

N°ordre: 41882

Université Lille1

Sciences et Technologies

École doctorale: ED SPI

*Discipline: Micro et Nanotechnologies,
Acoustique et Télécommunications*

THÈSE DE DOCTORAT

soutenue le 18/12/2015

par

Selina La Barbera

Development of Filamentary Memristive Devices for Synaptic Plasticity Implementation.

Composition du jury:

<i>Directeur de thèse:</i>	Dominique Vuillaume	HDR CNRS, IEMN - Université Lille1
<i>Co-Directeur de thèse:</i>	Fabien Alibart	CR CNRS, IEMN - Université Lille1
<i>Rapporteurs:</i>	Julie Grollier Damien Deleruyelle	HDR, Unité Mixté de Physique CNRS - Thales HDR, IM2NP - Polytech
<i>Examineurs:</i>	Alain Cappy Damien Querlioz Christian Gamrat	Professeur, IRCICA - Université Lille1 CR CNRS, IEF - Paris Sud DR, CEA-Saclay, Nano-INNOV

Dedicated to Prof. Gemma Fiore (1952-2005),
my Mother and my reason for being and
to Engr. Matteo Causo,
my Better Half and my motivation.

Abstract

Development of Filamentary Memristive Devices for Synaptic Plasticity Implementation.

Replicating the computational functionalities and performances of the human brain remains one of the biggest challenges for the future of information and communication technologies. In this context, neuromorphic engineering (i.e. replicating the brain properties and performances in hardware systems) appears a very promising direction. Among different directions in this field, memristive devices have been recently proposed for the implementation of synaptic functions, offering the required features and integration potentiality in a single component.

In this dissertation, we will present how advanced synaptic features can be implemented in memristive nanodevices. We first propose a review of the state of the art in the field of neuromorphic computing. Then, in a second part, by exploiting the physical properties of filamentary switching of electrochemical metallization cells, we successfully implement a non-Hebbian plasticity form corresponding to the synaptic adaptation. We demonstrate that complex filament shape, such as dendritic paths of variable density and width, can reproduce short- and long- term processes observed in biological synapses and can be conveniently controlled by achieving a flexible way to program the device memory state (i.e. the synaptic weights) and the relative state volatility. In a third part, we show that filamentary switching can be additionally controlled to reproduce Spike Timing Dependent Plasticity, an Hebbian plasticity form that corresponds to an increase of the synaptic weight when time correlation between pre- and post-neuron firing is experienced at the synaptic connection. In a fourth part, we show the analogy between biological synapses and our solid state memory device. More precisely, we interpret our results in the framework of a phenomenological model developed for biological synapses. Finally, we exploit this model to investigate how spike-based systems can be realized for memory and computing applications.

These results pave the way for future engineering of neuromorphic computing systems, where complex behaviors of memristive physics can be exploited.

Résumé

Développement des dispositifs memristifs filamentaires pour l'implémentation de la plasticité synaptique.

Reproduire les fonctionnalités et les performances du cerveau humain représente un défi majeur dans le domaine des technologies de l'information et de la communication. Plus particulièrement, l'ingénierie neuromorphique, qui vise à implémenter au niveau matériel les propriétés de traitement de l'information du cerveau, apparaît une direction de recherche prometteuse. Parmi les différentes stratégies poursuivies dans ce domaine, la proposition de composant memristif a permis d'envisager la réalisation des fonctionnalités des synapses et de répondre potentiellement aux problématiques d'intégration.

Dans cette dissertation, nous présenterons comment les fonctionnalités synaptiques avancées peuvent être réalisées à partir de composants mémoires memristifs. Dans un premier temps, nous présentons une revue de l'état de l'art dans le domaine de l'ingénierie neuromorphique. Dans une deuxième partie, en nous intéressant à la physique des composants mémoires filamentaires de type cellules électrochimiques, nous démontrons comment les processus de mémoire à court terme et de mémoire à long terme présents dans les synapses biologiques (STP et LTP) peuvent être réalisés en contrôlant la croissance de filaments de type dendritiques. Dans une troisième partie, nous implémentons dans ces composants une fonctionnalité synaptique basée sur la corrélation temporelle entre les signaux provenant des neurones d'entrée et de sortie, la STDP (Spike Timing Dependent Plasticity). Ces deux approches (STP/LTP et STDP) sont ensuite analysées à partir d'un modèle inspiré de la biologie permettant de mettre l'accent sur l'analogie entre synapses biologiques et composants mémoires filamentaires.

Finalement, à partir de cette approche de modélisation, nous évaluons les potentialités de ces composants mémoires pour la réalisation de fonctions neuromorphiques concrètes.

Acknowledgments

The relationship established with everybody, the friendly atmosphere and team spirit that live in the Nanostructures, nanoComponents and Molecules (NCM) group at the Institute of Electronics, Microelectronics and Nanotechnology (IEMN) have made flew fast these last three years and I would like to sincerely thank all those who contributed to this PhD thesis.

Firstly, I would like to express my sincere gratitude to my PhD Supervisor Dr. Fabien Alibart, for the opportunity of this project, for the continuous support of my PhD study and related research, for his patience, motivation, and excellent knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my PhD study.

Particular thanks to my PhD Director Dr. Dominique Vuillaume who has been an excellent supportive guide.

Many thanks to our collaborators at the Institute of Fundamental Electronics (IEF), specially to Dr. Damien Querlioz, Adrien Vincent and Christopher Bennett with whom we have had extremely fruitful collaboration and strong team work over the last three years.

I would like to thank the thesis jury members: Dr. Julie Grollier, Dr. Damien Deleruyelle, Dr. Damien Querlioz, Dr. Christian Gamrat and Prof. Alain Cappy for finding time to review this manuscript and for their interesting and constructive feedback.

A great thank also to all my PhD colleagues/friends specially to: Dr. François Vaurette, Dr. David Guerin, Dr. Stéphane Lenfant, Dr. Simon Desbief, Dr. Gilbert Sassine, Dr. Nabil Najjari, Dr. Yannick Viero and Dr. Jorge Trasobares.

I would like to thank my father Angelo for his support and many special thanks to Anna and Quintino for their comprehension and for their encouragements. Many thanks to Sara, Andrea, Simone, Claudia, Lucia, Valeria and all my italians friends. Many special thanks to Martina and Stella for having encouraged me and for their presence.

Finally I thank Matteo for having shared every little moment with me and for always giving me the force to fight and move on.

Contents

1	General Context and Motivation	13
1.1	Introduction	13
1.1.1	Why Neuromorphic?	14
1.1.2	Neuromorphic NNET for Information Computing	15
1.2	Neuromorphic NNET with Nano-devices	26
1.2.1	Synaptic Nano-devices	26
1.2.2	Basic Processing and Data Representation	28
1.3	Synaptic Plasticity for Information Computing	31
1.3.1	Causal approach: Synaptic Learning versus Synaptic Adaptation	32
1.3.2	Phenomenological approach: STP versus LTP	33
1.4	Synaptic Plasticity in Nano-devices	33
1.4.1	Causal implementation	34
1.4.2	Phenomenological implementation	45
1.5	Scope and Approach of this Work	51
1.6	Discussion and Perspectives	51
2	Neuromorphic NNET with Filamentary Switching	53
2.1	Introduction	53
2.1.1	Resistive Memories	54
2.2	Filamentary Switching: nano-device level	59
2.2.1	Experimental Evidences	66
2.2.2	Simulation Modeling	70
2.3	Integration strategies: circuit level	73
2.3.1	Top-down approach	73
2.3.2	Bottom-up approach	76
2.4	Neuromorphic NNET strategies: system level	79
2.4.1	The CMOL concept	79
2.4.2	The Reservoir Computing concept	80
2.5	Discussion and Perspectives	81

3	Filamentary Switching: Development and Characterization	83
3.1	Introduction	83
3.2	<i>Ag₂S</i> Thin Films Deposition	83
3.3	Millimeter-scale configuration	86
3.3.1	Fractal analysis of Dendritic Filaments	88
3.4	Nanoscale configurations	90
3.4.1	Top-down approach	90
3.4.2	Bottom-up approach	92
3.5	Discussion and Perspectives	95
4	Synaptic Plasticity with Filamentary Switching	97
4.1	Introduction	97
4.1.1	Plasticity Key Parameters	98
4.2	Synaptic Nano-devices: Phenomenological Implementation	101
4.2.1	Tunable Volatility regimes	102
4.2.2	STP to LTP Transition	104
4.3	Synaptic Nano-devices: Causal Implementation	105
4.3.1	Synaptic Adaptation Implementation	105
4.4	Conflict between Phenomenological and Causal approach	107
4.5	Discussion and Perspectives	108
5	Multiple Plasticity Mechanisms with Filamentary Switching	109
5.1	Introduction	109
5.2	Synaptic Nano-devices: Phenomenological Implementation	110
5.2.1	Tunable Volatility regimes	110
5.2.2	STP to LTP Transition	112
5.3	Synaptic Nano-devices: Causal Implementation	113
5.3.1	Synaptic Learning Implementation	113
5.4	Towards Multiple Plasticity Mechanisms	119
5.5	Discussion and Perspectives	121
6	Filamentary Switching Modeling and Circuit Simulations	123
6.1	Introduction	123
6.2	Synaptic Plasticity: Bio-inspired Model 1.0	124
6.3	Synaptic Adaptation Modeling	125
6.3.1	Bio-inspired Model 1.0	126
6.3.2	STP to LTP Transition for Memory Applications	127
6.4	Synaptic Learning Modeling	131
6.4.1	Bio-inspired Model 2.0	131
6.4.2	STP to LTP Transition for Information Computing	134
6.5	Discussion and Perspectives	138

7	Conclusions and Perspectives	139
7.1	Dissertation Research Work Conclusion	139
7.2	On-Going and Next Steps	140
A	List of publications	143
A.1	Book Chapter	143
A.2	Journal and Conference Papers	143
A.3	Seminar and Oral Presentations	144
A.4	Involvement in Journal and Conferences	145
B	Résumé en française	147
B.1	Chapitre I: Contexte général et motivation	147
B.2	Chapitre II: Réseaux neuronaux neuromorphiques avec com- mutation filamentaire	148
B.3	Chapitre III: Commutation filamentaire: Développement et Caractérisation	149
B.4	Chapitre IV: Plasticité synaptique avec commutation filamen- taire	149
B.5	Chapitre V: Plusieurs mécanismes de plasticité avec commu- tation filamentaire	151
B.6	Chapitre VI: Modélisation et simulations de circuit de la com- mutation filamentaire	152
B.7	Chapitre VII: Conclusions et perspectives	153
	Bibliography	154

Chapter 1

General Context and Motivation

*"If the human brain were so simple that we could understand it,
we would be so simple that we couldn't."*

Emerson M. Pugh, 1977

1.1 Introduction

This opening chapter will provide to the reader the general context in which this PhD thesis is placed by highlighting the motivation behind pursuing research in the field of Neuromorphic Systems.

We will present the state-of-the art of *Neuro-Inspired Computing*. The main objective is to design and fabricate artificial neural networks (ANNs) whose organizing principles are based on those of biological nervous systems (BNNs). For such purpose different research directions and implementation strategies are possible and we will discuss some of them.

Then, by focusing on the nanotechnology approach, the reader will be introduced to a comprehensive overview of current research on emerging nanoscale memories suitable to implement bio-inspired features such as *Synaptic Plasticity*. Without being exhaustive on the different forms of plasticity that could be realized, we propose an overall classification and analysis of few of them, that can be the basis for going into the field of Neuromorphic Computing.

Finally, we will briefly summarize the scope and the overall strategy adopted for the research conducted during this PhD thesis that is devoted to propose *Synaptic Plasticity*, in some of its different forms, as the key for future development of Neuromorphic Systems.

1.1.1 Why Neuromorphic?

The Big Data era brings an urgent demand for high-performance computing. Nowadays, the bridge between computing and communication technologies has given birth to Information Technology (IT) (Chandler and Munday, 2011) which is now the fastest growing industry in the world. The computing system and the human system have as their nerve center the computer (i.e. the Central Processing Unit (CPU) and the memory) and the human brain, respectively. Traditional computers are well adapted to manage data contained in relational databases and spreadsheets, the so-called 'structured data' being easily entered, stored, queried and analyzed. On the contrary, the human brain is very efficient in executing operations regarding a different nature of data, named 'unstructured data': photos, videos, streaming instrument data, web-pages, wikis or satellites data . . . (Chen et al., 2012). In the past years, reflecting the magnitude and the impact of the unstructured data in our society, the IT industry has attempted to make computer treating information in the same way that the brain does.

The most famous form of bio-inspired computing is the so-called Artificial Neural Networks (ANNs) in which the processing units are neurons and the dynamic interconnections are synapses. Software Engineering, Computer Science, Robotics and Machine Learning, all these fields have in common the goal to build ANNs architectures. On the contrary, Cognitive Neuroscience, Computational Neuroscience, Neurobiology and Psychology, are fields that aim to investigate Biological Neural Networks (BNNs), i.e. the human brain. In between these two research lines can be placed the *Neuro-Inspired Computing Paradigm* (also named 'Neuromorphic Engineering' (Mead, 1990)).

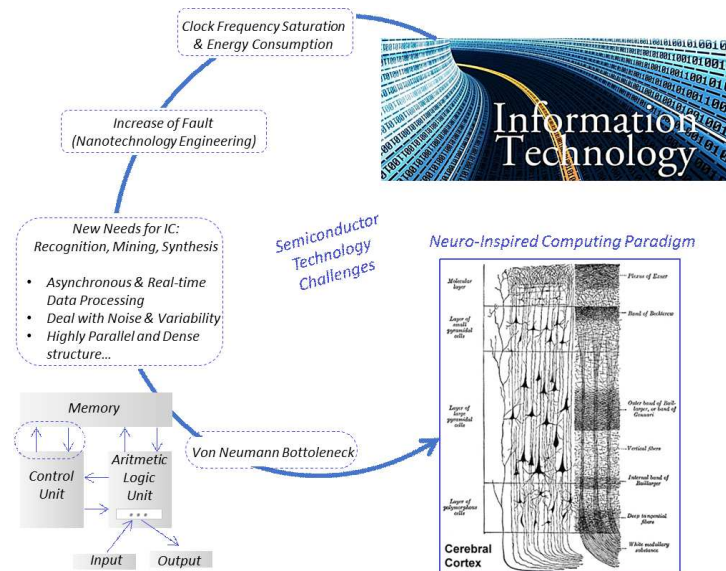


Figure 1.1: Schematic of the Motivations behind the *Neuro-Inspired Computing Paradigm*.

Since the 50s the computing industry has adopted the Von Neumann architecture as their platform (von Neumann, 1948). The major characteristics of this architecture rely on (i) computing system that operates on discrete signals; (ii) memory system that records discrete signals to be processed, a sequence of specific instructions that serially processes the signals and produces the output reports and (iii) computing system that operates by a continuous cycle of fetching instruction from the memory, executing the instruction and storing the result of the instruction in the memory. The predicted end of the Moore's Law due to physical limitation reached by the CMOS technology (Committee et al., 2013), the saturation of conventional computer performances due to material issues (i.e., clock frequency and energy limitations) and more fundamental constraints inherent in the Von Neumann bottleneck (Backus, 1978), i.e. the physical separation of computing units and memories, make the conventional processors inefficient for real-time processing of unstructured data.

IT has to face important challenges in providing suitable solutions for information processing and consequently researchers have started to investigate new computing paradigms that would allow for more powerful systems. The *Neuro-Inspired Computing Paradigm* seems a promising and realistic candidate. To achieve such an ambitious goal, research efforts are needed for understanding the computing principles of biological systems, elucidating how information is computed and stored in neuron and synapse assemblies, and exploring neuromorphic approaches that define hardware functionalities, performances, and integration requirements. Emerging nanotechnologies could play a major role in this context by offering devices with attractive bio-inspired functionalities and associated performances that would ensure the future development of Neuromorphic Hardware. Recent breakthroughs at the system (Merolla et al. (2014)), circuit (Prezioso et al. (2015)), and device levels (Strukov et al. (2008)) are very encouraging indicators for the development of computing systems that can replicate the brain's performances in tasks such as recognition, mining, and synthesis (Liang and Dubey (2005)).

1.1.2 Neuromorphic NNET for Information Computing

Neuromorphic NNET is a concept of information processing that is inspired by conventional ANNs as well as by the way biological nervous systems (i.e. BNNs) process information.

In this section we will present a comparative analysis of ANNs and BNNs at structural and functional levels, devoted to point out an hardware architecture and computing paradigms roadmap for neuromorphic NNET. One research direction will be devoted to implement conventional ANNs, while another one will aim at creating systems able to emulate BNNs behavior.

In between these two main directions, i.e. ANNs and BNNs, neuromorphic computing and engineering emerge as an intermediate solution: the objec-

tive is still oriented toward the development of new computing systems but with stronger analogy with biology with respect to ANNs. This classification should be carefully handled since the frontier between these different fields is far from being clear. In this contest, thanks to recent progress in nanotechnologies and material science can be envisioned to build new neuromorphic hardware architecture. Indeed, emerging nanotechnologies, able to mimic synaptic functionality, could play a key role for the future development of neuromorphic systems.

Basics

A driving force for *Neuro-Inspired Computing* is the state of the art performances reached by the ANNs, bio-inspired massively parallel systems that can implement a variety of challenging computational characteristics such as learning ability, adaptability, fault tolerance and low energy consumption. The historical trace of ANNs developments can be divided into three periods. The first one in the 40s with the McCulloch and Pitts's first neuron model (McCulloch and Pitts, 1943). The second one in the 60s with the introduction of the 'perceptron' by F. Rosenblatt (Rosenblatt, 1958), the so-called 'first neuromorphic engine', on which is still based the field of ANNs. Through the 70s, due to the limitations of computational machines that processed neural networks, the ANNs field was relatively stagnant. The third period starts in the 80s with the advent of greater processing power in computers, and advances with the backpropagation algorithm (Werbos, 1988) that brought back some interest in the ANNs field. In parallel, Hopfield proposed another concept for neuromorphic computing based on associative memory principle that were extended to speech recognition tasks or classification of pattern (Carpenter, 1989). In the 90s a fundamental milestone in the context of neuromorphic computing was accomplished by Mead with the first VLSI design of a silicon retina and neural learning chips in silicon (Mead, 1990). Nowadays, ANNs have seen the emergence of very complex systems with impressive performances in recognition tasks, for example. Along these lines, the deep neural networks (DNNs) and convolution neural networks (CNNs) are today the most promising candidates for new computing systems (Hinton et al., 2015).

ANNs are based on two fundamental components by analogy with biological systems: neurons and synapses (Fig. 1.2). The biological neuron (or nerve cell) consists of three main parts: a central cell body, called the soma, and two different types of branched, tree-like structures that extend from the soma, called dendrites and axons. A synapse is an elementary structure and functional unit between two neurons (i.e. an axon of a pre-neuron i and a dendrite of another post-neuron j). If a neuron is at rest, it maintains an electrical polarization (i.e., a negative electrical potential, around $-70mV$, inside the neuron's membrane with respect to the outside). When information

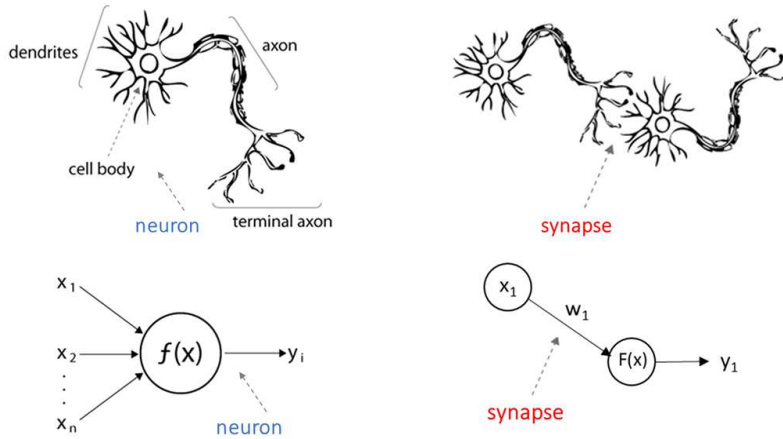


Figure 1.2: Schematics of biological and artificial neuron and synapse.

from other neurons, in the form of electrical pulses (named action potential), reaches the synapse's terminal (pre-neuron i), chemicals called 'neurotransmitters' are released and the post-neuron j goes from being polarized to being depolarized. The neurotransmitters diffuse across the synaptic gap, to enhance or inhibit, depending on the type of synapse, the receptor neuron's (post-neuron j) tendency to emit an electrical impulses. Once the neuron j reaches a certain threshold (firing level), an action potential is fired, sending the electrical signal down the axon. This is an all-or-none phenomenon. 'All-or-none' means that if a stimulus doesn't exceed the threshold level, no action potential results. After the neuron has fired, there is a refractory period in which another action potential is not possible. The synapse's effectiveness can be tuned by the synaptic activity so that the synapses can learn from the past activity history, acting as a memory.

An artificial neuron can be described from a computational view point, as a binary threshold unit. Given n input signals ($x_i, i = 1, 2, \dots, n$), the neuron computes a weighted sum of its n input signals and generates an output of 1 if this sum is above a certain threshold u . Otherwise, an output of 0 results:

$$y = \Theta \sum_{j=1}^n w_{ij} \cdot x_j - u \quad (1.1)$$

where Θ is a unit step function at 0, and w_{ij} is the synapse weight associated to the i^{th} -input of a pre-neuron i and a post-neuron j . A crude analogy between ANNs and BNNs can be done: wires and interconnections model axons and dendrites, connection weights represent synapses, and the threshold function approximates the activity in a soma. Positive weights correspond to excitatory synapses, while negative weights model inhibitory ones.

Once introduced the basic ingredients of both ANNs and BNNs, we now

present similarities and differences of such networks by focusing on key computing aspects on which Neuromorphic NNET are inspired.

ANNs vs. BNNs

The human cerebral cortex contains about 10^{11} neurons and approximately 10^{15} synaptic interconnections. From an architectural point of view, the BNNs architecture is constructed in a three dimensional way with a random organization from microscopic components, i.e. neurons that seem capable of nearly unrestricted interconnections with other neurons via dendrites and axons.

On the contrary, ANNs architecture can be viewed as weighted ordered topology in which artificial neurons are nodes and directed edges (i.e. weights) are connections between neuron outputs and neuron inputs. ANNs can be grouped into two categories: feed-forward networks, in which topologies have no loops and recurrent (feedback) networks, in which loops, occur because of feedback connections (Fig. 1.3). Feed-forward networks are defined static, that is, they produce only one set of output values rather than a sequence of values from a given input. Additionally, the response of Feed-forward networks to a certain input is independent of the previous network state. Recurrent networks, on the other hand, are dynamic systems. When a new input pattern is presented, the neuron outputs are computed. Because of the feedback paths, the inputs to each neuron are then modified, which leads the network to enter a new state. In ANNs architectures, the concept of memory relies in the combination of the network topology and the way how information is embedded in the updating weights history (i.e static feed-forward systems and dynamic recurrent ones).

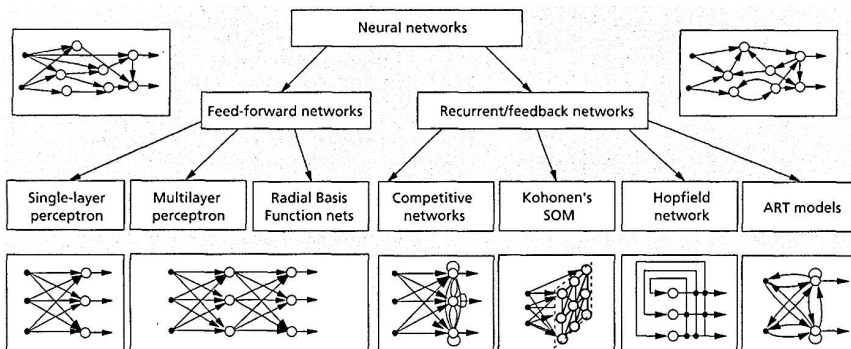


Figure 1.3: A taxonomy of feed-forward and recurrent network architectures (Jain et al., 1996).

As previously mentioned, what has attracted the most interest in ANNs is their ability to learn. A learning process can be roughly defined as the determination of the weights or mathematically, as an optimization prob-

lem. By following the way learning is performed, we can distinguish two major categories of ANNs: (i) fixed networks in which the weights cannot be changed $dw_{ij}/dt = 0$, i.e. the w_{ij} are fixed a priori according to the problem to solve; (ii) adaptive networks which are able to change their weights, i.e. $dw_{ij}/dt \neq 0$. Thus, by following a certain learning algorithm, the network is able to 'learn' by adjusting its connection weights. In other words, a learning algorithm, based on iterative corrections, aims at finding the optimal weight and the faster is its ability to reach the target (i.e. to converge) the higher is its performance. There are three main learning paradigms: supervised, unsupervised and reinforcement learning. In supervised learning the network is trained with a correct answer (output) for every input pattern. Reinforcement learning is a variant of supervised learning in which the network is provided with only a critique on the correctness of network outputs, not the correct answers themselves. In contrast, unsupervised learning, or learning without a teacher, does not require a correct answer associated with each input pattern in the training data set.

The well-known Rosenblatt's perceptron rule (Rosenblatt, 1958) belongs to the supervised learning. A perceptron consists of a single artificial neuron i with adjustable weights (w_{ij}), inputs ($x_i, i = 1, 2, \dots, n$), and threshold u , as previously described by the equation 1.1. The output y of the perceptron is +1 if $u > 0$, and 0 otherwise. In the simplest two-class classification problem, the perceptron assigns an input pattern to one class if $y = 1$, and to the other class if $y = 0$. The linear equation:

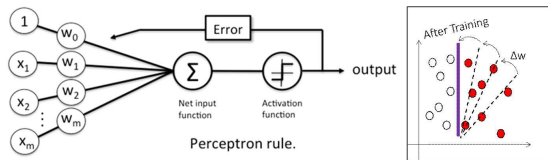
$$\sum_{j=1}^n w_{ij} \cdot x_j - u = 0 \quad (1.2)$$

defines the decision boundary (a hyperplane in the n -dimensional input space) that halves the space. A geometric interpretation is shown in Fig. 1.4a. Each unit in the first hidden layer forms a hyperplane in the pattern space; boundaries between pattern classes can be approximated by hyperplanes. A unit in the second hidden layer forms a hyperregion from the outputs of the first-layer units; a decision region is obtained by performing an AND operation on the hyperplanes. The output-layer units combine the decision regions made by the units in the second hidden layer by performing logical OR operations. As shown in Fig.1.4b, perceptron learning rules, based on the error-correction principle, are developed to determine the weights and threshold, given a set of training patterns. However, a single-layer perceptron can only separate linearly separable patterns as long as a monotonic activation function is used.

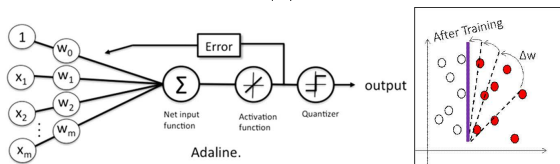
Another ANNs system was the ADALINE (ADAPtive LInear Element) which was developed in 1960 by Widrow and Hoff (Widrow et al., 1960). The memistor ADALINE was the first hardware implementation of the analogue synaptic weights. In contrast to the perceptron rule, the delta rule of the

Structure	Description of decision regions	Exclusive-OR problem	Classes with meshed regions	General region shapes
Single layer	Half plane bounded by hyperplane			
Two layer	Arbitrary (complexity limited by number of hidden units)			
Three layer	Arbitrary (complexity limited by number of hidden units)			

(a)



(b)



(c)

Figure 1.4: Learning Schematics. (a) A geometric interpretation of the role of hidden unit in a two-dimensional input space. (b) Schematics of Rosenblatt's perceptron rule. (c) Schematics the Adaline's rule.

adaline (also known as Widrow-Hoff rule) updates the weights based on a linear activation function rather than a unit step function (Fig. 1.4c).

The development of the back-propagation learning algorithm for determining weights in a multilayer perceptron (MLP) has made these networks the most popular ANNs (Jain et al., 1996). The backpropagation learning algorithm can be divided into two phases: (i) propagation and (ii) weight update. Each propagation (i) involves the following steps: forward propagation of a training pattern's input through the neural network in order to generate the propagation's output activations; backward propagation of the propagation's output activations through the neural network using the training pattern target in order to generate the deltas (the difference between the input and output values) of all output and hidden neurons. For each weight w_{ij} (ii) follow the following steps: multiply its output delta and input activation to get the gradient of the weight; subtract a ratio (percentage) of the gradient from the weight. This ratio, named learning rate η , influences

the speed and quality of learning. A common method for measuring the discrepancy between the expected output t and the actual output y is using the squared error measure:

$$E = (t - y)^2 \quad (1.3)$$

where E is the error. Using gradient-descent method the change in each weight in a back-propagation algorithm results:

$$\Delta w_{ij} = -\eta \frac{dE}{dw_{ij}} \quad (1.4)$$

The -1 is required in order to update in the direction of a minimum, not a maximum, of the error function. For a single-layer network, this expression becomes the Delta Rule.

BNNs communicate through pulses, the timing of the pulses to transmit information and perform computation while the ANNs are based on software computational model involving the propagation of continuous variable from one processing unit to the next. BNNs processing abilities follow highly parallel processes operating on representations that are distributed over many neurons. The relative slow processing speed for BNNs is due to the fact that neurons need several milliseconds to react to stimulus and the elementary 'cycle time' is of the order of one millisecond. For ANNs the processing speed can achieve switching times of a few nanoseconds. Silicon gate times are on the order of one nanosecond, that is, a million times faster than BNNs.

A very real difficulty of correlating ANNs with BNNs lies in the way weights and synaptic strengths were modified (i.e. their capability to learn). In the brain, we learn by creating (weighting) synaptic connections between neurons from different experiences. After, we can react and adapt to unknown situations which are similar to the learning ones by exploiting the information stored in the synaptic connections. Thus, biological synaptic strengths are modified in response to synaptic activity and learning is achieved as consequence of different experiences. On the other hand, weights in ANNs are altered mathematically in a software network, based on differences in values. Thanks to emerging nanoscale memories able to mimic biological synapses, this latter characteristic (i.e. the capability to learn), could be potentially directly integrated in new neuromorphic systems. As it will be explained in the next section, new nanotechnologies would ideally allow to change completely the conventional computing platform in the sense that the memory will be *IN* the processing unit.

Roadmap

In this context of new IT challenges and computing demands with higher complexity, a new device technology roadmap is required to continue scaling

the performance of old architectures and to implement new non-Von Neumann paradigms with enhanced and enriched computing capabilities (Fig. 1.5).

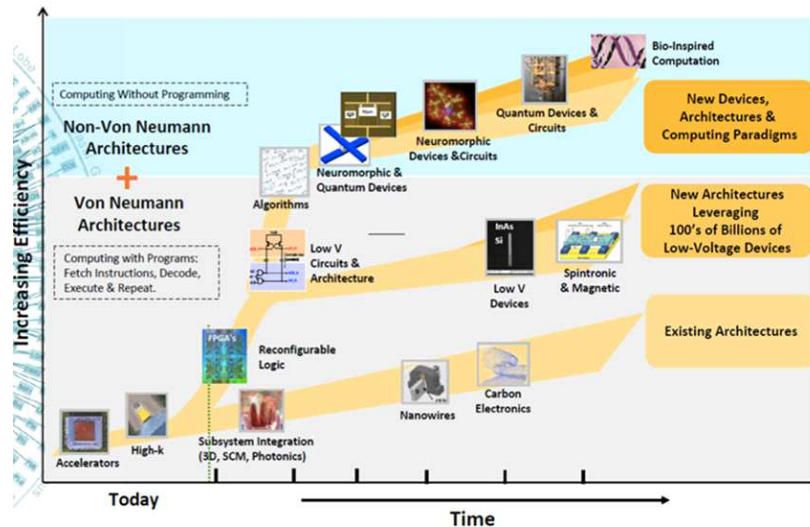


Figure 1.5: New Hardware Architectures and Computing Paradigms Roadmap (Dillenberger et al., 2011).

Recent multidisciplinary findings from neurobiology, computer science, machine learning and emerging nano-scale memory devices are creating the conditions for the efficient hardware implementation of ANNs (Temam, 2010). As mentioned before, we can distinguish different research directions and implementation strategies. One aims at fabricating ANNs by exploiting a 'purely digital approach'. Due to the continuous increase of computing performances of conventional computers, ANNs have been mainly developed in software, a convenient platform for their implementation. Software-based multi-layers perceptrons are capable of impressive performances in classification or recognition as illustrated by the state of the art classifier (Krizhevsky et al., 2012). Based on a convolutional network (Henaff et al., 2011) it can classify into 1000 classes more than 1 million pictures with a high capacity of generalization. Such software approaches are nowadays used by Google or Yahoo to realize complex classification tasks such as pictures or video classification. As a matter of comparison, supercomputers have today the capacity of ten of *petaflop/s* (with an energy consumption in the range of *MW*) when the biological brain is estimated to be in the range of *petaflop/s* (with an energy consumption around *10W*). Dedicated hardware are then required. In particular the next big challenges would be to allow an on-line intelligent computing, in other words the learning capability has to be realized on-line (i.e. directly on general purposed computer). Another ANNs hw implementation direction is 'the purely CMOS approach',

that is devoted to design efficiently neuromorphic circuits with conventional technology (i.e. analog computing based on subthreshold CMOS, for example). The first VLSI design of a silicon retina (Mead, 1990) is a successful example. One limitation is to have still the Von Neumann architecture as a foundation.

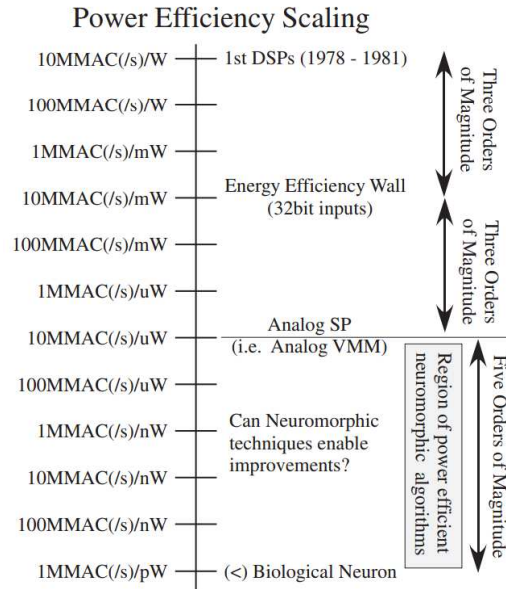


Figure 1.6: A spectrum showing the computational efficiency of various technologies, including digital technologies, analog Signal Processing (SP), as well as best estimate of biological neuron computation. (Hasler and Marr, 2013)

In parallel to research efforts focused in recreating ANNs, another one is devoted to build systems reproducing exactly the biological ones (BNNs). With an hybrid approach, the EU Flagship Human Brain Project (HBP) is targeting to replicate with high accuracy full cortical columns by using super computers resources and specific silicon chips implementing neuronal functionality. The American Synapse project funded by Darpa and supervised by IBM is targeting to match the density of components observed in the human brain (10^{10} neurons and 10^{15} synapses) by both super computers resources and dedicated multi cores CMOS chip (with lower accuracy on the neuronal dynamic with respect to the HBP). The SpiNNaker project is also developing specific neuromorphic core in order to reach about 1 million of neurons in terms of complexity. Fig.1.6 shows the estimated peak computational energy efficiency for digital systems, analog signal processing, and potential neuromorphic hardware-based algorithms. This comparison requires keeping communication local and low event rate, two properties seen in cortical structures. Computational power efficiency for biological systems is 8 – 9 orders of magnitude higher (better) than the power efficiency wall for digital

computation. Hasler et al. (Hasler and Marr, 2013) have recently proposed a new analog techniques at a $10nm$ node that can potentially reach the same level of biological computational efficiency and their conclusion states that with current research capabilities, reaching a system at the scale of the human brain is quite possible.

Along this research line a 'purely CMOS approach' can be also adopted. An area-efficient mixed-signal BNNs implementation realized in a VLSI model of a spiking neural network was performed by (Schemmel et al., 2006).

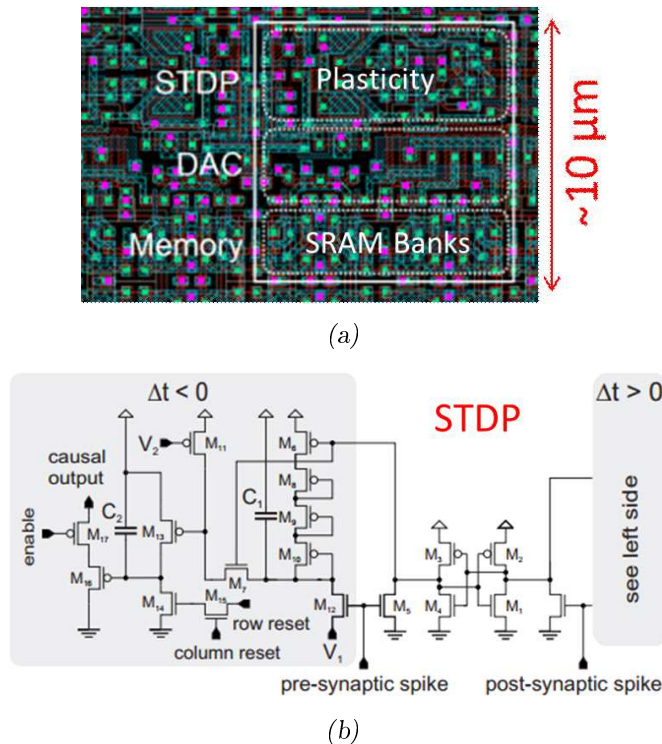


Figure 1.7: The custom IC Approach. (a) CMOS artificial synapse. (b) Circuit diagram of the STDP circuit located in each synapse. Adapted from (Schemmel et al., 2006)

This electronic implementation succeeds in emulating an emergent bio-inspired learning rule, the spike-timing dependent plasticity (STDP), that will be studied in more details in the next sections, maintaining a high level of parallelism and simultaneously achieves a synapse density of more than $9k$ synapses per mm^2 in a $180nm$ technology (Fig. 1.7b). This allows the construction of neural micro-circuits close to the biological specimen while maintaining a speed several orders of magnitude faster than biological real time. This BNNs implementation approach based on standard VLSI CMOS technology can be extremely useful for a large variety of applications, ranging from high speed modeling of large scale neural systems to real time behaving systems, to brain machine interfaces. For example, multi-chip spiking neural

networks comprising mixed analog/digital circuits can be used to validate brain inspired computational paradigms in real-world scenarios, and to develop a new generation of fault-tolerant event-based computing technologies (Indiveri et al., 2011). Nevertheless, the major limitation relative to this approach is still linked to the Von-Neumann bottleneck.

In between these research directions (i.e. ANNs or BNNs implementation strategies) we can place the neuromorphic NNET one (Fig.B.2).

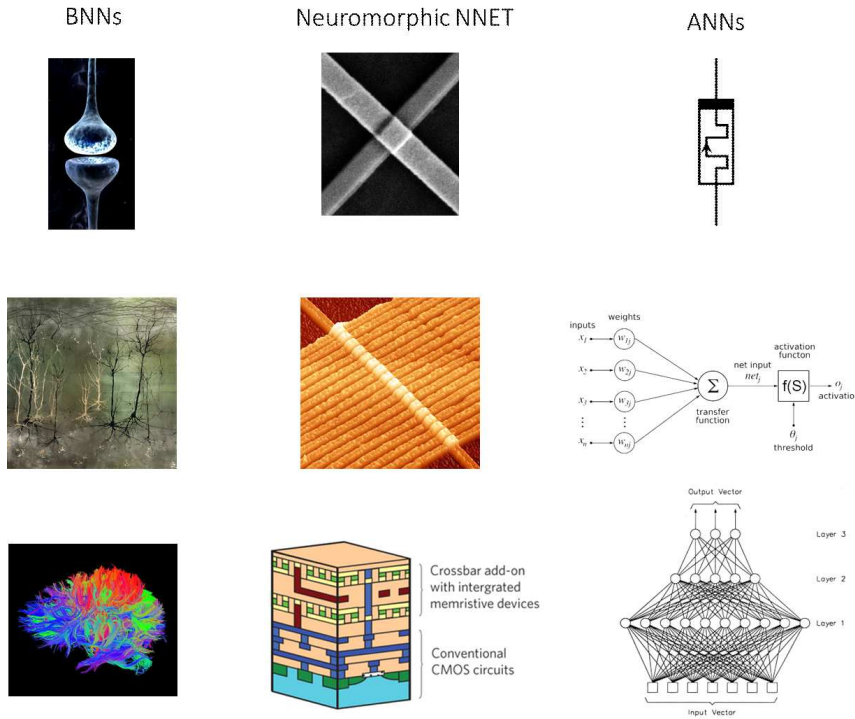


Figure 1.8: The Nanotechnology Approach: Neuromorphic NNET

The main objective is to design and fabricate hardware systems for computing whose organizing principles are based on those of BNNs and ANNs. In this manuscript we will refer to such configuration as neuromorphic NNET (or neuromorphic systems). In order to match the efficiency of biological systems (BNNs), synaptic functionalities should be realized with a dedicated technology well suited for its implementation in neuromorphic NNET. In this context, a promising implementation strategy is the 'nanotechnology approach'. Indeed, emerging nanoscale memory devices, able to mimic synaptic functionality, can be envisioned as ideal elements to provide new needs for information processing and storage. The main characteristic of the neuromorphic NNET relies in their ability to learn. As previously mentioned, as consequence of different experiences (i.e. synaptic activities) learning is achieved by tuning biological synaptic strengths. For such purpose, as it

will be explained in more details in the chapter, it is possible to consider synaptic nanoscale memory devices organized in a cross-bar like structure, well suitable to match the requirement of the high density of integration. A new architecture concept, the so-called CMOL architecture, developed by Likharev and Strukov (Likharev and Strukov, 2005), proposes a structure in which elementary logic units are interconnected in a crossbar topology with local nanoscale memory elements located at the node of the crossbar. Such system would ideally avoid the Von Neumann bottleneck by coupling in parallel memory and computing.

Even if such circuits will necessitate important efforts in material implementation, design, packaging and high level operation and control, Prezioso et al. (Prezioso et al., 2015)) have recently demonstrated the first memristive elements crossbar circuit for classification tasks that paves the way to promising computing systems.

This PhD work follows this approach by developing emerging nanotechnologies that could be the key elements for future realization of neuromorphic NNET. Specifically, the main objective is focused on a crucial aspect addressed by neuromorphic computing: the *Synaptic Plasticity* and how the nanotechnologies can be useful for information computing.

1.2 Neuromorphic NNET with Nano-devices

This section is dedicated to introduce emerging nanoscale memories, the memristive devices, and the basic ingredients necessary to use such components in the context of neuromorphic NNET systems.

1.2.1 Synaptic Nano-devices

Memristor or memristive nanodevices are two-terminal 'memory resistors' that retain internal resistance state according to the history of applied voltage and current. They are simple passive circuit elements, but their function cannot be replicated by any combination of fundamental resistors, capacitors and inductors. From symmetry arguments Chua (Chua, 1971) originally defined memristors as components that link charge and magnetic flux where the 'missing element' provides a functional relation between charge and flux, $d\Theta = Mdq$, where M is the memristance. In 2008, the proposition of physical implementation of a memristor by HP (Strukov et al., 2008) opened the way to an even more realistic implementation of neuromorphic functions: the equation describing memristors are effectively very similar to lots of computing processes observed in biological systems. The most basic mathematical definition of a current-controlled memristor for circuit analysis is the differential form:

$$V = R(W)i \tag{1.5}$$

$$\frac{dW}{dt} = f(W, i, t) \quad (1.6)$$

where W is the state variable of the device and R is a generalized resistance that depends upon the internal state of the device.

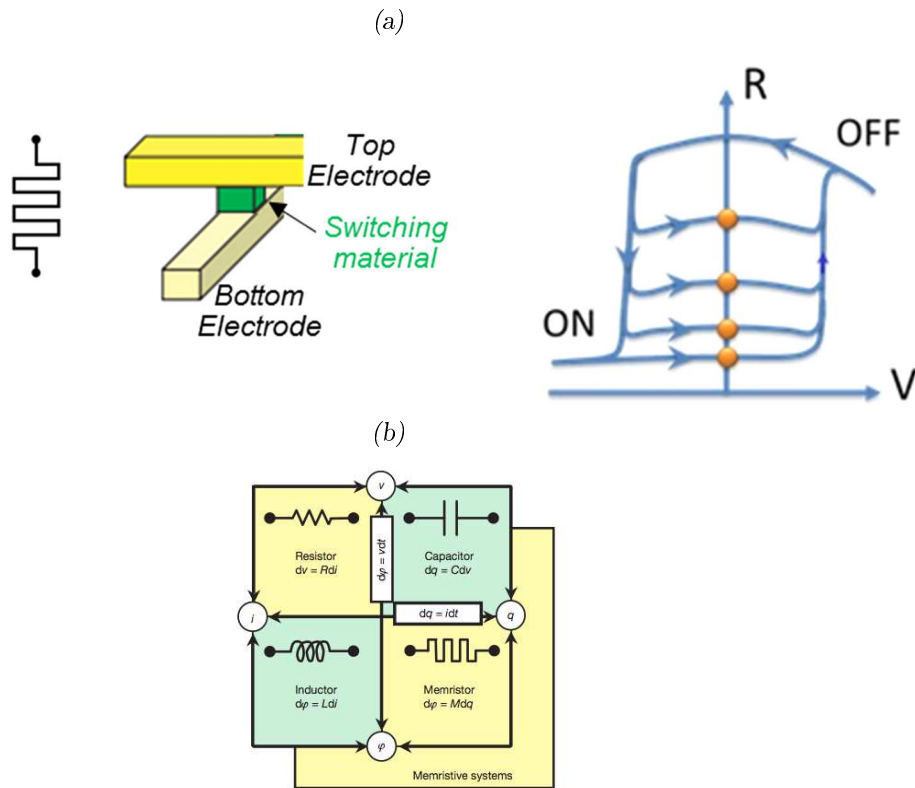


Figure 1.9: (a) Schematic representation of the two terminal memristive cross-point and its non-linear characteristics. (b) The four fundamental two-terminal circuit elements: resistor, capacitor, inductor and memristor.

Depending on the switching materials, different classes of memristive devices can be distinguished. Different physical phenomena that induce the resistance variations can be involved: fuse-antifuse, nano-ionic or thermal processes (RedOx RAMs, Phase Change memories) or 'purely electronic-effects' such as the ferroelectric- or the spin-based devices.

In the next chapter a more detailed ReRAM taxonomy will be presented. The major memristive devices property that will be used in this chapter focuses on their bio-mimetic aspect able to emulate the synaptic behavior at the nano-scale and their capability to learn.

1.2.2 Basic Processing and Data Representation

Neuromorphic Systems Analog Footprint

By analogy with biological systems, information in neuromorphic systems is carried by spikes of voltage with a typical duration in the range of milli-seconds. Starting from this simple observation, a first statement would be to consider neuromorphic networks as digital systems (spike being an all or nothing event). This direction was explored with the concept of neuron as logical unit performing logic operations in a digital way (McCulloch and Pitts, 1943). This short cut is of course hiding very important features observed in biological systems that present many analog properties of fundamental importance for computing. The first footprint of analog characteristics of biological systems can be simply emphasized by considering the analog nature of the synaptic connections bridging neurons. Analog synapses can be described in a first approximation as a tunable linear conductance, defining the synaptic weight between two neurons (this description is largely used in ANNs). Meanwhile, a more bio-realistic description should consider the analog synapse as a complex device-transmitting signal in a non-linear manner (i.e. frequency dependent, for example). The second footprint of analog property is somehow embedded in the time coding strategy used in BNNs: as the neuron is performing time integration of the digital spikes, the signal used for computing (the integrated value of the overall spiking activity) becomes an analog value regulating the spiking activity of the neuron. This second aspect is of particular relevance if we consider dynamical computing (i.e. natural data processing such as vision or sound that present a strong dynamical component). The temporal organization of spikes (or their time occurrence with respect to other spikes in the network) is carrying some analog component of the signal in biological networks. Now combining analog synapses with integrating neurons, the level of non-linearity used by the network for computing the analog signal can be strongly modify. Simple linear filters can be realized with linear synaptic conductance associated to simple integrate and fire (*I&F*) neurons or strongly non-linear systems can be built, based on non-linear synaptic conductance with complex integration at the neuron-level such as leaky integrate and fire (*LIF*) or sigmoid neurons.

Data Encoding in Neuromorphic Systems

Starting from the statement that neuromorphic systems are analog systems, we have to define the appropriate data representation that will match the function to be realized. It should be stressed that data representation in biological systems is still under debate and a detail understanding is still a major challenge that should open new avenues from both a basic understanding and practical computing point of views.

Based on these general considerations, we can now try to present a simplified vision of data-coding in biological systems that could be the basic ingredient for neuromorphic computing (i.e. hardware system implementation).

Rate-coding Scheme

The simplest data representation corresponds to a rate-coding scheme, i.e. the analog value of the signal carrying information (or strength of a stimuli) is associated to the average frequency of the train of pulse. The neuron can then transmit some analog signals through its mean firing rate. Rate-coding data representation is often used for static input stimuli representation but appears to be less popular for time varying stimuli. Indeed, the sampling time interval $\Delta_{sampling}$ used for estimating the mean firing rate imply that events with fast temporal variation (typically variation on a time scale smaller than $\Delta_{sampling}$) cannot be described accurately. For example, the brain's time response to visual stimuli is around $100ms$ and it cannot be accurately described in rate-coding systems that are typically in the range of frequencies from 1 to $100Hz$. A simple example of static data representation is to consider the representation of a static image from a $N \times M$ pixel array of black and white pixels into a $N \times M$ vector $X = (x_1, \dots, x_i, \dots, x_n)$ where x_i can be either 0 or 1 (i.e. min and max frequencies). Then, this concept can be simply extended to analog data (such as pictures with different level of grays) by choosing properly the average firing rate.

Temporal-coding Scheme

A second coding scheme is known as temporal-coding in which each individual pulse of voltage is carrying a logical +1 and a time signature. This time stamp, associated to a given spike, can carry some analog value if we now consider its timing with respect to the other spikes emitted in the network (Maass and Natschläger, 1997). The difficulty in this coding scheme is to precisely define the origin of time for a given spiking event that should depend on the event to be computed. A simple example is to consider a white point passing with a given speed in front of a detector with a black background and producing a pulse of voltage in each pixel of the detector when it is in front of it. By tracking both position of the activated pixel and time stamp attached to it, the dynamic of the event can be encoded. Fig.1.10 shows how the rate- and time-coding scheme can be used to encode an analog signal x_i .

Spike Computing for Neuromorphic Systems

In this chapter, we will use only these two simplified data encoding concepts but it should be stressed that other strategies such as stochastic-coding (i.e. the analog value of the signal is associated to the probability of a

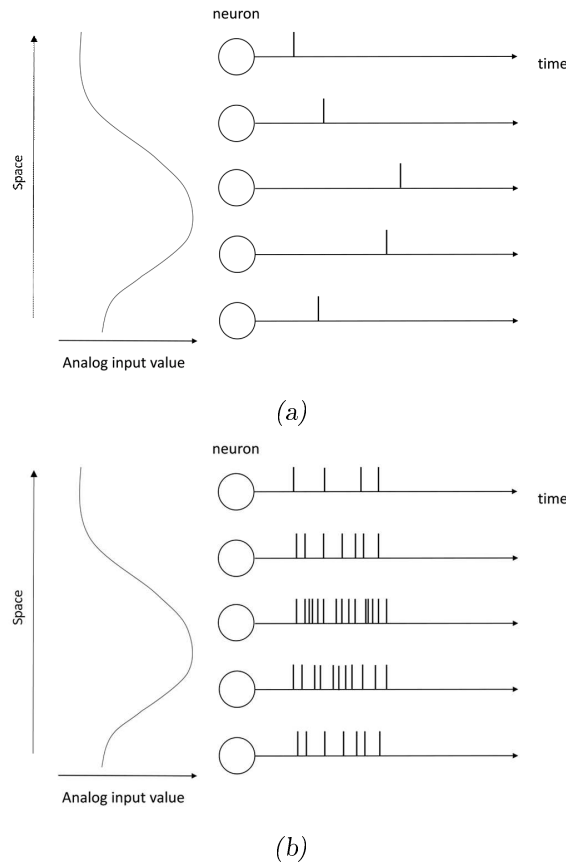


Figure 1.10: Schematic illustration of data encoding schemes. A natural stimulus (such as a visual or auditory cue) is encoded through an input neurons population that sends and encodes the information on time in (a) time-coding scheme and in (b) rate-coding scheme

spike) are potential directions that deserve attention. We should also be aware that both rate- and temporal-coding have been evidenced to coexist in biological systems and both coding strategies can be used for powerful computing implementation. In fact, spike computing has attracted a large attention since the low power performances of biological systems seem to be strongly linked to the spike-coding used in such networks. But it should be emphasized and we should be aware of that translating conventional representation (i.e. digital sequences as in video for example) into spiking signal would most probably miss the roots of low power computing in the biological system. Discretization of time and utilization of synchronous clock is in opposition with continuous time and asynchronous character of biological networks. Spike computing needs to be consider globally, i.e. by considering the full functional network and data encoding principle, from sensors to high level computing elements. In this sense, recent development of bio-inspired sensors such as artificial cochlea (sound detection) or artificial retinas (vi-

sual detection) with event-based representation opens many potentialities for fully spike-based computing where the dynamical aspect of spikes is naturally reproduced.

1.3 Synaptic Plasticity for Information Computing

By remaining in a computational spike-based context, we now focus on how a bio-inspired network, composed in a first approximation of neurons and synapses, can process information (other functional units have to be considered if we want to describe precisely a biological networks such as proteins, glial cells,...). We can roughly categorized spike processing into (i) how spikes are transmitted between neurons, (ii) how spikes propagate along neurons and (iii) how spikes are generated. These two last points can be attributed to 'neuron processing' and more precisely to the response of a biological membrane (the neuron membrane) to electrical or chemical signals. Many associated features such as signal integration, signal restoration or spike generation are of first importance for spike computing but these aspects are beyond the purposes of this chapter. The signal transmission will be the focus of this chapter and different processes involved at the synaptic connection between two neurons will be described. We will concentrate on the dynamical responses observed in chemical synapses that are of interest for spike processing. Such synaptic mechanisms are broadly described as synaptic plasticity: the modification of the synaptic conductance as a function of the neurons activity. The specific synaptic weight values stored in the network are a key ingredient for Neuromorphic Computing. Such synaptic weights distribution is reached through synaptic learning and adaptation and can be described by the different plasticity rules present in the network. Furthermore, it should be noted that all the processes observed in biological synapses and their consequences on information processing are still an ongoing activity and final conclusions are still out of reach. Most probably, the efficiency of biological computing systems lies in a combination of many different features (restricted to the synapse level in this chapter) and our aim is to expose few of them that have been successfully implemented and to discuss their potential interest for computing.

In biology, synaptic plasticity can be attributed to various mechanisms involved in the transmission of the signal between a pre- and post-synaptic neuron, such as neurotransmitter release modification, neurotransmitter recovery in the pre-synaptic connection, receptors sensitivity modification or even structural modification of the synaptic connection (see (Bliss et al., 1993)) for a description of the different mechanisms involved in synaptic plasticity).

It seems important at this stage to make a comprehensive distinction between different approaches used to describe the synaptic plasticity. The first

approach, used to describe the synaptic plasticity, can be identified as a '*causal description*' based on the origin of the synaptic conductance modification. A second one is based on a '*phenomenological description*', in which the temporal evolution (i.e. the dynamics) of the synaptic changes is the key element.

1.3.1 Causal approach: Synaptic Learning versus Synaptic Adaptation

By following the seminal idea of Hebb (Hebb, 1949), a first form of plasticity is the so-called *Synaptic Learning* (Hebbian-type Learning) and can be simply defined as an increase of the synaptic weight when the activity of its pre- and post-neuron increases. Many learning rules have been adapted following this simple idea of '*who fire together, wire together*'. Hebbian-type plasticity implies that the synaptic weight evolution dw_{ij}/dt depends on the product of the activity of the pre-neuron (a_i) and post-neuron (a_j), as follows:

$$\frac{dw_{ij}}{dt} \propto a_i \cdot a_j \quad (1.7)$$

This type of plasticity is defined in biology as *Homosynaptic Plasticity* (Sourdet and Debanne, 1999). Depending on the signal representation, i.e. rate- or temporal-coding, refinement (or particular cases) of Hebb's rule can be formulated such as Spike Rate Dependent Plasticity (SRDP) or Spike Timing Dependent Plasticity (STDP) with neuron activity defined as the mean firing rate or the spike timing, respectively.

A second form of synaptic plasticity can be referred to *Synaptic Adaptation* (where adaptation is in opposition with the notion of learning). In this case, synaptic weight modification depends on the activity of the pre- or post-neuron activity only or on the accumulation of both but in an additive process:

$$\frac{dw_{ij}}{dt} \propto a_i + a_j \quad (1.8)$$

In particular, if the synaptic plasticity depends only on post-activity, such mechanism is defined as *Heterosynaptic Plasticity* otherwise, if it is only pre-neuron activity dependent, it is named *Transmitter-Induced Plasticity*.

Practically, this distinction seems very useful to classify the different synaptic processes that will be implemented and to evaluate their efficiency and contribution to the computing network performances. One major difficulty is that both *Synaptic Learning* and *Synaptic Adaptation* can manifest simultaneously and it becomes much more complicated in practical cases to make a clear distinction between them. In fact, learning in its large sense (i.e. how a network can become functional based on its past experiences) may involve both processes. Also, activity independent weight modification can

also be included to describe synaptic plasticity (for example to describe the slow conductance decay of inactive synapses, as it will be presented in the following paragraph).

1.3.2 Phenomenological approach: STP versus LTP

Another important synaptic plasticity aspect that has to be considered is the *time-scale* involved in the synaptic weight modification. Thus, by focusing on the synaptic plasticity dynamics observed in biological systems, synaptic weight modification can be either permanent (i.e. lasting for months to years) or temporary (i.e. relaxing to its initial state with a characteristic time constant in the milliseconds to hours range). This observation leads to the definition of *Long-Term Plasticity* (LTP) and *Short-Term Plasticity* (STP), respectively. We can notice that the boundary classification into Long-Term (LT) and Short-Term (ST) effects is not well defined and should be considered with respect to the task to be realized. Both STP and LTP can correspond to an increase or decrease of the synaptic efficiency thus leading to the definition of facilitation (or potentiation) and depression, respectively. It is important to notice that there is no one to one equivalence between the concepts of STP, LTP and the notion of Short-Term Memory (STM) and Long-Term Memory (LTM) which corresponds to a higher abstraction level (i.e. memory is then used in the sense of psychology). In this latter case, the information can be recalled from the network (i.e. information that has been memorized) and it cannot be directly associated to a specific set of synaptic weight with a given lifetime and plasticity rule. In fact, how synaptic plasticity can be related to the memorization of the information as well as how it is involved in different time scale of memory (from milliseconds to years) still remains debated.

1.4 Synaptic Plasticity in Nano-devices

Many propositions of synaptic plasticity implementation with nanoscale memory devices have emerged these past years. By referring to the classification previously proposed, two main streams can be identified: the *Causal* description and the *Phenomenological* one. The first one relies on the implementation of the origin of the synaptic plasticity, without necessarily replicating the details of the spike transmission observed in biology. On the contrary, the second strategy has the aim to reproduce accurately the spike transmission properties observed in BNNs, by omitting the origin of the synaptic response, but rather by highlighting its temporal evolution. In this section, we will present examples of practical devices implementation by following these two lines. Of course, a global approach based on a combination of both descriptions (the causal and the phenomenological one),

would be the ideal solution to describe the synaptic weights distribution in ANNs for the future development of neuromorphic computing.

1.4.1 Causal implementation

In this first part, by following the *Causal* description, we will take into account the origin of the synaptic plasticity, without necessarily replicating the details of the spike transmission observed in biology.

Generality: Hebbian Learning

Hebbian Learning has been at the basis of most of the learning strategies explored in neuromorphic computing. Hebbian-type algorithms define how a synaptic weight evolves during the learning experience and set the final weight distribution after the learning experience. Starting from its simplest form, i.e. *'who fire together, wire together'*, a first limitation of Hebbian learning can be evidenced. Indeed, if all synapses of the network are subject to Hebbian learning (Fig.1.11), all synaptic connections should converge to their maximum conductivity after some time of activity since only potentiation is included in this rule, thus destroying the functionality of the network. A first addition to the Hebb's postulate is then to introduce Anti-Hebbian plasticity that would allow to decrease the synaptic weight conductance (i.e. depression) when activity of both pre and post neuron are present (Fig.1.11, green curve). One important consequence of this simple formulation (Hebbian and Anti-Hebbian) is that the final synaptic weight distribution after learning should become bimodal (or binary), i.e. some weights became saturated to their maximum conductance (i.e. fully potentiated) while all the others should saturate to their lowest conductance state (i.e. fully depressed).

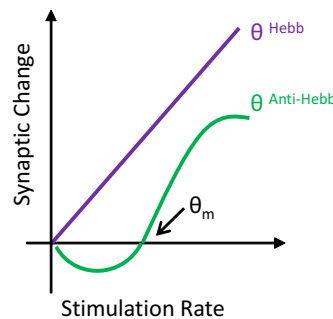


Figure 1.11: Representation of the Hebbian rule (purple) and Hebbian/Anti-Hebbian rule (green) for a constant post-neuron activity when pre-neuron activity is increased (stimulation rate). Addition of Anti-Hebbian learning is a prerequisite in order to prevent all the synaptic weight to reach their maximal conductance.

Time-based computing: Spike Timing Dependent Plasticity

Without reviewing all the different STDP implementation in nanoscale memory devices propositions, we want to highlight some general ideas that are at the origin of this plasticity mechanism. The STDP was introduced by (Abbott and Nelson, 2000) and (Senn et al., 2001) as a refinement of Hebb's rule. In this plasticity form (*Synaptic Learning*), the precise timing of pre- and post-synaptic spikes is taken into account as a key parameter for updating the synaptic weight. In particular, the pre-synaptic spike is required to shortly precede the post-synaptic one to induce potentiation, whereas the reverse timing of pre- and post-synaptic spike elicits depression. To understand how synaptic weights change according to this learning rule, we can focus on the process of synaptic transmission, depicted in Fig. 1.12.

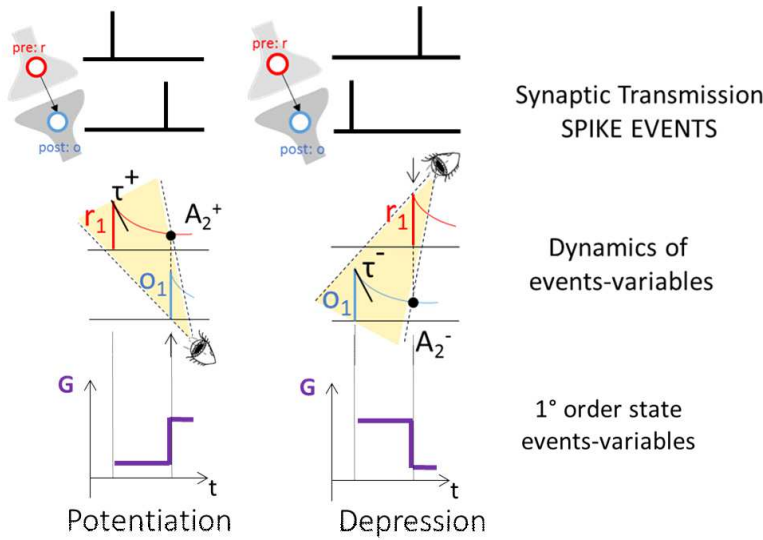


Figure 1.12: Pair-based STDP Learning rules: Long term Potentiation (LTP) is achieved thanks to a constructive pulses overlap respecting the causality principle (Pre-before-Post). On the contrary, if there is no causality correlation between pre and post synaptic spikes, Long term Depression (LTD) is induced.

Whenever a pre-synaptic spike arrives (t_{pre}) at an excitatory synapse, a certain quantity (r_1), for example glutamate, is released into the synaptic cleft and binds to glutamate receptors. Such detector-variable of pre-synaptic events r_1 , increases whenever there is a pre-synaptic spike and decreases back to zero otherwise with a time constant τ^+ . Formally, when $t = t_{pre}$ this gives the following:

$$\frac{dr_1}{dt} = -\frac{r_1(t)}{\tau_+} \quad (1.9)$$

We emphasize that r_1 is an abstract variable (i.e. state-variable). Instead of

glutamate binding, it could describe equally well some other quantity that increases after pre-synaptic spike arrival. If a post-synaptic spike arrives (t_{post}) at the same synapse, and the temporal difference with respect to the pre-synaptic one is not much larger than τ^+ , the interaction between these two spikes will induce potentiation (LTP). As a consequence the synaptic weight $w(t)$ will be updated as follows:

$$w(t) = w(t) + r_1 \cdot A_2^+ \quad (1.10)$$

If a pre-synaptic spike arrives after the post-synaptic one, another detector-variable will be taken into account, relative to post-synaptic events (o_1), as shown in Fig. 1.12. Similarly, we can consider that the dynamics of o_1 can be described by time constant τ_- . Formally, when $t = t_{post}$ this gives the following:

$$\frac{do_1}{dt} = -\frac{o_1(t)}{\tau_-} \quad (1.11)$$

If the temporal difference is not much larger than τ^- , the spikes interaction will induced depression (LTD). As a consequence the synaptic weight $w(t)$ will be updated as follows:

$$w(t) = w(t) - o_1 \cdot A_2^- \quad (1.12)$$

One of the important aspect of STDP is to present both Hebbian and Anti-Hebbian Learning. Replicating the exact biological STDP window (Fig. 1.13a) is not a mandatory condition for implementing interesting learning strategies (other shapes have been reported in biology) while balancing the Hebbian/Anti-Hebbian contribution remains a challenge in order to maintain STDP learning stable. It should be noted that synaptic weight distribution becomes bimodal after some time of network activity if this simple STDP window is implemented (Van Rossum et al., 2000).

The proposition of memristor (Strukov et al., 2008) provides an interesting framework for the implementation of synaptic weights (i.e. analog property of the memory) and for the implementation of STDP in particular. Nanoscale memories or 'memristive devices', as previously introduced, are electrical resistance switches that can retain a state of internal resistance based on the history of applied voltage and the associated memristive formalism. Using such nanoscale devices provide a straightforward implementation of this bio-inspired learning rule. In particular, the modulation of the memristive weight (i.e. the conductance change $\Delta G(W, V)$) is controlled by an internal parameter W , that depends on the physics involved in the memory effect. In most of the memory technologies used for such bio-inspired computational purpose, the internal state-variable W (and consequently the conductance) is controlled through the applied voltage or the current (and implicitly by

its duration). Mathematically, this behavior corresponds to a 1st - order memristor model:

$$\frac{dW}{dt} = f(W, V, t) \quad (1.13)$$

with $I = V \cdot G(W, V)$. Practically, by exploiting memristive devices as synapses, most of the STDP implementation relies on specific engineering of the spikes's shape that convert the time correlation (or anti-correlation) between pre- and post-spikes into a particular voltage that induces a modification of the memory element conductance. The time lag induced by pre-synaptic events, as the r_1 variable in Fig.1.12, determines that the potentiation is converted into a particular voltage across the memristor in order to induce an increase of conductance when a post-synaptic spike interact with it. Similarly, time lag induced by post-synaptic events in analogy with o_1 variable in Fig.1.12, will induce depression in form voltage across the memristor when interacting with a pre-synaptic spike.

First implementation was proposed by Snider (Snider, 2008) with time multiplexing approach (Fig. 1.13b), in which, although the spike signal is far from bio-realistic, the STDP window can be reproduced with high fidelity. Fig.1.13c shows another successful STDP implementation with non bio-realistic signal in a Phase Change Memory device (Kuzum et al., 2011a). Depending on the particular memory device considered, different encoding strategies were proposed with the same principle of input/output voltage correlation in which the STDP window mapped to bio-realistic observations. Recently, by going deeper in the memresistive switching behavior (i.e. by considering a higher order memristive model), STDP was proposed through even more bio-realistic pulse shape (Kim et al., 2015), as it will be explained in the section 1.4.1.

Rate based computing: the BCM learning rule

While the STDP learning rule has been largely investigated these past years, another refinement of the Hebb's rule can be formulated in the case of rate coding approaches. Bienenstock, Cooper and Munroe (Bienenstock et al., 1982) proposed in the 80s the BCM learning rule with the concept of '*sliding threshold*' that ensures to maintain the weight distribution bounded and thus avoiding unlimited depression and potentiation resulting from simple Hebbian learning implementation. The BCM learning rule can be simply formalize as follow:

$$\frac{dw_{ij}}{dt} = \varphi(a_j(t)) \cdot a_i(t) - \epsilon w_{ij} \quad (1.14)$$

Where w_{ij} is the synaptic conductance of the synapse bridging the pre-neuron i and post neuron j , a_i and a_j are the pre- and post-neuron activity,

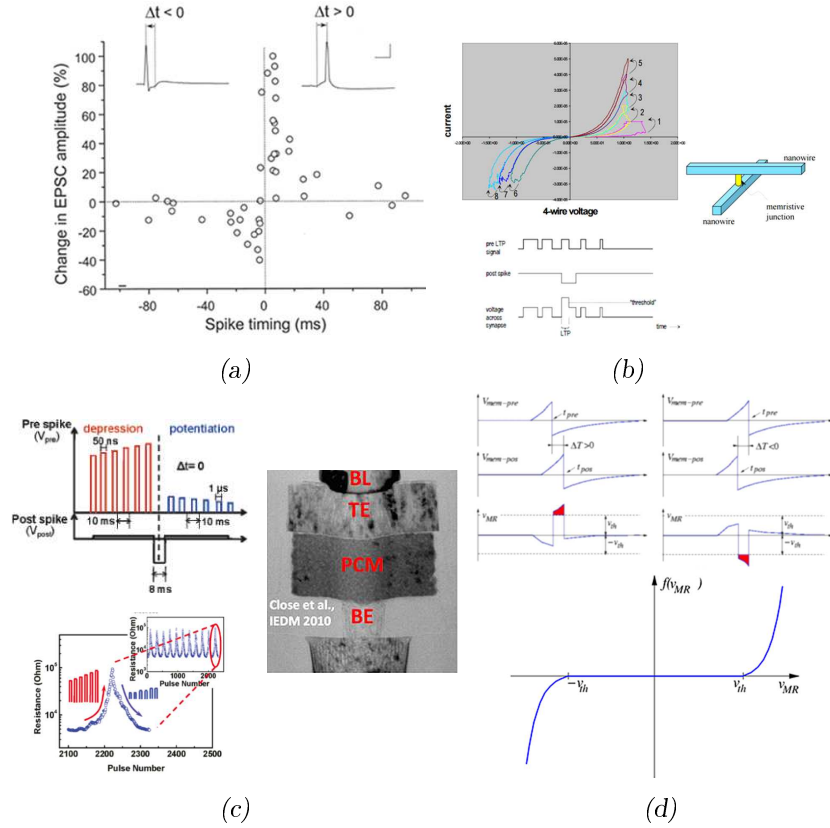


Figure 1.13: (a) Biological STDP window from (Bi and Poo, 1998a). In all three cases: (b), (c) and (d), the particular shape of the signal applied at the input (pre-neuron) and output (post-neuron) of the memory element induces a particular effective voltage that induces potentiation (increase of conductance) or depression (decrease of conductance) reproducing the STDP window of (a). (b) First proposition of STDP implementation in nanoscale bipolar memory devices where time multiplexing approach was considered. In this case, the STDP window can be reproduced with high fidelity while the spike signal is far from bio-realistic. (c) Implementation of STDP in unipolar PCM devices. Still the STDP window can be reproduced precisely while the signal is not bio-realistic. (d) Proposition of STDP implementation with bipolar memristor. Both the STDP window and pulse shape are mapped to bio-realistic observations.

respectively, ϵ is a constant related to a slow decaying component of all the synaptic weights (this term appears to become important in special cases, see (Bienenstock et al., 1982) but not mandatory) and φ a scalar function parametrized as follow:

$$\varphi(a_j) < 0 \text{ for } a_j < \theta_m \quad \& \quad \varphi(a_j) > 0 \text{ for } a_j > \theta_m$$

where θ_m is a threshold function that depends on the mean activity of the post neuron. A first order analysis can be realized on this simple learning rule. (i) Both Hebbian-type learning (product between a_i and a_j) and

adaptation (through the small decay function that is not related to pre- and post-neuron activity) are present in this rule. (ii) The threshold ensures that both Hebbian and Anti-Hebbian plasticity can be obtained through the scalar function φ that can take positive and negative values (potentiation and depression). (iii) Thus, the '*sliding threshold effect*' corresponds to the displacement of the threshold as a function of the post-neuron activity and is a key ingredient to prevent the synaptic weight distribution to become bimodal. Indeed, if the mean post-neuron activity is high, any pre-neuron activity should induce potentiation (most probably). If now θ_m is increased when the mean post-neuron activity increases, it will increase the probability of depression or at least reduce the magnitude of potentiation and consequently limit the potentiation of the weight.

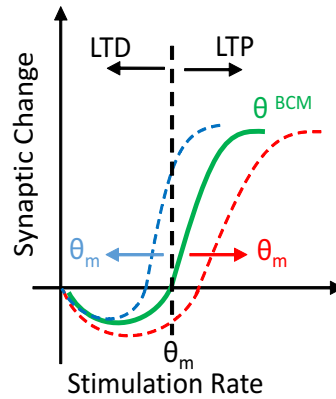


Figure 1.14: BCM learning rule representation. The synaptic weight modification is represented as a function of pre-neuron activity for a fixed post-neuron activity. The sliding threshold depends on the mean post neuron activity, i.e θ_m is increased if a_j increases while θ_m is decreased if a_j decreases, thus preventing unlimited synaptic weight modification.

The BCM learning rule was initially proposed for rate coding approaches and was measured in BNNs in the Long-Term regime of the synaptic plasticity. The BCM learning rule has been shown to maximize the selectivity of the post-neuron (Bienenstock et al., 1982). Only few works have demonstrated partially the BCM rule in nanoscale memory devices with some limitations. Lim et al. (Lim et al., 2013) proposed to describe the weight saturation in TiO_2 electrochemical cells subject to rate-based input. This work demonstrated the sliding threshold effect describing the saturation of the weight during potentiation and depression but did not reproduce the Hebbian/Anti-Hebbian transition. Ziegler et al. (Ziegler et al., 2015) demonstrate the sliding threshold effect in the Long-Term regime but without considering explicitly a rate coding approach, i.e. neuron activity was simply associated to the pre- and post-neuron voltages. Kim et al. (Kim et al., 2015) proposed an adaptation of the BCM rule in second order memristor, as it will

be presented in the next section, but in a transmitter-induced plasticity context, thus missing the Hebbian-type plasticity initially proposed in the BCM framework. Future works are expected to provide stronger analogy with BCM rule, both from a phenomenological point of view (i.e. bio-realistic rate coding implementation) and from a causal point of view (i.e. reproducing all the aspects of the BCM rule).

Reconciliation of BCM with STDP

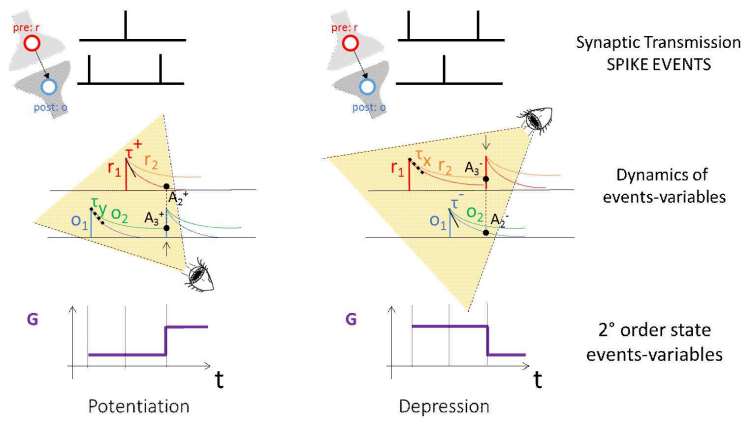
On the one hand, the importance of individual spikes and their respective timing can only be described in the context of STDP. The time response in the visual cortex being in the order of $100ms$, rate-coding approaches are unlikely to offer a convenient description of such processes while time-coding could. On the other hand, simple STDP function misses the rate coding property observed in BNNs and conveniently described in the context of the BCM. More precisely, in the case of pair-based STDP, both potentiation and depression are expected to decrease as the activity mean frequency of the network is increased while BNNs show opposite trend. Izhikevich et al. (Izhikevich et al., 2003) proposed that classical pair-based STDP, implemented with the nearest-neighbor spike interactions, can be mapped to the BCM rule. However, their model failed to capture the frequency dependence (Sjöström et al., 2001) if pairs of spikes are presented at different frequencies (Clopath et al., 2010).

From a neurocomputational point of view, Gjorgjieva et al. (Gjorgjieva et al., 2011) proposed a triplet STDP model based on the interactions of three consecutive spikes as generalization of the BCM theory. This model is able to describe plasticity experiments that the classical pair-based STDP rule has failed to capture and is sensitive to higher-order spatio-temporal correlations, which exist in natural stimuli and have been measured in the brain. As done for the pair-based case, to understand how synaptic weights change according to this learning rule, we can focus on the process of synaptic transmission, depicted in Fig. 1.16.

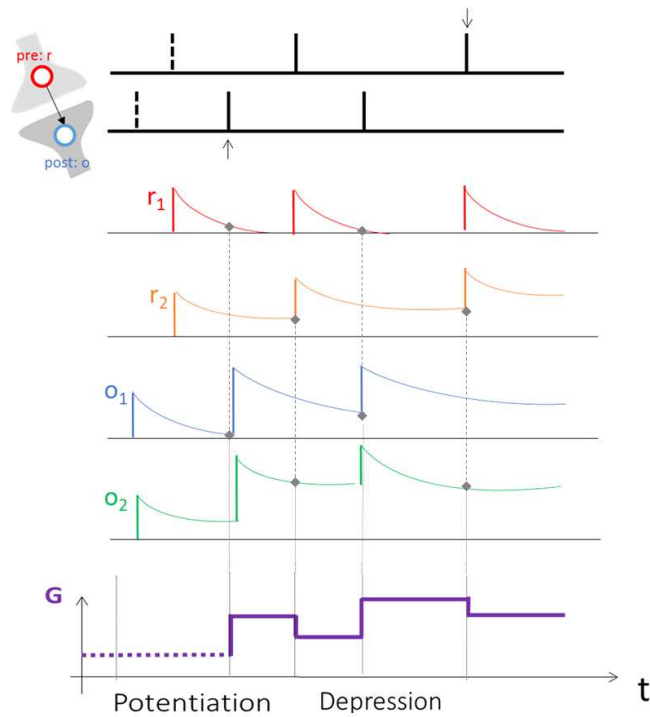
Instead of having only one process triggered by a pre-synaptic spike, it is possible to consider several different quantities, which increase in the presence of a pre-synaptic spike. We can thus consider, r_1 and r_2 two different detectors-variables of pre-synaptic events and their dynamics can be described with two time constant τ_+ and τ_x ($\tau_x > \tau_+$). Formally, when $t = t_{pre}$ this gives the following:

$$\frac{dr_1}{dt} = -\frac{r_1(t)}{\tau_+} \quad \& \quad \frac{dr_2}{dt} = -\frac{r_2(t)}{\tau_x} \quad (1.15)$$

Similarly, we can consider, o_1 and o_2 two different detectors-variables of post-synaptic events and their dynamics can be described with two time constant



(a)



(b)

Figure 1.15: Triplet-based STDP Learning rules. (a) Synaptic weight potentiation (LTP) is achieved thanks to a (post-pre-post) spike iterations, as a results the relative time lag of the detector-variables dynamics. Similarly a synaptic weight depression (LTD) is induced with a (pre-post-pre) spike interactions. (b) Synaptic weight evolution in function of time correlation of pre- and post- spikes.

τ_- and τ_y ($\tau_y > \tau_-$). Formally, when $t = t_{post}$ this gives the following:

$$\frac{do_1}{dt} = -\frac{o_1(t)}{\tau_-} \quad \& \quad \frac{do_2}{dt} = -\frac{o_2(t)}{\tau_y} \quad (1.16)$$

We assume that the weight increases after post-synaptic spike arrival by an amount that is proportional to the value of the pre-synaptic variable r_1 but depends also on the value of the second post-synaptic detector o_2 . Hence, post-synaptic spike arrival at time t_{post} triggers a change given by the following:

$$w(t) = w(t) + r_1(t) \cdot (A_2^+ + A_3^+ o_2(t)) \quad (1.17)$$

Similarly, a pre-synaptic spike at time t_{pre} triggers a change that depends on the post-synaptic variable o_1 and the second pre-synaptic variable r_2 as follows:

$$w(t) = w(t) - o_1(t) \cdot (A_2^- + A_3^- r_2(t)) \quad (1.18)$$

As done previously, we emphasize that r_1 , r_2 , o_1 , and o_2 are abstract variables that not identify with specific biophysical quantities. Biological candidates of detectors of pre-synaptic events are, for example, the amount of glutamate bound ((Buonomano and Karmarkar, 2002)) or the number of NMDA receptors in an activated state ((Senn et al., 2001)). Postsynaptic detectors o_1 , and o_2 could represent the influx of calcium concentration through voltage-gated Ca^{2+} channels and NMDA channels ((Buonomano and Karmarkar, 2002)) or the number of secondary messengers in a deactivated state of the NMDA receptor ((Senn et al., 2001)).

A possible solution to implement this generalized rule that embraces both BCM theory and STDP has been proposed by Mayr et al. (Mayr et al., 2012) for the first time in $BiFeO_3$ memristive devices. They succeeded in implementing triplet STDP through a more complex spikes's shape engineering that encodes the time interaction between more than two pulses into a particular voltage able to induce a modification of the memory element conductance. Triplet STDP rule has been also performed by Williamson et al. (Williamson et al., 2013) in asymmetric TiO_2 memristor in hybrid neuron/memristor system. Subramaniam et al (Subramaniam et al., 2013) have used triplet STDP rule in a compact electronic circuit in which neuron consists of a spiking soma circuit fabricated with nanocrystalline-silicon thin film transistors (ns-Si TFTs) with nanoparticle TFT-based Short-Term Memory device and HfO_2 memristor as synapse.

Another generalized description, in which both time- and rate-coding approaches are taken into account at the same time and implemented in an amorphous InGaZnO memristor, has been proposed by Wang et al. (Wang et al., 2012). In addition to the conventional ion migration induced by the application of pulse of voltage, another physical mechanism of the device

operation occurs: the gradient of the ions concentration, leading to the appearance of ion diffusion, resulting in an additional state-variable. Kim et al. (Kim et al., 2015) recently proposed a 2^{nd} - order memristor that offers an interesting solution towards this goal of reconciliation of various learning mechanisms in a single memory device.

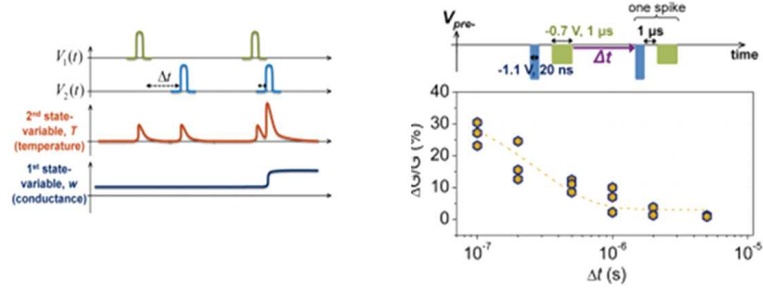
Mathematically, in analogy to the previous definition, a 2^{nd} - order memristor model can be described as:

$$\frac{dW_1}{dt} = f_1(W_1, W_2, V, t) \quad \& \quad \frac{dW_2}{dt} = f_2(W_1, W_2, V, t) \quad (1.19)$$

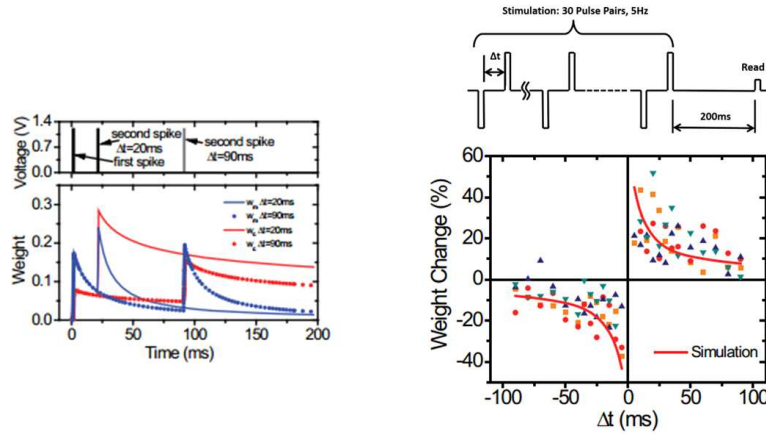
with $I = V \cdot G(W_1, W_2, V, t)$ and implemented with a simple non-overlapping pulses protocol for the synaptic weight modulation.

The interest behind this higher-order memristor description, is to provide additional parameters that will ensure some other higher-order interaction between pulses (i.e. more than two), while the pair-based interaction is preserved. More precisely, as shown in Fig. 1.16a, the temperature has been proposed as 2^{nd} - order state-variable that exhibits short-term dynamics and naturally encodes information on this relative timing of synapse activity. By exploiting these two state-variables (i.e. the conductance and the temperature), STDP has been implemented, as it is shown in Fig. 1.16a. Specifically, the first 'heating' spike elicits an increase in the device temperature by Joule effect regardless of the pulses polarity, which then tends naturally to relax after the removal of the stimulation, then temporal summation of the thermal effect can occur and can induce an additional increment in the temperature of the device if the second 'programming' spike is applied before T has decayed to its resting value. Longer time interval will induce a small conductance change because of the heat dissipation responsible to a lower residual T when the second spike is applied. Thus, the amount of the conductance change (long-term dynamics) can be tuned by the relative timing of the pulses encoded in the short-term dynamics of 2^{nd} *state-variable* (i.e. the temperature T).

Du et al. (Du et al., 2015) have proposed another 2^{nd} - order memristor model. Also in this case, two state variables are used to described an oxide-based memristor. The first one, as in the previous example, directly determines the device conductance (i.e. the synaptic weight). Specifically this first-state variable represents the area of the conducting channel region in the oxide memristor thus directly affecting the device conductance. The second-state variable represents the oxygen vacancy mobility in the film which directly affects the dynamics of the first-state variable (conductance) but only indirectly modulates the device conductance (Fig. 1.16a). Equivalently to T, the w is increased by application of a pulse and then tends to relax to an initial value and affects the 1^{st} -state variable by increasing the amount of conductance change in a short-time scale.



(a)



(b)

Figure 1.16: 2^{nd} order memristor model. (a) On the right: the modulated 2^{nd} order state-variable exhibits short-term dynamics and naturally encodes information on the relative timing and synapse activity. On the left: STDP implementation: memristor conductance change as a function of only two spikes (i.e. each spike consists of a programming pulse and a heating pulse) (Kim et al., 2015). (b) On the right: Simulation results illustrating how the short-term behavior affected long-term weight change. The difference in long-term weight is caused by the different values of residue of the second-state variable at the moment when the second pulse is applied. The first and the second state variable under two conditions (interval between two pulses $\Delta t = 20, 90$ ms) are shown. On the left: Memristor weight change as a function of the relative timing between the pre- and post-synaptic pulses without pulses overlapping (STDP implementation). (Du et al., 2015).

By exploiting this 2nd - order memristor model Du et al. (Du et al., 2015) have demonstrated that STDP can be implemented in oxide-based memristor by simple nonoverlapping pre- and post-synaptic spike pairs, rather than through the engineering of the pulse's shape (Fig 1.16b).

In neurobiology the timing information is intrinsically embedded in the internal synaptic mechanisms. Malenka et Bear (Malenka and Bear, 2004) have demonstrated that together with the neurotransmitter dynamics in the presynaptic connection, secondary internal state-variables, such as the natural decay of the post-synaptic calcium ion (Ca^{2+}) concentration, are involved in the synaptic weight modulation and the synaptic plasticity that can be achieved by simple nonoverlapping spikes and tuned by synaptic activity (i.e. rate- and timing-dependent spikes) which brings an interesting analogy between biological processes and material implementation described above (Gjorgjieva et al., 2011).

The hypothesis that several synaptic functions manifest simultaneously and are interrelated at synaptic level seems accepted by different scientific communities. Recent biological studies indicate that multiple plasticity mechanisms contribute to cerebellum-dependent learning (Boyden et al., 2004). Multiple plasticity mechanisms may provide the flexibility required to store memories over different time-scales encoding the dynamics involved. From a computational point of view, Zenke et al. (Zenke et al., 2015) have recently proposed the idea to use multiple plasticity mechanisms at different time scales. Instead of focusing on particular and local learning schemes, their strategy aims to create memory and learning functions through interplay of multiple plasticity mechanisms. By following this trend of multi-scale plasticity mechanisms Mayr et al. (Mayr et al., 2013) have realized a VLSI implementation in which short-term-, long-term-, and meta-plasticity interact each other at different timescales to tune the overall synapse weights distribution.

1.4.2 Phenomenological implementation

In this section, we will follow the second synaptic description approach: the *phenomenological* one. The spike transmission properties observed in BNNs will be presented as a function of the temporal evolution of the synaptic weight.

STP in a single Memristive Nano-Devices

As previously mentioned, the *transmitter-induced plasticity* is a particular form of synaptic adaptation that depends only on pre-neuron activity. From a phenomenological point of view, such plasticity is most often observed on short time scale, thus belonging to the class of STP. As shown in Fig. 1.17b this STP regime is frequency-dependent and can be used to modulate

the synaptic weights distribution as a function of network activity. From a biological view point, a phenomenological model of frequency-dependent synaptic transmission was used to describe such synaptic response in STP regime (Markram et al., 1997). The primary synaptic parameters are the absolute synaptic efficacy (A), the utilization of synaptic efficacy (U), recovery from depression (τ_{rec}) and the recovery from facilitation (τ_{facil}) (Fig. 1.17a). In this model, synaptic response is then dependent on the finite amount of neuro-transmitter resources in the pre-synaptic neuron and their respective dynamics (utilization and recovery) and on the absolute efficacy of the synaptic connection which could depends on post synaptic neuron receptors sensitivity or synaptic connection for example. The most likely biophysical mechanisms underlying changes in the value of these synaptic parameters were considered (Markram et al., 1997).

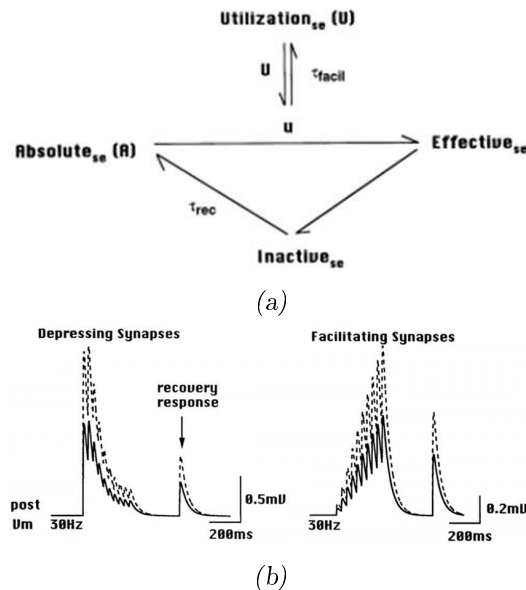


Figure 1.17: Phenomenological model of frequency-dependent synaptic transmission.(a) Each AP utilizes U a fraction of the available/recovered synaptic efficacy R . When an AP arrives, U is increased by an amplitude of u and becomes a variable, $U1$. (b)Phenomenology of changing absolute synaptic efficacy parameter A . On the left: synaptic responses of depressing synapses when A is increased 1.7-fold. On the right: synaptic responses of facilitating synapses when A is increased 1.7-fold. Adapted from (Markram et al., 1998).

If we consider a temporal-coding approach in which pulses are considered as discrete events, STP can be evidenced through the notion of Paired Pulse Facilitation (PPF) corresponding to the enhancement of a pulse transmission when this latter closely follows a prior impulse. The counter effect (i.e. corresponding to depression) is referred to as Paired Pulse Depression (PPD). If we now focus on rate-coding approaches, facilitation and depression can

be simply described as a high-pass and low-pass filter. Depending on the mean firing rate of the synapse, signal can be enhanced or depressed when pre-neuron frequency is increased. A simple material implementation of such mechanism can be realized through passive RC circuits. It turns out that RC circuits with time constants in the milliseconds to seconds range leads to very high capacity with large area (even at low current operation) that are a severe limitation for hardware implementation of STP. Different alternative approaches can realize more efficiently such dynamical effects by taking advantage of physical mechanisms present in nanoscale memory devices.

The first proposition of STP with nano-devices was realized in a nanoparticles/organic memory transistor (NOMFET) (Alibart et al., 2010). The basic principle of this device is equivalent to a floating gate transistor. Charges, stored in the nanoparticles, modify the channel conductivity via coulomb repulsion between the carriers (holes) and the charged nanoparticles. The particularity of this device relies on the leaky memory behavior: charges stored in the nanoparticles tend to relax with a characteristic time constant in the 100 to 200 ms range (Desbief et al., 2015). When the NOMFET is connected in a diode like configuration (Fig.1.18a), each input spike (with a negative voltage value) charges the nanoparticles and decrease the NOMFET conductivity. Between pulses, charges escape from the nanoparticles and the conductivity relaxes toward its resting value. By analogy with biology, this device mimics the STP observed in depressing synapses (Fig.1.18c and Fig.1.18b) and described by (Abbott et al., 1997). As a matter of comparison, this synaptic functionality is realized with a single memory transistor while its implementation in Si based technologies (i.e. CMOS) required 7 transistors (Boegerhausen et al., 2003).

STP has been also demonstrated in two terminal devices that would ensure higher devices density when integrated into complex systems. Equivalently, STP in two terminal devices is implemented by taking advantage of the volatility of the different memory technologies (i.e. low retention of the state that is often a drawback in pure memory applications). Redox systems based on Electro-Chemical Memory cell (ECM) (Ohno et al., 2011a) or Valence Change Memory (VCM)(Yang et al., 2013b) and (Chang et al., 2011a) have demonstrated STP with a facilitating behavior. In such devices, Short-Term Plasticity is ensured by the low stability of the conducting filaments that tends to dissolve, thus relaxing the device toward the insulating state. TiO_2 VCM cells have been reported with both facilitating and depressing behavior (Lim et al., 2013) with relaxation related to oxydo-reduction counter reaction. Protonic devices have demonstrated STP with depressing functionality due to proton recovery latency from atmosphere required to restore the proton concentration and conductivity (Deng et al., 2013).

In terms of functionality, (Abbott et al., 1997) has demonstrated that depressing synapses with STP act as a gain control device (at high frequency, i.e. high synaptic activity, the synaptic weight is decreased, thus leading

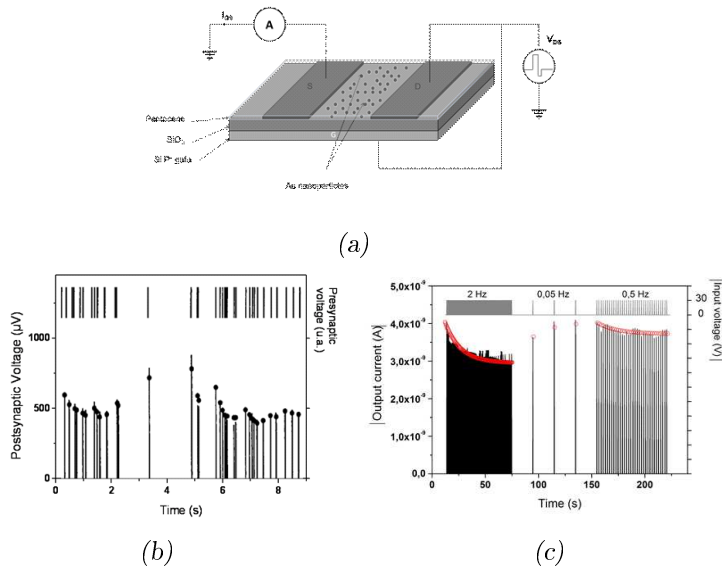


Figure 1.18: STP implementation in a NOMFET. (a) Schematic representation of the NOMFET and pseudo-two terminal connections of the device. (c) Comparison between the frequency-dependent post-synaptic potential response of a depressing synapse (lines) and the iterative model of Varela et al. (dots), adapted from (Varela et al., 1997), as a function of the frequency of the pre-synaptic input signal. (b) Response (drain current) of NOMFET with L/W ratio of $12\mu\text{m}/113\mu\text{m}$ and NP size of 5nm to sequences of spikes at different frequencies (pulse voltage $V_p = -30\text{V}$)

to a lowering of the signal when activity becomes too important). More generally, STP (both depressing and facilitating) provides a very important frequency coding property (as depicted in Fig.1.18c and Fig.1.18b that could play a major role in the processing of spike-rate coded information). Indeed, if a simple Integrate and Fire neuron ($I\&F$) is associated with static weight (with no dependence with spike frequency), the computing node (i.e. neuron and synapses) is only a linear filter (linear combination of the different input) while STP turns the node to non-linear. This property (i.e. locally induced non-linearity in spike signal transmission) has been used to implement reservoir-computing approaches as proposed by Maass (Buonomano and Maass, 2009) with the Liquid State Machine and could be an important property of biological systems for computing.

Co-existence of STP and LTP in the same Memristive Nano-Device

If the contribution of Short-Term and Long-Term processes to computing is not completely understood in biological systems, both STP and LTP effects in synaptic connections has been evidenced and should play a crucial role. A first approach is to consider that repetition of Short-Term effects should lead to Long-Term modification in the synaptic connections.

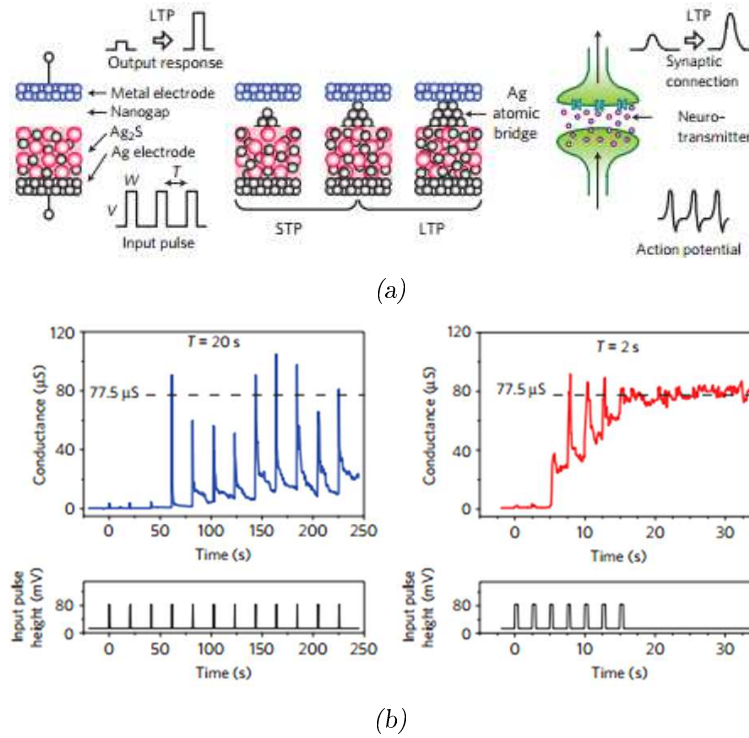


Figure 1.19: STP and LTP implementation in an ECM cell depending on input pulse repetition time. (a) Schematic representation of the Ag_2S ECM cell and the signal transmission of a biological signal. Application of input pulses causes the precipitation of Ag atoms from the Ag_2S electrode, resulting in the formation of an Ag atomic bridge between the Ag_2S electrode and a counter metal electrode. When the precipitated Ag atoms do not form a bridge, the ECM cell works in the STP regime. After an atomic bridge is formed, it works as LTP. (b) Frequent stimulation ($T = 2s$) causes long-term enhancement in the strength of the synaptic connection while short-term enhancement is induced at lower frequency ($T = 20s$) (Ohno et al., 2011a).

This behavior would explain the important hypothesis of memory consolidation in the sense of psychology (Lamprecht and LeDoux, 2004). Ohno (Ohno et al., 2011a) reported for the first time the transition from Short-Term to Long-Term Potentiation in atomic bridge technology (Fig.1.19). Considering again the *Transmitter-Induced plasticity* dependent on the pre-synaptic activity (associated to spike rate in this case), the synaptic conductivity is increased due to the formation of a silver (Ag) filament across the insulating gap. While for low frequency, the bridge tends to relax between pulses; higher frequencies lead to a strong filament that maintains the device in the ON state. These results suggest a critical size of the bridging filament in order to maintain the conductive state stable (i.e. providing a LTP of the synaptic connection).

Similar results have been obtained in a variety of memory devices where fil-

amentary switching displayed two regimes of volatility. Wang et al. (Wang et al., 2012) have shown that STP-to-LTP transition can occur through repeated 'stimulation' training. By stimulating sequentially an oxide-based memristive device with 100 positive pulses, the synaptic weight gradually increases with the number of pulses. Once the applied voltage is removed, a spontaneous decay of synaptic weight occurs in the case of no external inputs. The synaptic weight does not relax to the initial state, but stabilizes at a mid-state, which means that the change of synaptic weight consists of two parts: STP and LTP.

Chang et al. (Chang et al., 2011b) have evidenced a continuous evolution of the volatility as a function of the conductivity level of the device in WO_3 oxide cells attributed to the competition between oxygen vacancies drift (creation of conductive path across the device) and lateral diffusion (disruption of the conducting filament). Another description of these two regimes of volatility could be associated to a competition between surface and volume energies in the conductive filament (Yuan et al., 2010).

Conflict between causal and phenomenological description

If this concept of ST to LT transition has been well demonstrated in variety of nanoscale memory devices, we have to emphasize that they were all reported in the context of *Transmitter-Induced plasticity* (more precisely corresponding to the synaptic adaptation, a non-Hebbian plasticity form). In biology, the facilitating processes observed in short time scale (i.e. *transmitter-Induced STP*) and associated to an increase of neurotransmitter release probability during a burst of spike (i.e. corresponding to an increase of synaptic efficiency at high frequency spiking rate) is additive with LTP (Bliss et al., 1993) that could be associated to a Hebbian-type plasticity involving both pre- and post-neuron activity. In other words, a causal description will make a clear distinction between the origin of ST- and LT- plasticity while a phenomenological description (Fig. 1.17b) will not. Indeed, during high frequency burst of spikes associated to *Transmitter-Induced plasticity*, the firing of the post-neuron is favored and should lead to both pre- and post-activity, thus leading to Hebbian-type LTP. In the case of the neuromorphic implementation described above, the transition between STP and LTP is associated to a single parameter (such as the mean firing rate of the pre-neuron) and both ST and LT regimes cannot be uncorrelated (i.e. ST will lead to LT regime). The device state will move sequentially from one regime to another one via *Transmitter-Induced plasticity* only. It should be noted that this effect induces some restriction in terms of (i) network configurability, since non-Hebbian and Hebbian-type learning cannot be dissociated, and (ii) network functionality, since the synaptic connection moves from a non-linear conductance in its ST regime (i.e. frequency dependent) to a linear conductance in its LT regime. Alternative approaches are still needed as

proposed by Cantley et al. (Cantley et al., 2011) where Short-Term processes and Long-Term Processes are realized by two different devices (leaky floating gate transistor and non-volatile 2 terminals devices) in order to match the complexity of biological synapses.

1.5 Scope and Approach of this Work

In order to match the efficiency of biological systems, synaptic functionalities should be realized with a dedicated technology well suited for its implementation. In addition, going into the detail of the Synaptic Plasticity (processes observed in biological synapses corresponding to a modification of the synaptic weight as a function of its spiking history) requires more functionalities than an ideal non-volatile memory that will hardly implement these dynamical processes (or at the cost of additional overhead to emulate the dynamical functions). Indeed, computation in biological systems are a combination between long term synaptic processes (Long Term Potentiation and Depression, LTP and LTD) and short term mechanisms (Short Term Plasticity, STP) that contributes to the processing and storage of asynchronous spike signals.

In this multidisciplinary context can be placed the research of this PhD thesis that targets to develop specific nanoscale dynamic memory devices to replicate some of the key mechanisms observed in biological systems with a clear objective: bringing more functionality in a single component in order to reduce circuit overhead cost and improve circuit performances.

1.6 Discussion and Perspectives

In this chapter, we have presented the background and key motivations behind the research field in which this PhD is placed. The main objective of the *Neuro-Inspired Computing Paradigm* is to build ANNs whose organizing principles are based on those of BNNs. We looked at the state-of-the neuro-morphic NNET and different hw implementation directions.

We then focused on the functional aspect of the nanotechnology approach by highlighting the impact of emerging nanoscale memory devices, suitable to implement some aspect of *Synaptic Plasticity*, the key concept for the purpose of this work.

In the last section of the chapter, we briefly discussed the scope and the overall strategy adopted for the research conducted during this PhD thesis. In the following chapter we will focus on a practical aspect of such neuromorphic nanoscale devices and how they could be integrated in future neuromorphic NNET.

Chapter 2

Neuromorphic NNET with Filamentary Switching

*"Happiness is nothing more
than good health and a bad memory."*
Albert Schweitzer (1875-1965)

2.1 Introduction

In the previous chapter, the impact of emerging nanoscale memory devices has been presented from a functional point of view, in which their ability to implement some aspect of *Synaptic Plasticity* offers a promising and interesting way to enrich and enhance future bio-inspired information computing systems. The main objective of this chapter is to emphasize a practical aspect of such synaptic devices: how they can be used and integrated in neuromorphic systems.

In the first part of this chapter we will introduce resistive switching memories by focusing on a particular class of filamentary-type technology, the electro-chemical metallization (ECM) cells.

The second part of this chapter follows this research line at circuit and system level in which a review state of the art of integration strategies will be presented with structural and functional affinities between ANNs and BNNs.

Finally we will discuss about pros and cons of the integration approaches presented from a computational point of view by pointing out engineering efforts that have to be done and are required for future neuromorphic NNET hardware with emerging nanotechnologies.

2.1.1 Resistive Memories

Resistive memories (or RRAM cells) refer to any technology that uses varying cell resistance to store information. A resistive switching memory cell is a two terminal structure generally built with a MIM structure, composed of an insulating or resistive material 'I' sandwiched between two electron conductors 'M'. Information storage is based on multiples electrical resistance states. By applying an appropriate voltage, the 'MIM' cell can be switched between a high-resistance state (HRS or OFF-state) and a low-resistance state (LRS or ON-state). Switching from OFF-state to ON-state is called the SET process, while switching from ON-state to OFF-state is called the RESET process. These two states can represent the logic values '1' and '0', respectively. In other words, RRAM cell is able to induce a change of

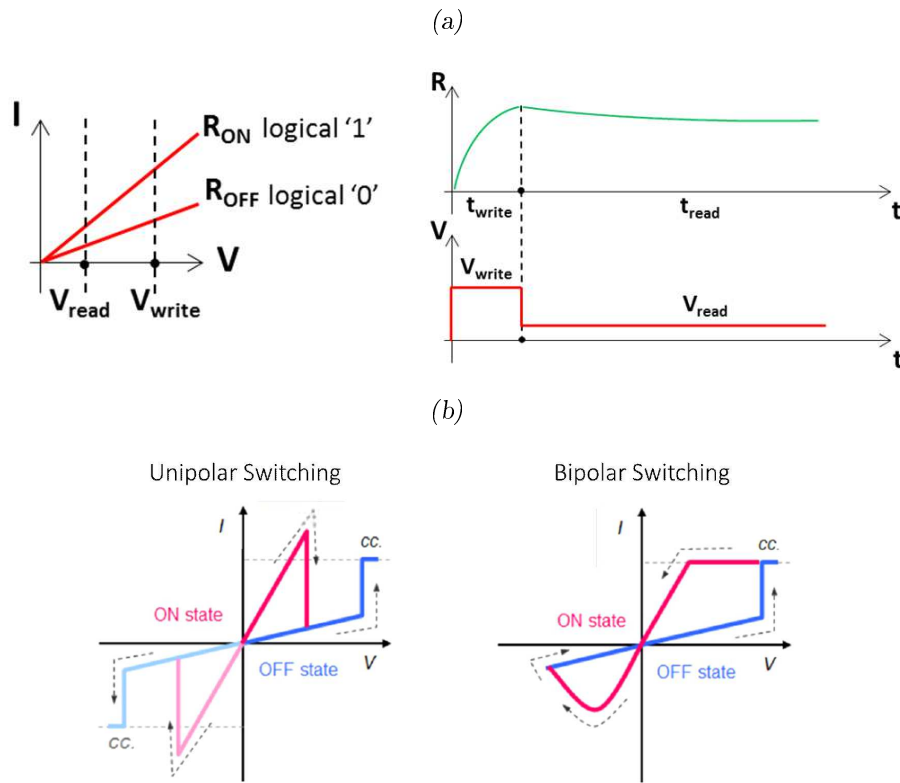


Figure 2.1: Resistive Memories. (a) Sketch of RRAM cells basic principle. (b) Unipolar and Bipolar Switching mechanisms. Adapted from (Kawai et al., 2012).

resistivity to discriminate two (or more) resistance states (1 bit of information, or more) by electrical stress V_{write} and to read the information stored at V_{read} (i.e. probing current). In principle, non-volatile memories (NVM) have to be able to store information fast (i.e. programming at $V_{write} \sim 1V$ in $\sim 1ns$), to induce a large change of resistance (R_{ON}/R_{OFF}) and to address information for a very long time without changing its state (i.e. good

retention: stable read at $V_{read} \sim 0.1V$ for ~ 10 years) (Fig. 2.1a). Such hard requirement represents the well-known time-voltage dilemma and whatever is the physical mechanism originating the change of resistance, non linearity in the $R(V)$ relationship is needed in order to prevent lost of information (i.e. $\Delta R(V) \sim e^V$). Depending on the switching mechanism, the resistive memories can be classified as unipolar and bipolar (Fig. 2.1b). In the unipolar case, the memory state of the system can be switched by successive application of electric stress of either the same or opposite polarities. In contrast, the bipolar memories can be toggled between the memory states by application of successive electric stress of alternate polarity, i.e. one polarity is used to switch from HRS to LRS, and the opposite one is used to switch back into HRS (Waser and Aono, 2007).

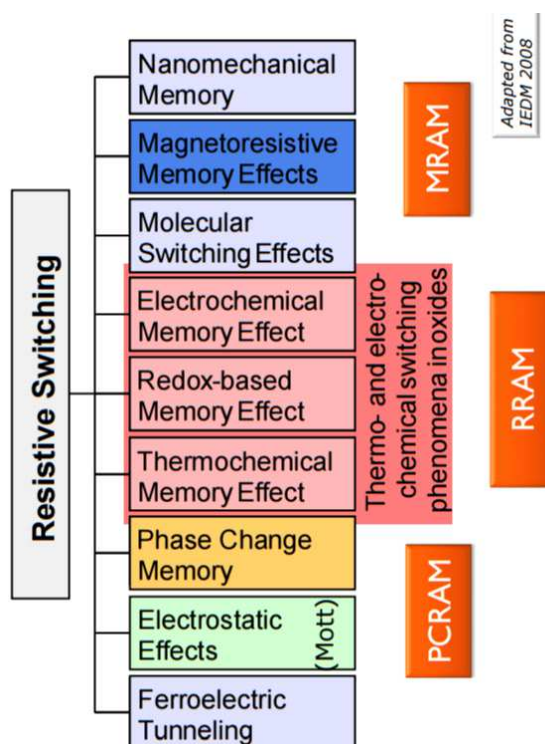


Figure 2.2: Taxonomy of Resistive Switching Memories which are considered for NVM applications. Adapted from (Waiser et al., 2008).

These last years RRAM technologies have experienced an increase of interest as a promising solution for storage and memory. Indeed, such emerging memory devices can offer potential alternative for flash technology or Dynamic Random Access Memory (DRAM) thanks to their fast switching performances ($< 1ns$) (Torrezan et al., 2011), high retention and cycling endurance (Miao et al., 2012), scalability ($< 10nm$) (Govoreanu et al., 2011) and Back End Of Line (BEOL) integration potential (Xia et al., 2009). A

large variety of physical phenomena lead to non-volatile resistive switching memory effects and Fig. 2.2 shows a taxonomy of RRAM technologies (Waser et al., 2009). By considering the spatial dimensions of different physical phenomena involved to induce the resistance variations, we can divided RRAM cells in three big classes (Fig 2.3): 1D 'Filamentary Switching', that will be studied in more details in the next section, 2D 'Interfacial Switching' and 3D 'Bulk Transition'.

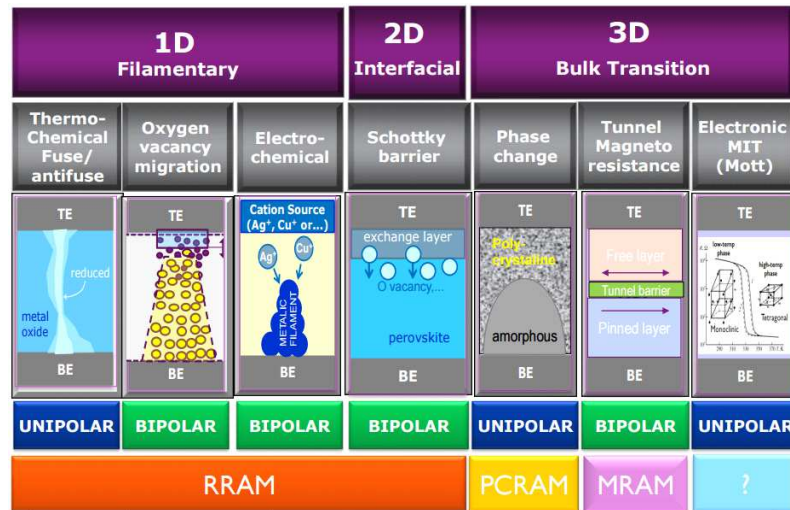


Figure 2.3: Resistive Switching Memories classification based on the modulation geometry. Adapted from (Wouters et al., 2012).

3D Bulk Transition Switching

One of the most 'mature' technology competing to replace flash memory is the phase change memory (PCM) (Wong et al., 2010). It belongs to the 3D 'Bulk Transition' class and uses a semiconductor alloy that can be changed between an ordered, crystalline phase having a low electrical resistance (LRS) to a disordered, amorphous phase with much higher electrical resistance (HRS). As fabricated, the PCM is in the crystalline, low-resistance state because the processing temperature of the metal interconnect layers is sufficient to crystallize the phase change material. To reset the PCM cell into the amorphous phase, the programming region is first melted and then quenched rapidly by applying a large electrical current pulse for a short time period. Doing so leaves a region of amorphous, highly resistive material in the PCM cell. To set the PCM cell into the crystalline phase, a medium electrical current pulse is applied to anneal the programming region at a temperature between the crystallization temperature and the melting temperature for a time period long enough to crystallize. PCM cells are programmed and read by applying electrical pulses which change temperature accordingly. PCM

allow to access intermediate-resistance states by controlling the dimensions of the least resistive current paths within the memory element, thus multi-bit operations can be implemented. PCM multi-layered architecture (3-D stackable memory) in which multiple layers of memory elements are stacked one above the other, sharing the addressing and sense-amplification circuitry among the memory layers, has also been demonstrated (Lu, 2009).

Another RRAM technology belonging to the 3D 'Bulk Transition' class is the magnetic tunnel junction (MTJ), which is a component consisting of two ferromagnets separated by a thin insulator. If the insulating layer is thin enough (typically few nanometers), electrons can tunnel from one ferromagnet into the other when a bias voltage is applied between the two metal electrodes. Here, the transition between HRS and LRS is controlled by the tunneling current that depends on the relative orientation of magnetizations of the two ferromagnetic layers, which can be changed by an applied magnetic field. This phenomenon is called tunneling magnetoresistance (TMR) which is a consequence of spin-dependent tunneling (Bibes et al., 2010).

2D Interfacial Switching

RRAM cells, in which the resistive switching takes place at the interface between the metal electrode and the semiconducting oxide, belong to the 2D interfacial switching class. In order to understand such switching mechanism, it can be useful to clarify the origin of the contact resistance, which can be changed by applying an electric field. Since the memory cell has a capacitor-like structure composed of insulating or semiconducting oxides sandwiched between metal electrodes, a Schottky barrier seems to be the most probable origin of the contact resistance.

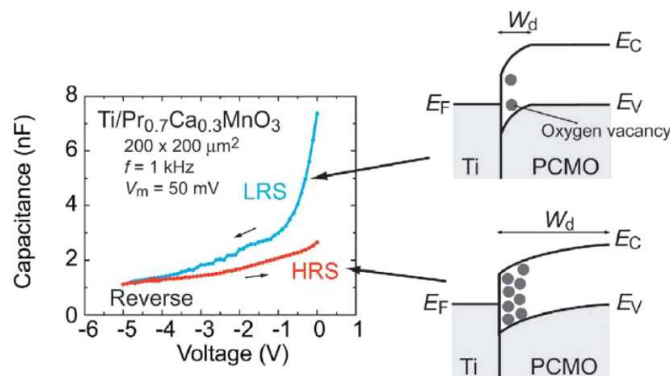


Figure 2.4: CV curves under reverse bias for a Ti/PCMO/SRO cell show hysteretic behavior. This indicates that the depletion layer width W_d at the Ti/PCMO interface is altered by applying an electric field. Adapted from (Sawa et al., 2005).

In the conventional Schottky model, the amplitude of the contact resistance

is attributed to the potential profile of the barrier, i.e. the depletion layer, and it can be determined from a capacitance-voltage (CV) curve (Fig. 2.4). In this case the change in the Schottky barrier height under voltages of different polarities due to the charge trapping/detrapping at the interface is responsible for the different resistance states. This switching mechanism is usually related to bipolar-type switching behavior observed in semiconducting perovskite oxides (Baikalov et al., 2002). A number of models have been proposed for the driving mechanism in resistive switching involving an interface-type switching, such as electrochemical migration of oxygen vacancies (Tsui et al., 2004), trapping of charge carrier (hole or electron) (Sawa, 2008), and Mott transition induced by carriers doped at the interface (Oka et al., 2003).

At this stage, an useful consideration relies on the difference between the interface and the filamentary resistive switching, that can be understood by considering the area dependence of the cell resistance (Fig.2.5).

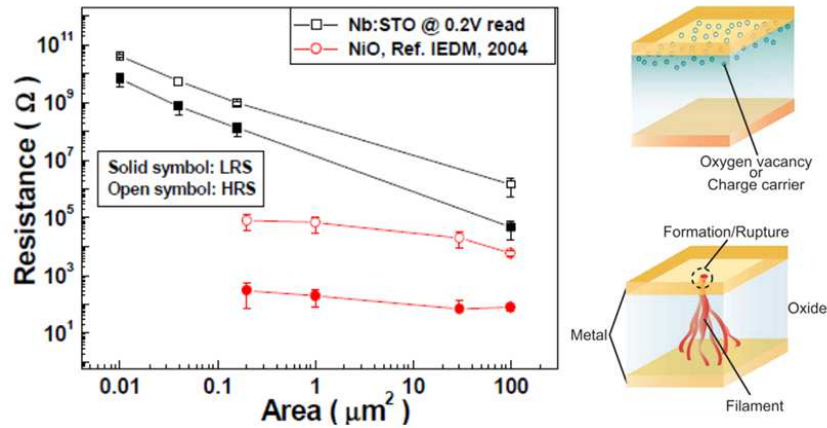


Figure 2.5: Area dependence of resistance values in high and low resistance states for Nb -doped $SrTiO_3$ ($Nb:STO$) and NiO memory cells. The resistance of $Nb:STO$ memory cells depends linearly on the area, suggesting that the resistive switching takes place over the entire area of the interface (interfacial-type). The resistance of NiO memory cells is almost independent of the area, suggesting that resistive switching is a local phenomenon (filamentary-type). Adapted from (Sawa et al., 2005).

A cell composed of semiconducting Nb -doped $SrTiO_3$ has a resistance that is inversely proportional to the cell area, whereas that of an insulating NiO cell is much less dependent on cell area (Sim et al., 2005). These results indicate that resistive switching in the Nb -doped $SrTiO_3$ cell takes place over the whole area of the cell, i.e. the entire interface, whereas switching occurs locally in the NiO cell through the formation of filamentary conducting paths.

2.2 Filamentary Switching: nano-device level

Filamentary Switching, as previously mentioned, is a particular resistive memory technology in which information storage is realized through two resistance states via formation (LRS) and rupture (HRS) of conducting filaments (CFs) across two metallic electrodes. There are various kinds of CFs which have different ingredients and origins. Generally CFs are metallic conductive channels formed through electrochemical metallization of the electrochemically active electrode metal, such as *Ag* and *Cu* in the case of CBRAM (Valov et al., 2011a) or through thermochemical metallization process in the insulator, such as *Ni* filament in *NiO* (Lee et al., 2009). Another CFs formation mechanisms is caused by the migration of oxygen ions, as in *TiO₂*, *Ta₂O₅*, and *Fe₃O₄* (OxRAM). An example of excellent scalability ($< 10\text{nm}$), superior switching speed/energy and high endurance ($> 10^{10}$) is given by the Hafnium dioxide *HfO₂*, one of the most promising candidate for memory applications (Govoreanu et al., 2011) and (Lee et al., 2010).

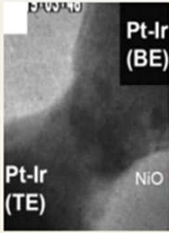
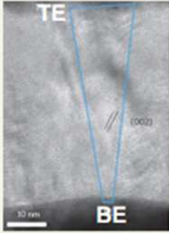
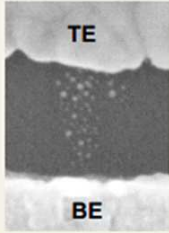
Category	Thermo-chemical RAM	OxRAM	CBRAM
Typical switching cause	Anion (oxygen ion) migration	Anion (oxygen ion) migration	Cation (metal ion) migration
Typical switching material	Transition metal oxide (TMO)	Transition metal oxide (TMO)	Solid electrolyte, semiconductor
Switching polarity	Unipolar	Bipolar	Bipolar
Device example			
Physical effect	1D Filament		

Figure 2.6: 1D RRAM filamentary switching classification based on the switching cause, material and polarity. Images from (Fujii et al., 2011), (Kwon et al., 2010), and (Yang et al., 2012b).

Thus, three different filamentary switching categories can be individualized: thermo-chemical RRAM based on fuse-antifuse mechanism (typically for unipolar switching), OxRAM based on migration of oxygen vacancies V_{O}^{2+} under electric field (typically for bipolar switching) and CBRAM in which the bipolar switching is controlled by the migration of metal ions under electric field (Fig. 2.6).

Thermo-Chemical Memories - TCM cells

In Thermo-chemical memories, the thermochemical process dominates over the electrochemical process and, hence the switching is inherently unipolar. During the switching, local temperature gradients occur and lead to local stoichiometry variations and redox reactions, which results in a change of the electronic conductivity. TCM switching is observed in all metal oxides which present a high resistivity in the most oxidized state and a much lower resistivity in reduced states, such as NiO , $ZrOx$, $TiOx$ and SiO_2 . Often the same metal is used for top and bottom electrodes, because, in contrast to OxRAM and CBRAM cells, an asymmetry is not required. Historically, the most prominent TCM cell is constituted by $Pt/NiO/Pt$ stack (Ielmini et al., 2011).

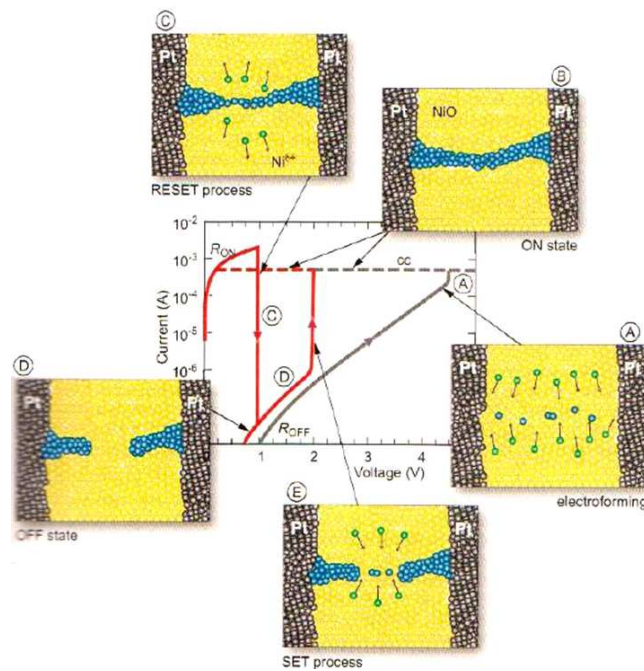


Figure 2.7: Unipolar IV characteristic of a $Pt/NiO/Pt$ cell. The states and processes are sketched in cross sections: (A) Electroforming process; (B) ON-state; (C) RESET process; (D) OFF-state; (E) SET process. Adapted from (Waser et al., 2012).

Typical IV characteristic of the electroforming and subsequent switching cycles in a NiO -based TCM cell are shown in Fig. 2.7. During the initial forming step a sudden increase occurs due to a thermoelectric breakdown and the creation of CF in the cell. During this electroforming cycle and during all SET operations, a current compliance I_c needs to be applied in order to limit the thermal effects and to establish the desired R_{ON} value, corresponding to a particular filament diameter. During RESET cycle, the I_c is released and the current overshoot ruptures and partially dissolves the

conductive filament. The ON- state of a TCM cell displays generally a linear (ohmic) current-voltage relation, that is consistent with the metal-rich conducting filament formed during the electroforming or SET process. On the contrary, the OFF- state present an exponential IV characteristic at high voltages. This behavior may be attributed either to a Schottky emission at the metal/oxide interface or to a thermally assisted conduction, due to a high density of defect states in the band gap (Jung et al., 2007). Regarding the scaling prospects, thermal engineering is a key issue in optimizing the cell design. Another challenge for TCM cells scaling is related to the reduction of the RESET current. This latter depends on the SET current and the relative R_{ON} value. Thus, as demonstrated by Ielmini et al. (Ielmini et al., 2011) for NiO systems, microscopically the filament diameter need to be thin to obtain sufficiently low RESET currents.

Valence-Change Memories - VCM cells

In the case of valence change memories, also called VCM or OxRAM cells, the MIM system consists of an active interface (active electrode (AE)) at which the switching takes places, a mixed ionic-electronic conducting (MIEC) layer and an ohmic counter electrode (CE).

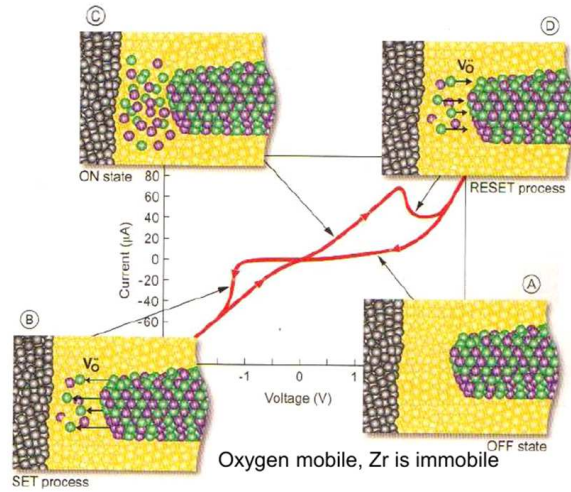


Figure 2.8: IV characteristic of a $Pt/ZrO_x/Zr$ cell. The Pt represents the active electrode (AE) and the Zr the ohmic electrode (CE). The green spheres indicates the oxygen vacancies, the purple ones indicate the Zr ions in a low valence state. The states and processes are sketched in cross sections: (A) OFF-state; (B) SET process; (C) ON-state; (D) RESET process. Adapted from (Waser et al., 2012).

A typical IV characteristics is shown in Fig. 2.8 where switching mechanisms are also depicted in the different steps. In the OFF-state the filament consists of n-conducting MIEC oxide and a potential barrier in front of the left electrode. Upon application of a negative bias, oxygen vacancies from the

filament are attracted into the barrier, which results in a significant decrease of the barrier height and width due to a local reduction process, which turns the cell into the ON-state. For the RESET, a positive bias is applied to the AE which repels the oxygen vacancies, leading to a local re-oxidation and turns the cell into the OFF-state. We can distinguish three different approaches relative to the OxRAM cell system. In any case, an electrode material with a low oxygen affinity (such as Pt, TiN, Ir) is used as AE. For the oxide-thin film, there are different strategies: (i) the oxide film is homogeneous monolayer (e.g. TiO_x, TaO_x, HfO_x). In the case of fully oxidized oxide, an electroforming is necessary to generate an oxygen deficiency at the side of the CE. (ii) Homogeneous bi-layer, in which an oxygen deficient layer is deposited on the ohmic electrode and a second one, fully oxidized of few nanometer thickness of the same oxide is processed on the side of the active electrode. Classical examples are: TiO_2/TiO_{x-2} Yang et al. (2012a) or $Ta_2O_5/TaO_{x<2.5}$ systems (Lee et al., 2011) systems.

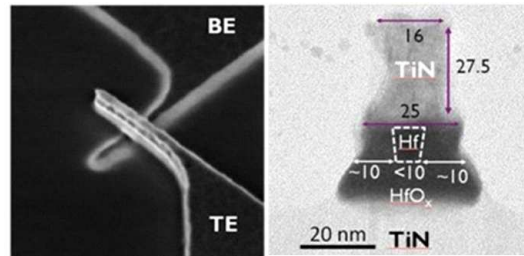


Figure 2.9: $TiN/HfO_x/Hf$ cell with a cross-section of $10nm \times 10nm$ (left) SEM-view of a crossbar resistive element and (right) high-resolution TEM cross-sections of the top-electrode. Adapted from (Govoreanu et al., 2010).

(iii) A heterogeneous bi-layer concept presents the second layer made from another oxide with a larger formation energy and/or larger band gap. Examples are Al_2O_3/TiO_{x-2} (Kwon et al., 2010) or HfO_2/TiO_{x-2} (Miao et al., 2012) (Fig.2.9).

Electro-Chemical Memories - ECM cells

The Electro-Chemical Metallization (ECM) cells, also called conductive bridging random access memory (CBRAM) cells, belongs to the 1D filamentary-family RRAM. The electrochemical metallization mechanism or memory effect relies on the dissolution and deposition of an active electrode metal such as Ag or Cu to perform the resistive switching operation (Waser and Aono, 2007). The ECM cell, similarly to the VCM cell, presents a MIM configuration that consists of an electrode made from an electrochemically active metal (AE), such as Ag or Cu , an electrochemically inert counter electrode (CE), such as $Pt, Ir, W,$ or Au , and a thin film of a solid electrolyte 'T', such as amorphous $GeSe_{2+x}$, disordered and amorphous sulfides Ag_2S

or oxides $a\text{-Si}$ and SiO_2 , sandwiched between both electrodes.

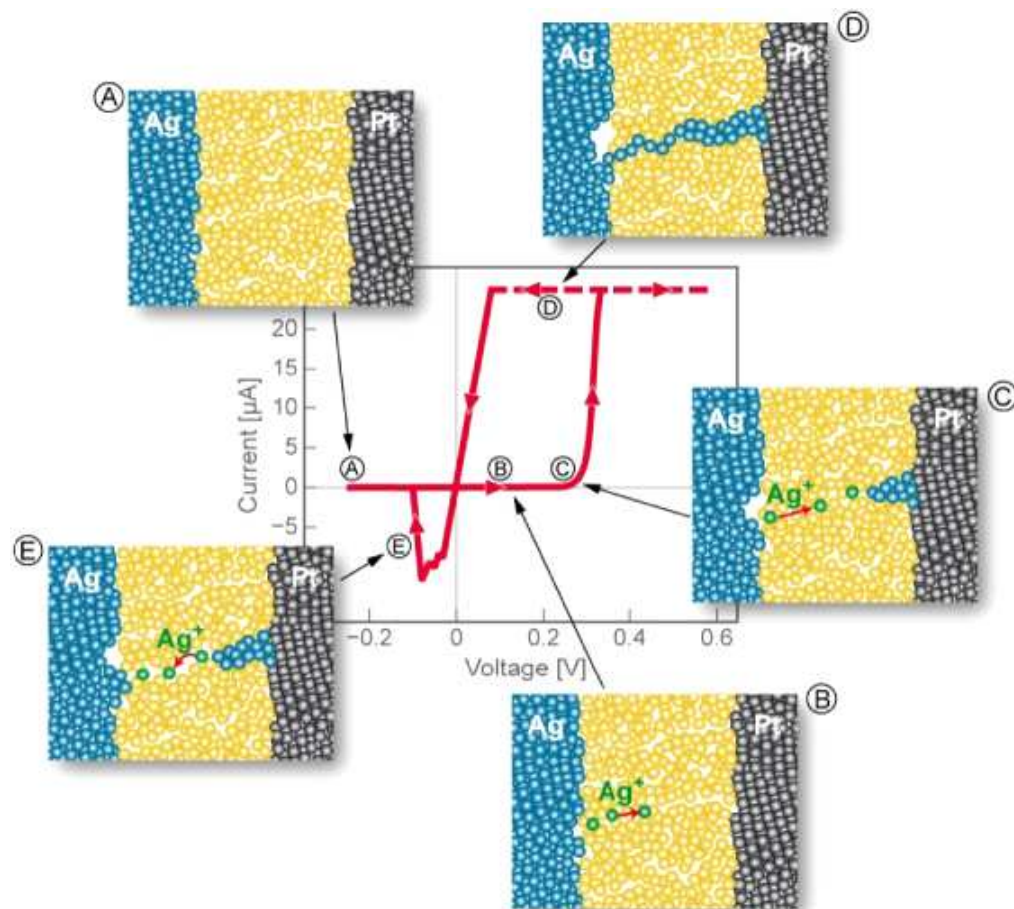


Figure 2.10: IV characteristic of a $\text{Ag}/\text{Ag} - \text{Ge} - \text{Se}/\text{Pt}$ electrochemical cell. The states and processes are sketched in cross sections: (A) OFF-state; (B) and (C) SET process; (D) ON-state; (E) RESET process. Adapted from (Valov et al., 2011).

The basic principle of operation of an ECM memory cell and the bipolar switching, i.e typical IV characteristic of ECM cell under a triangular voltage sweep, are shown in Fig.2.10. During the SET operation, a positive voltage is applied at the AE (in this case Ag electrode) and Ag is oxidized to Ag^+ ions which drift towards the opposite electrode because of the electric field. At the CE acting as a cathode, an electro-chemical reduction and an electro-crystallization of Ag on the surface of the inert electrode takes place. This process results in the formation of a Ag filament, which grows towards the active electrode until an electrical contact is established which defines the ON-state and where further filament growth is limited by a compliance current I_c . To RESET the cell a voltage with the opposite polarity is applied which leads to the dissolution of the filament (OFF-state).

A fundamental characteristic of ECM cells relies on the kinetics of the SET process: the SET speed depends strongly on voltage. In particular, the switching speed can be limited by several factors: (i) the anodic dissolution of the metal component (*Ag* or *Cu*); (ii) the transport of the metal ions through the electrolyte; (iii) the reduction of the ions to metal at the cathode (limited by charge transfer, diffusion or electro-crystallization); and (iv) the filament growth. The first three factors may lead to an exponential relationship between the switching speed and the applied voltage. As shown in Fig. 2.11a for the case of *Ag/Ge_xS_y/W* cell, a clear exponential relationship between the switching time and the switching voltage is observed for $V_{SET} > 0.4V$, while for long switching times a critical SET voltage seems to be approached. These results are complemented and confirmed by the results presented in Fig. 2.11b in which the switching experiments have been performed by a variation of the sweep rate in CV experiments. The pronounced exponential relation and a critical threshold voltage for the SET process, explain how the voltage-time dilemma is overcome in ECM cells.

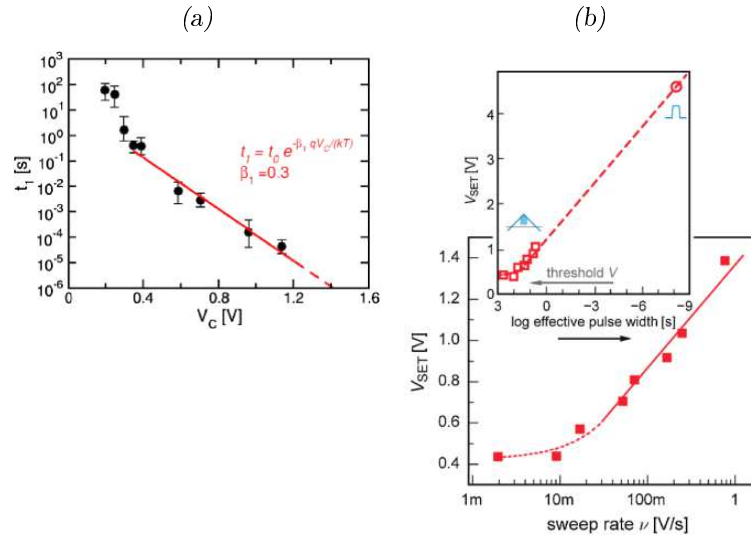


Figure 2.11: Nonlinear Switching Dynamics. (a) Illustration of the exponential dependence of the SET speed, t_1 , on the SET voltage, V_C for *Ag/Ge_xS_y/W* cell. Adapted from (Russo et al., 2009). (b) Switching voltage, V_{SET} as a function of sweep rate, measured on a *Cu/SiO₂/Ir* cell with an oxide thickness of $15nm$. The inset puts the data into relation with a pulse measurement (dot) using a pulse width of $10ns$. The sweep rates of the triangular sweep experiments are converted into effective pulse width defined by one quarter of a full period.

In order to analyze such rate limiting step, which finally controls the overall non-linear kinetic, let's now focus the attention on theoretical aspects of the SET switching speed of ECM cells. The kinetics of the electrode reactions (either at the anode or at the cathode) can often limit the overall reaction rate and therefore the ECM switching speed performance. Indeed, a rate-

limiting step is represented by an electron transfer reaction which occurs at the metal/ion conductor interface. Fig.2.13 shows the energy diagram of this process with and without applied voltage. The left potential well described the potential energy of a metal atom M at the metal surface. The right potential well is attributed to a metal ion M^{z+} close to the metal surface. To oxidize a surface metal atom, a free activation energy η_{ox} is required, whereas for the reduction of M^{z+} and succeeding deposition the free energy η_{red} is required. If a negative potential is applied to the electrode, its fermi energy is increased by $-ze\eta$, where z is the number of exchanged electrons and η represents the additional voltage applied at the interface (the so-called overpotential). Thus, the activation energies for redox-process are changed and reduction process is favored over the oxidation process. The change of the activation energy is proportional to the applied overpotential.

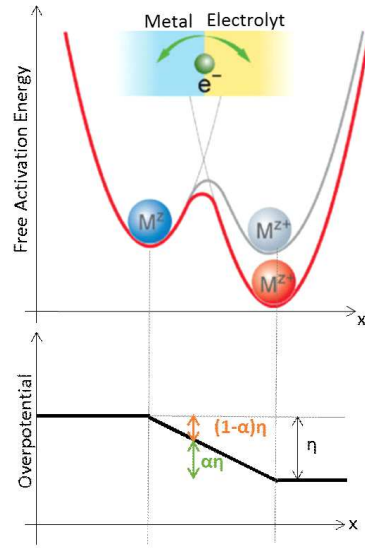


Figure 2.12: Sketch of the energy diagram of a charge transfer reaction at the interface between a metal cation at the surface of the metal electrode and a corresponding cation within the electrolyte as described by the Butler-Volmer equation. The grey line represents the situation with an overpotential η is applied (Waser and Aono, 2007).

This quantity limits the rate of the electrode reaction and can be defined as $\eta = V_{eq} - V$, difference between the equilibrium Nernst-potential V_{eq} of the metal M and the actual electrode potential V . The current density for the charge transfer across the electrode electrolyte interfaces during the cathodic reduction, leading to the metal deposition and the counter reaction representing the anodic oxidation and dissolution of M in the ECM cells, can be described by the so-called Butler-Volmer equation:

$$J = J_0 \left[\exp\left(\frac{(1-\alpha)ze\eta}{kT}\right) - \exp\left(-\frac{\alpha ze\eta}{kT}\right) \right] \quad (2.1)$$

where J_0 is the exchange current density which is strongly temperature dependent, α is the proportionality factor relative to the charge transfer coefficient and represents that part of the overpotential η being used for lowering the activation energy of the particular process. The right term of Eq.2.1 described the reduction, whereas the left term corresponds to the oxidation reaction. For low $\eta \ll k_B T / ze$ the current becomes linearly dependent on η , whereas this relation becomes exponential for high overpotential $\eta \gg k_B T / ze$.

To summarize, an important observation from this analysis is that speed and retention are related properties that are usually traded off against each other. Engineering the devices with a smaller activation energy for redox reaction will improve the switching speed but it will also reduce the retention time. High stress conditions (electric fields and/or elevated temperatures) seem crucial for nonlinear ionic transport, but they can be detrimental to other performance characteristics of memristive devices. Furthermore, strong non-linearity in ion transport may also lead to larger dispersion in switching dynamics. Identifying and engineering nonlinear ion transport mechanisms that do not impact endurance and variations in the memristive devices is therefore an important goal.

2.2.1 Experimental Evidences

As previously mentioned, a first indirect proof regarding the filamentary switching nature can be done by considering the independence of switching parameters (resistance in the two resistive states, SET/RESET currents) on the device area (Fig. 2.5). Due to the very localized nature of the CF, and to its reduced diameter (estimated to be around 10 - 100 nm), it is very hard to analyze its composition. Different techniques have been investigated to identify nanoscale CFs and their formation/rupture dynamics, thereby significantly enhancing the understanding of filamentary switching mechanisms. Several researchers attempted the task, trying to solve several doubts and debates. For instance, there is not yet a clear evidence of the formation of a single CF or multiple CFs, as well as the exact position of the filament. In this section we will present some examples of conductive filament experimental evidences.

Son and Shin (Son and Shin, 2008) have used a *Hg* drop top electrode to switch a *NiO* film (Fig. 2.13a). They removed the metal afterward and analyzed the surface of the oxide layer using a conducting atomic force microscopy (C-AFM). It is worth to note the granularity of the high conductive spots on the oxide layer, suggesting the formation and rupture of several filaments. Furthermore, it was shown that the CFs generally form at the grain boundaries of the *NiO* layer.

Combining delamination technique with C-AFM revealed spatially resolved

morphology and conductance changes in TiO_2 memristive junctions after electroforming and switching (Münstermann et al., 2010). The topography and the local current distribution of the sample, denoted by a red 'T' and 'C', respectively, are shown in Fig.2.13b. They demonstrated that electroforming results in the creation of localized conductance channels induced by oxygen vacancies evolution while subsequent resistive switching causes an additional conducting structure next to the forming spot.

Szot et al. (Szot et al., 2006) demonstrated that the local conductivity of $SrTiO_3$ thin films originates from nanoscale conducting filaments connecting the surface to the $SrRuO_3$ bottom electrode. By addressing individual filaments with the AFM tip as well as by scanning areas up to the μm -scale, filamentary switching was analyzed and the electrical conduction of the filaments resulted to be reversibly modulated over several orders of magnitude by application of an appropriate electrical field (Fig.2.13c).

Yasuhara et al. (Yasuhara et al., 2009) studied lateral cells showing the formation of a single percolation path through CuO , in a $Pt/CuO/Pt$ structure. Fig.2.13d shows their analysis, revealing that the CF is constituted by reduced Cu . So far, the most direct approach to study the nature of the CFs in resistive memories has been performed through transmission electron microscopy (TEM) combined with energy dispersive X-ray spectroscopy (EDS). Kwon et al. (Kwon et al., 2010) observed through ex-situ high resolution HRTEM a Ti_4O_7 filament in the ON-state of $Pt/TiO_2/Pt$ cell (Fig. 2.14a). Such conducting channel in the TiO_2 device was found to be made of a new conductive TiO phase with a stoichiometry of $Ti_{(n)}O_{(2n-1)}$, termed Magneli phase, as a product of local oxygen deficiency. The structure of the filament was determined by the selected area diffraction (SAED) pattern, in which the diffraction spot with a d-spacing of 0.62 nm can be identified as (002) diffraction of the Ti_4O_7 Magneli phase. Furthermore a darkfield TEM image obtained from the above mentioned Ti_4O_7 (002) diffraction, clearly shows the presence of the conducting filament bridging top and bottom electrodes. The fast Fourier transformation (FFT) of the filament region and the simulated diffraction pattern further verified that the filament was indeed made up of the Ti_4O_7 Magneli phase.

Chen et al.(Chen et al., 2013) performed in-situ TEM observation of anion migration based conducting filament growth and dissolution processes in ZnO that shows unipolar resistive switching. The real-time filament growth in ZnO during an electroforming process and the relative IV characteristics are displayed are shown in Fig.2.14b. Starting from the initial high resistance state, a conical shaped filament was found to form on the cathode upon application of a positive voltage, which later transformed to a dendrite shape probably due to the evolving electric field distribution during the growth process. A dramatic resistive change from the OFF-state to the ON-state occurred when a cylindrical filament was formed and connected the two electrodes. By applying a positive reset voltage, the filament grad-

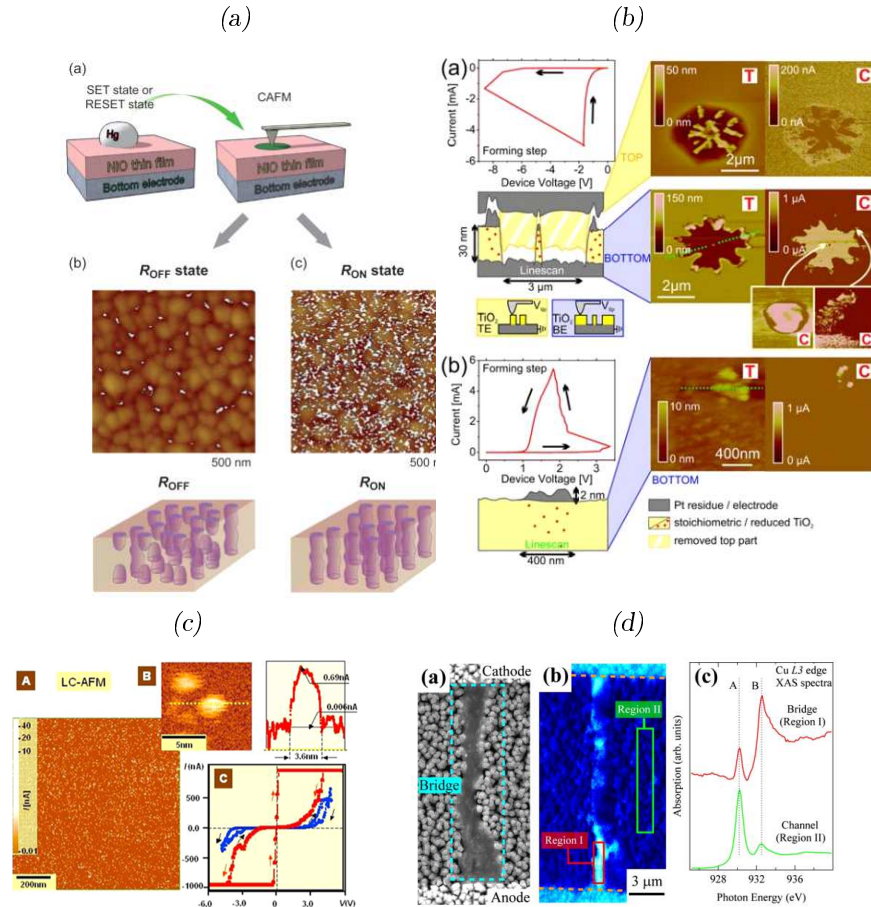


Figure 2.13: Experimental evidences. (a) A drop of Hg is used as top electrode. Once the cell is programmed, removing the Hg allows to analyze the NiO film via C-AFM and the distribution of the conductivity is shown for OFF and ON states. Adapted from (Son and Shin, 2008). (b) The IV and C-AFM data of a sample formed by a negative (positive) voltage sweep showing the influence of electroforming on the morphology and local conductivity of a sample. Adapted from (Munstermann et al., 2010). (c) Conductivity map ($1\mu m \times 1\mu m$) of a $10nm$ thick STO thin film recorded by LC-AFM and zoom of conductivity map ($10nm \times 10nm$) with line scan performed along the dotted line. IV characteristics of two conducting spots with different current load. Adapted from (Szt et al., 2007). (d) SEM image of the planar-type $Pt/CuO/Pt$ cell after forming process. A photoemission electron microscope (PEEM) image at the same region as the SEM image. The bright regions in the bridge structure correspond to the reduced region of the CuO channel. A x-ray absorption spectroscopy (XAS) spectra of $Cu L3$ absorption edge for bridge structure (Region I) and CuO channel (Region II) structures. Adapted from (Yasuhara et al., 2009).

ually dissolved near the anode. These observations are consistent with the thermochemical nature of unipolar switching, caused by oxygen vacancy/ion migration induced phase transition between $ZnO_{(1-x)}$ and ZnO phases. Another in situ TEM study of resistive memory structures and filament

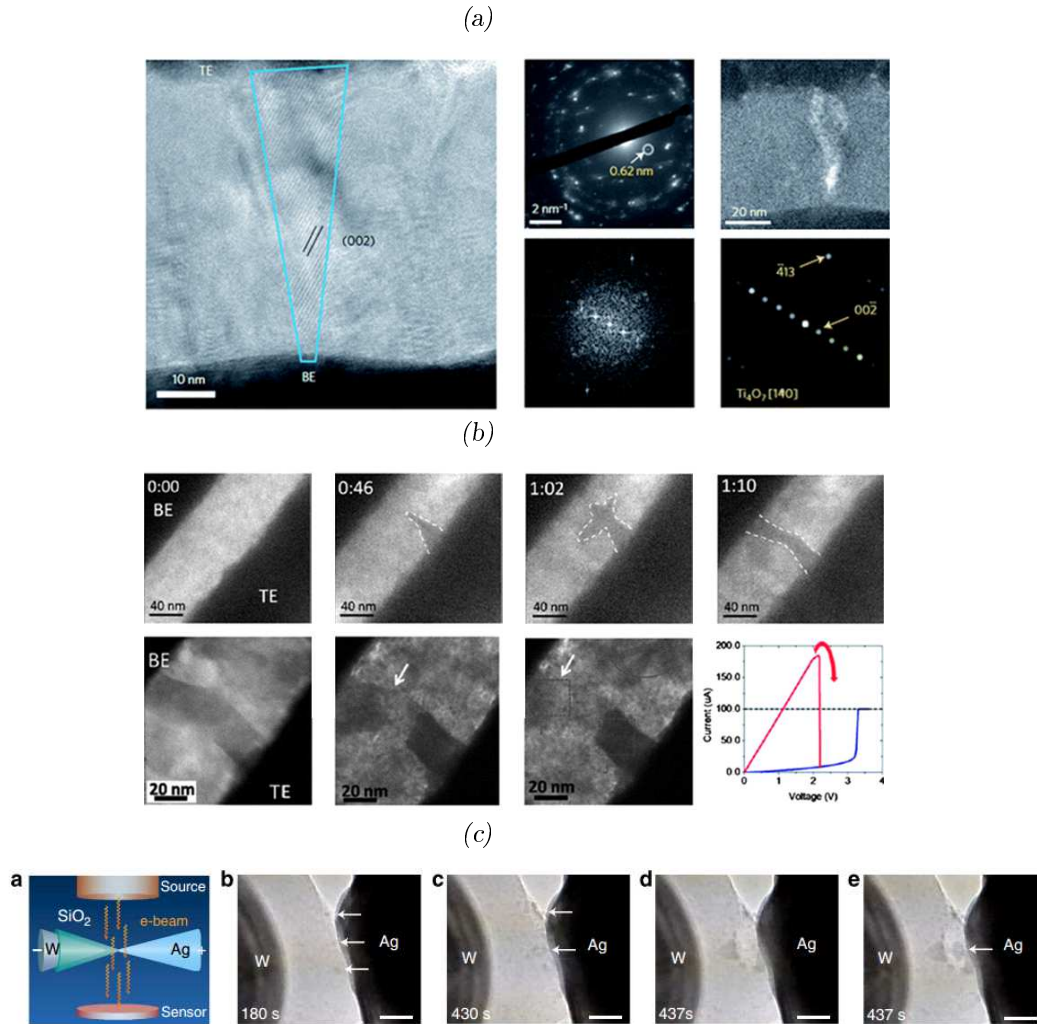


Figure 2.14: Experimental evidences. (a) Ex- situ TEM observation of a Ti_4O_7 filament formed in the device after being switched into the ON-state; SAED pattern of the TiO_x film with the Ti_4O_7 filament; dark-field TEM image obtained from the diffraction spot marked with a circle, showing the presence of the filament; Fast Fourier transformed micro-graph of the HRTEM image of the Ti_4O_7 phase; simulated diffraction pattern by the Bloch-wave method. Adapted from (Kwon et al., 2010). (b) In-situ TEM observation of filament formation and dissolution in ZnO . Adapted from (Chen et al., 2013). (c) (a) Schematic of the experimental setup. (b-e) Real time TEM images showing continuous filament growth within a $40nm$ thick SiO_2 film (applied voltage: $8V$). Scale bar, $20nm$. Adapted from (Yang et al., 2014).

growth has been performed by Yang et al. (Yang et al., 2014) (Fig.2.14c). The devices consist of a $Ag/SiO_2/W$ structure with an evaporated SiO_2 film covering a W probe. The device was directly formed inside the TEM column by connecting a high-purity Ag wire with a movable W probe coated with the SiO_2 film. The first visible filament growth in the SiO_2 was the

appearance of several *Ag* clusters near the *Ag* electrode after $\sim 3min$. Due to the higher concentration of Ag^+ ions near the *Ag* electrode and therefore the higher probability to overcome the nucleation barrier, more *Ag* clusters will be nucleated near the *Ag* electrode inside SiO_2 and the repeated nucleation and growth leads to the filament shape.

All of these different techniques for experimental evidences of conductive filaments are powerful tools to investigate the physics behind filamentary switching. To going deeper in the understanding of the physics behind filamentary switching, simulations modeling represent another useful strategy and it will be presented in the next section.

2.2.2 Simulation Modeling

The underlying physical mechanisms behind filamentary switching are very diverse and complex. Simulation modeling can provide a useful tool to gain deeper understanding on both, atomistic length scales and experimental time scales, to optimize device design and operation and to improve the performances.

A comprehensive study of the filament formation process in ECM cells has been performed through simulation methodology based on Kinetic Monte Carlo (KCM) by Pan et al. (Pan et al., 2011). Redox-based switching has been studied regarding the different relationships between the forming voltage, voltage sweep rate and forming time, as well as the combination of electrochemical processes involved such as adsorption, desorption, bulk and surface diffusion (Fig.2.15a). This work, unlike many others that are based on cylindrical approximations, takes into account the filament topography. The filaments shape has a crucial impact on the device forming time and forming voltage characteristics. They found that large filaments are obtained at low surface overpotentials (voltage applied at the electrode/ionic conductor interface) and long switching time while thin filaments results from large surface overpotentials and short switching time. These results can be explained by the fact that, when the voltage is small, the adsorbed cations at the cathode side tend to diffuse to and reduce at more stable step and hole sites rather than at adatom sites, and hence, the deposition is isotropic; thus, there is no effective gap shrinkage between the anode and cathode. A larger voltage, however, makes the adatom formation easier; thus, the filament grows faster, and its width decreases.

A planar configuration of an ECM cell ($Pt/H_2O/Ag$) is shown in Fig.2.15b. Guo et al. (Guo et al., 2007) have exploited such configuration for CFs in-situ observations during the switching mechanism. The filamentary signature of the redox-based memories results in a fractal dendrite morphology. Dendritic CFs growth can be observed in the Pt/Ag gap during the SET operation. After 1s, it is possible to observe the HRS characterized by shorter and smaller *Ag* dendrites while, after 4s, the cell is switched to the ON-operation. Once

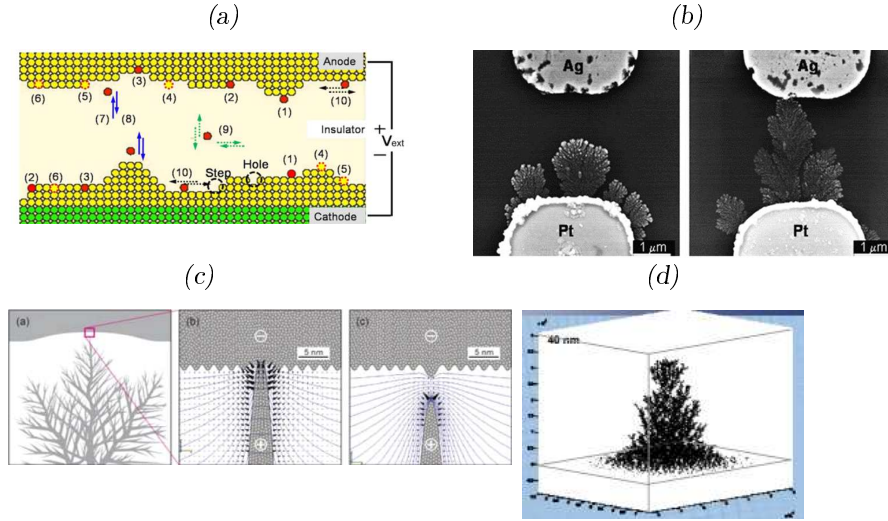


Figure 2.15: Simulation Modeling. (a) Processes described by the KMC simulation of ECM cell: Oxidation at (1) adatom site (0.65 eV), (2) step site (0.7 eV), and (3) hole site (0.75 eV). Reduction at (4) adatom site (0.35 eV), (5) step site (0.3 eV), and (6) hole site (0.25 eV). (7) Adsorption (0.15 eV). (8) Desorption (0.3 eV). (9) Bulk diffusion (0.15 eV). (10) Surface diffusion (0.2 eV). Adapted from (Pan et al., 2011). (b) SEM images of the switching ON process of a $Pt/H_2O/Ag$ cell, showing the Ag dendrite growth after applying -1 V. Adapted from (Guo et al., 2007). (c) Field simulation: (left-side) sketch of the Ag dendrite and the solid Ag electrode in contact. (middle) Tip of the Ag dendrite in higher magnification during the late ON-state. The lines represent equal potential lines after applying a switching-off voltage of 200mV. The cones represent the electrical field and point to the direction of the Ag^+ ion migration. (right-side) Early OFF-state. As soon as the electronic contact is disconnected, the field distribution changes, which accelerates the further dissolution of the dendrite tip. Adapted from (Guo et al., 2007). (d) Molecular dynamics simulation of dendrite growth in an ECM cell. Adapted from (Guo et al., 2007).

the CFs bridge the opposite electrode, the LRS is obtained, with longer and larger Ag dendrites. Unfortunately, AFM and high-resolution SEM failed to disclose the very fine fractal structure of the dendrite front. Due to the continuous dendrite growth, the Ag^+ -ions are depleted in the region immediately in front of the dendrite. Under the influence of the electrical field between the dendrite and the Ag electrode, the dendrite continues to grow. However, once the dendrite front comes into contact with the Ag electrode, the current compliance sets in, the electrical field between the dendrite and the solid inert electrode immediately drops to a significantly lower level. Then the driving force for the growth of the other dendrites decreases, and they almost stop growing. A numerical simulation of such system shows the situation immediately before the dendrite front contacts the Ag electrode (Fig. 2.15c). The contact point is considered of only a few atoms wide, because such a contact is sufficient to establish the low resistance state and to activate the current compliance. On the microscopic level, the contacting twig and the approximately planar Ag bulk electrode are extremely different

in their curvatures and this constitutes the reason for the bipolar switching of the cell. On a macroscopic level, this corresponds to a huge difference in volume density between the Ag dendrite and the solid Ag electrode. Another simulation modeling approach used to investigate dendrite growth in an ECM cell is Molecular Dynamics (Fig.2.15d). Such approach aims at simulating the Brownian motion of copper ions in an applied electric field. The interactions between different copper ions are modeled using suitable potentials. In addition the attraction of the negatively charged dendrite and the copper ions is taken into account.

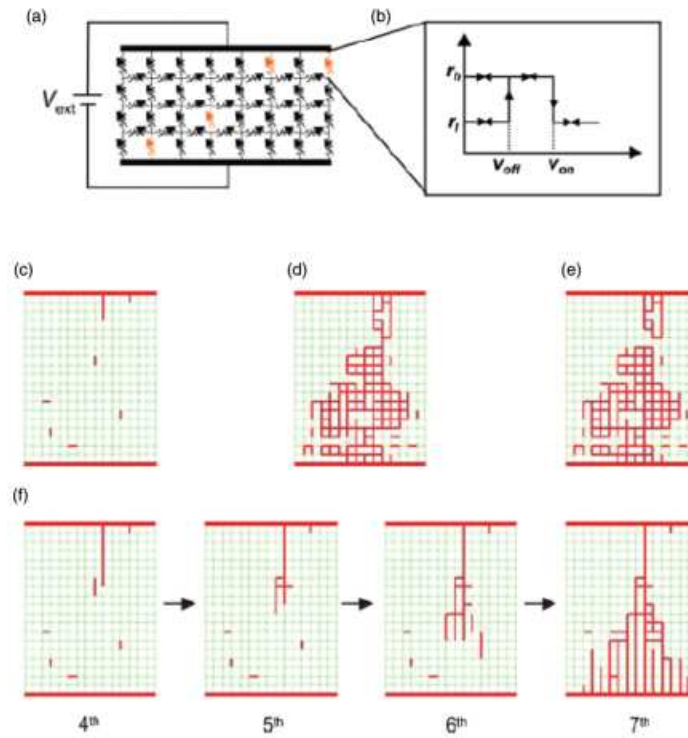


Figure 2.16: Percolation network model: circuit breaker configurations. (a) The red circuit breakers are in the ON state. (b) Detailed conditions for the switching between the two states. (c) Pristine state of a 2D (50 x 20) breaker array with some breakers initially set into the ON state. (d) Example for an ON state condition. (e) OFF state condition with a broken filament. (f) Avalanche like progress of a filament during the forming process. After only four iterations the filament is formed. Adapted from (Chae et al., 2008).

The filaments fractal morphology, as it will be explained in the next section, represents one of key feature of redox-based RRAM cell that we will use to enrich the capabilities that can be addressed by this device for alternative computing paradigms.

Another simulation approach that was proposed to describe unipolar switching of TCM cell is based on a percolation network model (Chae et al., 2008).

Such model is able to describe two switchable metastable states by exploiting the circuit breakers which are arranged in a network (Fig. 2.16). For each circuit breaker, two resistance values are defined, R_{OFF} (marked as black symbols) and R_{ON} (marked in red). It is assumed that switching depends on the magnitude of voltage ΔV applied across the circuit breaker. A RESET transition is observed when $\Delta V > V_{OFF}$ and a set transition for $\Delta V < V_{ON}$ where $V_{ON} \gg V_{OFF}$. Within the active device, the transitions may correspond to the formation or rupture of small segments of a filament. For the simulations ON state circuit breakers were chosen randomly with a given density. Then, the external voltage V_{ext} was increased and the simulation started. In general, it was found that a switching event in one circuit breaker created switching events in circuit breakers nearby. For the reset state the iterations were repeated until a stable state was achieved. For the forming and the set operation, the simulations were stopped when a certain current was flowing through the network. This criteria can be identified as the compliance current I_c in real measurements. Filamentary path during the forming process results in avalanche like progress and a complete filament is formed within only four iterations. More details about the model and its implications can be found in reference (Chae et al., 2008). This simulation approach is expected to provide more insights into the parameters which need to be controlled to improve the observed variations of the set and reset voltages. Tailored arrangements of defects within the device structure may serve to guide the growth of the conducting filament (Ielmini et al., 2011).

2.3 Integration strategies: circuit level

RRAM technology is the natural computing application of memristive nanodevices, two-terminal 'memory resistors' able to change their states of internal resistance state (i.e. conductance) depending on the history of applied voltage. Due to their dynamical nonlinear switching such emerging memories could be used to emulate biological synapses that change their strengths (i.e. weight) as a function of the synaptic activity.

In this section we will present two different approaches for their implementation and operation in the context of NNET: a top-down approach in which elementary cells can be precisely designed, controlled and organized and a bottom-up approach which is reminiscent of random organization in BNNs.

2.3.1 Top-down approach

As introduced in the previous chapter, a feedforward NNET in its simplest form can be represented by a directed acyclic graph (Fig. 2.17) in which neurons and synapses are nodes and edges of a graph, respectively. Each neuron applies a certain transfer function to the sum of its inputs and

then passes information forward to the next layer of neurons. A synapse multiplies its weight w_{ij} with the output of a pre-synaptic neuron and passes the resulting product to the input of the post-synaptic neuron.

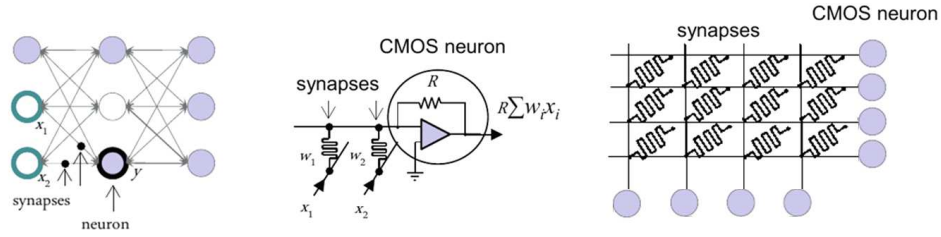


Figure 2.17: Schematic representation of ANNs (left side) and schematic of crossbar array formed by the two terminal memristive cross-points, nanowire electrodes and CMOS neurons.

Feedforward NNET, and in particular the multilayered perceptron MLP structure which is based on such networks maps naturally to the crossbar array circuit. From a circuit view point each neuron can be realized by complementary metal-oxide semiconductor (CMOS) with adjustable two-terminal resistive devices ('memristors') at each crosspoint.

RRAM technologies in its broad sense, as presented in the previous section, are ideal candidates for the implementation of dense memory arrays interconnected into crossbar. This is directly applicable to ANNs architecture that only required a static, non-volatile weight.

A conventional integration design where each RRAM cell has a dedicated MOSFET transistor is the '1T1R' structure (Fig. 2.18a). Similar to conventional DRAM (Udipi et al., 2010), when a row gets activated, the access transistors in the selected row provide exclusive access to the cells in that row without disturbing other cells in the array. However, unlike DRAM, resistive memories typically operate at a significantly higher current, requiring a large sized access transistor for each cell. The size of these transistors ultimately increases the area and hence the cost. However, due to perfect isolation provided by these access transistors, the '1T1R' design is more energy efficient and has superior access time compared to other alternatives.

Based on the characteristics of RRAM cross-point, an RRAM array can be designed as a dense crossbar architecture, configuration that has been proposed as a leading candidate for future memory and logic applications (Fig. 2.18a). In a crossbar architecture, all cells are interconnected to each other without transistors: RRAM cells are directly sandwiched between top and bottom electrodes. By eliminating access transistors, cells in a crossbar achieve the smallest theoretical size of $4F^2$ (Burr et al., 2010). Such design allows to access a single cell in an array by applying the proper potential across the wordline and bitline to which the cell is connected. However, as shown in Fig. 2.18b, as selected cells are no longer isolated from unselected cells, activating a wordline and a bitline will result in current flow across all

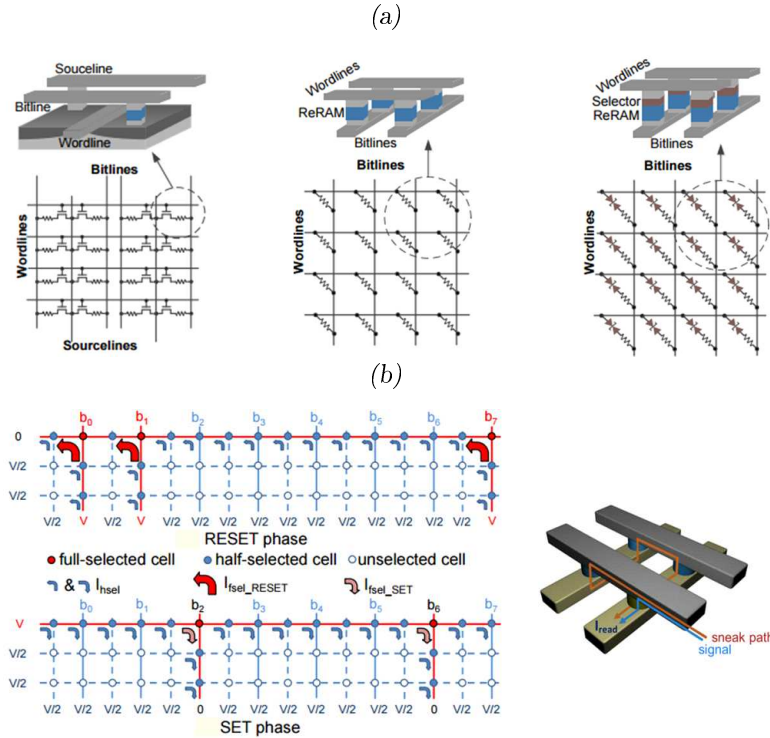


Figure 2.18: RRAM Array Architectures. (a) RRAM Array Architectures (from left to right): MOSFET accessed structure (grid with 1T1R cells); access-device-free crossbar structure and diode-accessed crossbar structure. (b) Two phase multi-bit write operation in a crossbar array: RESET and SET phase. On the left: sketch of the sneak path issue.

the cells in the selected row and column, i.e. other cells in the selected row and column also see partial voltage across them. These half-selected cells in the selected row and column leak current through them due to the partial write voltage across them, which is commonly referred to as sneak current. Several strategies can be adopted in order to reduce the sneak current and leakage currents issues. One of the most common solution is for instance to half biased at $V/2$ all of the other wordlines and bitlines that are not selected. This limits the voltage drop on the half selected cells to $V/2$ and voltage drop on the unselected cells to 0.

As previously mentioned, RRAM cells can exhibit a non-linear relationship between voltage and current. The current decreases significantly with a small drop in voltage and this could helps to confine the sneak current through half-selected cells. Thus, in a crossbar architecture, the ratio of the amount of current flowing through a fully-selected cell to a half-selected cell, referred to as non-linearity (κ), is one of the key parameter. The higher the κ , the lower the sneak current, and the higher the feasibility of a large crossbar array. Many recent RRAM prototypes employ a dedicated selector or bipolar

diode in each cell to improve κ . Since a selector can be built on top of the switching material, there is no extra area overhead required for the selector (Fig. 2.18a).

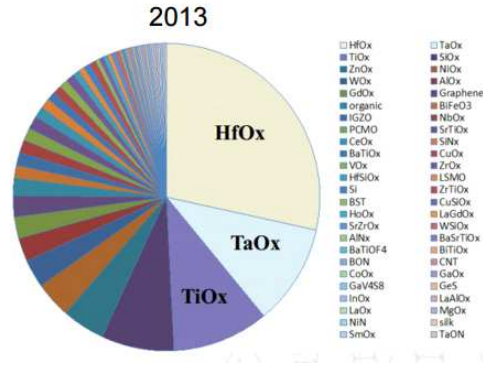


Figure 2.19: Materials for RRAM (from Google scholar Z. Wei, Flash Summit 2013).

Among the rich panel of emerging and prototypical filamentary devices, oxide-based resistive switching devices (OxRAM) can fulfill most of the requirements and have been developed with a large variety of materials (Wong et al., 2012). As reported in Fig.2.19, titanium dioxide (TiO_2), tantalum pentoxide (TaO_2) and hafnium dioxide (HfO_2) are the most popular materials used for memory applications. Especially this latter, the HfO_2 , is considered one of the most promising candidate thanks to its excellent scalability ($< 10nm$), superior switching speed/energy and high endurance ($> 10^{10}$) (Govoreanu et al., 2011) and (Lee et al., 2010). The TiO_2 -metal oxide technology, thanks to its high yield and low dispersion, is also considered a potential solution for future non-volatile memories (Xu et al., 2015). Furthermore, this OxRAM devices offer not only binary states but have been proposed for multi-level storage (Beck et al., 2000) or even analog memory implementation (Alibart et al., 2012a) and a precise analog control for TiO_2 devices can be obtained thanks to its gradual SET and RESET transition. Thus, due to these promising characteristics that could be exploited to be integrated in NNET architectures, not only for memory and logic applications, but also for alternative computing paradigms such as analog or neuromorphic computing.

2.3.2 Bottom-up approach

As discussed in the previous chapter, biological neural networks (BNNs), in contrast to the ordered layer configuration of the ANNs, are constructed in a three dimensional way with a random organization from microscopic components, i.e. neurons that seem capable of nearly unrestricted interconnections. Conventional fabrication techniques, well suitable for reproducing ANNs-like configurations, are unable to efficiently design structures

with the highly complex interconnectivity found in BNNs. Thus, alternative approaches such as bottom-up fabrication techniques and self-assembled of nano-objects could offer an ideal solution for building such bio-inspired complex network architecture.

Nanowires, due to their one-dimensional geometry and unique possibilities for engineering of electronic and optical properties, hold great promise for a variety of device applications including chemical and biological sensors (Cui et al., 2001) or field effect transistors (Dayeh et al., 2007). Furthermore, nanowires can be synthesized through a variety of techniques (Motohisa et al., 2004) and some of which allow for unique device geometries, such as axial or coaxial heterostructures, that are not easily realized in planar device fabrication schemes (Zhu et al., 2009). While significant advances have been made in nanowire synthesis and device characterization, post-growth manipulation and placement of nanowires in a coherent and useful fashion continues to be a considerable challenge, one that must be overcome to realize large-scale complex nanowire based systems. A number of schemes have been proposed to meet this challenge such as Langmuir-Blodgett films (Whang et al., 2003) or dielectrophoresis (DEP) (Raychaudhuri et al., 2009). All these techniques offer the ability to line up nanowires in parallel but do not allow for precise nanowire placement for functional systems and a way to make mass production feasible is still missing.

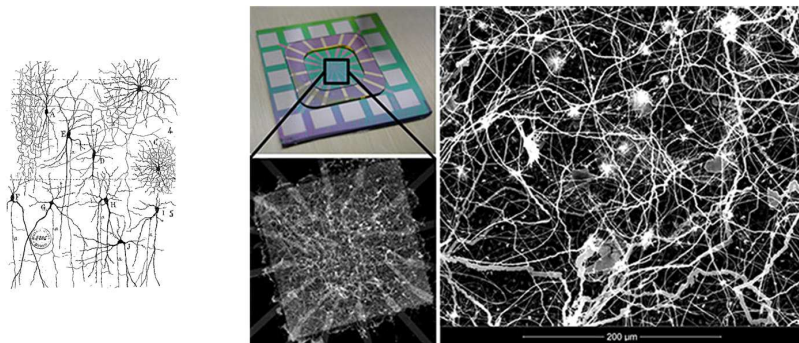


Figure 2.20: Schematic representation of BNNs (left side) and densely and randomly interconnected network of silver nanowires with patterned electrodes (Avizienis et al., 2012).

However, in the bio-inspired NNET context, complex nanowire networks are relatively simple to fabricate using self-assembly and would therefore be the ideal wiring architectures, as shown in Fig.2.20.

Avizienis et al. (Avizienis et al., 2012), have proposed to study the consequences of coupling many nanoscale synaptic memories together in a highly interconnected, recurrent structure to create an operational neuromorphic device that self-assembles into a functional state. The memristive device elements, also named 'atomic switches' at each point of contact between

silver nanowires will be presented in detail in the next section being the basis of our experimental work. In this context the collective interactions between these atomic switches has been investigated showing significant potential for neuromorphic computing. Specifically, interesting features have been demonstrated from the network properties, such as the distributed conductance and the recurrent dynamics from the frequency and dc networks response respectively, indicating a potential capacity for efficient information processing, thereby surmounting problems associated with wire delays and interconnect structures.

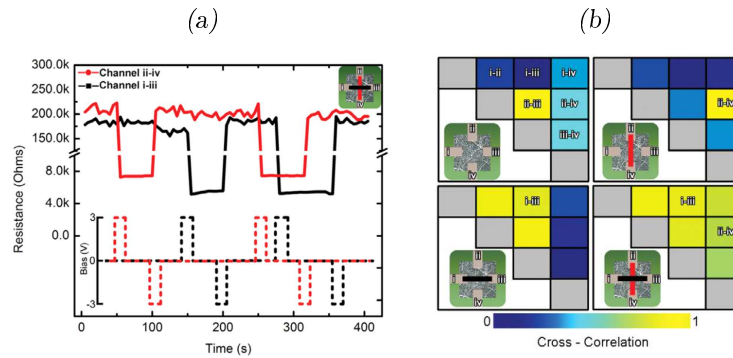


Figure 2.21: Distributed Memory Storage from Network scale Switching. (a) A 2 bit non volatile memory device operation by monitoring resistance states across two channels (i-iii and ii-iv). *ON/OFF* switching of each channel is induced using pulses (3V, 1s in duration) and resistances are measured every 5s with pulses (200 mV, 100 ms in duration). (b) The network's internal configurations show diverse correlated patterns, from no correlation (blue) to total correlation (yellow). Correlation coefficients of channel resistances are shown for all 6 pairwise electrode combinations. The correlation coefficients are calculated during each of the 4 network switching configurations; the black and red bars (insets) show the channels that are ON in the switching state(Avizienis et al., 2012).

Distributed memory storage has been also implemented from the network-scale switching by monitoring resistance states across two channels (Fig. 2.21). The conductive paths between the two channels that overlap spatially are switched independently, indicating that local subregions of the network can operate to distinct operational modes despite being embedded within a highly interconnected, largely metallic structure. By considering the BNNs, this is analogous to the presence of feedforward subnetworks within the recurrent architecture of the brain cortex. The distributed nature of the atomic switch array's dynamics makes it a candidate platform for efficient kernel design in the emerging field of 'Reservoir Computation' (Lukoševičius and Jaeger, 2009).

2.4 Neuromorphic NNET strategies: system level

2.4.1 The CMOL concept

The crossbar resistive memory array, in which the storage elements are two-terminal resistive switches, forming a passive interconnected network, and hybrid crossbar/CMOS systems have been identified as a leading candidate for future memory and logic applications (Strukov and Likharev, 2007), (Waser and Aono, 2007) and (Jo et al., 2010a).

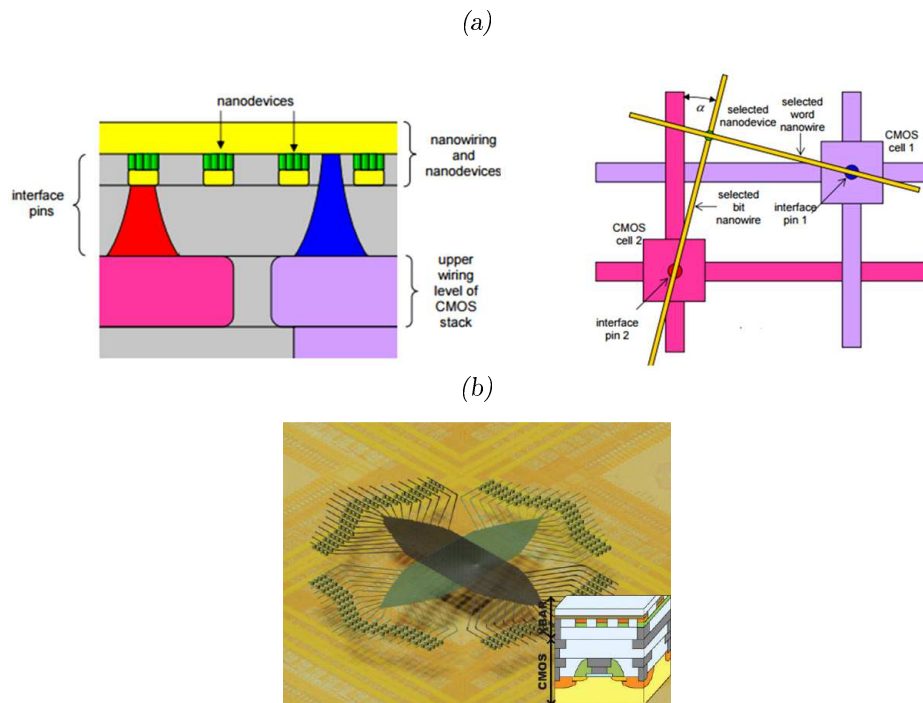


Figure 2.22: CMOL technology. (a) Hybrid circuit: CMOS/Nanotechnologies. A schematic side view (on the left) and a schematic top view showing the idea of addressing a particular nanodevice via a pair of CMOS cells and interface pins (on the right). (b) SEM image of a crossbar array fabricated on top of a CMOS chip realized by (Kim et al., 2011). Scale bar: $5\mu\text{m}$.

An efficient circuit implementation of such system has been proposed by Likharev and Strukov (Likharev and Strukov, 2005). This new technological circuits concept (CMOL) (Figure 2.22) provides a realistic solution to the following three technological points: (i) it allows an efficient interfacing between a CMOS platform and a crossbar of 2-terminal nanodevices. This point is of particular interest in the context of passive crossbar for neuromorphic systems where neuronal functions can be assigned to the CMOS platform and synaptic connections to the crossbar of nanodevices (Kim et al., 2011) (Figure 2.22b). (ii) CMOL architecture can be extended to 3D crossbar integration to increase the density of nanodevices. Experimental proof of concept of 3D

crossbar has been demonstrated in (Kügeler et al., 2009). (iii) This solution is compatible with the fabrication lines of semiconductor industry (Strukov and Williams, 2009b).

Even if the possibility of coupling nanodevices and CMOS (a 40x40 passif crossbar of memory nanodevices with a CMOS circuit used for addressing and signal restoration) for memory applications has been already demonstrated (Kim et al., 2011), the implementation of a functional system where neurons and synapses will realize a given function by interacting together represent a major challenge. Recently, a TiO_2 -based crossbar circuit have been successfully used for experimental demonstration of pattern classification by a single layer perceptron network implementation (Alibart et al., 2013) and (Prezioso et al., 2015) and such results paves the way to promising computing systems.

The CMOL concept is a promising integration strategy to realize NNET with emerging nanoscale devices. It requires a join effort from circuit, engineering at the fabrication level (i.e. CMOS processes are not flexible) and device optimization.

2.4.2 The Reservoir Computing concept

The Reservoir Computing (RC), a high-dimensional non-linear dynamical system driven by time-dependent inputs, is of particular interest nowadays. Liquid-state machines (LSM)(Kaminski and Wojcik, 2004), and echo state networks (ESN) (Tukker et al., 2012) represent two major types of reservoir computing (RC). In such a way, initial information contained in the input is spread into a space with many dimensions (states) and the readout layer is used to pick a particular set of states (Fig. 2.23). Generically, this means that the state configuration generated by the input signals can be regarded as an internal interference (correlation) pattern that can be read out by a generic 'image' processing device, typically a trained neural network. The spreading of the input signals over a large state space of the dynamical systems can be viewed as giving rise to a time dependent pattern in state space, corresponding to dynamical patterns in real space (e.g. wave patterns), frequency and time. RC does not require subtle control of internal network dynamics and is therefore simpler to execute, making it an appealing route to be used for complex networks of neuromorphic devices to perform computational tasks (Kulkarni and Teuscher, 2012) and (Burger and Teuscher, 2013). Thus, the complex network architectures, generated through self-assembly of functional nanoscale elements, like those described by Avizienis et al. (Avizienis et al., 2012), with its distributed collective nonlinear dynamics can be suitably described by RC concept.

RC has been also implemented with recurrent neural networks (RNNs) (involving feedback) of nonlinear memristive components (Konkoli and Wendin, 2013). Kulkarni et al. (Kulkarni and Teuscher, 2012) have implemented RC

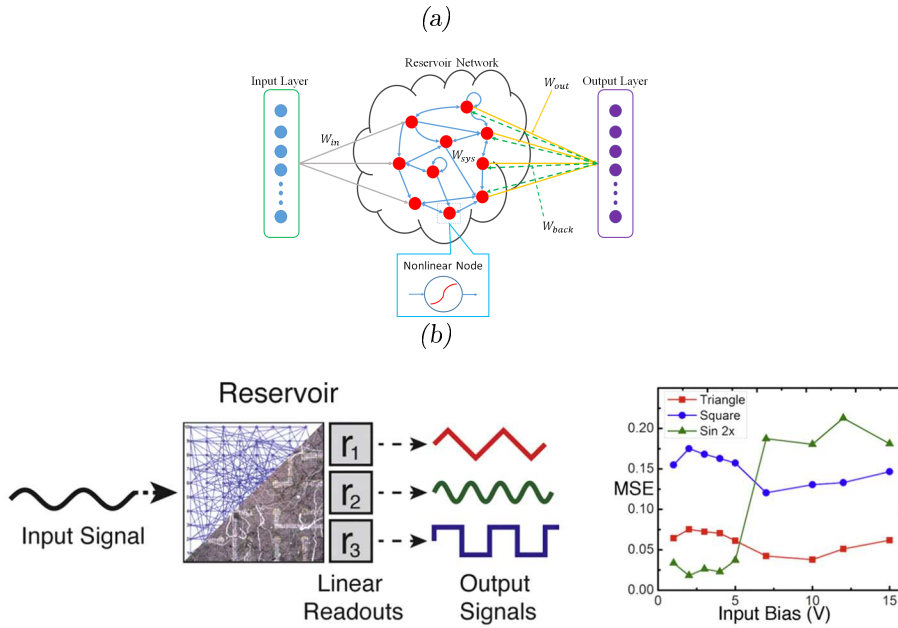


Figure 2.23: Reservoir Computing (RC). (a) Conceptual schematic of Reservoir Computing (RC). (b) Schematic of the network simulation used for RC implementation with a 5 V, 10 Hz sinusoidal input signal and tasked to produce 10 Hz triangle/square and 20 Hz sinusoidal waveforms and Mean-squared error (MSE) for each task with respect to driving amplitude showed minimal error in triangle/square waveform generation task at 10 V, corresponding to the onset of higher harmonic generation (Sillin et al., 2013).

in software for memristor-based networks with 5 – 40 nodes. The authors demonstrated two applications of memristor networks for information processing. In the first example a readout layer of neurons ‘Perceptron’ was trained to distinguish between sawtooth and square wave forms. In the second example a version of the Pavlov’s Dog problem has been implemented in which the output network is then able to learn to identify the Bell signal in the absence of the Food signal.

2.5 Discussion and Perspectives

In this chapter we presented a practical aspect of synaptic nano-devices: how they can be used and integrated in neuromorphic systems. By starting from the nano-device level, a review state of the art of resistive switching memories have been presented by focusing on a particular filamentary-type class, the ECM cell. Such nanoscale memory configuration has been developed and characterized during this PhD work and in the next chapter we will present the experimental details by motivating such technology choice in the context of neuromorphic computing.

The second part of this chapter at circuit and system level has been dedicated

to describe integration strategies with structural and functional affinities between ANNs and BNNs. By regarding the current technological status for the development of future neuromorphic hardware systems, 'ordered' memristive cross-bar architecture and CMOS-compatible seems the most promising and robust approach to an hardware implementation of ANNs. Random cross-bar architectures approach, even if several engineering challenging issues have to be addressed, presents promising and interesting peculiarities that could be exploit to implement complex neuromorphic functionalities and easiest way for an hardware implementation of BNNs.

Finally at computational level we proposed two different approaches (the CMOL and the Reservoir Computing) in line with the top-down and bottom-up integration strategies, respectively.

Chapter 3

Filamentary Switching: Development and Characterization

*"Perseverance is the hard work you do after
you get tired of doing the hard work you already did."*

Newt Gingrich, 1943

3.1 Introduction

By motivating the technology choice of a particular class of filamentary switching, the Ag_2S -ECM cell, in the context of neuromorphic computing, this chapter is dedicated to the experimental procedure, in terms of device fabrication techniques and electrical characterization, performed during this PhD work. In particular, we will present different nanofabrication technologies to realize filamentary memories: a top-down approach closer to the ANNs architecture and a bottom-up approach inspired by the BNNs one. In the next chapter, by going deeper into the expression of *Synaptic Plasticity* observed in biological synapses, we will demonstrate how complex plastic behavior can emerge from ECM cells, offering a promising and interesting way to enrich and enhance future bio-inspired information computing systems.

3.2 Ag_2S Thin Films Deposition

The filamentary memory device developed in this PhD work consists of a $Ag/Ag_2S/Pt$ cell and in this section we focus on the key-material element under-test: the silver sulfide (Ag_2S).

Ag_2S is a mixed conductor material, with a total conductivity due to the transport of both Ag^+ -ions and electrons. Ag_2S corresponds to the family

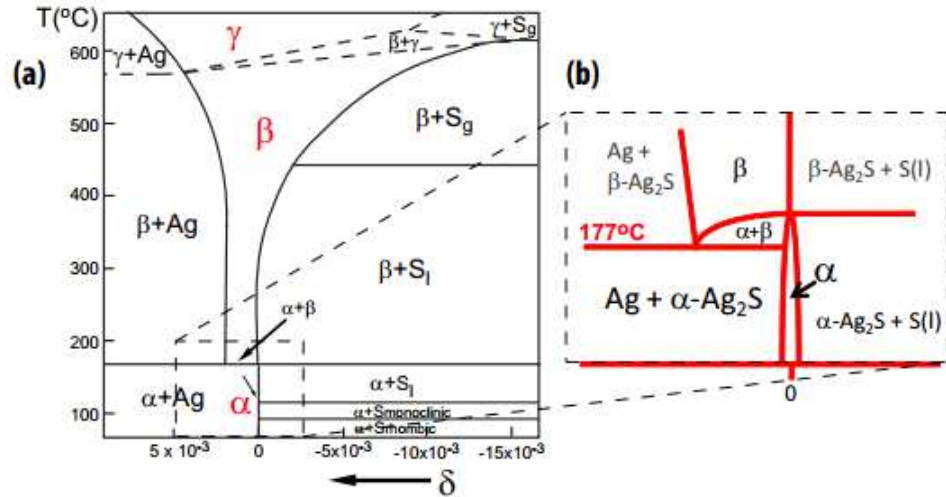


Figure 3.1: Phase diagram of the Ag - S binary system. From 0 to $177C^\circ$ the system is in α -phase; from $177C^\circ$ to around $590C^\circ$ is stable the β -phase; above $600C^\circ$ is in the γ -phase. Adapted from (Schmalzried, 1980).

of the silver chalcogenides and its structure is usually a rigid body-centered cubic (bcc) lattice, formed by covalently bonded chalcogen atoms. The Ag^+ -ions are distributed in octahedral and tetrahedral positions in the lattice. The number of octahedral and tetrahedral sites available is much larger than the number of Ag^+ -ions, and therefore there are always positions in the lattice available for the ions to move into. This results in a high ion mobility observed in all silver chalcogenides. The physical properties of Ag_2S , i.e. electronic and ionic conductivity, crystal structure and distribution of defects, are strongly modified with temperature, stoichiometry and composition (Ag/S ratio)(Schmalzried, 1980). It presents good chemical stability and exists in three stable phases, α , β and γ in order of increasing temperature. Fig.3.1 presents the phase diagram of Ag_2S , indicating the stable range for each of the three phases as a function of temperature and stoichiometry parameter δ which indicates the excess ($\delta > 0$) or deficit ($\delta < 0$) of Ag in $Ag_{2+\delta}S$. For our purpose, the α -phase is of special interest because it is the stable phase at room temperature, even if in the chapter 4, the temperature effect will be also taken into account.

Different methods have been adopted for the growth of Ag_2S thin films in literature for CBRAM, which include chemical vapor deposition (CVD) (Panneerselvam et al., 2008), chemical bath deposition (CBD) (Meherzi-Maghraoui et al., 1996) and (Rodríguez et al., 2005) and thermal evaporation (Lekshmi et al., 2008) and (Hasegawa et al., 2010).

In our work, thin films of Ag_2S were prepared by two different methods: (i) the sulfurization of a Ag thin film in vacuum and (ii) by thermal evaporation technique. (i) Silver sulfurization is the conversion of a Ag thin film to Ag_2S

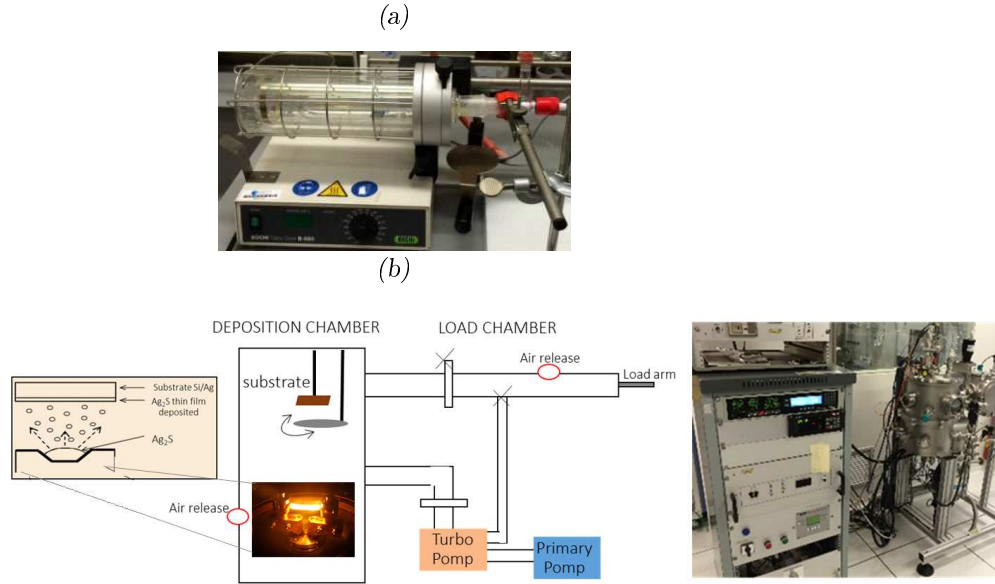
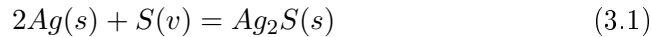


Figure 3.2: Ag_2S deposition techniques: (a) Crystal tube and furnace machine used for Ag sulfurization. (b) Thermal evaporation thin films deposition machine and schematic of the deposition principle.

by reaction with sulfur vapor:



The first step consists of a thin layer deposition of Ag ($35nm$) by electron beam evaporation onto a cleaned surface of silicon p -type $Si(100)$, which was covered with thermally grown $200nm$ thick SiO_2 . In the second step the synthesis of Ag_2S was performed by sulfurization of Ag film. Sulfur powder (reagent grade powder purified by sublimation) is loaded into a quartz tube ($18mm$ internal diameter) and the sample is held at $10cm$ horizontal distance facing the sulfur powder (as shown in Fig. 3.2a). Once the sulfur and the sample are loaded, the tube is evacuated to a pressure of $0.1mbar$. The temperature in the tube is then increased to $523K(\pm 3K)$ using a horizontal furnace with a programmable temperature control. The tube is kept under static vacuum to create a sulfur atmosphere, while the temperature remains constant at $523K(\pm 3K)$ for one hour. After one hour, the tube is evacuated but kept at $523K(\pm 3K)$ to anneal the samples during one more hour. Finally, the sample is slowly cooled down to room temperature at a rate of $1K/min$. Sulfurization starts with a direct reaction of Ag atoms on or near the surface of the film, with the S vapor forming a Ag_2S layer. The inconvenient of this simple method is that it does not allow a direct control of the Ag_2S thickness deposited.

(ii) The second Ag_2S deposition method is by thermal evaporation that consists of melting and evaporation of Ag_2S and consequently condensation

on the substrate ($Si/SiO_2(200nm)/Ag(35nm)$). Thanks to a quartz-crystal sensor integrated in the system (Figure 3.2b), it is possible to monitor the Ag_2S thickness deposited during the deposition (i.e. through the deposition-rate curve in function of thickness deposited). By tuning the current passing through the resistance integrated under the crucible containing the $Ag_2S(s)$, it is possible to induce Joule heating (i.e. the increase of the temperature responsible of the Ag_2S melting and evaporation) and the Ag_2S thin film deposition. The deposition machine presents a load and deposition chamber, and a system with primary and turbo pump (i.e. the vacuum reached during the deposition process is about $3.5 \times 10^{-5} mbar$). Ag_2S thin films thickness was verified by profilometer and ellipsometer by providing an optimized and reproducible sample preparation technique. The deposited thickness taken into account for the experiments described in this thesis is $60nm$.

3.3 Millimeter-scale configuration

A millimeter-scale $Ag/Ag_2S/Pt$ device configuration has been used (Fig.3.3a). A $25nm$ Ag bottom electrode was deposited by electron beam evaporation onto the cleaned surface of p-type silicon. A thin film of Ag_2S ($60nm$) was deposited by thermal evaporation, as explained in the previous section, onto the full substrate. Finally, a Pt top electrode, with a thickness of $25nm$ and electrode size of $(0.1, 0.3, 0.6, 0.9, 1, 2)mm$, was deposited on the Ag_2S layer by using a shadow mask (Fig.3.3b) and electron beam evaporation deposition technique.

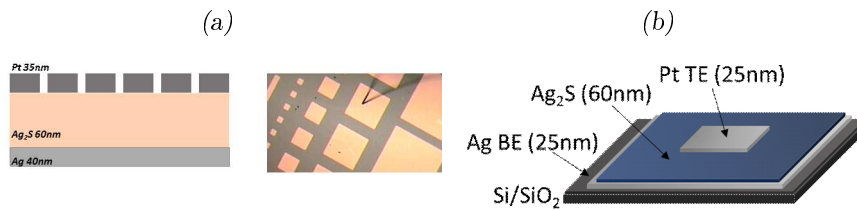


Figure 3.3: Millimeter-scale ECM cell configuration. (a) Schematic of the fabrication steps and optical microscope image of the mechanical mask (Sizes squares: $(0.1, 0.3, 0.6, 0.9, 1, 2)mm$). (b) Schematic of the device configuration with $(0.1mm \times 0.1mm)$ active area).

The basic principle of the device developed corresponds to a conventional ECM cell, as introduced in the previous chapter (Fig.3.4). Concerning the switching mechanism, a positive bias (with a grounded Pt electrode) induces the oxidation of Ag into Ag^+ ions at the Ag electrode, the migration of ions from the Ag anode to the Pt cathode, and the reduction of Ag^+ ions into Ag filaments across the insulating Ag_2S , thereby turning the device from an insulating OFF state to a conductive ON state (SET transition). A negative bias induces the oxidation of Ag from the filament into Ag^+ ions and reduc-

tion at the Ag electrode, leading to a disruption of the conductive path that turns the device OFF (RESET transition).

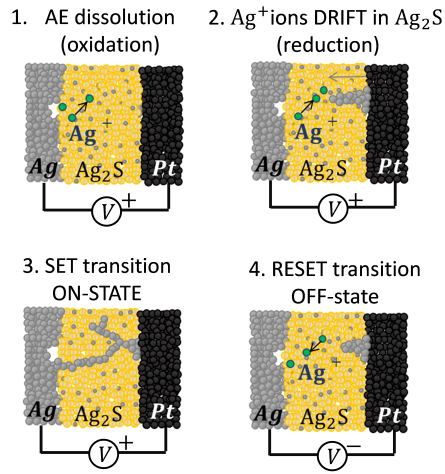


Figure 3.4: Basic switching mechanism of ECM cells.

The millimeterscale ECM cell developed, allowed us to gain insight into the filament shape and growth mechanism by performing optical microscopic imaging during the current voltage (IV) measurement with a square-shaped Pt electrode on top of a Ag/Ag_2S substrate. Consistent with the switching scenario described above, a positive bias induced the formation of Ag dendritic filaments from the the Ag anode toward the Pt cathode (SET transition, Fig.3.5a, snapshot 1 to 3). Application of a negative bias induced a partial destruction of the conducting paths, with remaining filament traces corresponding to preferential paths for subsequent switching (RESET transition, Fig. 3.5a, snapshot 4).

After an identical positive SET transition, an intermediate situation was observed, in which the device was kept grounded for 5 minutes with a slow dissolution of the metallic dendrites (Fig.3.5b, snapshot 4*). Such filament relaxation can be attributed to the Ag^+ ion diffusion in the Ag_2S ionic conductor and to the reverse oxidation-reduction process of the Ag filaments(Valov et al., 2013a).

A second analysis of the filament formation was realized by varying the compliance current (I_c) during the SET process. This approach is commonly used in ECM cells to tune the conductance of the ON state and to limit the formation of filaments (Russo et al., 2009). If tuning the conductance by limiting the growth of a single filament is considered straightforward (i.e., because the filament diameter corresponds directly to the conductance state), then a more complex picture was obtained for ECM cells that had complex dendritic filament morphologies. Increasing the density or width of the dendritic branch can correspond to an increase of conductance. Due to the

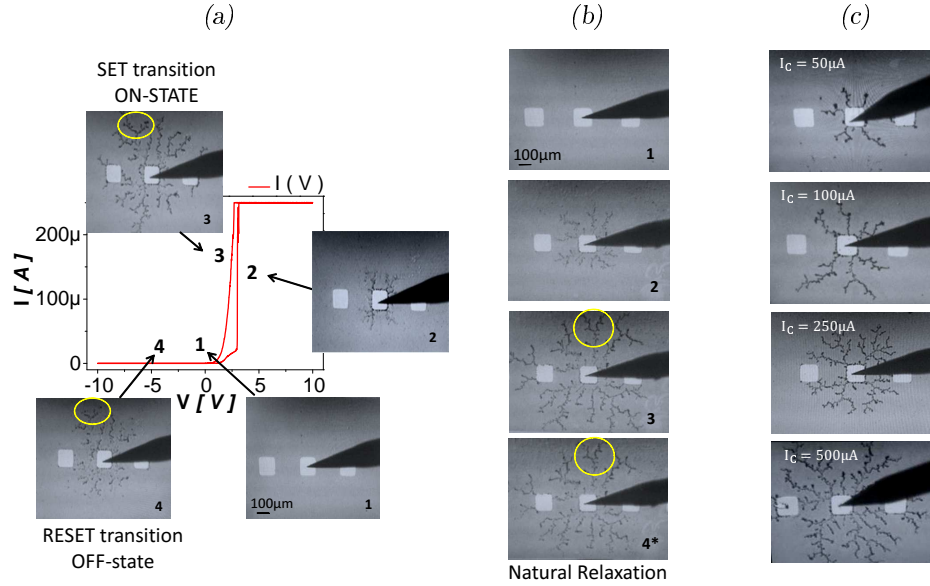


Figure 3.5: Millimeter-scale ECM cell configuration. (a) I-V characteristics and associated optical microscope imaging ($0.1\text{mm} \times 0.1\text{mm}$) of filament growth. (b) Natural relaxation of the filament. After a positive SET transition (1 – 3), the device was kept grounded for 5 minutes (4*). (c) Relationship between I_c and dendritic expansion/shape.

resolution of the optical microscope, it was not possible to obtain an accurate assessment of filament diameter. However, we effectively measured a larger filament expansion and dendritic tree density with a larger I_c (Fig.3.5c). This observation indicates a direct correlation between I_c and the fractal geometry of the dendritic filaments that will be investigated in the next section. Again, after RESET, the remaining filament traces corresponded to preferential paths for subsequent switching.

3.3.1 Fractal analysis of Dendritic Filaments

From optical imaging of the mm scale device configuration, the analogy between the filament growth morphology and fractal structure appears evident. Such pattern complexity is not well described by common euclidean measures (i.e. diameter or length). For this reason, an useful analysis to investigate the physics behind the filamentary switching could be done by exploiting the fractal geometry.

By properly choosing a region of an optical image ($60px \times 110px$) and by converting it in a binary image (Fig. 3.6) it is possible to estimate the fractal dimension (D) and its lacunarity (λ). These calculations are made through *ImageJ*, a software that allows to count the number of boxes of an increasing size needed to cover a one pixel binary object boundary and implements the fractal method as described in the ref. (Smith Jr et al., 1996). A plot is

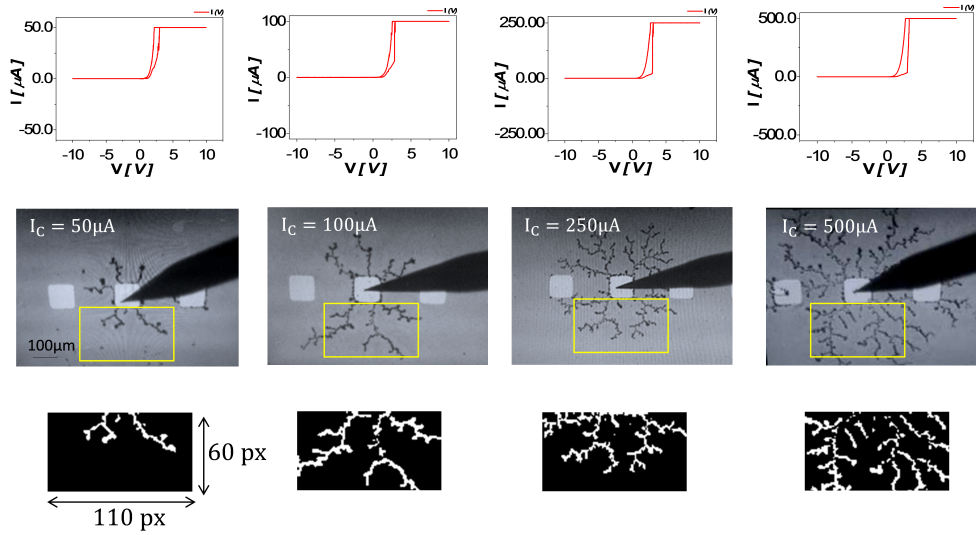


Figure 3.6: Filamentary switching analysis through fractal geometry: conditioning loops for $I_c = 50, 100, 250$ and $500 \mu A$, respective optical microscope imaging ($1mm \times 1mm$) of the filament growth and binary images of the selected yellow region ($60px \times 110px$).

generated with the log of size on the X-axis and the log of count on the Y-axis and the data is fitted with a straight line. The slope (S) of the line is

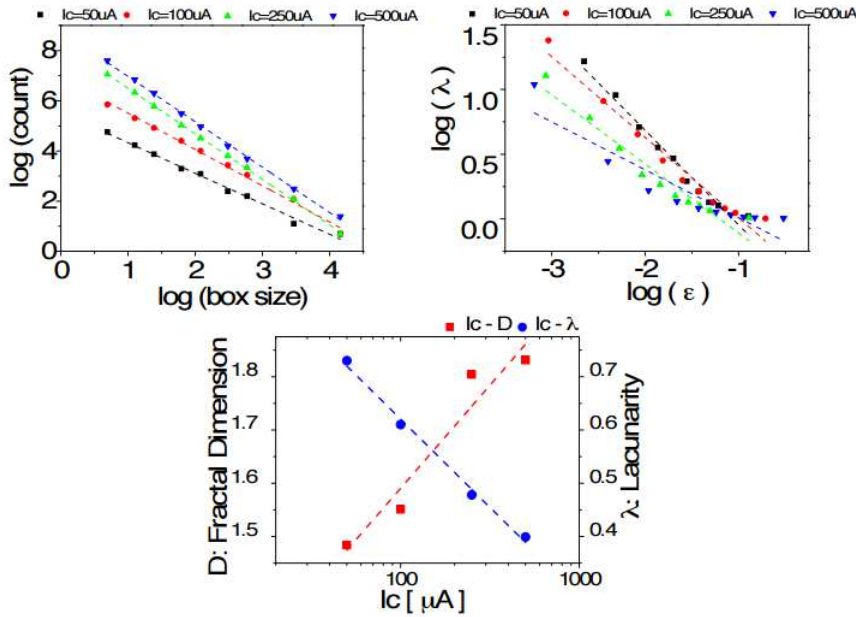


Figure 3.7: Filamentary switching analysis through fractal geometry: Fractal dimension D and Lacunarity λ parameter calculation and relation with I_c .

the negative of the fractal dimension, i.e., $D = -slope$. The lacunarity (λ) can be defined as the measure of the fractal structural variation or fractal texture. It is calculated from the standard deviation (σ), and mean (μ), for pixels per box, i.e. $\lambda = (\sigma/\mu)^2$. Thus, D and λ work together to characterize complex patterns extracted from digital images. Fig.3.7 presents the evolution of fractal parameters as a function of I_c during SET. A clear correlation and anti-correlation with I_c was obtained for λ and D , respectively. These latter parameters do not provide a direct description of dendritic branches density and width, but such evolution is in agreement with the proposed scenario. Further investigation will be carried to exploit fractal geometry description of filamentary switching.

3.4 Nanoscale configurations

Along the neuromorphic research line, two different fabrication approaches will be adopted: a top-down approach in which elementary cells can be precisely designed, controlled and organized and a bottom-up approach which is reminiscent of random organization in BNNs.

3.4.1 Top-down approach

A conventional Electron Beam Lithography (EBL) approach has been used to develop nanoscale ECM cells. The first nanofabrication step concerns the pattern design and for this purpose we used LayoutEditor, one of most popular software to edit designs for nanoscale devices, MEMS and IC fabrication. The visualization and edition is completely graphic and it allows flexible and fast manipulation, e.g. it allows to draw basic structures such as rectangles, circles or polygons in separated layers for multilevels exposure. One pattern design example, showing ECM cell design with cross-electrodes size of $200nm$ is shown in Fig.3.8.

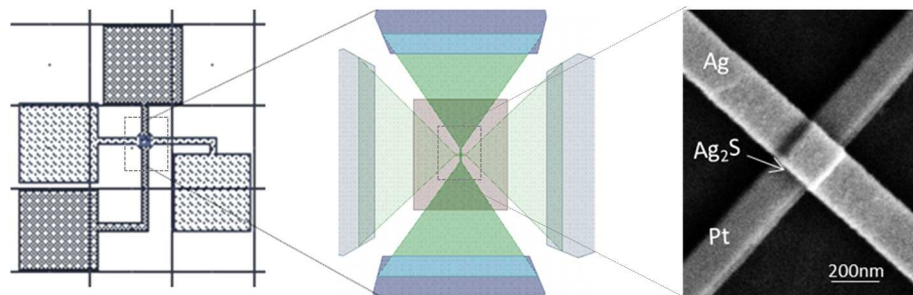


Figure 3.8: Nanoscale ECM cell configuration: Layout Editor device design and SEM image of the device realized ($200nm \times 200nm$ of cross-point active area).

Once the desired pattern is created, the software saves the project in a GDS format and it will be used for the electron beam exposure. the applied we

will not enter in the physical details regarding the EBL writing technique, we can mention some forethought required for a 'good' design for e-beam writing. For example, during the writing, the beam cannot be deflected along a whole wafer. For this reason, the pattern is cut into fields of maximum $512\mu m$, starting from the lower left corner of the gds file. This field size depends on the resolution we use for the writing, and if you need a $5nm$ -resolution, the maximum field size will be $320\mu m$. Another crucial point for the electron beam exposure concerns markers and alignment between layers and for such purpose additional patterns have been added to control the correctly aligned and other parameters such as dose ($\mu C/cm^2$) or resolution. A $20nm$ resolution can be reached through this nano-fabrication technique. We can summarize briefly the lithographic processes as follows: (i) spin coating step to deposit PMMA (polymethyl methacrylate) / Copolymer bilayer resist stack, resulting in a step-like (T-gate) profile, (ii) exposure (e-beam writing) step, (iii) developing to remove the resist from the exposed regions, (iv) metal deposition and (v) lift-off process. The substrate used is a p-type silicon, which was covered with thermally grown $200nm$ thick SiO_2 . A Ti/Pt ($5nm/25nm$) bottom electrode was deposited and patterned via EBL and lift-off. A thin film of Ag_2S ($60nm$) was deposited by thermal evaporation (as described in the previous section) and patterned via EBL and lift-off. Finally, a Ag/Pt ($10nm/70nm$) top electrode was deposited on the Ag_2S by direct electron beam evaporation and patterned via EBL and lift-off. A first generation of ECM cross-points has been designed with cross-electrodes size of $1000, 500, 200, 100nm$ and their switching behavior has been investigated by electrical characterization. As shown in Fig. 3.9, a bipolar switching (i.e. with a complete hysteresis loop) is achieved under low values of applied bias voltage ($300mV$) and low current. When the bias was swept from 0 to $300mV$, the current suddenly increased at about $200mV$ due to the formation of a bridge (SET process).

The linear decrease in current while the bias was swept from $300mV$ to $0V$, indicates that the Ag filament bridged the two electrodes, resulting in a non-volatile operation. When the bias was swept from 0 to $-300mV$, the current suddenly decreased at a bias value of about $-100mV$ due to the annihilation of the filament (RESET process).

A second generation of ECM cross-points has been developed by scaling the device size, and by optimizing all the lithographic parameters cross-electrodes size of $80, 40, 30nm$ have been realized. In Fig.3.10b it is interesting to observe a different switching behavior with respect to the previous one, characterized by volatile loops in both polarities. This behavior can be understood by considering the fact that the smaller the device structure, the higher the filaments instability, thus the device volatility. Fig. 3.10a shows 10 non-volatile loops obtained with ECM cell with cross-electrode size of $200nm$, resulting in reversible redox-processes that can be controlled by changing the bias polarity. Due to the high mobility of the Ag^+ ions in

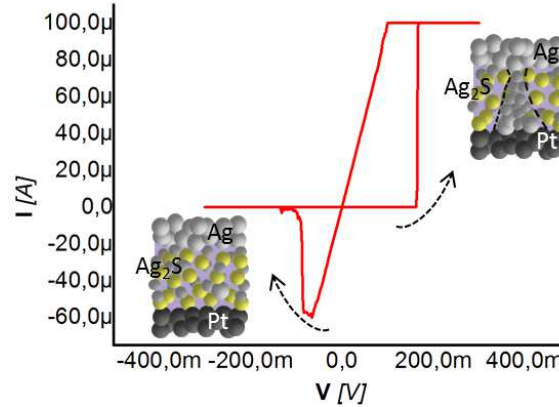


Figure 3.9: Nanoscale ECM cell configuration by EBL: I-V characteristics for cross-electrodes sizes of $200nm$. Positive applied bias induces Ag filaments formation resulting in the ON switch. Negative bias application causes dissolution of the precipitated Ag ions into the Ag_2S , resulting in the OFF switch.

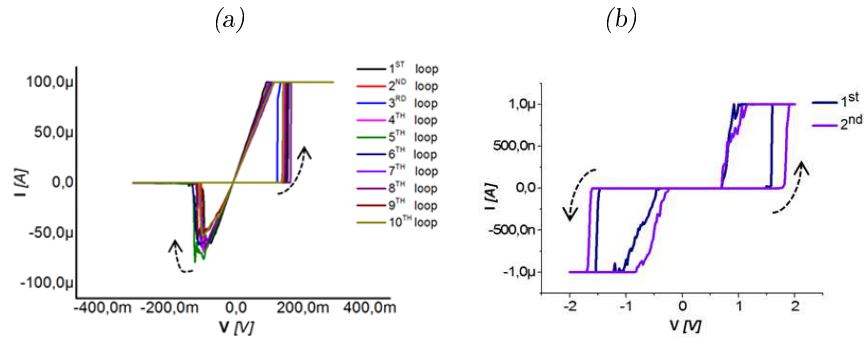


Figure 3.10: Nanoscale ECM cell configuration by EBL: (a) I-V characteristics for ECM cross-points with cross-electrodes sizes of: $1000, 500, 200, 100nm$. (b) I-V characteristics for ECM cross-points with cross-electrodes sizes of: $80, 40, 30nm$.

the Ag_2S ionic conductor, the device was operated at low voltages, close to the biological electrical potential recorded in neuronal cells during spiking ($200mV$ vs. $80mV$). This device configuration offers the potential for cross-bar integration (cross-point of metallic wires) and for the realization of dense synaptic arrays, as it will be discussed in the next section.

3.4.2 Bottom-up approach

Instead of precisely designed the ECM cross-points, a self-assembly of nanowires (NWs) has been adopted, following a bottom-up nano-fabrication approach.

Metallic nanowires (NWs) (Ag and Pt) are used as the bottom- and top-electrodes of the ECM cell, as shown in Fig. B.4, thus NW-NW cross-point is realized.

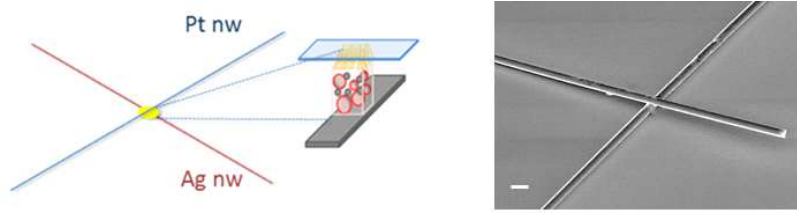


Figure 3.11: Nanoscale ECM cell configuration by self-assembly of NWs. Schematic and SEM images: scale-bar 200nm.

<i>Ag</i> NWs	<i>Pt</i> NWs
diameter \times length = 115nm \times 20 – 50 μ m	diameter \times length = 150nm \times 10 – 16 μ m
$2 * 10^{10}$ NW/mL in <i>Isopropanol</i> solution	$4 * 10^9$ NW/mL in <i>Ethanol</i> solution
total volume of 25mL	total volume of 4.5mL

Table 3.1: *Ag* and *Pt* NWs characteristics.

The building blocks of such nanoscale device configuration (i.e. the metallic NWs) were initially diluted in alcohol solvent and some of their characteristics, such as dimensions or concentration, are reported in table 3.1. Mixing and Sonication are the two techniques used to separate the aggregations of NWs in solution. After a careful tuning, it was possible to find a good compromise between NWs concentration and solvent dilution with mixing and sonication that allows us to deposit such nano-objects onto the silicon substrate. We used NWs transfer deposition by dip-coating (Fig. 3.12a). This technique is very simple, fast, not expensive and allows a good NWs orientation control. In fact, more than 200 NW-NW atomic switch cross-points can be obtained in a (0.84x0.84) cm^2 sample by performing a combine *Ag* and *Pt* NWs dip-coating in two orthogonal directions. After having localized the NW-NW cross-points by optical microscope, we designed electrodes by EBL. It should be noted that this strategy does not allow a precise and reproducible NW location control at large scale but a successful strategy for our purpose: an easy and fast way to investigate the I-V switching behavior of different NWs cross-point configurations, such as: *AgOx* or *Ag₂S*, as shown in Fig. 3.12b.

Figure 3.13a shows a volatile switching. This behavior seems reasonable if we take into account that by limiting the filament size during the forming, a lower stability is obtained, leading to a filament rupture.

There is a clear analogy between this switching behavior and the one in Fig. 3.10b relative to ECM cell cross-point with cross-electrodes sizes lower than 80nm, where the smaller volume/surface device confinement determines a

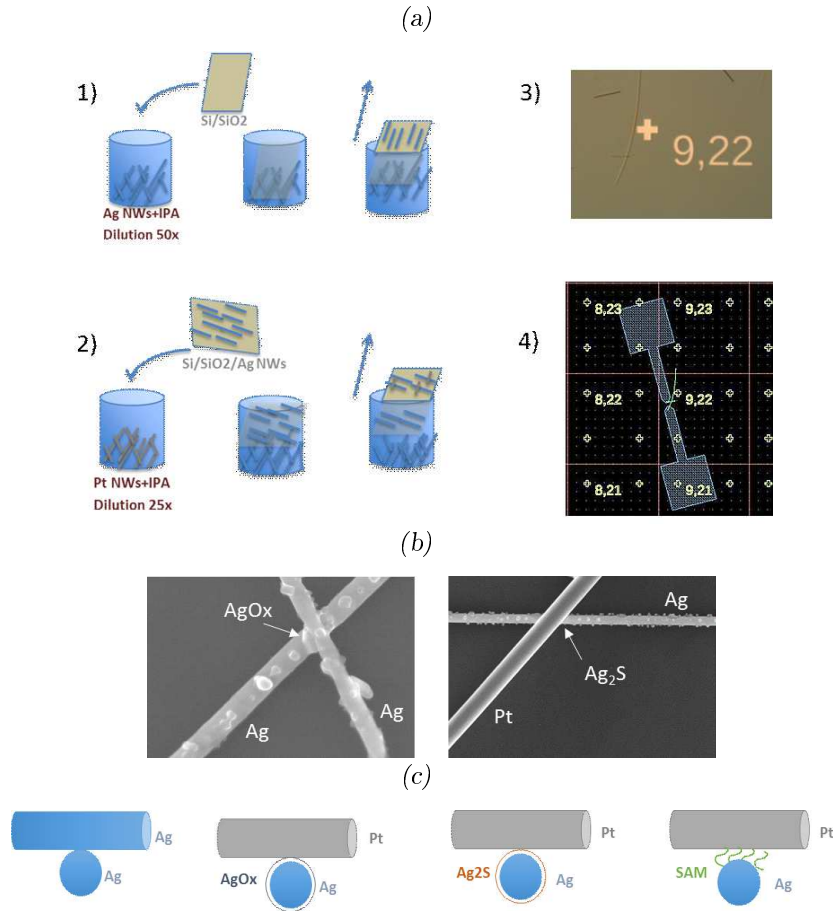


Figure 3.12: Nanoscale ECM cell configuration by self-assembly of NWs. (a) NWs transfer deposition by dip-coating steps. (b) Two NW-NW cross-point configurations examples: $Ag/AgOx/Ag$ and $Ag/Ag_2S/Pt$. (c) Schematics of different configuration under test.

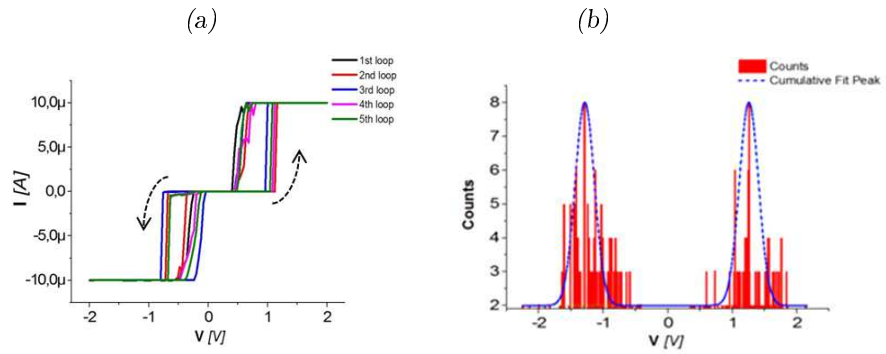


Figure 3.13: Nanoscale ECM cell configuration by EBL: (a) I-V characteristics showing volatile loops. (b) Variability in ON Voltage over 100 IV measurements.

filaments instability responsible of a completely volatile behavior.

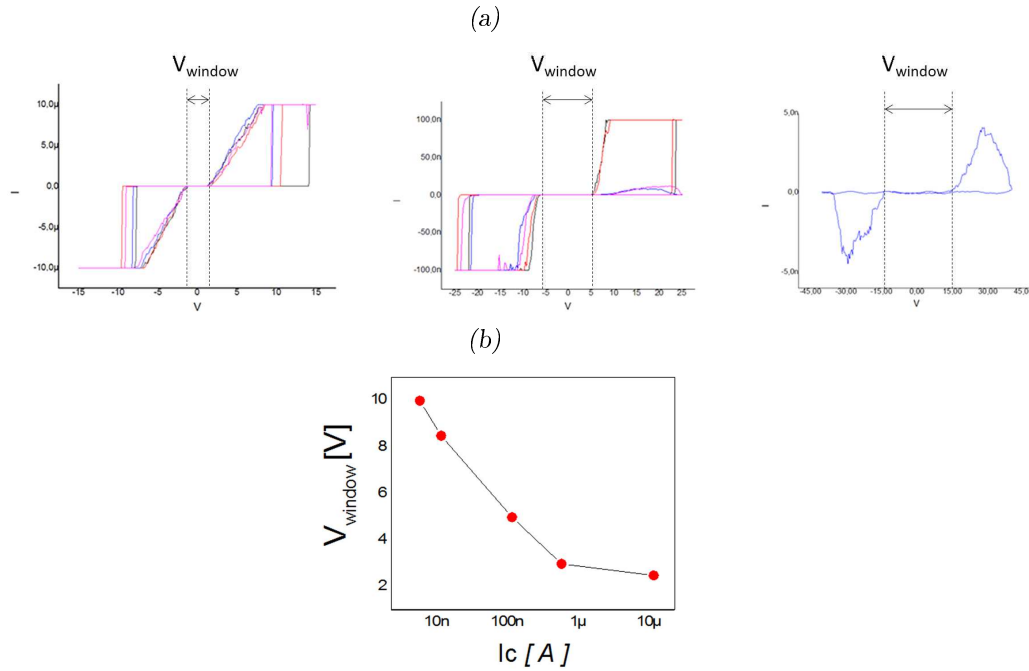


Figure 3.14: Nanoscale ECM cell configuration by EBL: (a) I-V characteristics showing volatile loops. (b) Stability Voltage window in function of the compliance current I_c .

Some preliminary investigations have been done concerning such unstable dynamics by considering the compliance current as key parameter to control the filament stability, technique that will be adopted and explain in more details in the next chapter. As it is possible to observe, all the IV characteristics in Fig.3.14a presents a sort of 'volatile' window, i.e. a voltage range in which the CF bridge is starting disappearing even if the applied bias is maintained. The higher is the instability window the lower is the compliance current adopted (Fig.3.14b).

Such very interesting volatile switching behavior required further research investigations from both physical and device operation point of view, and as far as we know there are not significant results in this regard.

3.5 Discussion and Perspectives

In this chapter, the technology used during this PhD work, the Ag_2S -ECM cell, has been introduced in the context of filamentary-type RRAM. Then, by motivating such technology choice in the context of neuromorphic computing, two different nanofabrication strategies, the top-down and the bottom-up approaches, have been performed to develop such nanoscale memory configuration.

By using a conventional top-down approach, we observed hysteresis loop in a range of $300mV$ applied bias, resulting in both a volatile and a non-volatile mode, depending on cross-electrodes sizes. By following a bottom-up approach of self-assembly of NWs, a volatile switching was observed.

In the next chapter we will demonstrate how to exploit such nanoscale memories to implement synaptic functionalities, additional properties that could enhance future computing paradigms.

Chapter 4

Synaptic Plasticity with Filamentary Switching

"Insanity is doing the same thing over and over again and expecting different results."

Albert Einstein (1879-1955)

4.1 Introduction

In order to match the efficiency of biological systems (BNNs), synaptic functionalities should be realized with a dedicated technology well suited for its implementation in neuromorphic NNET. In this context, the impact of emerging nanoscale memory devices has been presented in the first chapter from a functional point of view, in which *Synaptic Plasticity*, key element for information processing and storage, offers attractive functionalities embedded in a single component.

Thanks to their bio-mimetic aspect, in the second chapter, memristive devices have been described from practical point of view (i.e. how they can be used and integrated in neuromorphic NNET). In this bio-inspired computing context, we have adopted different nanofabrication technologies to realize filamentary-type memories (i.e. ECM cells), a top-down approach closer to the ANNs architecture and a bottom-up approach inspired by the BNNs one.

In this chapter, by going deeper into the expression of *Synaptic Plasticity* observed in biological synapses, we demonstrate that complex plastic behavior can emerge from ECM cells, offering a promising and interesting way to enrich and enhance future bio-inspired information computing systems.

4.1.1 Plasticity Key Parameters

Let's consider the nanoscale ECM cell configuration, consisting of Ag/Pt cross-points with a $200nm \times 200nm$ active area separated by Ag_2S , realized by top-down approach (Fig.3.8). Modification of the device's resistance (i.e. conductance) as a function of the bridging filament can show a direct analogy with biological synaptic processes observed during synaptic adaptation and learning. Growth of the conductive filament by electrical stimulation is associated to activity dependent synaptic potentiation (i.e. increase of the synaptic weight).

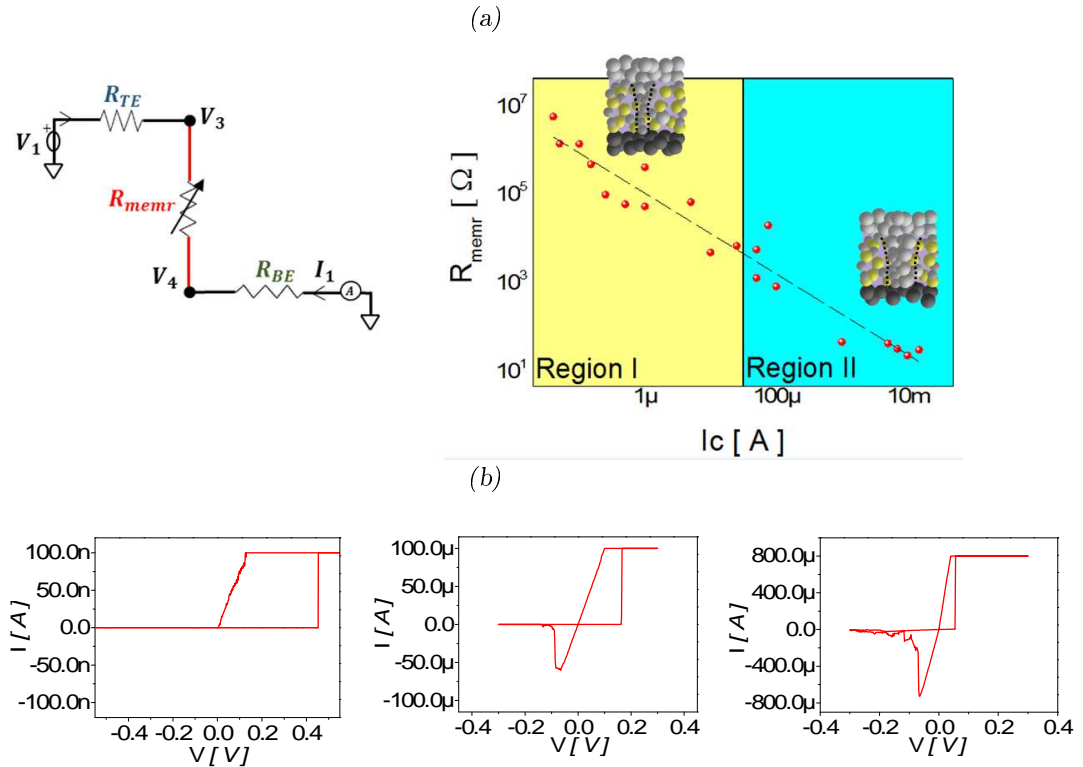


Figure 4.1: Synaptic Nano-device. (a) Schematic of the four-probes electrical measurements set-up (on the left): depending on the compliance current I_c , is possible to induce different ON states corresponding to different resistance values. (b) I-V switching characteristics for different values of the compliance current, I_c . When $I_c = 100nA$, the ON state is unstable and tends to relax very quickly (OFF transition is not measurable). When $I_c = 100\mu A$ or $800\mu A$, conventional bipolar switching hysteresis loops are obtained, corresponding to the stable ON state.

So far, the parameter associated to the CFs stability in RRAM has been the compliance current I_c , i.e. the higher is the current passing through a two-terminals device, the thicker is the filament diameter formed producing an higher stability. As presented in the previous chapter, by investigating the morphology of the filamentary switching with a fractal analysis, is possible to

extend the effects induced by such parameter in which larger filament expansion and dendritic tree density corresponds to a larger I_c (Fig. 3.6). For the moment let's focus the attention on the fact that, depending on the I_c used during the IV characterization, we can describe different ON states, where the resistance can be modulated via the filament size. As shown in Fig.4.1a, the higher is the I_c applied (range between $100nA$ to $10mA$) the lower is the resistance of the ON states (range between 100Ω to $10M\Omega$). Filament stability can be studied from the I-V switching characteristics (Fig.4.1b) and it is possible to distinguish two different regions: when $I_c = 100nA$ to $50\mu A$ (region I), the bridging filaments are very thin showing a high volatility; when $I_c > 50\mu A$ (region II), thicker filaments induce stable ON states. As expected, controlling the I_c value during SET transition limited the filament growth and tuned the ON conductance state. ON states at I_c values of $100nA$ to $50\mu A$ were strongly volatile, whereas ON states at I_c values above $50\mu A$ were stable, with RESET transition observed at a negative bias.

In order to investigate the conductive filaments stability we performed pulses measurements (Fig.4.2). After a first pre-condition step which consist of a IV sweep with a fixed I_c , pulses measurements are applied to the cross-point device. This second step is made of a first excitation part in which writing pulses V_{WRITE} induce SET transition and a second 'relaxation' part characterized by V_{READ} pulses, that without modifying the switching state, are used to study the resistance (i.e. conductance) state evolution over six decades of time.

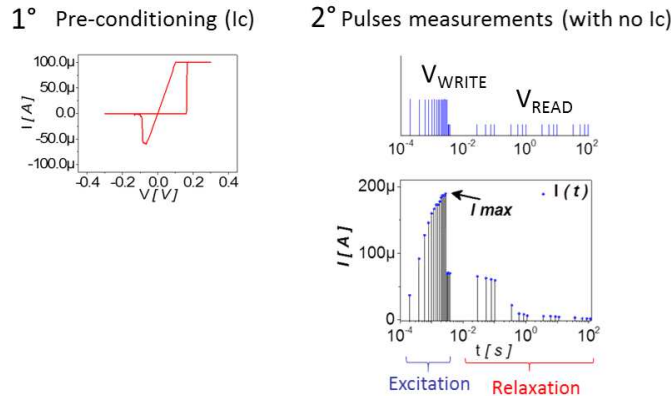


Figure 4.2: Schematics of the pulses measurements protocol: 1° pre-conditioning step: IV characteristic with a I_c ; 2° pulses measurements (free from I_c) consist of excitation part, that induce the SET transition and a relaxation part to investigate over 6 decades of time the evolution of the ON state.

By following this measurements protocol, we studied the parameters that allow us to control and tune the device volatility regimes. The first parameter that affects the device volatility is the I_c . By using pulses electrical stimulation (15 pulses with $V_{WRITE} = 0.21V$ in the excitation part and two V_{WRITE}

pulses after 10s and 100s to study the relaxation part, according to schematics shown in Fig.4.3a), it is possible to observe how, by changing the I_c , the resistance of the ON state can be modulated. By using $I_c = 800\mu A$ the resistance of the ON state is the same after 100s, showing a high filament stability (non-volatile behavior). On the contrary, with a lower compliance current $I_c = 100\mu A$ the resistance of the ON state after 100s is lower, that means a lower CFs stability (volatile behavior).

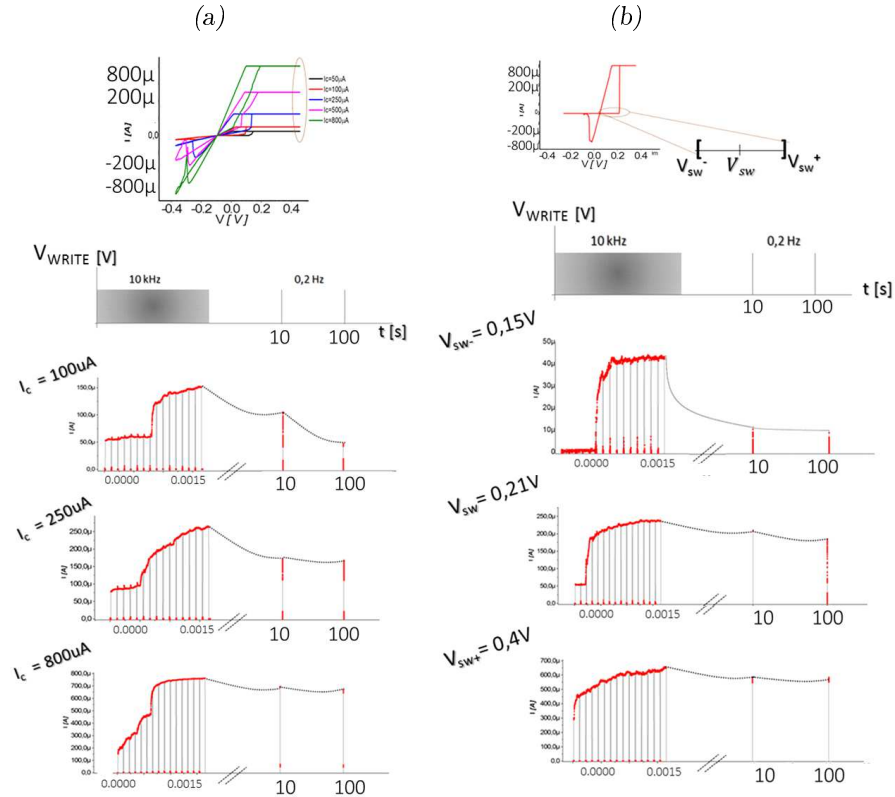


Figure 4.3: Device Volatility key parameters. (a) I_c : on top, different conditioning loops with different I_c ; on the bottom device electrical response to pulses measurements, showing the tunable device volatility. (b) V_{sw} : on top, a conditioning loop in which is highlighted the switching threshold voltage range; on the bottom device electrical response to pulses measurements, showing the tunable device volatility.

Another key parameter that can be used to control the device volatility is the switching voltage V_{sw} . As shown in Fig. 4.3b, we have taken into account the switching voltage range (i.e. between $V_{sw} = 0,15V$ and $V_{sw} = 0,4V$). By applying a sequence of pulses (15 pulses as in the previous case) at the same current compliance $I_c = 250\mu A$, it is possible to distinguish different volatile regimes. For $V_{sw} = 0,15V$, we can observe a decrease in the resistance of the ON state after 100s, showing a high filament instability (volatile behavior) or for $V_{sw} = 0,4V$ the same resistance of the ON state after 100s means an

higher CFs stability (non-volatile behavior).

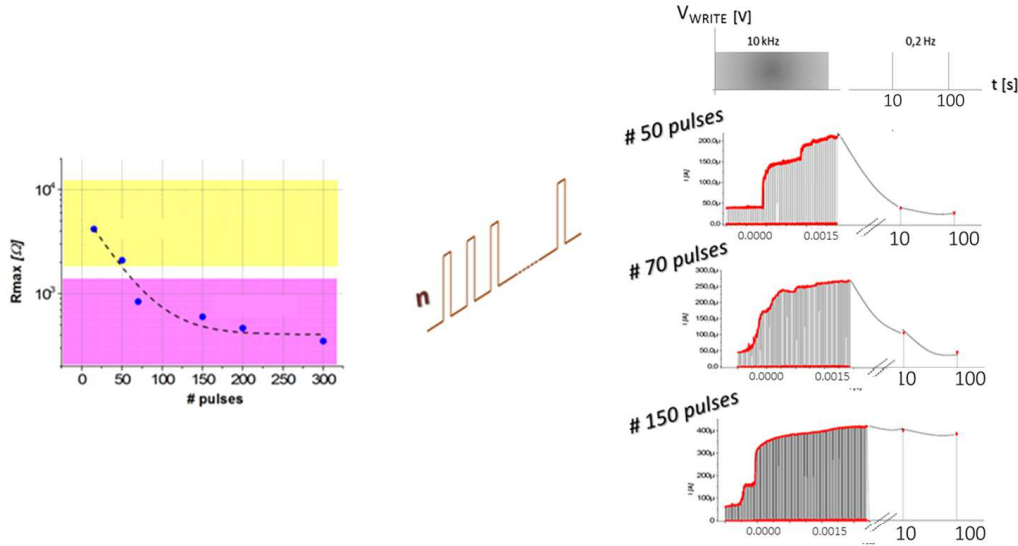


Figure 4.4: Device Volatility key parameter: Number of Pulses. On the left: 'the cumulative effect' ON state Resistance as a function of the number of pulses used in the electrical stimulation. On the right: device electrical response to pulses measurements, showing the tunable device volatility.

Similarly, by setting $V_{WRITE} = 0.21V$ (as done in the first case) and $I_c = 250\mu A$ (as done in the second case), it is possible to tune the device volatility by changing the number of pulses (i.e. spikes). As shown in Fig. 4.4, at the same frequency ($5kHz$), with an high number of stimulation pulses (150 pulses) the switching behavior induced is non-volatile, while, a low number of spikes (15 pulses), in the same conditions, is not enough to induce CFs stability inducing a volatile behavior.

4.2 Synaptic Nano-devices: Phenomenological Implementation

Modification of the synaptic weight as a function of neuronal activity (i.e., spiking activity) is widely recognized as a key mechanism for information processing and storage in neuromorphic NNET.

Inspired by such plastic behavior, in this section we will present how to tune ECM cell device conductance (in analogy with the synaptic strength) as a function of electrical stimulation. In particular, we will present key parameters that allow to induce different device volatility regimes. Then, from a physical point of view, we will describe the memory dynamical aspect of ECM cells in terms of time constant, parameter in which is integrated the device 'past history' and through which it is possible to extract information

concerning the STP to LTP transition.

4.2.1 Tunable Volatility regimes

After having introduced the key volatility elements that allow to tune the CFs stability, in this section we show how to achieve different volatility regimes in ECM cell cross-points. One of the aspects of major impact of this PhD work concerns the demonstration that more complex plastic behaviors can emerge from nanoscale memristive devices, thus allowing a greater number of features to be embedded in a single component and potentially permitting more complex computing systems.

A linear IV relationship, defining the ON conductance state G_{ON} , was obtained in all ON states (Fig. 4.1b), indicating that the filaments bridged the gap between the electrodes. Consequently, the large dynamic range of ON states presented in Fig.4.5 namely, from high resistance at low I_c (i.e. $1M\Omega$ at $100nA$, corresponding to a switching power $< 100nW$), to low resistance at high I_c (i.e., $1k\Omega$ at $1mA$, corresponding to a switching power of $300\mu W$) can be attributed to a modification of the bridging filament morphology, rather than to a modulation of the tunnel barrier length (which is a plausible mechanism in the case of a non-bridging filament). As a first level of

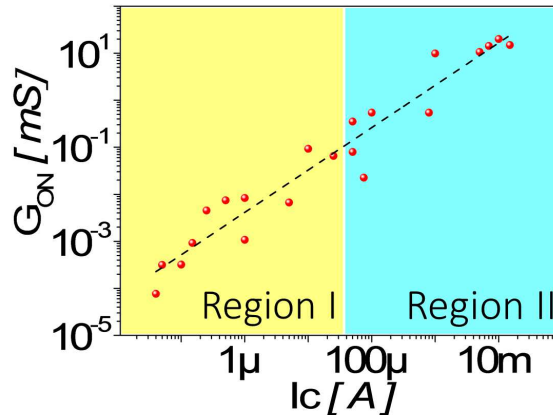


Figure 4.5: ON state conductance as a function of I_c . Limiting the current during SET limits filament formation. When $I_c = 100nA$ to $50\mu A$ (region I), the bridging filaments show a high volatility; when $I_c > 50\mu A$ (region II), the ON states are stable.

interpretation, the low I_c region can be reasonably described by weak filaments that tend to dissolve very quickly once the voltage is removed. The high I_c region can be considered to correspond to strong bridging filaments with slower relaxation. This effect has been described thermodynamically in Ag filaments (Hsiung et al., 2010) as a competition between the surface and volume energies: thin filaments tend to be disrupted because the sur-

face energy is higher than the volume energy, whereas thick filaments tend to stabilize because the volume energy is higher than the surface energy.

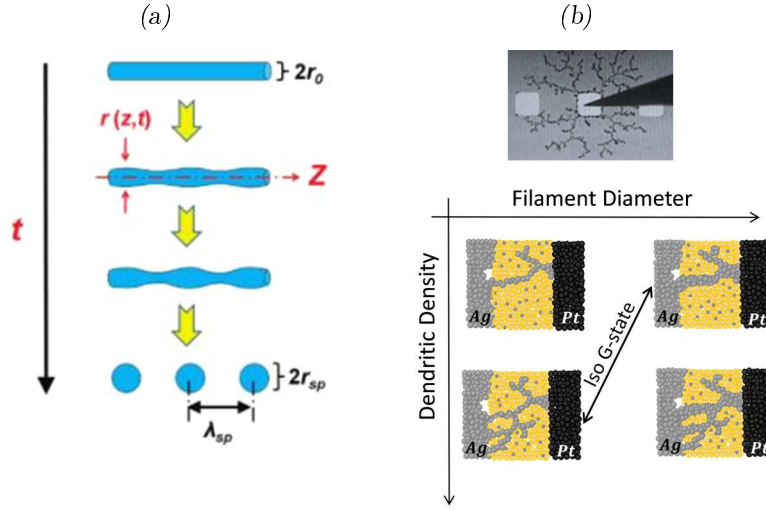


Figure 4.6: Nanoscale ECM cell configuration. (a) Rayleigh instability of the *Ag* filament induced by structure evolution of *Ag* filament. (b) Schematic of the proposed scenario describing switching in ECM cells. Both the density and diameter of the dendritic branches can induce an increase in the ON state. The isoconductance state can be obtained with two different filament configurations. On top optical image of *Ag* dendritic paths obtained with millimeter scale cross-point configuration.

In particular, Hsiung et al. (Hsiung et al., 2010) investigated the exact mechanism of the *Ag* filament structural evolution during the forming process when filament tends to stay very thin. It is so thin that it breaks up into a chain of nanospheres (according to Rayleigh instability) right after the formation has been completed, as depicted in Fig.4.6a. Let's assume that *Ag* spheres with the number of n have evolved from a *Ag* cylinder in which the length and radius of the cylinder are L and r_0 , respectively. The aspect ratio α is introduced and defined as L/r_0 . The average radius of the *Ag* sphere is r_{sp} and the volume is kept the same during evolution. Thus,

$$r_{sp} = \frac{3\alpha^{1/3}}{4n} r_0 \quad (4.1)$$

and the total free energy change $\Delta G = \Delta G_V + \Delta G_A$, where ΔG_V is the volume free energy, ΔG_A is the surface free energy. If the structural evolution is spontaneous, ΔG should be negative and thus, it follows that $r_0 < r_{sp}/1.5$ indicating that the *Ag* spheres are formed via the structural evolution from the *Ag* cylinder (filament) by reducing the surface energy (surface tension). Such relaxation of the conductive paths has been reported in nanoscale devices (Ohno et al., 2011b; Yang et al., 2012a) and was the basis for the implementation of *STP* and the *STP* to *LTP* transition.

After the conductive filament forms via a strong stimulation, the filaments tend to dissolve and the device relaxes toward its insulating state, leading to *STP* behavior. Stronger stimulation of the device during the SET transition leads to stronger filaments and higher conductance states with more stable characteristics, resulting in *LTP*. In this case, the conductance state is correlated directly with the volatility. Assuming that similar dendritic processes occur at the nanometer and millimeter scales (Fig. 3.5 and on top of Fig.4.6b), we can draw a more complex picture for the interpretation of filament stability. Specifically, the different ON states can be described by dendritic trees, in which the resistance is modulated equally by the density and diameter of the branches. At the nanoscale, the same ON state can be obtained by filaments with dense and thin branches as can be obtained by filaments with less dense and thick branches (Fig.4.6b). Both configurations should lead to different volatilities, emulating different plasticity properties, as it will be demonstrated in the following section.

4.2.2 STP to LTP Transition

To evaluate the plasticity properties of our electronic synapses, we performed pulsed measurements, as done previously, with simplified pulses equivalent to the spike rate-coding scheme observed in biological networks. First, a full SET and RESET cycle was realized by voltage sweeping and limiting the current in the SET transition, with the conditioning loop resulting in an initial OFF state equivalent to Figure 4.1b. Then, the device was exposed to a train of pulses ($5kHz$) with fixed amplitude ($0.42V$) and width ($100\mu s$), resulting in potentiation of the device (i.e., conductance increase). Relaxation of the synaptic efficiency was sampled over six decades of time by short read pulses with lower voltage ($0.1V$) and short duration ($100\mu s$), to minimize the effect on the relaxation mechanism (Figure 4.7a). Different excitatory bursts, obtained by varying the number of pulses, were used to modulate the potentiation obtained at the end of the pulse sequence, corresponding to the conductance at the end of a burst of pulses, G_{max} . These bursts were fitted by a simple exponential function, ($y = Ae^{-x/t}$, Figure 4.7b). Consistent with our previous observation that low stability is obtained at a low ON state due to the thinner filaments, we obtained a short relaxation time constant for the lowest ON state. Increasing G_{max} led to a higher time constant and more stable filaments. When we analyzed the evolution of the relaxation time as a function of G_{max} for different I_c values during the conditioning loop (Figure 4.7c), a second parameter for volatility control emerged. At high I_c values, there was a sharp transition between the low and high time constants. A smoother transition was obtained as G_{max} increased when lower I_c values were used.

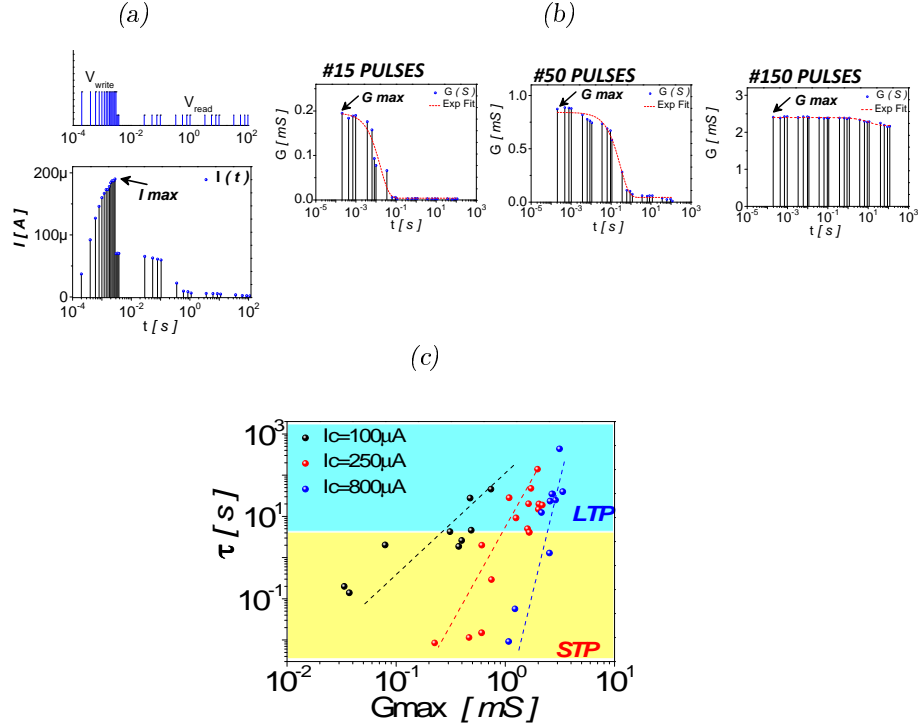


Figure 4.7: Device Volatility Characterization. (a) Protocol for the measurement of pulse relaxation. A burst of pulses at $5kHz$ ($V_{write} = 0.42V$) induced potentiation. Current relaxation was measured at a lower voltage ($V_{read} = 0.1V$) over six decades of time. (b) Measurements of conductance relaxation (blue points) and fitting (red line) on six time decades for different potentiation (G_{max}) values, obtained by varying the number of pulses (15, 50, and 150 pulses). Low and high G_{max} values led to STP (complete relaxation over time) and LTP (no relaxation over time), respectively. (c) STP to LTP Transition: relaxation time constant as a function of I_c and conductance state at the end of the burst of pulses, G_{max} .

4.3 Synaptic Nano-devices: Causal Implementation

In this section we show that a more complex filament shape, such as dendritic paths of variable density and width, can permit the short- and long- term processes to be tuned independently, by achieving a flexible way to program the device memory (i.e. the synaptic weights) and the relative device volatility.

4.3.1 Synaptic Adaptation Implementation

Another formulation of our results describing the STP to LTP Transition implementation is presented in Figure 4.8. If we consider the conductance state 100s after the end of the excitatory burst, then different transitions from STP (relaxation of the conductance state after 100s; $G_{max} > G_{100s}$)

to *LTP* (no relaxation of the conductance state after 100s; $G_{max} \simeq G_{100s}$, blue area in Figure 4.8a) can be identified as a function of I_c . This behavior can be attributed to the combination of two effects. Namely, both I_c and the strength of the excitatory burst (i.e., number of pulses) contribute to the definition of the conductive paths. After the conditioning loop, the device is in its OFF state. Traces for the remaining dendritic branches (defined by I_c) correspond to preferential paths for filament formation during the excitatory burst. By analogy with filament formation obtained on millimeter-scale

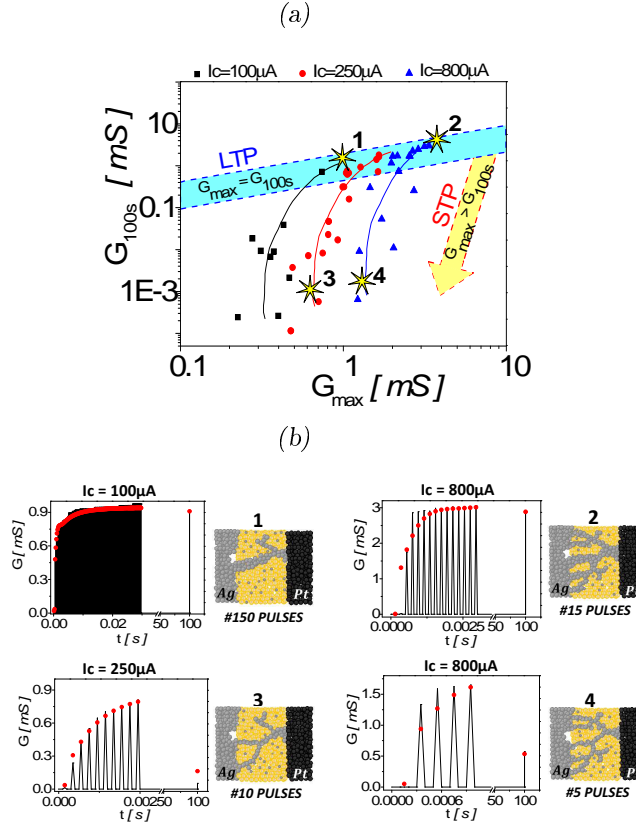


Figure 4.8: Implementation of the Synaptic Adaptation through STP to LTP transition. (a) After a conditioning loop (full SET and RESET cycle with current compliance, I_c), the device is stressed with a burst of spikes, which induce a potentiation from the OFF state to a final conductive ON state, G_{max} . Device conductance is measured 100s after the end of the burst to evaluate the relaxation. Different transitions from *STP* to *LTP* are obtained with different conditioning I_c values ($I_c = 100, 250, 800 \mu A$). (b) Two examples of *LTP* (cases 1 and 2) and *STP* (cases 3 and 4), for the case in which the number of pulses is set as the key plasticity factor and the I_c value is set as the dendritic path definition. The density (through I_c) and diameter (through burst excitation) of the dendritic branches can be tuned independently to reproduce various *STP/LTP* combinations.

devices, higher I_c should lead to denser dendritic trees. Thus, the first parameter for plasticity tuning is the I_c value used during conditioning. This

value controls the average conductance of the filament during switching in pulse mode, by defining the switching path (i.e., dendrite density). The second parameter that controls the *STP* to *LTP* transition is the excitation strength (i.e., number of pulses, which controls G_{max}). This parameter can be associated with an increase of the branch diameter. These two parameters, the past history of the device through the conditioning loops, and the spiking activity during potentiation can be changed independently of each other to modify the device conductance and the filament volatility.

To illustrate the improved functionality obtained with our approach, we used the biological model of synaptic plasticity developed by (Markram et al., 1998) to fit our different synaptic potentiation experiments (Fig.4.8b). This model describes the excitatory postsynaptic potentiation response produced by a train of presynaptic action potentials (APs) and it will be described in more details in the Chapter 4. Two examples of *LTP* (cases 1 and 2) and *STP* (cases 3 and 4) are shown, for the case in which the number of pulses is set as the key plasticity factor and the I_c value is set as the dendritic path definition. The density (through I_c) and diameter (through burst excitation) of the dendritic branches can be tuned independently to reproduce various *STP/LTP* combinations. From a practical perspective, we believe that developing devices that are more functional (i.e., have properties closer to biological synapses) will allow the construction of more complex systems (La Barbera et al., 2015).

4.4 Conflict between Phenomenological and Causal approach

In a previous report describing the *STP* to *LTP* transition (Ohno et al., 2011b; Kim et al., 2013a), the transition was controlled by a single parameter (i.e., device conductance). We argue that the rate-coding property obtained in the *STP* regime, as observed in the facilitation of synaptic signal transmission during a high frequency burst of spikes and the subsequent relaxation at lower frequencies, disappears once the device enters into its *LTP* regime and, thus, becomes a linear resistor. From a circuit perspective, if we consider a simple integrate-and-fire neuron associated with linear synapses, the node (neuron and synapses) is equivalent to a simple linear filter (if the variable is the average spiking rate). The node is a nonlinear filter in the *STP* regime with frequency-dependent synaptic conductance. The overall network functionality is reduced when learning moves synapses from their *STP* to their *LTP* domain. For the device presented in this chapter, synaptic adaptation can be realized by modifying the dendritic filament density. The frequency coding property can be ensured by controlling the filament diameter and relaxation.

4.5 Discussion and Perspectives

In this chapter, we demonstrated that the basic physics involved in the filamentary switching of ECM cells can reproduce important biological synaptic functions that are key mechanisms for information processing and storage. We report a single synaptic device that highly resembles its biological counterpart, opening the field to more complex neuromorphic systems.

By referring to the plasticity mechanisms classification proposed in the first chapter, the *transmitter-induced plasticity* corresponds to the synaptic adaptation, a non-Hebbian plasticity form. In this context, the *STP* to *LTP* transition has been well demonstrated in variety of nanoscale memory devices. In particular, the transition between *STP* to *LTP* was so far associated to a single parameter (such as the mean firing rate of the pre-neuron) and both ST and LT regimes cannot be uncorrelated (i.e. ST will lead LT regime). The device state will move sequentially from one regime to another one via *Transmitter-Induced plasticity* only. In this chapter we demonstrated that a more complex filament shape, such as dendritic paths of variable density and width, can permit the short- and long- term processes to be tuned independently, by achieving a flexible way to program the device memory (i.e. the synaptic weights) and the relative device volatility. These results represent an original solution to the conflict between the causal and phenomenological plasticity description being closer to the complexity of biological synapses.

Synaptic Adaptation has been successfully implemented in our nanoscale memristive device by considering the filament stability of ECM cells, in terms of competition between the density and diameter of the dendritic branches. *STP* and *LTP* regimes can be controlled by tuning the device volatility. The first parameter for plasticity tuning, I_c , is used during conditioning and controls the average conductance of the filament during switching in pulse mode. The second parameter handles the *STP* to *LTP* transition and corresponds to the excitation strength (number of pulses), which controls G_{max} . The second parameter can be associated with an increase of the branch diameter. These two parameters can be tuned independently of each other to modify the device conductance and filament volatility.

Future work should investigate how such synaptic properties can be advantageous for large-scale neuromorphic circuits.

Chapter 5

Multiple Plasticity Mechanisms with Filamentary Switching

*"Working hard for something we don't care about is called stress.
Working hard for something we love is called passion."*
Simon Sinek, 1973

5.1 Introduction

In the first chapter, we have classified some forms of *Synaptic Plasticity* well known in BNNs. In the second chapter, we have described how nanoscale devices can be exploited and integrated in neuromorphic NNET. In the third chapter, we have presented the experimental part of this PhD work relative to the development and characterization of a particular class of filamentary switching, the ECM cell. By exploiting the physical properties and the dynamic device volatility of such nanoscale memory device, we have demonstrated in the fourth chapter, how to successfully reproduce and control fundamental processes observed in biological synapses.

In BNNs a combination between long term synaptic processes (Long Term Potentiation and Depression, LTP and LTD) and short term mechanisms (Short Term Plasticity, STP) contributes to the processing and storage of information. Individually such forms of synaptic plasticity such as Short Term Plasticity, Short Term to Long Term Plasticity transition or STDP have already been successfully implemented in this class of filamentary switching devices.

In this chapter, we demonstrate that ECM cells can be controlled and programmed to reproduce advanced bio-inspired features in which all these synaptic features can be realized and independently controlled in a single memory element thus providing a more general solution for the development of bio-inspired circuits.

5.2 Synaptic Nano-devices: Phenomenological Implementation

By referring to the analysis done in the first chapter, *Synaptic Plasticity* can be implemented by following two different strategies. The first one is by a phenomenological approach, devoted to replicate the spike transmission observed at the synaptic connection in BNNs without considering necessarily the physical origin of the synaptic weight modification. Demonstration of STP corresponding to synaptic weight potentiation or depression followed by a relaxation on short time scales (from milli-seconds to seconds) has been proposed in various systems, (Alibart et al., 2010), (Yang et al., 2013b) and (Chang et al., 2011a), while the importance of such mechanism for computing was not put forward. Following this line, STP to LTP transition (i.e. corresponding to a decrease of the volatility of the synaptic weight when its conductance is increased) was also demonstrated and interpreted as a possible signature of memory consolidation mechanism (i.e. in the sense of psychology) while non-trivial functions based on it, are still unclear (Ohno et al., 2011b), (Wang et al., 2012) and (La Barbera et al., 2015).

Another implementation strategy is based on a causal approach, that focuses on the origin of the synaptic weight modification without necessarily implementing bio-realistic signals and synaptic dynamics. For example, STDP was successfully implemented in various memory devices based on the principle of overlapping pulses that converts conveniently the time correlation of pre- and post-spike signals into voltages applied across the memory element and induces a synaptic weight modification replicating the STDP window of biology (or some variation of it). If attractive function can be realized with this learning algorithm, its implementation was mostly deterministic in a non-volatile regime that do not reproduce the richness of plastic behaviors observed in BNNs such as STP or STP to LTP transition.

In this chapter, we will present how, by taking advantages of both approaches, *Synaptic Learning* can be implemented in ECM cell cross-points by reproducing multiple plasticity mechanisms with different volatility regimes.

5.2.1 Tunable Volatility regimes

As presented in the third chapter, let's consider a filamentary memristive device, fabricated in a cross-point configuration of $(200 \times 200) \text{ nm}^2$ with *Ti/Pt* bottom electrode, *Ag₂S* ionic conductor and *Ag* top electrode (inset Fig.5.1a). The basic switching mechanism during SET (ON transition) is based on the oxidation of *Ag* into *Ag⁺* at the top electrode, reduction of *Ag⁺* ions into conductive *Ag* filaments across the ionic conductor while RESET (OFF switching) corresponds to *Ag* oxidation from the filaments and reduction to the top electrode. Such reversible switching effect present bipolar switching characteristics (Fig.5.1a).

In the fourth chapter, we reported a detail analysis of filament stability: the evolution of the conductance of the device under pulse stimulation at various frequency corresponds to a competition between filament growth induced by pulses of voltage with positive polarity and *Ag* filament dissolution associated to a competition between surface and volume energy in the filament (i.e. natural relaxation when the device is at rest). Based on this mechanism both STP and LTP were successfully realized (Fig.5.1c).

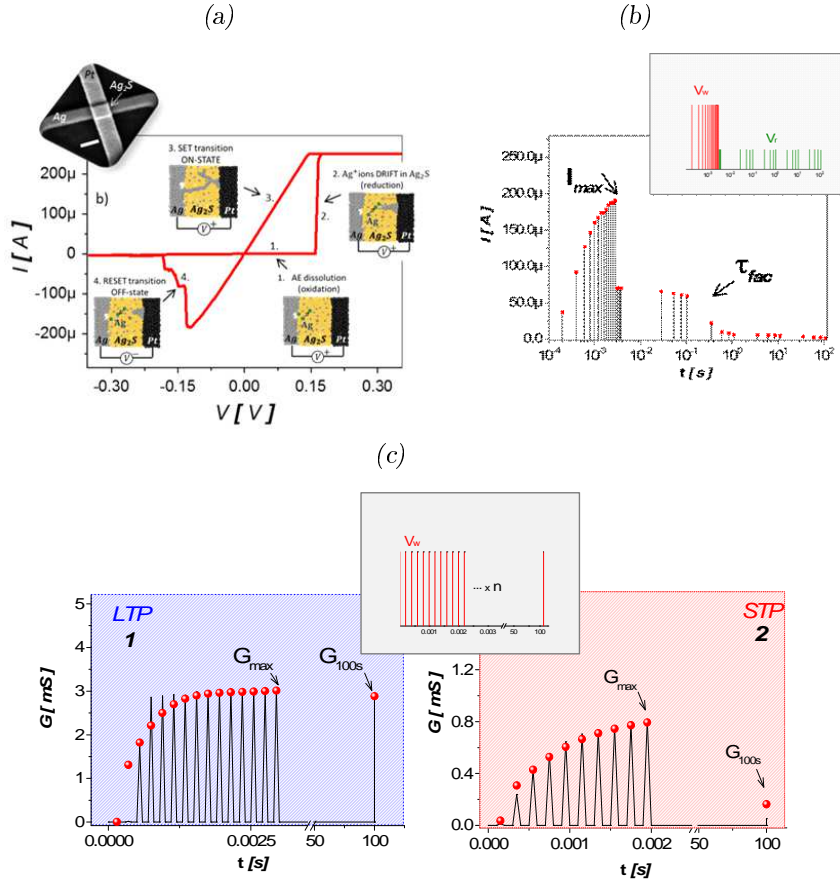


Figure 5.1: Phenomenological Synaptic Plasticity in ECM cells. (a) Device configuration (SEM images: $200nm \times 200nm$ cross-point active area) and I-V characteristics. (b) Protocol for the measurement of pulse relaxation. A burst of pulses at $5kHz$ ($V_{write} = 0.42V$) induced potentiation. Current relaxation was measured at a lower voltage ($V_{read} = 0.1V$) over six decades of time. (c) Spike-train based measurement protocol for Synaptic Plasticity implementation and two examples of LTP (case 1) and STP (case 2).

STP corresponds in this case to a pulse induced potentiation (i.e. increase of conductance G) followed by a decrease of conductance (i.e. device relaxation) with a characteristic time constant τ_{fac} (Fig.5.1b). This phenomenological description reproduce the plasticity observed in BNNs in facilitating synapses. It corresponds to a transmitter-induced form of plasticity that

depends only on the pre-neuron spiking activity and thus belongs to a non-Hebbian form of plasticity referred to as synaptic adaptation rather than synaptic learning. Markram proposed a phenomenological model describing such STP in BNNs (Markram et al., 1997). As it will be explained in more details in the next chapter, this model was adapted to our memristive device (red points, Fig.5.1c) to describe the conductance evolution during constant frequency pulse potentiation and subsequent relaxation. Such non-Hebbian synaptic adaptation plasticity induces interesting features for computing. Indeed, a burst of activity at the pre-neuron will induce strong potentiation (i.e. increase of the synaptic conductance) and increase the probability of post-neuron firing. More generally, facilitating STP provide non-linear synaptic response (i.e. frequency dependent response) that should play a key role in spike-based computing.

5.2.2 STP to LTP Transition

As described in the third chapter, since the filament relaxation is due to a competition between surface and volume energy in the filament (Hsiung et al., 2010), different levels of volatility can be obtained by modifying the filament morphology. Thin metallic filaments associated to a low conductance state (weak potentiation) presented strong volatility and short relaxation time constant τ_{fac} , while thick filaments with high conductance state (strong potentiation) were more stable and presented long τ_{fac} (Fig.5.2a). These two behaviors originate the STP and LTP, respectively.

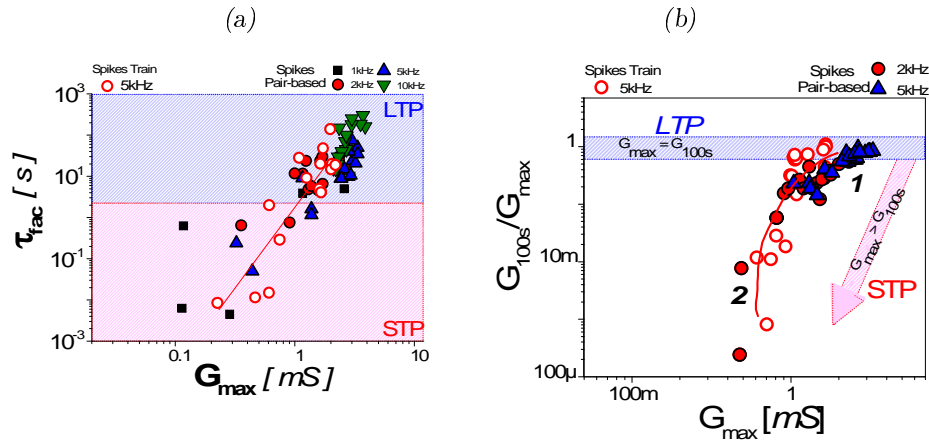


Figure 5.2: STP to LTP Transition in ECM cells. (a) Relaxation time constant τ_{fac} as a function of the conductance state at the end of the burst of pulses, G_{max} . (b) Synaptic Plasticity as a function of the device conductance measured 100s after the end of the burst to evaluate the G relaxation.

We previously reported an additional feature embedded in ECM cells corresponding to STP to LTP transition observed when the device potentiation

was increased. Practically, this plastic feature depends on the conductance state reached after potentiation, and can be obtained by controlling the switching parameters such as pulse amplitude (i.e. pulse voltage will define the amount of potentiation induced by each pulse), pulse train frequency (interval between pulses define the amount of relaxation between two successive pulses) or total number of pulses (accumulation of potentiation obtained after application of a given number of pulses). A convenient representation of the different relaxation time was proposed by considering the ratio G_{100s}/G_{max} as a metric for STP/LTP evaluation with G_{100s} the conductance of the device after 100s of rest and G_{max} the conductance immediately after the end of the potentiation (Fig.5.2b). LTP was associated to G_{100s}/G_{max} close to 1 while STP corresponded to $G_{100s}/G_{max} \ll 1$.

STP to LTP transition in nanoscale memory devices reported to date (Ohno et al., 2011a), (Yang et al., 2012a), was induced by controlling the pre-neuron activity only. Thus, this synaptic changes can be referred to as a non-Hebbian form of plasticity. In BNNs, LTP has been first evidenced through the application of high frequency stimulation (tetanus-induced potentiation) at the ore-neuron side that was associated to the opening of NMDA (N-methyl-D-aspartate) receptors. Thus, from a phenomenological point of view, STP to LTP transition reported previously was consistent with biology. If we now consider the physical origin of LTP induction in BNNs, we have to consider that high frequency stimulation also induces a strong depolarization of the post neuron membrane which is a key element for the opening of the NMDA receptors. In other words, not only the high frequency stimulation from the pre-neuron has to be taken into account but also the post-neuron state.

Along this line, from a causal description of LTP induction, later results in BNNs have evidenced that LTP was induced by learning (Whitlock et al. (2006) in its large sense, or more particularly by STDP (Markram et al., 1997). In this case, LTP is induced by an hebbian form of plasticity (i.e. correlation of pre and post-neuron activity) and the proposed implementations of STP to LTP transition in nanoscale memory devices fail to reproduce LTP induction. At the computational level, it would be highly valuable to be able to dissociate non-hebbian plasticity such as pre neuron induced plasticity from hebbian plasticity involving both pre and post-neuron activity correlation. In this case, it would be possible to dissociate synaptic adaptation from learning, as it will be explain in the next section.

5.3 Synaptic Nano-devices: Causal Implementation

5.3.1 Synaptic Learning Implementation

Hebbian STDP corresponds to an increase of the synaptic weight when time correlation between pre- and post-neuron firing is experienced at the

synaptic connection, i.e. firing of both pre- and post-neuron happens during a time correlation window called 'STDP window'. In our case, not only the instantaneous potentiation should increase during STDP events, but also the LTP characteristic of this synaptic weight modification. The first aspect (i.e. potentiation) is well described by studying G_{max} evolution during STDP events while the second one (i.e. LTP induction) is measured by recording the G_{100s}/G_{max} value after an STDP experiment. Experimentally, we developed a STDP protocol based on the repetition of 10 STDP events, i.e. pre- and post-spike correlation (Fig.5.3a). The spike used for this protocol were simple square-shaped pulses.

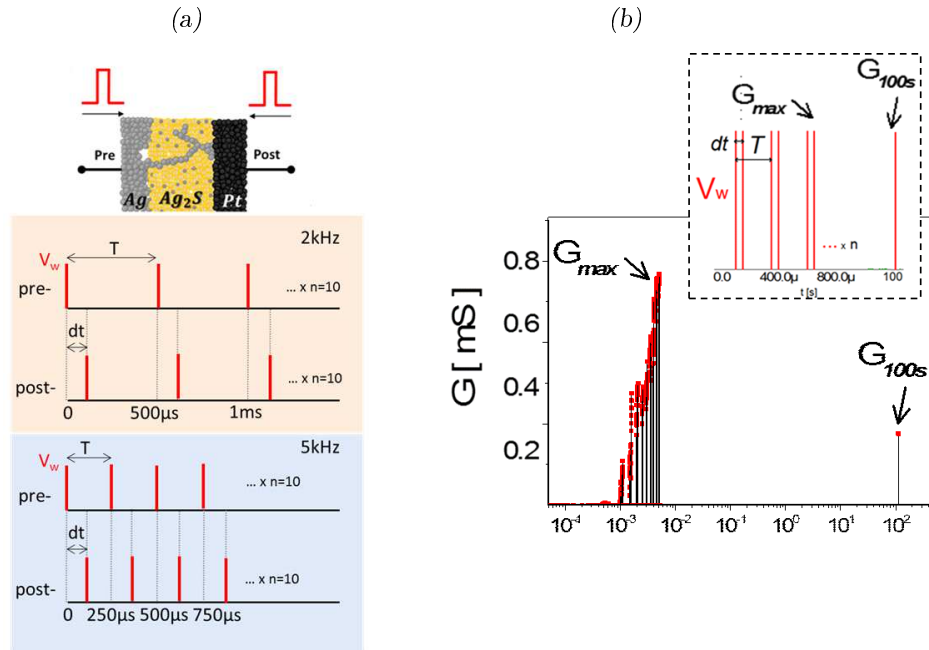


Figure 5.3: Causal Synaptic Plasticity Implementation in ECM cells. (a) STDP protocol based on the repetition of 10 STDP events (pre-before-post). (b) Protocol for the measurement of pulse relaxation after learning. A burst of pulses at a mean frequency $\langle f \rangle$ of pre-neuron firing and different dt ($V_{write} = 0.42V$) induced potentiation. Current relaxation was measured at a lower voltage ($V_{read} = 0.1V$) over six decades of time.

Two parameters were tuned during the STDP experiment: (i) the time correlation between pre- and post- pulses dt and (ii) the mean frequency $\langle f \rangle$ of pre-neuron firing associated to a period T . All the experiments started from a resting state of the ECM cell (i.e. low conductance or OFF state). G_{max} corresponds to the final conductance state at the end of the STDP protocol. After each STDP protocol, LTP-induction was evaluated by applying a single pre-pulse after 100s of rest and measuring the conductance G_{100s} . In order to evaluate the device volatility response to STDP experiment (i.e. extrapolate the characteristic time constant τ_{fac}), as done before, current

relaxation was measured at a lower voltage ($V_{read} = 0.1V$) over six decades of time (Fig.5.3b).

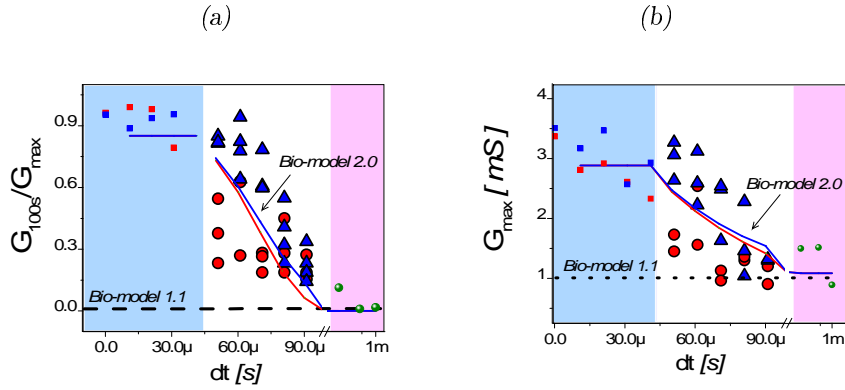


Figure 5.4: Synaptic Learning Implementation in ECM cells. (a) G_{100s}/G_{max} conductance change as a function of the time correlation between pre- and post-spike dt . (b) Similarly, G_{max} conductance change as a function of the time correlation between pre- and post-spike dt .

The STDP results obtained from Ag_2S -based ECM cells are shown in Fig.5.4. A clear increase of potentiation from $1mS$ to $3.5mS$ and LTP induction (from 0 to 1) is measured for time correlation such as $dt < 100\mu s$. In addition, when dt is decreased, this effect is strengthened, thus reproducing gradual STDP windows observed in biology. Time correlation dt smaller than $50\mu s$ resulted in pre- and post-pulse overlapping (pulse width was $50\mu s$). Since large voltages are obtained in this case (i.e. $2 \cdot V_{write}$), fully potentiated weights ($G_{max} = 3.5mS$, squared points (blue region) in Fig.5.4a and Fig.5.4b in the LTP regime ($G_{100s}/G_{max} = 1$)) were measured. Control experiments (green points in the pink regions) with pre-neuron spikes only were performed and showed weak potentiation ($G_{max} = 1mS$) and no LTP ($G_{100s}/G_{max} \ll 1$). Interestingly STDP measurements also show a rate based effect corresponding to higher LTP induction when the STDP protocol was realized at higher frequencies (i.e. $5kHz$ vs. $2kHz$). If the STDP-induced LTP when $dt < 50\mu s$ is straightforward and reminiscent of conventional STDP implementation in memristive devices, based on pulses overlapping, the LTP induction observed for non overlapping pulses suggest the presence of internal dynamics at short time scale (i.e. below $100\mu s$ range).

The STDP implementation demonstrated in this chapter corresponded to an hebbian form of plasticity (i.e. no anti-Hebbian corresponding to synaptic weight depression). If it is well known that Hebbian only potentiation should lead to network failure (i.e. potentiation only leading to saturation of all the synaptic weight to their max conductance state), this effect is balanced in our case by a natural relaxation of the weights (i.e. natural depression) that

tends to decrease the synaptic weights of weak synapses toward their low conductance state. After learning, the network should present a bimodal distribution of the weight while during learning and operation, all the intermediate values of conductance can be reached.

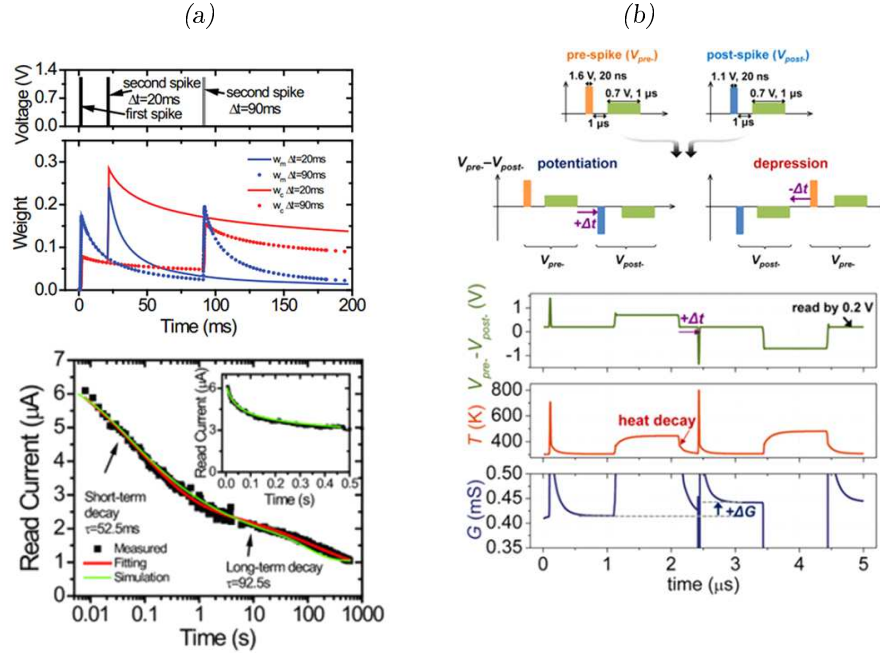


Figure 5.5: Origin of Synaptic Learning in memristive devices. (a) (On the top) Simulation results illustrating how the short-term behavior affected long-term weight change. The difference in long-term weight is caused by the different values of residue w_m at the moment when the second pulse is applied. State variable w_c and w_m are shown with interval between pulses $\Delta t = 20, 90\text{ms}$. (On the bottom) Relaxation measurements showing short- and long- decays constants. Adapted from (Du et al., 2015). (b) Simulation results showing the spike pair, in internal temperature evolution and the device conductance evolution during a spike pair with $\Delta t = 300\text{ns}$. Adapted from (Kim et al., 2015).

In the next chapter, we will explain how the Markram biological model of *Synaptic Plasticity* (Markram et al. (1997) provides a good qualitative description of the evolution of potentiation and LTP-induction ('*Bio-inspired model 2.0*') by considering physical phenomena at short-time scale.

The origin of the STDP function could be mainly explained by two physical effects reported in ECM memory devices. The first one (i) relies on the non-linear conductance relaxation in filamentary devices that was recently proposed by Du et al. (Du et al., 2015). In such systems, different slope of conductance relaxation in time after potentiation were reported (Fig. 5.5a). Each region of the conductance relaxation was attributed to short term plasticity and long term plasticity while their connection was conveniently associated to STP to LTP transition. This model was able to describe both STP

and STDP measurements without pulse overlapping.

Following a similar approach, we performed measurement of conductance relaxation in time from $500ns$ to $100s$. Short time scale regime were not observable in our setup. Since current (i.e. conductance) measurement in short time scale (i.e. below $10\mu s$) becomes challenging and questionable, we cannot completely rule out this possibility but absence of short time scale relaxation is a first indication of other effect involved in short time scale interactions between two successive pulses.

A second effect (ii) that could reasonably explain the short time scale interaction is based on recent works from Kim et al. (Kim et al., 2015) in which a second order memristor model is introduced to describe temperature effects in phase change materials (Fig. 5.5b). In their experiments, the correlation between pre- and post-pulses was implemented by adding in the pre-spike signal an additional heating pulse that strengthened the effect of the post-pulse on the conductance when overlapping between heating pulse and post-pulse occurred.

A possible explanation of short time scale interactions in our devices could be attributed to similar heating effects and subsequent heat dissipation after switching. A second pulse following a prior impulse can benefit from local heating in the switching region of the filament and increases the effect of this second excitation on potentiation.

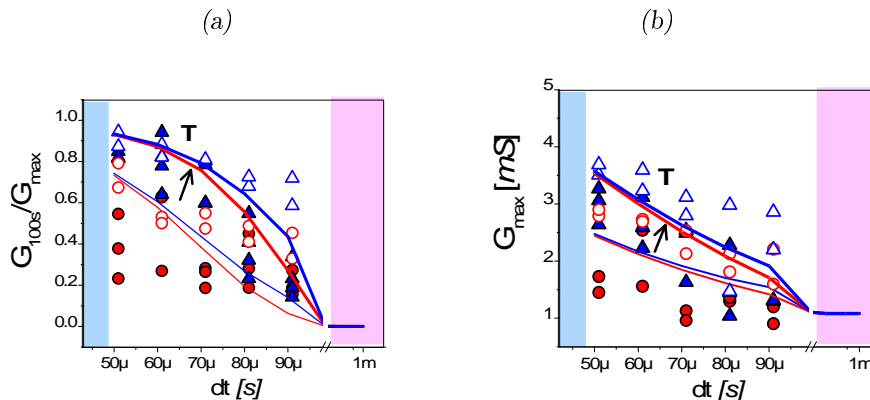


Figure 5.6: Synaptic Learning Implementation and Temperature effects in ECM cells. (a) G_{100s}/G_{max} conductance change as a function of the time correlation between pre- and post-spike dt . (b) Similarly, G_{max} conductance change as a function of the time correlation between pre- and post-spike dt .

In order to evaluate the temperature effects in ECM cells, we performed STDP measurements while the sample was heated at $420K$. Resulting STDP measurements are presented in Fig.5.6. A clear shift of both potentiation and LTP-induction was measured with respect to room temperature measurements. By controlling the physical parameters, as it will be explained in the next chapter, the '*Bio-inspired model 2.0*' is able to describe the STDP

measurements at 420K.

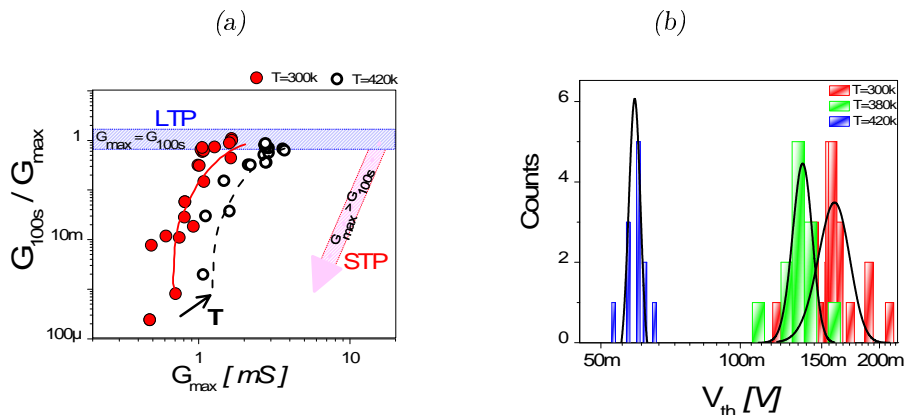


Figure 5.7: Synaptic Learning Implementation and Temperature effects in ECM cells.(a). STP to LTP transition and Temperature effects.(b). Switching threshold voltage range distributions as a function of the Temperature.

To validate the temperature effects on the switching dynamics of nanoscale memory under-test, we adopted the potentiation protocol corresponding to pre-pulses potentiation at fixed frequency similar to the previous results (La Barbera et al., 2015), described in the fourth chapter. We investigated the evolution of G_{100s}/G_{max} as a function of G_{max} for two different temperature (i.e. room temperature and 420K) as it is shown in Fig.5.7a. A clear shift toward higher G_{max} is obtained for potentiation at 420K. A second analysis was realized by considering the evolution of the switching threshold during conventional sweeping measurements. By increasing the temperature from 300K to 420K, a clear decrease of the threshold voltage is obtained (Fig.5.7b). Thus, for fixed pulse amplitude, increasing T corresponds to an increase of the amount of switching induces by a given pulse.

The temperature effects on the ECM cells can also be described in analogy to the neurocomputational triplet STDP model (Gjorgjieva et al., 2011). Triplet STDP rule not only reproduces STDP window but also the rate-based plasticity described by the BCM rule (this latter aspect was not described by pair-based STDP). As mentioned in the first chapter, instead of having only one process triggered by a pre- and post-synaptic spikes, it is possible to consider more quantities, which increase in the presence of a pre- and post-synaptic spike (i.e. r_1, r_2, o_1 and o_2) with their relative dynamics described by time constants such as $\tau_x, \tau_y > \tau_-, \tau_+$. The weight increases after post-synaptic spike o_1 arrival by an amount that is proportional to the value of the pre-synaptic variable r_1 but depends also on the value of the second post-synaptic detector o_2 . In such way two different state-variables can be distinguished depending on the time-scale. Equivalently, in our experiment the conductance G with time constant τ_{fac} can be used to describe

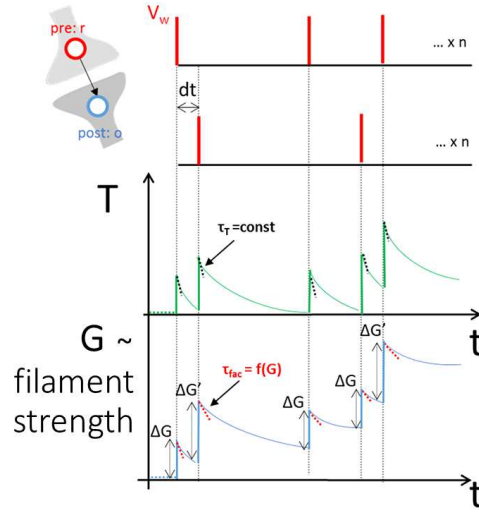


Figure 5.8: Synaptic Learning Implementation and Temperature effects in ECM cells and second order states variables.

pre- and post- pulses interaction while the temperature T can be associated to the second order state variable. We can associated the filament strength (i.e. the device conductance) to the first-order state variable. ΔG is affected by the temperature T (i.e. $\Delta G < \Delta G'$). The accumulation in T achieved with a time correlation between pre- and post- spikes with $\Delta T < 100\mu s$ well explains the synaptic learning implementation through the STP to LTP transition.

5.4 Towards Multiple Plasticity Mechanisms

First, we demonstrated in this chapter STDP induced LTP. Our devices presented STP characteristic that can be conveniently controlled by adjusting the mean firing rate in the network $\langle f \rangle$. By limiting $\langle f \rangle$, the device response can be hold in the short- term regime. The potentiation induced in the synaptic connection is then volatile and depends only on the pre-neuron firing rate. This form of non-Hebbian plasticity provides to the network a non-linear response of the synaptic connection as a function of the mean frequency $\langle f \rangle$. Such feature should be of particular interest for implementing asynchronous spiking networks since this form of synaptic adaptation will enhanced the potentiation when high frequency events are detected. Secondly, since learning in neural networks is mostly associated to Hebbian-type plasticity, we implemented a bio-realistic protocol in order to demonstrate Hebbian STDP corresponding to an increase of potentiation when correlated events (i.e. spiking) between pre- and post-neurons are detected. Not only potentiation was increased during STDP events but also

the LTP characteristic of the synaptic weight modification. In other words, Hebbian plasticity was conveniently associated to the STP to LTP transition. Such combination of both STP and STDP-induced LTP in the same memory device is highly valuable since it offers the possibility to deal with rate coding strategies as in the case of BCM concept and with temporal coding approaches where meaningful information are encoded through the precise timing of neurons (note that both rate coding and temporal coding has been evidenced to coexist in BNNs). The hypothesis that several synaptic functions manifest simultaneously and are interrelated at synaptic level seems accepted by different scientific communities. Recent biological studies indicate that multiple plasticity mechanisms contribute to cerebellum-dependent learning (Boyden et al., 2004). From a computational point of view, Zenke et al. (Zenke et al., 2015) have recently proposed the idea to used multiple plasticity mechanisms at different time scales. Multiple plasticity mechanisms may provide the flexibility required to store memories over different time-scales and to encode the dynamics involved. These plasticity mechanisms could act in combination with appropriate information-coding strategies for learning systems.

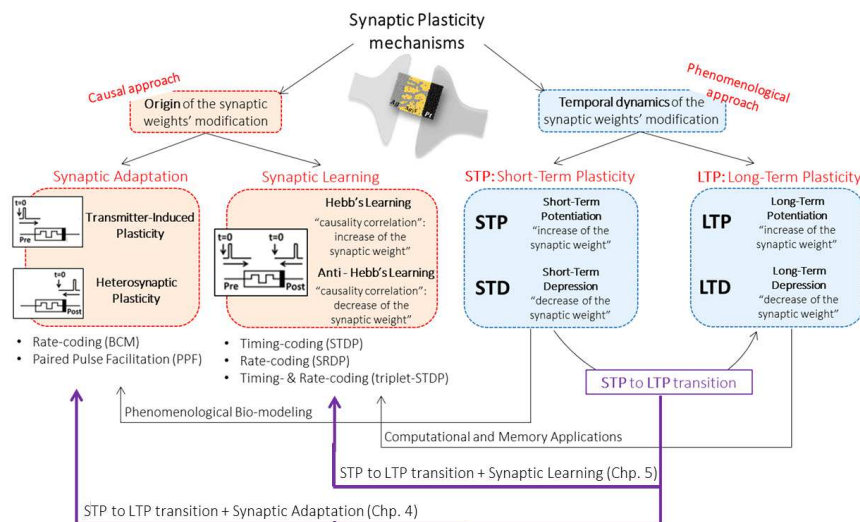


Figure 5.9: Towards Multiple Plasticity Mechanisms: schematic of the proposed Synaptic Plasticity mechanisms scenario. By following a casual approach we can distinguish Synaptic Adaptation from Synaptic Learning while by following a phenomenological one we can have short- or long term plasticity. The originality of this PhD work is linked to the STP to LTP transition, through which we demonstrated how it coexists with Synaptic Adaptation (in the previous chapter) or with Synaptic Learning (in this chapter).

To summarize the overall picture presented in this PhD manuscript concerning our approach to conceive the *Synaptic Plasticity* and its implementation in filamentary memristive devices, a schematic is presented in Fig.5.9. Depending if we consider the origin of the synaptic weight's modification, i.e.

the cause or the dynamics of the synaptic weight's modification, i.e. the effects, *Synaptic Plasticity* can be described by a causal or phenomenological approach, respectively. Along this line, as presented in the fourth chapter, we have successfully implemented in ECM cells transmitter-induced plasticity (Synaptic Adaptation) through the STP to LTP transition. For computational applications, LTP is generally associated to Synaptic Learning. The impact of the results presented in this chapter, relies another time in the STP to LTP Transition able to reproduce Synaptic Learning. Indeed, we have demonstrated that ECM cells can be engineered and programmed to reproduce different synaptic features in a dynamic volatility regime. Thus, the novelty of our approach focuses on the implementation of the STP to LTP Transition in a nanoscale component to reproduce multiple plasticity mechanisms at synaptic-level: promising key tool to exploit the BNNs efficiency for future neuromorphic NNET systems.

5.5 Discussion and Perspectives

In this chapter, we demonstrated that ECM cells can be engineered and programmed to reproduce STP-to-LTP transition and Synaptic Learning by taking into account the time correlation between pre- and post- spikes (dt) and the mean frequency of pre-neuron firing ($\langle f \rangle$) with a very simple protocol made of squared-shaped pulses without overlapping. By going deeper in the filamentary switching analysis, we studied short time scale interactions in our devices that seem reasonably linked to the temperature effects. Such dependence has been evaluated on the synaptic implementation behavior of our ECM cells and effectively the time correlation between pre- and post-spikes (dt) results in an accumulation in temperature which is responsible for a greater increase in conductance ($\Delta G' > \Delta G$).

Advanced bio-inspired features in which multiple plasticity mechanisms can be implemented and independently controlled in a single memory element could provide a general solution for the development of bio-inspired circuits. To improve the efficiency of future bio-inspired computing systems, interdisciplinary research is needed to obtain a better understanding of the contributions of *STP* and *LTP* mechanisms to memory construction and spike-coding information processing.

Chapter 6

Filamentary Switching Modeling and Circuit Simulations

"We are all agreed that your theory is crazy. The question which divides us is whether it is crazy enough to have a chance of being correct."
Niels Bohr (1885-1962)

6.1 Introduction

The main objective of this PhD work is to propose the *Synaptic Plasticity* (i.e. processes observed in biological synapses corresponding to a modification of the synaptic weight as a function of its spiking history) for information storage and computing in neuromorphic NNET. For such purpose, according to the 'nanotechnology approach', we have developed filamentary memristive devices. By exploiting the physical properties and the dynamic volatility regimes of such nanoscale device, we have demonstrated in the fourth chapter how it is possible to successfully reproduce and control fundamental processes observed in biological synapses. In the fifth chapter, we have demonstrated that ECM cells can be additionally programmed to reproduce advanced bio-inspired features in which multiple plasticity mechanisms can be implemented at the same time and independently controlled in a single memory element thus providing a general solution for the development of bio-inspired circuits.

This chapter describes the filamentary switching modeling and circuits simulations. Specifically, we will present a biological model of synaptic plasticity (Markram et al., 1998) that represents the starting point for analyzing the behavior of our synaptic devices. Then, we will present how such bio-inspired model can be used to describe our results for both synaptic adaptation and synaptic learning implementation. Once the consistency between this biolog-

ical model and the physical properties of our device will be validated, we will exploit it in two different forms: for memory and computing applications.

6.2 Synaptic Plasticity: Bio-inspired Model 1.0

The biological model of *Synaptic Plasticity*, developed by Markram et al. (Markram et al., 1998), is depicted in Fig.6.1a. This model describes

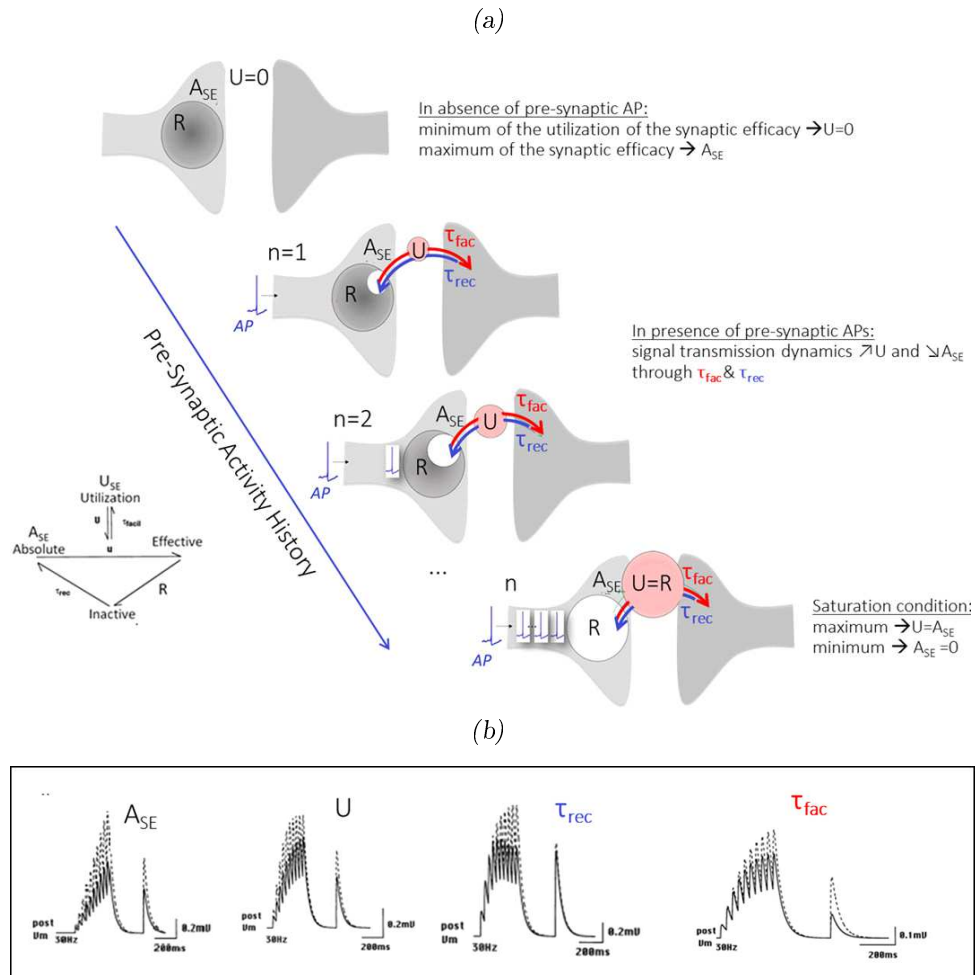


Figure 6.1: Bio-inspired Model 1.0: (a) Model schematic in which all the biological parameters: A_{SE} , U , τ_{rec} and τ_{fac} are shown as a function of the incoming APs. (b) Effect of each biological parameters involved in the synaptic transmission. Synaptic responses of facilitating synapses: when A_{SE} is increased 1.7-fold (simulation with $U = 0.01$, $\tau_{rec} = 60ms$ and $\tau_{fac} = 3000ms$ and initial $A = 2$); when U is increased from 0.05 to 0.1 (simulation with $A = 1$, $\tau_{rec} = 60ms$ and $\tau_{fac} = 3000ms$); when τ_{rec} is increased from 60ms to 600ms (simulation with $A = 1$, $\tau_{fac} = 3000ms$ and $U = 0.01$) and when τ_{fac} is increased from 1000ms to 3000ms (simulation with $A = 2$, $\tau_{rec} = 60ms$ and $U = 0.01$). Adapted from (Markram et al., 1998).

the Excitatory Post-Synaptic Potentiation (*EPSP*) response produced by a train of pre-synaptic Action Potentials (*APs*). After a number n of *APs*, the post-synaptic current response to the $n + 1^{th}$ *AP* is given by:

$$I_{n+1} = A_{SE} \cdot R_{n+1} \cdot U_{n+1} \quad (6.1)$$

where the absolute synaptic efficiency, A_{SE} , corresponds to the maximum possible synaptic efficiency; the fraction of available synaptic resources, R , corresponds to the neurotransmitter resources that are available in the pre-synaptic connection ($0 < R < 1$); and the utilization of the synaptic efficacy, U , corresponds to the amount of neurotransmitter that is released from the pre- to the post-synaptic connection ($0 < U < 1$). Thus, R_{n+1} and U_{n+1} are given by:

$$\begin{cases} R_{n+1} = R_n(1 - U_{n+1})e^{-\Delta t/\tau_{rec}} + (1 - e^{-\Delta t/\tau_{rec}}) \\ U_{n+1} = U_n e^{-\Delta t/\tau_{fac}} + U_{SE}(1 - U_n)e^{-\Delta t/\tau_{fac}} \end{cases} \quad (6.2)$$

The facilitating behavior observed during a burst of spikes is associated with the parameter U_{SE} , which is modified with the characteristic time τ_{fac} and applied to the first *AP* in a train (i.e., $R_1 = 1 - U_{SE}$). Recovery of the synaptic efficiency (or available neurotransmitters) is associated to the characteristic time τ_{rec} .

This biological model allows to reproduce different kind of synaptic plasticity mechanisms. Plasticity can be controlled through the neurotransmitter dynamics in the pre-synaptic connection (i.e., recovery of the available neurotransmitters or increase in the neurotransmitter release probability), by the improvement of neurotransmitter detection in the post-synaptic connection or even by a structural modification of the synaptic connection (i.e., increase in the size of a given synapse or the overall number of synapses connecting two neurons). To investigate the frequency-dependent signal transmission behavior of facilitating or depressing synapses Markram et al. (Markram et al., 1998) studied the effect of each biological parameters involved in the signal transmission (Fig. 6.1b). For a detailed review of synaptic plasticity, see (Zucker and Regehr, 2002; Collingridge et al., 2010). The synaptic efficiency of a given spike is determined by a combination of parameters that lead to different synaptic responses and expressions of *Synaptic Plasticity*.

6.3 Synaptic Adaptation Modeling

To illustrate the improved functionality obtained with our approach, we will present how the biological model of *Synaptic Plasticity* developed by Markram et al. (Markram et al., 1998) is able to fit our different synaptic potentiation experiments.

6.3.1 Bio-inspired Model 1.0

By considering the filaments stability of ECM cells and through a detail control of the device past history and electrical pulses stimulation, we succeeded in implementing Synaptic Adaptation in such cross-point devices. By accounting for the parameters of the bio-inspired model 1.0 by dividing with respect the applied bias (Eq.6.1), the conductance evolution can be described as follows:

$$G_n = A_{SE} \cdot U_n \cdot R_n \quad (6.3)$$

where, as previously described, A_{SE} is the absolute synaptic efficiency, U_{SE} is the utilization of the synaptic efficiency, τ_{fac} and τ_{rec} are the facilitating and recovery time constants, respectively. Four different cases as described in the fourth chapter (Fig.4.8b) and re-presented in Fig. 6.2, can be analyzed as a function of the number of pulses and I_c (table 6.1).

LTP		STP	
case 1: 150 pulses $I_c = 100\mu A$	case 2: 15 pulses $I_c = 800\mu A$	case 3: 10 pulses $I_c = 250\mu A$	case 4: 5 pulses $I_c = 800\mu A$
$U_{SE} = 0.0279$	$U_{SE} = 0.0279$	$U_{SE} = 0.0251$	$U_{SE} = 0.0279$
$A_{SE} = 6 \text{ mS}$	$A_{SE} = 25 \text{ mS}$	$A_{SE} = 6.5 \text{ mS}$	$A_{SE} = 16 \text{ mS}$
$\tau_{rec} = 0.0013 \text{ s}$	$\tau_{rec} = 0.0013 \text{ s}$	$\tau_{rec} = 0.0010 \text{ s}$	$\tau_{rec} = 0.0012 \text{ s}$
$\tau_{fac} = 11.5500 \text{ s}$	$\tau_{fac} = 18.5500 \text{ s}$	$\tau_{fac} = 0.0150 \text{ s}$	$\tau_{fac} = 1.5500 \text{ s}$

Table 6.1: Fitting parameters used for *Synaptic Plasticity* modeling

If we consider experiments 1 and 3, the same potentiation (i.e., $G_{max} = 0.9mS$) can lead to LTP (case 1 with 150 pulses and $I_c = 100\mu A$) or STP (case 3 with 10 pulses and $I_c = 250\mu A$).

The STP to LTP transition is mainly associated with an increase of the facilitating time constant, τ_{fac} . This increase is obtained by increasing the number of pulses during the excitatory burst. Slightly increasing I_c is mostly represented by an increase in A_{SE} . This observation is also evident by comparing case 2 with case 4. The difference in conductance level between cases 1 and 2, which showed qualitatively equivalent LTP responses, is mainly attributed to an increase of A_{SE} , from $6mS$ (case 1) to $25mS$ (case 2).

Synaptic Plasticity can be implemented by different burst configurations that modulates the potentiation obtained at the end of the pulse sequence (corresponding to the conductance at the end of a burst of pulses, G_{max}). We cannot establish a one-to-one correspondence between biological processes (e.g., neurotransmitter dynamics, structural modifications, etc.) and filament growth or relaxation in our experiments because most of the parameters are coupled in both cases. However, in the next section, by exploiting

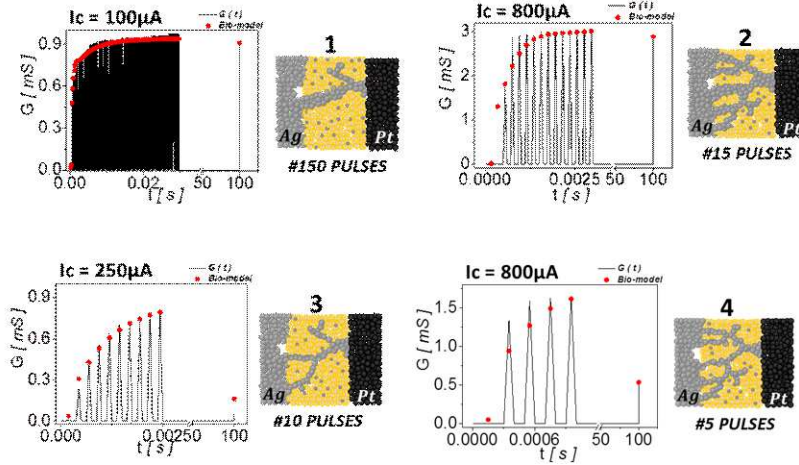


Figure 6.2: Two examples of LTP (cases 1 and 2) and STP (cases 3 and 4), for the case in which the *number of pulses* is set as the key plasticity factor and the I_c value is set as the dendritic path definition. The density (through I_c) and diameter (through burst excitation) of the dendritic branches can be tuned independently to reproduce various STP/LTP combinations.

the memristive synaptic bio-inspired model of this original behavior, we will show how this device can modulate its weight in a STP to LTP transition, and how this can be harnessed in a neuromorphic memory applications.

6.3.2 STP to LTP Transition for Memory Applications

By referring to the relation between G_{max} , the maximum conductance state induced after the potentiation, and the facilitation time constant τ_{fac} , as described in the fourth chapter and shown in Fig.6.3, we fitted the experimental results with an allometric function:

$$\tau_{fac} = a \cdot (G_n)^b \quad (6.4)$$

where fitting parameters a and b are function of the compliance current I_c as follows: $a = 6.25 \cdot 10^8$ and $b = 2$ for $I_c = 100\mu A$; $a = 3.40 \cdot 10^{12}$ and $b = 4$ for $I_c = 250\mu A$; $a = 2.35 \cdot 10^{21}$ and $b = 7.7$ for $I_c = 800\mu A$.

Practically, since STP to LTP transition depends on the conductance state reached after potentiation, such transition can be obtained by controlling the switching parameters such as pulse amplitude (i.e. higher voltage will lead to higher conductance), pulse train frequency (decrease the relaxation between pulses) or total number of pulses (accumulation of potentiation). We modified the bio-inspired model 1.0 to describe the STP to LTP transition by simply adding this relation into the model. The recovery time constant τ_{rec} , that was constant in all simulations to reproduce our measurements, as reported in the table 6.1, has been neglected. Thus, by referring to the

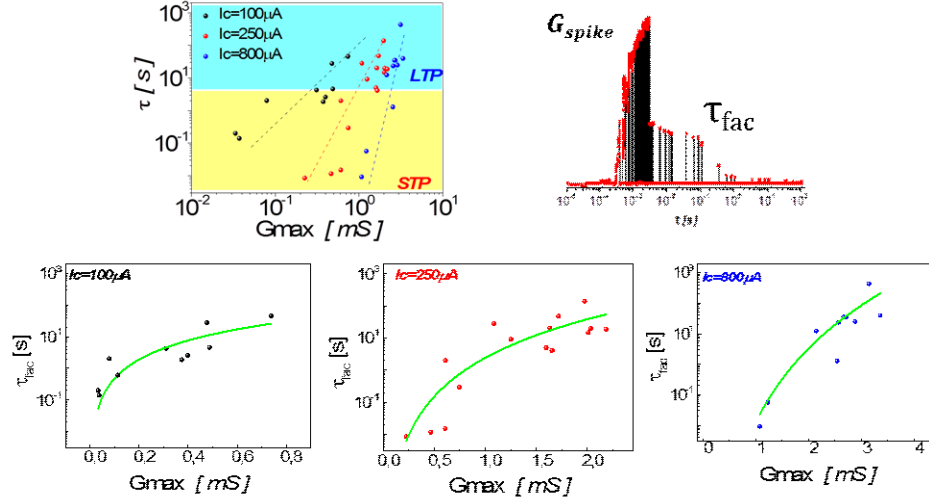


Figure 6.3: STP to LTP transition through the relation between G_{max} (i.e. conductance state reached after potentiation) and the facilitation time constant τ_{fac} depending on the compliance current I_c .

Equ.6.3, the conductance evolution can be written as follows:

$$G_n = A_{SE} \cdot [U_{n-1}e^{-dt/\tau_{fac}} + U_{SE}(1 - U_{n-1})e^{-dt/\tau_{fac}}] \quad (6.5)$$

where, also in this case, the absolute synaptic efficiency is fixed by the parameter A_{SE} and the utilization of the synaptic efficacy is termed U_{SE} . Thus, with simple substitutions the post-synaptic current response (i.e. conductance G_n) after n APs, that is related to the previous $(n-1)$ AP can be written as:

$$G_n = G_{n-1} \cdot e^{-dt/\tau_{fac}} + U_{SE} \cdot (A_{SE} - G_{n-1} \cdot e^{-dt/\tau_{fac}}) \quad (6.6)$$

G_n , as previously described is associated to the facilitating behavior observed during a burst of spikes, which is modified with a characteristic time τ_{fac} that leads to different synaptic responses and different *Synaptic Plasticity* expression. We will refer to such modified model as bio-inspired model 1.1.

We demonstrated how this model can be useful for future neuromorphic memory applications by implementing it in a spike-based system. In collaboration with Adrien Vincent, Christopher Bennett and Dr. Damien Querlioz from the Institute of Fundamental Electronics (IEF) of Paris Sud, we performed a filamentary-type memristive devices cross-bar system. In a first work we simulated an architecture in which a binary 'target' pattern ('1' if there is a spike, '0' otherwise) is fed into a (6×6) crossbar of memristive nanodevices corresponding to exactly one column for each class (Fig. 6.4).

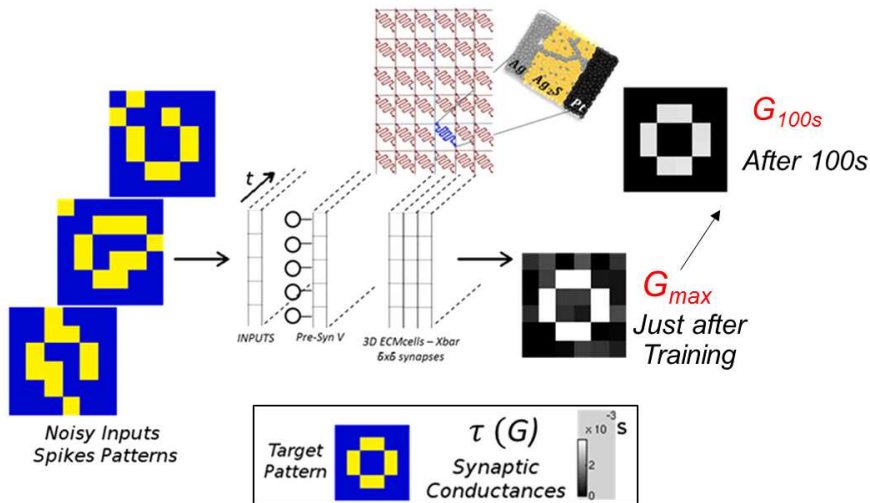


Figure 6.4: ECM cells Cross-bar implementation for memory applications: a crossbar system (6×6) based on pre-synaptic activities is able to store clean patterns despite the inputs presented in the programming/learning phase are noisy by exploiting the *STP* to *LTP* transition.

Based on pre-synaptic activities, the bio-inspired model 1.1 is able to evaluate the synaptic conductance time-evolution at each ECM cell cross-points of the system.

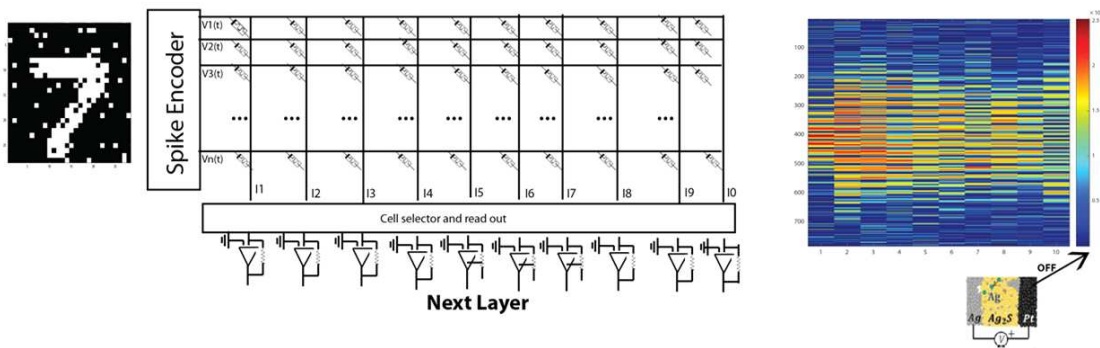


Figure 6.5: Synaptic Adaptation Implementation in ECM cells cross-bar. On the left: schematic of Memristive Nanodevice cross-bar architecture in which an instance of the MNIST database, once encoded into a binary stream of spikes, is fed into a crossbar of memristor ECM nanodevices corresponding to exactly one column for each class. As shown the row inputs are voltage spikes in time and the column outputs are currents. On the right: corresponding conductance evolution map for this scheme once the spikes are applied over an entire conditioning period. Blue values represent low conductance and red values represent high, as numerically illustrated in the scale bar.

By exploiting the non-linear transformations of the input data (separation) due to the intrinsic relaxation time constant τ_{fac} , the current state of the

network is only affected by the previous states up to a certain time. Therefore, based on pre-synaptic activities (i.e. performing Synaptic Adaptation), the spike-based system is able to store clean patterns, despite the inputs presented in the programming/training phase are noisy. Just after the training phase (i.e. corresponding to G_{max}), the system recognizes the target pattern and the feature extracted persists after 100s if the synapses was programmed in a LTP (non-volatile) regime.

Similarly, in a second work, we performed a larger cross-bar system (28×28) as shown in Fig.6.5, in which the non-linear state change in the synaptic connection (i.e. the time evolution of the conductance associated at each pixel of the input pattern) with the state relaxation described by the dynamic time constant $\tau_{fac} = f(G)$ regulating the STP to LTP transition, have been used to reproduce successfully our results relative to Synaptic Adaptation in ECM cells. In particular, we have simulated an architecture in which an in-

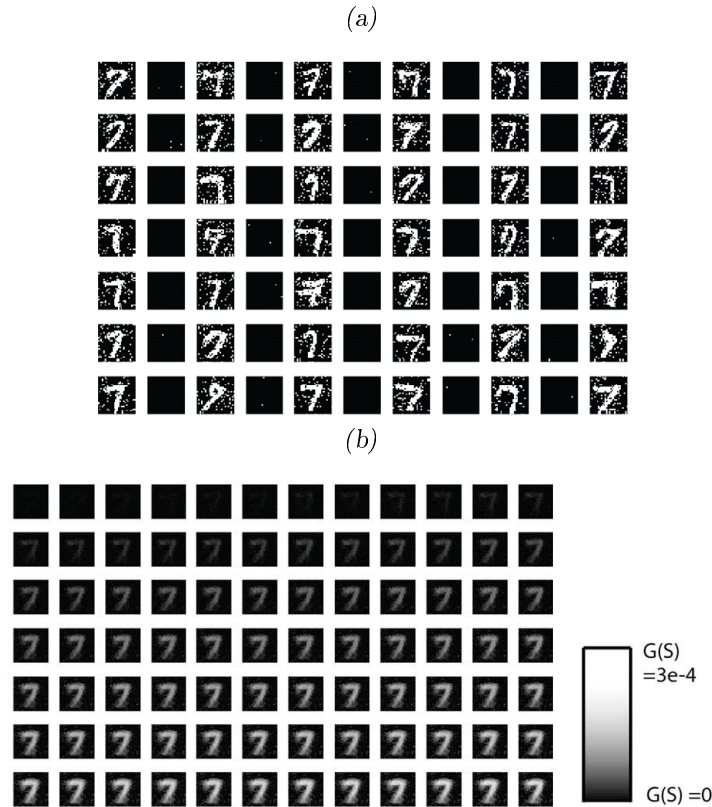


Figure 6.6: Binary Spike Encoding Scheme and corresponding Conductance Map. (a) Images of the MNIST database are fed to the network via binary spike encoding, where white represents a '1' and '0' is none (background). Note the significant noise visible in the surroundings of the image. (b) The same effect of (a) spike train in a progressive conditioning of one column of memristive devices to its respective class, in this case, a '7'. As visible, the noise is eliminated and image averaged.

stance of the MNIST database, once encoded into a binary stream of spikes, is fed into a crossbar of memristive nanodevices corresponding to exactly one column for each class. As shown in Fig.6.5 the row inputs are voltage spikes in time and the column outputs are currents. For our purposes we focus on the current read outs, yet once fed through a simple CMOS neuron that may be converted to voltage for the next row. The non-linear state change in the synaptic connection (i.e. the time evolution of the conductance associated at each pixel of the input pattern) with the state relaxation, described by the dynamic time constant τ_{fac} regulating the *STP* to *LTP* transition, has been used to recognize noisy images of the MNIST database as shown in the conductance map in Fig.6.6b. It is interesting to emphasize that such memory application works as a pre-filter creating projections of the data that could help for classification task on multi-layer perceptron (MLP), for example and could be exploited to enrich future neuromorphic computing.

6.4 Synaptic Learning Modeling

In the fifth chapter, we have demonstrated how, by considering short-time scale interactions and a second-order memristor model, another form of *Synaptic Plasticity* can be implemented in ECM cells: the Synaptic Learning. In this section we will present how the bio-inspired model 2.0, that provides a good qualitative description of our results, can be exploited for computational applications.

6.4.1 Bio-inspired Model 2.0

By taking into account the time correlation between pre- and post- spikes (dt) and the mean frequency of pre-neuron firing ($\langle f \rangle$) with a very simple protocol made of squared-shaped pulses, we have programmed ECM cells to reproduce STP to LTP transition and Synaptic Learning.

As described in the fifth chapter, the STDP results at $2kHz$ and $5kHz$ obtained from Ag_2S -based ECM cells are re-presented in Fig.6.7. A clear increase of potentiation from $1mS$ to $3.5mS$ and LTP induction (from 0 to 1) is measured for time correlation such as $dt < 100\mu s$. In addition, when dt is decreased, this effect is strengthened, thus reproducing gradual STDP windows observed in biology. Time correlation dt smaller than $50\mu s$ resulted in pre- and post-pulse overlapping (pulse width was $50\mu s$). Since large voltages are obtained in this case (i.e. $2 \cdot V_{write}$), fully potentiated weights ($G_{max} = 3.5mS$, squared points (blue region) in Fig.6.7c and Fig.6.7d in the LTP regime ($G_{100s}/G_{max} = 1$)) were measured. Control experiments (green points in the pink regions) with pre-neuron spikes only were performed and showed weak potentiation ($G_{max} = 1mS$) and no LTP ($G_{100s}/G_{max} \ll 1$). As mentioned in the previous chapter, the impossibility to the Markram bio-inspired model 1.1, as shown in Fig.6.7 in dashed line, to describe our results

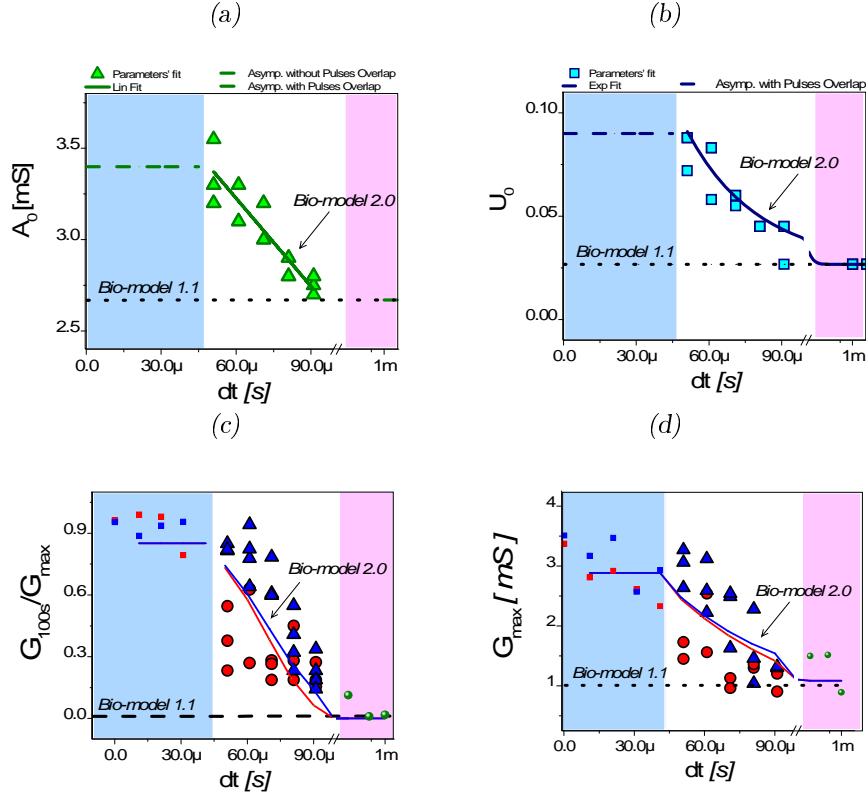


Figure 6.7: Synaptic Learning Implementation in ECM cells. (a) Evolution of the absolute synaptic efficiency, A_0 , and (b) the utilization of the synaptic efficiency U_0 for fitting the STDP measurements. (c) G_{100s}/G_{max} conductance change as a function of the time correlation between pre- and post-spike dt . (d) Similarly, G_{max} conductance change as a function of the time correlation between pre- and post-spike dt . Red dots correspond to measurements at $2kHz$ while blue triangles at $5kHz$.

can be explained by considering the origin of STDP in ECM devices. Since correlated pulses can result in accumulation of potentiation, we calculated the expected G_{max} and G_{100s}/G_{max} values for similar spike protocol with the bio-inspired model 1.1. Since pre- and post-pulses are similar, a first conclusion is that short-time scale interactions between two successive pulses are not captured by the bio-inspired model 1.1. In order to describe such mechanisms, the first substantial modification of the model relies on the absolute synaptic efficiency, A_{SE} , and the utilization of the synaptic efficiency U_{SE} considered as free parameters (i.e. A_0 and U_0) for fitting the STDP measurements (Fig. 6.7a and Fig. 6.7b). For large dt (i.e. $dt > 90\mu$ s), $U_0 = U_{SE} = 0.0267$ and $A_0 = A_{SE} = 2.7mS$ presented similar values as the one extracted from control experiment and pre-neuron only excitations (pink region in Fig.6.7) and corresponding to the bio-inspired model 1.1. For 50μ s $< dt < 90\mu$ s a good fitting of STDP experiments is obtained when

both U_0 and A_0 are increased when dt is decreased. We described A_0 and U_0 evolution with linear fitting and with exponential decay in the short dt regime, respectively as follows:

$$A_0 = a + m \cdot dt \quad (6.7)$$

$$U_0 = u_0 + u_1 \cdot e^{-dt/\tau_T} \quad (6.8)$$

where fitting parameters are: $a = 0.00432$, $m = -18$, $u_0 = 0.00267$, $u_1 = 0.2717$ and $\tau_T = 3.41 \cdot 10^{-5}$. The conductance evolution can be described by the bio-inspired model 2.0, as follows:

$$G_n = G_{n-1} \cdot e^{-dt/\tau_{fac}} + U_0 \cdot (A_0 - G_{n-1} \cdot e^{-dt/\tau_{fac}}) \quad (6.9)$$

providing a good qualitative evolution of LTP-induction and STDP measurements, as shown by the red (at $2kHz$) and blue (at $5kHz$) lines in Fig.6.7. A possible explanation of short-time scale interactions in our devices, as described in the previous chapter, could be attributed to heating effects and subsequent heat dissipation after switching. A second pulse following a prior impulse can benefit from local heating in the switching region of the filament and increases the effect of this second excitation on potentiation.

In order to check for temperature effects in ECM cells, we performed STDP measurements while the sample was heated at $420K$. Resulting STDP measurements are presented in Fig.6.8. A clear shift of both potentiation and LTP-induction was measured with respect to room temperature measurements. Fitting of the STDP measurement at $420K$ with bio-inspired model 2.0 was possible by increasing the A_0 ($a = 0.00588$ and $m = -35$ in Fig. 6.8a) and U_0 ($u_0 = 0.0027$ and $u_1 = 0.45$ and $\tau_T = 3.73 \cdot 10^{-5}$ Fig.6.8b) dependency with dt . Thus, for fixed pulse amplitude, the increase in T corresponds to an increase of the amount of switching induces by a given pulse, i.e. U_0 , consistent with the reported evolution of U_0 with dt during STDP measurements at $420K$. If this experiment is not sufficient to attribute short-time scale interaction between pulses to heating effects only, it is a strong indication in favor of this possibility.

We can remark that the dependence of STDP measurements with mean frequency $\langle f \rangle$ is only slightly captured by model 2.0, considering only $2kHz$ and $5kHz$ that have been tested experimentally and fitted with the same A_0 and U_0 function. Refinement in the fitting (which required more intensive measurements in order to average variability observed in STDP measurements) with detail analysis of A_0 and U_0 evolution as a function of $\langle f \rangle$ is a possible direction to improve the rate-dependent effect in STDP (i.e. effect of $\langle f \rangle$ on the STP to LTP transition and potentiation). As describe in the fourth chapter, conductance relaxation in response to 20 writing pulses lead to STP regime while a LTP regime can be induced with 150 pulses. Thus, in order to evaluate how the frequency affects the bio-inspired model 2.0,

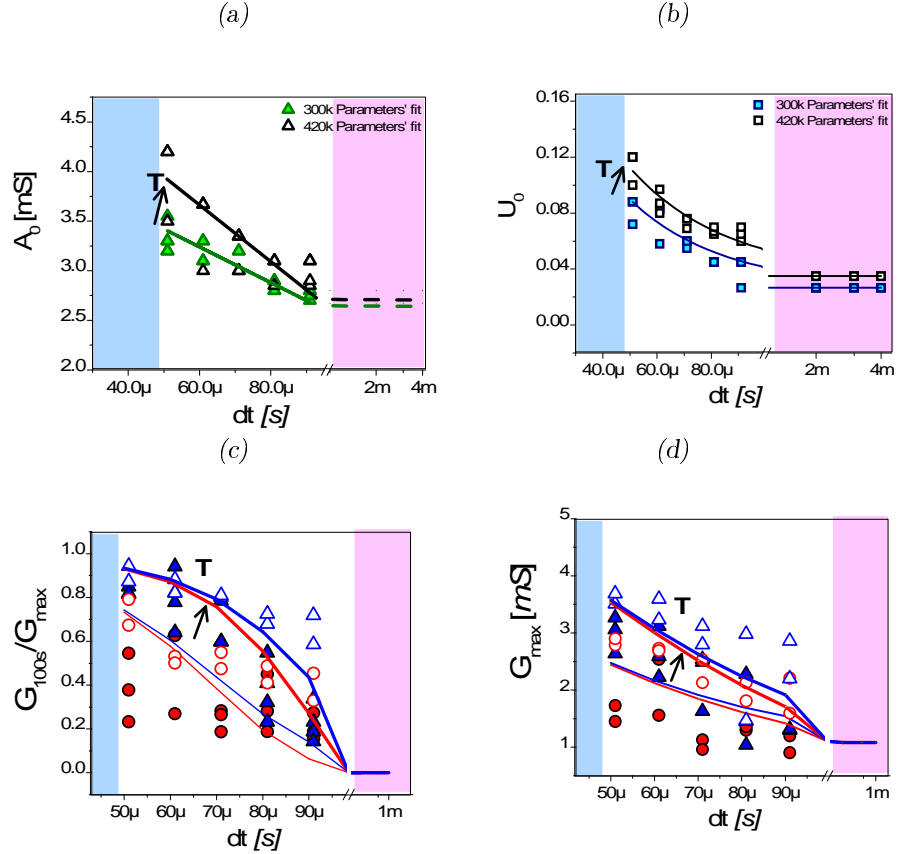


Figure 6.8: Temperature effects on Synaptic Learning Implementation in ECM cells. (a) Evolution of the absolute synaptic efficiency, A_0 at 300K green triangles and linear fit and at 420k black triangles and linear fit. (b) Evolution of the utilization of the synaptic efficiency U_0 for fitting the STDP measurements at 300K blue squares and exponential fit and at 420k black squared and exponential fit. (c) G_{100s}/G_{max} conductance change as a function of the time correlation between pre- and post-spike dt at 300K red filled dots (at 2kHz) and blue filled triangles (at 5kHz) and at 420K red empty dots (at 2kHz) and blue empty triangles (at 5kHz). (d) Similarly, G_{max} conductance change as a function of the time correlation between pre- and post-spike dt at 300K and at 420K.

we have simulated the evolution of G_{max} and G_{100s}/G_{max} as function of the frequency (with $U_0 = 0.0267$ and $A_0 = 2.7mS$) in these two cases (Fig.6.9a). As expected, the STP to LTP transition induced with an excitation sequence of 150 pulses is more abrupt with respect the one with 15 pulses.

6.4.2 STP to LTP Transition for Information Computing

By referring to the bio-inspired model 2.0 (Equ.6.9), let's consider dt the time difference between pre- and post-spike and G_{\setminus} the relaxation of the

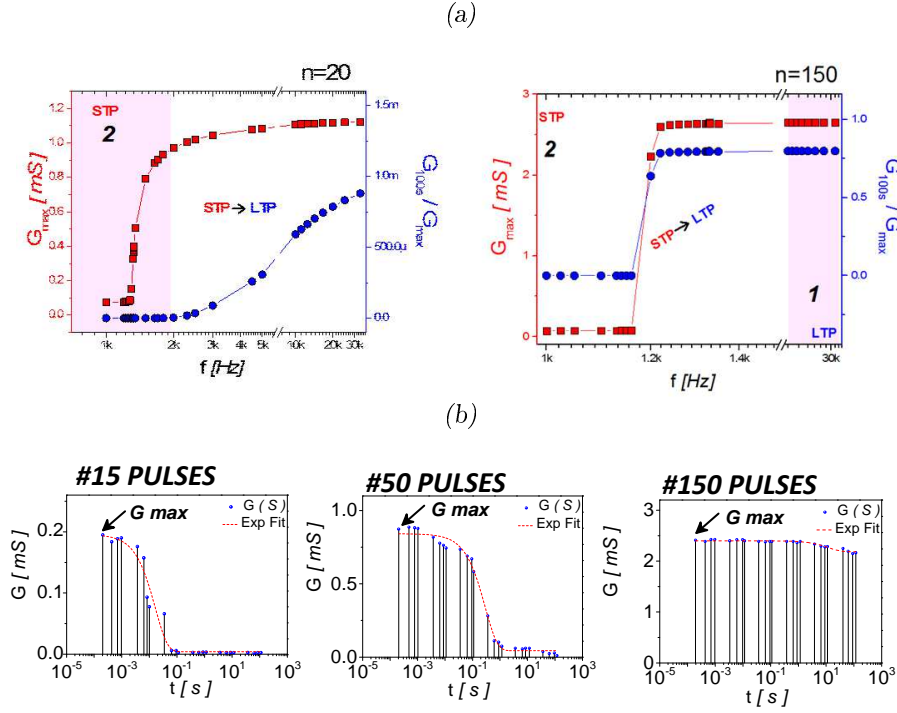


Figure 6.9: Frequency and number of pulses effects on bio-inspired model 2.0. (a) Model simulation of G_{max} and G_{100s}/G_{max} as function of the frequency for 20 pulses (on the left) and 150 pulses (on the right). (b) Measurements of conductance relaxation (blue points) and fitting (red line) on six time decades for different potentiation (G_{max}) values, obtained by varying the number of pulses (15, 50, and 150 pulses), results reported in the fourth chapter.

conductance after a spike (i.e. a pulse):

$$\forall dt \geq 0, \quad G_{\searrow} = (G(t) - G_{min}) \cdot e^{-\frac{dt}{\tau_{fac}}} + G_{min} \quad (6.10)$$

where the facilitation time constant τ_{fac} is related to the increase in conductance induced by a spike and it depends on the compliance current I_c (i.e. according to the allometric Equ.6.4). G_{min} is the minimum conductance value (i.e. asymptotic value in the pink region of Fig. 6.7). The conductance evolution in function of time can be re-written as follows:

$$G(t + dt) = \begin{cases} G_{\searrow} & \text{if no spike} \\ G_{\searrow} + U(A - G_{\searrow}) & \text{if spike} \end{cases} \quad (6.11)$$

where, in absence of a spike the conductance dynamics can be described by the facilitation time constant while after a spike the dynamics follows the learning rule that depends on the time correlation between pre- and post-spikes. At the boundary conditions we can write: $G(t + dt) \rightarrow G(t)$ for

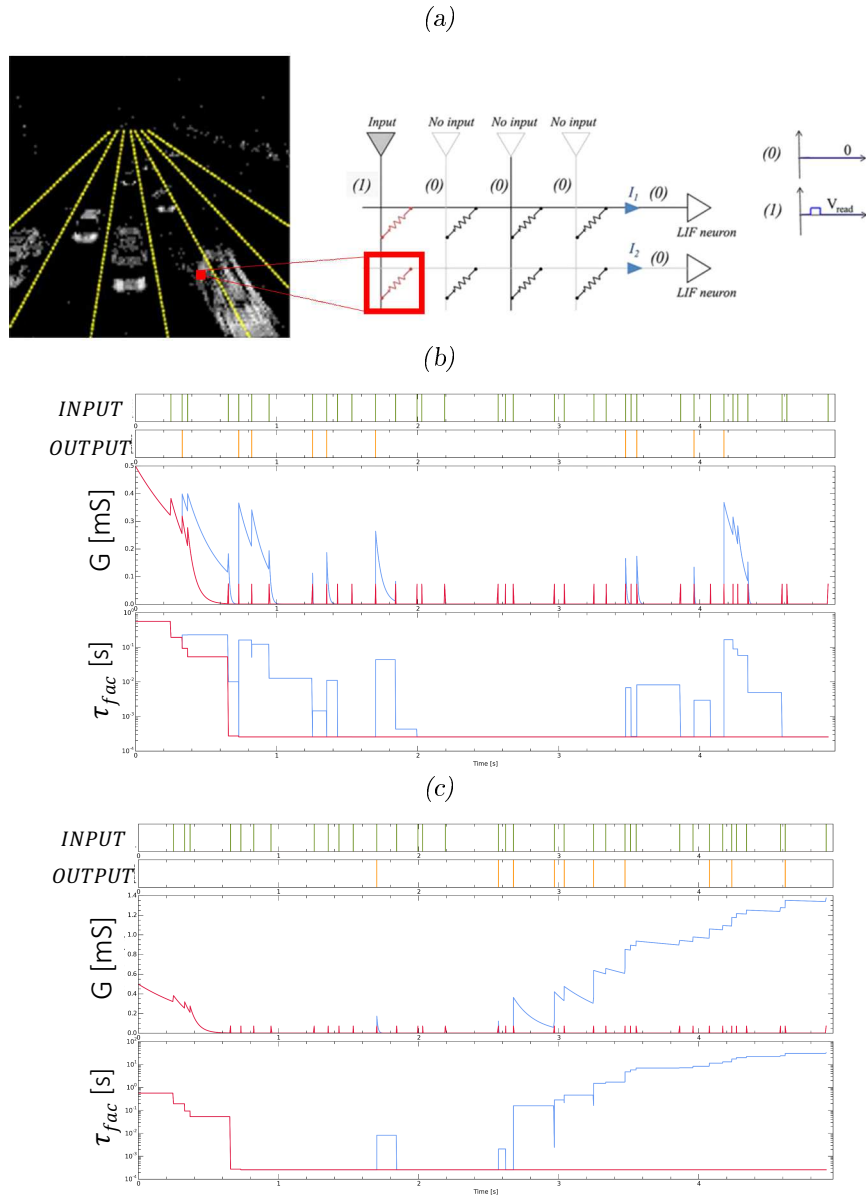


Figure 6.10: Simulation of the pulse train of a pixel of DVS cars video through Synaptic Learning in crossbar system of ECM cells. (a) DVS image in which the pixel considered is marked with red dot and on the right schematics of the crossbar architecture (1R case) for the learning system. (b) Conductance G and time constant τ_{fac} evolution of a pixel. Red curves for the case where only the input pulses (green events) are applied to the synapse; blue curves for the case where the input pulses (green events) and output pulses (orange events) are applied to the synapse. This configuration is not able to induce synaptic potentiation. (c) Similarly study but in a pulses configuration that is able to induce synaptic potentiation.

$dt \rightarrow 0$ and $G(t + dt) \rightarrow G_{min}$ for $dt \rightarrow \infty$.

In collaboration with Adrien Vincent and Dr. Damien Querlioz from the Institute of Fundamental Electronics (IEF) of Paris Sud, we used this bio-inspired model of the synaptic learning in ECM cells in a spike-based system. A spiking neural networks simulators time-step based allows to perform large-scale systems in which CMOS circuits of the neurons are functionally simulated, connected to memristive cross-bar in which at each interconnection a physical synaptic model can be used (i.e. allowing to take also into account devices' imperfections and variability). Specifically, in such system the inputs are the pixels of a bio-inspired dynamic vision sensor (DVS), which naturally produces asynchronous spikes, analogous to our retina and the outputs are LIF neurons. Each input is connected to each output by a synaptic device (in our case ECM cell described by the bio-inspired model 2.0) in a 'all-to-all' configuration. Beyond image classification, the time-dependent nature of the plasticity form considered makes it particularly appropriate to learn features on dynamic data. In particular, we considered as inputs the pulse train of a pixel (marked by a red dot in Fig.6.10a) of file cars video that are applied to the synaptic ECM cell. The video time scale was accelerated by a factor 15. In order to mimic an output neuron firing attached to the synapse considered, 10 artificially correlated pulses are added randomly to the inputs. The initial conductance is not presumed at the minimum value. In Fig.6.10b it is possible to observe the conductance and time constant evolution in which the pre- and post- spikes pairs are in a configuration that is not able to induce a potentiation of the synapse while in Fig.6.10c with the same number of pre- and post- spikes pairs in a different configuration results in the potentiation of the synaptic device. It is interesting to note how the effect of an incoming pulse is to increase the conductance G that subsequently evolves with the time constant τ_{fac} . When a second pulse (a post-spike) arrives immediately after a pre- one (in $dt < 100\mu s$ in our case), it induces a stronger increase in the conductance G , as it is visible from $t = 2.6s$ in Fig.6.10c suggesting that the synaptic potentiation observed in this scenario is the result of the dynamics at short time scales (i.e. process that could be explained as a temperature effects). These preliminary results pave the way for large scale circuit simulations exploiting ECM cell bio-inspired model.

On going works aim at investigating how to simulate a spiking based system in which, thanks to this dynamic bio-inspired model, the synapses who see only the input pulses remain generally depressed, while those that would see some pre- and post- pairs in short dt (i.e. temporally correlated) are potentiated. These results demonstrate how such complex behaviors of memristive physics can be exploited for computing applications.

6.5 Discussion and Perspectives

In this chapter we demonstrated how a rich panel of functionalities can be embedded in a single filamentary memristive element described by a biological model of *Synaptic Plasticity* that has been used for circuit simulations. By exploiting plasticity form of Synaptic Adaptation based on pre-synaptic activities, we have performed a spike-based system (a (6×6) and (28×28) ECM cells cross-bar) able to detect clean patterns, although the inputs presented in the programming/training phase were noisy.

By considering short-time scale interactions another form of *Synaptic Plasticity* has been implemented in ECM cells: the Synaptic Learning. Thanks to the intrinsic time-dependent nature of this plasticity form, we have used the bio-inspired model to learn features on dynamic data.

These results pave the way for future engineering of neuromorphic computing systems, where complex behaviors of memristive physics can be exploited.

Chapter 7

Conclusions and Perspectives

7.1 Dissertation Research Work Conclusion

During this PhD research activity, we developed specific nanoscale technologies to replicate some of the key mechanisms observed in biological systems, such as the *Synaptic Plasticity*, with a clear objective: bringing more functionality in a single component in order to reduce circuit overhead cost and improve circuit performances for future engineering of neuromorphic systems.

We proposed an overall classification of different forms of *Synaptic Plasticity*, i.e. processes observed in biological synapses corresponding to a modification of the synaptic weight as a function of its spiking history, that can be implemented in emerging memristive devices. In particular, depending if we consider the origin of the synaptic weight's modification or their dynamics, *Synaptic Plasticity* can be described by a causal or phenomenological approach, respectively.

By exploiting the physical properties and the dynamic volatility regimes of filamentary memristive devices, we successfully implemented the transmitter-induced plasticity that corresponds to the *Synaptic Adaptation* (causal description), a non-Hebbian plasticity form that depends only on pre-neuron activity. We demonstrated that complex filament shape, such as dendritic paths of variable density and width, can permit the short- and long-term processes (phenomenological description) to be tuned independently, by achieving a flexible way to program the device memory (i.e. the synaptic weights) and the relative device volatility. In particular for plasticity tuning we used the compliance current I_c during pre-conditioning that regulates the average conductance of the filament during switching in pulse mode. We used also the excitation strength (number of pulses or pulse amplitude) that handles the *STP* to *LTP* transition which can be associated to an increase of the branch diameter. These two parameters can be tuned independently of each other to modify the device conductance and filament volatility.

In BNNs a combination between long term synaptic processes (Long Term Potentiation LTP and Depression LTD) and short term mechanisms (Short Term Plasticity, STP) contributes to the processing and storage of information. Individually such forms of synaptic plasticity such as Short Term Plasticity, Short Term to Long Term Plasticity transition or STDP have already been successfully implemented in this class of filamentary switching devices. The novelty of our work consists in demonstrating that ECM cells can be programmed to reproduce advanced bio-inspired features in which all these synaptic features can be realized and independently controlled in a single memory element thus providing a more general solution for the development of bio-inspired circuits. Specifically, *Synaptic Learning* (causal description) has been implemented in filamentary switching by considering the Hebbian STDP rule, that corresponds to an increase of the synaptic weight when time correlation between pre- and post-neuron firing is experienced at the synaptic connection. Not only the instantaneous potentiation should increase during STDP events, but also the LTP characteristic (phenomenological description) of this synaptic weight modification. The first aspect (i.e. potentiation) has been well described by studying G_{max} evolution during STDP events while the second one (i.e. LTP induction) has been measured by recording the G_{100s}/G_{max} value after an STDP experiment. To illustrate the improved functionality obtained with our approach, we have taken into account the biological model of *Synaptic Plasticity* to fit our different synaptic potentiation experiments. By considering the analogy between the biological and the device parameters, the model provided a good description of the synaptic functionality implemented in our nanoscale memory device and it has been used for circuit simulations. We performed a spike-based system (6x6 and 28x28 ECM cells cross-bars) that, by adopting the *Synaptic Adaptation* based on pre-synaptic activities, are able to detect clean patterns, although the inputs presented in the programming/training phase were noisy.

By considering short-time scale interactions in ECM cells, the *Synaptic Learning* has been implemented in a spike-based systems. Thanks to the intrinsic time-dependent nature of this plasticity form described by the bio-inspired model the simulation results demonstrate how to learn features on dynamic data.

These results pave the way for future engineering of neuromorphic computing systems, where complex behaviors of memristive physics can be exploited.

7.2 On-Going and Next Steps

The main challenge addressed by future neuromorphic engineering relies on the realization of bio-mimetic hardware system, i.e. ANNs whose organizing principles are based on those of BNNs. In order to achieve such an

ambitious goal, research efforts have to be done from:

1. "*Single-device level*": a deep investigation of the intrinsic biomimetic aspect of memristive nano-devices related to the basic physics of resistive switching.
2. "*Circuit level*": engineering and integration solutions to implement such nano-devices in massively parallel and ultra-dense architectures, i.e. neuromorphic NNET.
3. "*System level*": looking for learning strategies, to model different circuit topologies and level of processing devoted to improve and enrich features extraction and mining of future computing systems.

By regarding the current technological status, memristive cross-bar architecture that is CMOS-compatible, seems the most promising and robust approach to an hardware implementation of ANNs. Along this integration strategy, the CMOL concept is of particular interest where neuronal functions can be assigned to the CMOS platform and synaptic connections to the crossbar of nanodevices.

A bottom-up approach of random cross-bar architectures, reminiscent of random organization in BNNs, can be also envisioned for future realization of neuromorphic NNET. Even if several engineering challenging issues have to be addressed, promising and interesting characteristics could be exploited by reservoir computing to implement complex neuromorphic functionalities.

In this context, the research activity presented in this PhD manuscript is mainly centered at the device level, by proposing the *Synaptic Plasticity* as key element for future Information Computing. On going projects aim at investigating how such synaptic properties can be advantageous for large-scale neuromorphic circuits and preliminary results are promising indicators for future research directions.

The 'exotic' characteristics of the filamentary memories technology realized in this PhD work (i.e. poor retention of state (ms to s), ON/OFF ratio (10^3), analog programmable and high device variability), suggest the idea that a bottom-up approach could be the most promising integration strategy. In such direction the main challenge would be how to control such random networks of devices even if it could offer a material solution for Reservoir Computing implementation. Furthermore, from a physical point of view, it would be very interesting to investigate deeper the filamentary switching, by comparing the amorphous sulfides Ag_2S with other insulator layers (such as amorphous $GeSe_{2+x}$, ordered or disordered oxides $a-Si$, SiO_2 , TaO_5) and by coupling fractal geometry with a percolation network model such as the circuit breaker configurations.

In this emerging research direction behind the Neuro-inspired Computing, the multidisciplinary interactions, from biology, computational neuroscience,

mathematics, computer architecture and computer systems, microelectronics, nanotechnology and physics, are of paramount importance for a future development of neuromorphic hardware systems. Future works should investigate the strategy in order to emphasize such aspect in which different point of views, competences, efforts could converge towards a common objective: by improving the understanding of the mechanisms regulating the human brain, to create chips based on natural computation.

Appendix A

List of publications

A.1 Book Chapter

1. *Synaptic Plasticity with Memristive Nano-Devices*
Selina La Barbera and Fabien Alibart
to appear in *Advances in Neuromorphic Hardware using Emerging Nanodevices*, M. Suri, ed., Springer, 2015.

A.2 Journal and Conference Papers

1. *Filamentary Switching: Synaptic Plasticity through Device Volatility*
Selina La Barbera, Dominique Vuillaume and Fabien Alibart
ACS Nano, 2015, 9 (1), pp 941-949.
2. *Multiple Time-scale Plasticity Mechanisms through Short-Term to Long-Term Plasticity Transition in Filamentary Switching*
Selina La Barbera, Adrien F. Vincent, Dominique Vuillaume, Damien Querlioz and Fabien Alibart
Submitted, 2015.
3. *Plasticity in memristive devices for Spiking Neural Networks*
Sylvain Saighi, Christian G. Mayr, Teresa Serrano-Gotarredona, Heidemarie Schmidt, Gwendal Lecerf, Jean Tomas, Julie Grollier, Soren Boyn, Adrien Vincent, Damien Querlioz, Selina La Barbera, Fabien Alibart, Dominique Vuillaume, Olivier Bichler, Christian Gamrat and Bernabe Linares-Barranco
Front. Neurosci, 2015, 9(51), pp 1-33
4. *Short-Term to Long-Term Plasticity Transition in Filamentary Switching for Memory Applications*
Selina La Barbera, Adrien F. Vincent, Dominique Vuillaume, Damien Querlioz and Fabien Alibart

- International Conference on Memristive Systems (MEMRISYS) 2015, Cyprus
5. *Short-Term to Long-Term Plasticity Transition by Filamentary Switching Volatility Control*
Selina La Barbera, Dominique Vuillaume and Fabien Alibart
W03 2nd International Workshop on Neuromorphic and Brain-Based Computing Systems (NeuComp 2015), Grenoble, France
 6. *Atomic switch: synaptic functionalities and integration strategies*
Selina La Barbera, David Guerin, Dominique Vuillaume and Fabien Alibart
17emes Journees Nationales du Reseau Doctoral en Micro-Nanoelectronique, JNRDM 2014, Villeneuve d'Ascq, France
 7. *Short-Term to Long-Term Plasticity Transition by Filamentary Switching Volatility Control*
Selina La Barbera, Dominique Vuillaume and Fabien Alibart
(Poster) W03 2nd International Workshop on Neuromorphic and Brain-Based Computing Systems (NeuComp 2015), Grenoble, France
 8. *Filamentary Switching: Synaptic Plasticity through Device Volatility*
Selina La Barbera, Dominique Vuillaume and Fabien Alibart
(Poster) IEEE Swiss CAS/ED Workshop 2014 on Memristive Devices and Neuromorphic Applications, Zurich, Switzerland
 9. *Atomic switch: fonctionnalites synaptiques et strategies d'integration*
Selina La Barbera, David Guerin, Dominique Vuillaume and Fabien Alibart
(Poster) Journees NeuroSTIC 2014, Cergy-Pontoise, France
 10. *Atomic switch: Synaptic Functionalities and Integration Strategies*
Selina La Barbera, David Guerin, Dominique Vuillaume and Fabien Alibart
(Poster) 17emes Journees Nationales du Reseau Doctoral en Micro-Nanoelectronique, JNRDM 2014, Villeneuve d'Ascq, France

A.3 Seminar and Oral Presentations

1. *Short-Term to Long-Term Plasticity Transition in Filamentary Switching for Memory Applications*
Selina La Barbera, Adrien F. Vincent, Dominique Vuillaume, Damien Querlioz and Fabien Alibart
International Conference on Memristive Systems (MEMRISYS 2015), November 8, 2015, Cyprus

2. *Memristive Nanodevices for Neuromorphic Computing*
Selina La Barbera, Dominique Vuillaume and Fabien Alibart
Bio-inspired Workshop, IRCICA - USR 3380 , October 16, 2015, Villeneuve-d'Ascq, France
3. *Filamentary Switching: Synaptic Plasticity through Device Volatility*
Selina La Barbera, Dominique Vuillaume and Fabien Alibart
Intermolecular Inc., September 8, 2015, San Jose, California
4. *Short-Term to Long-Term Plasticity Transition by Filamentary Switching Volatility Control*
Selina La Barbera, Dominique Vuillaume and Fabien Alibart
IBM Research - Almaden, August 25, 2015, San Jose, California
5. *Filamentary Switching: Synaptic Plasticity through Device Volatility*
Selina La Barbera, Dominique Vuillaume and Fabien Alibart
IBM Research - Zurich, November 27, 2014, Rueschlikon, Switzerland
6. *Atomic switch: Synaptic Functionalities and Integration Strategies*
Selina La Barbera, David Guerin, Dominique Vuillaume and Fabien Alibart
Strukov Research Group UCSB, August 18, 2014, Santa Barbara, California
7. *Atomic Switch: Synaptic Functionalities and Integration Strategies* *Selina La Barbera, David Guerin, Dominique Vuillaume and Fabien Alibart
European Materials Research Society Spring Meeting, E-MRS Spring 26/05/2014, Lille, France

A.4 Involvement in Journal and Conferences

1. *Journal Reviewer*,
Advanced Electronic Materials
2. *TPC member - Reviewer*,
International Conference on Memristive Systems (MEMRISYS) 2015,
Cyprus
3. *Reviewer*,
IEEE International Symposium on Circuits and Systems, ISCAS 2016
4. *Organizer Sponsor section*,
17emes Journées Nationales du Réseau Doctoral en Micro-Nanoelectronique,
JNRDM 2014, Villeneuve d'Ascq (France)

Appendix B

Résumé en française

B.1 Chapitre I: Contexte général et motivation

Ce chapitre d'ouverture fournit au lecteur le contexte général dans lequel cette thèse est placée en mettant en évidence la motivation de la recherche dans le domaine des systèmes neuromorphiques. Nous présentons l'état de l'art du codage de l'information neuro-inspiré. L'objectif principal est de concevoir et de fabriquer des réseaux de neurones artificiels dont les principes d'organisation sont basés sur ceux des systèmes nerveux biologiques.

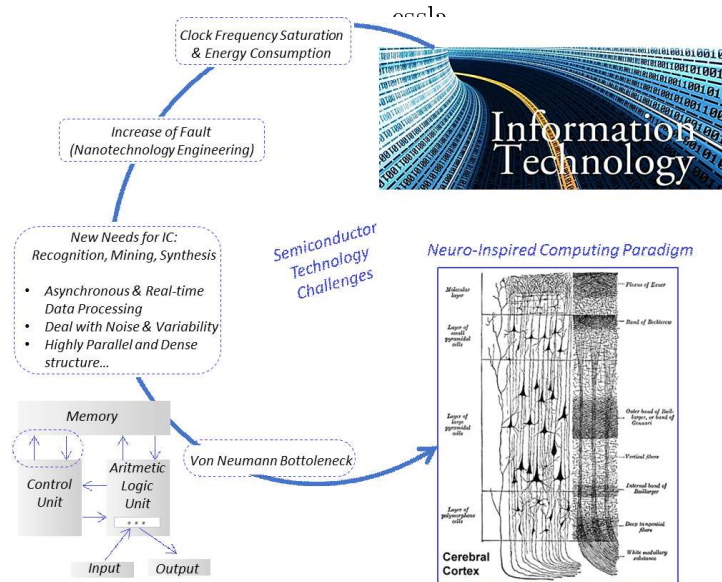


Figure B.1: Schematic of the Motivations behind the Neuro-Inspired Computing Paradigm.

À cette fin, nous discutons différentes directions de recherche. En mettant l'accent sur l'approche des nanotechnologies, le lecteur est introduit dans un aperçu de la recherche actuelle sur les mémoires à l'échelle nanométrique

aptes à l'implémentation des fonctionnalités bio-inspirées tels que la plasticité synaptique.

B.2 Chapitre II: Réseaux neuronaux neuromorphiques avec commutation filamentaire

Dans la première partie de ce chapitre, nous décrivons les mémoires à commutation résistive en nous concentrant sur une classe particulière: la technologie filamentaire et plus particulièrement les cellules à métallisation électrochimiques. La deuxième partie de ce chapitre poursuit cette ligne de recherche, au niveau circuits et systèmes, en présentant l'état de l'art des stratégies d'intégration.

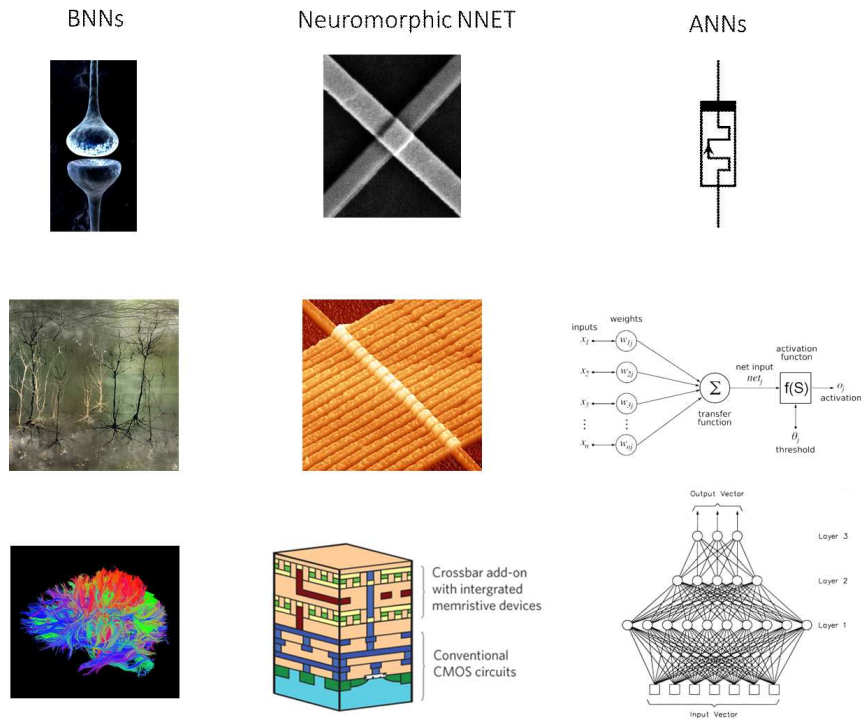


Figure B.2: The Nanotechnology Approach: Neuromorphic NNET

Enfin, nous discutons des avantages et des inconvénients des approches d'intégration présentées d'un point de vue du codage de l'information en soulignant les efforts d'ingénierie qui doivent être faits et qui sont nécessaires pour l'avenir des architectures matérielles neuromorphiques.

B.3 Chapitre III: Commutation filamentaire: Développement et Caractérisation

En motivant le choix de la technologie des cellules à métallisation électrochimiques a base de sulfure d'argent, dans le contexte de l'ingénierie neuro-morphique, ce chapitre est consacré à la procédure expérimentale, en termes de techniques de fabrication et de caractérisation électrique du dispositif, réalisée pendant ce travail de thèse.

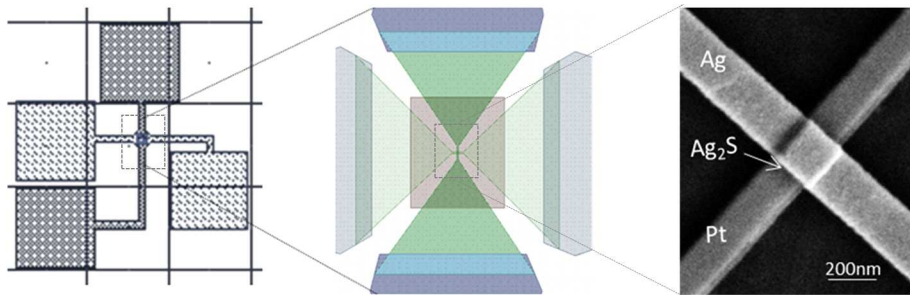


Figure B.3: Nanoscale ECM cell configuration: Layaout Editor device design and SEM image of the device realized ($200nm \times 200nm$ of cross-point active area).

En particulier, nous présentons les différentes technologies de nanofabrication utilisées pour réaliser des mémoires filamenteuses: une approche top-

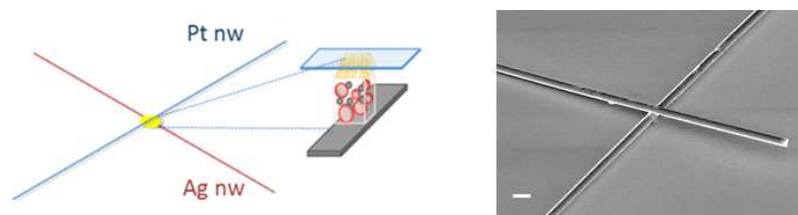


Figure B.4: Nanoscale ECM cell configuration by self-assembly of NWs. Schematic and SEM images: scale-bar $200nm$.

down plus proche de l'architecture des reseaux de neurones artificiels et une approche bottom-up plus proche de la biologie.

B.4 Chapitre IV: Plasticité synaptique avec com- mutation filamentaire

Dans ce chapitre, en allant plus loin dans l'expression de la plasticité synaptique observée dans les synapses biologiques, nous démontrons qu'un comportement plastique complexe peut émerger à partir de cellules mémoire, offrant une voie prometteuse et intéressante pour enrichir et améliorer les futurs systèmes de calcul bio-inspiré.

En nous intéressant à la physique des composants mémoires filamentaires de type cellules électrochimiques, nous démontrons comment les processus de mémoire à court terme et de mémoire à long terme présents dans les synapses biologiques (STP et LTP) peuvent être réalisés en contrôlant la croissance de filaments de type dendritiques. Nous avons démontré que la forme complexe des filaments, telles que les chemins dendritiques de densité et de largeur variables, peut permettre un contrôle indépendant des processus à long et à court terme en proposant une manière flexible de programmer le dispositif mémoire (i.e. les poids synaptiques) et la relative volatilité du dispositif.

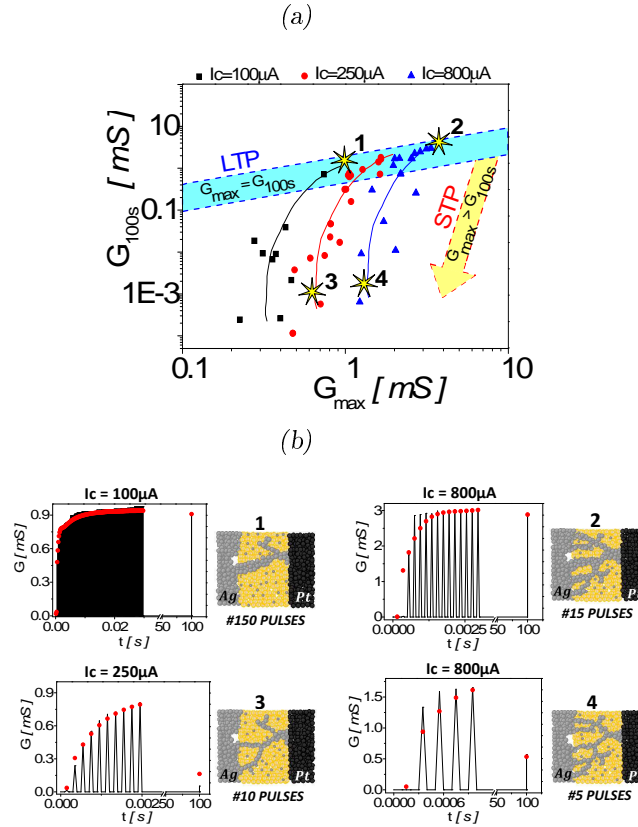


Figure B.5: Implementation of the Synaptic Adaptation through STP to LTP transition. (a) After a conditioning loop (full SET and RESET cycle with current compliance, I_c), the device is stressed with a burst of spikes, which induce a potentiation from the OFF state to a final conductive ON state, G_{max} . Device conductance is measured 100s after the end of the burst to evaluate the relaxation. Different transitions from STP to LTP are obtained with different conditioning I_c values ($I_c = 100, 250, 800 \mu A$). (b) Two examples of LTP (cases 1 and 2) and STP (cases 3 and 4), for the case in which the number of pulses is set as the key plasticity factor and the I_c value is set as the dendritic path definition. The density (through I_c) and diameter (through burst excitation) of the dendritic branches can be tuned independently to reproduce various STP/LTP combinations.

En particulier pour le réglage de la plasticité, nous avons utilisé: (i) la limita-

tion du courant au cours du pré-conditionnement qui régule la conductance moyenne du filament lors de la commutation en mode d'impulsion et (ii) la force d'excitation (nombre d'impulsions ou des impulsions) qui gère le transition de la plasticité à court terme et de l'état de la conductance maximale atteinte après la potentiation qui peut être associée à une augmentation du diamètre du filament.

B.5 Chapitre V: Plusieurs mécanismes de plasticité avec commutation filamentaire

Dans ce chapitre, nous implémentons dans ces composants une fonctionnalité synaptique basée sur la corrélation temporelle entre les signaux provenant des neurones d'entrée et de sortie, la STDP (Spike Timing Dependent Plasticity). Ces deux approches (STP/LTP et STDP) sont ensuite analysées à partir d'un modèle inspiré de la biologie permettant de mettre l'accent sur l'analogie entre synapses biologiques et composants mémoires filamenteuses.

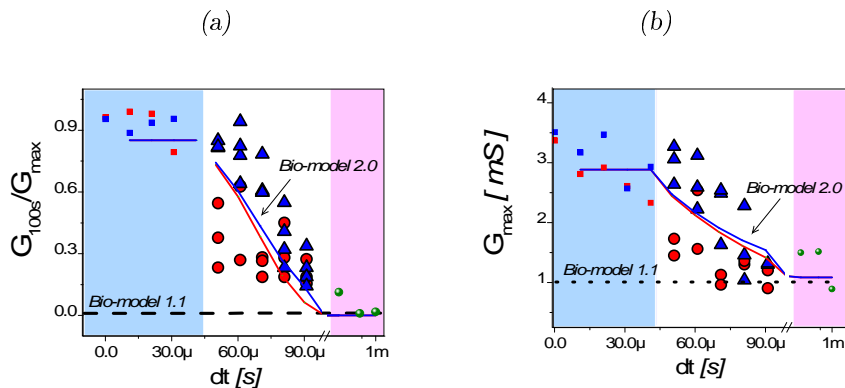


Figure B.6: Synaptic Learning Implementation in ECM cells. (a) G_{100s}/G_{max} conductance change as a function of the time correlation between pre- and post-spike dt . (b) Similarly, G_{max} conductance change as a function of the time correlation between pre- and post-spike dt .

En tenant compte de la corrélation temporelle entre les impulsions et la fréquence moyenne de pré-neurone avec un protocole simple composé d'impulsions de forme de carré. Nous avons étudié les interactions à l'échelle des temps courts dans nos dispositifs qui semblent raisonnablement liées aux effets de la température. Cette approche a permis de démontrer des fonctions bio-inspirées avancées dans lesquels les mécanismes de plasticité multiples peuvent être implémentés et contrôlés indépendamment dans un élément de mémoire unique. Ces résultats pourraient fournir une solution générale pour le développement de circuits bio-inspirés.

B.6 Chapitre VI: Modélisation et simulations de circuit de la commutation filamentaire

Dans ce chapitre, nous démontrons comment un riche panel de fonctionnalités peut être intégré dans un seul élément memristif filamentaire décrit par un modèle biologique de la plasticité synaptique qui a été utilisé pour des simulations de circuit. En exploitant une forme de plasticité synaptique de type adaptation synaptique basée sur les activités pré-synaptiques, nous avons modélisé en collaboration avec Adrien Vincent, Christopher Bennett et Dr. Damien Querlioz de l’Institut d’Electronique Fondamental (IEF) de l’Université de Paris Sud, un système (6×6 at 28×28 cross-bar) capables de mémoriser des motifs propres, bien que les entrées présentées dans la phase de programmation étaient bruitées.

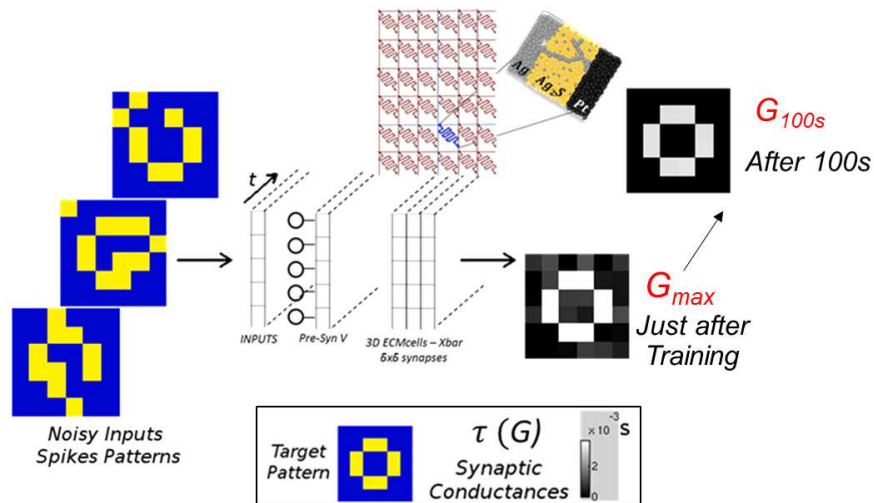


Figure B.7: ECM cell Cross-bar implementation for memory applications: a crossbar system (6×6) based on pre-synaptic activities is able to store clean patterns despite the inputs presented in the programming/learning phase are noisy by exploiting the *STP* to *LTP* transition.

En considérant les interactions aux temps courts une autre forme de plasticité synaptique a été implémenté dans nos composants mémoires filamentaires: l'apprentissage synaptique. Nous proposons d'utiliser le modèle bio-inspiré pour l'apprentissage des données dynamiques. Ces résultats ouvrent la voie à l'ingénierie future des systèmes de calcul neuromorphiques, où les comportements complexes basés sur la physique des composants memristifs peuvent être exploités.

B.7 Chapitre VII: Conclusions et perspectives

Dans cette direction de recherche émergente du codage de l'information neuro-inspiré les interactions multidisciplinaires, des la biologie, neurosciences computationnelles, des mathématiques, de l'architecture de circuit, des systèmes informatiques, de la microélectronique, les nanotechnologies et de la physique, sont d'une importance primordiale pour un développement futur de systèmes matériels neuromorphiques.

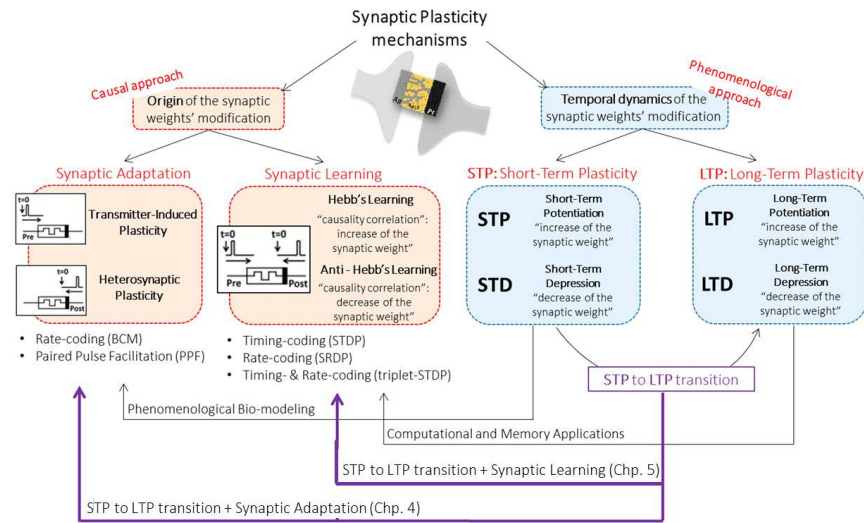


Figure B.8: Towards Multiple Plasticity Mechanisms: schematic of the proposed Synaptic Plasticity mechanisms scenario. By following a casual approach we can distinguish Synaptic Adaptation from Synaptic Learning while by following a phenomenological one we can have short- or long term plasticity. The originality of this PhD work is linked to the STP to LTP transition, through which we demonstrated how it coexists with Synaptic Adaptation (in the previous chapter) or with Synaptic Learning (in this chapter).

Les travaux futurs devraient se concentrer sur comment améliorer la compréhension des mécanismes du cerveau humain et proposer des realitation innovant de puces bio-inspirées.

Bibliography

- Abbott, L., Varela, J., Sen, K., and Nelson, S. (1997). Synaptic depression and cortical gain control. *Science*, 275(5297):221–224.
- Abbott, L. F. and Nelson, S. B. (2000). Synaptic plasticity: taming the beast. *Nature neuroscience*, 3:1178–1183.
- Ai, N., Zhou, Y., Zheng, Y., Chen, H., Wang, J., Pei, J., and Cao, Y. (2013). Achieving high sensitivity in single organic submicrometer ribbon based photodetector through surface engineering. *Organic Electronics*, 14(4):1103–1108.
- Alibart, F., Gao, L., Hoskins, B. D., and Strukov, D. B. (2012a). High precision tuning of state for memristive devices by adaptable variation tolerant algorithm. *Nanotechnology*, 23(7):075201.
- Alibart, F., Pleutin, S., Bichler, O., Gamrat, C., Serrano-Gotarredona, T., Linares-Barranco, B., and Vuillaume, D. (2012b). A memristive nanoparticle organic hybrid synapstor for neuroinspired computing. *Advanced Functional Materials*, 22(3):609–616.
- Alibart, F., Pleutin, S., Gu erin, D., Novembre, C., Lenfant, S., Lmimouni, K., Gamrat, C., and Vuillaume, D. (2010). An organic nanoparticle transistor behaving as a biological spiking synapse. *Advanced Functional Materials*, 20(2):330–337.
- Alibart, F., Zamanidoost, E., and Strukov, D. B. (2013). Pattern classification by memristive crossbar circuits using ex situ and in situ training. *Nature communications*, 4.
- Ananthanarayanan, R., Esser, S. K., Simon, H. D., and Modha, D. S. (2009). The cat is out of the bag: cortical simulations with 109 neurons, 1013 synapses. In *High Performance Computing Networking, Storage and Analysis, Proceedings of the Conference on*, pages 1–12. IEEE.
- Atkinson, R. C. and Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. *The psychology of learning and motivation*, 2:89–195.

- Avizienis, A. V., Sillin, H. O., Martin Olmos, C., Shieh, H. H., Aono, M., Stieg, A. Z., and Gimzewski, J. K. (2012). Neuromorphic atomic switch networks. *PloS one*, 7(8):e42772.
- Backus, J. (1978). Can programming be liberated from the von neumann style?: a functional style and its algebra of programs. *Communications of the ACM*, 21(8):613–641.
- Baikalov, A., Wang, Y., Shen, B., Lorenz, B., Tsui, S., Sun, Y., Xue, Y., and Chu, C. (2002). Field driven hysteretic and reversible resistive switch at the ag pr0 7ca0 3mno3 interface. *arXiv preprint cond-mat/0212464*.
- Bard, A. J. and Faulkner, L. R. (1980). *Electrochemical methods: fundamentals and applications*, volume 2. Wiley New York.
- Bear, M. F. (1995). Mechanism for a sliding synaptic modification threshold. *Neuron*, 15(1):1–4.
- Beck, A., Bednorz, J., Gerber, C., Rossel, C., and Widmer, D. (2000). Reproducible switching effect in thin oxide films for memory applications. *Applied Physics Letters*, 77(1):139–141.
- Bi, G.-q. and Poo, M.-m. (1998a). Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *The Journal of neuroscience*, 18(24):10464–10472.
- Bi, G.-q. and Poo, M.-m. (1998b). Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *The Journal of neuroscience*, 18(24):10464–10472.
- Bibes, M., Grollier, J., Barthélémy, A., and Mage, J. (2010). Ferroelectric device with adjustable resistance. *Patent WO*, 2010142762:A1.
- Bichler, O., Querlioz, D., Thorpe, S. J., Bourgoin, J. P., and Gamrat, C. (2012). Extraction of temporally correlated features from dynamic vision sensors with spike timing dependent plasticity. *Neural Networks*, 32:339–348.
- Bienenstock, E. L., Cooper, L. N., and Munro, P. W. (1982). Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. *The Journal of Neuroscience*, 2(1):32–48.
- Bishop, C. M. (1995). *Neural networks for pattern recognition*. Oxford university press.
- Bliss, T. V., Collingridge, G. L., et al. (1993). A synaptic model of memory: longterm potentiation in the hippocampus. *Nature*, 361(6407):31–39.

- Boegerhausen, M., Suter, P., and Liu, S.-C. (2003). Modeling short-term synaptic depression in silicon. *Neural Computation*, 15(2):331–348.
- Boyden, E. S., Katoh, A., and Raymond, J. L. (2004). Cerebellum-dependent learning: the role of multiple plasticity mechanisms. *Neuroscience*, 27.
- Buonomano, D. V. and Karmarkar, U. R. (2002). Book review: How do we tell time? *The Neuroscientist*, 8(1):42–51.
- Buonomano, D. V. and Maass, W. (2009). State-dependent computations: spatiotemporal processing in cortical networks. *Nature Reviews Neuroscience*, 10(2):113–125.
- Burger, J. R. and Teuscher, C. (2013). Variation tolerant computing with memristive reservoirs. In *Nanoscale Architectures (NANOARCH), 2013 IEEE/ACM International Symposium on*, pages 1–6. IEEE.
- Burnet, S. F. M. et al. (1959). *The clonal selection theory of acquired immunity*. University Press Cambridge.
- Burr, G. W., Breitwisch, M. J., Franceschini, M., Garetto, D., Gopalakrishnan, K., Jackson, B., Kurdi, B., Lam, C., Lastras, L. A., Padilla, A., et al. (2010). Phase change memory technology. *Journal of Vacuum Science & Technology B*, 28(2):223–262.
- Cantley, K. D., Subramaniam, A., Stiegler, H. J., Chapman, R., Vogel, E. M., et al. (2011). Hebbian learning in spiking neural networks with nanocrystalline silicon tfts and memristive synapses. *Nanotechnology, IEEE Transactions on*, 10(5):1066–1073.
- Carpenter, G. A. (1989). Neural network models for pattern recognition and associative memory. *Neural networks*, 2(4):243–257.
- Celano, U., Goux, L., Opsomer, K., Belmonte, A., Iapichino, M., Detavernier, C., Jurczak, M., and Vandervorst, W. (2013). Switching mechanism and reverse engineering of low-power cu-based resistive switching devices. *Nanoscale*, 5(22):11187–11192.
- Chae, S. C., Lee, J. S., Kim, S., Lee, S. B., Chang, S. H., Liu, C., Kahng, B., Shin, H., Kim, D.-W., Jung, C. U., et al. (2008). Random circuit breaker network model for unipolar resistance switching. *Advanced materials*, 20(6):1154–1159.
- Chandler, D. and Munday, R. (2011). *A dictionary of media and communication*. Oxford University Press.
- Chang, S. H., Lee, S. B., Jeon, D. Y., Park, S. J., Kim, G. T., Yang, S. M., Chae, S. C., Yoo, H. K., Kang, B. S., Lee, M.-J., et al. (2011a). Oxide

- double-layer nanocrossbar for ultrahigh-density bipolar resistive memory. *Advanced Materials*, 23(35):4063–4067.
- Chang, T., Jo, S.-H., Kim, K.-H., Sheridan, P., Gaba, S., and Lu, W. (2011b). Synaptic behaviors and modeling of a metal oxide memristive device. *Applied physics A*, 102(4):857–863.
- Chen, H., Chiang, R. H., and Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS quarterly*, 36(4):1165–1188.
- Chen, H., Lee, Y., Sun, G., Lee, H., Maxwell, T., and Giles, C. L. (1986). High order correlation model for associative memory. In *Neural networks for computing*, volume 151, pages 86–99. AIP Publishing.
- Chen, J.-Y., Hsin, C.-L., Huang, C.-W., Chiu, C.-H., Huang, Y.-T., Lin, S.-J., Wu, W.-W., and Chen, L.-J. (2013). Dynamic evolution of conducting nanofilament in resistive switching memories. *Nano letters*, 13(8):3671–3677.
- Choi, S.-J., Kim, G.-B., Lee, K., Kim, K.-H., Yang, W.-Y., Cho, S., Bae, H.-J., Seo, D.-S., Kim, S.-I., and Lee, K.-J. (2011). Synaptic behaviors of a single metal oxide metal resistive device. *Applied Physics A*, 102(4):1019–1025.
- Chua, L. O. (1971). Memristor the missing circuit element. *Circuit Theory, IEEE Transactions on*, 18(5):507–519.
- Clopath, C., Büsing, L., Vasilaki, E., and Gerstner, W. (2010). Connectivity reflects coding: a model of voltage-based stdp with homeostasis. *Nature neuroscience*, 13(3):344–352.
- Collingridge, G. L., Peineau, S., Howland, J. G., and Wang, Y. T. (2010). Long term depression in the cns. *Nature Reviews Neuroscience*, 11(7):459–473.
- Committee, I. R. et al. (2013). International technology roadmap for semiconductors: 2013 edition executive summary. *Semiconductor Industry Association, San Francisco, CA, available at: <http://www.itrs.net/Links/2013ITRS/2013Chapters/2013ExecutiveSummary.pdf>*.
- Cui, Y., Wei, Q., Park, H., and Lieber, C. M. (2001). Nanowire nanosensors for highly sensitive and selective detection of biological and chemical species. *Science*, 293(5533):1289–1292.
- Dayeh, S. A., Soci, C., Paul, K., Edward, T. Y., and Wang, D. (2007). Influence of surface states on the extraction of transport parameters from inas nanowire field effect transistors. *Applied Physics Letters*, 90(16):162112.

- Deng, Y., Josberger, E., Jin, J., Rousdari, A. F., Helms, B. A., Zhong, C., Anantram, M., and Rolandi, M. (2013). H⁺-type and oh-type biological protonic semiconductors and complementary devices. *Scientific reports*, 3.
- Desbief, S., Kyndiah, A., Guerin, D., Gentili, D., Murgia, M., Lenfant, S., Alibart, F., Cramer, T., Biscarini, F., and Vuillaume, D. (2015). Low voltage and time constant organic synapse transistor. *Organic Electronics*, 21:47–53.
- Dillenberger, D. E., Gil, D., Nitta, S. V., and Ritter, M. B. (2011). Frontiers of information technology. *IBM Journal of Research and Development*, 55(5):1–1.
- Dirkmann, S., Ziegler, M., Hansen, M., Kohlstedt, H., Trieschmann, J., and Mussenbrock, T. (2015). Kinetic simulation of filament growth dynamics in memristive electrochemical metallization devices. *arXiv preprint arXiv:1509.00208*.
- Du, C., Ma, W., Chang, T., Sheridan, P., and Lu, W. D. (2015). Biorealistic implementation of synaptic functions with oxide memristors through internal ionic dynamics. *Advanced Functional Materials*, 25(27):4290–4299.
- Duch, W., Adamczak, R., and Grkabczewski, K. (1998). Extraction of logical rules from neural networks. *Neural Processing Letters*, 7(3):211–219.
- Fujii, T., Arita, M., Takahashi, Y., and Fujiwara, I. (2011). In situ transmission electron microscopy analysis of conductive filament during solid electrolyte resistance switching. *Applied Physics Letters*, 98(21):212104.
- Ghirardi, M., Montarolo, P. G., and Kandel, E. R. (1995). A novel intermediate stage in the transition between short and long term facilitation in the sensory to motor neuron synapse of aplysia. *Neuron*, 14(2):413–420.
- Gjorgjieva, J., Clopath, C., Audet, J., and Pfister, J.-P. (2011). A triplet spike timing dependent plasticity model generalizes the bienenstock-cooper-munro rule to higher-order spatiotemporal correlations. *Proceedings of the National Academy of Sciences*, 108(48):19383–19388.
- Govoreanu, B., Kar, G., Chen, Y., Paraschiv, V., Kubicek, S., Fantini, A., Radu, I., Goux, L., Clima, S., Degraeve, R., et al. (2011). 10 × 10nm² hf/hfo x crossbar resistive ram with excellent performance, reliability and low energy operation. In *Electron Devices Meeting (IEDM), 2011 IEEE International*, pages 31–6. IEEE.
- Guo, X., Schindler, C., Menzel, S., and Waser, R. (2007). Understanding the switching-off mechanism in ag⁺ migration based resistively switching model systems. *Applied Physics Letters*, 91(13):133513.

- Hasegawa, T., Ohno, T., Terabe, K., Tsuruoka, T., Nakayama, T., Gimzewski, J. K., and Aono, M. (2010). Learning abilities achieved by a single solid state atomic switch. *Advanced Materials*, 22(16):1831–1834.
- Hasler, J. and Marr, B. (2013). Finding a roadmap to achieve large neuromorphic hardware systems. *Frontiers in neuroscience*, 7.
- Hebb, D. O. (1949). The first stage of perception: Growth of the assembly. *The Organization of Behavior*, pages 60–78.
- Henaff, M., Jarrett, K., Kavukcuoglu, K., and LeCun, Y. (2011). Unsupervised learning of sparse features for scalable audio classification. In *ISMIR*, volume 11, page 298.
- Hinton, G., Vinyals, O., and Dean, J. (2015). Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*.
- 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.
- Hsiung, C.-P., Liao, H.-W., Gan, J.-Y., Wu, T.-B., Hwang, J.-C., Chen, F., and Tsai, M.-J. (2010). Formation and instability of silver nanofilament in ag based programmable metallization cells. *ACS nano*, 4(9):5414–5420.
- Ielmini, D., Bruchhaus, R., and Waser, R. (2011). Thermochemical resistive switching: materials, mechanisms, and scaling projections. *Phase Transitions*, 84(7):570–602.
- Indiveri, G., Linares-Barranco, B., Hamilton, T. J., Van Schaik, A., Etienne-Cummings, R., Delbruck, T., Liu, S.-C., Dudek, P., Häflicher, P., Renaud, S., et al. (2011). Neuromorphic silicon neuron circuits. *Frontiers in neuroscience*, 5.
- Izhikevich, E. M. et al. (2003). Simple model of spiking neurons. *IEEE Transactions on neural networks*, 14(6):1569–1572.
- Jain, A. K., Mao, J., and Mohiuddin, K. (1996). Artificial neural networks: A tutorial. *Computer*, (3):31–44.
- Janek, J. (2009). Mixed conductors: The bridge to redox switches. *Nature materials*, 8(2):88–89.
- Jo, S. H., Chang, T., Ebong, I., Bhadviya, B. B., Mazumder, P., and Lu, W. (2010a). Nanoscale memristor device as synapse in neuromorphic systems. *Nano letters*, 10(4):1297–1301.

- Jo, S. H., Chang, T., Ebong, I., Bhadviya, B. B., Mazumder, P., and Lu, W. (2010b). Nanoscale memristor device as synapse in neuromorphic systems. *Nano letters*, 10(4):1297–1301.
- Josberger, E. E., Deng, Y., Sun, W., Kautz, R., and Rolandi, M. (2014). Two terminal protonic devices with synaptic like short term depression and device memory. *Advanced Materials*.
- Jung, K., Seo, H., Kim, Y., Im, H., Hong, J., Park, J.-W., and Lee, J.-K. (2007). Temperature dependence of high and low resistance bistable states in polycrystalline nio films. *Applied physics letters*, 90(5):052104.
- Kaminski, W. A. and Wojcik, G. M. (2004). Liquid state machine built of hodgkin huxley neurons. *Informatica, Lith. Acad. Sci.*, 15(1):39–44.
- Kim, K., Chen, C.-L., Truong, Q., Shen, A. M., and Chen, Y. (2013a). A carbon nanotube synapse with dynamic logic and learning. *Advanced Materials*, 25(12):1693–1698.
- Kim, K., Chen, C.-L., Truong, Q., Shen, A. M., and Chen, Y. (2013b). A carbon nanotube synapse with dynamic logic and learning. *Advanced Materials*, 25(12):1693–1698.
- Kim, K.-H., Gaba, S., Wheeler, D., Cruz-Albrecht, J. M., Hussain, T., Srinivasa, N., and Lu, W. (2011). A functional hybrid memristor crossbar array-cmos system for data storage and neuromorphic applications. *Nano letters*, 12(1):389–395.
- Kim, S., Choi, S., and Lu, W. (2014). Comprehensive physical model of dynamic resistive switching in an oxide memristor. *ACS nano*, 8(3):2369–2376.
- Kim, S., Du, C., Sheridan, P., Ma, W., Choi, S., and Lu, W. D. (2015). Experimental demonstration of a second-order memristor and its ability to biorealistically implement synaptic plasticity. *Nano letters*, 15(3):2203–2211.
- Konkoli, Z. and Wendin, G. (2013). Toward bio inspired information processing with networks of nano scale switching elements. *arXiv preprint arXiv:1311.6259*.
- Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105.
- Krzysteczko, P., Münchenberger, J., Schäfers, M., Reiss, G., and Thomas, A. (2012). The memristive magnetic tunnel junction as a nanoscopic synapse neuron system. *Advanced Materials*, 24(6):762–766.

- Kügeler, C., Meier, M., Rosezin, R., Gilles, S., and Waser, R. (2009). High density 3d memory architecture based on the resistive switching effect. *Solid-State Electronics*, 53(12):1287–1292.
- Kulkarni, M. S. and Teuscher, C. (2012). Memristor based reservoir computing. In *Nanoscale Architectures (NANOARCH), 2012 IEEE/ACM International Symposium on*, pages 226–232. IEEE.
- Kuzum, D., Jeyasingh, R. G., Lee, B., and Wong, H.-S. P. (2011a). Nanoelectronic programmable synapses based on phase change materials for brain-inspired computing. *Nano letters*, 12(5):2179–2186.
- Kuzum, D., Jeyasingh, R. G., Lee, B., and Wong, H.-S. P. (2011b). Nanoelectronic programmable synapses based on phase change materials for brain inspired computing. *Nano letters*, 12(5):2179–2186.
- Kwon, D.-H., Kim, K. M., Jang, J. H., Jeon, J. M., Lee, M. H., Kim, G. H., Li, X.-S., Park, G.-S., Lee, B., Han, S., et al. (2010). Atomic structure of conducting nanofilaments in tio2 resistive switching memory. *Nature nanotechnology*, 5(2):148–153.
- La Barbera, S., Vuillaume, D., and Alibart, F. (2015). Filamentary switching: Synaptic plasticity through device volatility. *ACS nano*, 9(1):941–949.
- Lamprecht, R. and LeDoux, J. (2004). Structural plasticity and memory. *Nature Reviews Neuroscience*, 5(1):45–54.
- Lee, H., Chen, Y., Chen, P., Gu, P., Hsu, Y., Wang, S., Liu, W., Tsai, C., Sheu, S., Chiang, P.-C., et al. (2010). Evidence and solution of over reset problem for hfo x based resistive memory with sub ns switching speed and high endurance. In *Electron Devices Meeting (IEDM), 2010 IEEE International*, pages 19–7. IEEE.
- Lee, M.-J., Han, S., Jeon, S. H., Park, B. H., Kang, B. S., Ahn, S.-E., Kim, K. H., Lee, C. B., Kim, C. J., Yoo, I.-K., et al. (2009). Electrical manipulation of nanofilaments in transition-metal oxides for resistance-based memory. *Nano letters*, 9(4):1476–1481.
- Lee, M.-J., Lee, C. B., Lee, D., Lee, S. R., Chang, M., Hur, J. H., Kim, Y.-B., Kim, C.-J., Seo, D. H., Seo, S., et al. (2011). A fast, high endurance and scalable non volatile memory device made from asymmetric ta2o5 x/tao2 x bilayer structures. *Nature materials*, 10(8):625–630.
- Lekshmi, I., Berera, G., Afsar, Y., Miao, G., Nagahama, T., Santos, T., and Moodera, J. (2008). Controlled synthesis and characterization of ag2s films with varied microstructures and its role as asymmetric barrier layer in trilayer junctions with dissimilar electrodes. *Journal of Applied Physics*, 103(9):093719.

- Liang, B. and Dubey, P. (2005). Recognition, mining and synthesis. *Intel Technology Journal*, 9(2).
- Likharev, K. K. and Strukov, D. B. (2005). Cmol: Devices, circuits, and architectures. In *Introducing Molecular Electronics*, pages 447–477. Springer.
- Lim, J., Ryu, S. Y., Kim, J., and Jun, Y. (2013). A study of tio₂/carbon black composition as counter electrode materials for dye-sensitized solar cells. *Nanoscale research letters*, 8(1):1–5.
- Lu, J.-Q. (2009). 3d hyperintegration and packaging technologies for micro nano systems. *Proceedings of the IEEE*, 97(1):18–30.
- Lukoševičius, M. and Jaeger, H. (2009). Reservoir computing approaches to recurrent neural network training. *Computer Science Review*, 3(3):127–149.
- Maass, W. and Natschläger, T. (1997). Networks of spiking neurons can emulate arbitrary hopfield nets in temporal coding. *Network: Computation in Neural Systems*, 8(4):355–371.
- Maijenburg, A., Maas, M., Rodijk, E., Ahmed, W., Kooij, E., Carlen, E., Blank, D., and Ten Elshof, J. (2011). Dielectrophoretic alignment of metal and metal oxide nanowires and nanotubes: A universal set of parameters for bridging prepatterned microelectrodes. *Journal of colloid and interface science*, 355(2):486–493.
- Malenka, R. C. and Bear, M. F. (2004). Ltp and ltd: an embarrassment of riches. *Neuron*, 44(1):5–21.
- Markram, H. (2006). The blue brain project. *Nature Reviews Neuroscience*, 7(2):153–160.
- 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.
- Markram, H., Pikus, D., Gupta, A., and Tsodyks, M. (1998). Potential for multiple mechanisms, phenomena and algorithms for synaptic plasticity at single synapses. *Neuropharmacology*, 37(4):489–500.
- Mayr, C., Partzsch, J., Noack, M., and Schüffny, R. (2013). Live demonstration: Multiple-timescale plasticity in a neuromorphic system. In *ISCAS*, pages 666–670.
- Mayr, C., Stärke, P., Partzsch, J., Cederstroem, L., Schüffny, R., Shuai, Y., Du, N., and Schmidt, H. (2012). Waveform driven plasticity in bifeo₃ memristive devices: model and implementation. In *Advances in Neural Information Processing Systems*, pages 1700–1708.

- McCulloch, W. S. and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133.
- McGaugh, J. L. (2000). Memory a century of consolidation. *Science*, 287(5451):248–251.
- Mead, C. (1990). Neuromorphic electronic systems. *Proceedings of the IEEE*, 78(10):1629–1636.
- Meherzi-Maghraoui, H., Dachraoui, M., Belgacem, S., Buhre, K., Kunst, R., Cowache, P., and Lincot, D. (1996). Structural, optical and transport properties of ag₂s films deposited chemically from aqueous solution. *Thin Solid Films*, 288(1):217–223.
- Merolla, P. A., Arthur, J. V., Alvarez-Icaza, R., Cassidy, A. S., Sawada, J., Akopyan, F., Jackson, B. L., Imam, N., Guo, C., and Nakamura, Y. (2014). A million spiking-neuron integrated circuit with a scalable communication network and interface. *Science*, 345(6197):668–673.
- Miao, F., Yi, W., Goldfarb, I., Yang, J. J., Zhang, M.-X., Pickett, M. D., Strachan, J. P., Medeiros-Ribeiro, G., and Williams, R. S. (2012). Continuous electrical tuning of the chemical composition of tao x based memristors. *ACS nano*, 6(3):2312–2318.
- Morales Masis, M., Wiemhöfer, H.-D., and Van Ruitenbeek, J. M. (2010). Towards a quantitative description of solid electrolyte conductance switches. *Nanoscale*, 2(10):2275–2280.
- Motohisa, J., Noborisaka, J., Takeda, J., Inari, M., and Fukui, T. (2004). Catalyst-free selective area movpe of semiconductor nanowires on (111) b oriented substrates. *Journal of crystal growth*, 272(1):180–185.
- Münstermann, R., Yang, J. J., Strachan, J. P., Medeiros-Ribeiro, G., Dittmann, R., and Waser, R. (2010). Morphological and electrical changes in tio₂ memristive devices induced by electroforming and switching. *physica status solidi (RRL)-Rapid Research Letters*, 4(1-2):16–18.
- Nadel, L. and Moscovitch, M. (1997). Memory consolidation, retrograde amnesia and the hippocampal complex. *Current opinion in neurobiology*, 7(2):217–227.
- Neves, G., Cooke, S. F., and Bliss, T. V. (2008). Synaptic plasticity, memory and the hippocampus: a neural network approach to causality. *Nature Reviews Neuroscience*, 9(1):65–75.
- Ngezahayo, A., Schachner, M., and Artola, A. (2000). Synaptic activity modulates the induction of bidirectional synaptic changes in adult mouse hippocampus. *The Journal of Neuroscience*, 20(7):2451–2458.

- Ohno, T., Hasegawa, T., Tsuruoka, T., Terabe, K., Gimzewski, J. K., and Aono, M. (2011a). Short-term plasticity and long-term potentiation mimicked in single inorganic synapses. *Nature materials*, 10(8):591–595.
- Ohno, T., Hasegawa, T., Tsuruoka, T., Terabe, K., Gimzewski, J. K., and Aono, M. (2011b). Short term plasticity and long term potentiation mimicked in single inorganic synapses. *Nature materials*, 10(8):591–595.
- Oka, T., Arita, R., and Aoki, H. (2003). Breakdown of a mott insulator: a nonadiabatic tunneling mechanism. *Physical review letters*, 91(6):066406.
- Pan, F., Yin, S., and Subramanian, V. (2011). A detailed study of the forming stage of an electrochemical resistive switching memory by kmc simulation. *Electron Device Letters, IEEE*, 32(7):949–951.
- Panneerselvam, A., Malik, M. A., O’Brien, P., and Raftery, J. (2008). The cvd of silver sulfide and silver thin films from a homoleptic crystalline single source precursor. *Journal of Materials Chemistry*, 18(27):3264–3269.
- Pickett, M. D., Medeiros-Ribeiro, G., and Williams, R. S. (2013). A scalable neuristor built with mott memristors. *Nature materials*, 12(2):114–117.
- Prezioso, M., Merrih Bayat, F., Hoskins, B., Adam, G., Likharev, K. K., and Strukov, D. B. (2015). Training and operation of an integrated neuromorphic network based on metal oxide memristors. *Nature*, 521(7550):61–64.
- Querlioz, D., Bichler, O., and Gamrat, C. (2011). Simulation of a memristor-based spiking neural network immune to device variations. In *Neural Networks (IJCNN), The 2011 International Joint Conference on*, pages 1775–1781. IEEE.
- Raychaudhuri, S., Dayeh, S. A., Wang, D., and Yu, E. T. (2009). Precise semiconductor nanowire placement through dielectrophoresis. *Nano letters*, 9(6):2260–2266.
- Rodríguez, A. N., Nair, M., and Nair, P. (2005). Structural, optical and electrical properties of chemically deposited silver sulfide thin films. *Semiconductor science and technology*, 20(6):576.
- Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386.
- Russo, U., Kamalanathan, D., Ielmini, D., Lacaita, A. L., and Kozicki, M. N. (2009). Study of multilevel programming in programmable metallization cell (pmc) memory. *Electron Devices, IEEE Transactions on*, 56(5):1040–1047.

- Sawa, A. (2008). Resistive switching in transition metal oxides. *Materials today*, 11(6):28–36.
- Schemmel, J., Grübl, A., Meier, K., and Mueller, E. (2006). Implementing synaptic plasticity in a vlsi spiking neural network model. In *Neural Networks, 2006. IJCNN'06. International Joint Conference on*, pages 1–6. IEEE.
- Schmalzried, H. (1980). Ag₂s the physical chemistry of an inorganic material. *Progress in Solid State Chemistry*, 13(2):119–157.
- Senn, W., Markram, H., and Tsodyks, M. (2001). An algorithm for modifying neurotransmitter release probability based on pre-and postsynaptic spike timing. *Neural Computation*, 13(1):35–67.
- Sharad, M., Augustine, C., Panagopoulos, G., and Roy, K. (2012). Spin based neuron model with domain-wall magnets as synapse. *Nanotechnology, IEEE Transactions on*, 11(4):843–853.
- Sillin, H. O., Aguilera, R., Shieh, H.-H., Avizienis, A. V., Aono, M., Stieg, A. Z., and Gimzewski, J. K. (2013). A theoretical and experimental study of neuromorphic atomic switch networks for reservoir computing. *Nanotechnology*, 24(38):384004.
- Sim, H., Choi, D., Lee, D., Seo, S., Lee, M.-J., Yoo, I.-K., and Hwang, H. (2005). Resistance switching characteristics of polycrystalline nb 2 o 5 for nonvolatile memory application. *Electron Device Letters, IEEE*, 26(5):292–294.
- Sjöström, P. J., Turrigiano, G. G., and Nelson, S. B. (2001). Rate, timing, and cooperativity jointly determine cortical synaptic plasticity. *Neuron*, 32(6):1149–1164.
- Sköldberg, J. and Wendin, G. (2007). Reconfigurable logic in nanoelectronic switching networks. *Nanotechnology*, 18(48):485201.
- Smith Jr, T., Lange, G., and Marks, W. (1996). Fractal methods and results in cellular morphology dimensions, lacunarity and multifractals. *Journal of neuroscience methods*, 69(2):123–136.
- Snider, G. S. (2008). Spike-timing-dependent learning in memristive nanodevices. In *Nanoscale Architectures, 2008. NANOARCH 2008. IEEE International Symposium on*, pages 85–92. IEEE.
- Son, J. and Shin, Y.-H. (2008). Direct observation of conducting filaments on resistive switching of nio thin films. *Applied Physics Letters*, 92(22):2106.
- Sourdet, V. and Debanne, D. (1999). The role of dendritic filtering in associative long-term synaptic plasticity. *Learning & Memory*, 6(5):422–447.

- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15(1):1929–1958.
- Stieg, A. Z., Avizienis, A. V., Sillin, H. O., Martin Olmos, C., Aono, M., and Gimzewski, J. K. (2012). Emergent criticality in complex turing b type atomic switch networks. *Advanced Materials*, 24(2):286–293.
- Strukov, D. B. (2011). Nanotechnology: smart connections. *Nature*, 476(7361):403–405.
- Strukov, D. B. and Likharev, K. K. (2007). Defect tolerant architectures for nanoelectronic crossbar memories. *Journal of Nanoscience and Nanotechnology*, 7(1):151–167.
- Strukov, D. B., Snider, G. S., Stewart, D. R., and Williams, R. S. (2008). The missing memristor found. *nature*, 453(7191):80–83.
- Strukov, D. B. and Williams, R. S. (2009a). Exponential ionic drift: fast switching and low volatility of thin film memristors. *Applied Physics A*, 94(3):515–519.
- Strukov, D. B. and Williams, R. S. (2009b). Four dimensional address topology for circuits with stacked multilayer crossbar arrays. *Proceedings of the National Academy of Sciences*, 106(48):20155–20158.
- Subramaniam, A., Cantley, K. D., Bersuker, G., Gilmer, D., and Vogel, E. M. (2013). Spike-timing-dependent plasticity using biologically realistic action potentials and low-temperature materials. *Nanotechnology, IEEE Transactions on*, 12(3):450–459.
- Sudholter, E., de Smet, L., Mescher, M., and Uliien, D. (2011). Organic surface modification of silicon nanowire based sensor devices, nanowires implementations and applications. *ISBN*, pages 978–953.
- Suri, M., Querlioz, D., Bichler, O., Palma, G., Vianello, E., Vuillaume, D., Gamrat, C., and DeSalvo, B. (2013). Bio-inspired stochastic computing using binary cbram synapses. *Electron Devices, IEEE Transactions on*, 60(7):2402–2409.
- Szot, K., Dittmann, R., Speier, W., and Waser, R. (2007). Nanoscale resistive switching in srtio3 thin films. *physica status solidi (RRL)-Rapid Research Letters*, 1(2):R86–R88.
- Szot, K., Speier, W., Bihlmayer, G., and Waser, R. (2006). Switching the electrical resistance of individual dislocations in single crystalline srtio3. *Nature materials*, 5(4):312–320.

- Temam, O. (2010). The rebirth of neural networks. In *International Symposium on Computer Architecture*.
- Torrezan, A. C., Strachan, J. P., Medeiros-Ribeiro, G., and Williams, R. S. (2011). Sub nanosecond switching of a tantalum oxide memristor. *Nanotechnology*, 22(48):485203.
- Tsui, S., Baikalov, A., Cmaidalka, J., Sun, Y., Wang, Y., Xue, Y., Chu, C., Chen, L., and Jacobson, A. (2004). Field induced resistive switching in metal oxide interfaces. *Applied physics letters*, 85(2):317–319.
- Tukker, R., van Rossum, A., Haselager, P., and Frank, S. (2012). *Echo state networks for hierarchical cognitive control*. PhD thesis, Master’s thesis, Donders Institute, Radboud University Nijmegen.
- Udipi, A. N., Muralimanohar, N., Chatterjee, N., Balasubramonian, R., Davis, A., and Jouppi, N. P. (2010). Rethinking dram design and organization for energy constrained multi cores. *ACM SIGARCH Computer Architecture News*, 38(3):175–186.
- Valov, I., Linn, E., Tappertzhofen, S., Schmelzer, S., van den Hurk, J., Lentz, F., and Waser, R. (2013a). Nanobatteries in redox-based resistive switches require extension of memristor theory. *Nature communications*, 4:1771.
- Valov, I., Linn, E., Tappertzhofen, S., Schmelzer, S., van den Hurk, J., Lentz, F., and Waser, R. (2013b). Nanobatteries in redox-based resistive switches require extension of memristor theory. *Nature communications*, 4:1771.
- Valov, I., Waser, R., Jameson, J. R., and Kozicki, M. N. (2011a). Electrochemical metallization memories fundamentals, applications, prospects. *Nanotechnology*, 22(25):254003.
- Valov, I., Waser, R., Jameson, J. R., and Kozicki, M. N. (2011b). Electrochemical metallization memories fundamentals, applications, prospects. *Nanotechnology*, 22(25):254003.
- Van Rossum, M. C., Bi, G. Q., and Turrigiano, G. G. (2000). Stable hebbian learning from spike timing-dependent plasticity. *The Journal of Neuroscience*, 20(23):8812–8821.
- Varela, J. A., Sen, K., Gibson, J., Fost, J., Abbott, L., and Nelson, S. B. (1997). A quantitative description of short-term plasticity at excitatory synapses in layer 2/3 of rat primary visual cortex. *The Journal of neuroscience*, 17(20):7926–7940.
- von Neumann, J. (1948). Electronic methods of computation. *Bulletin of the American Academy of Arts and Sciences*, 1(3):2–4.

- Wang, Z. Q., Xu, H. Y., Li, X. H., Yu, H., Liu, Y. C., and Zhu, X. J. (2012). Synaptic learning and memory functions achieved using oxygen ion migration diffusion in an amorphous ingazno memristor. *Advanced Functional Materials*, 22(13):2759–2765.
- Waser, R. and Aono, M. (2007). Nanoionics based resistive switching memories. *Nature materials*, 6(11):833–840.
- Waser, R., Dittmann, R., Staikov, G., and Szot, K. (2009). Redox based resistive switching memories nanoionic mechanisms, prospects and challenges. *Advanced Materials*, (21):2632–2663.
- Wei, Y., Wu, W., Guo, R., Yuan, D., Das, S., and Wang, Z. L. (2010). Wafer scale high throughput ordered growth of vertically aligned zno nanowire arrays. *Nano letters*, 10(9):3414–3419.
- Werbos, P. J. (1988). Generalization of backpropagation with application to a recurrent gas market model. *Neural Networks*, 1(4):339–356.
- Whang, D., Jin, S., Wu, Y., and Lieber, C. M. (2003). Large scale hierarchical organization of nanowire arrays for integrated nanosystems. *Nano letters*, 3(9):1255–1259.
- Whitlock, J. R., Heynen, A. J., Shuler, M. G., and Bear, M. F. (2006). Learning induces long term potentiation in the hippocampus. *science*, 313(5790):1093–1097.
- Widrow, B., HOFF, M. E., et al. (1960). Adaptive switching circuits.
- Williamson, A., Schumann, L., Hiller, L., Klefenz, F., Hoerselmann, I., Husar, P., and Schober, A. (2013). Synaptic behavior and stdp of asymmetric nanoscale memristors in biohybrid systems. *Nanoscale*, 5(16):7297–7303.
- Wong, H. P., Raoux, S., Kim, S., Liang, J., Reifenberg, J. P., Rajendran, B., Asheghi, M., and Goodson, K. E. (2010). Phase change memory. *Proceedings of the IEEE*, 98(12):2201–2227.
- Wong, H. S. P., Lee, H. Y., Yu, S., Chen, Y. S., Wu, Y., Chen, P.-S., Lee, B., Chen, F. T., and Tsai, M.-J. (2012). Metal oxide rram. *Proceedings of the IEEE*, 100(6):1951–1970.
- Xia, Q., Robinett, W., Cumbie, M. W., Banerjee, N., Cardinali, T. J., Yang, J. J., Wu, W., Li, X., Tong, W. M., Strukov, D. B., et al. (2009). Memristor cmos hybrid integrated circuits for reconfigurable logic. *Nano letters*, 9(10):3640–3645.

- Xu, C., Niu, D., Muralimanohar, N., Balasubramonian, R., Zhang, T., Yu, S., and Xie, Y. (2015). Overcoming the challenges of crossbar resistive memory architectures. In *High Performance Computer Architecture (HPCA), 2015 IEEE 21st International Symposium on*, pages 476–488. IEEE.
- Xu, D., Subramanian, A., Dong, L., and Nelson, B. J. (2009). Shaping nanoelectrodes for high precision dielectrophoretic assembly of carbon nanotubes. *Nanotechnology, IEEE Transactions on*, 8(4):449–456.
- Yang, J. J., Pickett, M. D., Li, X., Ohlberg, D. A., Stewart, D. R., and Williams, R. S. (2008). Memristive switching mechanism for metal oxide metal nanodevices. *Nature nanotechnology*, 3(7):429–433.
- Yang, J. J., Strukov, D. B., and Stewart, D. R. (2013a). Memristive devices for computing. *Nature nanotechnology*, 8(1):13–24.
- Yang, R., Terabe, K., Liu, G., Tsuruoka, T., Hasegawa, T., Gimzewski, J. K., and Aono, M. (2012a). On demand nanodevice with electrical and neuromorphic multifunction realized by local ion migration. *ACS nano*, 6(11):9515–9521.
- Yang, Y., Choi, S., and Lu, W. (2013b). Oxide heterostructure resistive memory. *Nano letters*, 13(6):2908–2915.
- Yang, Y., Gao, P., Gaba, S., Chang, T., Pan, X., and Lu, W. (2012b). Observation of conducting filament growth in nanoscale resistive memories. *Nature communications*, 3:732.
- Yang, Y., Gao, P., Li, L., Pan, X., Tappertzhofen, S., Choi, S., Waser, R., Valov, I., and Lu, W. D. (2014). Electrochemical dynamics of nanoscale metallic inclusions in dielectrics. *Nature communications*, 5.
- Yao, J., Zhong, L., Natelson, D., and Tour, J. M. (2012). In situ imaging of the conducting filament in a silicon oxide resistive switch. *Scientific reports*, 2.
- Yasuhara, R., Fujiwara, K., Horiba, K., Kumigashira, H., Kotsugi, M., Oshima, M., and Takagi, H. (2009). Inhomogeneous chemical states in resistance-switching devices with a planar type ptcuopt structure. *Applied Physics Letters*, 95(1):012110.
- Yuan, P., Leonetti, M. D., Pico, A. R., Hsiung, Y., and MacKinnon, R. (2010). Structure of the human bk channel ca2 activation apparatus at 3.0 Å resolution. *Science*, 329(5988):182–186.
- Zeng, F., Li, S., Yang, J., Pan, F., and Guo, D. (2014). Learning processes modulated by the interface effects in a ti conducting polymer ti resistive switching cell. *RSC Advances*, 4(29):14822–14828.

- Zenke, F., Agnes, E. J., and Gerstner, W. (2015). Diverse synaptic plasticity mechanisms orchestrated to form and retrieve memories in spiking neural networks. *Nature communications*, 6.
- Zhu, L. Q., Wan, C. J., Guo, L. Q., Shi, Y., and Wan, Q. (2014). Artificial synapse network on inorganic proton conductor for neuromorphic systems. *Nature communications*, 5.
- Zhu, Y., Xu, F., Qin, Q., Fung, W. Y., and Lu, W. (2009). Mechanical properties of vapor liquid solid synthesized silicon nanowires. *Nano letters*, 9(11):3934–3939.
- Ziegler, L., Zenke, F., Kastner, D. B., and Gerstner, W. (2015). Synaptic consolidation: from synapses to behavioral modeling. *The Journal of Neuroscience*, 35(3):1319–1334.
- Zucker, R. S. and Regehr, W. G. (2002). Short term synaptic plasticity. *Annual review of physiology*, 64(1):355–405.

*Development of Filamentary Memristive Devices
for Synaptic Plasticity Implementation*

Selina La Barbera

18 December 2015