entities. Their fundamental function is relatively straightforward: neurons receive input from other neurons, and if that input is sufficiently stimulating, they will fire an action potential (also known as a spike), which propagates to other neurons.

Figure 2.1 illustrates the basic structure of a neuron. Neurons can be subdivided into three principal parts: the dendrites, the cell body (soma), and the axon. Neurons receive input currents via their dendrites, which then transmit or channel this into the cell body, called the soma. When a neuron spikes, it sends current down its axon, which results in the release of neurotransmitter(s) at the synapses. These are connections from a neuron's axon to the dendrites of other neurons, and the neurotransmitter release causes dendritic input currents in these other connected neurons.

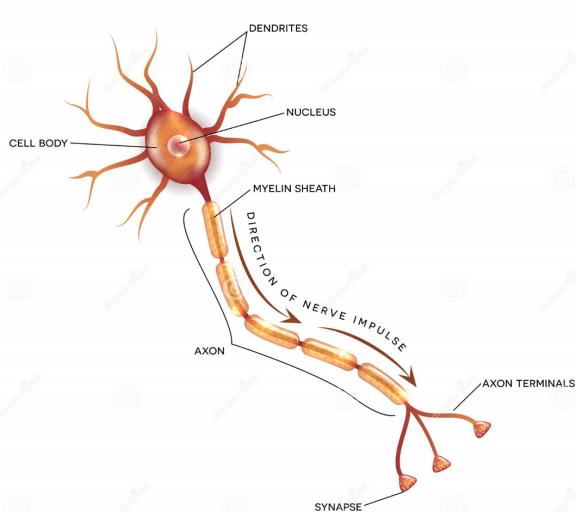


Fig. 2.1 Labeled diagram of the neuron, nerve cell that is the main part of the nervous system. A neuron's dendrites include synapses that allow it to accept input from other neurons. The dendrites carry current to the soma, which is where electrical charge is integrated. If the neuron membrane gets sufficiently polarised, an action potential (also known as a spike) travels down the axon. This causes neurotransmitters to be released at synapses, resulting in currents in the dendrites of postsynaptic neurons.

From a computational perspective, the soma represents the integration point for all incoming currents from dendrites [57], marking the initiation of the action potential generation

process (Figure 2.2). When a neuron is at rest, the soma exhibits a negative charge. This is referred to as the resting voltage and is maintained by ion pumps that regulate the concentration of ions (predominantly sodium, Na^+ , potassium, K^+ , and calcium, Ca^{+2}) within the cell.

As the currents arrive from the dendrites, they initiate a process of depolarisation of the cell [31]. Once the voltage within the soma reaches a sufficient level, it initiates the opening of voltage-activated sodium channels, which permit the influx of sodium ions into the cell, further depolarising it. This process persists until the electrical gradient resulting from the accumulation of sodium ions reaches a point where it is no longer in equilibrium with the chemical gradient caused by the imbalance of sodium within and outside the cell. This leads to a notable increase in the neuron's positive charge, exceeding the resting voltage.

Furthermore, this substantial depolarisation also activates voltage-gated potassium channels, which subsequently permit the release of potassium ions from the cell, thereby facilitating repolarisation. Concurrently, the sodium channels undergo inactivation. The opening of potassium channels ultimately results in the cell reaching a voltage below its resting level, a state known as hyperpolarisation. The sodium channels remain inactivated and the potassium channels remain open for a period of time following the spike.

The combination of these factors renders it almost impossible for the neuron to fire during this time; this is referred to as the absolute refractory period. The change in ionic concentrations within the cell is relatively minor during a single spike, but over the course of numerous spikes, the ion pumps are required to maintain the optimal concentrations of sodium and potassium. Other currents, most notably calcium currents, are present in some neurons.

The rapid depolarisation associated with an action potential not only causes an increase in the somatic voltage potential, but also results in partial depolarisation of the axon segments situated in closer proximity to the soma. This results in the opening of sodium channels in that part of the axon, which in turn causes further depolarisation and the opening of sodium channels in the subsequent section of the axon. In this way, the somatic spike triggers a voltage wave that travels down the axon, eventually leading to the release of neurotransmitter(s) from synaptic vesicles situated near the ends of the axon.

It has been established that all synapses located along a neuron's axon are responsible for the release of a singular neurotransmitter or a combination of neurotransmitters. This phenomenon is commonly referred to as Dale's principle. At the time of its development,

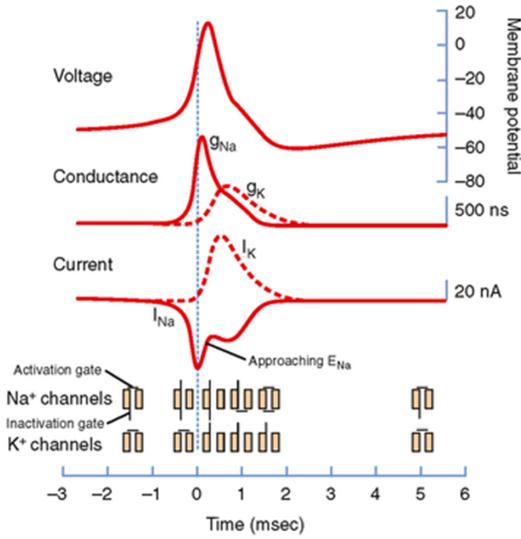


Fig. 2.2 The action potential and the underlying conductance and currents with respect to time [59]. It should be noted that the increased conductance for Na^+ (and its inward flow) is associated with the rising phase of the action potential, whereas the slower increase in conductance for K^+ (and its outward flow) is associated with repolarisation of the membrane and with afterhyperpolarisation. The reduction in I_{Na} before the peak of the action potential (even though G_{Na} is still high) is due to inactivation of the Na^+ channels.

Dale's principle was based on the assumption that each neuron produced a single type of neurotransmitter. Nevertheless, evidence of cotransmission was only discovered subsequently [3]. It was understood that neurotransmitters can only be either excitatory or inhibitory, in relation to different postsynaptic cells.

2.2.2 Spiking Neuron Dynamics

In recent times, the number of available neuron models has proliferated. The models currently in use in the literature range from the simplest possible rate-neuron model, namely binary threshold units [62], to complex multi-compartmental models that account for detailed dendritic morphologies [50]. In the context of large-scale neural models aiming to reproduce high-level behaviours, single-compartment neuron models remain the prevailing approach.

These models treat the neuron as a single electrical compartment, combining the dendrites, soma, and axon. In contrast, multi-compartmental models represent the neuron as comprising multiple electrical compartments, with equations that describe the influence of activity in one compartment on that of another. By modelling the spike separately from the rest of the neural dynamics, it is possible to separate time scales, thereby avoiding the need for additional

computational resources to model the spike trajectory [1].

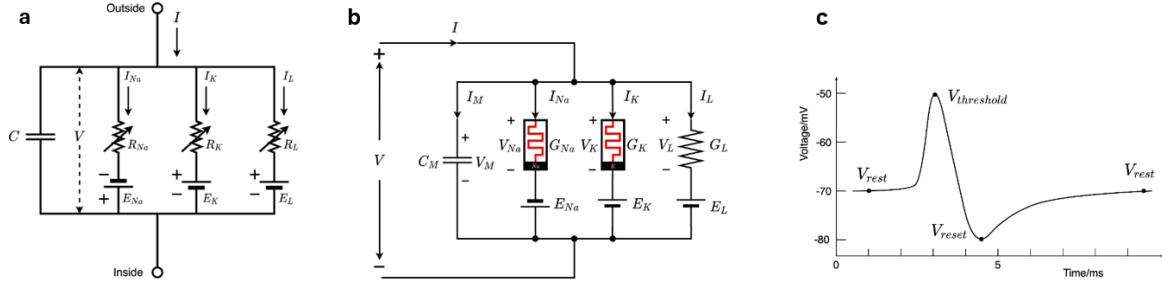


Fig. 2.3 Hodgkin-Huxley neuron model. (a) An equivalent circuit for the HH models [27]. (b) An equivalent circuit for memristive HH model [11]. (c) An action potential waveform, which demonstrates the resting, threshold, and reset potentials.

A significant proportion of the most influential findings in computational neuroscience are based on mathematically detailed models of neuronal functioning. One of the most renowned of these is the Hodgkin-Huxley model of the squid giant axon [27]. The Hodgkin-Huxley (HH) model is one of the most widely used, comprising a set of nonlinear differential equations that accurately approximate the electrical signals of neurons [11].

Figure 2.3(a) depicts the HH neural model, wherein the time-varying nonlinear conductor $R_{Na}(G_{Na})$ and $R_K(G_K)$ represent the sodium and potassium channels, respectively, while the linear conductor $R_L(G_L)$ simulates leak channels and C models the membrane of a neuron. The equations of the HH model are presented below:

$$C \frac{dV_m(t)}{dt} = I_C(t) + \sum_k I_k(t) \quad (2.1)$$

In this context, V_m represents the membrane potential. $\sum_k I_k(t)$ denotes the sum of the ionic currents flowing into the neuron. This can be formulated by three ion currents, as follows:

$$\sum_k I_k = C_m \frac{dV_m}{dt} + G_K n^4 (V_m - V_K) + G_{Na} m^3 (V_m - V_{Na}) + G_L (V_m - V_L) \quad (2.2)$$

$$\frac{dn}{dt} = \alpha_n(V_m)(1-n) - \beta_n(V_m)n \quad (2.3)$$

$$\frac{dm}{dt} = \alpha_m(V_m)(1-m) - \beta_m(V_m)m \quad (2.4)$$

$$\frac{dh}{dt} = \alpha_h(V_m)(1-h) - \beta_h(V_m)h \quad (2.5)$$

The reversal potentials V_K , V_{Na} , and V_L are the three parameters in question. The rate constants α_i and β_i , which depend on the membrane potential, describe the behaviour of the i^{th} ion channel. The maximal value of the conductance is represented by G_K , G_{Na} , and G_L .

Finally, the dimensionless quantities n , m , and h , which lie between 0 and 1, are associated with three ion channels. In order to achieve the optimal fit for human action potentials, the HH model is reduced by setting the leakage channel conductance to $G_L = 0$ [55]. It has been demonstrated that G_{Na} and G_K are memristors [12], in the equivalent circuit in figure 2.3(b).

The integrate-and-fire (IF) neuron [42] constituted one of the earliest computational models of a neuron. This model was developed prior to the ability of researchers to measure the electrical and chemical changes occurring in a functioning neuron. It is based on the premise that the neuron membrane can be modelled as a capacitor that stores charge over time [1].

As the name suggests, the IF model exhibits two principal behaviours: The model integrates current over time, as would be expected of a capacitor, and fires when the voltage reaches a threshold. Furthermore, the model may or may not incorporate a leak term, which represents a resistor in parallel with the capacitor that permits the dissipation of charge over time. The model with a leak term is typically designated as the leaky integrate-and-fire (LIF) model. While the term "integrate-and-fire (IF) model" can be used interchangably.

To identify how the neuron's membrane voltage evolves over time and, based on this, to determine when the neuron spikes, the charge Q across a capacitor is represented by $Q = V \times C$, where V is the voltage across the capacitor and C is the capacitance. By differentiating this with respect to time, the membrane voltage $V(t)$ of the neuron is:

$$C \frac{dV(t)}{dt} = J(t) \quad (2.6)$$

In this context, $J(t)$ represents the input current to the neuron over time, whereas C denotes the membrane capacitance. The current here is the time derivative of charge. Equation 2.6 demonstrates that the IF neuron simply integrates the input current over time. It is still necessary to identify the point at which the neuron spikes.

This is achieved by defining a threshold voltage, V_{th} , which is exceeded when the voltage passes this threshold, resulting in the neuron firing. This is a fundamental principle in neurophysiology: once the neuron voltage passes a threshold, the neuron begins firing a

spike, and once this firing process begins, it is almost impossible to reverse.

Once a neuron has fired a spike, the membrane voltage is reset to the resting potential, V_{rest} . This phenomenon can be attributed to physiological resetting procedures. Following the occurrence of a spike in a neuron, other ionic currents, typically potassium, are initiated, leading to a restoration of the membrane voltage towards the resting potential.

The leaky integrate-and-fire (LIF) model [39] incorporates an additional physiological factor: Neuron membranes are not perfect capacitors; rather, they slowly leak current over time, pulling the membrane voltage back to its resting potential. Therefore, the membrane is modelled as a capacitor and resistor in parallel, which allows for the neuron to exhibit a degree of "forgetting": in the absence of any input, the membrane voltage will return to its resting potential [40]. The LIF dynamics are captured by the following equation:

$$C \frac{dV(t)}{dt} = J(t) - \frac{1}{R}(V - V_{rest}) \quad (2.7)$$

In this model, R represents the membrane resistance, and the remaining parameters are consistent with those of the IF model, with identical resetting procedure.

The LIF model comprises a number of parameters, including C, R, V_{rest} and V_{th} . It is possible to normalise the model in order to reduce the number of parameters while maintaining the full dynamics of the original model. In particular, the model can be manipulated so that the normalised voltage lies within the range $[0, 1]$, with a normalised resting potential of zero and a normalised firing threshold of one. Initially, Equation 2.7 is multiplied by R to give:

$$\tau_{rc} \frac{dV}{dt} = RJ(t) - V + V_{rest} \quad (2.8)$$

$$\tau_{rc} = R \times C \quad (2.9)$$

$$\bar{V} = \frac{V - V_{rest}}{V_{th} - V_{rest}} \quad (2.10)$$

$$\bar{V}_{rest} = \frac{V_{rest}}{V_{th}} \quad (2.11)$$

By substituting \bar{V} and \bar{V}_{rest} into equation 2.3 to give:

$$\tau_{RC}(V_{th} - V_{rest}) \frac{d\bar{V}}{dt} = RJ(t) - \bar{V}(V_{th} - V_{rest}) \quad (2.12)$$

$$\tau_{rc} \frac{d\bar{V}}{dt} = \frac{R}{V_{th} - V_{rest}} J(t) - \bar{V} \quad (2.13)$$

$$\tau_{rc} \frac{d\bar{V}}{dt} = \bar{J}(t) - \bar{V} \quad (2.14)$$

When the firing threshold for the new equation $\bar{V}_{th} = 1$, the voltage resets to $\bar{V}_{rest} = 0$, and $\bar{J}(t) = \frac{R}{V_{th} - V_{rest}} J(t)$. It can be observed that $\bar{J}(t)$ is merely a linear transformation of $J(t)$. Consequently, (2.14) retains the full dynamics of (2.7) for a scaled input, but with only one parameter, τ_{RC} .

It should be noted that both \bar{V} and \bar{J} are unitless quantities. Conventionally, the unitless space is employed exclusively, and the quantities are often referred to simply as V and J , despite the fact that they are not voltages or currents. This simplifies the mathematical representation, without limiting the generality of the models.

(2.14) provides an exact description of the circumstances under which the model neuron will spike in response to a given input current, $J(t)$. However, in some cases, it is sufficient to consider only the spike rate, that is, the number of spikes per second that the neuron will produce in response to a given input current.

In the case of the LIF model, it is possible to determine the analytical firing rate for a constant input current. This is achieved by calculating the inter-spike interval (ISI), which is the time between one spike and the next. The firing rate is then given by the inverse of the ISI. When a constant input current, $J(t) = j$, is provided, it is possible to solve (2.14) in order to find the neuron voltage over time.

$$V(t) = (V(0) - j)e^{\frac{-t}{\tau_{rc}}} + j \quad (2.15)$$

In the absence of spikes, the objective is to ascertain the time required for the voltage to increase from $V(0) = 0$ to $V(t) = 1$. This property will only occur if $j > 1$. Substitution into (2.15) and subsequent solution for t yields:

$$t = -\tau_{RC} \log \left(-\frac{1}{j} \right) \quad (2.16)$$

Incorporating the refractory period and performing the inversion, the spike rate r for the LIF neuron is given by:

$$r = \begin{cases} \frac{1}{t_{ref} - \tau_{RC} \log\left(1 - \frac{1}{j}\right)} & \text{if } j > 1 \\ 0 & \text{otherwise} \end{cases} \quad (2.17)$$

The LIF model is one of the most widely utilised simplified neuron models [43]. The simple equivalent model is illustrated in Figure 2.4(a). In this model [60], a resistor R , connected in series with a *DC* source V_{rest}/V_{reset} , is connected in parallel with a capacitor C . A postsynaptic neuron receives a synaptic current $I(t)$, generated by presynaptic spikes.

A proportion of the current $I(t)$ flowing into C results in an increase in the membrane potential $V(t)$. The charge leakage occurs via resistor R . When $V(t)$ reaches a threshold value, the neuron generates a spike. Following the generation of a spike, the membrane potential is reset to the reset value.

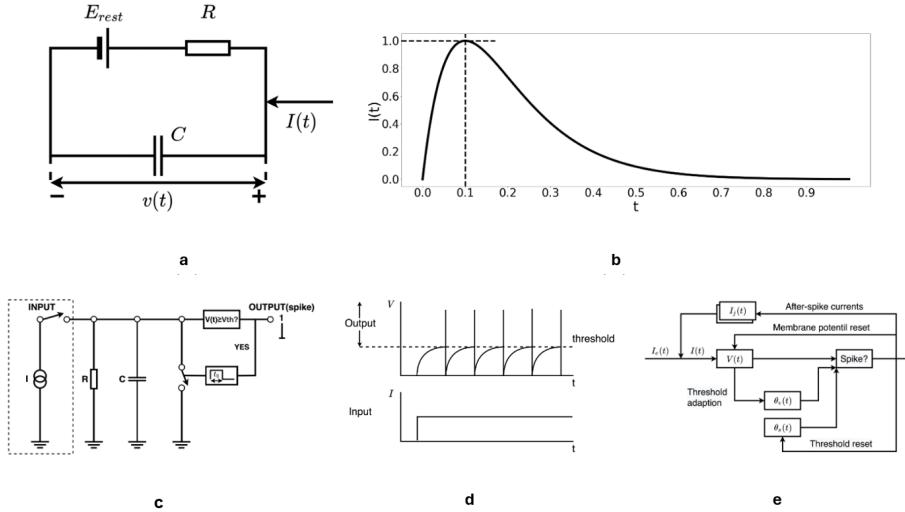


Fig. 2.4 The LIF neuron model. (a) Schematic diagram of the LIF electrical model. (b) Input current in the form of an alpha function, $\tau_\alpha = 0.1, I_0 = 1$ (c) The LIF model allows for the control of spiking behaviour through a comparison of membrane potential and threshold at each time step. Upon the triggering of a spike, a voltage-controlled switch discharges C for a duration corresponding to the refractory period t_0 [66]. (d) A simulation of constant firing frequency for DC current input in which t_0 is hidden. DC input current and output spikes are both shown. (e) A generalised LIF model with threshold control [67].

In the absence of $I(t)$, the voltage across C is eventually settled at V_{rest} , representing the cell's resting potential. During the refractory period t_0 , a neuron is incapable of spiking. Figure 2.4(c,d) illustrates the LIF neuron dynamics for the case of a *DC* input current and a zero rest and reset potential, $E_{reset} = E_{rest} = 0$ [66].

The input synaptic current, $I(t)$, can be described by a time-varying alpha function. However, alternative functions may be employed, including "Instantaneous Rise and Single Exponential Decay," "Biexponential Functions," "Sawtooth," and "Pulse Function." The alpha synaptic current is modeled by Equation 2.15, and the resulting plot is shown in Figure 2.2(b).

Nevertheless, Figure 2.4(a) lacks a circuit for resetting the system when the threshold is reached. In order to evaluate the inequality $V > V_{threshold}$, it is necessary to use an active circuit, such as a comparator. Upon reaching the threshold, the membrane potential must be reset in accordance with the illustration in Figure 2.4(d). Therefore, the LIF model's generalized version necessitates additional overhead, as illustrated in Figure 2.4(e).

It is evident that neurons manifest considerable heterogeneity with regard to their dynamics, morphology and connectivity. They can summarily be categorised as follows: Excitatory neurons are responsible for promoting activity in connected neurons. In contrast, inhibitory neurons are responsible for suppressing activity, a process that is crucial for stability and rhythm. Finally, interneurons connect local circuits, thereby enabling complex computations.

2.2.3 Synaptic Transmission and Plasticity

Prior to model internal neural dynamics, it is essential to model the dynamics of the synapses that connect neurons to one another. Synapses exert a significant functional effect as a low-pass filter on the spikes that pass through them. A spike in the presynaptic neuron elicits an extended current pulse in the postsynaptic neuron. This pulse can be conceptualised as a low-pass filtered version of the presynaptic spike.

The simplest model of a synapse is that of a first-order low-pass filter. The impulse response of a filter describes the manner in which the filter responds to an infinitesimally short input of unit integral, which is called an impulse. This idealised impulse is also a reasonable model of a spike, and thus the impulse response also describes what the postsynaptic current will look like in response to a presynaptic spike. The impulse response of the first-order low-pass

filter is as follows:

$$h(t) = \frac{1}{\tau_s} e^{\frac{t}{\tau_s}} \quad (2.18)$$

The synaptic time constant, denoted by τ_s , is defined as the length of time over which the postsynaptic current is spread. Given that the impulse response is an exponential function, the exponential synapse model is therefore a suitable description.

$$h(t) = \frac{1}{\tau_s^2} e^{\frac{t}{\tau_s}} \quad (2.19)$$

It was determined that a second-order lowpass filter is a superior model for a synapse [48]. The impulse response of this filter is defined by (2.19). This function is referred to as the alpha function, and thus the model is designated as the alpha synapse model. Both of these models are based on the current generated by a spike in the postsynaptic neuron, which is a current-based synapse model.

One of the primary objectives of computational neuroscience is to ascertain the manner in which the brain represents—or encodes—information. To this end, researchers have put forth a multitude of potential coding schemes that neurons could utilise for information encoding. One key distinction between rate coding and temporal coding is the following dichotomy.

In a rate code, the sole pertinent measure is the firing rate (i.e. the number of spikes) of a neuron over a given period of time. An example of a rate code is motor neurons in the peripheral nervous system. The contraction of a muscle is contingent upon the number of spikes per unit time; thus, only the rate of motor neuron spikes is significant [20]. In a temporal code, the time of individual spikes is also a factor. For example, in the early auditory system, precise spike timing facilitates the localisation of sounds [7].

The precise definitions of rate and temporal codes remain contentious, with differing interpretations presented by various authors [14]. To illustrate, a neuron may discharge a number of spikes in rapid succession, followed by a period of quiescence. A second neuron may be observed to fire the same number of spikes, but in a more evenly distributed manner over a given period.

Both rate codes and temporal codes describe the encoding properties of individual neurons. Additionally, one may inquire about the coding properties of a group (also known as a population) of neurons. The concept of population coding pertains to instances where a

representation is distributed across numerous neurons within a population, such that the represented value cannot be decoded from the activities of a limited number of neurons.

The simplest method of extrapolating the concept of rate or temporal coding to multiple neurons would be to have numerous neurons all implementing the same code. In other words, all neurons will exhibit a similar firing pattern when representing a given value, due to their comparable tuning properties. This results in a significant degree of redundancy between neurons.

In contrast, population coding entails each neuron representing a distinct aspect of the represented value. To illustrate, if the objective is to represent head direction, there are neurons that represent a head that is fully turned to the left, others that represent a head that is fully turned to the right, and still others that represent a centred head. Additionally, there are neurons that represent values in between these three head directions. The direction in which a neuron is most active is referred to as its preferred direction.

Synapses are therefore known to play a dual role in the nervous system. They facilitate communication between neurons and serve as the primary locus of learning and memory. The magnitude of this influence, or 'synaptic strength' [2], is determined by the relative strength of the connection between the presynaptic and postsynaptic neurons. This phenomenon is known as synaptic plasticity.

It is important to note that plasticity can be categorised into several distinct types. Short-Term Plasticity (STP): Transient changes that last from milliseconds to seconds. Long-Term Potentiation (LTP) and Long-Term Depression (LTD): The phenomenon of sustained increases or decreases in synaptic strength over time.

The most significant model of synaptic plasticity is Hebbian learning, which can be succinctly summarised as follows: The hypothesis that neurons that fire together wire together has been proven to be accurate. A more precise, temporally-sensitive rule is Spike-Timing Dependent Plasticity (STDP) [73].

Mathematical Models can capture the effect of precise spike timing on synaptic weight updates. If a presynaptic neuron fires before a postsynaptic neuron within a short window, the synapse is strengthened; if the order is reversed, the synapse weakens. A common

representation for this is:

$$\Delta w = \begin{cases} A_+ \cdot e^{-\Delta t / \tau_+}, & \Delta t > 0 \\ -A_- \cdot e^{-\Delta t / \tau_-}, & \Delta t < 0 \end{cases} \quad (2.20)$$

Where $\Delta t = t_{post} - t_{pre}$ is the timing difference, A_+, A_- are learning rates, τ_+, τ_- are time constants for potentiation and depression. This asymmetric window is indicative of experimental observations and provides a biologically plausible basis for synaptic learning in hardware.

As a small primer, memristive devices offer an electronic analogue to synapses due to their tunable conductance and memory of past activity. When configured in crossbar arrays, these devices have the capacity to implement synaptic weight matrices directly in hardware, with updates governed by local voltage or current pulses.

Memristive STDP implementations frequently exploit device physics, where conductance change is contingent on pulse overlap:

$$\Delta G = f(V_{pre}, V_{post}, \Delta t) \quad (2.21)$$

where f is a device-specific function determined by material properties and pulse shapes.

It is important to note that memristors have the capacity to inherently facilitate the nonlinear, history-dependent behaviour that is characteristic of biological plasticity rules, such as STDP. To illustrate this point, the application of carefully timed voltage pulses to a memristor has been demonstrated to result in an increase or decrease in conductance, respectively, reminiscent of LTP and LTD.

2.3 Foundations of Neuromorphic Computing

Extensive research has been conducted in device physics and material science to explore innovative materials and techniques for memories and prolonged retention objectives [28]. The term "neuromorphic" was created by researchers to describe new technologies and systems that, in addition to being essential for the construction of massive AI computer networks, exhibit certain behaviours that can be compared to those of real synapses [16].

Soon after, the notion of using these novel nanoscale components as "memristors" gained popularity, with the underlying notion being that they could be utilised to produce synapses in deep neural networks and sustain their synaptic weights locally [30]. The hardware and technology described could enable neural networks to perform "in-memory computing" and exhibit advanced non-linear properties, mimicking the physics of biological synapses [58].

The research in this field aims to develop various types of volatile and non-volatile memristive electronics. Additionally, spike or pulse-based control systems are being created to elicit biologically realistic learning behaviours in memristive cross-bar arrays. The challenging task is to find the perfect artificial synapse, which requires investigation into different materials, tools, and techniques.

2.3.1 Memristor Fundamentals

Memristive devices, also known as memory resistors, are emerging as foundational elements in neuromorphic computing due to their ability to retain resistive states based on electrical history, thereby emulating biological synapses. The central purpose of these components is to facilitate both memory storage and computation within a unified, compact structure, thereby enabling the co-location of memory and processing elements that is vital for brain-inspired architectures.

The presence of symmetry in nature, which is believed to arise from a common origin, is remarkable. However, the traditional electromagnetic passive circuit components of resistor, capacitor, and inductor are inadequate for describing the characteristics connected by the symmetry of circuit theory. Leon Chua addressed this issue by introducing the concept of a memristor in 1971 [9], which couples flux linkage and charge as a circuit device:

$$M(q) = \frac{d\phi}{dq} \quad (2.22)$$

where M denotes the memristance, a quantity whose value is known to be dependent on the history of the current that has previously passed through the device. This phenomenon gives rise to a form of resistive memory, wherein the device retains a memory of its previous state. However, proof of resistive switching in the memristor model was not established until 2008 [64].

There are three fundamental circuit elements and four essential circuit variables in basic electrical circuit theory. It is evident that one component is absent to achieve symmetry. This device ought to function in such a way that charge and magnetic flux are interconnected,

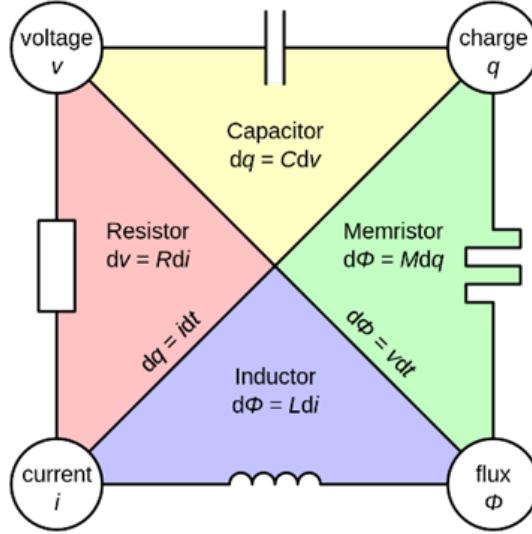


Fig. 2.5 Conceptual symmetries of resistor, capacitor, inductor, and memristor [17]. These four fundamental variables in circuit theory are depicted with their relationships. Each variable can be related to another via either a passive component or a well-known equation.

as illustrated in Figure 2.5. The link between the mathematical memristive model and a two-terminal resistive switching device is pivotal in this instance.

The concept of memristance is distinct from that of resistance in that it is dependent on charge, rather than being a constant value. The current is defined as the amount of charge flowing per unit time. Therefore, the expression for current can be written as:

$$q = \int_{-\infty}^{t_0} i(t) dt \quad (2.23)$$

where charge is the sum of current at a given time t_0 .

This indicates that the memristance, being dependent on charge, is determined by the historical currents that have previously passed through the device. In the event of interruption to the current flowing through the device, the memory state persists until the current flow is restored. The device is evidently equipped with a type of memory known as a "memristor".

In physical terms, memristors are often modeled as two-terminal devices whose resistance varies due to the drift of ions or vacancies in a dielectric medium. The state-dependent

resistance can be written as:

$$V(t) = R(w(t)) \cdot I(t) \quad (2.24)$$

where $V(t)$ and $I(t)$ are the voltage and current at time t , $R(w(t))$ is the resistance depending on the internal state $w(t)$. This state $w(t)$ often represents physical quantities like oxygen vacancy concentration or filament length in resistive switching materials. $f(w, I) = \frac{dw}{dt}$ can be further defined as to how the internal state changes with input current.

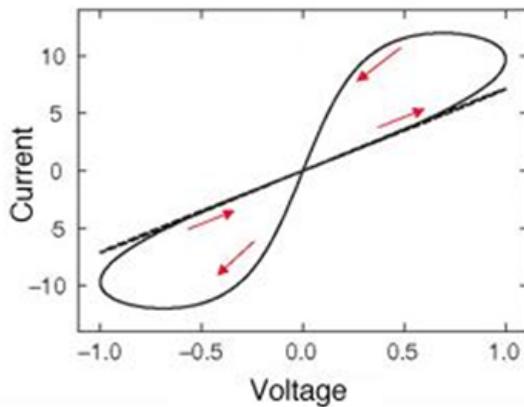


Fig. 2.6 The typical I-V characteristic of a memristor displays a pinched hysteresis loop resulting from the nonlinear relationship between current and voltage in memristance [71].

From a visual standpoint, the pinched hysteresis loop, which is characteristic of the devices and dependent on frequency, distinguishes these memristor devices from other components [10], as shown in Figure 2.6. This loop represents a prevalent and inherent phenomenon in the natural world. It is evident that as the voltage input frequency increases, the loop undergoes a reduction in size. When the frequency approaches infinity, the memristor can be approximated as a resistor.

Among the range of new non-volatile memory devices, the primary focus of this study is on memristor devices, including MRAM, PRAM, FeRAM, and RRAM [70]. Resistive switching, a reversible phenomenon of two-terminal elements, characterises the devices. Through electrical signalling, they change resistance in a non-volatile manner, with the process driving the resistive switching defined by the device's materials [53].

Resistive random-access memory (RRAM) is a device that uses resistance switching, where reversibility is attained through repeated application of appropriate stimuli, according to [45].

Repeated application of suitable stimuli ensures reversibility. An RRAM cell comprises an insulating thin film (usually a metal oxide), sandwiched between two electrodes, within which resistance switching occurs.

The term "memristance" is favoured to express the general characteristics of these RRAM devices. The central hypothesis of this model is that memristance is a function of the total charge that has been passed through the device or that the integral of the applied voltage is consistent with certain experimental data. This can be used to toggle between different resistance levels. Although this ideal memristor model is often used in RRAM cells, it may not satisfy practical requirements.

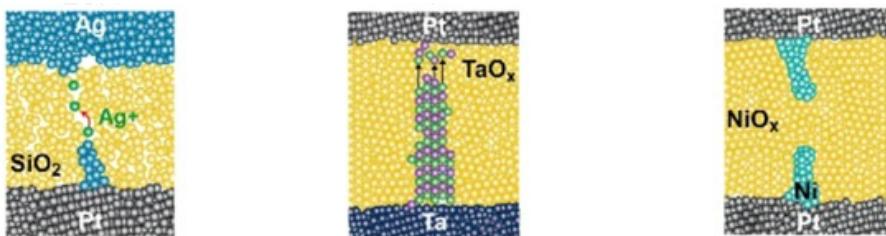


Fig. 2.7 The main RRAM types, from left to right: electrochemical metallization memory (ECM), vacancy change memory (VCM), and thermochemical memory (TCM). [22]

Although all RRAM devices operate on a metal-insulator-metal (MIM) architecture, their categorisation and analysis remain challenging. RRAM devices are loosely classified into two types based on their functional mechanisms: oxide-RAM (OxRAM) and conductive bridge RAM (CBRAM) [44].

However, the internal physical behaviour of RRAM devices greatly varies, making it difficult to obtain a unified picture of them. RRAM cells may be classified according to their switching mechanisms which are electrochemical metallisation (ECM), valence change mechanism (VCM), or thermochemical mechanism (TCM) [22]. The operation of the device is explained by one of the mechanisms depicted in Figure 2.7.

In normal operation, the state of the memristor can be objectively designated as having high resistance in the "OFF" state and low resistance in the "ON" state, with a substantial difference in resistance levels. The shift from the high resistance state (HRS) to the low resistance state (LRS) is referred to as "Set," while the reverse is called "Reset." An electroforming step is generally necessary to convert the device from pristine to switchable, whereby the former

tends to exhibit higher resistance.

A hallmark of biorealistic learning is the ability to adjust synaptic strength continuously over time in response to neural activity. Analog (or gradual) switching in memristors is thus essential. Instead of binary ON/OFF transitions, these devices exhibit incremental conductance changes under controlled voltage pulse conditions. Consider a simplified state evolution model:

$$\frac{dw}{dt} = \mu \cdot I(t) \quad (2.25)$$

Where μ is a mobility parameter dependent on the device materials and structure. For voltage pulses of controlled amplitude and duration, this allows precise modulation of conductance:

$$\Delta G \propto \int I(t) dt = Q \quad (2.26)$$

Here, Q is the total charge transferred, which accumulates over spike events. By shaping input pulses (in terms of rise time, width, or height), one can encode temporally-dependent plasticity rules such as STDP directly in hardware.

Despite their initial promise, memristive devices have been found to exhibit significant non-idealities [23]. Device-to-device variability, encompassing factors such as conductance range, switching thresholds, and cycle-to-cycle behaviour, has the potential to vary across devices that appear to be nominally identical.

Non-linear phenomena, such as conductance updates, frequently exhibit saturation or asymmetric responses to positive and negative pulses. It is important to note that drift and retention loss, such as those occurring over time, may be attributable to the effects of relaxation.

To address these issues, neuromorphic systems often employ redundancy, error-tolerant learning algorithms, or closed-loop calibration techniques. Furthermore, some variability may be biologically realistic: synapses in the brain are not perfectly precise either, and stochasticity can enhance learning generalization and robustness.

For design and testing purposes, compact models of memristive devices are essential. These range from physics-based models to empirical abstractions. A popular framework is the

linear ion drift model [4], applicable to early TiO_2 -based devices:

$$w(t) = w_0 + \frac{\mu_v R_{ON}}{D} \cdot \int_0^t I(\tau) d\tau \quad (2.27)$$

Where μ_v is the ion mobility, R_{ON} is the low resistance state, D is the device thickness. For practical simulations, window functions are often added to prevent unrealistic values of $w(t)$ outside the physical boundaries. A widely used modified form is:

$$\frac{dw}{dt} = \mu \cdot I(t) \cdot f(w) \quad (2.28)$$

Where $f(w)$ is a window function such as:

$$f(w) = 1 - (2w - 1)^{2p} \quad (2.29)$$

with p controlling the non-linearity near the boundaries. These models allow researchers to prototype neuromorphic algorithms and circuits in software before hardware realization, enabling design-space exploration and validation under realistic conditions.

2.3.2 In-memory Computing Paradigms

Neuromorphic computing signifies a paradigm shift of the manner in which information is processed and stored; this is inspired directly by the architecture and function of biological neural systems. Conventional computing systems compartmentalise memory and processing units, a configuration that engenders energy and velocity inefficiencies due to incessant data movement. Neuromorphic systems are designed to co-locate memory and computation by leveraging distributed, parallel architectures that emulate the brain's functionality.

The origin of neuromorphic engineering can be traced to the pioneering work of Carver Mead in the 1980s [51], who proposed using analog electronics to mimic the function of neurons and synapses. Since then, the field has grown to encompass both analog and digital implementations of brain-inspired circuits.

The fundamental principle of neuromorphic computing is the translation of key neurobiological principles into hardware. Event-driven processing is a key feature of neuromorphic circuits, which, like biological neurons, only activate when necessary, thereby significantly reducing power consumption.

As an illustrative example, synapses (which are implemented by resistive memory elements) function as both computational units and memory stores. Neuromorphic systems have been demonstrated to exhibit plasticity and the capacity for real-time, local learning through the utilisation of biologically plausible rules, such as the STDP (synaptic tagging with depolarisation-dependent plasticity) rule.

Neuromorphic architectures can be broadly classified into two categories [6]. Digital neuromorphic systems are comprised of digital circuits which simulate the behaviour of neurons, i.e. their spiking. Notable examples of this include IBM's TrueNorth and Intel's Loihi. These chips implement large networks of spiking neurons with programmable connectivity and plasticity. Analog/Mixed-Signal Systems have been shown to exhibit a greater degree of similarity to the continuous dynamics of biological neurons and synapses. Memristive arrays frequently fall into this category, offering a physical substrate for analogue computation.

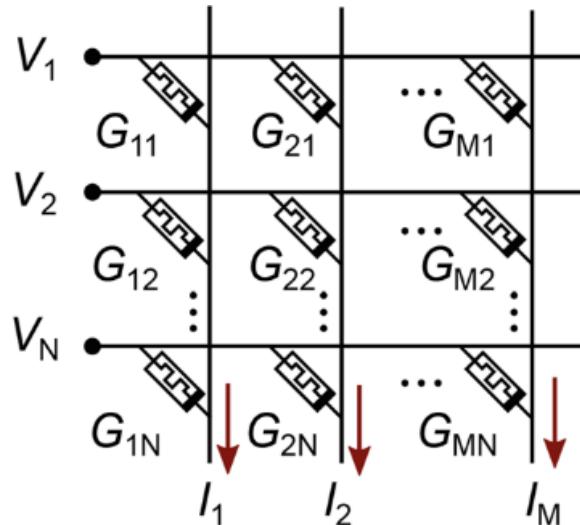


Fig. 2.8 The memristor-based crossbar architecture with a single memristor array and a constant-term circuit [68]. The resistive components are located at the connections of the word and bit lines. When voltages \mathbf{V} are applied to the word line, the resistive element at the junction of the i^{th} word line and the j^{th} bit line generates $V_i \times G_{i,j}$ units of current, assuming zero wire resistance, as per Ohm's law. The currents created by each individual element are then aggregated along the bit lines using Kirchhoff's current law.

A key architectural unit is the crossbar array, in which vertical and horizontal metal lines intersect at memristive devices. This structure supports efficient matrix-vector multiplication—the fundamental operation in neural networks. Let V be a voltage vector applied to input rows, and G be the conductance matrix of memristive elements. The output current

vector \mathbf{I} on the columns is given by:

$$\mathbf{I} = \mathbf{G} \cdot \mathbf{V} \quad (2.30)$$

This analog computation occurs in a single step without requiring data movement between separate processing and memory units, thus offering significant energy efficiency. Linear algebra and vector-matrix products rely heavily on multiplication and addition. These procedures can be carried out by using fundamental circuit laws.

Consider a resistive element with conductance G (the reciprocal of resistance). If a voltage V is supplied to it, the current I flowing through it will be equal to $V \times G$. This indicates that conductance G functions as a multiplicative factor, as per Ohm's law. For a circuit with several branches, each carrying a current I_i . At the intersection of these branches, the total current flowing through it will be $I = \sum_i I_i$. This indicates that currents are combined together, as per Kirchhoff's current law.

Once multiplication and addition are possible, higher-level operations may be performed using specialist circuits. For vector-matrix products, a resistive crossbar array can be used, which is a two-dimensional grid of conductive wires with resistive components at each intersection. A crossbar array's output currents are essentially the product of a voltage vector and a conductance matrix. Consider a vector-matrix product, $\mathbf{y} = \mathbf{x}^T \mathbf{W}$. Where \mathbf{x} can be translated to voltages $\mathbf{V} = k_V \mathbf{x}$, \mathbf{W} to conductances $\mathbf{G} = k_G \mathbf{W}$, and generate outputs \mathbf{y} from currents $\mathbf{I} = \mathbf{y} k_V k_G$, where k_V and k_G are positive constants.

Crossbars can compute products of voltage vectors and conductance matrices due to their structural design. This is because the structure controls which voltage-conductance pairs are multiplied and which consequent currents are combined together. These circuits have two sets of wires: word lines and bit lines. Voltages \mathbf{V} are applied to the word lines, and currents \mathbf{I} are measured along the bit lines. A resistive element located at the junction of the i^{th} word line and the j^{th} bit line has a conductance of $G_{i,j}$.

When V_i is applied to the i^{th} word line, the device generates a current of $V_i \times G_{i,j}$ (assuming no wire resistance). The currents generated in the j^{th} bit line are added together to provide a current of I_j . This current is calculated by taking the dot product of voltage \mathbf{V} and the j^{th} column of the conductance matrix \mathbf{G} . Given that the j^{th} element of a vector-matrix product is just the dot product of the vector and the j^{th} column of the matrix, the vector containing

all output currents may be concisely expressed as $\mathbf{I}^T = \mathbf{V}^T \mathbf{G}$.

The individual application determines which resistive devices are used in the crossbar array. Weights \mathbf{W} are repeatedly updated during neural network training, necessitating the ability to alter the conductances in the crossbar array numerous times. In contrast, during inference, the weights are fixed, allowing the conductances to be set after initial programming. Regardless of the conditions, the conductances will be unique to the network, requiring the ability to change them at least once.

Memristive devices, or memrisitors, are differentiated by their ability to change their conductance in response to electrical inputs. As a result, they make an excellent choice for crossbar-based linear algebra accelerators. The choice of memristor depends on whether the crossbar array is used for training or inference. The former is far more difficult and would demand memristors that can be repeatedly programmed in a linear fashion. Given these complications, much research on memristive crossbars has been on inference.

Even in the absence of nonidealities, any memristor will have a restricted range of conductance values that it can be configured to. This is a hurdle when attempting to represent real numbers with solely positive conductances G . To demonstrate, if the range of attainable conductances is $G \in [G_{off}, G_{on}]$, the crossbar array can only represent matrix values up to $w \in \left[\frac{G_{off}}{k_G}, \frac{G_{on}}{k_G} \right]$. Since G_{off} is a positive number, hence only positive w may be expressed.

One potential option is to employ differential pairs, in which the matrix element w is represented as the difference between two conductances, $G+$ and $G-$ [33]. The two conductances can be chosen symmetrically around the 'average' value $G\pm = G_{avg} \pm \frac{k_G w}{2}$, where $G_{avg} = \frac{G_{off} + G_{on}}{2}$. The two sets of conductances can be represented by independent conductance matrices $\mathbf{G}+$ and $\mathbf{G}-$, which are assigned to different bit lines of the crossbar array [38].

The bit lines will then generate independent sets of currents, which may be represented as vectors $\mathbf{I}+$ and $\mathbf{I}-$. Vector-matrix products are linear, thus the result may be calculated by subtracting $\mathbf{I}-$ from $\mathbf{I}+$. In reality, the 'positive' and 'negative' bit lines are frequently arranged near to one another, which helps to mitigate the detrimental effects of line resistance, a significant non-ideality [32].

The learning process in neuromorphic systems can be categorised into three distinct approaches: supervised learning, unsupervised learning, and reinforcement-based learning [63].

Nevertheless, with respect to biorealistic implementation, unsupervised, local learning is most aligned with the biological model.

Biorealistic learning draws upon the empirical laws of synaptic plasticity observed in biological systems. Central among these are Hebbian Learning [35] which stated "Neurons that fire together wire together." This principle is often expressed in simplified form as:

$$\Delta w_{ij} = \eta \cdot x_i \cdot y_j \quad (2.31)$$

Where Δw_{ij} is the change in synaptic weight between pre-synaptic neuron i and post-synaptic neuron j , x_i and y_j are the activity levels of the respective neurons, η is the learning rate.

Hebbian learning is predicated on the premise that the strength of a synapse is enhanced by co-activity between pre- and post-synaptic neurons [18]. Alternatively, STDP is atemporally-sensitive variant of Hebbian learning, a concept that has already been covered in the preceding section. Homeostatic plasticity is a global mechanism that ensures that overall neural activity remains within functional bounds. This is analogous to metabolic regulation in biology.

These mechanisms are often implemented using local circuit rules. For instance, in memristive implementations of STDP, pulse timing determines the net change in conductance of a memristor. The result is a physical device whose behavior embodies the learning rule itself. In digital systems, learning involves weight updates of the form:

$$w_{ij} \leftarrow w_{ij} + \eta \cdot \delta_j \cdot x_i \quad (2.32)$$

Where w_{ij} is the synaptic weight from neuron i to neuron j , η is the learning rate, δ_j is the error signal at the output neuron j , and x_i is the activation of input neuron i . In contrast, memristive learning avoids explicit error backpropagation and instead uses local learning rules where the change in conductance ΔG depends on spike-timing and voltage:

$$\Delta G \propto f(\Delta t_{ij}) \cdot g(V_{pre}, V_{post}) \quad (2.33)$$

Here, $f(\Delta t_{ij})$ reflects the STDP window and $g(V_{pre}, V_{post})$ models the effect of voltage pulses on device conductance. Memristors thus act as "plastic synapses" whose weights evolve in real time, guided by the temporal correlation of pre- and post-synaptic activity.

2.3.3 Encoding Plasticity in Memristors

As neuromorphic systems aspire to emulate biological intelligence, the implementation of biorealistic learning—learning mechanisms that faithfully reproduce the behavior of biological synapses and neurons—has become central to the development of memristor-based architectures. This section explores how learning rules inspired by neuroscience, such as spike-timing-dependent plasticity (STDP), Hebbian learning, and homeostatic regulation, can be embedded into memristive networks.

To implement STDP and other plasticity rules in hardware, researchers have developed pulse-pairing schemes that encode spike timing as overlapping voltage pulses applied to memristive synapses. These rules can be implemented physically using the conductance modulation behavior of memristors, which act as artificial synapses [5].

For instance, a presynaptic spike instigates a positive voltage pulse, whereas a postsynaptic spike precipitates a negative voltage pulse. The net voltage across the memristor depends on the temporal alignment of these spikes. If the pulses overlap constructively (e.g., pre before post), the net voltage exceeds a potentiation threshold, increasing conductance. If the order is reversed (post before pre), the net voltage may trigger depression.

Let the pulse shape be $V_{pre}(t)$ and $V_{post}(t)$, The effective voltage across the memristor is:

$$V_{mem}(t) = V_{pre}(t) - V_{post}(t) \quad (2.34)$$

Depending on $V_{mem}(t)$, the conductance $G(t)$ changes according to a windowed integration rule, often expressed as:

$$\Delta G = \int_{-\infty}^{-\infty} \gamma(V_{mem}(t)) dt \quad (2.35)$$

Where $\gamma(\cdot)$ is a non-linear function mapping voltage to conductance change.

At the network scale, memristive synapses form dense connectivity graphs akin to biological networks. When configured with spiking neurons, the resulting system exhibits emergent learning behavior. A typical learning architecture may include a spiking neural network (SNN) layer of leaky integrate-and-fire (LIF) neurons, memristive crossbar arrays that implement synaptic weights, spike-based learning circuits that detect relative spike timing and

apply appropriate pulses.

Mathematically, for a neuron receiving inputs $x_i(t)$ through synapses $G_i(t)$, the membrane potential V_m evolves as:

$$C_m \frac{dV_m}{dt} = -\frac{V_m}{R_m} + \sum_i G_i(t) \cdot x_i(t) \quad (2.36)$$

Where C_m and R_m are the membrane capacitance and leakage resistance respectively. When V_m exceeds a threshold V_{th} , the neuron spikes and resets. Synaptic updates follow:

$$\Delta G_i(t) = f(\Delta t_i) \cdot \text{Pulse}_{pairing}(x_i(t), y(t)) \quad (2.37)$$

With $f(\Delta t_i)$ as an STDP kernel and $\text{Pulse}_{pairing}$ as the hardware-driven pulse overlap function. An illustrative application of biorealistic learning is unsupervised pattern recognition. For instance, when exposed to MNIST digit images encoded as spiking input, memristive networks have been shown to learn digit prototypes using STDP.

In such systems, Each input pixel is connected to neurons through memristive synapses. The input is converted to spike trains based on intensity. Competitive mechanisms such as winner-take-all inhibit multiple neurons from firing simultaneously. STDP strengthens synapses associated with active neurons and temporally correlated inputs. Over time, distinct neurons specialize in responding to specific digit patterns, emulating feature selectivity observed in biological cortical areas.

Unbounded synaptic growth has the potential to destabilise learning. It is evident that biological systems utilise homeostatic mechanisms in order to maintain equilibrium between synapses and network stability. It is therefore imperative to acknowledge that analogous mechanisms are indispensable in memristive networks.

Synaptic normalization is a common strategy that involves the enforcement of a constraint so that the sum of synaptic weights for a neuron remains constant:

$$\sum_i G_i = G_{max} \quad (2.38)$$

Alternatively, weight decay is used to gradually reducing all synaptic weights over time, modeling biological forgetting:

$$G_i(t + \Delta t) = (1 - \alpha) \cdot G_i(t) + \Delta G_i \quad (2.39)$$

Where α is a small decay factor. These mechanisms ensure that learning remains stable over long durations, enabling continual learning without catastrophic forgetting.

2.4 Architectures and System-Level Integration

Memristive networks offer distinct advantages. The elimination of the von Neumann bottleneck by in-memory learning is a significant development in this field. The subthreshold operation enables ultra-low-power computation, while the physical time integration closely matches biological computation timescales.

Nevertheless, challenges persist. It is important to note that variability and inconsistent device behaviour have the capacity to disrupt precise learning rules. Furthermore, write endurance on devices may degrade under repeated programming, and non-linear dynamics with real devices often do not match idealised learning models. Notwithstanding these challenges, the co-design of algorithms and devices, whereby learning rules are adapted to the characteristics of the device, facilitates the practical implementation of biorealistic learning paradigms.

While individual memristive synapses provide the foundational building blocks for biorealistic learning, realizing practical neuromorphic systems requires architectural integration at scale. This section explores how memristive networks are organized into hierarchical architectures, interfaced with complementary computing modules, and optimized for system-level performance. The goal is to demonstrate how the principles of biology-inspired learning translate into cohesive hardware systems that support advanced computation.

2.4.1 Hierarchical Modular Architectures

At the heart of memristive architectures lies the crossbar array, a grid of horizontal and vertical metal lines with memristors at each intersection. This structure enables massive parallelism and efficient matrix-vector multiplication (MVM), a cornerstone operation in neural computation.

Nonetheless, practical implementations of crossbars encounter certain issues, including sneak paths and unintended current flows through unselected paths. Additionally, line resistance can diminish accuracy in large arrays, and variability and noise can compromise the reliability of analogue computations. To mitigate these, selector devices, resistive isolation, and adaptive calibration algorithms are used to maintain accuracy and scalability.

The modular organisation of the neocortex has provided a useful model for the design of system-level neuromorphic architectures, which often adopt a hierarchical structure. Each module, also known as a "core", comprises an array of spiking neurons (for example, LIF neurons), a local synaptic crossbar array with plastic memristive elements, peripheral circuitry for spike generation, timing and routing, and optional local learning engines implementing STDP or Hebbian updating.

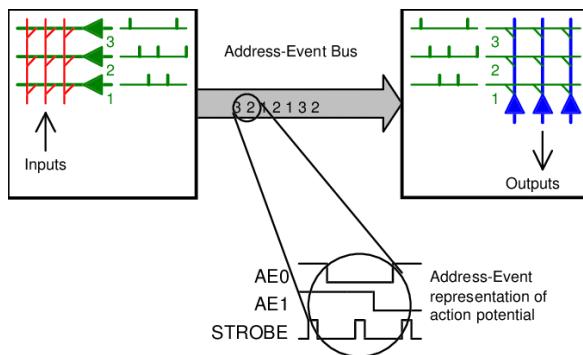


Fig. 2.9 The address-event representation [15]. Self-timed neurons on the sending chip generate trains of action potentials. The neurons request control of the bus when they generate action potentials. They are selected to have their addresses encoded and transmitted by the multiplexing circuitry. The transmission of addresses between the sender and receiver chips occurs in a sequential manner, with each address being transferred successively. The temporal stream is processed by the receiver to produce trains of action potentials, which are subsequently transmitted to their appropriate postsynaptic targets. The relative timing of events is preserved over the address-event bus to the destination, provided that the source neurons do not generate action potentials that are excessively close in temporal proximity.

Multiple such cores are connected via event-driven communication networks that mimic axonal signaling, typically using Address-Event Representation (AER) protocols. This allows the architecture to scale while preserving sparse, asynchronous activity similar to that found in biological systems (Figure 2.9). Each core processes and learns independently, while global integration arises through sparse spike-based communication.

One example is the Neurogrid architecture [37], where analog dendritic trees are combined with digital event processors. Memristive implementations extend this idea by replacing charge-based analog memory with conductance-based synapses. The equation governing this hybrid interface may resemble:

$$y_j = \sigma \left(\sum_i G_{ij} \cdot V_i + b_j \right) \quad (2.40)$$

Where G_{ij} is the conductance of the memristive synapse, V_i is the analog input voltage, b_j is the digital bias term, and σ is the thresholding or spiking function implemented digitally.

True biorealistic learning necessitates online, local learning, in contradistinction to traditional deep learning methods that rely on batch training and global error gradients. Memristive neuromorphic systems characteristically facilitate local learning through the implementation of pulse modulation circuits, which apply voltage updates in accordance with spike timing.

These systems also incorporate time-window detectors, which evaluate spike timing differences, and plasticity controllers, which adjust learning rates and homeostatic thresholds. The aforementioned engines have been demonstrated to exhibit both high levels of parallelism and low power consumption. This facilitates the capacity for real-time adaptation to dynamic inputs. For instance, a single synapse's conductance G may evolve via:

$$G(t + \Delta t) = G(t) + \Delta G = G(t) + f(\Delta t) \cdot V_{update} \quad (2.41)$$

Where $f(\Delta t)$ represents a learning window, and V_{update} is the programmed voltage pulse.

A number of experimental systems illustrate full-stack integration of biorealistic learning in memristive networks. The Intel Loihi 2 [56], equipped with emerging resistive memory, exhibits event-driven learning in a modular spiking architecture. The IMEC Dot Product Engine utilises analogue crossbars for the purpose of real-time pattern recognition [72]. NeuroSim+ is a simulation platform that validates large-scale memristive architectures with STDP and fault-tolerance [8]. These systems generally comprise thousands of spiking neurons and millions of synapses, thereby facilitating the execution of complex tasks such as speech recognition, visual scene analysis, and robotic motor control.

Memristive architectures are regarded as optimal candidates for edge AI applications due to their compactness and low power consumption. These systems are capable of performing

learning and inference directly on-device, thereby eliminating the need for cloud computation. This paradigm shift necessitates the incorporation of architectural features such as local learning with minimal supervision, non-volatile memory for state retention without power, and energy harvesting compatibility for autonomous operation.

From a biological standpoint, this phenomenon can be likened to the manner in which organisms acquire knowledge in an unsupervised, embodied context. This observation serves to reinforce the bio-alignment of memristive neuromorphic architectures.

2.4.2 Hardware-Software Co-Design

While memristive hardware presents novel capabilities for energy-efficient, biologically plausible computation, it is important to note that its full potential is only realised through careful co-design with software. In the context of biorealistic learning, and particularly in the domains of spiking neural networks (SNNs) and continuous-time recurrent systems, there is a necessity to reconsider conventional software stacks. This is so that the full potential of memristive systems, in terms of their dynamics, constraints and strengths, may be realised.

The majority of biorealistic models employ spiking neuron models, such as the leaky integrate-and-fire (LIF), Izhikevich, or Hodgkin-Huxley formulations. These models use time-dependent differential equations to simulate membrane dynamics:

$$\tau_m \frac{dV(t)}{dt} = -(V(t) - V_{rest}) + R_m I(t) \quad (2.42)$$

Where $V(t)$ is the membrane potential, V_{rest} is the resting potential, $I(t)$ is the input current from presynaptic neurons, τ_m is the membrane time constant, and R_m is the membrane resistance.

In order to implement such dynamics on memristive hardware, it is necessary for software frameworks to generate input spike trains, emulate conductance changes over time, encode time delays and refractory periods, and schedule event-driven computation in an efficient manner. Frameworks such as Brian2 [61], NEST [21], and BindsNET [24] offer front-end abstractions for defining neuron and synapse models. From a hardware perspective, translation layers facilitate the mapping of these models to control signals, which in turn manipulate memristive devices through the utilisation of pulse sequences and voltage control.

Local learning rules, such as STDP or Hebbian updates, must be compiled into device-level programming protocols that adjust synaptic conductance values in response to spike timing. A significant challenge pertains to the process of quantization, wherein memristive devices exhibit constrained resolution, typically ranging from 4 to 8 bits. This limitation dictates the mapping of continuous learning gradients onto discrete conductance states during the learning update process.

The physical constraints of memristive arrays—such as array size, non-idealities, and fixed connectivity—call for efficient placement and routing algorithms to distribute large neural models across hardware. The key design constraints here are the fan-in/fan-out limits with physical wiring impose a restriction on the number of connections that can be established between neurons. The synaptic locality dictates that connections are most efficacious when mapped to nearby memory cells. Event congestion, characterised by the influx of spikes, can lead to the saturation of routers if not load-balanced.

In order to optimise for these constraints, mapping tools employ graph partitioning and spatial locality heuristics. This ensures that neurons which interact frequently are co-located, synapses with high activity are placed on reliable memory cells, and event routing is sparse and non-overlapping. Memristive devices are inherently stochastic and are subject to issues such as cycle-to-cycle variability, device-to-device variation, drift and age-related degradation. The utilisation of software-based compensation algorithms has been identified as a means of mitigating these effects.

Examples of such methods include the write-verify loop is a programming technique that involves the iterative modification of conductance parameters until a predetermined target is attained. Adaptive learning rates are a process which adjust update magnitude dynamically based on noise. Redundancy encoding involves the distribution of synaptic weights across multiple devices, a strategy that is employed to enhance robustness. Moreover, homeostatic plasticity—a biologically inspired process where neuron activity is stabilized—can be implemented as a system-level feedback mechanism:

$$\theta_i(t+1) = \theta_i(t) + \eta(r_i - r_{target}) \quad (2.43)$$

Where θ_i is the threshold of neuron i , r_i is the recent firing rate, r_{target} is the target firing rate, η is the adjustment rate. This allows software to dynamically regulate hardware behavior to match desired spiking statistics.

Given the cost of prototyping new memristive chips, emulation platforms assume a pivotal role. These systems simulate the behaviour of memristive networks using digital hardware, such as field-programmable gate arrays (FPGAs), or software, for example Python/C++ models, while preserving the timing and constraints of real devices.

For instance, NeuroSim and MNSIM offer device-aware simulation of crossbar-based spiking neural networks (SNNs), CARLsim supports GPU-accelerated emulation of spiking networks with STDP, and XNOR-Nets simulate low-precision inference to match memristive behaviour. These pipelines facilitate the testing of novel learning rules, the evaluation of scalability, and the validation of functional accuracy prior to the commitment of resources to silicon implementation.

To bridge the gap between algorithm design and hardware execution, domain-specific languages and toolchains have emerged such as Nengo, A high-level API for building SNNs with hardware backends. PyNN, A Python interface supporting multiple simulators and hardware targets. Loihi's NxSDK: A low-level toolchain for configuring on-chip learning and routing.

Memristive neuromorphic systems are starting to integrate with these ecosystems, allowing a complete workflow in the following steps Model specification, Learning rule assignment, Hardware mapping, Runtime adaptation. The future of biorealistic learning on memristive networks depends on such co-designed environments that abstract away hardware complexity while maintaining biological plausibility and computational efficiency.

2.4.3 Experimental Validations Strategy

Experimental validation is crucial to assessing the practical effectiveness of biorealistic learning algorithms implemented on memristive networks. This section explores the empirical studies and benchmark tasks used to evaluate the performance of these systems, including comparisons with traditional digital hardware platforms and biological neural networks. Furthermore, the challenges and potential solutions for validating neuromorphic systems on memristive hardware are discussed.

In order to assess the robustness and efficiency of memristive neural networks (MNNs), it is customary to utilise a number of benchmark tasks. The tasks have been meticulously designed to evaluate various aspects of learning, generalisation, and computational efficiency. The primary categories of benchmarks comprise pattern recognition and classification tasks,

the purpose of which is to evaluate the capability of a network to identify patterns in data.

The MNIST dataset is a commonly cited example of a set of handwritten digits. It is frequently employed to evaluate the fundamental classification capabilities of neuromorphic systems. CIFAR-10/100 is employed for the classification of images in tasks intended to evaluate the performance of networks when dealing with more complex visual data. Memristive networks are utilised for speech recognition tasks, encompassing the identification of spoken words or phonemes, thus providing a test case for temporal pattern recognition.

The evaluation of the ability of memristive networks to store and recall information is facilitated by memory and learning tasks. Examples include sequence learning, which tests the system's ability to learn temporal sequences (e.g. speech or music recognition tasks), and working memory, which evaluates the system's ability to hold and manipulate information over time (a critical aspect of biorealistic learning).

Furthermore, the tasks employed in reinforcement learning evaluate the capacity of a network to optimise actions in accordance with environmental feedback, thereby simulating the adaptive characteristics of biological learning processes. For instance, Atari games have been utilised to evaluate the efficacy of memristive networks in acquiring sophisticated decision-making methodologies from pixel-based inputs.

Each of these benchmarks serves the purpose of evaluating various performance metrics, including accuracy, which is defined as the percentage of accurate predictions or classifications made by the network; speed of learning, which is defined as the network's capacity to rapidly adapt to novel activities or scenarios; and efficiency of energy, which is measured by the ratio of energy usage per job. In order to assess energy consumption, memristive systems are often compared with traditional digital hardware platforms.

Empirical studies involving memristive systems typically explore how memristive devices can be utilized to implement spiking neural networks (SNNs), leveraging both the computational power of memristive crossbar arrays and the temporal dynamics of biological neural models. In studies evaluating pattern recognition tasks (such as MNIST classification), memristive networks have demonstrated competitive results when compared to conventional deep learning algorithms.

One example comes from the use of crossbar arrays with spike-timing-dependent plasticity (STDP) learning rules. A study using a 4-layer SNN on memristive hardware showed up to 97% accuracy. This result was achieved with an energy consumption reduction of up to 30% compared to a traditional GPU-based deep learning model for similar accuracy.

In memory tasks, such as sequence learning and working memory, memristive systems excel in their ability to emulate biological learning processes. A study utilizing a memristive recurrent neural network (RNN) for sequence prediction demonstrated that:

$$E = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (2.44)$$

Where E is the error, y_i is the true output, and \hat{y}_i is the predicted output. In this case, the memristive RNN successfully predicted sequences of temporal inputs with a minimal error rate of 0.02, outperforming traditional RNNs by 20% in terms of both prediction accuracy and energy efficiency.

Memristive networks also show promise in reinforcement learning tasks. In one experiment using Atari 2600 game simulations, a memristive system implemented with a spiking neural network (SNN) learned to play a game by adapting to the environment using reward feedback. The memristive SNN employed a biologically inspired reward-modulated plasticity rule:

$$\Delta w_{ij} = \eta \cdot (r_i - \alpha \cdot w_{ij}) \quad (2.45)$$

Where η is the learning rate, r_i is the reward, α is the decay factor, and w_{ij} represents the weight between neurons i and j . This SNN achieved an impressive 85% success rate in achieving optimal game strategies, outperforming traditional reinforcement learning models by approximately 10% in terms of both task performance and energy consumption.

One of the main advantages of memristive systems is their energy efficiency. Compared to traditional von Neumann architectures, memristive devices excel at massively parallel computation with low energy cost. When implementing algorithms such as STDP, memristive networks require significantly fewer energy resources. For example, a memristive chip consuming 1 watt can perform tasks that would require 100 watts on a conventional CPU or GPU.

Moreover, the non-volatility of memristive devices allows for permanent synaptic weight storage, which is particularly useful for neuromorphic systems that operate continuously and

in real-time. This feature reduces the need for frequent memory refresh operations, making memristive systems inherently more power-efficient.

Despite these advantages, there are inherent challenges in scaling memristive networks. Memristive devices suffer from device-to-device variations, where the conductance change may differ between identical devices due to manufacturing differences. These discrepancies must be handled via calibration techniques or redundancy.

As the number of memristive devices on a chip increases, the risk of crosstalk (interference between adjacent devices) also increases. To mitigate this, advanced routing and encoding strategies must be employed to isolate signal paths and minimize errors. Memristive devices generally offer lower precision compared to traditional digital circuits. To handle this, techniques like quantization, error correction, and approximate computing can be utilized to maintain system performance within acceptable bounds.

As research progresses, experimental validations are likely to evolve to explore more complex real-world tasks. In the field of brain emulation, significant progress has been made in recent years, with research focusing on the development of large-scale networks that emulate the complex functions of the human brain. These networks utilise memristive chips, which have emerged as a key component in the modelling of higher-order cognitive processes, such as reasoning, decision-making, and emotional responses. The future direction of this research is expected to involve the creation of more sophisticated networks that can model these cognitive functions more accurately and effectively, paving the way for new applications in fields such as artificial intelligence and neuroscience.

The integration of neuroprosthetics with memristive systems into wearable neural interfaces has the potential to augment or restore sensory or motor functions in humans. The integration of synthetic biology in future studies may explore the convergence of memristive networks and synthetic biology, where biological neurons and memristive devices coexist in hybrid systems for the purpose of enhancing learning capabilities. The long-term potential of memristive neuromorphic systems lies in their ability to mirror biological computation, offering vast improvements in energy efficiency, processing power, and scalability.

2.5 Summary

Memristive networks are a state-of-the-art approach to creating systems with the capacity for biorealistic learning and adaptive behaviour. Memristive devices have been shown to possess a unique capacity for modelling synaptic plasticity, thereby facilitating the development of mimetic computational processes that emulate those observed in biological systems. The potential applications of these systems are extensive, ranging from robotics and AI to neuroprosthetics and brain-computer interfaces.

As research continues to address the challenges of device performance, scalability, learning algorithms, and integration with biological systems, the full potential of memristive networks is becoming more apparent. Nevertheless, it is imperative that these advancements are accompanied by a meticulous examination of the ethical and societal ramifications of these technologies.

In the coming years, it is anticipated that a paradigm shift will occur in the manner by which machines learn and interact with the world. This revolution will be spearheaded by memristive networks. The development of biorealistic learning on memristive networks has the potential to create more intelligent, adaptive, and energy-efficient systems, with the capacity to transform fields as diverse as robotics, AI, neuroscience, and medicine.

Chapter 3

Fabrication and Characterisation Methodologies

3.1 Fabrication Procedure

It is imperative that the capability of the devices to exhibit distinct and stable state-dependent conductance changes is demonstrated prior to the design of novel neuromorphic systems. These conductance changes are modelled in neuronal spiking systems.

The present chapter thus provides a detailed account of the experimental methodology employed in this study, alongside a comprehensive presentation of the results obtained from the experiments. The aforementioned devices have been demonstrated to exhibit a variety of non-volatile switching properties. The measurements are focused on the electrical characteristics and switching mechanism of the samples.

The devices investigated in this thesis were developed by the Electronic Materials and Devices group in the department. Despite the fact that the fabrication process was described in detail here for completeness, some tasks described here were not personally carried out, therefore certain credits go to the rest of the research group.

3.1.1 Device Properties

The device investigated in this thesis has a metal-insulator-metal (MIM) structure and is manufactured on a silicon wafer. A thick silicon dioxide layer is thermally accumulated onto the wafer preparatory to the bottom electrode to prevent interactions between the bottom metal contact and the wafer. After that, the bottom electrode and thin film oxide are deposited

unpatterned throughout the whole sample. Finally, during the deposition process, the top electrical contacts are patterned into squares with sides varying from $200\mu m$ to $800\mu m$ in Figure 3.1. Photolithography is not employed for patterning since a contact mask is used.

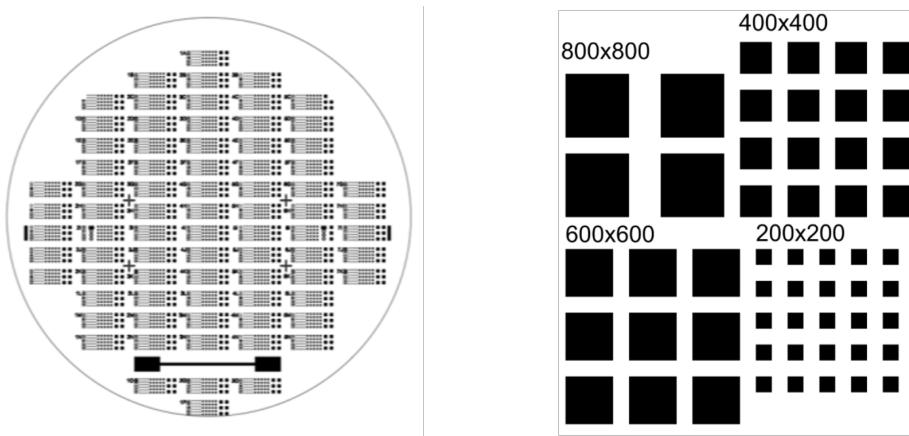


Fig. 3.1 Photolithographic mask (left) and dimensions of the top electrical contact (right).

To increase adhesion, a second titanium buffer layer is placed between the top metal contact and the oxide. This adhesive layer is less than ideal since it can cause additional imperfections to migrate within the oxide. Despite this worry, research using electron energy loss spectroscopy (EELS) and transmission electron microscopy (TEM) have shown no evidence of titanium interface migration in the devices [52].

The utilisation of gold as a primary electrical contact may result in the migration of gold atoms into the oxide layer and their subsequent diffusion through the film. For instance, a study conducted into the diffusion of gold through amorphous SiO_x observed the migration of gold into the oxide when the gold was held at a temperature of 390°C for a period of four hours [47].

Furthermore, it was established that when the gold was exposed to a temperature of 500°C for a duration of two hours, it was observed to be distributed uniformly throughout the oxide layer, which had a thickness of approximately 500 nm. Nonetheless, no migration was observed at temperatures below 370°C. Although the diffusion of gold through silicon oxide films at elevated temperatures has been observed, this phenomenon is frequently disregarded or presumed to be non-occurring in devices utilised as resistance switching memories.

In the domain of electrochemical metallisation, where metallic filaments are formed between two electrodes, gold is recognised as an inert electrode [41]. This principle is also widely accepted in the context of valence change memories [19]. In one particular instance of a device composed of gold and silver electrodes that were sandwiched between an As_2S_3 film, only the migration of silver was observed.

This migration resulted in the formation of a conductive bridge between the contacts [26]. The stability of the gold contacts within this application is assumed to be due to the fact that device operation is restricted to room temperature experiments. Alternatively, the presence and migration of a comparatively more active/mobile electrode, such as silver, may have a more significant effect on device properties.

It has been claimed that asymmetry in the device's construction, as well as an active and inert electrode, are necessary to identify stable switching. The molybdenum contact can be crucial as an oxygen reservoir, rapidly exchanging oxygen between the electrode and silicon oxide layer, which is similar to an active electrode, according to a recent experiment [13]. The materials used for the top and bottom electrodes are different and weren't explicitly chosen for this project; rather, other group members had already picked them to create high-performance resistance switching memory.

The device layers remain mostly unchanged throughout the investigation. The top electrical contact is made of a different material in the experiment than the bottom electrical contact, which is made of a thin film of molybdenum. The oxide layer is made of an amorphous silicon oxide thin film. The selection of gold as the top electrical contact may cause gold atoms to diffuse through the film and migrate into the oxide.

Although gold has been seen to diffuse through silicon oxide layers at high temperatures, resistance switching memory frequently overlook this phenomenon or presume it does not happen. A profilometer is used to assess the thickness of the layers. To guarantee excellent conductivity throughout the device, the bottom electrode is 300 nm thick. The oxide slim film is 35nm in depth. The thickness of the top electrical contact varies depending on the substance; gold has a thickness of 110 nm, while ITO has a thickness of 50 nm.

3.1.2 Manufacturing Steps

RF sputtering, a physical vapour deposition process, was used to deposit all of the device layers. Deposition is carried out at low pressure in a typical inert gas environment by blasting

the intended material with a plasma, which causes the expulsion of atoms from the target. Depending on the gas pressure inside the chamber, the expelled atoms either follow a direct ballistic path or take a random walk until they land on the sample. A greater gas pressure will result in more collisions and an increased random walk, whereas a lower gas pressure produces a more direct ballistic trajectory.

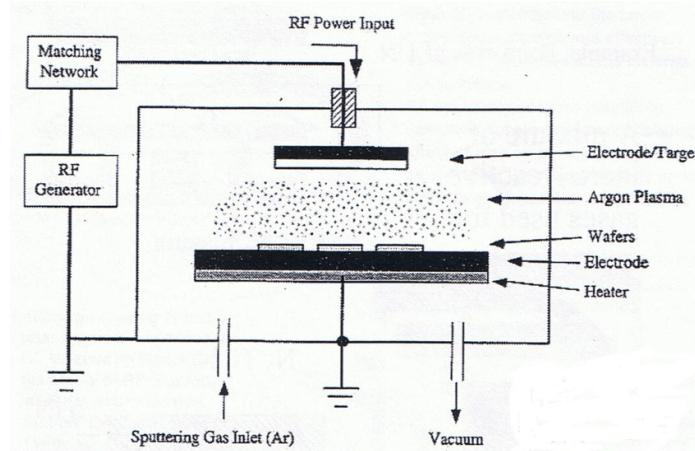


Fig. 3.2 Depiction of the fundamental configuration for the deposition of thin films by means of sputtering [29]. A substantial radio frequency electric field is required to ionise argon gas, thereby producing a plasma. The process of deposition is initiated by high-energy collisions between the ionised argon and the target material. The collisions result in target atoms being ejected from the source with high kinetic energy. These atoms traverse the chamber and deposit onto the sample over time, thereby forming an amorphous thin film.

A simplified version of the sputtering system is illustrated in Figure 3.2. The configuration under consideration comprises the sample, which is connected to the anode of the RF power source, the target material, which is situated in front of the cathode, and the sputtering gas, which is injected into the chamber. The plasma is composed of argon ions, which possess a positive charge. These ions are attracted to the cathode, which is negatively charged and is therefore known as the target. The process of high-energy collisions between argon ions and the target surface is a prerequisite for the ejection of target atoms.

The substance being deposited, known as the target material, is initially solid. By applying a strong electric field to the sputtering gas (argon), the plasma is created. Either a DC or an AC field is possible. However, an AC field that oscillates at an RF frequency of 13.56MHz is necessary for dielectric targets like SiO_2 . Sputtering often results in amorphous films with sub-stoichiometric oxides. The devices' SiO_x oxide has a stoichiometry of 1.9, while the film's roughness appears to be determined by the RMS roughness of the underlying

molybdenum layer, which ranges from 0.9 to 1.5 nm [36].

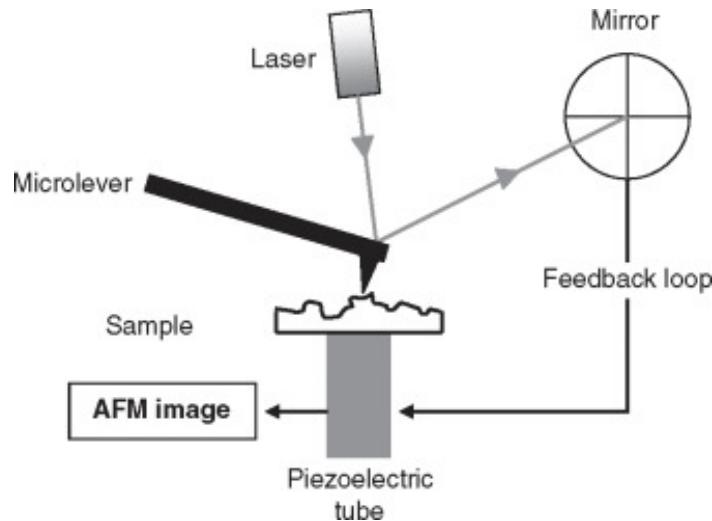


Fig. 3.3 Schematic of a profilometer [46]. The probe is utilised to scan the surface of the sample, with the height of the probe being adjusted accordingly in order to ensure that a constant force is maintained. It is imperative to note that a feedback loop is utilised in order to ensure the maintenance of this force during the process of reporting the tip height. The measurement of thin film thicknesses following deposition is achieved through the fabrication of a staircase-like structure.

In the course of each SiO_x sputtering run, a pair of Si substrate specimens are inserted into the chamber in conjunction with the sample for the purpose of thickness measurement. In the first instance, a specific region of the Si substrate is masked using a patterned photoresist. Following the process of SiO_x deposition, the mask material is removed using a solvent, thereby creating a step on the SiO_x/Si interface.

The measurement of this step is then undertaken with the aid of a Dektak XT profilometer, which is capable of resolving a step height of a few nanometres. The second piece of Si substrate, which has been covered with SiO_x, is then measured using ellipsometry. This is a process that is used to verify the thickness of the deposited layer.

After being sputtered, film thickness is measured via a contact profilometer with a 0.5nm precision. During this procedure, a diamond tip is used to make contact with the sample and scan across the surface. Utilising a feedback loop, the tip's height is adjusted to maintain a consistent force against the sample's surface as it scans, giving the measurement of the sample height. The sample's surface height changes in direct proportion to the change in tip

height. Layer thicknesses of a device stack are measured in relation to one another using a staircase-like pattern that is created during production.

3.1.3 Experimental Setup

The amount of current passing through the device is the significant observable. This includes details on the oxide layer's bulk conductivity as well as the interface barrier heights. The difficulty, however, is in minimising any deviations or nonlinearities brought on by the measuring apparatus itself, with probe contact resistance serving as one such example. It is necessary to choose how to make contact with the device electrodes before conducting current measurements. There are essentially two methods: either the circuit is wire bonded inside a chip carrier, or the contacts are directly probed with tiny metallic probes using micromanipulators.

The direct probing method utilising tungsten probes has been adopted instead due to the devices' design and susceptibility to break from the wire bonding procedure. The tip of the probe must be brought down carefully to prevent damage. When placing the probe into contact, a low voltage is often supplied as a test signal.

To determine if the probe has made contact, the current is watched for a spike in the device current. Initially, because there is no measurable electrical current while the probe is not in touch with the device, the current oscillates around positive and negative currents at 0 amps. Once the probe makes contact with the device, the voltage that has been applied across it now causes a detectable current that matches the polarity of the applied voltage.

In contrast to the probe method, which can be vulnerable to sample damage brought on by the experimentalist, the wire-bonded approach has the advantage which the position of the electrical connection does not change between experiments, thermal expansion while temperature measurements will not significantly affect the contact, and there is less risk of deteriorating the device throughout characterisation.

However, there is a chance that the component will be broken during the bonding procedure with wires. An ultrasonic pulse is utilised to melt a gold or aluminium wire to the device contact while applying pressure to help fuse the two metals together. It has been regularly observed that this pressure can cause internal layers to compress, leading to electric shorts between the two metal contacts and ultimately damaging the device.

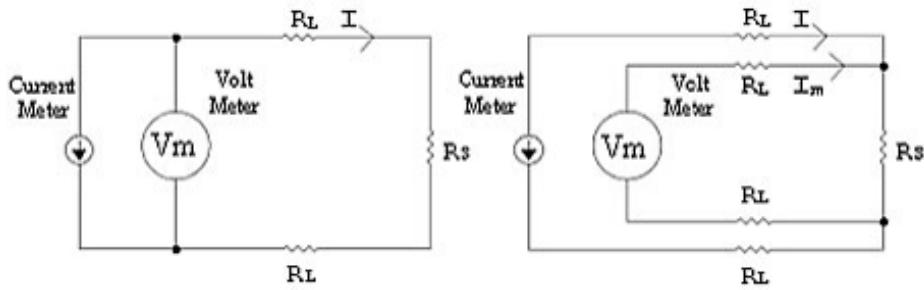


Fig. 3.4 Two wire or Four wire (Kelvin) testing. A schematic of the 2-wire measurement setup is provided herewith. The voltage (VSource) is applied to the sample using two probes. It is important to note that each probe introduces a series resistance (R_{probe}) with the sample's resistance (R_{sample}). The device current (I_{meas}) is measured by an ammeter in series. Conversely, a constant current (I_{source}) is supplied through the sample by two probes. The voltage induced across the device (V_{meas}) by the current is measured with a voltmeter in parallel with the sample.

After deciding on a contact technique, the next choice is how currents will be monitored. Both a 2-wire measurement and a 4-wire measurement are frequently available as options. The sample's conductivity serves as the basis for the decision. The easiest way to measure electrical resistance is to apply a set voltage and track the total current passing through the object. Only two electrical connections are formed, thus the term "2-wire measurement" for this procedure. Ohm's Law is used to determine the device resistance by connecting a voltage source, an ammeter, and the device in series.

Due to the assumption that the electrical resistance is only determined by the test device, which is not true in reality where there are several sources of electrical resistance connected in series with the device, this measurement is often not correct. These include the wires connecting the test object and the voltage source, the internal resistance of the voltage source or the ammeter, and, especially, the contact resistance that develops at the point where the electrical probes and the test object meet.

One of the most crucial parameters to take into account when describing thin films is contact resistance, which may be reduced by placing metal contacts on the sample during manufacturing. Fortunately, the device resistance usually outweighs the electrical resistance, making this method valid in the majority of instances. However, when resistance is small, the parasitic resistances of the measuring circuit become notable and must be eliminated by using a 4-wire resistance measurement.

Ohm's law is still used in this configuration to calculate resistance. Instead of sourcing a voltage and monitoring a current, the device is subjected to a steady current that induces a voltage across it. Through two extra probes connected in parallel to the device, a voltmeter measures the potential decrease. It is crucial to recognise that the same contact resistance and wire resistances that plagued the 2-wire method continue to exist for all four connections. However, in this case, the high impedance of the voltmeter causes a substantially lesser current to pass through the measuring contacts.

The voltage recorded by the voltmeter is thought to more precisely represent the voltage drop across the device since the voltage dip across the parasitic resistance is insignificant. This occurs because the voltage produced across the contact resistance is lowered as a result of the reduced current flowing through the probes, which detect the voltage across the device. By lowering these voltages, which are induced across each probe's contact resistances and contribute mistakes into the voltage measurements, it is possible to measure the voltage across the device with more accuracy.

Thus, the device resistance determines whether to use a 2-wire or 4-wire resistance measurement. The devices examined in this work have high resistance, ranging from kilo-ohms to mega-ohms. The parasitic resistances of the measuring circuit, like the contact resistances, are insignificant at this level. The issue of measuring device currents must now be solved once the device has been attached. Again, there are a variety of techniques that might be applied; the one selected will often depend on the size of the current being measured.

The most elementary method of measuring current is to use a digital multimeter (DMM) ammeter. The device functions in accordance with Ohm's law, utilising the principle of electrical resistance to measure the voltage drop across a fixed resistor, commonly referred to as the shunt resistor. Whilst the validity of this approach is indisputable for currents within the milliamp range and above, issues arise for lower current levels due to the noise induced by the shunt resistor. In order to measure smaller currents, larger shunt resistors are required. This, however, gives rise to two problems.

Firstly, it is important to note that larger resistors are known to introduce greater thermal noise, which has the capacity to disturb the voltage being measured. Secondly, an increase in resistance results in an increase in the voltage drop across the ammeter. The voltage drop, termed the 'voltage burden', becomes problematic when its magnitude is no longer negligible in comparison to the voltage applied to the device under test. The combination of voltage

burden and the thermal noise of the shunt resistor invariably imposes a lower limit on current measurements when a DMM is employed.

The average current range for the devices is 100nA to 1mA, therefore a picoammeter is required to detect considerably lower currents on the order of picoamps to nanoamps. Picoammeters minimise current readings by a number of methods that differ across manufacturers. The majority of them employ a transimpedance amplifier to magnify the signal while an op-amp converts the input current to a voltage. Once again, how this is implemented differs from manufacture to manufacture and is frequently protected intellectual property that is not revealed.

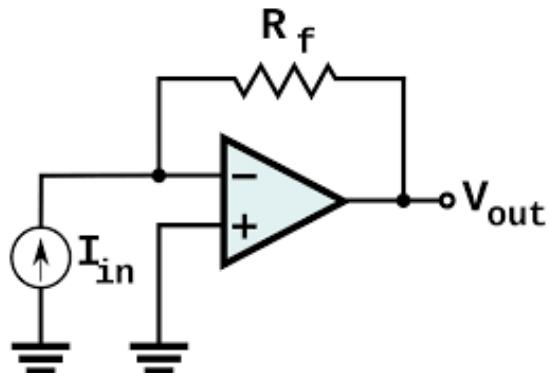


Fig. 3.5 Transimpedance amplifier circuit. The input current (I_{in}) is amplified and converted into an output voltage (V_{out}) via the operational amplifier. The gain is defined by the feedback resistor (R_f). The voltage/current (V_{out}/I_{in}) is equivalent to negative resistance ($-R_f = V_{out}/I_{in}$).

Nevertheless, the general operation can be comprehended with the aid of the circuit in Figure 8. The circuit under consideration is a transimpedance amplifier, the function of which is to convert the input current (I_{in}) to an output voltage (V_{out}). The operational amplifier is known to adjust its output voltage in order to reduce the voltage difference between its two input pins, designated as '-' and '+'. In this circuit, the non-inverting input pin (+) is grounded. This action causes the operational amplifier (op-amp) to adjust its output voltage, thereby ensuring that the voltage at the inverting input pin (-) is also zero volts.

The application of an input current to the circuit results in a transient voltage offset at the input pin. The op-amp rapidly adjusts the output voltage, thereby inducing a current of equal and opposite magnitude through the feedback resistor. This, in turn, results in the cancellation of the input current. The voltage at the inverting pin (-) is rapidly returned to

zero by the feedback from the operational amplifier, thereby creating a virtual ground.

The generation of this inverse current (I_{inv}) is accompanied by the definition of the voltage at the output of the operational amplifier in accordance with Ohm's law: It can thus be demonstrated that the voltage is $V_{out} = -R_f \cdot I_{inv}$, resulting in a voltage that follows the input current. The amplification of this voltage is defined by the feedback resistor, R_f . The virtual ground is a key advantage of this technique. The consequence of this is a significant reduction in the voltage burden, since the shunt resistor that was previously connected in series with the device under test has now been removed. This facilitates the measurement of smaller currents, which would not have been possible using a DMM due to the significant voltage burden caused by the sensing resistor.

It is evident that, in view of the aforementioned factors, the utilisation of a picoammeter constitutes the optimal instrument for the execution of current-time measurements or current-voltage sweeps on our devices. The equipment used in this instance is the Keithley 6430 sourcemeter, which combines a picoammeter and a low noise voltage source into a single device.

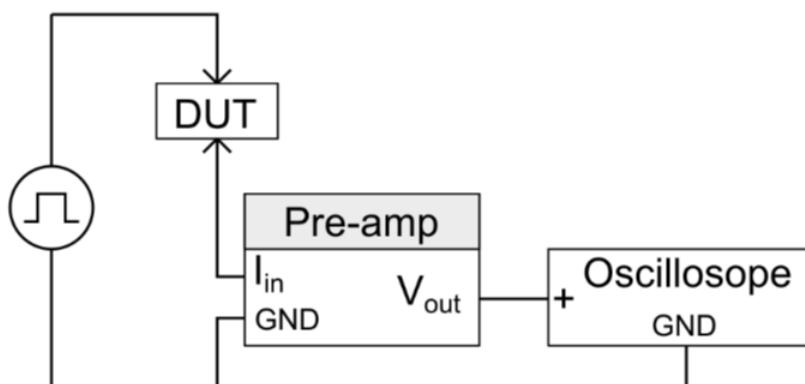


Fig. 3.6 Experimental setup of spike train measurements. Spike trains are generated using an arbitrary signal generator. The device current is amplified and converted into an output voltage via the preamplifier, which is connected in series with the device. The output of the current preamplifier is then captured by an oscilloscope.

In some cases, the device requires the application of voltage transients that are more complex than step potentials, such as pulses or custom spike trains. In these cases, the Keithley's sampling frequency is insufficient to generate such signals. An arbitrary signal generator is

used in its place to create voltage transients, and a current preamplifier is connected in series with the device to amplify device currents. In particular, the oscilloscope (Rigol DS4024) and current preamplifier (SR570) are used.

3.2 Electrical Characterisation

Resistive switching is defined as a reversible phenomenon that occurs in two-terminal elements. In a non-volatile manner, these devices undergo a change in resistance when subjected to electrical stimuli. In the case of ReRAM devices, it is a local redox process that dictates the resistive switching mechanism. The reversibility of the process is achieved by the repeated application of suitable stimuli. This mechanism governs the resistance values between two or more levels.

The predominant phenomenon observed in these devices is resistive switching. For the sake of convenience, the switching states of the memristor can be defined. The assignment of high resistance to the "OFF" state and low resistance to the "ON" state is intuitive, with a contrast in resistance by a few orders of magnitude. The transition from the high resistance state (HRS) to the low resistance state (LRS) is defined as "Set", while the reverse is defined as "Reset". In many cases, an initial electroforming process is required to transform the device from a pristine state to a switchable state. It is generally accepted that the pristine device exhibits a higher degree of resistance than the HRS.

The majority of metal oxide devices exhibit either unipolar or bipolar switching. In contrast, both unipolar and bipolar switching can be observed in our silicon oxides. The preliminary characterisation of these devices encompassed the fundamental I-V characteristics. The experimental procedure involved the execution of the tests utilising the dual sweep functionality of the Keithley 4200-SCS, employing two tip probes with a diameter of $10\ \mu m$. Testing was performed on both sets of samples across all electrode pad sizes.

3.2.1 Unipolar Switching Mode

Initial electroforming is a prerequisite for switching in these devices. It is generally accepted that fresh samples are in a very HRS, which necessitates the application of a significant electrical stimulus to enable the cell to transition into LRS for the first time. Subsequent to this preliminary phase of formation conditioning, the apparatus may be reversibly switched

between two bi-stable states.

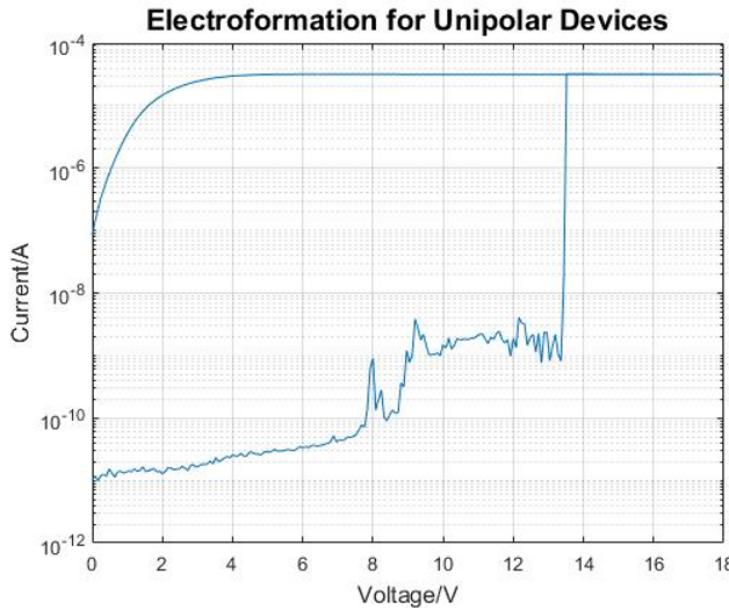


Fig. 3.7 Initial Electroformation step for unipolar switching via a double sweep curve.

The electroforming operation is widely regarded as a form of electrical breakdown, which is critically dependent on current-limiting mechanisms to ensure the subsequent switching functionality of the cell. Current limitations may be addressed by leveraging the existing compliance functionality of the Keithley-SCS. In certain instances, the analyser may exhibit a slower response rate than the formation process itself, resulting in overshoot phenomena during practical applications. It is imperative to note that this electroformation step is only performed once to pristine devices.

The operation is conducted through the programming of the Keithley-SCS to sweep at an elevated voltage of up to 18V, as illustrated in Figure 3.7. During the process of sweeping, it is possible to observe a number of current peaks with the I-V curve displaying an unstable state. Once a sufficiently high voltage is reached, approximately 14V in this case, the device abruptly switches into LRS. Subsequent sweeps are found to be of a more even and refined nature when compared with the preceding sweep.

The observed change in conductance may be attributed to structural changes occurring during the forming process, possibly resulting from a reduction step that involves the removal of oxygen from silicon oxide, thereby forming oxygen vacancies. Following the electroforma-

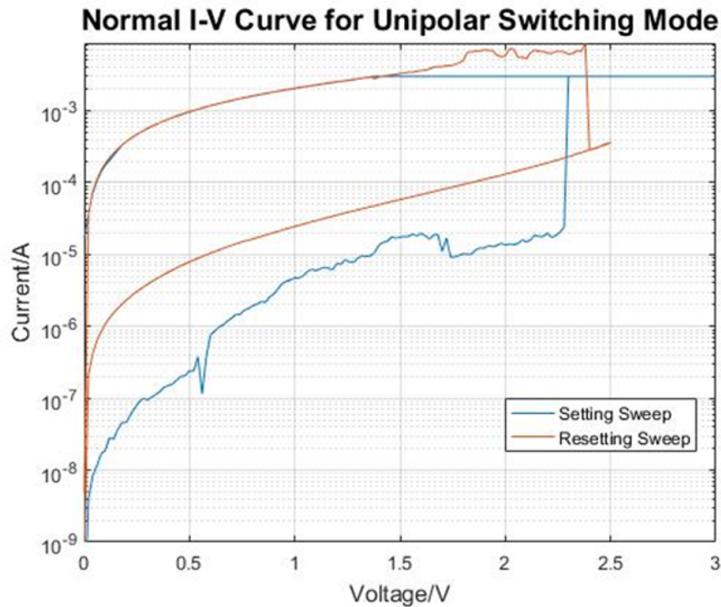


Fig. 3.8 Non-volatile switching behaviour for unipolar device under current compliance.

tion step, the LRS demonstrates stability. The device maintains its state subsequent to the removal of the electrical stimulus, thereby exhibiting non-volatile switching characteristics. It is important to note that the initial very HRS is never recovered.

As demonstrated in Figure 3.8, the unipolar switching mechanism is evident in the initial set of symmetrical MIM devices. The blue plot indicates the Set process, whereby the device transitions from the "Off" state to the "On" state at a specific threshold voltage, approximately 2.3V in this instance.

In this instance, the sweeping voltage has been configured to 3V with 3 mA current compliance, a setting sufficient for the switching process to occur. It has been demonstrated that a reduction in voltage below the threshold does not result in the device transitioning to its previous state.

It is evident that a critical current must be attained for the purpose of resetting the device. The Reset process can be observed in the orange plot, which displays a larger current, approximately one order of magnitude greater than the setting current compliance. This results in the device being restored to HRS.

The phenomenon of Joule heating is induced by high-current flow, resulting in localised heating and device reset. In the absence of current compliance, the device may undergo a hard breakdown or exhibit multiple transitions between the two states. It is important to note that the switching sequence can be performed repeatedly.

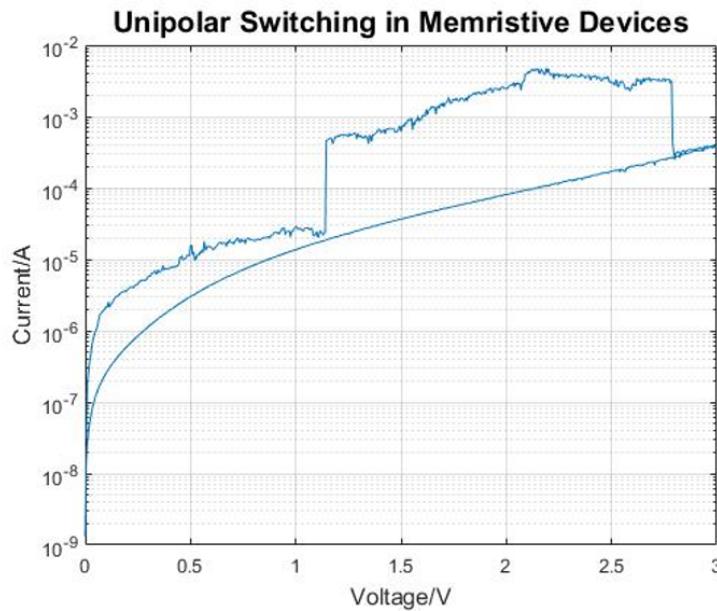


Fig. 3.9 Observation of Set and Reset process under the same sweep.

The HRS conductance remains in a state between that of the LRS and the pristine state. In both cases, the transition is found to be abrupt and independent of the sweeping parameters, in contrast to the ideal pinched hysteresis loop suggested in the previous chapter. In summary, an elevated magnetic field is likely to set the device in the LRS, whereas high Joule heating is likely to reset the device to the HRS.

An alternative mechanism for unipolar switching can be observed in Figure 3.9. Devices with a setting voltage lower than the reset voltage will transition to LRS at a lower voltage. Subsequently, these devices will return to HRS at a higher voltage, which in this case is 1.15V and 2.85V, respectively. There is no current compliance requirement for this type of unipolar switching with reset occurring when the current has reached a critical value.

As illustrated in Figure 3.10, the unipolar device undergoes cycling under conditions of stress testing. The blue spikes in the diagram represent voltage pulses that are utilised to switch the devices in positive bias. The device is set using a short voltage pulse of 4.5 V, with a

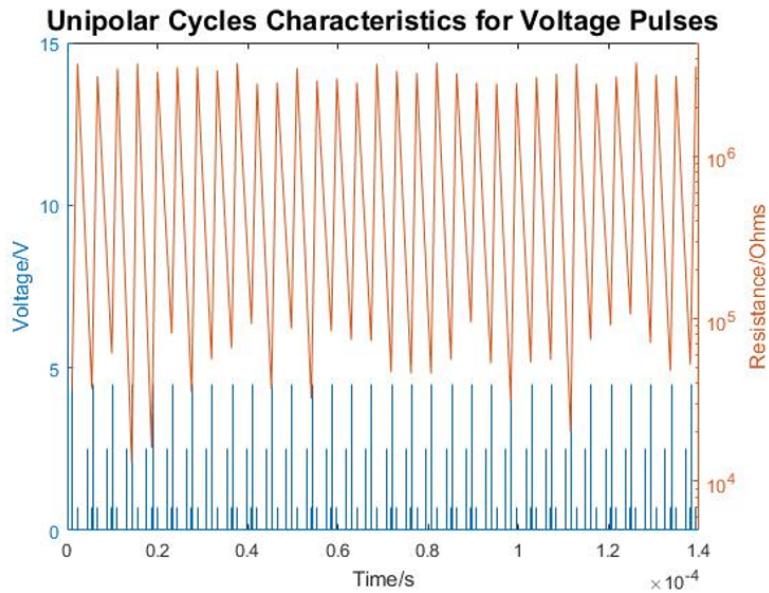


Fig. 3.10 Observation of Set and Reset process under the same sweep.

duration of 100 ns. In order to effect a reset of the device, a longer voltage pulse of 2.5 volts at 2 milliseconds was utilised in order to accommodate Joule heating.

It was observed that each setting and resetting pulse was succeeded by a subsequent reading pulse of 0.7 V at 1 ms. The amplitude of this reading pulse is sufficiently small to avoid interfering with the set and reset process, while providing a clear reading that can be seen in the orange plot. It is evident that under typical operating conditions, the cycling resistance readings exhibit a discrepancy that is at least two orders of magnitude apart.

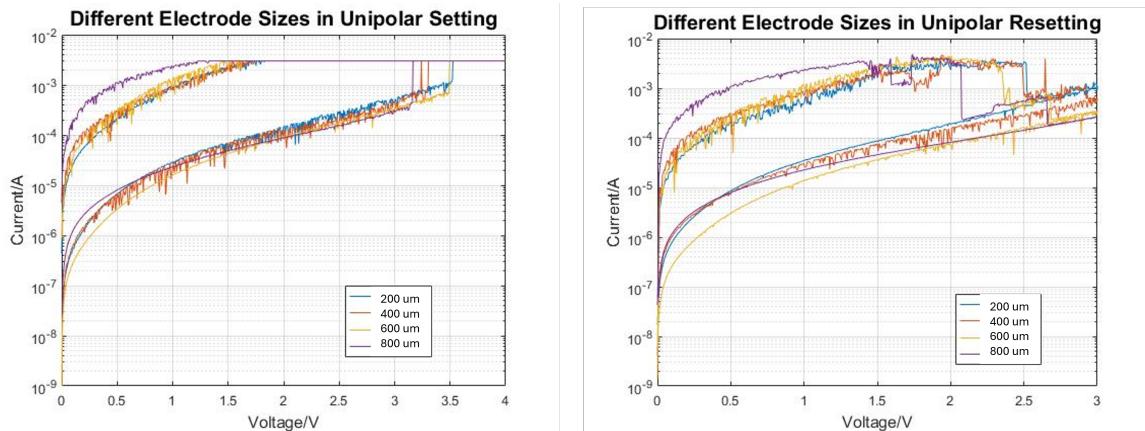


Fig. 3.11 Switching in unipolar devices across different electrode sizes.

It was also demonstrated that switching in unipolar devices is independent of electrode size. Figure 3.11 shows that the switching processes for square contacts ranging from $200 \times 200 \mu\text{m}$ to $800 \times 800 \mu\text{m}$ are comparable. The devices consistently switch at around 3.5 V and 3 mA of current compliance. Similarly, the reset process is consistent when the samples reach the critical current threshold of approximately 5 mA.

3.2.2 Bipolar Switching Mode

Bipolar switching results were obtained from a set of asymmetric devices with Mo/SiO_x/TiAu construct. As with unipolar devices, asymmetric bipolar devices require an initial electro-forming step before the samples can be cycled between two distinct states. Figure 3.12 shows the electroforming process in bipolar devices. A dual voltage sweep is applied to the sample up to -10 V at a current compliance of 0.1 mA. As with the unipolar devices, the sample exhibits some unstable spiking activity as the voltage sweeps from a pristine HRS to a LRS.

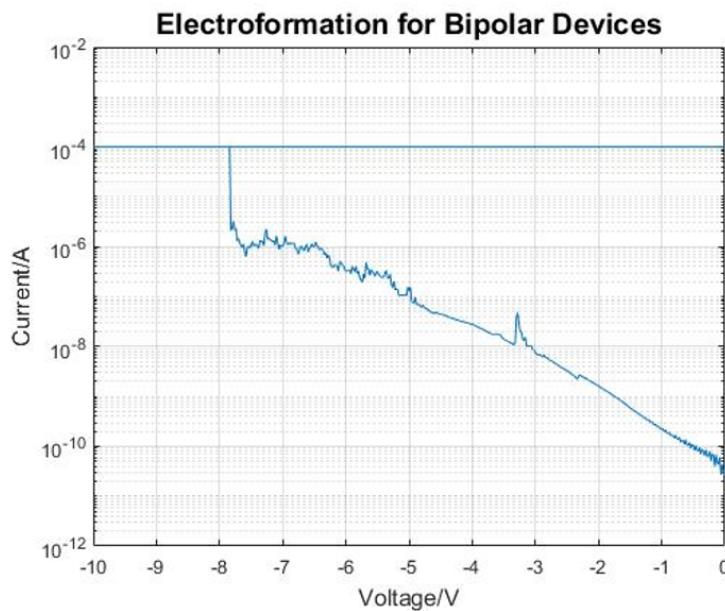


Fig. 3.12 Initial electroformation step for unipolar sample via a double sweep curve.

Following the conclusion of the preliminary electroforming process, the apparatus is capable of reliably transitioning between two stable states. As illustrated in Figure 3.13, the device transitions between two distinct resistance states through the application of voltage stimuli of opposite polarity. The device is set using a negative voltage sweep up to -2V.

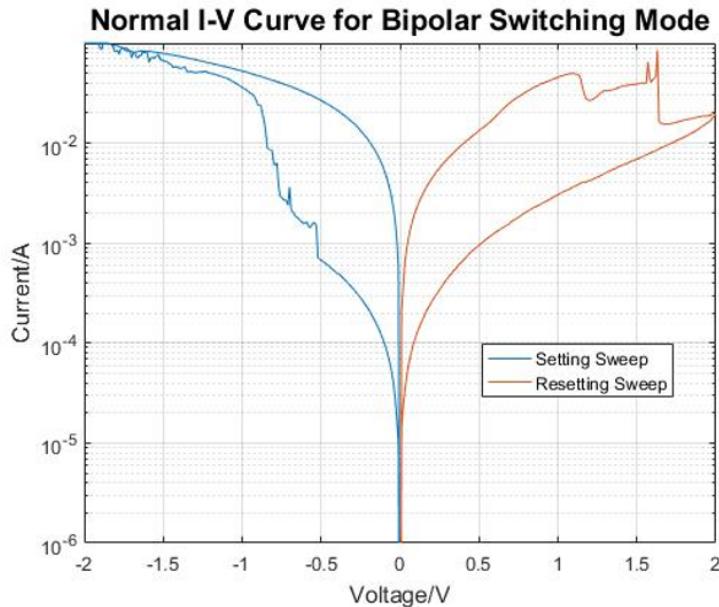


Fig. 3.13 Observation of bipolar switching in asymmetric device with -2V Set and 2V Reset sweeps.

The current compliance was set at 100mA in order to demonstrate a clear transition between the two states, with the conductance changing by two orders of magnitude. It is important to note that a reduced current compliance should be employed in order to achieve a balance between the device's lifespan and its conductivity. The device is reset by means of an opposing 2V dual sweep of positive polarity, a process known as bipolar switching.

From a physical perspective, this particular type of bipolar switching mechanism is intrinsic and can be categorised as belonging to the valence change mechanism class. In this category of memory devices, the electroforming process typically leads to local reduction, thereby forming a conductive pathway. It is hypothesised that this channel is composed of oxygen vacancies, which permit oxygen ions to migrate in and out of the channel in response to an applied electric field.

The location of the local redox process is hypothesised to be in proximity to a filament-to-electrode interface. The effective tunnelling barrier height at this interface is indicative of the resistance state of the device. The height of this barrier is subject to variation under different applied voltage biases, thereby inducing the movement of oxygen ions and resulting in a corresponding alteration to the resistance state.

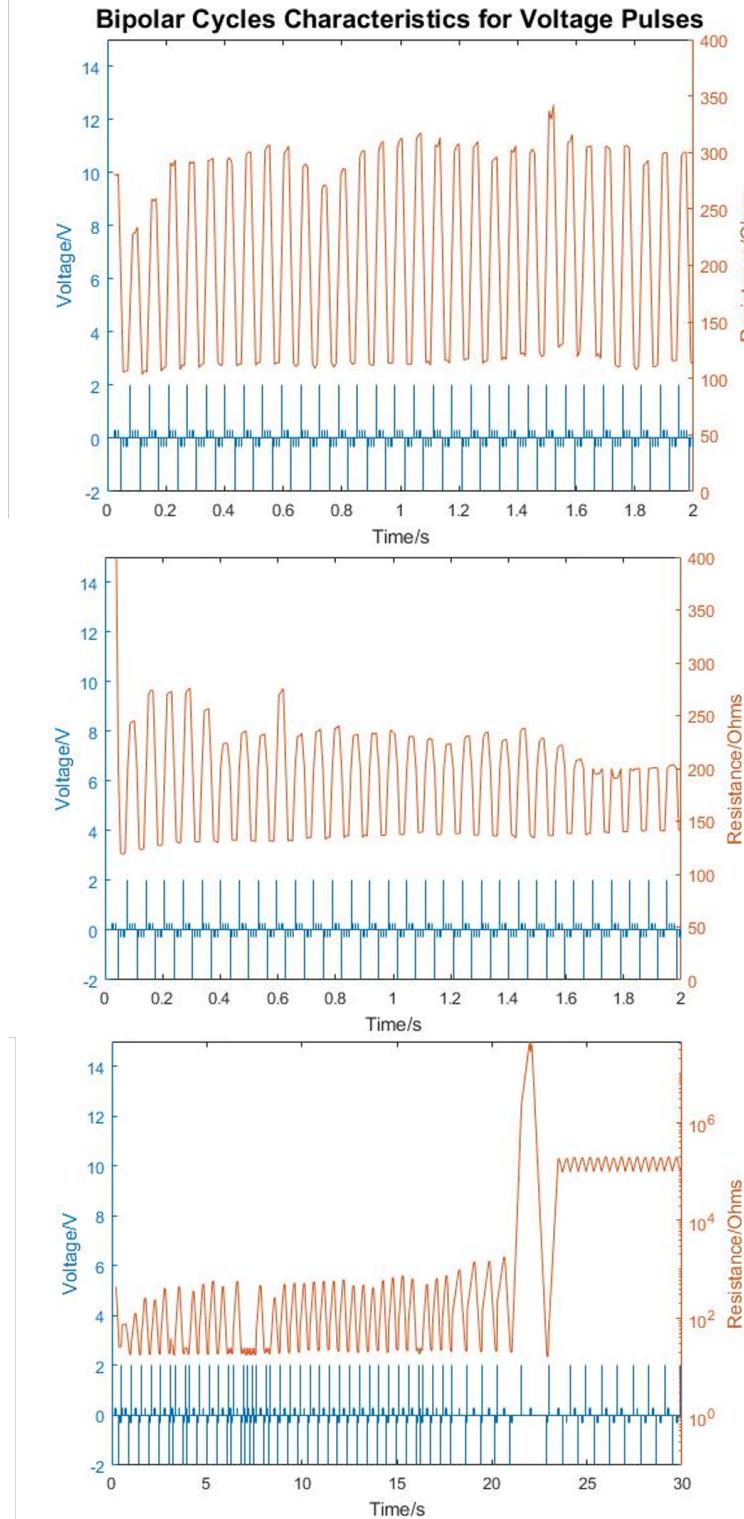


Fig. 3.14 Cycling stress test for bipolar devices. Initial device cycles (top) followed by LRS convergence (middle) and hard reset breakdown (bottom).

The transition between these two resistive states can be facilitated by the application of suitable voltage pulses. As demonstrated in Figure 3.14, the device exhibits a high degree of reliability in its switching capability when utilising a -2 V setting, in conjunction with 2 V resetting pulses.

It was recorded that each setting or resetting pulse is succeeded by five 0.1V or 0.1V reading pulses for the resistive state that the device is purportedly in. In this configuration, the HRS is approximately 300Ω , while the LRS is about 100Ω . It is noteworthy that the selection of these voltage pulses was made with the objective of accurately measuring the resistance, without causing the switching mechanisms of the device to be triggered.

In the experiment, the device demonstrated a minimum of 4500 cycles of operational longevity when subjected to a current bias of 10mA. This was followed by a convergence towards LRS, as evidenced by the switching between 200Ω and 150Ω , as depicted in Figure 3.14. As an alternative scenario, when the device is operating at a higher current compliance of 100mA, the stress test sustains approximately 40 cycles before the device experiences irreversible failure, entering the HRS state at $200k\Omega$.

Finally, Figure 35 demonstrates switching behaviours for bipolar devices across a range of contact sizes, from $200 \times 200\mu m$ to $800 \times 800\mu m$. All the setting sweeps were programmed up to -2V with 5mA current compliance for the purpose of facilitating clear transition observation. It is evident that the resetting sweep has been configured to a voltage of 2V, with a current compliance of 100mA.

The results obtained demonstrate some variations in the switching voltages and contrast ratio between HRS and LRS. This finding suggests the potential necessity for further statistical analysis in subsequent devices. However, it is evident that all samples demonstrate consistent switching activities within the range of voltage stimuli applied during the testing process.

3.2.3 Alternate Operating Modes

As demonstrated in previous observations, the switching of both unipolar and polar samples is reliable under specific, correctly configured, programming conditions. Furthermore, it appears that the switching does not scale in proportion to the electrode contact size. This finding indicates that carrier transport occurs for individual conducting filaments. However, it should be noted that certain devices exhibit alternative switching modes, namely gradual and multi-level switching modes. The presence of parallel conductive pathways within the

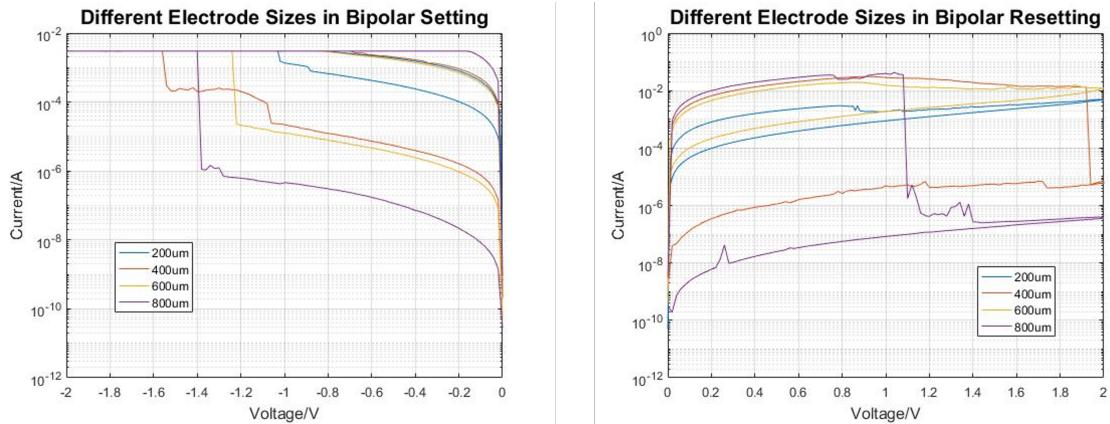


Fig. 3.15 Switching in bipolar devices across different electrode sizes.

same insulating layer is a potential cause of this phenomenon.

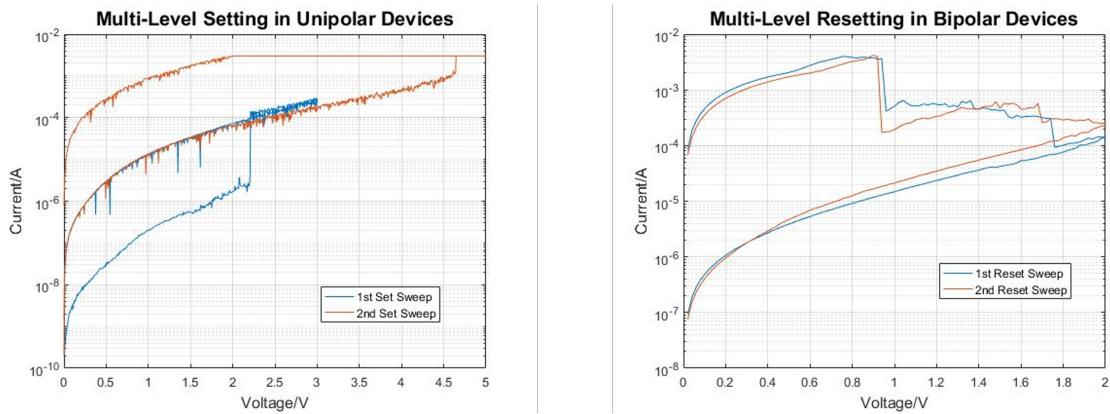


Fig. 3.16 Multi-levels I-V characteristics in MIM devices during the Set (left) and Reset (right) process.

In both unipolar and bipolar samples, there are some devices exhibiting the multilevel switching characteristic, as illustrated in Figure 3.16. The initial transition process during the set stage can be followed by a subsequent stable transition to an even lower LRS, thereby providing a minimum of three or more switchable states.

As demonstrated above, both setting states are found to be stable, with the HRS of the second sweep coinciding with the LRS of the first sweep. The device can be configured at the first or second LRS, with two separate voltage sweep levels available for this purpose. As illustrated in the aforementioned example, the generation of the primary and secondary LRS

was achieved through the utilisation of 3V and 5V sweeps, respectively.

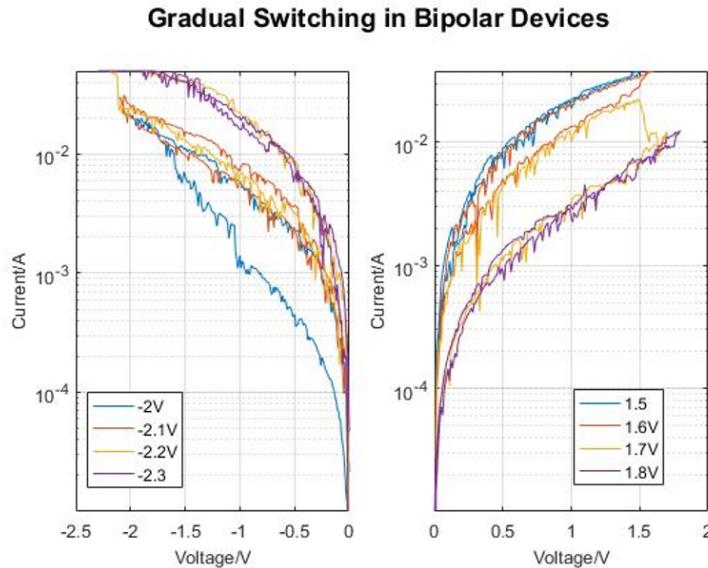


Fig. 3.17 Gradual increase (left) and decrease (right) in conductivity for bipolar device.

In a similar manner, multilevel reset can be observed in bipolar samples at 1V and 1.8V, respectively. It is important to note that, in this case, the transitions window is smaller than that of the unipolar devices. In both cases, the switching is stable, with a contrast ratio between each state that is at least one order of magnitude.

In the case of bipolar switching samples, it is possible to observe not only the abrupt changes in resistance that are normally observed, but also gradual changes in conductivity (see Figure 3.17). As indicated by HRS, the gradual increase in conductance is achieved by sequentially sweeping the device with increasing setting voltage levels, ranging from -2V to -2.3V, under an appropriate stepping current compliance of 100mA in this case.

It has been demonstrated that a gradual decrease in conductivity is generated during the reset process, with this decrease commencing from LRS. This gradual change is obtained by progressively sweeping the device at higher potential, from 1.5V to 1.8V. The concluding phase of these procedures is characteristically sudden, thereby impeding the attainment of further transitions. The outcomes obtained were found to vary in terms of their gradualness or abruptness, suggesting the potential for further statistical analysis with reduced voltage steps in subsequent characterisations.

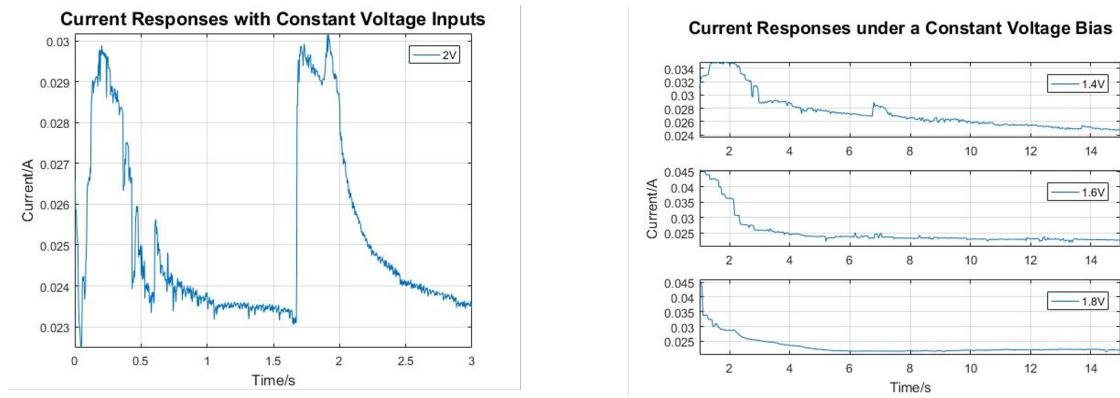


Fig. 3.18 Current-time plots showing transitions between resistive states (left) and time constant comparison (right).

As information processing is concerned with the manner in which data is processed over time, it is imperative to assess the performance of devices over time. In the preceding section, the switching mechanisms have been investigated with regard to time via the cycling stress tests. In this experiment, the characteristics of the device under constant voltage and current bias will be observed.

As illustrated in Figure 3.18, the current time plot for different voltage biases in unipolar samples is demonstrated. Applying a voltage of 2V to the device results in discernible transitions between the set and reset processes, with these transitions exhibiting an inverse relationship to one another.

It is evident that alterations in the resistive state can be observed when there is a rapid increase in the current (set), which is then shortly followed by an exponential decay (reset). The rate of recovery is indicated by the time constant of the exponential decay. A comparison of the individual inputs reveals that the time constant varies in proportion to the voltage bias. When the voltage bias increases from 1.4V to 1.8V, the time constant decreases from 7s to 1s.

When a constant current is applied to the samples, volatile spiking activities can be observed in Figure 3.19. It is evident that an increase in current from $5\mu\text{A}$ to $10\mu\text{A}$, as observed for unipolar devices, results in a corresponding rise in volatility. This phenomenon can be attributed to the constant current input, which has a significant impact on the device's response. These instabilities can manifest as amplitude variations, with spikes ranging from 0.02V to 0.45V, or as timing variations, with spikes occurring more frequently at higher

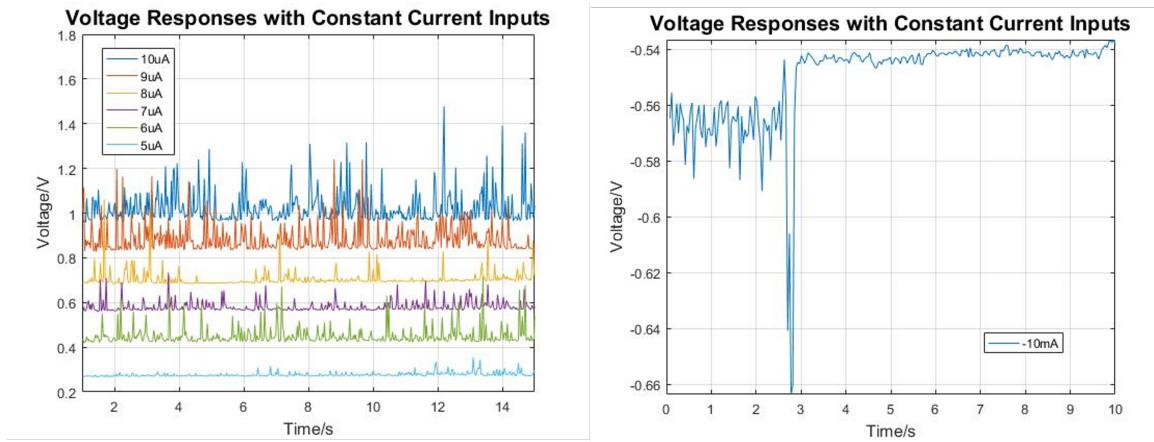


Fig. 3.19 Volatile activities observed under different constant current inputs for unipolar device (left) and switching at sufficiently high input current bias in bipolar device (right).

current biases.

The device exhibits spiking activities until the current bias is sufficiently large to trigger a switching transition. When the bipolar device is critically biased at -10mA, it switches shortly after exhibiting volatile activities, following an uncharacteristically large spike. Following the transition, the spiking behaviours become less predominant.

3.3 Resistive Switching in Silicon Oxide

The present section aims to propose a phenomenological model that governs the switching activities in silicon-rich silica of RRAM devices. The model under discussion will be based on the theory obtained from the literature review in the preceding chapter, with a particular emphasis on the distinction between unipolar and bipolar modes of switching.

In the context of oxide ReRAM devices, two commonly employed switching settings are identified: unipolar and bipolar mode [74]. In the context of unipolar switching, it is notable that the alteration in resistance state is independent of the electrical stimuli polarity. The configuration of these devices is typically characterised by a symmetrical design, incorporating electrodes of equal dimensions on both the top and bottom surfaces.

The present compliance is utilised for the purpose of averting any impairment to the device that may be occasioned by a hard breakdown during the switching process. Conversely,

bipolar devices necessitate the application of electrical stimuli of contrasting polarity to execute switching operations.

3.3.1 Conduction Mechanisms

For the devices and samples referenced in this study, the primary material utilised for the insulating layer is silicon dioxide, SiO_x . It has been reported that silicon dioxide has been doped with conducting ions, such as silver or copper, during the fabrication process in order to behave like ECM cells.

However, diffusion of metallic ions is generally undesirable in CMOS processing, as it can compromise the operations of neighbouring electronics. The present study is concerned with the intrinsic resistive switching property, irrespective of the electrode materials. Given that silicon-rich silica is predominantly employed in the insulating layer, its capacity for complete CMOS-compatible processing is deemed to be highly favourable.



Fig. 3.20 : Conductive regions for filamentary switching (left) and interface switching (right).

In the context of bulk silicon oxide, the formation of a conductive filament within the insulating layer typically occurs during the electroforming process. This conductive filament is generally independent of electrode size, with the switching mechanism being dominated by a single filament. The switching process instigates a minor alteration to the filament. It has been established that this is independent of the insulating layer thickness. This is due to the fact that changes in resistance usually take place in a localised region.

The surface switching mechanism is comparatively under-researched in comparison to filamentary switching. The conductivity of this mechanism is found to be predominantly contingent upon the dimensions of the electrode. The primary driving force behind this mechanism is the formation of a Schottky tunnel barrier across the entire electrode interface and the insulating layer. Consequently, a switching layer is formed at the interface.

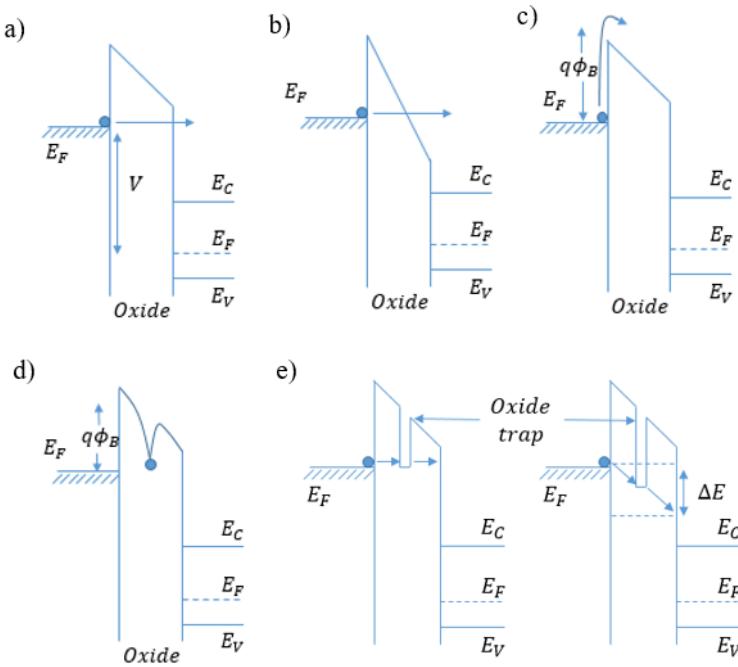


Fig. 3.21 Energy-band diagrams showing different conduction mechanisms: (a) direct tunnelling, (b) Fowler-Nordheim tunnelling, (c) thermionic emission, (d) Poole-Frenkel emission, (e) trap-assisted tunnelling [65]

Electrical conductivity is defined as an intrinsic property that determines the extent to which a given material can oppose a flow of charge. The ideal insulator is characterised by a complete absence of conductance and infinite resistance. It is evident that the conductivity of the silicon oxide thin film, which is measured in several hundred nanometres, is suboptimal due to the presence of a finite amount of conductance. The conductivity of the semiconductor material is also contingent on a variety of external conditions. The aforementioned parameters encompass specific frequencies of light, temperature dependency and applied electric field.

$$E = \frac{V}{d} \quad (3.1)$$

In (3.1), the electric field strength, E , can be expressed as a function of the applied voltage, V , and the distance that the voltage is being applied across, d . Nevertheless, it is important to note that this fundamental estimate may not be applicable to the actual devices. The validity of assumptions regarding negligible oxide charges, voltage flat-band and small band bending may be called into question. In order to facilitate a more thorough analysis and identification of the switching procedure in the samples, additional conduction mechanisms in the insulator

are considered.

$$J \propto E^2 \exp \left[-\frac{4\sqrt{2m^*}(q\phi_B)^{3/2}}{3q\hbar E_i} \right] \quad (3.2)$$

$$J \propto V^2 \exp \left(-\frac{b}{V} \right) \quad (3.3)$$

(3.2) displays the tunnelling current density's dependence on the electric field and voltage, applied as appropriate, while being independent of temperature. In the context of the aforementioned equation, ϕ_B denotes the tunnelling barrier height, E represents the insulator electric field, m^* is defined as 0.42m, which is the carrier effective mass for silicon oxide, \hbar is the reduced Planck constant, q is the electric charge, and b is a constant of proportionality.

In the presence of a strong electric field, conventional tunnelling is the predominant conduction mode for insulating materials. The tunnelling process is a consequence of quantum mechanical effects, with the electron wave function having a finite probability of penetrating through a potential barrier of finite height. Conventional quantum tunnelling refers to the direct passage of an electron through the entire width of a barrier. Alternatively, Fowler-Nordheim tunnelling refers to the electron tunnelling through only part of this height.

$$J \propto E \cdot \exp \left(-\frac{\Delta E_a}{kT} \right) \quad (3.4)$$

$$J \propto \frac{V}{T} \cdot \exp \left(-\frac{c}{T} \right) \quad (3.5)$$

(3.4) illustrates the ohmic current density as a function of electric field, as well as the voltage applied and the temperature. In this equation, ΔE_a denotes the activation energy, k is the Boltzmann constant, T is the temperature in Kelvin and c is a constant of proportionality. In the context of low fields and elevated temperatures, ohmic conduction exerts a predominant influence. This phenomenon entails the thermally induced excitation of carriers, thereby facilitating their transition between conductive states.

$$J \propto \frac{E}{T} \cdot \exp \left(-\frac{\Delta E_a}{kT} \right) \quad (3.6)$$

$$J \propto \frac{V}{T} \cdot \exp \left(-\frac{d}{T} \right) \quad (3.7)$$

Ionic conduction exhibits a comparable expression to ohmic conduction, yet it possesses a distinct activation energy and constant of proportionality. This process is typically charac-

terised by the movement of ions across a material via defects in the crystal lattice of a solid. For an ideal insulator, ions cannot readily travel into and out of the material.

However, an applied electric field will result in the build-up of ionic carriers at the metal-to-insulator interfaces, thereby modifying the voltage distribution across the region. The elimination of the applied electric field will result in the retention of a significant internal field, thereby enabling the flow of an ionic current until equilibrium is achieved.

$$J = \frac{9\epsilon_i \mu V^2}{8d^3} \quad (3.8)$$

$$J \propto V^2 \quad (3.9)$$

The phenomenon of space charge can be attributed to the injection of charge from the electrodes into the insulator, in the absence of compensating charges. The process involves the injection of charges into the dielectric from one electrode and their subsequent capture by the other. The Mott–Gurney law is delineated in (3.8) for space charge limited current in solid and in the velocity-saturation regime accordingly.

In this equation, ϵ denotes the dielectric permittivity, μ is the carrier mobility, L is the material thickness, and $v = \mu E$ is the electron drift velocity. This conduction mechanism is predicated on the presence of a single type of charge carrier, the absence of intrinsic conductivity, and an electric field of zero magnitude at the cathode responsible for the injection of charge.

Further exploration will be directed towards other conduction mechanisms, including Schottky emission and Poole-Frenkel conduction, which will be examined in greater detail. Schottky emission, otherwise known as thermionic emission, occurs when the carriers receive thermal energy in excess of the potential barrier height. The phenomenon of Poole-Frenkel conduction occurs when trapped electrons are thermally excited into the conduction band.

3.3.1.1 Fowler-Nordheim Tunnelling

In the presence of elevated electric fields, quantum mechanical tunnelling emerges as the predominant conduction mechanism in insulating materials. This is a consequence of the process inherent in quantum mechanics, whereby the electron wave function is capable of penetrating a potential barrier.

This process is typically contingent on the electric field, irrespective of temperature. The phenomenon of direct tunnelling occurs when carriers traverse the entire width of the barrier. It has been established that, in the context of Fowler-Nordheim tunnelling, carriers only traverse a proportion of this width.

$$J = \frac{q^2 E^2}{8\pi\hbar\phi_B} \exp\left[-\frac{4\sqrt{2m^*}(q\phi_B)^{3/2}}{3q\hbar E_i}\right] \quad (3.10)$$

$$J \propto \frac{4\pi q m^* k T}{\hbar^3} \quad (3.11)$$

The phenomenon of Fowler-Nordheim tunnelling is contingent upon the trapezoidal configuration of the potential barrier. In the presence of a substantial application of an electric field, an increased incidence of band-bending is observed. This results in a significant reduction in the effective width required for carriers to tunnel through. In the context of a thick oxide layer, this is the prevailing conduction mechanism for a metal oxide structure. Subsequent to the tunnelling process, the carriers are able to move freely between the conduction and valence bands.

The identification of the mechanism for the device is possible through the rearrangement of (3.10) and the graphical representation of the Fowler-Nordheim plot of $\ln(J/E^2)$ against $\frac{1}{E}$ for experimental I-V characterisations. The gradient of this straight-line plot is equivalent to $-\frac{4\sqrt{2m^*}(q\phi_B)^{3/2}}{3q\hbar}$, which can be rearranged to obtain the barrier height ϕ_B . The y-intercept, on the other hand, describes the geometrical efficiency of electron-field emission. The occurrence of this mechanism is contingent upon the product of the electric field and layer thickness exceeding the barrier height.

3.3.1.2 Poole-Frenkel Conduction Mechanism

Conduction may also occur in the absence of quantum tunnelling through the insulator. In the context of materials characterised by a high density of structural defects, the movement of carriers is constrained in a manner that is distinct from the behaviour exhibited by tunnelling mechanisms. The presence of these structural defects also gives rise to the appearance of additional energy states, also known as traps, in the vicinity of the energy band edges. The function of these traps is to restrict the flow of current, and they achieve this by means of a capture and release process.

The Poole-Frenkel conduction mechanism is concerned with electrons trapped in these states. These trapped electrons can eventually amass sufficient energy via thermal fluctuations in the

material to escape from the localized trap states. It is imperative to note that, in the absence of being captured in an alternative trap state, the electrons can ultimately reach the conduction band. It is evident that this mechanism is contingent on two factors: the application of an electric field and the presence of thermal energy. The electron derives its total energy from two sources: the electric field and thermal fluctuations.

$$J \propto E_i \cdot \exp \left[-\frac{q(\phi_B - \sqrt{qE_i/\pi\epsilon_i})}{kT} \right] \quad (3.12)$$

$$J \propto V \cdot \exp \left[\frac{q}{kT} \left(2a\sqrt{V} - \phi_B \right) \right] \quad (3.13)$$

This conduction mechanism is typically driven by electron drift current, $J = qn\mu E$, where q is the electric charge, n is the carrier density, μ is the carrier mobility and E is the electric field. This current may be expanded into (3.12) with dependence on the trap depth ϕ_B , the permittivity of the insulator ϵ and the temperature T . A non-ideality factor m , varying between 1 and 2, may be introduced to the equation to account for the fabrication process and the semiconductor materials used.

3.3.1.3 Thermionic Emission

3.3.1.4 Trap Assisted Tunnelling

3.3.2 Switching Model Analysis

3.4 Summary

References

- [1] Abbott, L. F. (1999). Lapicque's introduction of the integrate-and-fire model neuron (1907). *Brain research bulletin*, 50(5-6):303–304.
- [2] Bastos, P. and Barbosa, R. (2022). Motor reserve: How to build neuronal resilience against ageing and neurodegeneration? *Revue Neurologique*, 178(8):845–854.
- [3] Burnstock, G. (2004). Cotransmission. *Current opinion in pharmacology*, 4(1):47–52.
- [4] Cai, W., Ellinger, F., Tetzlaff, R., and Schmidt, T. (2011). Abel dynamics of titanium dioxide memristor based on nonlinear ionic drift model. *arXiv preprint arXiv:1105.2668*.
- [5] Campbell, K. A., Drake, K. T., and Barney Smith, E. H. (2016). Pulse shape and timing dependence on the spike-timing dependent plasticity response of ion-conducting memristors as synapses. *Frontiers in bioengineering and biotechnology*, 4:97.
- [6] Ceolini, E., Frenkel, C., Shrestha, S. B., Taverni, G., Khacef, L., Payvand, M., and Donati, E. (2020). Hand-gesture recognition based on emg and event-based camera sensor fusion: A benchmark in neuromorphic computing. *Frontiers in Neuroscience*, 14:637.
- [7] Chase, S. M. and Young, E. D. (2006). Spike-timing codes enhance the representation of multiple simultaneous sound-localization cues in the inferior colliculus. *Journal of Neuroscience*, 26(15):3889–3898.
- [8] Chen, P.-Y., Peng, X., and Yu, S. (2017). Neurosim+: An integrated device-to-algorithm framework for benchmarking synaptic devices and array architectures. In *2017 IEEE International Electron Devices Meeting (IEDM)*, pages 6–1. IEEE.
- [9] Chua, L. (1971). Memristor-the missing circuit element. *IEEE Transactions on circuit theory*, 18(5):507–519.
- [10] Chua, L. (2019). Resistance switching memories are memristors. In *Handbook of memristor networks*, pages 197–230. Springer.
- [11] Chua, L., Sbitnev, V., and Kim, H. (2012). Hodgkin–huxley axon is made of memristors. *International Journal of Bifurcation and Chaos*, 22(03):1230011.
- [12] Chua, L. O. and Kang, S. M. (1976). Memristive devices and systems. *Proceedings of the IEEE*, 64(2):209–223.

- [13] Cox, H. R., Buckwell, M., Ng, W. H., Mannion, D. J., Mehonic, A., Shearing, P. R., Fearn, S., and Kenyon, A. J. (2021). A nanoscale analysis method to reveal oxygen exchange between environment, oxide, and electrodes in reram devices. *APL Materials*, 9(11).
- [14] Dayan, P. and Abbott, L. F. (2005). *Theoretical neuroscience: computational and mathematical modeling of neural systems*. MIT press.
- [15] Deiss, S. R., Douglas, R. J., Whatley, A. M., and Maass, W. (1999). A pulse-coded communications infrastructure for neuromorphic systems. *Pulsed neural networks*, 6:157–78.
- [16] Di Ventra, M., Pershin, Y. V., and Chua, L. O. (2009). Circuit elements with memory: memristors, memcapacitors, and meminductors. *Proceedings of the IEEE*, 97(10):1717–1724.
- [17] Du, C. (2017). *Metal Oxide Memristors with Internal Dynamics for Neuromorphic Applications*. PhD thesis, The University of Michigan, US.
- [18] Frenkel, C., Legat, J.-D., and Bol, D. (2019). Morphic: A 65-nm 738k-synapse/mm² quad-core binary-weight digital neuromorphic processor with stochastic spike-driven online learning. *IEEE transactions on biomedical circuits and systems*, 13(5):999–1010.
- [19] Ge, N., Zhang, M., Zhang, L., Yang, J. J., Li, Z., and Williams, R. S. (2014). Electrode-material dependent switching in taox memristors. *Semiconductor Science and Technology*, 29(10):104003.
- [20] Gerstner, W., Kreiter, A. K., Markram, H., and Herz, A. V. (1997). Neural codes: firing rates and beyond. *Proceedings of the National Academy of Sciences*, 94(24):12740–12741.
- [21] Gewaltig, M.-O. and Diesmann, M. (2007). Nest (neural simulation tool). *Scholarpedia*, 2(4):1430.
- [22] Goux, L. and Valov, I. (2016). Electrochemical processes and device improvement in conductive bridge ram cells. *physica status solidi (a)*, 213(2):274–288.
- [23] Govoreanu, B., Redolfi, A., Zhang, L., Adelmann, C., Popovici, M., Clima, S., Hody, H., Paraschiv, V., Radu, I., Franquet, A., et al. (2013). Vacancy-modulated conductive oxide resistive ram (vmco-rram): An area-scalable switching current, self-compliant, highly nonlinear and wide on/off-window resistive switching cell. In *2013 IEEE International Electron Devices Meeting*, pages 10–2. IEEE.
- [24] Hazan, H., Saunders, D. J., Khan, H., Patel, D., Sanghavi, D. T., Siegelmann, H. T., and Kozma, R. (2018). Bindsnet: A machine learning-oriented spiking neural networks library in python. *Frontiers in neuroinformatics*, 12:89.
- [25] Hebb, D. O. (2005). *The organization of behavior: A neuropsychological theory*. Psychology Press.
- [26] Hirose, Y. and Hirose, H. (1976). Polarity-dependent memory switching and behavior of ag dendrite in ag-photodoped amorphous as₂s₃ films. *Journal of Applied Physics*, 47(6):2767–2772.

- [27] Hodgkin, A. L. and Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of physiology*, 117(4):500.
- [28] Indiveri, G. (2021). Introducing ‘neuromorphic computing and engineering’. *Neuromorphic Computing and Engineering*, 1(1):010401.
- [29] Jeong, J. H., Yang, H. W., Park, J.-S., Jeong, J. K., Mo, Y.-G., Kim, H. D., Song, J., and Hwang, C. S. (2008). Origin of subthreshold swing improvement in amorphous indium gallium zinc oxide transistors. *Electrochemical and Solid-State Letters*, 11(6):H157.
- [30] Jo, S. H., Chang, T., Ebong, I., Bhadviya, B. B., Mazumder, P., and Lu, W. (2010). Nanoscale memristor device as synapse in neuromorphic systems. *Nano letters*, 10(4):1297–1301.
- [31] Johnston, D., Magee, J. C., Colbert, C. M., and Christie, B. R. (1996). Active properties of neuronal dendrites. *Annual review of neuroscience*, 19(1):165–186.
- [32] Joksas, D., Freitas, P., Chai, Z., Ng, W. H., Buckwell, M., Li, C., Zhang, W., Xia, Q., Kenyon, A., and Mehonic, A. (2020). Committee machines—a universal method to deal with non-idealities in memristor-based neural networks. *Nature communications*, 11(1):4273.
- [33] Joksas, D., Wang, E., Barmpatsalos, N., Ng, W. H., Kenyon, A. J., Constantinides, G. A., and Mehonic, A. (2022). Nonideality-aware training for accurate and robust low-power memristive neural networks. *Advanced Science*, 9(17):2105784.
- [34] Kandel, E. R., Schwartz, J. H., Jessell, T. M., Siegelbaum, S., Hudspeth, A. J., Mack, S., et al. (2000). *Principles of neural science*, volume 4. McGraw-hill New York.
- [35] Kempter, R., Gerstner, W., and Van Hemmen, J. L. (1999). Hebbian learning and spiking neurons. *Physical Review E*, 59(4):4498.
- [36] Kenyon, A. J., Munde, M. S., Ng, W. H., Buckwell, M., Joksas, D., and Mehonic, A. (2019). The interplay between structure and function in redox-based resistance switching. *Faraday discussions*, 213:151–163.
- [37] Khodagholy, D., Gelinas, J. N., Thesen, T., Doyle, W., Devinsky, O., Malliaras, G. G., and Buzsáki, G. (2015). Neurogrid: recording action potentials from the surface of the brain. *Nature neuroscience*, 18(2):310–315.
- [38] Kim, H., Mahmoodi, M., Nili, H., and Strukov, D. B. (2021). 4k-memristor analog-grade passive crossbar circuit. *Nature communications*, 12(1):5198.
- [39] Knight, B. W. (1972). Dynamics of encoding in a population of neurons. *The Journal of general physiology*, 59(6):734–766.
- [40] Koch, C. (2004). *Biophysics of computation: information processing in single neurons*. Oxford university press.
- [41] Kozicki, M. N., Mitkova, M., and Valov, I. (2016). Electrochemical metallization memories. *Resistive switching: from fundamentals of nanoionic redox processes to memristive device applications*, pages 483–514.

- [42] Lapicque, L. É. (1907). Louis lapicque. *J. physiol*, 9:620–635.
- [43] Lapique, L. (1907). Researches quantatives sur l'excitation electrique des nerfs traitee comme une polarization. *Journal of Physiology, Pathology and Genetics*, 9:620–635.
- [44] Liu, J.-C., Hsu, C.-W., Wang, I.-T., and Hou, T.-H. (2015). Categorization of multilevel-cell storage-class memory: an rram example. *IEEE Transactions on Electron Devices*, 62(8):2510–2516.
- [45] Liu, Q., Long, S., Lv, H., Wang, W., Niu, J., Huo, Z., Chen, J., and Liu, M. (2010). Controllable growth of nanoscale conductive filaments in solid-electrolyte-based reram by using a metal nanocrystal covered bottom electrode. *ACS nano*, 4(10):6162–6168.
- [46] Mabilleau, G. and Sabokbar, A. (2008). In vitro biological test methods to evaluate bioresorbability. In *Degradation Rate of Bioresorbable Materials*, pages 145–160. Elsevier.
- [47] Madams, C., Morgan, D., and Howes, M. (1974). Migration of gold atoms through thin silicon oxide films. *Journal of Applied Physics*, 45(11):5088–5090.
- [48] Mainen, Z. F. and Sejnowski, T. J. (1995). Reliability of spike timing in neocortical neurons. *Science*, 268(5216):1503–1506.
- [49] Markram, H., Lübke, J., Frotscher, M., and Sakmann, B. (1997). Regulation of synaptic efficacy by coincidence of postsynaptic aps and epsps. *Science*, 275(5297):213–215.
- [50] Markram, H., Muller, E., Ramaswamy, S., Reimann, M. W., Abdellah, M., Sanchez, C. A., Ailamaki, A., Alonso-Nanclares, L., Antille, N., Arsever, S., et al. (2015). Reconstruction and simulation of neocortical microcircuitry. *Cell*, 163(2):456–492.
- [51] Mead, C. and Ismail, M. (1989). *Analog VLSI implementation of neural systems*, volume 80. Springer Science & Business Media.
- [52] Mehonic, A., Munde, M. S., Ng, W., Buckwell, M., Montesi, L., Bosman, M., Shluger, A., and Kenyon, A. (2017). Intrinsic resistance switching in amorphous silicon oxide for high performance siox reram devices. *Microelectronic Engineering*, 178:98–103.
- [53] Mehonic, A., Shluger, A. L., Gao, D., Valov, I., Miranda, E., Ielmini, D., Bricalli, A., Ambrosi, E., Li, C., Yang, J. J., et al. (2018). Silicon oxide (siox): A promising material for resistance switching? *Advanced materials*, 30(43):1801187.
- [54] Mel, B. W. (1994). Information processing in dendritic trees. *Neural computation*, 6(6):1031–1085.
- [55] Noble, D. (1962). A modification of the hodgkin—huxley equations applicable to purkinje fibre action and pacemaker potentials. *The Journal of physiology*, 160(2):317.
- [56] Orchard, G., Frady, E. P., Rubin, D. B. D., Sanborn, S., Shrestha, S. B., Sommer, F. T., and Davies, M. (2021). Efficient neuromorphic signal processing with loihi 2. In *2021 IEEE Workshop on Signal Processing Systems (SiPS)*, pages 254–259. IEEE.
- [57] Polsky, A., Mel, B. W., and Schiller, J. (2004). Computational subunits in thin dendrites of pyramidal cells. *Nature neuroscience*, 7(6):621–627.

- [58] Saighi, S., Mayr, C. G., Serrano-Gotarredona, T., Schmidt, H., Lecerf, G., Tomas, J., Grollier, J., Boyn, S., Vincent, A. F., Querlioz, D., et al. (2015). Plasticity in memristive devices for spiking neural networks. *Frontiers in neuroscience*, 9:51.
- [59] Squire, L., Berg, D., Bloom, F. E., Du Lac, S., Ghosh, A., and Spitzer, N. C. (2012). *Fundamental neuroscience*. Academic press.
- [60] Stein, R. and Hodgkin, A. L. (1967). The frequency of nerve action potentials generated by applied currents. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 167(1006):64–86.
- [61] Stimberg, M., Brette, R., and Goodman, D. F. (2019). Brian 2, an intuitive and efficient neural simulator. *elife*, 8:e47314.
- [62] Stocks, N. (2001). Information transmission in parallel threshold arrays: Suprathreshold stochastic resonance. *Physical Review E*, 63(4):041114.
- [63] Stone, J. V. (2019). *Artificial intelligence engines: a tutorial introduction to the mathematics of deep learning*. Sebtel Press Warszawa, Poland.
- [64] Strukov, D. B., Snider, G. S., Stewart, D. R., and Williams, R. S. (2008). The missing memristor found. *nature*, 453(7191):80–83.
- [65] Sze, S. M., Li, Y., and Ng, K. K. (2021). *Physics of semiconductor devices*. John wiley & sons.
- [66] Tal, D. and Schwartz, E. L. (1997). Computing with the leaky integrate-and-fire neuron: logarithmic computation and multiplication. *Neural computation*, 9(2):305–318.
- [67] Teeter, C., Iyer, R., Menon, V., Gouwens, N., Feng, D., Berg, J., Szafer, A., Cain, N., Zeng, H., Hawrylycz, M., et al. (2018). Generalized leaky integrate-and-fire models classify multiple neuron types. *Nature communications*, 9(1):709.
- [68] Truong, S. N. and Min, K.-S. (2014). New memristor-based crossbar array architecture with 50-% area reduction and 48-% power saving for matrix-vector multiplication of analog neuromorphic computing. *JSTS: Journal of Semiconductor Technology and Science*, 14(3):356–363.
- [69] Von Neumann, J. (1993). First draft of a report on the edvac. *IEEE Annals of the History of Computing*, 15(4):27–75.
- [70] Wang, Z., Joshi, S., Savel'ev, S. E., Jiang, H., Midya, R., Lin, P., Hu, M., Ge, N., Strachan, J. P., Li, Z., et al. (2017). Memristors with diffusive dynamics as synaptic emulators for neuromorphic computing. *Nature materials*, 16(1):101–108.
- [71] Wen, S. and Zeng, Z. (2012). Dynamics analysis of a class of memristor-based recurrent networks with time-varying delays in the presence of strong external stimuli. *Neural processing letters*, 35(1):47–59.
- [72] Zha, Y. and Li, J. (2017). Imec: A fully morphable in-memory computing fabric enabled by resistive crossbar. *IEEE Computer Architecture Letters*, 16(2):123–126.

- [73] Zheng, N. and Mazumder, P. (2018). Learning in memristor crossbar-based spiking neural networks through modulation of weight-dependent spike-timing-dependent plasticity. *IEEE Transactions on Nanotechnology*, 17(3):520–532.
- [74] Zhuge, F., Fu, B., and Cao, H. (2013). Advances in resistive switching memories based on graphene oxide. *New Prog. Graphene Res*, 7:185–206.