Lithium-based nano-ionic synaptic transistors for neuromorphic computing

Ngoc Anh Nguyen

université PARIS-SACLAY

Lithium-based Nano-ionic Synaptic Transistors For Neuromorphic Computing

Transistors synaptiques nano-ioniques à base de lithium pour le calcul neuromorphique

Thèse de doctorat de l'Université Paris-Saclay

École doctorale n° 575 : Electrical, optical bio : physics and engineering (EOBE)
Spécialité de doctorat : Electronique et Optoélectronique, Nano- et Microtechnologies
Graduate School : Sciences de l'ingénierie et des systèmes . Référent : CentraleSupélec

Thèse préparée dans les unités de recherche Laboratoire de Génie Électrique et Électronique de Paris (Université Paris-Saclay, CentraleSupélec, CNRS) et CEA-LETI, sous la direction de **Olivier SCHNEEGANS**, Chargé de Recherche CNRS, le co-encadrement de **Sami OUKASSI**, Ingénieur R&D CEA-LETI

Thèse soutenue à Paris-Saclay, le 19 décembre 2022, par

Ngoc Anh NGUYEN

Composition du Jury

Ahmad BSIESY Professeur, Grenoble INP - UGA	Président
Fabien ALIBART Chargé de recherche HDR, CNRS, Université Lille (IEMN)	Rapporteur & Examinateur
Marc BOCQUET Professeur, Université Aix-Marseille (IM2NP)	Rapporteur & Examinateur
Marie-Paule BESLAND Directrice de Recherche CNRS, Nantes Université (IMN)	Examinatrice
Damien QUERLIOZ Chargé de recherche HDR, CNRS, Université Paris-Saclay (C2N)	Examinateur
Olivier SCHNEEGANS Chargé de recherche HDR, CNRS, Université Paris-Saclay (GeePs)	Directeur de thèse

THESE DE DOCTORAT

NNT : 2022UPAST164

université PARIS-SACLAY

DOCTORAL SCHOOL
Electrical, Optical, Bio: Physics and Engineering (EOBE)

Titre : Transistors synaptiques nano-ioniques à base de lithium pour le calcul neuromorphique

Mots clés : nanoelectronique, calcul neuromorphique, transistors synaptiques, films minces, $LiCoO_2$, $LiTiO_2$

Résumé : En informatique, l'architecture actuelle de Von Neumann est confrontée à d'importantes difficultés dans la réalisation de tâches cognitives (de reconnaissance ou de classification d'images ou de sons par exemple). Pour surmonter cet obstacle, l'architecture neuromorphique représente une piste prometteuse vers la réalisation de traitements cognitifs performants avec une faible consommation énergétique. La conception de tels systèmes nécessite cependant le développement de synapses artificielles dont le comportement se rapproche de leurs analogues biologiques. À l'heure actuelle, de nombreuses recherches se concentrent sur des nanodispositifs spécifiques (memristors) dont la conductance électrique peut être modulée aisément afin d'émuler le comportement de liaisons synaptiques biologiques. Pour ces composants électroniques, deux configurations sont possibles (à 2 terminaux et à 3 terminaux). Parmi les synapses artificielles à 3 terminaux, les transistors ioniques apparaissent comme de bons candidats potentiels. Leur fonctionnement repose sur un empilement {canal/conducteur ionique} qui permet d'injecter/extraire des ions (via le conducteur ionique) dans la partie active du transistor (le canal), et de moduler ainsi finement la conductance électrique du composant.

Dans cette thèse, nous explorons de nouveaux types de transistors nano-ioniques pour la réalisation de synapses artificielles. Nous avons d'abord élaboré des transistors synaptiques tout-solide à l'échelle d'un wafer en utilisant des techniques de microfabrication compatibles CMOS : une première génération de composants (deux types d'empilements possibles : $LiCoO_2$/LiPON, Li_xTiO_2/LiPON) a été réalisée. Les propriétés physiques et structurales de tels transistors ont été caractérisées par différentes techniques de microscopie et de spectroscopie (MEB, MET, spectroscopie Raman). Leurs performances en termes de comportement synaptique (modulation de la conductance, stabilité des états, non-linéarité, consommation d'énergie et endurance) ont été démontrées. Une étude électrochimique systématique (focalisée sur le matériau constituant le canal du transistor) a été réalisée, afin de proposer une explication sur l'origine des performances de ces composants. À partir des résultats expérimentaux, des réseaux de calcul neuromorphique (ANNs et SNNs) ont été simulés. En particulier, un réseau de neurones artificiels (ANN : artificial neural network) composés de matrices de transistors synaptiques a été simulé et testé sur différentes tâches de reconnaissance de formes. Le comportement cognitif de conditionnement classique (expérience de Pavlov) a également été simulé, montrant l'applicabilité potentielle de nos transistors synaptiques aux réseaux de neurones à impulsions (SNNs : spiking neural networks). Enfin, diverses approches (nouveaux designs, architectures et matériaux) ont été envisagées pour améliorer encore les performances globales de nos transistors synaptiques, vers une seconde génération de composants.

Title: Lithium-based nano-ionic synaptic transistors for neuromorphic computing

Keywords: nanoelectronics, neuromorphic computing, synaptic transistors, thin films, $LiCoO_2$, $LiTiO_2$

Abstract: The present Von Neumann computing architecture faces huge problems in achieving complex tasks such as recognition and classification. To overcome such bottleneck, neuromorphic computing represents an innovative and promising architecture towards performing intelligent and energy-efficient computation. The construction of such systems requires however the development of artificial synapses with biorealistic behavior. At present, a high research interest focuses on specific devices (memristors) whose electrical conductance can be tuned to emulate the evolution of biological synaptic weights. Two different device configurations (2-terminal and 3-terminal) are being explored. Among 3-terminal artificial synapses, ion-gated transistors appear as promising candidates. They rely on using an ion conductor as a gate dielectric to intercalate or extract ions into/from the channel layer following electrochemical reactions, thus, modifying the analog conductance states of the transistor.

In this thesis, we explored novel nano-ionic transistors as synaptic components. We first managed to elaborate wafer-scale, all-solid-state synaptic transistors using CMOS-compatible microfabrication techniques: a first generation of transistors with two possible gate stacks ($LiCoO_2$/LiPON and Li_xTiO_2/LiPON) has been successfully realized. The physical and structural properties of these transistors have been characterized by different microscopy and spectroscopy techniques (SEM, TEM, Raman spectroscopy). Subsequently, synaptic behaviors such as conductance modulation, state retention, nonlinearity, energy consumption, and endurance have been demonstrated. We carried out a systematic electrochemical study (focused on the active channel material), and proposed an explanation on the performance of the synaptic transistors. Furthermore, from the experimental results, neuromorphic computing networks (ANN and SNN) have been simulated. Specifically, neural ANN cores composed of crossbar arrays including our synaptic transistors have been simulated, trained, and tested with different pattern recognition tasks. Besides, the Pavlovian conditioning experiment has been simulated, showing the potential applicability of our synaptic transistors to spiking neural networks (SNN). Finally, various approaches (new designs, architectures, and materials) have been considered to improve further the overall transistor performance, towards a second generation of synaptic transistors.

ACKNOWLEDGEMENTS

I would like to express my gratitude to all the people who have supported me throughout the three years of my Ph.D. in this part.

First and foremost, I'm deeply indebted to my supervisors, Dr. Sami Oukassi and Dr. Olivier Schneegans. Thank you, Sami, for having always spent at least five minutes of your precious time on me and my problems. Your countless pieces of advice on scientific, technical, and professional aspects were essential to completing this work and my personal growth. Olivier, your supervision style was the warmest I have ever experienced. Weekly discussions with you always brought me lots of motivation and encouragement. Your pedagogical way of teaching and limitless patience helped me a lot in positioning and presenting research findings more efficiently. Your mentorship truly inspired me.

I would also like to show my deep gratitude to Dr. Fabien Alibart and Prof. Marc Bocquet, reviewers of my thesis, for their insightful comments and questions on the scientific content and presentation of my work. I am also very grateful to the committee members: Prof. Ahmad Bsiesy, Dr. Marie-Paule Besland, and Dr. Damien Querlioz, for their constructive suggestions and criticism during the thesis defense.

I would like to extend my sincere thanks to my entire team (LSME and LCRE at Leti-CEA) for your support on different aspects of my work. I very much appreciated the practical career advice from my managers, Dr. Raphael Salot and Dr. Yann Lamy. Thank you, Sylvain and Jouhaiz, for your aid on thin-film deposition. Thanks to Severine, Valentin, Jean-Marc, Jordan, Marjolaine, and Clemence for your assistance in the microfabrication process. Special thanks to Isabelle, Denis, and Anne-Marie, who have supported me in the device and material characterization.

Last but not least, I would like to express my deepest gratitude to my family in Vietnam, especially my mother, Pham Thi Hoa. Her daily care messages motivated me to work harder and be more resilient in this foreign country. I cannot end this part without saying thank you to my girlfriend, Nhat Anh, for her invaluable love and support. Thanks for always believing in me, for the hours of confiding, and the meals with the taste of home.

TABLE OF CONTENTS

INTRODUCTION 8

CHAPTER 1: **Bibliographical study** 11

CHAPTER 2: **Microfabrication techniques of synaptic transistors** 53

CHAPTER 3: **First generation of electrochemical synaptic transistors** 97

CHAPTER 4: **Simulation of neuromorphic computing systems composed of our Li_xTiO_2-based transistors** 135

CHAPTER 5: **Optimization towards a second generation of synaptic transistors** 169

CONCLUSIONS AND PERSPECTIVES 187

Summary of the thesis in French (5 pages) 191

Publications during the thesis 197

INTRODUCTION

The conventional von Neumann computer plays a critical role in solving problems with different levels of complexity in almost every field of life. However, this computing architecture presents a bottleneck: a significant amount of time and energy are required to transmit data between processors and memory units. This barrier will inevitably limit computational efficiency, especially for solving data-intensive tasks such as pattern recognition, real-time speech, and visual computing. Taking the inspiration from the human brain, neuromorphic computing systems are expected to mitigate the aforementioned limit by performing the computations on the structured memory arrays with massive parallelism.

Neuromorphic computing systems are composed of neural networks (representing neurons) connected by artificial synapses. Developing brain-like computers requires artificial synapses that mimic the behaviors of their biological counterparts. Efforts have been made to simulate synaptic functions with CMOS analog circuits. However, they face a critical challenge in large-scale integrations, as tens of components are required to mimic one synapse. This mismatch in efficiency necessitates the search for single electronic devices that can emulate the synaptic functions. In recent years, specific devices (memristors) whose electrical conductance can be tuned to emulate the evolution of biological synaptic weights, have attracted much interest. These components can be subdivided in 2-terminal devices and 3-terminal devices. Each configuration has its own strengths and weaknesses. Specifically, in the 3-terminal configuration, the Write operation (synaptic weight modulation) is decoupled from the Read operation, allowing a better control of conductance tuning.

Among 3-terminal artificial synapses, electrochemical synaptic transistors (whose working principle is similar to that of biological synapses) appear as promising candidates. The electrical conductance of the channel is modified upon the intercalation of ions following redox reactions controlled by the gate potential, creating a robust mode of analog weight programming.

The large-scale integration of the electrochemical synaptic devices at wafer-scale fabrication is inexorable to develop the hardware for neuromorphic computing. There exist devices composed of liquid and polymer electrolytes and manually exfoliated channels, which prohibits however their further integration into dense computing chips. To overcome these drawbacks, the goals of my thesis are to (i) propose innovative solid-state synaptic transistors to optimize the overall performance (ii) elaborate and

characterize such solid-state electrochemical synaptic transistors with CMOS compatible processes, and (iii) demonstrate their required synaptic functionalities. The work in this thesis has been realized with the facilities of Laboratoire Composants pour la RF et l'Energie (LCRE), CEA-LETI, Grenoble and the Laboratoire Génie Électrique et Électronique de Paris (GeePs), Gif sur Yvette.

This dissertation is divided into 5 chapters. In the first chapter, we present a brief review of neuromorphic computing, its motivation, and its emerging artificial synapse solutions with a focus on electrochemical synaptic transistors. A global view of state-of-the-art electrochemical devices and their reported performance parameters is also presented.

In chapter 2, we will introduce the processes in microfabrication, including thin-film deposition, patterning, and characterization techniques employed in elaborating the electrochemical synaptic transistors in this thesis. The detailed process flow to realize the synaptic transistors based on thin films with the progressive optimization progress is subsequently discussed.

Chapter 3 is dedicated to showing the first functional generation of electrochemical synaptic transistors with two gate stacks: $LiCoO_2$/LiPON and Li_xTiO_2/LiPON. We will first present briefly the properties of the materials composing the transistors. We then investigate the electrochemical and electrical properties of $LiCoO_2$/LiPON devices with some preliminary tests. It is followed by a systematic study of Li_xTiO_2/LiPON systems in which we could demonstrate excellent synaptic behaviors of the electrochemical transistors and their correlation to the electrochemical phenomena.

With the results obtained, we initiate in chapter4 the simulation of neuromorphic computing systems which take into account the properties of our synaptic transistors. An Artificial Neural Network (ANN) with analog synaptic transistor crossbar arrays, considering the nonlinearities of realistic devices, is simulated, trained, and tested with the MNIST pattern recognition task. A benchmark among the available artificial synapses is given. In addition, the Pavlovian conditioning experiment is examined with a neural electronic circuit, which includes our synaptic transistor as a learning element: this allows showing the potential applicability of our transistors to spiking neural networks (SNN).

In chapter 5, we will discuss measures to improve further the electrical performance of the synaptic transistors. Feasible approaches come from shrinking both lateral and vertical dimensions of the transistors to facilitate fast, energy-efficient programming pulses and linear, stable analog states. Besides, new ultrathin materials will be considered as an important axe for the next steps.

CHAPTER I

BIBLIOGRAPHICAL STUDY

ABSTRACT

The first chapter will be dedicated to introducing the field of neuromorphic computing and neuromorphic hardware. This chapter is divided into 5 main sections. In the first section, we will present the urge to search for and develop a new computing architecture that can help digest the accumulated immense amount of data generated in an energy-efficient way.

Section 2 of this chapter introduces a bio-inspired solution - neuromorphic computing. We will briefly cover the functions of biological synapses and subsequently describe the methods implemented to mimic these brain functionalities using electronic components. With the main arguments of energy and space consumption, we argue the need for emerging synaptic devices to construct novel neuromorphic computing structures.

In section 3, we address the types of available electronic solutions for artificial synapses, including two-terminal devices (Resistive random-access memory – ReRAM, Phase-change memory – PCM, and Magnetic random-access memory – MRAM) and three-terminal devices (Ferroelectric field-effect transistor – FeFET and Electrochemical synaptic transistor – SynT). The operation principles of each type of device will be discussed.

With the similarity to the biological synapses, SynT will be the focus of this thesis and will be explicitly presented in section 4. We first summarize recent works and the trends in the field of electrochemical synaptic transistors. Subsequently, the SynT figures of merit are discussed with methods and examples demonstrated on realistic synaptic transistors, paving the way to systematically characterize future SynTs. In section 5, we will comment on the overall state-of-the-art of the artificial synapses with the input from SynTs and other candidates, and main approaches to improve the performance of these devices. From there, this thesis's outline and scope are presented in section 6.

Table of Contents

1 THE NEED FOR A NEW COMPUTING PARADIGM 14
2 NEUROMORPHIC COMPUTING 17
2.1 Biological synapses 17
2.2 Neuromorphic computing – the brain emulation using hardware 19
3 THE ARTIFICIAL SYNAPSES 21
3.1 Two-terminal memristive devices 21
3.2 Three-terminal synaptic transistors 22
3.2.1 Ferroelectric field-effect transistor (FeFET) 23
3.2.2 Electrochemical Synaptic Transistor (SynT) 24
4 STATE OF THE ART IN ELECTROCHEMICAL SYNAPTIC TRANSISTORS 27
4.1 Summary on the recent works 27
4.2 Performance parameters of SynTs as artificial synapses 32
4.2.1 Conductance modulation 32
4.2.2 Analog states modulation 35
4.2.3 Write nonlinearity 37
4.2.4 State retention 40
4.2.5 Energy consumption 41
4.2.6 Endurance 43
4.2.7 Temporal dynamics for SNN applications 44
5 CONCLUSIONS 46
6 REFERENCES 49

1 THE NEED FOR A NEW COMPUTING PARADIGM

Ever since being conceptualized, Artificial Intelligence (AI) has revolutionized the world we are living in a way that no one could imagine. The ability to refine and analyze an enormous amount of datasets and then turn them into fantastic use cases of data that were not previously obvious to the naked eye makes it a genuine game-changer in this "Big Data" era. More and more data is being processed in various fields, such as health, business, and transportation, helping firms make critical data-driven decisions. However, such kind of cognitive operations of AI requires remarkable computing resources, owing to the current constraints posed by the conventional computing hardware platform, namely the termination of Moore's law and the Von Neumann (VN) bottleneck [1].

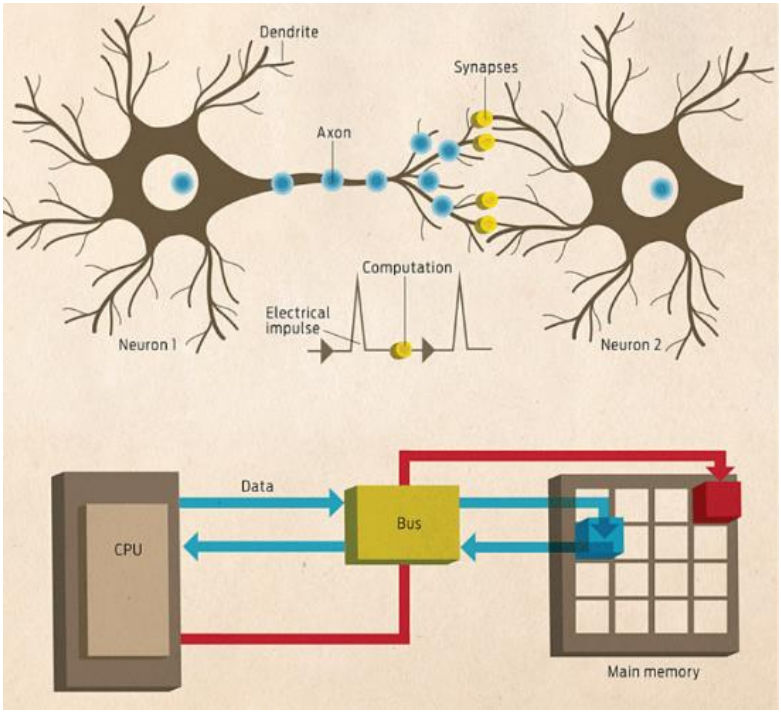

Figure I. 1: Brain versus Computer. (Top) Neurons in human brain are connected via billions of neural connections – synapses, creating a biological computer with energy- and time-efficient computing ability. (Bottom) Conventional computer structure with memory and processing units separated. This structure requires data to be shuttled between the two units and thus, the data latency is unavoidable [2].

Over the past few decades, together with the rapid evolution of micro-fabrication technology, computer chips' performances have been pushed to the edge. They could no

longer be appropriately described by the prediction of Moore in 1965 [3]. As the miniaturization continues, the transistors will eventually reach their atomic limits, which results in critical problems concerning current leakage, overheating, and thus, limiting the processing power of the nano-chips. For these reasons, it is complicated and extremely costly to ensure the ideal functionality across billions of devices. On the other hand, VN architecture implies the usage of a shared data bus between program and data memories, therefore, prohibiting the parallelism of instruction processing. Since the central processing unit (CPU) speed and memory size have increased much faster than the throughput between them, the bottleneck has become an increasingly severe problem with every new generation of CPU. As a result, to accelerate the development of AI, extensive research efforts have been made to optimize an alternative brain-inspired computing paradigm that is both energy and time-efficient – the neuromorphic computing systems (Figure I. 1).

The human brain exhibits an appealing Non-Von Neumann (Non-VN) computing paradigm for future computing systems that can complement or even replace the current VN architecture. Characterized by its massively parallel architecture connecting numerous low-power computing elements (neurons) and adaptive memory elements (synapses), the brain can outperform modern processors on many cognitive tasks involving unstructured data classification and pattern recognition. Neuromorphic computing is an active multidisciplinary branch of research, combining physics, nanoelectronics, nanomaterials, neuroscience, and computer science. It serves the purpose of building and developing the materials, devices, and systems that closely mimic the human brain, targeting a new generation computing system that inherits the parallelism and low power operation of the mammalian nervous system.

By far, there are two main approaches to realizing brain-like computing, namely software simulation and hardware implementation. However, current software simulation often requires a huge amount of energy consumption and great physical space. For example, IBM's Blue Gene supercomputer is used for executing software simulations consuming roughly 10 MW of power [4]. Moreover, the software in digital computation performs calculations in series, and for these reasons, it is not suitable to mimic the parallelism of neural networks effectively. Such problems could be addressed if we can actualize a massively parallel neural network at a hardware level [5]. Hardware implementation aims to construct artificial neuron networks by using electronic devices. Since synapses are the functional connections of neurons and serve as the basic units of computing and learning, designing physical synaptic devices that exhibit synaptic behaviors is the key step to build brain-like computers. To achieve this, artificial neural networks (ANNs) have been developed and successfully applied in various fields. Despite the fact that these recent favorable outcomes in neuromorphic computing [6]–[8], the hardware implementation of these ANNs has been hindered by the fact that multiple CMOS transistors are required to mimic the behaviors of

the analog synapses. Furthermore, their energy consumption is much higher than the biological synapses to perform the same tasks [9]. Non-volatile electronic memory devices, including memristors and synaptic transistors, are rising as promising technologies to complement the conventional Si CMOS systems with a better energy and space tradeoff.

2 NEUROMORPHIC COMPUTING

2.1 Biological synapses

The human brain is by far the most efficient computing system, which is not very surprising as it is the result of millions of years of evolution. The brain combines various types of cells, but the primary functional unit is a cell called a neuron. These neurons are responsible for generating and analyzing signals that control our emotions, memories, movements, thinking, and feelings; these are the traits that make us humans. A human brain contains approximately 10^{11} neurons [10]. Each of them is made up of a cell body called soma, an axon, and multiple dendrites. Axon carries information from soma to a junction, where it is collected by dendrites of other neurons. The intersection is called a synapse, and the strength of the synapse (synaptic weight) decides the connection strength between two neurons, which can be altered by neural activities. This process is known as synaptic plasticity and is believed to be the backbone of human learning ability.

The synaptic inputs picked up by the dendrites of other neurons are then integrated by their own cells, encoded in the form of action potentials, and distributed to even more neurons from their axon terminals. Typically, each neuron is connected with about 10^4 other neurons, resulting in a large number (about 10^{15}) of biological synapses in the human brain. With its high energy efficiency, the consumes averagely around 20 W of power for its functioning [11], yielding an energy consumption of approximately 1–10 fJ per synaptic event.

Various neuron interactions should be taken into consideration to understand how a neuron works. Each neuron carries out five essential functions as described in Figure I. 2.a: (1) Generate intrinsic membrane activity in the neuron; (2) Receive synaptic inputs in dendrites; (3) Combine synaptic inputs with the intrinsic membrane activity; (4) Generate outputs in the form of action potentials; (5) Distribute the outputs from axon terminals [12].

There are two fundamental types of synapses: electrical and chemical. Electrical synapses mainly exist in invertebrates – animals lacking a backbone, while chemical synapses are found in humans and other vertebrates. We will only focus on the chemical synapses in this study. The chemical synapse between two neurons is illustrated in Figure I. 2.b. At these synapses, information transfer from one neuron to another occurs through the release of neurotransmitters by one neuron (pre-synaptic neuron) and the detection of the neurotransmitters by an adjacent neuron (post-synaptic neuron).

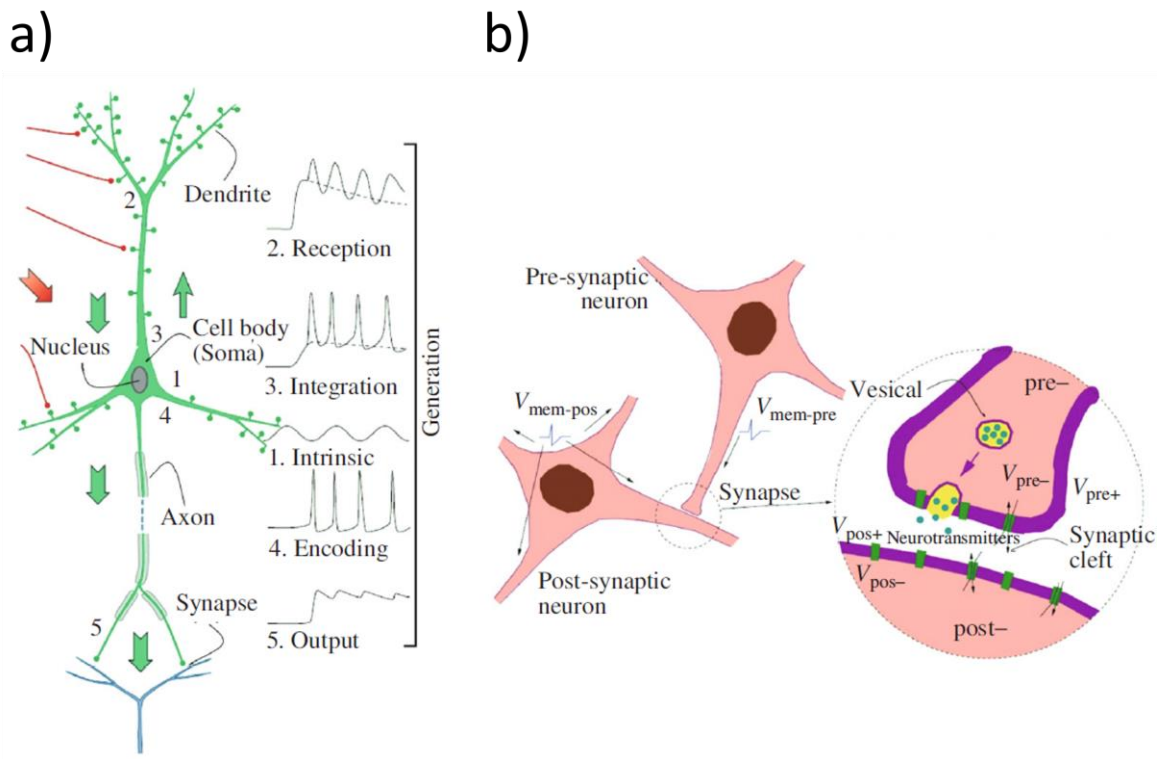

Figure I. 2: Schematic of neurons and chemical synapses. a. Different functional regions of a neuron. b. A synaptic connection between a pre-synaptic neuron and a post-synaptic neuron [13].

The synaptic plasticity shown by chemical synapses is one of the most important neurochemical foundations of learning and memory in the brain. Chemical synapses consist of two types: excitatory or inhibitory [13]. In the brain, the synapse receptors for glutamate neurotransmitters are typically excitatory, whereas the receptors for GABA (Gamma-Aminobutyric acid) neurotransmitters are generally inhibitory. An excitatory receptor results in an Excitatory Postsynaptic Potential (EPSP) and drives the postsynaptic neuron closer to the depolarization threshold, which makes the cell "fire" an action potential. An inhibitory receptor results in an Inhibitory Postsynaptic Potential (IPSP) and drives the postsynaptic neuron further from the depolarization threshold. Axon terminals from many neurons can connect to a given neuron and release a variety of neurotransmitters, which impinge on excitatory and inhibitory receptors to produce EPSPs and IPSPs.

The postsynaptic neuron behaves like a tiny computer, integrating all the EPSPs and IPSPs, which later determines whether it will "fire" or not. When a neuron fires, the resulting action potential travels towards the synaptic cleft through its axon. The arrival of the action potential at the axon terminal results in the merging of neurotransmitter vesicles with the presynaptic membrane, and a subsequent release of the neurotransmitters into the synaptic cleft. The neurotransmitter diffuses through the synaptic cleft, binds to and activates a

receptor in the postsynaptic membrane, modifying the plasticity of the connection, i.e., the synaptic weight. In the learning phase, these synaptic weights are updated in an analog and parallel fashion based on multiple learning rules [13].

2.2 Neuromorphic computing – the brain emulation using hardware

Modern computers are more capable of the enormous amount of information storage as well as fast numerical computation than that of human brains. Still, even the biggest and fastest supercomputers in the world cannot match the overall processing power of the human mind in performing cognitive and adaptive tasks, such as pattern recognition, perception, motor control, flexibility, and learning [14].

Due to these fantastic capabilities of the brain, it is very appealing to study its structure and working mechanisms in order to mimic it using electronic circuits. The study of the brain and its inspiration in developing computing systems has led to a new form of computer architecture, known as neuromorphic architecture. This field combines knowledge from multi-disciplines in order to design artificial neural systems. The physical architecture, design principles, and computing algorithms of artificial systems are based on those of biological systems (Figure I. 3) [15], [16].

Although the behavior and connections between neurons can be partially simulated on a Von Neumann-architecture computer, such a system will consume excessive power, for example "MilkyWay-2" supercomputer [17] consumes a normal power of 20 MW compared to 20 W ultralow consumption of the human brain on real time processing tasks [18]. Furthermore, such a computing system is not capable of exploiting the architecture of the brain due to the fundamental differences between these two systems. As a result, a race is on to develop new types of devices and hardware architectures that can better resemble bio-intelligent systems at the physical level, and thus more efficiently emulate the brain at the functional level. For instance, the so-called neuromorphic circuit is built from devices that behave like neurons, transmitting and responding to information sent in the form of spikes rather than continuously varying voltages. Carver Mead, dating back to 1980s, first developed the concept of neuromorphic engineering. Mead described it as "using VLSI systems containing electronic analog circuits to mimic neurobiological architectures present in the nervous system [19]."

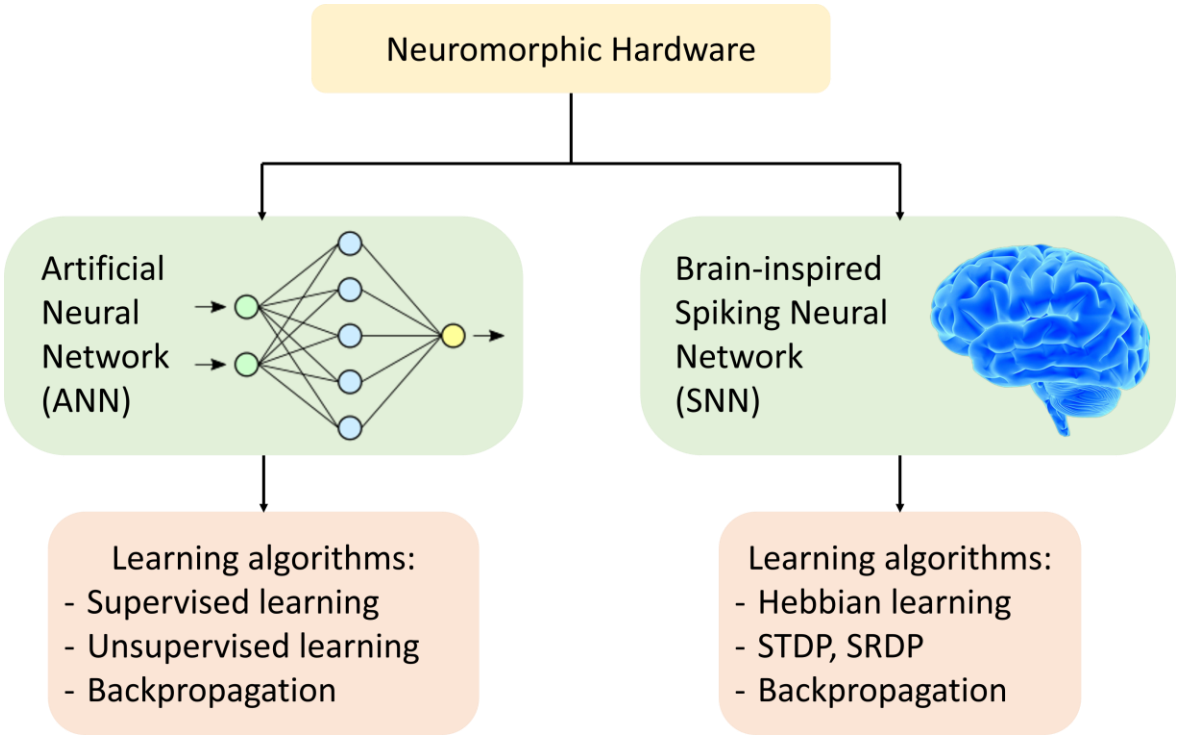

Figure I. 3: Neuromorphic hardware can be implemented via both artificial neural networks (ANNs) and brain-inspired neural networks. These two approaches rely on a different set of algorithms and learning rules, e.g. the backpropagation algorithm in ANNs, or the temporal dynamics in brain-inspired SNN concepts.

As the brain consumes extremely low energy, the first crucial step in realizing these hardware systems is to achieve a suitable device that can function as a synapse with low power consumption and desired plasticity. Accordingly, researchers have explored a variety of device systems with programmable conductance, also known as synaptic devices.

3 THE ARTIFICIAL SYNAPSES

In 1971, Chua first proposed the concept of memristor (memristive device) [20], which was defined as a two-terminal system with resistive switching (RS) effect that could demonstrate a variable and non-volatile resistance depending on the history of applied voltage and current. Later in 2008, the HP lab realized a device that has characteristics that described by Chua [21]. With this progress, huge research attention was dragged into developing and optimizing neuromorphic computing systems with memristors as the primary functional units [22]–[24] as well as the post-CMOS memory components themselves. Over the years, many types of memristive devices were developed for the artificial synapse application, including two-terminal devices (e.g. resistive random-access memory (ReRAM) [25], [26], phase-change memory (PCM) [27], magnetoresistive random-access memory (MRAM) [28]), and three-terminal devices (e.g. ferroelectric field-effect transistor (FeFET) [29], and electrochemical synaptic transistors (SynT) [30], [31]).

3.1 Two-terminal memristive devices

Memristors usually have a simple metal/insulator/metal structure (see Figure I. 4) [32]. ReRAM memristors are two-terminal resistance switches that can retain their internal resistance states depending on the history of applied voltages/currents [33]. Based on the filament (conducting bridge) rupture mechanism, there are two types of memristive devices, namely, drift and diffusive memristors. PCM devices rely on the resistivity difference between two phases of a chalcogenide material (phase change material): the crystalline phase with low electrical resistivity and the amorphous phase with high electrical resistivity [34]. PCM devices are currently among the most mature devices for artificial synapses [35].

Spintronics-based MRAM consists of a tunneling oxide layer sandwiched by two metallic ferromagnetic layers: the free layer and the pinned layer. This structure is also called a magnetic tunnel junction (MTJ) [36]. The spin polarization of the pinned layer is fixed in a specific direction, while the free layer magnetization can be altered using an external current or magnetic field. Depending on whether the magnetization directions in the two layers are parallel or not, devices can exhibit a high resistance (opposite or antiparallel) state or a low resistance (parallel) state. The artificial synapses built on two-terminal devices have several advantages, such as low power consumption, simple device structure, small cell size, and easy large-scale integration with crossbar structure. However, the device variability and operation instability of the two-terminal synaptic devices may hinder their further applications in advanced artificial intelligent systems [37].

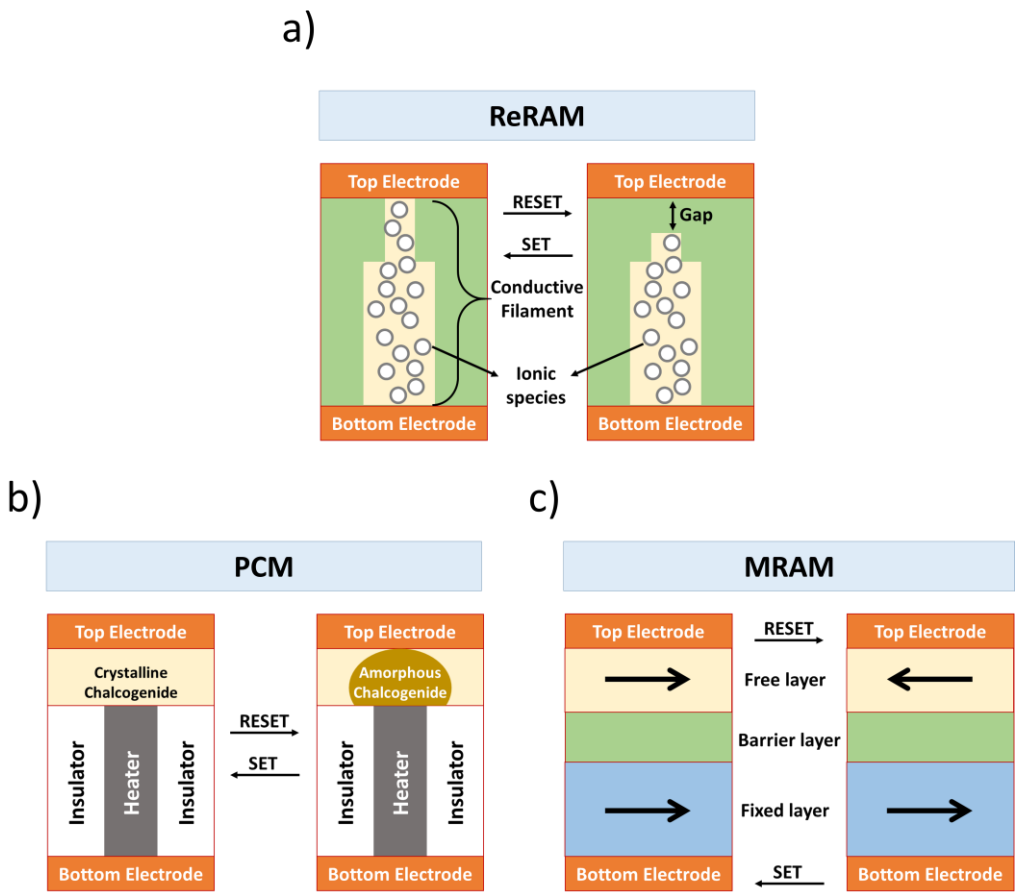

Figure I. 4: Memristor devices. a) Resistive random access memory (ReRAM). b) Phase change memory (PCM). c) Magnetoresistive random access memory (MRAM).

3.2 Three-terminal synaptic transistors

By adding a third terminal to these devices, quite a several problems are mitigated. In comparison with two-terminal synaptic devices, three or multi-terminal synaptic transistors have the advantages of excellent stability, relatively controllable testing parameters, and robust operation mechanism [38]. Through proper material selection and structural design, transistors can convert external stimuli (light, pressure, temperature, etc.) into the electrical signal, which provides the possibility to achieve artificial synapses that can directly respond to the external environment. In addition, synergistic control of one device can be easily implemented in a transistor-based artificial synapse, which opens up the possibility of developing stable neuron networks with significantly fewer neural elements. More importantly, signal transmission and self-learning can be performed simultaneously in multi-terminal transistor-based artificial synapses. Therefore, transistors may be more suitable for simulating synaptic functions than other types of devices, especially for simulating concurrent

learning and dendrites integration that require a multi-terminal operation. In this section, the working mechanisms of the main synaptic transistor types, FeFET and SynT, are discussed with the focus on the central theme of the thesis as electrolyte-gated electrochemical synaptic transistors. A State-of-the-art review will follow to highlight the trends in this topic.

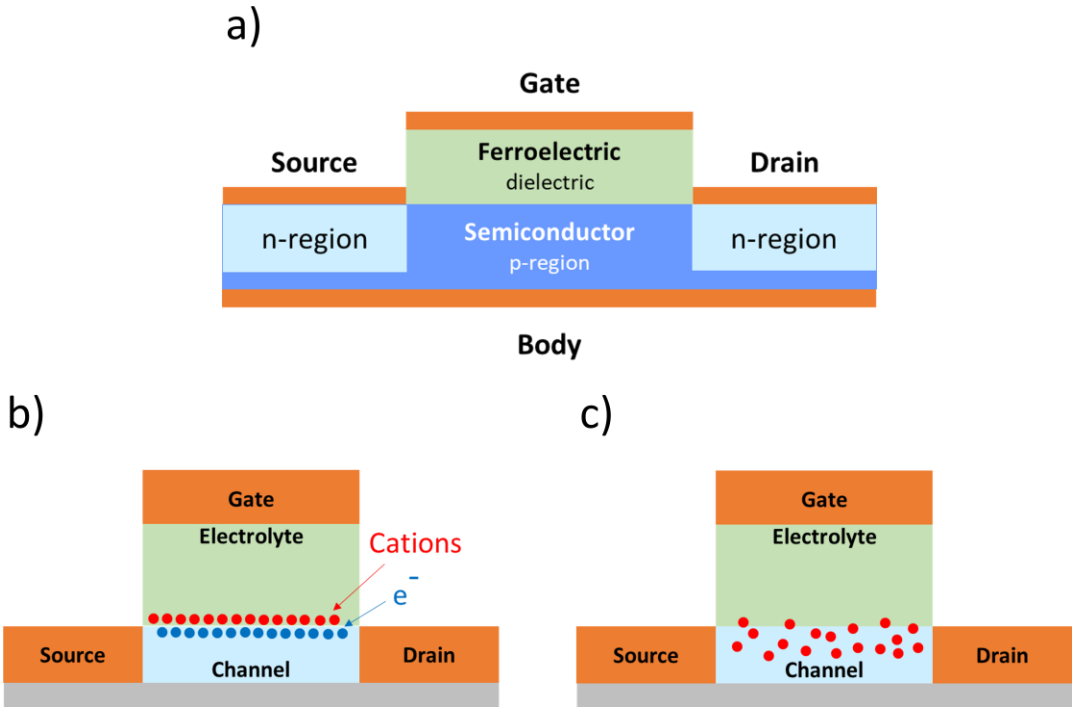

Figure I. 5: Synaptic transistors: a) Ferroelectric field-effect transistor (FeFET). b) Electric double-layer electrochemical transistor. c) Ion intercalation electrochemical transistor [39].

3.2.1 Ferroelectric field-effect transistor (FeFET)

FeFETs have been actively studied for practical nonvolatile memory applications due to their non-destructive readout, low power consumption, and high operating speed. The ferroelectric insulator layer, whose polarization states are spontaneous, is the main component of this technology. Upon gating, the carrier concentration of FeFETs can be precisely and gradually modulated by changing the polarization state of ferroelectric materials [29]. For traditional memory applications, the ferroelectric insulator switches between two polarization states representing two digital states of the memory. Recently, FeFETs have attracted considerable attention as a promising platform for mimicking biological synapses thanks to the excellent multi-domain polarization switching capability of ferroelectric materials, which can be used to obtain multilevel FeFET channel conductance.

The multilevel channel conductance of FeFETs can be utilized to record synaptic weight. A general schematic of a FeFET is shown in Figure I. 5. [40].

Ferroelectric synaptic transistors show some promising features, such as high stability, large ON/OFF ratio, fast programming operations, as well as fewer variations in the weight update curve [41]. However, they also suffer from scaling issues as floating-gate synaptic transistors because of their similar charge-based memory characteristics. In addition, they have difficulties implementing excellent short-term synaptic plasticity. Therefore, further research focused on addressing these limitations of FeFET-based synaptic devices is urgent.

3.2.2 Electrochemical Synaptic Transistor (SynT)

An additional gate terminal and an ion-conducting electrolyte layer surrounding the channel material constitute the gate-stack of an electrochemical synaptic transistor. The ions inside the electrolyte could be driven toward and even into the channel material, leading to its conductance change. Such ionic dynamics much resemble the pre-synapse process emits synaptic transmitters, which then move across the synaptic cleft, pass the ion channels on the post-synapse, and finally enhance the post-synapse signal. There are two types of working mechanisms in SynTs: electric double-layer formation and ion intercalation.

3.2.2.1 Electric double-layer formation

A constant gate voltage or a low-frequency gate pulse would drive the ions with opposite charges, e.g., cations, toward the channel material surface, leading to a cation accumulation. According to Electrostatics, such an accumulation would also call for a thin layer of electrons inside the channel material, near the material surface Figure I. 5.b. The cations and electron layer would form an ultra-thin capacitor, which is named the electric double-layer capacitor (EDLC). The ion-induced electron layer also functions as a conductive layer, which would significantly improve the conductance of the channel material, leading to the resistance change of the channel material. By engineering the intensity of the gate pulse, the ion density, and the type of ion inside the electrolyte, the formation of the EDL could be changed, which provides an approach to tuning the temporal behaviors of the channel conductance.

EDL-based devices possess many advantages compared to other mainstream mechanisms. The most important one is the extremely large electric field formed between the double layers, which can easily be more than 10 MV/cm due to the extremely short distance

between the two electrical layers. Such a strong electric field can hardly be achieved in conventional solid-state capacitors due to the dielectric breakdown or tunneling. It can lead to better gate control in artificial neuromorphic devices. In other words, the device can work at lower operation voltage compared to conventional devices. Moreover, in most cases, the electrolyte is only conductive to specific types of ions and is insulating to electrons. Hence, there only exists a very low leakage current between the resistive switching material and the controlling terminal. As a result, devices based on EDL formation can often work at a voltage lower than 0.3 V and possess energy consumption as low as 1.23 fJ/spike [33]. Moreover, after the controlling voltage is removed, the ions accumulated on the resistive material surface spontaneously diffuse back into the electrolyte and are re-distributed uniformly from a few milliseconds to a few seconds. This phenomenon leads to a significant current drop from the stimulated state to the original state, which can be read out by the source and drain terminals and greatly resembles the short-term plasticity inside biological systems.

3.2.2.2 Ion intercalation

Another phenomenon that takes place inside electrochemical transistor systems is ion intercalation. The mobile ions in the electrolyte would migrate into the target material under the influence of the gate voltage following electrochemical redox reactions. As a result, the electrical property of the channel layer changes, see Figure I. 5.c. The intercalation effect has been widely studied in electrode materials used for battery cells, with $LiCoO_2$ and graphene nanosheets being the few types that have been thoroughly studied. Hence, for some layered structure materials, such as WSe_2, MoS_2, graphene, and some sub-stoichiometric salts, ions can get into the crystal structure and get stored inside the channel materials. Recent research studies have showed that intercalations inside neuromorphic devices are of two main types, namely electronic redistribution [30], [42] and phase transition of channel materials [43] induced by intercalation of external mobile ions.

Unlike the devices based on EDL, the intercalated ions could remain inside the channel materials even after the controlling voltage is removed, leading to a constant memory effect. By repeatedly applying voltage pulses to the gate relative to the open-circuit voltage (OCV) of the vertical cell, relatively the same amount of ions would be intercalated into the channel material, leading to a linear increase in channel conductance. The OCV measured from Gate-Source electrodes indicates the ionic concentration of the mobile ions inside the channel. Therefore, by controlling the gradual increase of this quantity, one can monitor the conductance change of the channel with precision. With the amount of inserted ions being controlled to the minimum, one can program analog conductance states with a large number of conductance states and small energy consumption of femto-Joule or less per WRITE action,

which is comparable to the biological synapses' operations. These features of ion intercalation SynTs indicate a great fit for bio-inspired computing applications such as artificial neural network (ANN), in which these transistors are the constituting elements.

Furthermore, desirable synaptic functions have been demonstrated with ion intercalation types of electrochemical synaptic transistors, such as short-term plasticity (LTP), long-term plasticity (LTP), spiking-time dependent plasticity (STDP), and spiking-rate dependent plasticity (SRDP) [31], [44]. The listed functions are the requirements for the next generation of the neural networks inspired by the biological nervous system, spiking neural network (SNN). They employ spiking neurons as computational units that process information with the timing of spikes. Therefore, SNNs provide the potential for spatiotemporal information processing with high time and energy efficiency.

4 STATE OF THE ART IN ELECTROCHEMICAL SYNAPTIC TRANSISTORS

4.1 Summary on the recent works

Electrochemical synaptic transistors have attracted much attention from researchers as a great candidate for artificial synapses, and this device will be my main subject throughout this thesis. Two types of electrolytes are used in these devices: ionic liquid (IL) and solid-state electrolytes. ILs have been widely studied for the use of electrolyte gating to induce extremely large modulations in the carrier densities at the interfaces and also a faster switching speed compared to solid-state electrolytes [45]. However, this electrolyte's main drawbacks are environmental dependence, e.g., humidity and unscalability [31]. In fact, from a technological point of view, it is challenging to incorporate liquid phase and environmental factors in the device fabrication and encapsulation, limiting the high-density integration of the devices. Therefore, an all-solid-state electrolyte-gated transistor has a great potential for becoming the best candidate for a synaptic transistor with the application of neuromorphic computing. In this section, I will present some selected progress in the field of synaptic transistors. Table 1 will sum up the characterizations of the devices.

Among the current ions used for modulating the synaptic strength in the electrolyte-gated transistors, such as Oxygen O^{2-} and protons H^{+}, Lithium ions Li^{+} show their merits in such tasks. Lithium-ion-based devices possess high reversibility and ultra-high stability of Li^{+} ion under electrolyte gating. Fuller and coworkers demonstrated the modulation of the conductivity of the $LiCoO_2$ channel by intercalating (de-intercalating) Li^{+} via LiPON electrolyte in their Li-ion synaptic transistor for analog computation (LISTA) (see Figure I. 6.a) [30]. $LiCoO_2$ is well-known for its reliability and endurance in electrochemical cycling. The removal of Li oxidizes Co^{3+} to Co^{4+} and generates positively charged polarons. As the fraction x in $Li_{1-x}CoO_2$ is varied from 0 to 0.5, the material undergoes an insulator to metal transition, and this process is highly reversible. The LiPON electrolyte was chosen for its scalability, approximately down to 20nm in thickness, high ionic mobility (1.2×10^{-6} S/cm), and low electronic conductivity (8×10^{-14} S/cm) at room temperature. This pair of materials is also well studied in the field of all-solid-state micro-batteries. The electrical measurements of LISTA have shown promising characteristics of a synaptic transistor. It possesses a very dynamic conductance range in the order of micro Siemens, from 180 – 220 µS, with 200 nonvolatile states on average. In addition, high linearity, good endurance and low noise weight updates have been demonstrated. Achieving all this at a small cost of a few femto-Joule or even less

has been very promising and a driving factor toward brain-inspired energy-efficient neuromorphic computing.

Yang and *et al*. exploited the stacked structure of α-MoO_3 to illustrate short/long-term plasticity (STP/LTP) with very low channel conductance (Figure I. 6.b) [31]. α-MoO_3 is a layered 2D material allowing the reversible intercalation of Lithium via a Faradaic reaction involving the reduction/oxidation of Mo ions. This process tunes the oxide electrical properties with minimal structural changes compared to that of the filament formation of memristors. The short and long-term plasticity, bidirectional analog, and near-symmetric weight update between LTP and LTD have been shown with ultralow channel conductance values (<75 nS) and picojoules operation energy. These traits are essential for high-energy efficiency. Another promising technology involving 2D van der Waals layered crystals, or quasi-2D transition metal-oxide whose properties can be tuned with ion gating has been shown to exhibit both STP and LTP owning to high-frequency gating. This phenomenon is accounted by the lithium-ion intercalation into the channel that yields a nonvolatile state before they can diffuse back into the electrolyte (volatile behavior).

At the device level, synapses are still implemented by dozens of digital complementary metal-oxide-semiconductor (CMOS) devices in today's artificial neural networks. Thus, developing synaptic elements with CMOS-friendly materials is highly desirable. Tang *et al*. presented an Electro-Chemical Random-Access Memory (ECRAM) employing the reversible Li-ion intercalation in the Tungsten oxide WO_3 channel via LiPON solid-state electrolyte (Figure I. 6.c) [46]. This nonvolatile transistor can have more than 1000 stable conductance levels, prolonged retention of programmed states, and high-speed writing at 5 ns. All of these characteristics are notably demonstrated at under 100 fJ. Li and colleagues show a high-performance, α-Nb_2O_5-based single transistors and system (Figure I.6.d) [44]. The devices exhibit quasi-linear update, good endurance (10^6) and retention, a high switching speed of 100 ns, ultralow readout conductance (<100 nS), and ultralow areal switching energy density (20 fJ/μm^2). The prominent analog switching performance is leveraged for hardware implementation of an SNN with the capability of spatiotemporal information processing, where spike sequences with different timings are able to be efficiently learned and recognized by a 32x32 crossbar array.

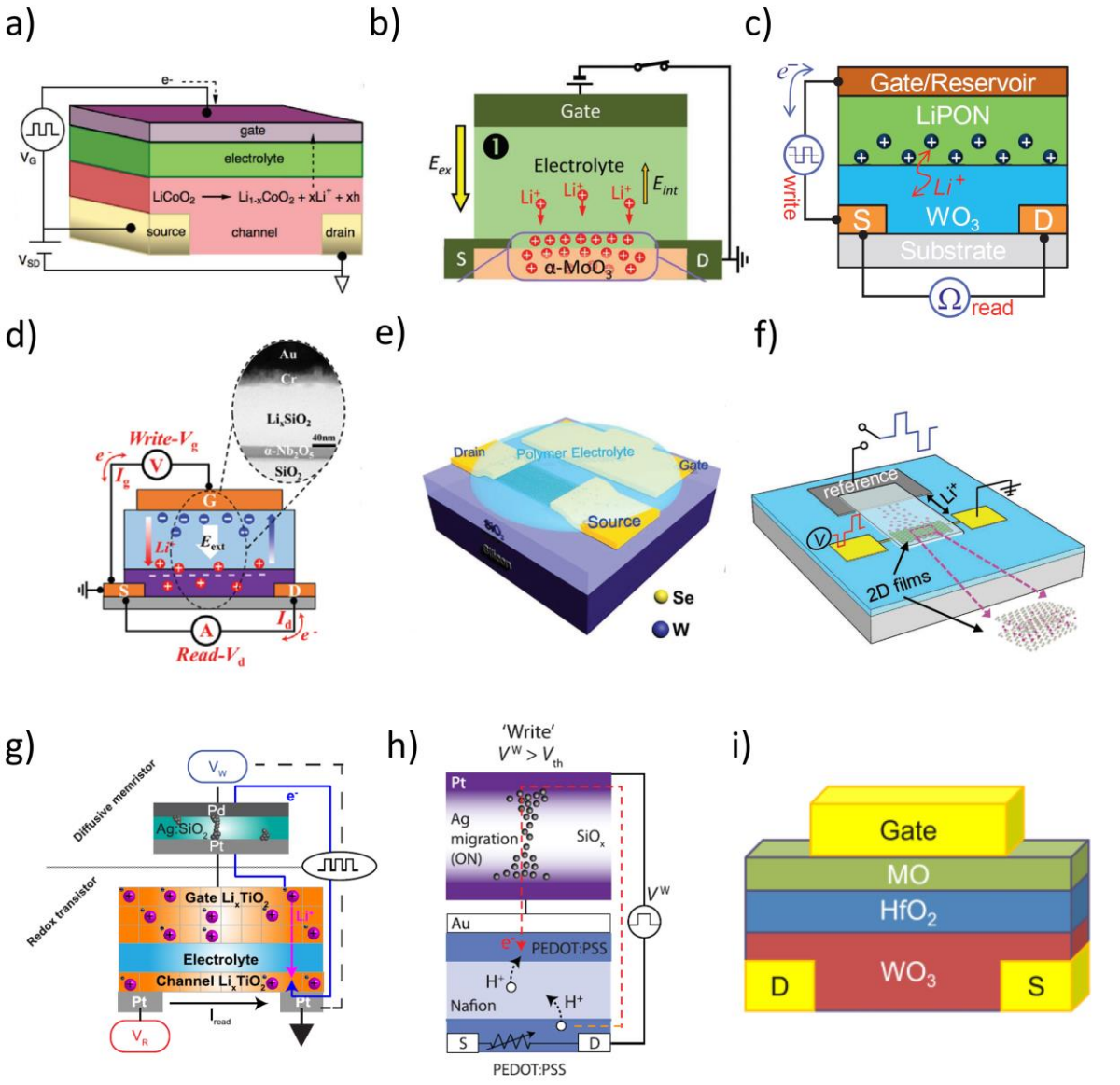

Figure I. 6: Recent electrochemical synaptic transistors. a) E. J. Fuller's SynT in 2017, composing of $LiCoO_2$ channel and LiPON electrolyte films [30]. b) C-S Yang's SynT in 2018, composing of α-MoO_3 channel and PEO:$LiClO_4$ polymer electrolyte films [31]. c) J. Tang's SynT in 2018, composing of WO_3 channel and LiPON electrolyte films [46]. d) Li's SynT in 2020, composing of α-Nb_2O_5 channel and $LixSiO_2$ electrolyte films [44]. e) J. Zhu's SynT in 2018, composing of 2D WSe_2 channel and PEO:$LiClO_4$ polymer electrolyte films [47]. f) M. T. Sharbati's work in 2017, composing of 2D Graphene channel and PEO:$LiClO_4$ polymer electrolyte films [48]. g) Y. Li's 1S1T in 2019, composing of a memristive Pd/Ag:SiO_2/Pt selector, anatase-phase LixTiO2 channel and PEO:$LiClO_4$ polymer electrolyte films [49]. h) E. J. Fuller's 1S1T in 2019, composing of a CBM Pt/Ag/SiO_xN_y/Ag/Pt selector, PEDOT:PSS polymer channel and Nafion polymer electrolyte films [50]. i) . S. Kim's selectorless SynT in 2019, composing of WO_3 channel and HfO_2 electrolyte films [51].

Novel approaches involving two-dimensional (2D) materials have demonstrated high performance of synaptic functions based on the exotic physics of these layered materials and the potential for further scaling for synaptic devices. Zhu *et al*. presented an ionic-gating-modulated synaptic transistor in which the channel is made-up of a 2D van der Waals layered crystal such as WSe_2, $NiPS_3$, and $FePSe_3$ (Figure I. 6.e) [47]. These devices have shown almost linear potentiation and depression with multiple conductance states (<200 nonvolatile states), as well as very low energy consumption of 30 fJ per spike. However, a challenge with this device might stem from its tiny conductance change (approximately 300 pS), which is too subtle for realistic measurements and applications. In 2017, Sharbati and colleagues realized an energy-efficient electrochemical transistor with a layered Graphene channel (Figure I. 6.f) [48]. While the electrochemical behaviors of graphite with Li ions have been thoroughly characterized since this material has already been widely used as an anode in Li-ion batteries, Li ions also have been reported to have an unusually high diffusion coefficient (7×10^{-5} cm^2s^{-1}) in bilayer graphene at room temperature. This combination has resulted in a well-performed transistor with linear, precise, and reversible conductance change, together with scalability in switching speed and operation energy. However, devices based on low-dimensional materials show certain integration issues due to the immature deposition techniques. Thus, they are unsuitable for the further development of neural networks composed of a large number of synaptic components.

Another recent concept that has caught the attention of researchers is combining an additional two-terminal memristive device connected in series with the gate of the synaptic transistor. This memristor acts as a switch to selectively program the transistor and prevent further unattended conductance changes through the gate [52]. Li *et al*. demonstrated a cell, i.e., 1 selector & 1 transistor, in which they utilized a two-terminal low-voltage threshold switch based on Ag filament formation (Figure I. 6.g) [49]. Concerning the transistor, they used a stack of $LiTiO_2$/$LiClO_4$:PEO/$LiTiO_2$ to reduce the built-in open-circuit voltage. Because of this combination, they were able to demonstrate linear potentiation and depression of the channel conductance in more than 250 states with long retention. In addition to those, the write noise can be as low as 200 mV, corresponding to roughly 0.3 fJ per weight update. Similarly, Fuller and colleagues realized a 1S1T cell with the same memristor, but proton (H^+) was utilized instead of Li^+ ion (Figure I. 6 h) [50]. Another design worth mentioning in this section is the selector-free synaptic transistor presented by Kim *et al.* in his 2019 paper (Figure I. 6.i) [51]. Their idea is to create an additional oxide layer between the gate electrode and the electrolyte. This layer generates a supplementary nonlinearity in the IV characteristic that can act as a transistor threshold switch. Even though clear evidence behind this phenomenon was not provided in the article, the "selector-free" approach will gradually be a requirement for fully developed synaptic transistor technology.

Table 1: Summary of current SynTs' device characterizations (Figure I. 6)

Devices / Characs.		**a**[30]	**b**[31]	**c**[46]	**d**[44]	**e**[47]	**f**[48]	**g**[49]	**h**[50]	**i**[51]
Mobile Ion		Li^+	Li^+	Li^+	Li^+	Li^+	Li^+	Li^+	H^+	Li^+
Channel		$Li_{1-x}CoO_2$	α-MoO_3	WO_3	α-Nb_2O_5	WSe_2	Graphene	$Li_{1-x}TiO_2$	PEDOT:PSS	WO_3
Dimensions		SD: Pt Gate: Si Electrolyte : LiPON 400 nm Channel : Li_xCoO_2 120nm thick 2µm length	SD and Gate: Cr/Au 60/5 nm Electrolyte: PEO:$LiClO_4$ Channel: α-MoO_3 16.8 - 28 nm thick (12-20 layers) 7µm length 4µm width	Electrolyte: LiPON Channel: WO_3 0.3-10-80 µm length 0.3-60-100µm width	SD: TiN 40 nm Gate: Cr/Au Electrolyte: Li_xSiO_2 80 nm Channel: α-Nb_2O_5 20 nm thick	Gate: Pd/Au 50/10nm Electrolyte: PEO:$LiClO_4$ Channel: WSe_2 (5 – 48layers) 1 µm length 3µm width	Electrodes: Cu 80nm Electrolyte: PEO:$LiClO_4$ Channel: Graphene 3 – 20 nm thick 15µm length 4µm width	SD: Pt (50nm) Gate: Li_xTiO_2 (90nm) Electrolyte: PEO:$LiClO_4$ Channel: Li_xTiO_2 8µm thick 200 µm length 10 µm width	Electrodes: Ti/Au 5/100 nm Electrolyte: Nafion Channel: PEI/PEDOT:PSS 125 µm length 45 µm width	Channel: WO_3 W/L from 100/100 µm to 10/4 µm
Physical characterizations		SEM	AFM	TEM	TEM, SIMS, GIXRD, AFM, XPS	FIB-HRTEM, SAED, AFM	Optical imaging, Raman Spec	Optical imaging, SEM	-	TEM
Electrical Characterizations	**Common**	• Conductance range : G_{SD} or I = f(V) using gate sweep (V_G or I_G) • Number of states: G_{SD} = f(pulse number) for one cycle. • Linearity: ΔG = f(G_0), G = f(pulse number) for one cycle • Retention: G or I = f(t) • Endurance: cycling cycles • Energy consumption: ΔG = f(t_w), E = f(channel area)								
	Unique	-	C = f(Freq) STDP	-	STDP	SRDP, STDP	STDP	-	-	-

4.2 Performance parameters of SynTs as artificial synapses

The non-volatile synaptic transistors are developed for constructing novel neuromorphic computing architectures to achieve memory density, energy efficiency, and massive parallelism for data-centric tasks. Crossbar arrays built from SynTs reduce the computing cost by alleviating the need to shuttle data from memory storage to the central processing unit. To do so, the connected SynTs, in the role of artificial synapses, act as an embedded memory by storing information in a non-volatile manner in their channels' conductance. Subsequently, the crossbar systems perform neural computing operations such as vector-matrix multiplication and parallel weight updates. For SynTs to become good candidates for artificial synapses, they must meet some electrical characteristics, such as conductance modulation with linear and symmetric profile, analog state retention, endurance, and low energy consumption per operation. I will discuss these figures of merit in detail in the following subsections.

4.2.1 Conductance modulation

When studying SynT, the most important figure of merit is the hysteresis change of channel conductance, between the lowest conductance to the highest conductance (see Figure I. 7). The source electrode is grounded to measure this, and a linearly ramped voltage is applied on Gate-Source (GS) electrodes while recording the current flowing through Drain-Source (DS) electrodes. Observing this bidirectional change of the DS conductance is considered proof of the concept of electrochemical synaptic transistors. The scan speed (scan rate) is of great importance in determining the conductance modulation of the SynTs. The scan rate signifies how fast the applied potential is linearly varied. It is often presented in the unit of [V/s] or constant Gate current [A] by controlling an unchanged current flowing out of the electrodes.

The electrical property of the channel layer is modified upon the change of its ion content via the charge (ion extraction) and discharge (ion insertion) processes. Fuller *et al.* demonstrate the change of $Li_{1-x}CoO_2$ conductivity by monitoring the gate current to be $|I_G|$ = 350 nA for a full charge-discharge cycle and record the DS conductance (Figure I. 7.a) [30]. When the gate voltage (Open circuit potential – OCP) decreases from 0 V to -4.2 V (the charge process), the conductance of the channel increases from 4.5 to 270 µS. The discharge process drives the conductance back to the initial low conductance state and the gate voltage from -4.2 V to 0 V. To avoid the saturation region, where the modulation of the conductance is low under the sweep of gate voltage, a voltage window of [-4.1 V, -3.0 V] is selected. The shown

profiles of cycles 1, 10, 20, 30, and 40 on this voltage window on the inset prove that these bidirectional scans are stable.

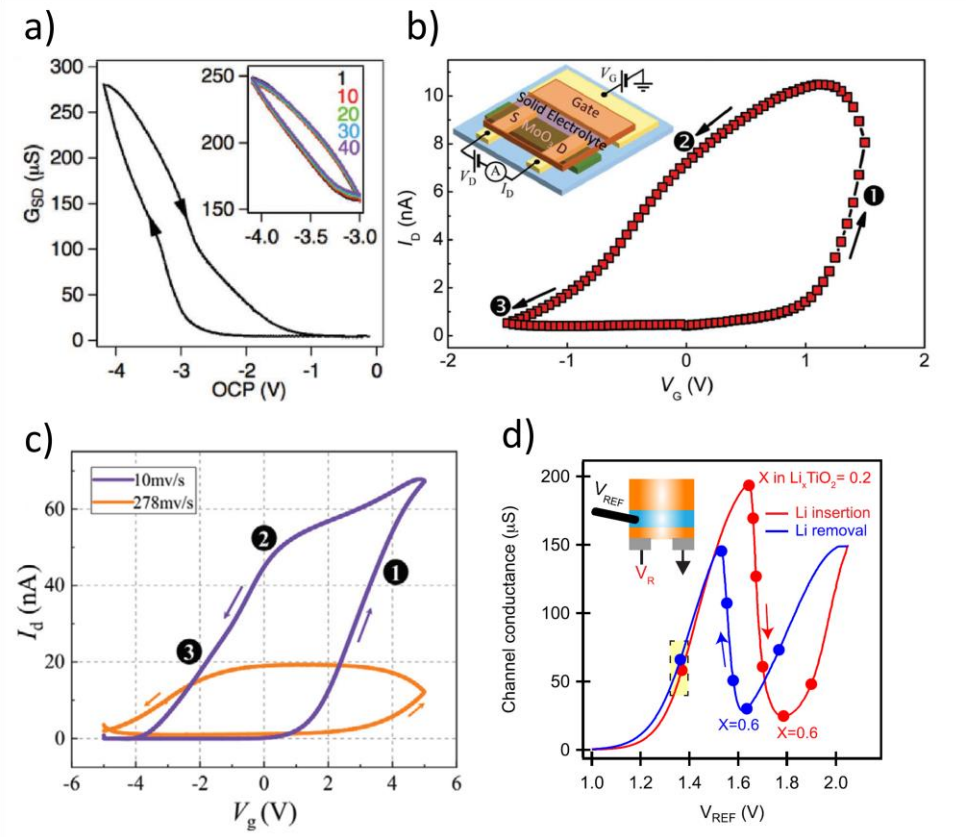

Figure I. 7: Conductance modulation of channel layer under Gate voltage sweeps. a) Li_xCoO_2 channel SynT works in the range of under 300 µS [30]. b) α-MoO_3 channel SynT works in the range of under 240 nS [31]. c) α-Nb_2O_5 channel SynT works in the range under 700 nS or 200 nS depending on the sweep rate [44]. d) Anatase Li_xTiO_2 channel SynT works in the range of under 200 µS [49].

A similar hysteresis conductance curve was also demonstrated on an α-MoO_3-based SynT (Figure I. 7.b) [31]. The voltage window used for this scan is from -1.5 V to 1.5 V at 20 mV/s. Lithium (Li^+) ions intercalate into the 2D α-MoO_3 film and induce an important high/low conductance ratio of 17 at V_G = 0 V. As we can observe from the curve, there is no significant change in DS current until the V_G reaches 1 V. This is explained by the accumulation of mobile Li ions on the topmost layer of the α-MoO_3 nanosheets before actually intercalating into the lattice of the channel with the increase of Gate potential. In contrast to the rapid increase of I_{DS} from the forward scan, the decrease of channel conductivity under Li extraction is rather steady from the high conductance to the initial low conductance. This gradual decline of conductance can be explained by the fact that Lithiated α-MoO_3 (Li_xMoO_3) is thermodynamically stable, and a high negative Gate voltage is required to extract all the

inserted Li ions. The difference in dynamics between forward and reversed scans creates the large hysteresis curve as observed.

Intercalation processes of mobile ions into various types of channel material have different timescales, and these timescales of the experiment are determined by scan rate. Li *et al.* show the bidirectional curves at different 2 scan rates of 10 mV/s and 278 mV/s with their α-Nb_2O_5-based SynT (Figure I. 7.c) [44]. The V_G is swept from –5 to 5 V, creating a counter-clockwise loop with a large high/low current ratio. It is clear from graph that with smaller scan rate (10 mV/s) the ratio of high/low conductance is 3.5 times higher than that of higher scan rate (278 mV/s). This explains the fact that with slow scan rates, Li ions have more time to intercalate into α-Nb_2O_5, and thus the modulation of electrical property is of higher importance. However, one should keep in mind that the scan rates partially represent the realistic resistive switching speed of SynTs. Therefore, we have to select carefully a balanced trade-off between a high ON/OFF ratio and fast switching devices.

With the help of a reference electrode (REF), monitoring how channel conductance is modified under the intercalation of ions is possible using the voltage sweep method. Li *et al.* show how the anatase phase Li_xTiO_2 channel layer changes its conductance by electrochemically inset and remove Li ions using $Li_{0.7}FePO_4$ reference (Figure I. 7.d) [49]. The controlled charge and discharge constant current is 1 nA. Here, higher V_{REF} means that the channel Li_xTiO_2 is more reduced or more Li ions are inserted, and vice versa. The conductance of the anatase Li_xTiO_2 film increases significantly with x in Li_xTiO_2 rising from 0 to 0.2 by enhancing the Ohmic contact between DS electrodes and channel layer, before dropping sharply with x being from 0.2 to 0.6. This drop is associated with an anatase-to-Li-titanate phase transition. As more ions are intercalated into the structure, the re-increase of channel's conductance is observed. The reversed scan shows the same pattern but with a smaller amplitude and a hysteresis, as seen previously in other electrochemical systems. The operating region for this Li_xTiO_2-based device is between x = 0 and x = 0.65 for linearity reason.

Studying these conductance hysteresis curves induced by Gate voltage sweep as a kick-start is more than obligatory for three reasons: (i) Observing the trend of conductance modulation, (ii) Determining the working voltage window to obtain a linear conductance profile, (iii) Estimating the switching speed of the device to balance with the high/low conductance ratio. However, it is important to note that the conductance modulation measured by the sweeping the GS voltage does not match the operating conditions of realistic transistors, which work with short, squared pulses followed by a relaxation time.

4.2.2 Analog states modulation

In biological terms, long-term synaptic plasticity characteristics, including long-term potentiation (LTP) and long-term depression (LTD) of synaptic weight, are fundamental to accomplishing neuromorphic functions. Long-term potentiation is behind the mechanism of fundamental learning and memory in biological systems. On the contrary, long-term depression is applied to selectively weaken specific synapses and prevent encoding new information [53]. To mimic these features, the SynTs to be implemented as artificial synapses in the artificial neural network accelerators need to demonstrate long-term plasticity by modifying their conductance levels. We refer to this characteristic as the analog states modulation.

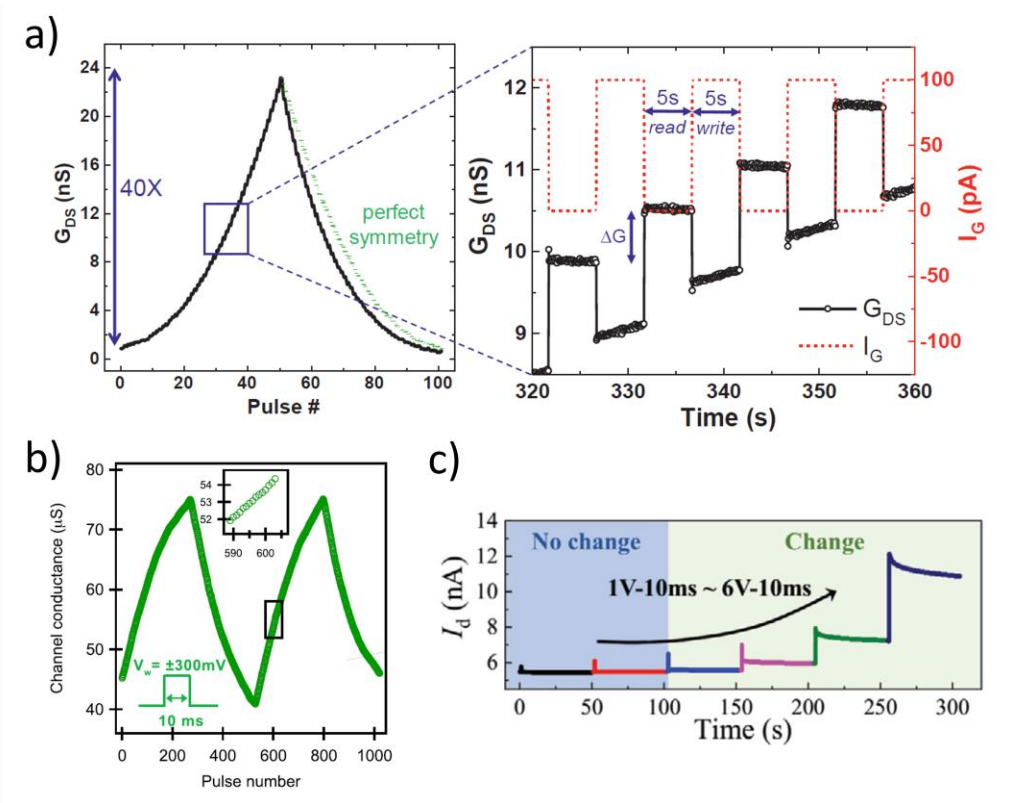

Figure I. 8: Analog states modulation of SynTs. a) Gate current pulses (50 up then 50 down pulses with amplitudes I_G = ± 100 pA and width t_w = 5 s), showing good symmetry and a large high/low conductance ratio ~ 40. Zoom-in shows the behaviors of channel conductance during and after pulses [46]. b) More than 500 pulses per cycle realized with V_G = ± 300 mV and short pulses of t_w = 10 ms, yielding a 2X high/low conductance ratio [49]. c) A threshold at 3 V between volatile and non-volatile analog state modulation with voltage amplitude as a parameter varying from 1 to 6 V and a constant pulse duration of 10 ms [54].

Analog states modulation is the ability to switch among different conductance levels of the channel induced by the electric field from the gate. The increment (potentiation) or deduction (depression) of the channel's conductance is realized by inserting or extracting mobile ions into the channel material host. Short and squared pulses of either potential difference or controlled current are applied to Gate-Source electrodes to perform the doping reactions (WRITE operations). Then they are followed by short pauses of relaxation. After applying writing pulses, the gate probe is usually switched to floating mode (no GS current exchange). The Drain-Source current is sampled during the pause period to record the analog change between the programming pulses (READ operations). The READ pulses, however, are kept as small as possible to avoid disturbing the programmed states, usually in the range of 50 – 100 mV.

In Figure I. 8.a, Tang et al. demonstrate their long-term plasticity experiment on a WO3/LiPON gate stack SynT [46]. In this case, the writing pulses are current monitored, i.e. controlling the amount of charge inserted/extracted into the WO3 channel. A series of 100 pulses of IG = 100 pA, and tw = 5s duration are used to program the device for a full cycle, from the low conductance state of less than 1 nS to the high conductance state of 24 nS, yielding a large (40X) dynamic range. It should be noted that the conductance G and its change per pulse ΔG can be tuned by modifying pulse width/amplitude parameters, device geometry, and material engineering. In spite of some presented excellent merits of linearity and stability, the time spent for writing and settling for this device is quite long due to low ion kinetics.

Li and colleagues present a fast and low voltage programming scheme on their thin-film, Anatase-LixTiO2-based SynT (Figure I. 8.b) [49]. The weight update (or conductance modulation) for more than 500 states is conducted with writing pulses of VG = ± 300 mV and duration tw = 10 ms, giving a steep conductance change with 2X dynamic range, from 40 µS to 80 µS. The device can be programmed at low write voltage because the gate metal is of the same material as the channel, creating a lower chemical potential difference and providing an advantage in energy consumption. The relation between the applied voltages and the energy consumption will be discussed in the following section.

The conductance change per write pulse is modifiable by playing pulses parameters. However, if not enough energy is provided for the intercalation reactions, the change of the channel's conductivity is temporary. This transient modification is referred to as short-term plasticity. In Figure I. 8.c, Li et al. present the time evolution of their SynT's DS current under different stimulus parameters (from VG = 1 to 6 V, 1 V-space) and the same duration time (tw = 10 ms). The threshold voltage is at about 3 V, where we can observe volatile and non-volatile operation regions. When the applied pulses VG is less than 3 V, the DS current IDS first spikes and then quickly decays back to the initial value, resulting in zero net conductance

change. By increasing the V_G to 3 V or higher, the Id gradually relaxes to a stable level and does not decay back to its initial value for a long period. The same story goes for the pulse duration variation, i.e., shorter pulses will lead to minor conductance modification and are likely less stable than longer pulses, but this feature is system-dependent. Therefore, researchers have to do experiments on their devices to find the threshold values for pulses' amplitude and duration to balance the need for energy recession and stable analog states.

4.2.3 Write nonlinearity

Programmable SynTs are the essential hardware to build artificial neural network processors. In these crossbar systems, ideally, the SynTs would be programmed in a perfectly linear and controllable way, allowing them to be assigned to any arbitrary analog value. Unfortunately, realistic devices exhibit write nonlinearity, which stems from material properties or device architecture that one needs to consider when modelling artificial synapses with these transistors. There are two subclasses of write nonlinearity: Asymmetric nonlinearity, characterized by the asymmetric ratio AR, and symmetric nonlinearity, illustrated by nonlinearity parameter α.

The asymmetric nonlinearity can reflect how much the potentiation slope differs from the depression slope (Figure I. 9.a). Usually, this nonlinearity is asymmetric with regard to the direction of the pulse. For example, near the maximum conductance level, a given pulse on the upward curve will contribute lightly to increasing the conductance. Still, it can decrease the conductance significantly on the downward curve. This is particularly true for electrochemical systems with an EDL charging types. In other words, the last pulses will accumulate ions on the surface and contribute faintly to the increase of channel conductance; a pulse with opposite polarity, however, will purge this layer, resulting in a massive drop in conductance. The AR of the modulation is calculated by the following equation:

$$AR = \left[\frac{\max\left|G_p(n) - G_d(n)\right|}{G_p(n_{max}) - G_d(n_{max})}\right] \; for \; n = 1 \; to \; n_{max} \qquad \text{Eq. 1}$$

where $G_p(n)$ and $G_d(n)$ are the channel conductance values after the n^{th} potentiation and depression pulses, respectively, and n_{max} is the maximum pulse number for potentiation/depression. For ideal symmetry, the AR should be 0.

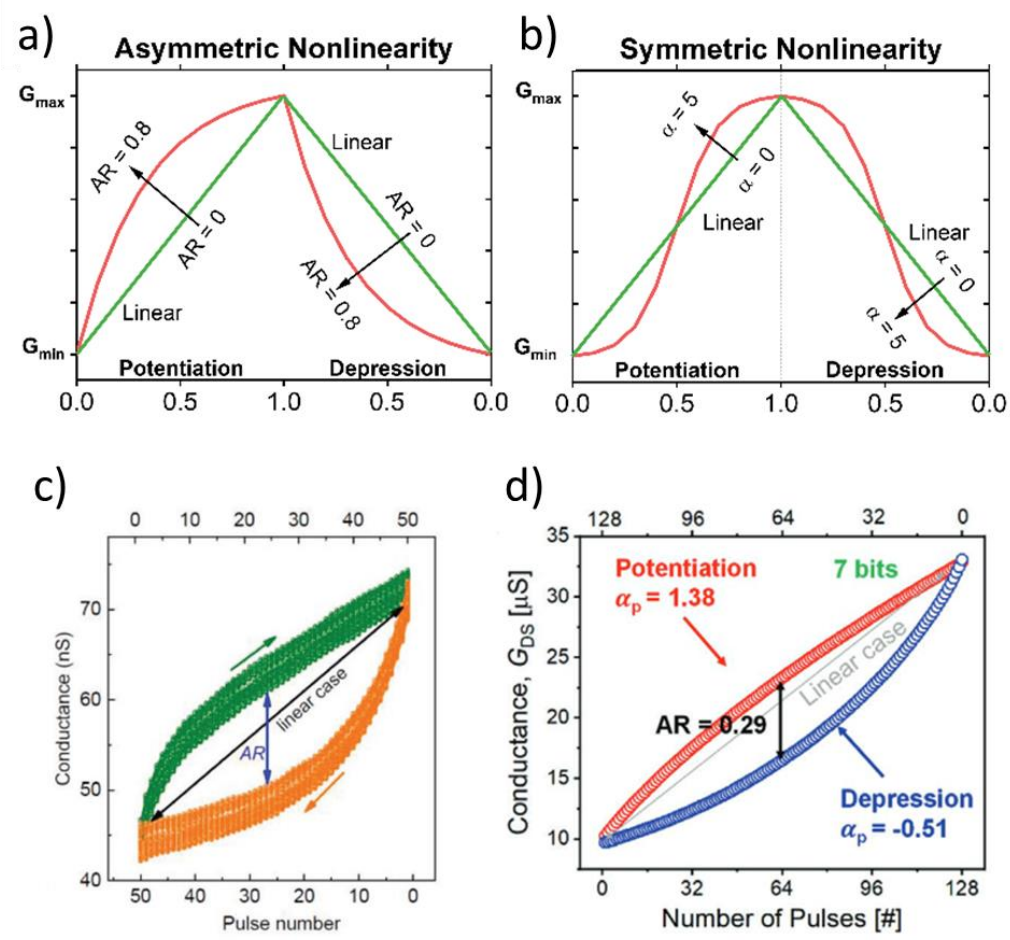

Figure I. 9: Write nonlinearity of synaptic devices. a) Asymmetric nonlinearity graph between a linear profile (AR = 0), and an asymmetric profile (AR = 0.8). b) Symmetric nonlinearity graph between a linear profile (α = 0) and a nonsymmetric profile (α = 5). c) 50 cycles plotted in a graph to demonstrate nonlinearity properties in a reaslistic SynT, for this device, the calculated AR = 0.31 ± 0.12 [31]. d) Conductance modulation profile of 256 states (7 bits) with quantified nonlinearity [54].

On the other hand, some devices display a symmetric nonlinearity (Figure I. 9.b). This type of symmetry happens on electrochemical SynTs where the working voltage range covers different doping mechanisms. Some doping reactions are more significant (more mobile ions are intercalated or extracted) than others. Faradaic reactions and non-Faradaic reactions during a discharge curve (ions insertion) can be taken as an example. At the beginning of the cycle, a non-Faradaic reaction prevails. Thus, not a lot of ions are extracted from the channel creating a slow increase in channel conductance. Gradually, when the gate potential reaches the redox potential, a massive amount of ions are inserted into the matrix, driving a surge in the conductivity of the channel. As the potential sweep continues, fewer ions can intercalate into the channel, so we observe a saturation trend at the end. The reversed sweep has a similar

pattern. The conductance G of devices with symmetric nonlinearity can be modelled using the following Sigmoid-type equation:

$$G = A \times \frac{1}{1 + e^{-2\alpha(p-0.5)}} + B \quad \text{Eq. 2}$$

where:

$$A = (G_{max} - G_{min}) \times \frac{e^{\alpha} + 1}{e^{\alpha} - 1} \quad \text{Eq. 3}$$

$$B = G_{min} - \frac{G_{max} - G_{min}}{e^{\alpha} - 1} \quad \text{Eq. 4}$$

G_{min} is the minimum conductance, G_{max} is the maximum conductance, p is pulse number, and α is a parameter characterizing the nonlinearity. α is defined such that the symmetric and asymmetric models have the same slope at the center conductance: $(G_{min} + G_{max})/2$.

For the devices whose conductance modulation profiles do not clearly resemble a logistic function, we can fit the conductance values by using the below equations:

$$G = \left((G_{max}^{\alpha} - G_{min}^{\alpha}) \times p + G_{\min}^{\alpha}\right)^{\frac{1}{\alpha}} \text{ if } \alpha \neq 0, \text{ and} \quad \text{Eq. 5}$$

$$G = G_{min} \times \left(\frac{G_{max}}{G_{min}}\right)^{\omega} \text{ if } \alpha = 0 \quad \text{Eq. 6}$$

Two examples of write nonlinearity in realistic electrochemical transistors are presented in Figure I. 9.c, d. In their work, Yang *et al.* demonstrate a 50-cycle condensed plot of programming an α-MoO_3-based SynT (Figure I.9.c). We can observe that their device shows asymmetric nonlinearity property as the first depression pulses deduce the conductance rapidly. The calculated AR (using Eq. 1) for this profile is 0.31 ± 0.12, which is reasonably linear. Li et al. illustrate a linear conductance modulation profile with 0.29 AR using their α-Nb_2O_5 channel SynT (Figure I.9.d). In this programming scheme, 7-bit (128 analog states) are shown with the doping of Li ions. The symmetric linearity parameters α of the potentiation curve and depression curve are calculated (using Eq. 5) to be 1.38 and -0.51, respectively. Nonlinearity is an indispensable property of realistic nano-devices, and the impact of these features on the performance of neural network systems will be discussed in a following chapter.

4.2.4 State retention

Studying state retention of a nonvolatile SynT device consists of checking if the programmed states are stable over time after the WRITE operations. After inducing the conductance modulation (either potentiation or depression), the ability to hold the states can be analyzed by sampling the channel using a low (around 100 mV) read voltage over time, and subsequently measuring a current versus time (I_{DS} = f(t)) curve for each state. An example of this strategy is showed in Figure I. 10.a. Authors have programmed the devices using 9 identical pulses for potentiation and depression. Following each pulse is a relaxing and sampling time of 100 seconds. Receiving pulses from the gate, the conductance rises temporarily to a high value, and then decreases to a stable state, which can be considered as a long-term effect.

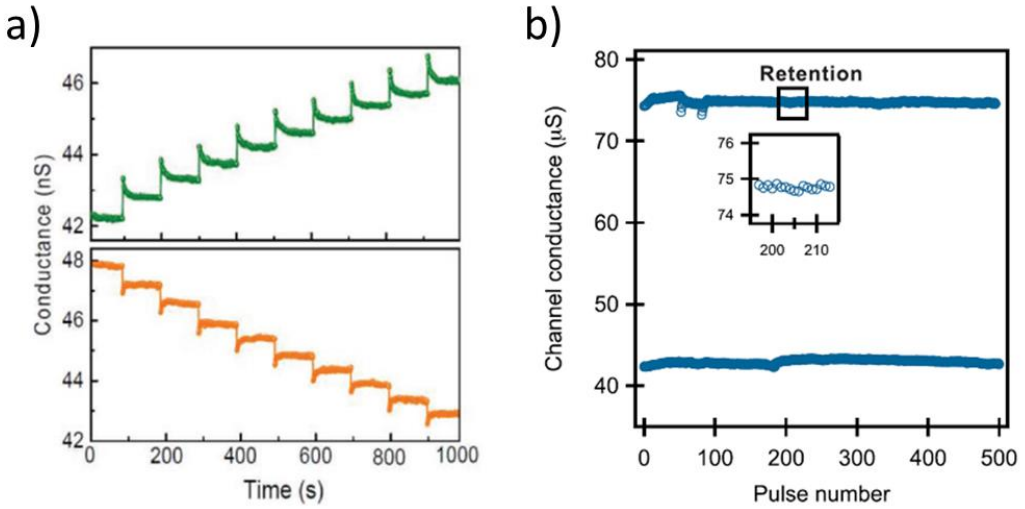

Figure I. 10: Analog state retention of synaptic transistors. a) Retention of the α-MoO_3-based channel conductance states for LTP (top) and LTD (bottom) in periods of 100 seconds [31]. b) The retention of more than 500 seconds of single states (highest and lowest) measured on a 1S1T device after the selector is switched off (each pulse represents 1 s) [49].

For electrochemical transistors, certain phenomena can disturb the programmed states and diminish the state retention of the devices, including redistribution/rearrangement of ions inside the channel host, surface self-relaxation, and ions migration between the gate and the channel. In order to prevent such effects, scientists have implemented a memristive selector acting as a switch. This selector is turned off if the applied voltage is under a threshold voltage, thus, preventing the potential disturbance or noise from the electrical system. Li and colleagues demonstrate the state retention of a 1S1T device with a Pd/Ag:SiO2/Pt memristor

as the selector (Figure I. 10.b). Here, after programming the device to a high conductance state (75 µS), the memristive selector switches by relaxation, and DS electrodes are sampled by pulses of 1-second duration. This stability of over 500s proves that bulk ion migration is the critical phenomenon of nonvolatile conductance modulation in this electrochemical SynT.

Further experiments show that this device can hold this high conductance state for more than 7 hours with the selector in the OFF state. Long retention is definitely desired for a memory system. However, a recent study demonstrated that a potential resistive device that will be implemented as an artificial synapse in a neuromorphic architecture does not require very long retention to function well [15].

4.2.5 Energy consumption

Achieving high computation yield with minimum consumption of energy is the ultimate goal of the field of neuromorphic computing [55]. It is desirable to develop a crossbar architecture having higher energy efficiency than CMOS systems, and then gradually surpassing the brain itself [30], [56].

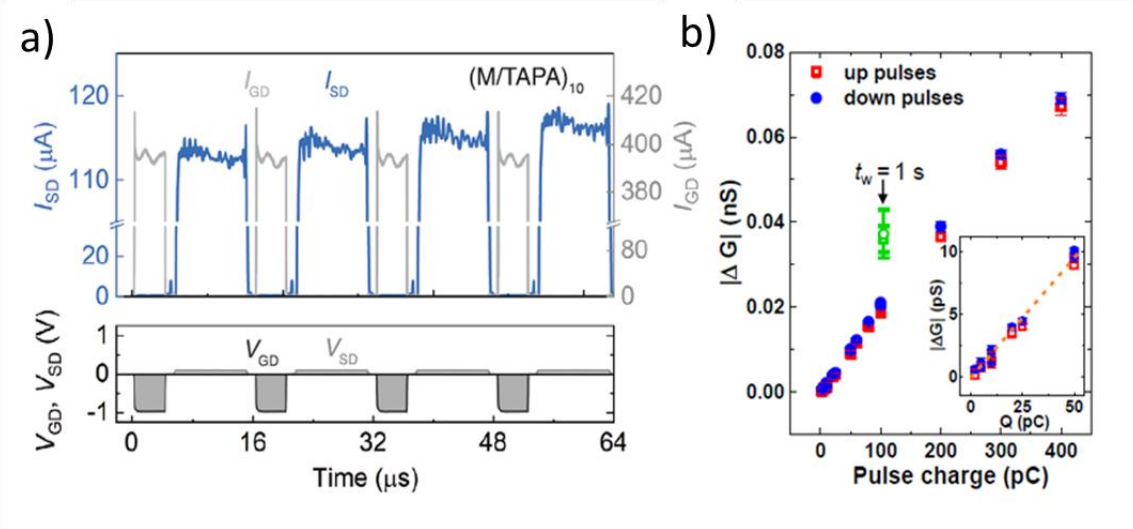

Figure I. 11: Current measurements for energy calculation. a) The pulses scheme for GS and DS electrodes (bottom), and the resulted current measurement. The I_{SD} represents the analog states and I_{GD} represents the flow of charge transferred [57].

Characterizing writing (switching) energy requires measuring the charge transferred upon each weight update. This charge quantity is the outflow/inflow of electrons

compensating for the voltage-induced ionic intercalation into the channel, creating an electrostatic equilibrium. The total charge recorded for potentiation (depression) pulses in a cycle is divided by a total number of states, yielding the average charge required for programming from one state to an adjacent one. The following equation then calculates the energy for each WRITE operation:

$$E_W = V_G \times \Delta Q \qquad \text{Eq. 7}$$

where E_w is the energy consumption per operation, V_G is the potential applied on GS electrodes, and ΔQ is the average charge transferred.

In Figure I. 11.a, we can observe the potential pulses applied to modified the analog states and the channel sampling of a SynT (bottom graph), and the subsequent current readouts (top graph) [57]. With approximately 400 µA each pulse, the channel witnesses an increase of ΔG = 1 µS. The energy consumption per writing of this device then equals 1.6 nJ by Eq. 7. We can see the relation between the charge transferred and conductance gap ΔG in Figure I. 11.b. [46]. With a certain charge injected/extracted, the gap of potentiation/depression nearly superimposes. In addition, the conductance gap ΔG linearly scales with the pulse charge ΔQ. The amount of charge transferred controls the conductance gap ΔG and the writing energy. Therefore, efforts have been made to decrease this quantity, either by reducing the volume of the channel (reducing gate area, channel area and thickness) [46], [48], or deducting the amount of ions injected for each pulses but still maintaining the stable of the analog states. While the former method requires engineering endeavor, the latter solution relies on finding the right pulse parameters to control single ions intercalation.

Several reported SynTs demonstrate a high, temporary current increase after the pulse application and then relax rapidly. This current is referred to as the excitatory postsynaptic current (EPSC). The energy consumption for a single pulse event for these devices is calculated as following:

$$E_W = I_p \times V_{DS} \times \Delta t \qquad \text{Eq. 8}$$

in which I_p, V_{DS}, and Δt represent the peak value of the EPSC, the reading voltage, and the pulse duration, respectively.

Crossbar arrays' working mechanism is reading the conductance of the elemental SynTs' channel following Ohm law and Kirchhoff circuit law. Thus, minimizing the conductivity of the channel layer will lower the overall power consumed for operating neuromorphic

networks [56]. In addition to reducing the energy consumption to operate ANN systems, low conductance synaptic devices are favorable for dense crossbar arrays. Low device conductance is highly required for use in crossbar arrays. For instance, to support a 1000x1000 crossbar with a fully parallel write/read operation, each synaptic device can load a maximum current of 10 nA (or 200 nS at 50 mV) because the scaled wires at 10 nm half-pitch can only handle 10 μA to avoid electromigration issues. Furthermore, a higher operating current on the system will induce unacceptable parasitic voltage drops and excessive energy dissipation on connection lines [56]. For these reasons, scientists are focusing on materials with intrinsic or engineered low electrical conductivity materials to construct SynTs.

4.2.6 Endurance

If SynTs are to be integrated into ANN accelerators, they must demonstrate stable operation during extensive cycling. In a SynT, device endurance is defined as the number of cycles and times it can be switched among the intermediate states keeping a defined ratio between the highest conductance state/lowest conductance state and the conductance gap among each other. Therefore, an endurance test consists of finding out the maximum number of operations (or cycles) that the intermediate analog states are still distinguishable. The common figure of merit is a saw-like graph representing a series of conductance modulation cycles (see Figure I. 12). Several devices are demonstrated to survive several hundreds of cycles (Figure I. 12.a) [44], while others can keep their functionalities after 2.7x107 writing operations (Figure I. 12.b) [51]. The failure of the resistive device may not happen in one specific cycle, but it may be progressive (Figure I. 12.c) [58].

Here, the device has been cycled for 105 operations, and during the course of this endurance test, the Gmax loses 12%. It is worthy to note that 105 operations of weight updates are largely sufficient for training the MNIST data set because not every synapse is updated in each training cycle [59]. There are many possible phenomena associated with the gradual decrease of Gmax in electrochemical transistors, namely material degradation (phase change), mobile ions loss due to trapping or being oxidized, etc. [60], [61]. The endurance characterization is performed in a similar scheme to the conductance modulation test, but it takes long time to complete. Thus, environmental testing conditions such as temperature, humidity, and air exposure must be considered when performing such tests for high precision.

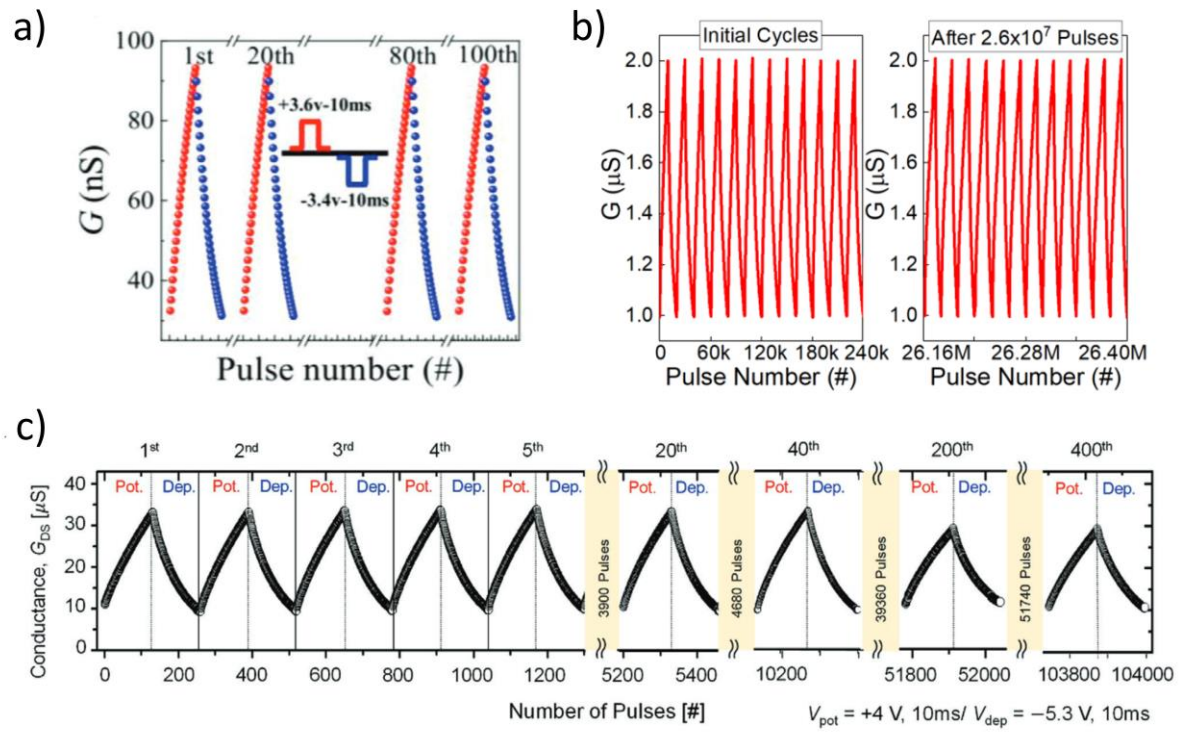

Figure I. 12: Endurance test of SynTs. a) 100 cycles of channel conductance switching with 32 discrete states induced by asymmetric V_G pulses [44]. b) Demonstration of endurance for 1300 cycles and 2.6×10^7 pulses on a SynT using a conditioned cycling measurement, and the device still works after the last cycle of this test [51]. c) A SynT demonstrated with $>10^5$ writing operations and the evolution of the highest conductance value G_{max} over the cycles [44].

4.2.7 Temporal dynamics for SNN applications

The aforementioned merits, including stable, numerous conductance states and linear and reversible modulation profiles, are more adapted to the analog-based ANN applications. In this subsection we will consider a SNN-oriented properties demonstrated on certain SynTs [43], [47], [58]. SNNs have recently attracted much attention due to their close similarity to biological systems. In SNNs, the input signals are spikes, and being able to process spikes is believed to be one of the reasons that the brain is so spectacular at sequence recognition as well as memory [62]. Developed for several decades, the Hebbian learning rule is mainly reflected in two kinds of plasticity: spike-timing-dependent plasticity (STDP) and spike-rate-dependent plasticity (SRDP). STDP is defined as the change of synaptic weight as a function of the time difference between the pre-synaptic spikes[63]. In general, the weight change increases nonlinearly as the time difference decreases, and this feature can be quantified by calculating the paired-pulse facilitation (PPF) ratio (Figure I. 13.a) [31].

$$PPF\ ratio = \frac{A_2 - A_1}{A_1} \qquad \text{Eq. 9}$$

where A1, A2, as denoted in the plot, are the peak current after the first and the pre-synaptic pulses, respectively. The experimental results can be presented with the PPF ratio as a function of pulse interval (Δt) well approximated using the double exponential decay function ($f(t) = ce^{\frac{-\Delta t}{\tau}}$) that is verified in biology [64].

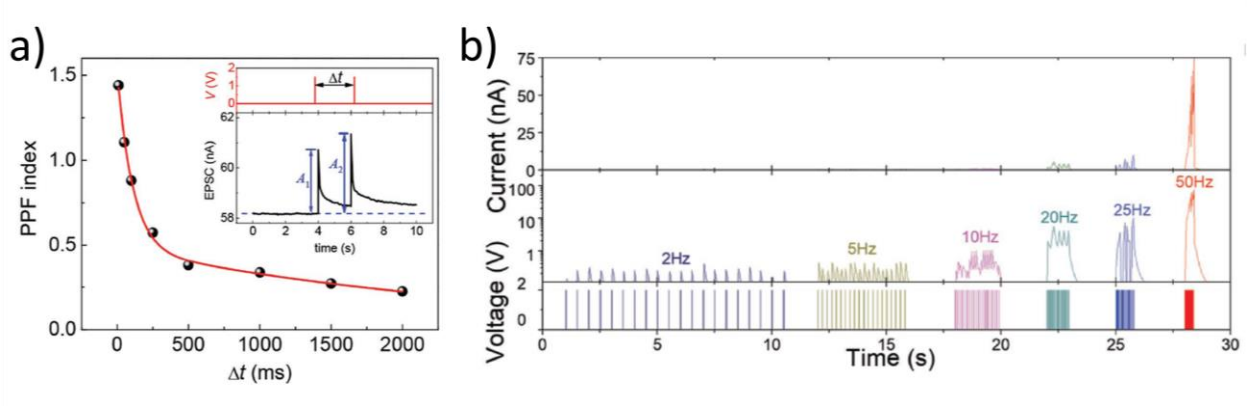

Figure I. 13: Temporal dynamics demonstrated on SynTs. a) Paired-pulse facilitation (PPF) ratio measurement on an α-MoO_3-based SynT represents STDP characteristics [31]. b) The current (weight) change in linear scale (top panel), log scale (middle panel) in WSe_2 based synaptic transistors as a function of the frequency of gate pulses, implying SRDP behavior [47].

SRDP states that the synaptic weight change is a function of the frequency of the pre-synaptic spikes [65]. According to SRDP learning rules, pre-synaptic pulses with high frequency result in a significant change in weight value, and that with low frequency will lead to small and short-term potentiation STP (Figure I. 13.b) [47].

The realization of "brain-like computing" can start from simulating the structure and function of neural networks and artificial synapses without waiting for neuroscientists and cognitive scientists to understand fully the brain's mechanism, which may even need the exploration process to be longer.

5 CONCLUSIONS

In summary, we have discussed the necessity of developing a novel computing paradigm stemming from Von Neumann bottleneck of conventional computer architecture. Taking the inspiration from human brain, the neuromorphic computing systems have great potential in performing data-intensive tasks in an energy- and time-efficient way.

Different types of physics are employed to realize electronic devices that will be used for constructing artificial synapses in neuromorphic computing hardware. Several of them are described in this chapter, including the two-terminal devices (ReRAM, PCM, and MRAM), and three-terminal devices (FeFET and SynT). Each of these technologies have both advantages and drawbacks when being implemented as artificial synapses. A comparison radar chart to summarize the performance of synaptic devices can be found in Figure I. 14. The two-terminal devices in general have fast speed, high endurance, great dynamic range and retention. However, they also exhibit nonlinear behavior and high operation (WRITE/READ) power that can limit learning accuracy and power efficiency in large-scale neural networks.

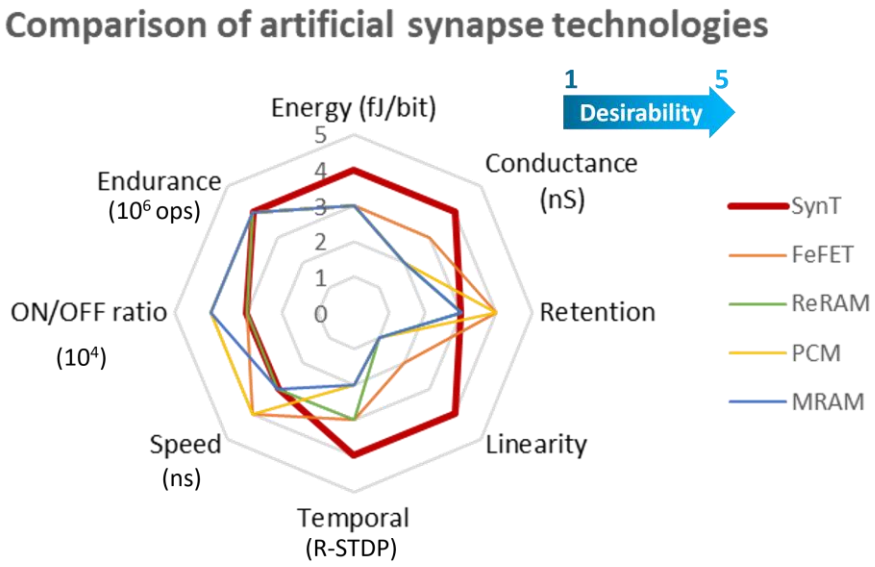

Figure I. 14: Radar graph comparing the device metrics among SynT (bold red line), FeFET (orange line), ReRAM (green line), PCM (yellow line), and MRAM (blue line).

FeFET synapses, on the other hand, may offer fast programming operations, high stability, and less update variations. Nonetheless, this type of three-terminal device suffers from the similar scaling problems as DRAM and floating gate memories because in essence they all are charge-based memories. Furthermore, temporal dynamics can be difficult to

demonstrate on FeFET components. Nanoscale SynT devices are shown to have a great precision, ultralow operation power, linear analog-state modulation, and potentially low device variations with robust electrochemical working mechanism. Nonetheless, the low programming speed, mediocre state retention, and ON/OFF ratio are key performance merits that SynT needs to work on to become more appealing for neural network applications.

State-of-the-art of SynTs before the beginning of this work is discussed in the section 4 based on several performance figures of merits such as conductance modulation, analog state modulation, nonlinearity, retention, energy consumption, endurance, and temporal dynamics. A summary of selected SynTs can be found in Table 2. To optimize further the performance of SynTs, several approaches have been suggested, including novel material screening and design (deposition techniques, phase engineering), dimensions shrinking (ultrathin, small active area layers) and advanced process integration (CMOS-compatible elaboration techniques). Furthermore, understanding the underlying physics of resistive switching of Li-based SynTs, either by advanced physical characterization or atomistic simulation, plays an important role in advancing further the existing and implementing new gate stacks.

In this thesis, we focus on three goals: (i) proposing innovative solid-state gate-stacks and designs to optimize the overall performance of SynTs (ii) elaborating and characterizing such solid-state electrochemical synaptic transistors with CMOS compatible processes, (iii) demonstrating their required synaptic functionalities thanks to the fast dynamics of Li-ion intercalation leading to rapid and ultralow power analog switching.

Table 2: SynTs' Performance figures of merits

Devices / Merits	**a** [30]	**b** [31]	**c** [46]	**d** [44]	**e** [47]	**f** [48]	**g** [49]	**h** [50]	**i** [51]
Conductance modulation range	4 - 270 µS	8 - 240 nS	-	0 - 700 nS	-	-	0 - 200 µS	-	0 - 50 µS
Analog states modulation (# states, operation G range, programming pulses)	200 states/cycle 180 – 230 µS ±400 nA/±100 mV t_w = 2s	100 states/cycle 40 - 80 nS ±2.5 V t_w = 10 ms	100 states/cycle 0 - 24 nS ±100 pA t_w = 5 s	32 states/cycle 30 - 100 nS +3.6 V, - 3.4 V t_w = 10 ms	120 states/cycle 250 - 570 pS 1.2 V, -0.4 V t_w = 100 ms	250 states/cycle 100 - 1100 µS ±50 pA tw = 10 ms	250 states/cycle 40 - 80 µS ±300 mV t_w = 10 ms	100 states/cycle 50 - 100 nS -0.95 V, 1.2V t_w = 50 µs	1000 states/cycle 1 - 2 µS ±2.5 V t_w = 100 ns
Write nonlinearity	-	AR = 0.31	α_p = 0.347, α_d = 0.268	α_p = 0.6 α_d = 1.58	-	-	-	-	-
State retention	-	100 s	-	> 1000 s	-	13 h	7 h	-	15 h
Energy consumption	10 aJ (projected)	0.16 pJ	1 fJ (projected)	200 aJ (projected)	30 fJ	500 fJ	30 pJ	-	100 fJ/nS (projected)
Endurance	40 cycles	50 cycles	1000 cycles	100 cycles	-	500 cycles (2 states)	4000 cycles	10^6 cycles	$2.6x10^7$ ops
Temporal dynamics	-	STDP, τ_1 = 110 ms and τ_2 = 2624 ms	-	STDP, τ1 = 17.27 s	STDP, SRDP	STDP, τ_1 = 22 ms, τ_2 = 315 ms, τ_3 = 19 s	-	-	-

6 REFERENCES

[1] T. M. Conte, E. Track, and E. DeBenedictis, "Rebooting Computing: New Strategies for Technology Scaling," *Computer*, vol. 48, no. 12, pp. 10–13, 2015, doi: 10.1109/mc.2015.363.

[2] M. Versace and B. Chandler, "The brain of a new machine," *IEEE Spectr.*, vol. 47, no. 12, pp. 30–37, Dec. 2010, doi: 10.1109/MSPEC.2010.5644776.

[3] G. M. Moore, "Moore's Law ,Electronics," *Electronics*, vol. 38, no. 8, p. 114, 1965.

[4] H. Markram, "The Blue Brain Project," *Nature Reviews Neuroscience*, vol. 7, no. 2, pp. 153–160, 2006, doi: 10.1038/nrn1848.

[5] C. S. Poon and K. Zhou, "Neuromorphic silicon neurons and large-scale neural networks: Challenges and opportunities," *Frontiers in Neuroscience*, vol. 5, no. SEP, pp. 2009–2011, 2011, doi: 10.3389/fnins.2011.00108.

[6] S. Li, X. Zuo, Z. Li, and H. Wang, "Applying deep learning to continuous bridge deflection detected by fiber optic gyroscope for damage detection," *Sensors (Switzerland)*, vol. 20, no. 3, 2020, doi: 10.3390/s20030911.

[7] Z. Ansari and S. A. Seyyedsalehi, "Toward growing modular deep neural networks for continuous speech recognition," *Neural Computing and Applications*, vol. 28, pp. 1177–1196, 2017, doi: 10.1007/s00521-016-2438-x.

[8] D. Silver *et al.*, "Mastering the game of Go with deep neural networks and tree search," *Nature*, vol. 529, no. 7587, pp. 484–489, 2016, doi: 10.1038/nature16961.

[9] D. Kuzum, S. Yu, and H.-S. Philip Wong, "Synaptic electronics: materials, devices and applications," *Nanotechnology*, vol. 24, no. 38, p. 382001, Sep. 2013, doi: 10.1088/0957-4484/24/38/382001.

[10] N. C. S. Larry Squire, Darwin Berg, Floyd E. Bloom, Sascha du Lac, Anirvan Ghosh, *Fundamental Neuroscience*. Elsevier, 2012.

[11] L. A. Hart, *How the brain works*. Basic Books, 1975.

[12] C. Zamarreño-Ramos, L. A. Camuñas-Mesa, J. A. Pérez-Carrasco, T. Masquelier, T. Serrano-Gotarredona, and B. Linares-Barranco, "On Spike-Timing-Dependent-Plasticity, Memristive Devices, and Building a Self-Learning Visual Cortex," *Front. Neurosci.*, vol. 5, 2011, doi: 10.3389/fnins.2011.00026.

[13] N. K. Upadhyay, S. Joshi, and J. J. Yang, "Synaptic electronics and neuromorphic computing," *Sci. China Inf. Sci.*, vol. 59, no. 6, p. 061404, Jun. 2016, doi: 10.1007/s11432-016-5565-1.

[14] V. Massimiliano and C. Ben, "MoNETA: A Mind Made from Memristors," 2010. https://spectrum.ieee.org/robotics/artificial-intelligence/moneta-a-mind-made-from-memristors (accessed May 06, 2020).

[15] D. Ielmini and S. Ambrogio, "Emerging neuromorphic devices," *Nanotechnology*, vol. 31, no. 9, p. 092001, Feb. 2020, doi: 10.1088/1361-6528/ab554b.

[16] Q. Wan, M. T. Sharbati, J. R. Erickson, Y. Du, and F. Xiong, "Emerging Artificial Synaptic Devices for Neuromorphic Computing," *Adv. Mater. Technol.*, vol. 4, no. 4, p. 1900037, Apr. 2019, doi: 10.1002/admt.201900037.

[17] X. Liao, L. Xiao, C. Yang, and Y. Lu, "MilkyWay-2 supercomputer: system and application," *Front. Comput. Sci.*, vol. 8, no. 3, pp. 345–356, Jun. 2014, doi: 10.1007/s11704-014-3501-3.

[18] Z. Wang, L. Wang, M. Nagai, L. Xie, M. Yi, and W. Huang, "Nanoionics-Enabled Memristive Devices: Strategies and Materials for Neuromorphic Applications," *Adv. Electron. Mater.*, vol. 3, no. 7, p. 1600510, Jul. 2017, doi: 10.1002/aelm.201600510.
[19] C. Mead, "Neuromorphic Electronic Systems," *Proceedings of the IEEE*, vol. 78, no. 10, pp. 1629–1636, 1990.
[20] L. O. Chua, "Memristor—The Missing Circuit Element," *IEEE Transactions on Circuit Theory*, vol. 18, no. 5, pp. 507–519, 1971, doi: 10.1109/TCT.1971.1083337.
[21] D. B. Strukov, G. S. Snider, D. R. Stewart, and R. S. Williams, "The missing memristor found," *Nature*, vol. 453, no. 7191, pp. 80–83, 2008, doi: 10.1038/nature06932.
[22] X. Zhu, Q. Wang, and W. D. Lu, "Memristor networks for real-time neural activity analysis," *Nature Communications*, vol. 11, no. 1, 2020, doi: 10.1038/s41467-020-16261-1.
[23] P. Yao *et al.*, "Fully hardware-implemented memristor convolutional neural network," *Nature*, vol. 577, no. 7792, pp. 641–646, 2020, doi: 10.1038/s41586-020-1942-4.
[24] J. Moon *et al.*, "Temporal data classification and forecasting using a memristor-based reservoir computing system," *Nature Electronics*, vol. 2, no. 10, pp. 480–487, 2019, doi: 10.1038/s41928-019-0313-3.
[25] J. J. Yang, M. D. Pickett, X. Li, D. A. A. Ohlberg, D. R. Stewart, and R. S. Williams, "Memristive switching mechanism for metal/oxide/metal nanodevices," *Nature Nanotech*, vol. 3, no. 7, pp. 429–433, Jul. 2008, doi: 10.1038/nnano.2008.160.
[26] Z. Wang *et al.*, "Memristors with diffusive dynamics as synaptic emulators for neuromorphic computing," *Nature Materials*, vol. 16, no. 1, pp. 101–108, 2017, doi: 10.1038/nmat4756.
[27] G. W. Burr *et al.*, "Phase change memory technology," *Journal of Vacuum Science & Technology B, Nanotechnology and Microelectronics: Materials, Processing, Measurement, and Phenomena*, vol. 28, no. 2, pp. 223–262, Mar. 2010, doi: 10.1116/1.3301579.
[28] A. D. Kent and D. C. Worledge, "A new spin on magnetic memories," *Nature Nanotech*, vol. 10, no. 3, pp. 187–191, Mar. 2015, doi: 10.1038/nnano.2015.24.
[29] N. Yu, K. Yukihiro, U. Michihito, F. Eiji, and T. Ayumu, "Dynamic Observation of Brain-Like Learning in a Ferroelectric Synapse Device," *Japanese Journal of Applied Physics*, vol. 52, no. 42, 2013.
[30] E. J. Fuller *et al.*, "Li-Ion Synaptic Transistor for Low Power Analog Computing," *Advanced Materials*, vol. 29, no. 4, pp. 1–8, 2017, doi: 10.1002/adma.201604310.
[31] C. Sen Yang *et al.*, "All-Solid-State Synaptic Transistor with Ultralow Conductance for Neuromorphic Computing," *Advanced Functional Materials*, vol. 28, no. 42, pp. 1–10, 2018, doi: 10.1002/adfm.201804170.
[32] Daniele Ielmini and Rainer Waser, *Resistive Switching: From Fundamentals of Nanoionic Redox Processes to Memristive Device Applications*, 25 vols. Wiley-VCH Verlag GmnH & Co. KGaA, Weinheim, Germany, 2016. [Online]. Available: 10.1002/9783527680870
[33] C.-L. Tsai, F. Xiong, E. Pop, and M. Shim, "Resistive Random Access Memory Enabled by Carbon Nanotube Crossbar Electrodes," *ACS Nano*, vol. 7, no. 6, pp. 5360–5366, Jun. 2013, doi: 10.1021/nn401212p.
[34] S. R. Nandakumar, M. Le Gallo, I. Boybat, B. Rajendran, A. Sebastian, and E. Eleftheriou, "A phase-change memory model for neuromorphic computing," *Journal of Applied Physics*, vol. 124, no. 15, p. 152135, Oct. 2018, doi: 10.1063/1.5042408.

[35] G. W. Burr and R. M. Shelby, "29.5 Experimental Demonstration and Tolerancing of a Large-Scale Neural Network (165,000 Synapses), Using Phase-Change Memory As the Synaptic Weight Element," p. 4.
[36] A. F. Vincent *et al.*, "Spin-Transfer Torque Magnetic Memory as a Stochastic Memristive Synapse for Neuromorphic Systems," *IEEE Trans. Biomed. Circuits Syst.*, vol. 9, no. 2, pp. 166–174, Apr. 2015, doi: 10.1109/TBCAS.2015.2414423.
[37] V. Milo, G. Malavena, C. Monzio Compagnoni, and D. Ielmini, "Memristive and CMOS Devices for Neuromorphic Computing," *Materials*, vol. 13, no. 1, p. 166, Jan. 2020, doi: 10.3390/ma13010166.
[38] S. Jiang, S. Nie, Y. He, R. Liu, C. Chen, and Q. Wan, "Emerging synaptic devices: from two-terminal memristors to multiterminal neuromorphic transistors," *Materials Today Nano*, vol. 8, p. 100059, Dec. 2019, doi: 10.1016/j.mtnano.2019.100059.
[39] J. Zhu, T. Zhang, Y. Yang, and R. Huang, "A comprehensive review on emerging artificial neuromorphic devices," *Applied Physics Reviews*, vol. 7, no. 1, p. 011312, Mar. 2020, doi: 10.1063/1.5118217.
[40] A. Cano and D. Jiḿnez, "Multidomain ferroelectricity as a limiting factor for voltage amplification in ferroelectric field-effect transistors," *Applied Physics Letters*, vol. 97, no. 13, pp. 52–55, 2010, doi: 10.1063/1.3494533.
[41] M. Jerry *et al.*, "Ferroelectric FET analog synapse for acceleration of deep neural network training," *Piscataway, NJ: IEEE*, vol. 6, no. c, 2017.
[42] J. Zhu *et al.*, "Ion Gated Synaptic Transistors Based on 2D van der Waals Crystals with Tunable Diffusive Dynamics," *Advanced Materials*, vol. 30, no. 21, 2018, doi: 10.1002/adma.201800195.
[43] X. Zhu, D. Li, X. Liang, and W. D. Lu, "Ionic modulation and ionic coupling effects in MoS 2 devices for neuromorphic computing," *Nature Materials*, vol. 18, no. 2, pp. 141–148, 2019, doi: 10.1038/s41563-018-0248-5.
[44] Y. Li *et al.*, "Oxide-Based Electrolyte-Gated Transistors for Spatiotemporal Information Processing," *Adv. Mater.*, vol. 32, no. 47, p. 2003018, Nov. 2020, doi: 10.1002/adma.202003018.
[45] L. Mracek, T. Syrovy, S. Pretl, S. Nespurek, and A. Hamacek, "Comparison of quasi-solid state and liquid electrolytes for organic electrochemical transistor," *Proceedings of the International Spring Seminar on Electronics Technology*, vol. 2016-Septe, pp. 66–70, 2016, doi: 10.1109/ISSE.2016.7563163.
[46] J. Tang *et al.*, "ECRAM as Scalable Synaptic Cell for High-Speed, Low-Power Neuromorphic Computing," *IEEE Electron Device Letters*, vol. 38, no. 7, pp. 997–997, 2017, doi: 10.1109/led.2017.2718158.
[47] J. Zhu *et al.*, "Ion Gated Synaptic Transistors Based on 2D van der Waals Crystals with Tunable Diffusive Dynamics," *Advanced Materials*, vol. 30, no. 21, 2018, doi: 10.1002/adma.201800195.
[48] M. T. Sharbati, Y. Du, J. Torres, N. D. Ardolino, M. Yun, and F. Xiong, "Low-Power, Electrochemically Tunable Graphene Synapses for Neuromorphic Computing," *Advanced Materials*, vol. 30, no. 36, pp. 1–6, 2018, doi: 10.1002/adma.201802353.
[49] Y. Li *et al.*, "Low-Voltage, CMOS-Free Synaptic Memory Based on Li XTiO2 Redox Transistors," *ACS Applied Materials and Interfaces*, vol. 11, no. 42, pp. 38982–38992, Oct. 2019, doi: 10.1021/acsami.9b14338.

[50] E. J. Fuller *et al.*, "Parallel programming of an ionic floating-gate memory array for scalable neuromorphic computing," *Science*, vol. 364, no. 6440, pp. 570–574, 2019, doi: 10.1126/science.aaw5581.
[51] H. Kim *et al.*, "Zero-shifting Technique for Deep Neural Network Training on Resistive Cross-point Arrays," pp. 847–850, 2019.
[52] G. W. Burr *et al.*, "Access devices for 3D crosspoint memory," *Journal of Vacuum Science and Technology B: Microelectronics and Nanometer Structures*, vol. 32, no. 4, pp. 1–23, 2014, doi: 10.1116/1.4889999.
[53] A. Citri and R. C. Malenka, "Synaptic Plasticity: Multiple Forms, Functions, and Mechanisms," *Neuropsychopharmacol*, vol. 33, no. 1, pp. 18–41, Jan. 2008, doi: 10.1038/sj.npp.1301559.
[54] Y. Li *et al.*, "One Transistor One Electrolyte-Gated Transistor Based Spiking Neural Network for Power-Efficient Neuromorphic Computing System," *Adv. Funct. Mater.*, vol. 31, no. 26, p. 2100042, Jun. 2021, doi: 10.1002/adfm.202100042.
[55] T. M. Taha, R. Hasan, C. Yakopcic, and M. R. McLean, "Exploring the design space of specialized multicore neural processors," in *The 2013 International Joint Conference on Neural Networks (IJCNN)*, Dallas, TX, USA, Aug. 2013, pp. 1–8. doi: 10.1109/IJCNN.2013.6707074.
[56] J. Hasler and B. Marr, "Finding a roadmap to achieve large neuromorphic hardware systems," *Front. Neurosci.*, vol. 7, 2013, doi: 10.3389/fnins.2013.00118.
[57] A. Melianas *et al.*, "High-Speed Ionic Synaptic Memory Based on 2D Titanium Carbide MXene," *Adv Funct Materials*, vol. 32, no. 12, p. 2109970, Mar. 2022, doi: 10.1002/adfm.202109970.
[58] J. Yu *et al.*, "All-Solid-State Ion Synaptic Transistor for Wafer-Scale Integration with Electrolyte of a Nanoscale Thickness," *Adv. Funct. Mater.*, vol. 31, no. 23, p. 2010971, Jun. 2021, doi: 10.1002/adfm.202010971.
[59] S. Yu, "Neuro-Inspired Computing With Emerging Nonvolatile Memorys," *Proc. IEEE*, vol. 106, no. 2, pp. 260–285, Feb. 2018, doi: 10.1109/JPROC.2018.2790840.
[60] V. J. Ovejas and A. Cuadras, "Effects of cycling on lithium-ion battery hysteresis and overvoltage," *Sci Rep*, vol. 9, no. 1, p. 14875, Dec. 2019, doi: 10.1038/s41598-019-51474-5.
[61] L. Zhang *et al.*, "Capacity Fading Mechanism and Improvement of Cycling Stability of the SiO Anode for Lithium-Ion Batteries," *J. Electrochem. Soc.*, vol. 165, no. 10, pp. A2102–A2107, 2018, doi: 10.1149/2.0431810jes.
[62] Q. Yu, R. Yan, H. Tang, K. C. Tan, and H. Li, "A Spiking Neural Network System for Robust Sequence Recognition," *IEEE Trans. Neural Netw. Learning Syst.*, vol. 27, no. 3, pp. 621–635, Mar. 2016, doi: 10.1109/TNNLS.2015.2416771.
[63] S. Dai *et al.*, "Recent Advances in Transistor-Based Artificial Synapses," *Adv. Funct. Mater.*, vol. 29, no. 42, p. 1903700, Oct. 2019, doi: 10.1002/adfm.201903700.
[64] R. S. Zucker and W. G. Regehr, "Short-Term Synaptic Plasticity," *Annu. Rev. Physiol.*, vol. 64, no. 1, pp. 355–405, Mar. 2002, doi: 10.1146/annurev.physiol.64.092501.114547.
[65] C. R. Noback, *The human nervous system: structure and function*, 6th ed. Totowa, N.J: Humana Press, 2005.

CHAPTER II

MICROFABRICATION TECHNIQUES OF SYNAPTIC TRANSISTORS

ABSTRACT

In the second chapter, we will introduce the processes in microfabrication, including thin-film deposition, patterning, and characterizations employed in elaborating the electrochemical synaptic transistors. This chapter is divided into 4 main sections.

Section 1 of this chapter introduces thin-film processes, including deposition and patterning. We will focus on deposition techniques and their parameters that allow the coating of controllable uniform films on 200 mm Si wafers, such as atomic layer deposition and sputtering. Details on the steps of thin-film patterning, such as photolithography and etching, will be discussed to create a comprehensive view of the process development to fabricate synaptic transistors.

Section 2 addresses multiple types of thin film and device characterizations, such as photon/electron microscopy and spectroscopy. These tests serve as quality checkpoints for devices after microfabrication. Furthermore, micrometrology reveals valuable information on the stack's appearance, elemental composition, material phase, and crystallography of the films presented in the devices, assisting in understanding their electrical performance.

We will present the process flow to elaborate SynTs in section 3. The details of different steps are shown, facilitating the reproduction or further developments of synaptic transistors based on thin films. In addition, the work to optimize the procedures, the design and the materials is also included, which gives an example of the process development for a device. Finally, some ideas to summarize the chapter can be found in section 4.

TABLE OF CONTENTS

1 MICROFABRICATION PROCESSES 56
1.1 Introduction 56
1.2 Thin-film materials and deposition techniques 56
1.2.1 Physical vapor deposition (PVD) 58
1.2.2 Chemical vapor deposition (CVD) 61
1.3 Pattern generation 63
1.4 Lithographic photoresist pattern 66
1.4.1 Photoresist application 66
1.4.2 Photoresist properties and processes 67
1.5 Etching 69
1.5.1 Wet etching 70
1.5.2 Dry etching 72
2 THIN-FILMS AND MATERIALS CHARACTERIZATIONS 74
2.1 Microscopy and visualization 74
2.1.1 Scanning electron microscopy (SEM). 74
2.1.2 Transmission electron microscopy (TEM) 77
2.2 Physicochemical analyses 79
2.2.1 Raman spectroscopy 79
2.2.2 Energy dispersive X-ray spectroscopy (EDX-EDS) 81
3 ELECTROCHEMICAL SYNAPTIC TRANSISTOR PROCESS FLOW 84
3.1 Process flow 85
3.2 Progressive optimization of microfabrication steps 87
3.2.1 SD patterning 87
3.2.2 Channel processes 88
3.2.3 Gate and electrolyte patterning 91
4 CONCLUSIONS 93
5 REFERENCES 94

1 MICROFABRICATION PROCESSES

1.1 Introduction

A schematic view of our elaborated synaptic transistors during my thesis is illustrated in Figure II. 1 below. For the bottom electrodes (source and drain, in yellow), two different materials have been tested: Ti and Pt (deposited by DC-sputtering). Concerning the channel (in red), two materials have been examined: $LiCoO_2$ (deposited by RF-sputtering) and TiO2, (deposited by ALD). The electrolyte (in blue) was made of LIPON, deposited by RF-sputtering. The top electrode (in yellow) is made of Ti, deposited by DC sputtering (more details concerning the process flow will be given in section 3).

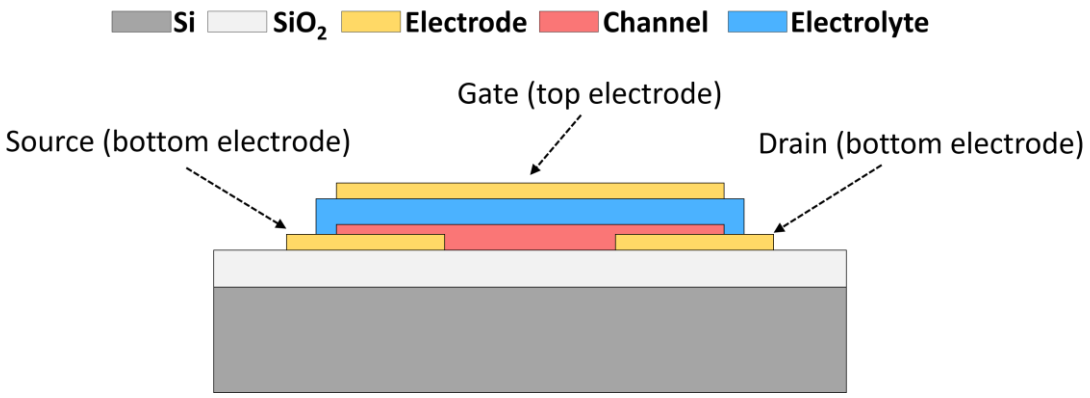

Figure II. 1: Schematic view of the synaptic transistors elaborated.

In the following, I will show a global view of the microfabrication processes (deposition, patterning, etching), with a special focus on the methods we used for our synaptic transistors elaboration.

1.2 Thin-film materials and deposition techniques

A thin film is a material layer with a thickness ranging from a fraction of a nanometer (atomic monolayer) to several micrometers. It is to state that thin films are building blocks of the nano-world as they are responsible for multiple functions in nanometric devices. They are not only part of the finished devices but also employed during wafer processing as protective films or sacrificial layers in etching or as diffusion masks.

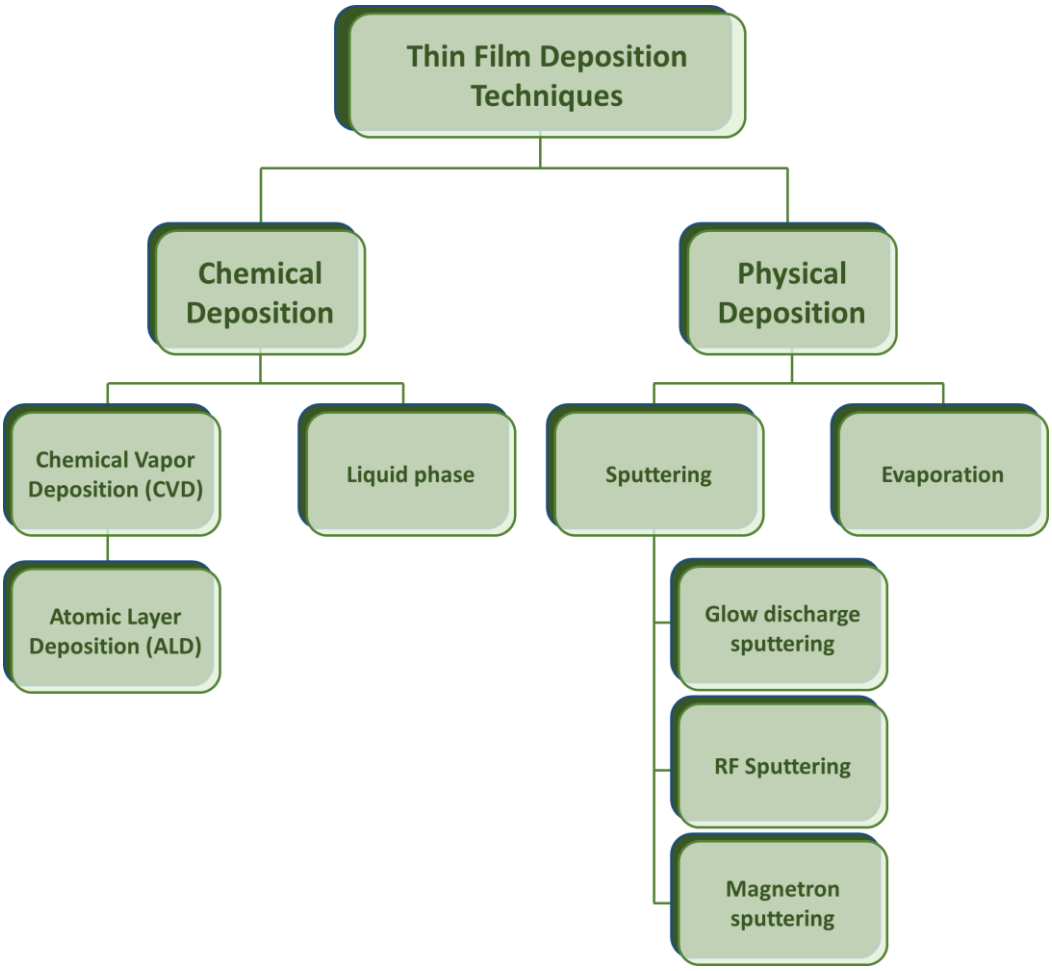

Figure II. 2: Schematic view of thin-film production methods. Adapted from [1].

Thin films are available in different electrical resistivity, for instance, conducting (Pt, Ti, W), semiconducting (Si), and insulating (SiO_2, Al_2O_3) films. For their usage, metallic films can serve as electrodes to make contact with non-conducting parts of the devices. Semiconducting layers can change under different conditions to act as a switch in some applications. Insulating layers can block the flow of current as a dielectric in field effect transistors, and they can be used as passivation to protect components from environmental factors. Electrolyte layers are electrical insulators, but they can conduct mobile ions (Li^+, Na^+, O^{2-}, etc.). These films are widely used in energy and information storage applications.

There is a wide range of microfabrication techniques that can deposit these thin functional layers using physical or chemical ways (See Figure II. 2). While physical deposition of films relies on the material extraction from a source using physical excitation (heating, bombarding, etc.) and coating it on a substrate, chemical deposition methods use chemical reactions and their products to form desired films.

In my thesis, we wanted to produce controlled, uniform solid-state thin films with a thickness that varies from a few tens of nanometers on Si wafers, hence we consider only the sputtering technique for Physical Vapor Deposition (PVD) and atomic layer deposition (ALD) technique for Chemical Vapor Deposition (CVD).

Many thin-film properties (resistivity, refractive index, density, thermal expansion, crystal orientation) are thickness dependent and can be monitored during deposition steps. In this part, I will discuss the principles of selected deposition techniques in microfabrication and how different parameters are employed to control the characteristics of the grown films.

1.2.1 Physical vapor deposition (PVD)

Physical vapor deposition is the semiconductor industry's primary method for metallic thin-film deposition. PVD is a versatile deposition technique allowing depositing metals, metal alloys, and compounds like Ti, W, Au, ZnO, AlN, and so on. The deposition rate is fast, ranging from 1 – 100 angstrom per second. The principle of PVD is material ejection from a condensed target material, transport in a vacuum in a vapor phase, and then to the substrate in the condensed phase again (Figure II. 3).

Based on the types of excitation, we can generally divide the PVD method into various subclasses, as described in Table 1 below.

Table 1: PVD classification by excitation types

Deposition method	Excitation type
Thermal evaporation	Resistive heating
E-beam evaporation	Electron beam heating
Sputtering	Argon ion bombardment

Evaporation, in general, is a standard method for depositing thin films. The source material is excited to a threshold and evaporated in a vacuum environment. The vacuum assists these vapor particles in being transferred directly to the target substrate, where they change phase to become a solid-state film. Evaporation is usually used for the deposition of metal layers. The evaporation deposition technique, however, offers a poor step coverage of evaporated films because of the directional nature of the ejected material. Heating and rotating the substrate can help increase the coverage area of the film, but they cannot form a continuous film for an aspect ratio greater than 1 (with $AR = \frac{step\ height}{step\ width}$). Therefore, a method for less directional coating is required. Sputtering is a primitive

alternative to evaporation, and its working mechanisms will be presented in the next section.

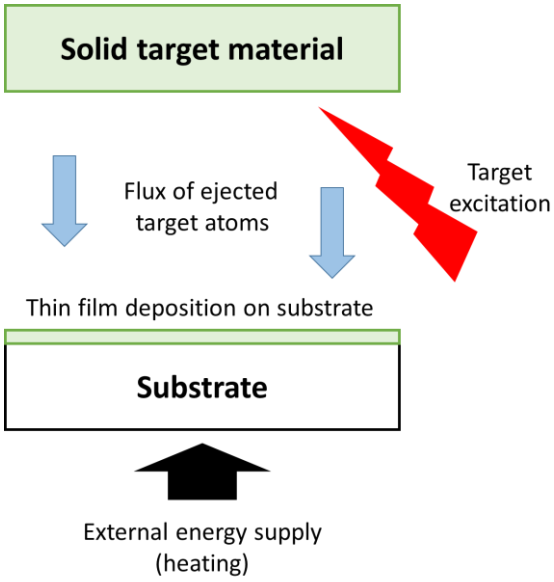

Figure II. 3: Physical vapor deposition working principle. Adapted from [2].

Sputtering

Sputtering (used to deposit the electrodes, active material and the electrolyte in my thesis) is an essential PVD technique. In principle, the excitation comes from Argon ions (Ar^+) (generated from a glow discharge plasma), which will hit the negatively biased target and slow down by collisions. Upon the impact, one or more target atoms are ejected towards the substrate wafers in vacuum. The schematic of DC and AC sputtering methods is presented in Figure II. 4.

The glow discharge is produced by an applied electric field between two electrodes in a flowing gas (usually inert) at low pressure. The gas breaks down to conduct electricity when a certain minimum voltage is reached. Ions of the plasma are accelerated towards the target under the influence of a large electric field. When the ions reach the target, atoms are ejected from the target into the plasma, where they are carried away and then deposited on the substrate. This type of sputtering is called DC sputtering (Figure II. 4.a). As the plasma ions hit the target, they become neutralized and then return to the process as atoms. In the case the target is an insulator, this neutralization process will positively charge the target surface. As a result, bombarding ions are partially repelled and the sputtering process will be gradually halted. To facilitate the process, the polarity must be reversed to attract enough electrons from the plasma to eliminate surface charge. This

periodic switch of polarity is done by applying a radio-frequency (typically 13.56 MHz) voltage on the electrodes, thus the name RF sputtering (Figure II. 4.b).

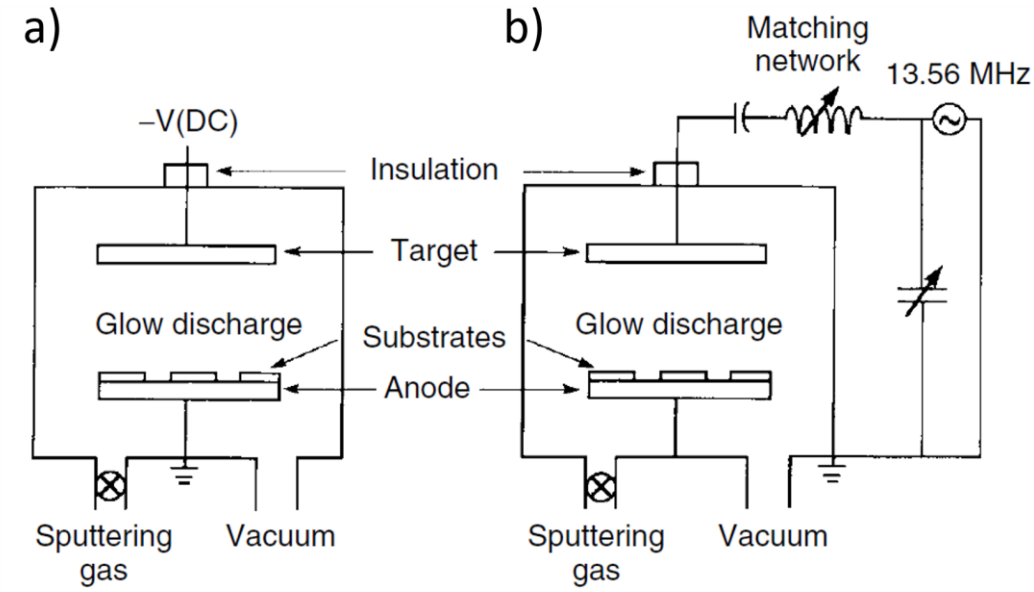

Figure II. 4: Schematic illustration of sputtering reactors using a) DC voltage source, and b) AC voltage source to generate plasma [2].

In order to increase the efficiency of the sputtering process, it is common for the sputtering source to have some magnetic confinement through a magnetron source. The electrons are driven inside a magnetic field so that they have more chance of undergoing an ionizing collision, thus enabling the plasma to be operated at a higher density. This type of sputtering is called "magnetron sputtering" and it can be used with DC or RF sputtering.

As an example, RF-sputtering Lithium phosphorus oxynitride (LiPON) is a very common solid-state electrolyte [3], [4], which we used in our elaborated synaptic transistors. The LiPON film is deposited by sputtering a Li_3PO_4 target in a nitrogen plasma (Figure II. 5). Nitrogen is incorporated into the layer, leading to the formation of LiPON. The film now improves its stability and ionic conductivity, enhancing the cross-linking of the phosphate chains [5]. Furthermore, deposition parameters such as RF power and N_2 flow can affect the properties of this thin film. It is observed that with an increase in RF power, ionic conductivity is increased and for increased nitrogen flow, there is an increased ionic conductivity for 10 to 30 sccm but reduces for higher N_2 flow of 40 sccm. This amorphous ion-conducting layer has been deposited by other techniques, such as electron beam evaporation and pulsed laser deposition [6], [7]. However, LiPON synthesized by RF sputtering has the highest ionic conductivity of $3.3 \times 10^{-6} S/cm$. In addition, RF sputtering offers high repeatability for multi-elemental compounds, formation of pinhole-free films with good contact, and high particle energy leading to dense layers [4].

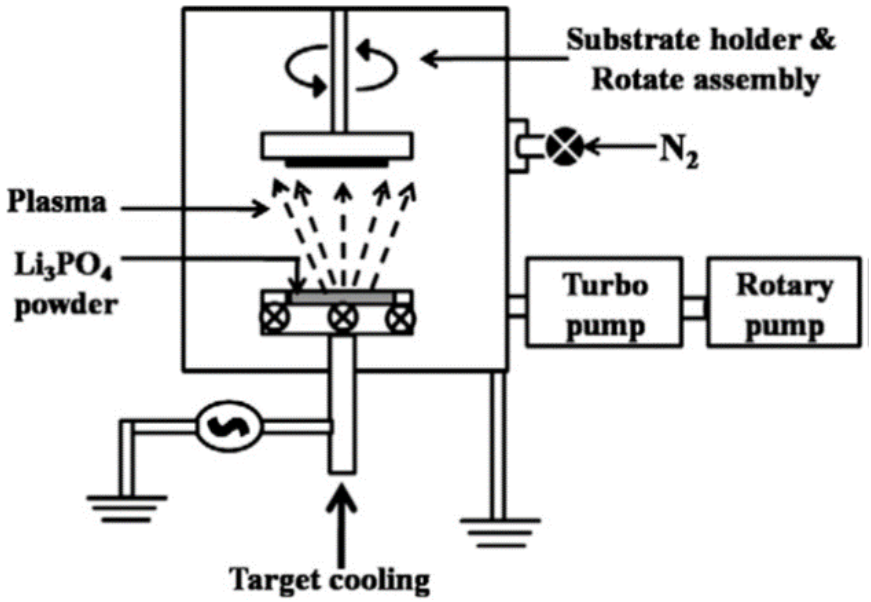

Figure II. 5: Schematic view of magnetron sputtering system using Li_3PO_4 powder target for LiPON deposition [4].

1.2.2 Chemical vapor deposition (CVD)

Chemical vapor deposition is a general name for deposition techniques that occur due to a chemical reaction among reactants in the vicinity of the substrate. In principle, reactant gases (precursors) are pumped into a reaction chamber (reactor). Under the right conditions of temperature and pressure, reactants undergo a reaction at the substrate. One of the products of the reaction gets deposited and accumulated on the substrate. The volatile by-products are pumped out with the gas flow. In several CVD types, the chemical reactions take place simultaneously in the gas phase and on the substrate surface, which complicates the deposition processes, potentially leading to the loss of film conformality. A solution to such problems is a form of CVD where the gas phase precursors are introduced to the system separately, eliminating the gas phase reactions. This form of CVD is called atomic layer deposition (ALD). We used ALD for the TiO_2-based channels in our synaptic transistors, hence it is described in more details in the following.

Atomic layer deposition (ALD)

Atomic layer deposition has become recognized as a reliable method to deposit dielectric (for example, Al_2O_3, TiO_2, etc.) and metal (TiN) thin films in the microelectronics industry. This technique is able to coat uniformly ultrathin films for narrow and three-dimensional structures with complex surfaces, for example, high-k dielectric gate materials in MOSFET [8] or energy storage (supercapacitor and battery) development [9], [10].

For each ALD cycle, there are at least two half-cycles, as illustrated in Figure II. 6. The half-cycles consist of an introduction of gaseous precursors or co-reactants, subsequently followed by purge steps with inert gas to remove any unreacted precursors or volatile byproducts. The reactions between gaseous precursor molecules and substrate groups are self-limiting, thus, the film thickness growth per cycle is ideally the same. Therefore, by repeating a certain number of ALD cycles, the targeted thickness of depositing film can be achieved.

There are different parameters that one should consider for the development of an ALD process. The precursors and co-reactants for a desired film's material composition are selected so that they are reactive toward the substrate surface groups. The target chemical composition of the film after growth is consistently verified by X-ray photoelectron spectroscopy (XPS), Rutherford backscattering spectroscopy (RBS), or a quick check by refractive index. The thickness control is essential in this deposition technique because it can provide the amount of material growth in each cycle. This characteristic is referred to as growth per cycle (GPC). GPC can be monitored using spectroscopic ellipsometry *in situ* by following the depositing film during the process or *ex-situ* by measuring the deposited film after a number of cycles. Uniformity of the deposited film (thickness, resistivity, chemical composition, etc.) over a large substrate can be assessed manually or using an automated mapping stage.

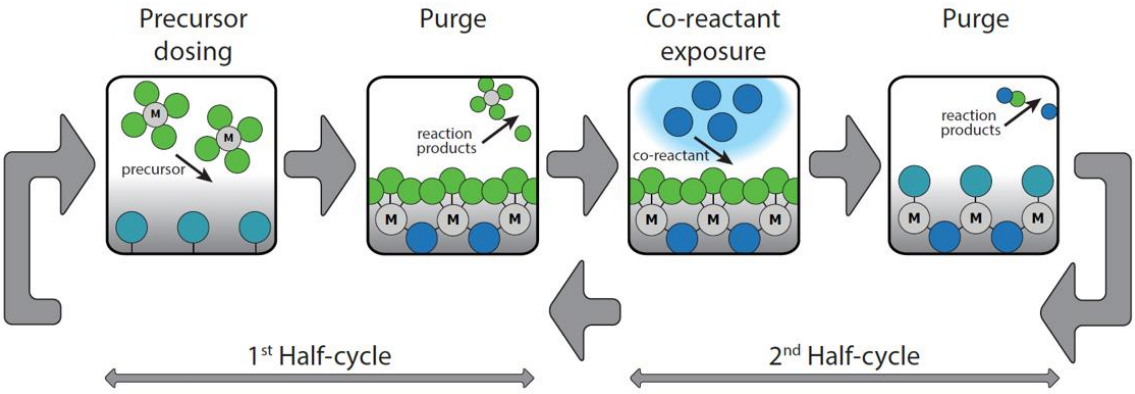

Figure II. 6: Illustration of a ALD cycle with two half-cycles. Precursor and co-reactant doses are introduced at the beginning of each half-cycle and followed by purge steps, leading to self-limiting film growth. "M" represents the metal center, blue reactant atoms are typically O, N, S, etc., and precursor ligands are colored green [11].

ALD of titanium dioxide – TiO_2 has been realized for a wide range of applications from device passivation coat to photovoltaic and catalyst fabrication. In my thesis, ultrathin layers of TiO_2 are deposited to serve as the electrical bridge (channel) between the source and drain electrodes. The deposition tool (ALD Savannah) is placed inside a glove box to minimize air exposure to deposited film (Figure II. 7.a). The thin film is grown with TDMAT

(tetrakis (dimethylamino) titanium – $Ti(N(CH_3)_2)_4$) and water precursors. In the first half of the cycle, pre-heated at under 100 ˚C, TDMAT precursor is introduced into the wafer vicinity, leading to the reactions with Hydroxyl (OH) groups at the surface. The remained TDMAT and the reaction products will be purged by a gas flow to conclude the half cycle. In the second half, H_2O precursor is pumped into the chamber, and the water molecules will react with surficial TDMAT, creating new OH groups on the substrate and bonding Ti atoms with Oxygen bridge. The unreacted H_2O and the byproducts are then purged, thus, completing a cycle. At the end of the deposition, the film is inspected by ellipsometry mapping (Figure II. 7.b).

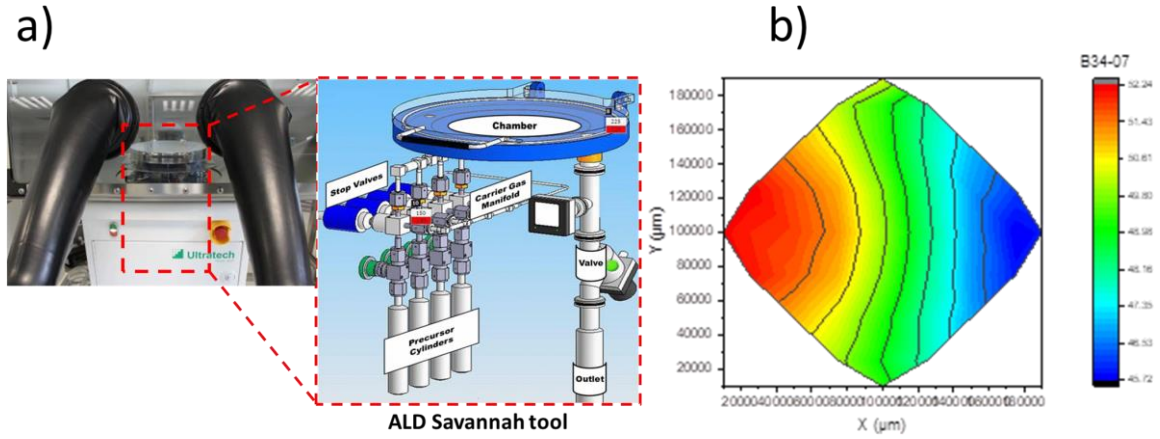

Figure II. 7: a) ALD deposition tool and the schematic of the parts. Adapted from [12]. b) A thickness mapping by ellipsometry on 50 nm TiO_2 film grown on SiO_2.

1.3 Pattern generation

The microstructure designs are transcribed to physical structures by pattern generation tools. Such machines must obtain the ability to expose single pixels in a fast manner since the designs can be of the order of millions of pixels. In the early days of microfabrication, patterns were generated by using mainly the optomechanical shuttle systems with a flash bulb, yielding limited linewidths resolution. Drawing features with a focused beam of electrons, ions, or photons is considered the most precise way of delineating devices nowadays. Two main applications of direct writing lithography are considered in this section: Direct electron beam writing and optical photomask writing.

Electron beam lithography has been considered one of the most flexible methods to realize sub-micrometric devices. The versatility of EBL has been obtained thanks to the successive development of different components and elements involved in the process, including the beam generation, the system, the resist, and the operation system. While the effect of diffraction is detrimental in photolithography, it is not a concern for electron radiation. The resolving power of an optical system is defined as:

$$R = \frac{0.61 \times \lambda}{NA} \qquad \text{Eq. 1}$$

where R is the resolving power or the smallest separation of two closely-placed structures which still permits them to be distinguished as separate, λ is the wavelength of the light source, and NA is the numerical aperture of the lens. For photons, the wavelength is fixed, and it depends on the light source. For electrons, on the other hand, the wavelength can vary, and it follows the de Broglie equation:

$$\lambda = \frac{h}{mv} = \frac{h}{\sqrt{2 \times m \times e \times V_a}} \qquad \text{Eq. 2}$$

where h is the Plank's constant, m, e, v are respectively the mass, charge and velocity of electron, and V_a is the acceleration voltage. An accelerating voltage of 1V leads to a wavelength of λ = 1.2 nm, and 1000 V to λ = 0.03 nm. Thus, yielding a much more pronounced resolution compared to, for example, a violet light source with λ = 380 nm.

In spite of having the ability to draw with ultimate resolution, EBL is not recommended for high throughput microfabrication due to several drawbacks, including excessive write time, electron scattering inside resist, space charge effect, and proximity effect. Electron beam writing, hence, is mainly considered for R&D or low volume productions, while optical lithography is and will be the mainstay of large-scale microlithography (this is why this method only has been used in my thesis).

Photomask optical lithography

Instead of direct writing millions of pixels from one wafer to another, beam writers can be used to write photomasks for optical lithography. Photomasks are glass plates coated with chromium (ca. 100 nm thick). Soda-lime glass is used for larger linewidths (>3 µm), and quartz is the material of choice for micron and submicron work. The photomasks can be dark field (DF) or light field (LF) depending on the chrome coverage area over the plate. Mass microfabrication of electronic devices uses optical lithography with photomasks because of its speed: illumination through a photomask exposes up to 10^{10} pixels in a one-second exposure.

With the substrate and the photomasks ready, the lithography workflow consists of the following steps:

1. Photosensitive film (photoresist) application
2. Alignment of mask and wafer
3. Exposure of the photoresist
4. Development of patterns.

The goal of this process is to transfer the patterns of the photomask onto the photoresist film coated on a wafer. The photoresist matters will be discussed thoroughly in a following section. To begin, the mask is first set in a mask-aligner/exposure tool. It is then aligned to the desired wafer, and exposed by different radiation sources (Figure II. 8).

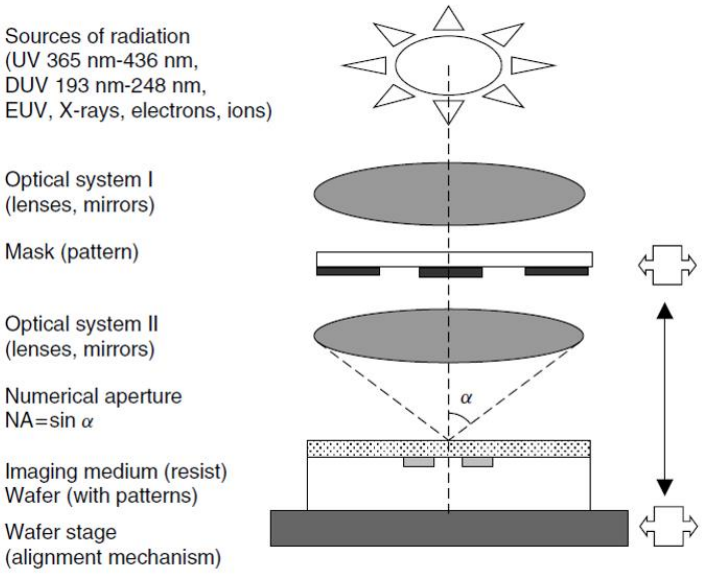

Figure II. 8: Schematic of an optical lithography aligner/exposure tool [2].

The gap between the photomask and resist-coated wafer can immensely affect the final patterns after exposure and resist development. There are two modes of lithography: contact and proximity. Contact mode lithography is done simply by bringing the mask and the wafer into intimate contact and then expose. Then, mask dimensions and diffraction at mask edges determine the resolution of transferred patterns. Extremely small patterns can be made in theory using photomasks with submicron features; however, realizing such masks is prohibitively expensive. The other mode is proximity lithography, in which a small gap, i.e., 3 to 50 µm, is kept between the mask and the wafer. The light traversing the mask is diffracted by the mask patterns, and Fresnel diffraction formulae have to be used to estimate resolution:

$$2b_{min} = 3\sqrt{\frac{\lambda}{n} \times (g + \frac{d}{2})} \qquad \text{Eq. 3}$$

where $2b_{min}$ is the minimum resolvable period, λ is the wavelength of radiation source (436 nm for mercury lamp g-line), g is the gap between mask and photoresist (0 – 50 µm), d is the thickness of photoresist (1 – 10 µm), and n is the refractive index of photoresist (around 1.6). Following the formula, we can achieve higher resolution by using extreme ultraviolet (EUV) or even X-Ray sources. The resolution generated by proximity mode is lower than

that of contact mode, but the photomasks are better conserved in this case. Both contact and proximity lithography are conducted in the same machine, with the gap being an adjustable parameter.

The previous methods of lithography are used for 1X optical systems, in which the size of generated and the original patterns are the same. Another approach to improve the performance of lithography step is to use reduction optics, marked as system II in Figure II. 8. For example with 5X reduction projection optics, the original photomask features can be made rather large, for example, 1μm for 0.2 μm final feature size. Fraunhofer far-field diffraction governs the optics of projection systems. Projection optics is often used for the step-and-repeat approach, where one die is exposed, the wafer is moved to a new position, and another die is exposed. The systems are known as steppers, and their photomasks are termed reticles. Exposing each chip (4X, 5X) is certainly longer than exposing the whole wafer (1X), but we can have a better critical dimension at a lower cost.

Since microfabricated devices are formed by building up thin films, the ability to align micro-patterns precisely with successive layers, which is termed overlay, is a critical factor in defining the resolution. The overlay is affected by lens aberrations, wafer chuck irregularities (equipment-related problems), mask pattern misplacement (mask fabrication problems), wafer alignment, or distortions on the wafer itself, such as thermal expansion or site flatness. Because of the variety of error sources in the lithography process, controlling the structures regularly with micrometrology (optical/electron microscopy) is necessary to ensure the high yield of fabricated wafers.

1.4 Lithographic photoresist pattern

The properties of photoresists as an optical material will be discussed in this subsection. Patterns of resists will serve as the etch mask for the underneath layer(s), so they strongly affect the whole lithography process. Photoresists have exposure threshold energy, finite contrast, and finite selectivity in developers. Furthermore, they are parts of the optical system in which they exhibit optical reflection, interference, and absorption properties. These features are enhanced when one patterns photoresists on topology, which is very likely in the context of semiconductor device.

1.4.1 Photoresist application

The application process starts with surface preparation. The wafer is annealed in a short time to remove moisture. Then, some substrates require a wafer-priming step known as adhesion promotion. Adhesion promotion is essential to protect wafers from cleanroom humidity variations and an equalizer for wafers with different storage times.

After the surface preparation, the photoresist is commonly applied to the wafer by spin coating (an illustration of a spin coater can be found in Figure II. 9). A quantified amount of photoresist is slowly distributed on a static or rotating wafer. The wafer rotation is then accelerated up to several thousand rotations per minute (rpm), spreading the resist uniformly over the wafer. Rotation speed can tune down the thickness of the applied resist to a specific limit, for example, from 0.5 to 5 µm. If we want to flatten further, a new photoresist is required. After the coating, the resist film typically has a residual solvent concentration of 10 - 35%, depending on the film thickness and the solvent type. To get rid of the high solvent content, the wafer is annealed for a short period on a hot plate or in an oven, which is termed soft bake. In addition to drying the substance, this process promotes the resist adhesion to the substrate and prevents photomask contamination with resist sticking.

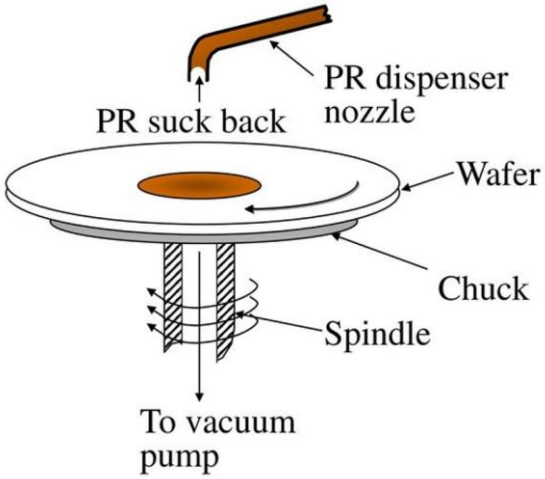

Figure II. 9: Schematic of a photoresist automatic dispenser [13].

There are other resist coating techniques: electrochemical coating, spray coating, and casting. These methods of resist covering are highly preferred on topographic surfaces, where spin coating performs poorly. The electrochemical coating requires special resist formulations, and spray is applicable to thin resists. Casting is suitable for thick resists only. Thin resists are preferred for better resolution. However, they are prone to particle defects, and pinhole density rapidly increases when resist thickness is scaled down.

1.4.2 Photoresist properties and processes

A photoresist is an organic polymer with three main components: (1) a film-forming agent or resin to control the mechanical and thermal properties, (2) a photoactive compound to determine the radiation sensitivity, and (3) a solvent to control viscosity. There are two types of photoresists: negative and positive. In a negative resist, the polymers

change from an un-polymerized state (soluble in developer solvents) to polymerized (insoluble in developer solvents) state after exposure to a light source and vice versa for a positive resist (see Figure II. 10). The terms positive and negative stem from photographic processes, in which the results of the lithographic process is either positive or negative images of the original patterns. Negative photoresists are considered tougher than positive photoresists and can withstand more rigorous etching processes.

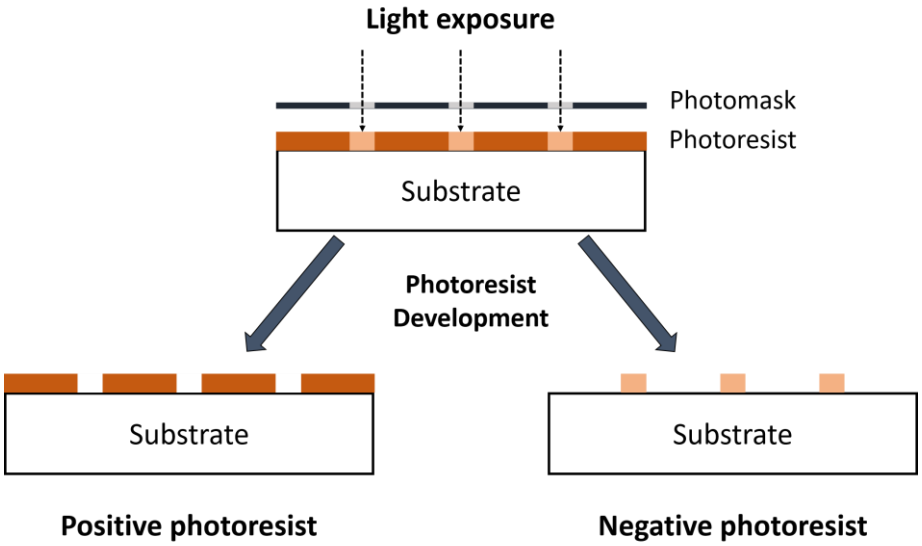

Figure II. 10: Two types of photoresist: positive and negative.

Exposure is important to consider while doing photolithography. As described above, the goal of this step is to decompose inhibitors in the resist using radiation. The theoretical calculation of exposure dose depends on different parameters such as exposure-dependent, exposure-independent absorption, and sensitivity to exposing radiation. These are called Dill parameters, and they depend on the intensity and wavelength of the light source [14]. The exposure dose or energy (normally presented in mJ/cm^2) for a unit of thickness can be found in the technical data sheet of the photoresist. With this information, the time required for each exposure session is calculated by:

$$t_{exposure}(s) = I \times Dose \qquad \text{Eq. 4}$$

where I is the light source intensity in W/cm^2. However, this dose value serves as a reference only as the lithographic conditions and tools might differ from case to case. Thus, in order to find the good exposure time for a specific process, we need to conduct a study with different exposure times and controlled experimental conditions (light intensity of the aligner, prebake temperature, moisture, etc.).

After exposure, the resist mask exists as a latent image in the photoresist: the exposed and the non-exposed areas are chemically different. The development step is

performed to dissolve the depolymerized areas from the prior irradiation step, resulting in the final desired resist structures. Aqueous alkaline developers are frequently employed to develop photoresists in microfabrication. These developers have different bases, for example, diluted sodium hydroxide or potassium hydroxide solution or on an aqueous solution of the metal ion-free organic TMAH (TetraMethylAmmoniumHydroxide). A good development process is performed in a short period resulting in minimum pattern distortion or swelling. The development process can be carried out using different methods, including immersion developing, spray developing, or puddle developing. Immersion of wafers into a solvent bath is a common method used for Si wafers. Subsequently, the wafers are rinsed with deionized water and dried out using air or N2 gas to terminate the development reactions. Depending on the demand of the next step, the developed film can go through an annealing process, termed hardbake, to increase the thermal, chemical, and physical stability of developed resist structures by cross-linking.

The lithographic patterns are then inspected for errors using microscopy before the wafer is ready for the next lithographic steps of etching, implantation, or depositions. Each step has its different properties and requirements for the resist film. For example, in wet etching, the resist has to have good adhesion to the material underneath, and it must tolerate hot, strongly acidic, or alkaline etch solutions. On the other hand, in plasma etching, the resist must be thick enough because it will be consumed/damaged in the process. The etch properties will be covered in the next part of this section.

After serving its role as a protective layer, the polymerized film has to be removed. In some cases, the photoresist can be easy to remove by being dissolved in a strong solvent like acetone. However, it is highly recommended to rinse the wafer with isopropanol to get rid of the contaminated solution before it vaporizes and creates stubborn streaks. If the photoresist has undergone a plasma etch or a hard bake, acetone cannot fully remove the polymer film; in this case, aqueous alkaline remover must be used.

1.5 Etching

In the previous part, we discuss the process of creating a protective layer with desired patterns. In this section, we will cover the etching techniques to transfer the patterns from the photoresist to the underlying materials. There are two classes of etching techniques: wet and dry. In principle, wet etching solvents remove the uncovered area of materials by turning them into soluble products, while gaseous etchants turn solid films into volatile products and purge them. Note that I use the term dry etching, as opposed to wet etching, to indicate plasma etching. However, other dry, physical etching techniques do not involve a plasma creation process. For example, ion milling and etching using XeF_2 gas. A short summary of the two techniques can be found in Table 2. Here, we can observe that both techniques have their pros and cons, and thus, selecting the right method for

each step is case-dependent. In the following parts, I will explain the working mechanisms of wet and dry etching and give examples from my thesis work.

Table 2: Comparison between wet and dry etching methods.

	Wet etching	Dry etching
Profile		
Etchant	**Chemicals**	**Ionized gas**
Controllable features	**> 2 µm**	**> Single nm**
Selectivity	**High >100x**	**Low <10x**
Throughput	**High (Batch)**	**Low (Single wafer)**

1.5.1 Wet etching

Wet etching is a common and theoretically easy method in microfabrication. The basic process is illustrated in Figure II. 11. The simplified reaction for the wet etching is as follows:

$$Solid + liquid\ etchant \rightarrow soluble\ products \qquad \text{Eq. 5}$$

The reaction is then subdivided into two main mechanisms: electron transfer for metal etching, and acid-base reaction for insulator etching. The etching rate is dependent on the surface reaction: (1) if the surface reaction is slow, the reaction determines the rate and (2) if the reaction is fast, etchant availability in the solution determines the rate.

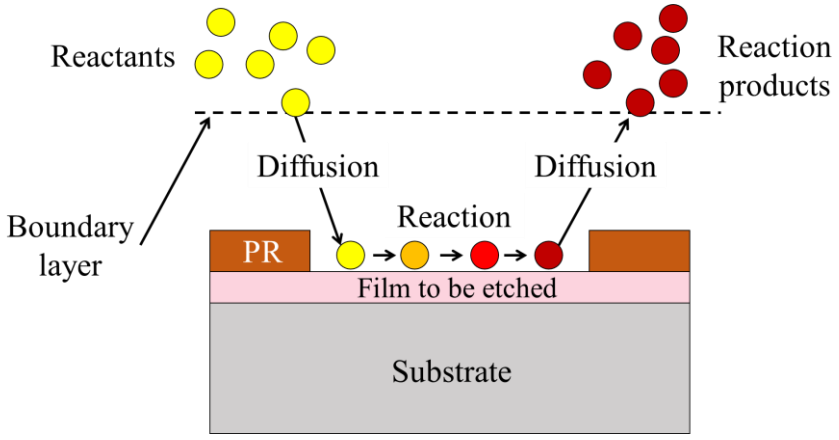

Figure II. 11: Schematic illustration of wet etching process. Adapted from [15].

Submersion-type wet etching is conducted in a tank (or bath) made of quartz in a heating and temperature control. The vessel is filled with water and chemicals, and the wafers are submersed in liquid for the required time and then transferred to a similar sink to rinse/dry and stop the reactions. A programmable single-wafer tool is an alternative to bath etching, in which, instead of immersion, the etchants are sprayed to rotating wafers through stationary nozzles. The rinsing and drying steps follow the etching in the same chamber. Thus, the single-wafer tool has more advantages over tanks in controllable processes.

Platinum (Pt) is an important current conductor in my technology. This noble transition metal, however, can be tricky to etch. The film can be etched using a mixed solution of water, nitric acid, and hydrochloric acid following the ratio of 4H2O:1HNO3:7HCl. The etching solution is heated to 57 ˚C inside a glass tank using the water heating system of a chemical bench. The wafer is submerged in the tank for 7 minutes and 30 seconds for Pt to be completely removed, and then it is rinsed using DI water. The time used here is approximate because this etching step happens at a high rate inside an acidic and corrosive solution, so it is hard to control manually with bare eyes. Thus, most of the time, the patterns will be slightly over-etched after this step. Several etching etchant compositions are presented in Table 3.

Table 3: Wet etchants for some materials.

Materials	Wet etchants
Al	H_3PO_4:HNO_3:H_2O (80:4:16), water can be changed to acetic acid
W	H_2O_2:H_2O (1:1)
Pt	H_2O:HNO_3:HCl (4:1:7)
Cu	HNO_3:H_2O (1:1)
Ti	HNO_3:H_2O_2:H_2O
Au	KI:I_2:H_2O; KCN:H_2O

Most wet etchants result in an isotropic profile due to the same rate of etching reactions in all directions (horizontal and vertical). Such phenomenon creates an undercutting with the same dimension of the etched depth, preventing this method from creating sub-micrometric features. As opposed to an isotropic etching profile, an anisotropic etching profile is generated by etchants that react more favorably in the vertical direction than lateral direction, creating vertical/near vertical structures of the sidewall. Anisotropy profile is a result of directional ion bombardment in the plasma reactor. This etching profile is commonly observed in dry etching processes, which will be discussed in the following subsection.

1.5.2 Dry etching

Dry or plasma etching has been a popular method for more than 40 years. It has always been able to transfer with high precision those lithographic patterns printed in photoresist to the underlying materials. There are three main types of dry etching with different properties: physical/sputtering, reactive ion etching (RIE), and chemical plasma etching. A brief comparison of the three types can be found in Figure II. 12.

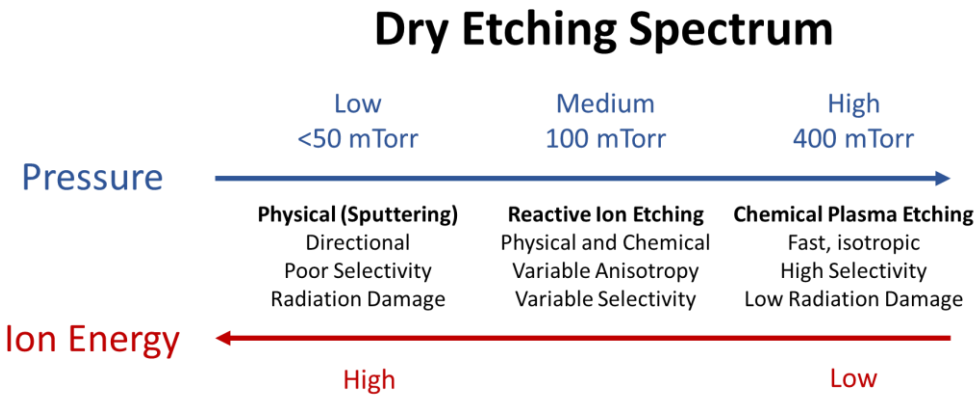

Figure II. 12: Dry etching spectrum with pressure and ion energy as parameters.

Plasma etching is conducted in a vacuum chamber by reactive gases excited under radio frequency (RF) fields. The excited and ionized species are equally important for this process. Excited molecules like CF^*_4 are very reactive, and ionic species like CF^+_3 are accelerated by the RF field, and they impart energy directionally to the surface. Plasma etching is, thus, a combination of chemical (reactive) and physical (bombardment) processes. The subsequent microscopic steps of RF plasma etching are presented in Figure II. 13.

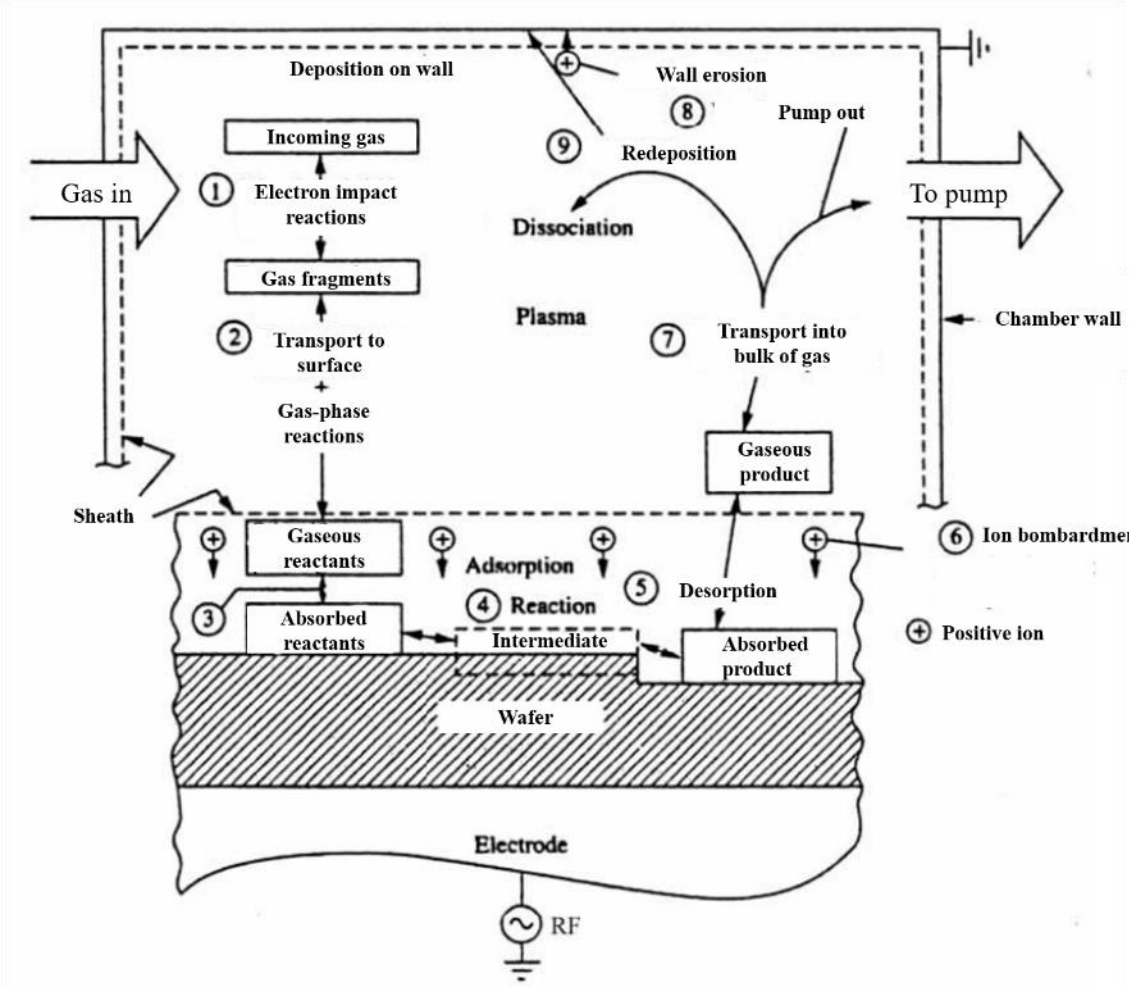

Figure II. 13: Processes that happen inside a plasma etching system [16].

Ion bombardment supplies energy to horizontal surfaces. These surfaces experience ion-induced desorption, ion-induced damage, and ion-activated chemical reactions. Thus, etching processes are accelerated or favored. Sometimes etchant gases (together with resist erosion products) form films on the sidewalls, and these films prevent etching laterally. Sidewalls do not experience ion bombardment, and, therefore, an anisotropic etching profile is a critical feature of dry etching. However, plasma etching suffers from several disadvantages of thin-film selectivity, surface ion damage, and residue formation on the sidewalls that require engineering solutions, including increased etching/additive gas content [17].

In many applications, the choice of wet versus plasma etching is a question of convenience: certain equipment or etch bath is available, or some suitable masking material is handy. When sloped etch profiles are required, or when undercutting is needed, wet etching with an isotropic profile must be used. On the other hand, dry etching with the anisotropic profile is preferred when vertical walls are needed or etching with solution-sensitive bottom layers.

2 THIN-FILMS AND MATERIALS CHARACTERIZATIONS

2.1 Microscopy and visualization

Coming out of microfabrication, the patterns, the thin films, or the devices usually have micrometric and sub-micrometric features that eyes and normal magnifiers cannot resolve. To characterize and control the quality of these processes, measurement tools should be ready to observe even smaller details.

The first characterization tool that comes in handy is optical microscopy. Optical microscopy resolution is similar to wavelength, that is, in the micrometer range. This is useful in many applications because we can always observe the sample in its real colors with no need for pre-treatment. Furthermore, the system is fast and adaptable to all kinds of sample systems, in any shape or geometry. However, due to the light diffraction limit, this type of microscopy can only resolve down to a few hundreds of nanometers. The electron-based tool can solve this problem. Scanning electron microscopy (SEM) has a minimum resolution down to 10 nm, which makes it applicable to almost all microfabricated structures. With this resolution power, SEM can provide details of electronic devices such as deposition step coverage and interlayer defect inspections via their cross-sections. Transmission electron microscopy (TEM) provides ultimate image resolution down to the atomic scale.

In my thesis, SEM and TEM techniques have often been used. Thus, in this part, I will cover the fundamentals of these electron microscopy techniques and their applications in microfabrication imaging.

2.1.1 Scanning electron microscopy (SEM).

Electron waves can be used in imaging. By accelerating the electrons into a high-energy beam (with high voltage), the generated wavelength is far shorter than that from photon light sources. Therefore, the diffraction limit is not a big issue for electron microscopy. In an SEM system, an electron beam is focused into a small probe and scanned in a raster pattern across the surface of a specimen. Several electron-matter interactions with the sample result in the emission of electrons or photons as the electrons penetrate the surface. These emitted particles can be collected with different detectors to yield valuable information about the materials (Figure II. 14.a).

SEM can produce multiple signals, including secondary electrons, back-scattered electrons (BSE), characteristic X-rays, light (cathodoluminescence), and transmitted electrons. Secondary electrons are electrons generated as ionization products. Secondary

electrons result from the inelastic collision and scattering of incident electrons with specimen electrons. They are helpful in revealing the surface structure of a material with high resolution. Another type of electron that contributes to the visual inspection of the sample surface is BSE. Back-scattered electrons are a result of an elastic collision and scattering event between incident electrons and specimen nuclei or electrons. Back-scattered electrons can be generated further from the surface of the material and help to resolve topographical contrast and atomic number contrast with a resolution of >1 µm. Furthermore, BSE is used in analytical SEM along with the spectra made from the characteristic X-rays to provide valuable information about the distribution of different elements in the sample.

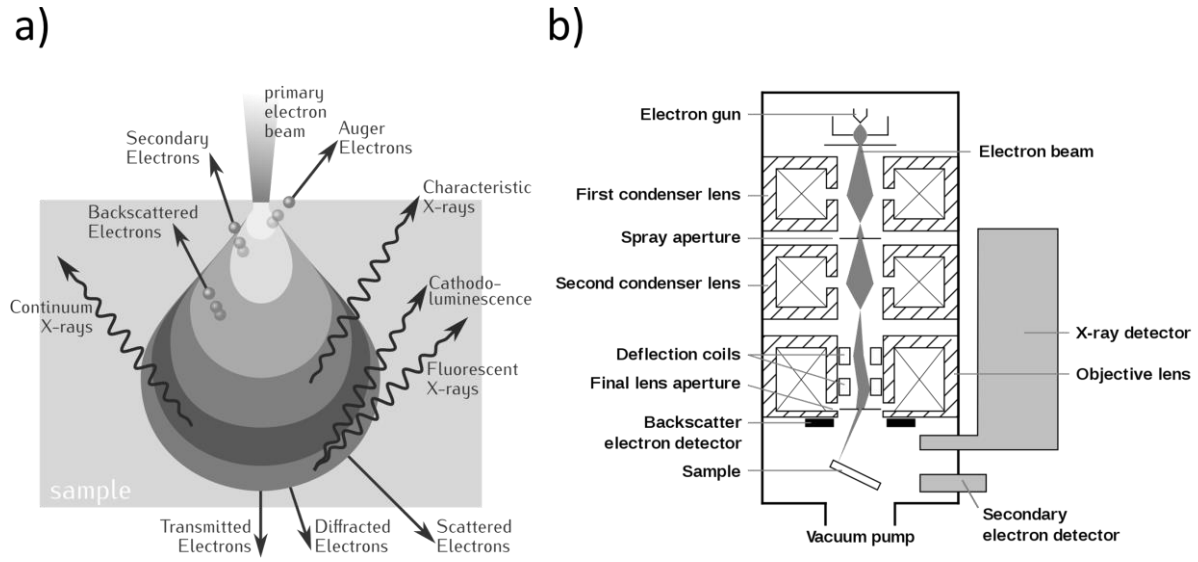

Figure II. 14: a) Illustration of electron-matter interaction volume and the generated signals [18]. b) Schematic of scanning electron microscopy instrument [19].

Characteristic X-rays are emitted when the electron beam removes an inner shell electron from the sample, causing a higher energy electron to fill the shell and release energy. The X-ray signal can originate from further down into the surface of the specimen surface and allows for the determination of elemental composition through EDS (energy dispersive x-ray spectroscopy) analysis of characteristic X-ray signals, which will be discussed in the physiochemical characterization part following.

The SEM instrument comprises two main components, the electronic console, and the electron column. The electronic console provides control knobs and switches that allow for instrument adjustments, while the electron column is where the electron beam is generated, focused, and scanned across the surface of a specimen. A schematic of the electron column can be found in Figure II. 14.b.

In principle, the electron gun generates free electron beams by thermionic emission from a tungsten filament at ~2700 °C. The voltage applied to accelerate electrons are

adjustable in the range of 200 V to 30 kV. After that, the condenser lenses guide the beam to converge and pass through a focal point at which the electron beam is focused down to 1000 times its original size. The apertures (spray and final lens) reduce and exclude extraneous electrons in the lenses and decrease the beam spot size at the specimen. A small spot size will allow for an increase in resolution and depth of field with a loss of brightness in imaging. Images are formed by rastering the electron beam across the specimen using deflection coils inside the objective lens. The astigmatism corrector is located in the objective lens and uses a magnetic field in order to reduce aberrations of the electron beam. The lower portion of the column is called the specimen chamber. Specimens are mounted and secured onto the stage, which is manually controlled by a goniometer for x, y, and z translation, 360° rotation, and 90° tilt.

SEM is commonly used to inspect the appearance of the device from the top view and the details of the composition of thin films from the cross-section observation. Here, the surface of the grid-like array architecture of a transparent thin film battery is examined using SEM with BSE enhancement (Figure II. 15.a). The cross-section of the battery illustrated in Figure II. 15.b demonstrates good conformal coverage of the top electrode and the lateral etching control of $LiCoO_2$ and LiPON films. [20].

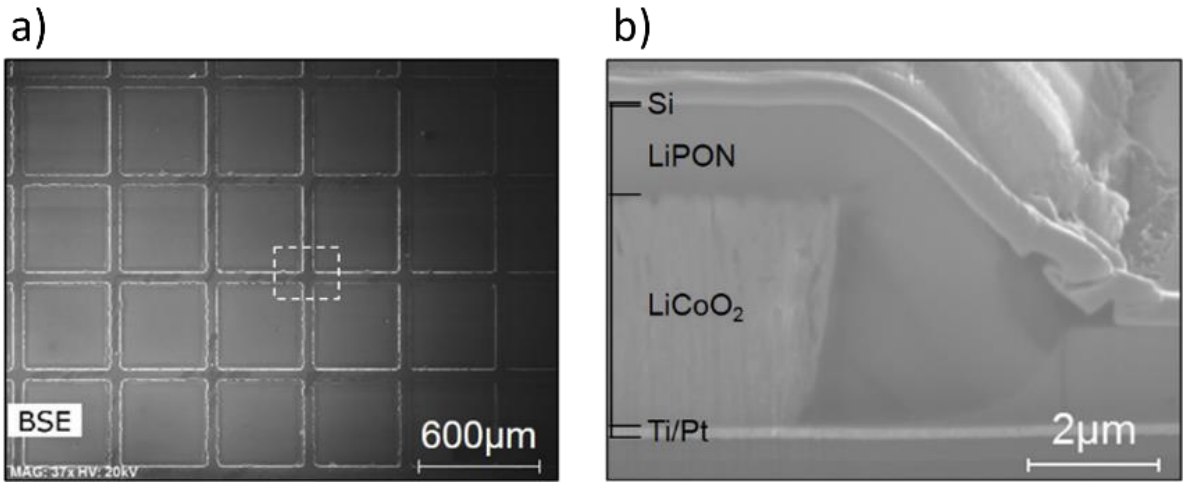

Figure II. 15: SEM visualization of a thin-film battery. a) Top grid structure view using BSE, and b) Cross-section SEM of the device [20].

Even with high power resolution, sample imaging with SEM requires attentive control to avoid image disturbances such as lack of sharpness, low image quality, noises, and image distortion & deformation, especially for samples constructed by ultrathin layers of tens of nanometer thickness or less. In this case, a more powerful technique is employed, which is transmission electron microscopy.

2.1.2 Transmission electron microscopy (TEM)

Transmission Electron Microscopy (TEM) is widely used for the study of semiconductor devices due to its ability to offer an atomic resolution. TEM can provide information about the microstructures, crystal structures, and defect concentrations in a sample. Diffraction patterns reveal crystal symmetry and lattice parameter data, and the elemental composition of the specimen can be obtained from TEM-coupled techniques such as energy dispersive X-ray analysis, Auger spectroscopy, and Electron Energy Loss Spectroscopy (EELS).

A beam of electrons (typically 100- 1000 keV in energy) is generated in an electron gun and sent down an evacuated column through a series of lenses. The specimen itself is inserted through an airlock. The transmitted electrons, which may be unscattered (direct), elastically scattered (diffracted), or inelastically scattered - are used to form an image, which is observed and analyzed by various detection devices.

The TEM optics are built so that the full thickness of the specimen is simultaneously in focus, and no refocusing is required when switching between image collection systems. The TEM is usually operated in one of two fundamental modes: imaging mode or diffraction mode with the schematic ray diagrams illustrated in Figure II. 16.

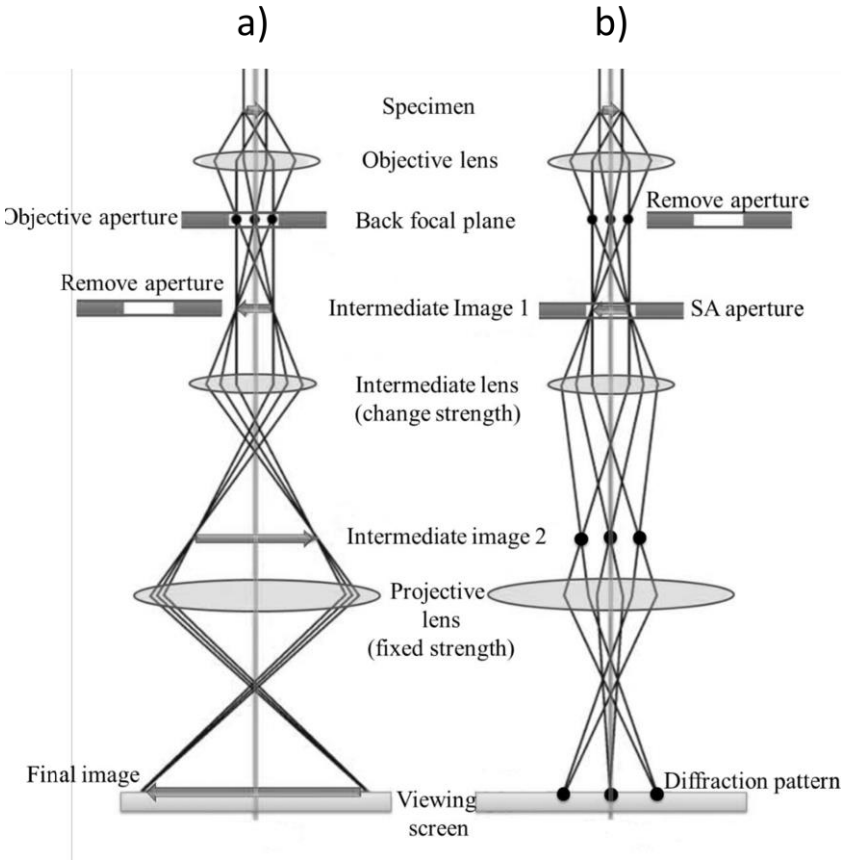

Figure II. 16: Illustration of the transmission electron microscopy operation modes. a) Imaging mode. b) Diffraction mode [18].

TEM imaging mode (Figure II. 16.a) functions based on two kinds of image contrast: amplitude contrast and phase contrast. Amplitude contrast is the combined result of mass-thickness contrast and diffraction contrast which alters the scattered electron amplitude and thus the intensity. Phase contrast comes from the phase change of the electron wave as it is transmitted through a sample. This contrast mechanism is hard to interpret due to many factors such as the objective lens focus and astigmatism, the orientation, and the changes in the thickness of the sample. All of these factors can be exploited to produce atomic resolution TEM images arising from the interference between different diffracted beams and the unscattered beam.

TEM diffraction mode (Figure II. 16.b) functions based on the interaction of the electron wave and the microstructure of the sample. When electron waves propagate across the periodic crystal planes in the sample, elastic scattering of the electron wave is generated at various different angles according to Bragg's Law, forming an electron diffraction pattern in the back focal plane. A selected area (SA) aperture is inserted in the first intermediate image zone to define the diffracting area. Samples with a single crystal will generate spot diffraction patterns, whereas polycrystalline materials will create ring diffraction patterns on the viewing screen. The SA electron diffraction (SAED) technique can provide information about the material crystallography, such as the crystallographic lattice spacing, the crystal growth direction, and the exact crystalline phase. High-Resolution TEM (HRTEM) images are very advantageous for studying the microstructures at the atomic resolution of samples. In principle, no objective aperture is used, and both the scattered and unscattered electron waves are employed to produce HRTEM images.

In the field of nano-devices, a cross-section view of the stack can provide useful information on layer appearance, interlayer diffusion, and film defects. A synaptic transistor with a schematic view of the device can be found in Figure II. 17.a [21]. The gate stack of the transistor comprising of 25 nm $WO_{2.7}$ channel, 40 nm $ZrO_{1.7}$ electrolyte, and 20 nm GdO_x is confirmed with a TEM observation (Figure II. 17.b).

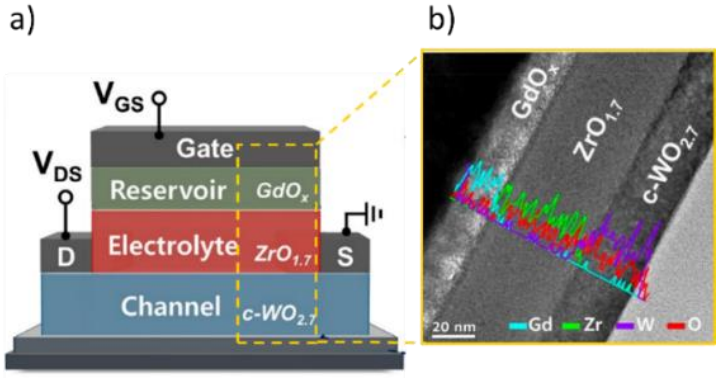

Figure II. 17: TEM imaging of a synaptic transistor. a) Schematic view of the device. b) Cross-section TEM and EDS measurements of the gate-stack. Adapted from [21].

2.2 Physicochemical analyses

2.2.1 Raman spectroscopy

Raman spectroscopy is a widely used optical technique in chemistry and material science for materials identification and the characterization of their properties. This spectroscopic technique studies the inelastic scattering of monochromatic light, known as Raman scattering, named after the physicist C. V. Raman.

An energy diagram illustrating the processes involved in light-matter scattering is shown in Figure II. 18.a. The majority of photon scattering are elastically scattered (or Rayleigh scattering), resulting in scattered photons whose energy is conserved (same frequency, wavelength, and color as the incident photons) but with different traveling directions. The intensity of Rayleigh scattering is around 0.1% to 0.01% compared to that of a radiation source. Scattered photons whose energy is shifted after the interaction, on the other hand, are naturally much rarer (approximately 1 in 1 million). This phenomenon is called inelastic scattering or Raman scattering. In the Raman spectroscopy experiment, a visible laser will be used as the monochromatic light source. As the laser photons interact with molecular vibrations, the energy of the laser photons will be shifted down or up (Stokes or Anti-stoke Raman scattering). This energy shift provides information about the vibrational modes in the materials (fingerprints).

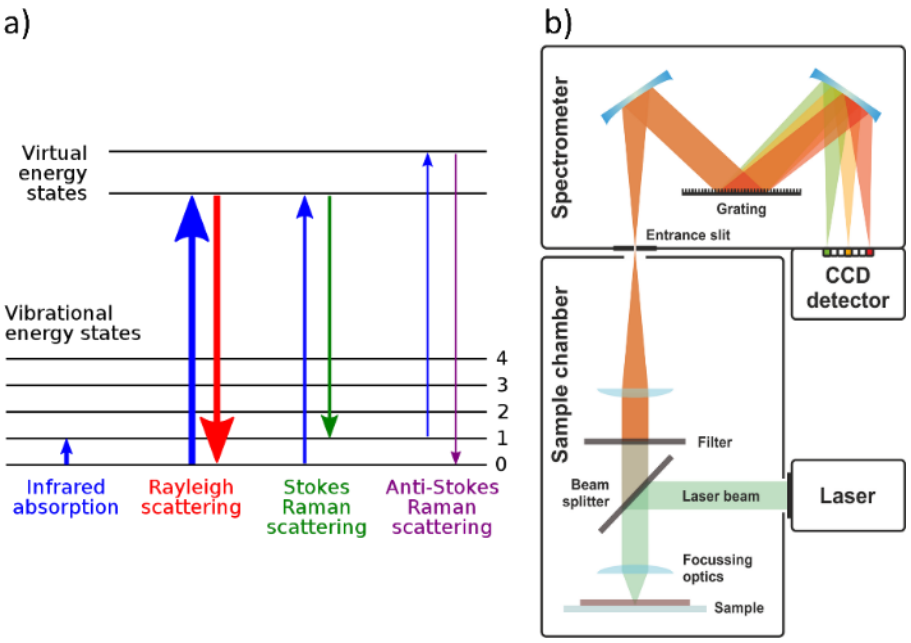

Figure II. 18: Raman spectroscopy. a) Energy diagram of Rayleigh scattering [22]. b) Diagram of a Raman spectroscopy setup [23].

A schematic view of a Raman spectrometer is presented in Figure II. 18.b. We begin by focusing a polarized monochromatic laser into a specimen, and the scattered light at 90° to the incident laser beam is collected. The scattered light is guided to the spectrometer, where it is dispersed by a high-resolution monochromator (grating component) and then detected by a CCD camera to obtain the scattered frequency spectrum. The scattered light yields a very low count ($<10^{-7}$ of the incident power). Thus, monochromators with excellent undesired photons rejection and sensitive detectors are required for the experiment. Signals acquired from the CCD will be fed to a built-in program to show the photon counts directly as a function of their Raman shift.

The Raman shift of the scattered photons is presented in wavenumbers, whose unit is length inversed. To convert from spectral wavelength to wavenumbers, we can use a formula as follow:

$$\Delta\tilde{\nu} = \frac{1}{\lambda_0} - \frac{1}{\lambda_1} \qquad \text{Eq. 6}$$

where $\Delta\tilde{\nu}$ is the Raman shift, λ_0 and λ_1 are the wavelengths of the incident radiation and the scattered one, respectively. The Raman shift is normally presented in [cm^{-1}], thus, the wavelengths are converted from [nm] to [cm] by default.

Raman spectrum measured on our ALD-deposited TiO_2 film is demonstrated in Figure II. 19. The vibration dynamics of anatase phase TiO_2 are schematically illustrated in Figure II. 19.a, in which the arrows show the actual displacement of the corresponding atoms [24]. We deposited 50 nm TiO_2 film on a 200 mm Si wafer. The as-deposited film is of amorphous phase. Subsequently, we annealed the whole wafer under air at 400 ˚C for 30 minutes using an RTP tool to transfer this film into the anatase phase. Raman scattering on the layer was excited by a 532 nm laser, and the spectrum was acquired using the inVia spectrometer (Renishaw, UK) coupled to a Leica microscope at room temperature. The spectrum reveals all the vibration modes available in the TiO_2 film, and they correspond correctly to the theoretical calculation of anatase phase. Thus, it proves the phase transition of an as-deposited 50 nm amorphous TiO_2 film to anatase TiO_2 by thermal annealing (Figure II. 19.b). As a remark, we have not observed the described phase transition for 10 nm amorphous TiO_2, which is integrated as the channel material for first generation of SynT, with the same thermal treatment.

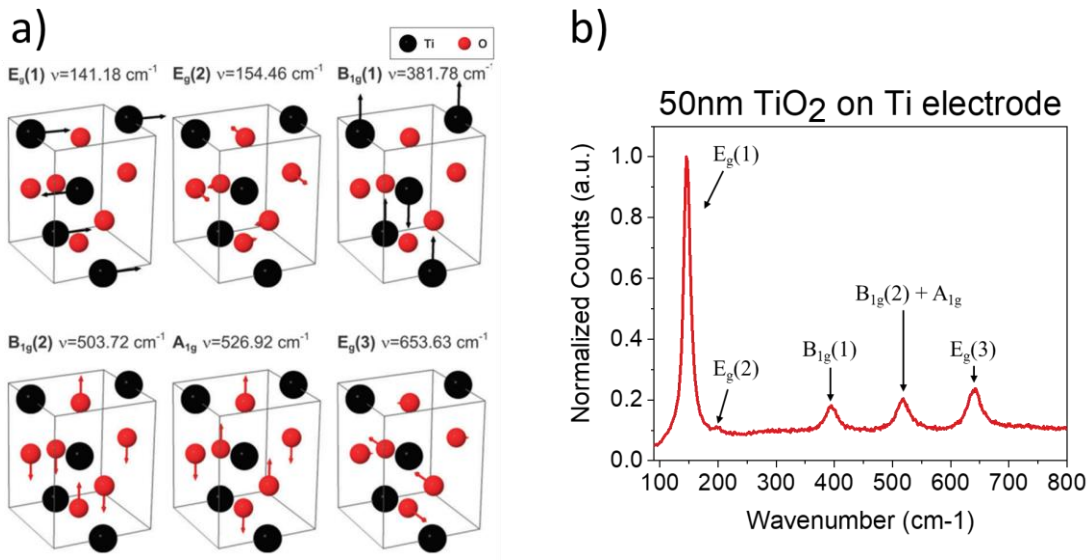

Figure II. 19: Raman spectroscopy experiment on anatase-phase TiO_2. a) Scheme of Raman active atomic vibrations in anatase TiO_2 (the arrows represent the amplitudes of vibrations) [24]. b) Raman spectrum of 50 nm TiO_2 anatase measured at 25 ˚C.

2.2.2 Energy dispersive X-ray spectroscopy (EDX-EDS)

Energy dispersive X-ray spectroscopy (EDX or EDS) is a powerful technique when it comes to the elemental or chemical analysis of a solid-state sample. EDS functions based on the X-ray spectrum emitted because of the interaction of a focus electron beam and a sample. The resulted X-ray carries the information on the atomic structure of the elements in the specimen, yielding a unique set of peaks on its emission spectrum. EDS can provide either qualitative analysis (identification of the elemental lines in the spectrum) or quantitative analysis (determination of the concentrations of the elements based on the Standards of known composition). The range of detectable elements is wide, from Beryllium (atomic number = 4) to Uranium (atomic number = 92), making it a standard tool for micrometrology. EDS can produce element distribution images or maps when coupled with SEM, which is extremely versatile for analyzing the interfacial processes on cross-sections of thin-film devices (defects, anomalies, diffusion, etc.).

In principle, to obtain the emission of characteristic X-rays from a sample, a focused beam of electrons is guided into the specimen. At the rest state, an atom contains ground (unexcited) state electrons in discrete energy levels (electron shells) bound to the nucleus. The incident beam provides enough energy to excite an electron in an inner shell, ejecting it from the shell while creating an electron hole in the atom. An electron residing on an outer, higher-energy shell then fills the hole, and the energy difference between the higher and lower shells is released in the form of an X-ray. An energy-dispersive spectrometer records the count and energy of the emitted particles from the sample.

As measured X-rays are characteristic of the emitting elements constituting the sample (difference in energy between the two shells and the atomic structure), EDS allows the elemental composition measurement of the specimen (Figure II. 20).

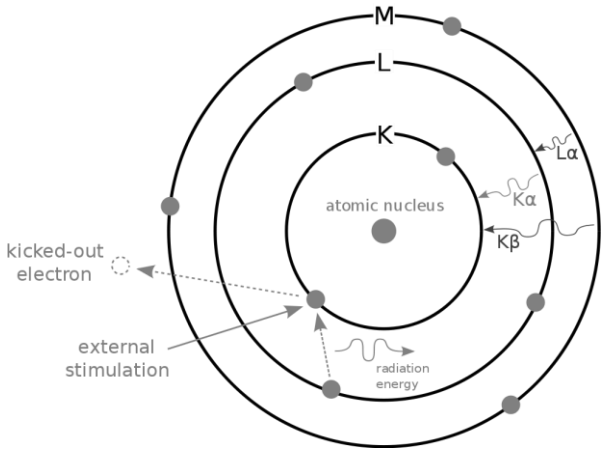

Figure II. 20: Illustration of the electron-matter interaction resulting in electron and characteristic X-ray emission [25].

Some steps have to be considered when performing an EDS experiment. As the electron beam can only penetrate a shallow depth, samples should be well prepared so that the morphology of the surface does not affect the results. Similar to SEM, the insulating samples have to be coated with a conducting material to avoid the charging effect. For EDS, the common coating material is vacuum-evaporated carbon, which has minimal effect on X-ray intensity thanks to its low atomic number.

In Figure II. 21, I present a cross-section of our SynT with the gate stack Ti/TiO_2/LiPON/Ti microfabricated on thermally grown SiO_2 on a Si wafer. The device was prepared with the focused ion beam (FIB) technique using Gallium ions. A portion of the cross-section was observed using SEM (Figure II. 21.a). Here we can see the coverage of the LiPON film on the bottom Ti electrode and the overall topology of the device via the contrast. An EDS measurement was performed on the selected region, and the presented elements can be identified by their emission spectrum. Besides the detectable principle elements, such as Si, P, O, N, and Ti, we can also find the byproducts of the FIB process that are C from the protective film and Ga from the milling ions (Figure II. 21.b). The previous SEM image was coupled with the detected elements to obtain an EDS MAP (Figure II. 21.c). The gate stack is well-defined with LiPON electrolyte with P and O as the representative elements and Ti electrodes. We can see from the MAP that the information on the far side of the sample is also illustrated, causing the image to be confusing. On the other hand, when we observed the sample with the secondary electron mode, mainly the signals from the interesting cross-section (near the electron beam) were detected and mapped (Figure II. 21.d).

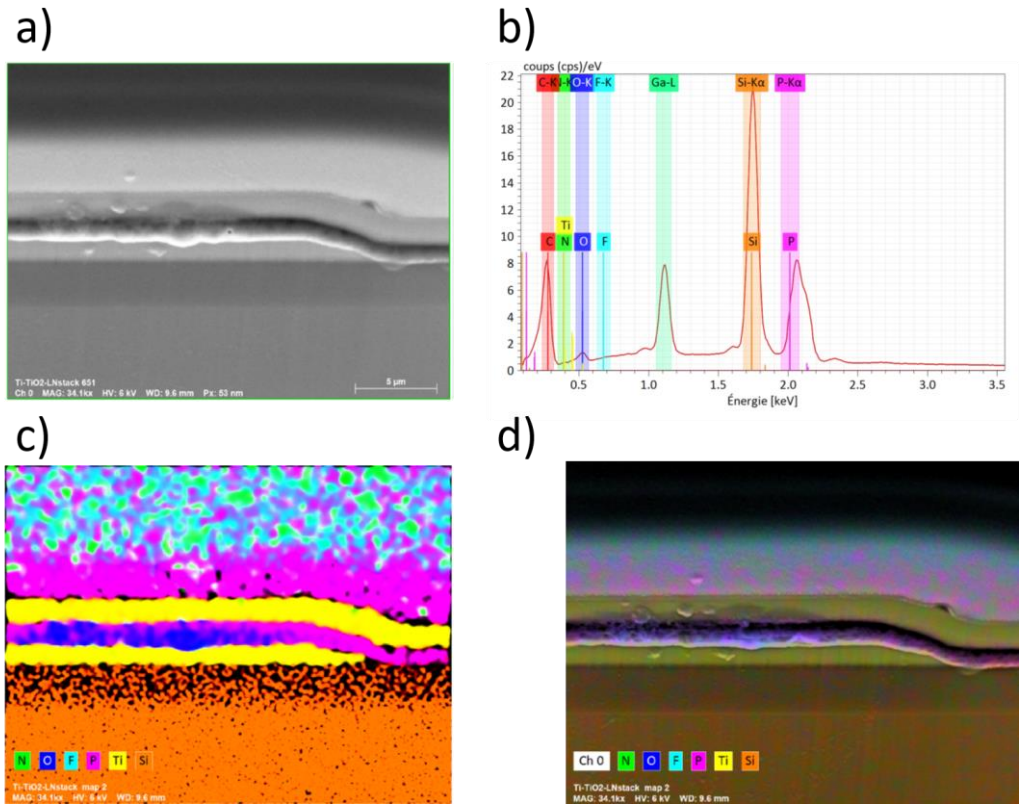

Figure II. 21: SEM coupled with EDS analysis on transistor structure. a) SEM cross-section of the transistor. b) EDS spectrum of the presenting elements in the transistor sample. c) Elemental mapping without secondary electron (SE) mode, and d) with SE mode.

3 ELECTROCHEMICAL SYNAPTIC TRANSISTOR PROCESS FLOW

An example of a completed wafer after microfabrication can be found in Figure II. 22.a. There are a number of different devices on this wafer, with the SynTs lying on the dices on the edge of the wafer. Elaborated electrochemical synaptic transistors are three-terminal devices comprising Gate, Source, and Drain electrodes, a semiconductor channel connecting the Source-Drain electrodes, and an electrolyte layer separating the Gate electrode and the channel (See Figure II. 22.b). The sandwich structure of thin films on silicon wafers can be realized by a sequence of microfabrication techniques, including deposition and patterning steps. Devices' performance depends directly on the several details of the fabrication process. For example, the micrometric gap between SD electrodes defines the channel length and, thus, the channel current. This gap dimension can be undesirably widened from the designed feature by overexposed or overdeveloped photoresist patterning and overetching with high-temperature wet etchants. Precise alignment is critical for this multiple-layer device. For instance, a slight misalignment on the gate electrode can add an extra capacitance (created by metal/electrolyte/metal in parallel to the transistor gate stack).

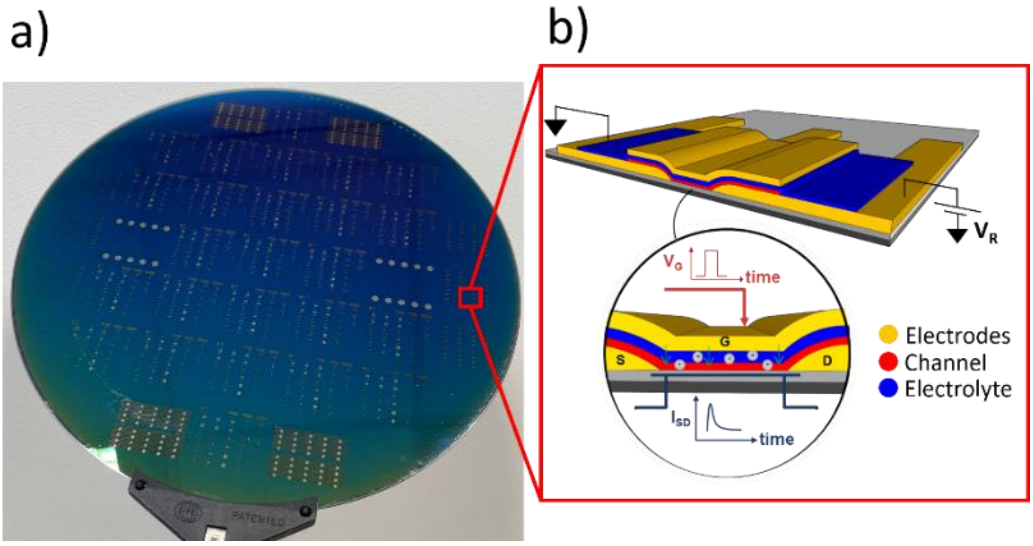

Figure II. 22: a) Photo of an elaborated SynT 200 mm substrate with multiple test structures. b) Schematic illustration of an electrochemical synaptic transistor from the wafer.

In this section, I will spend time describing the details of the process flow to microfabricate SynTs at LCRE.

3.1 Process flow

The completed process of SynT microfabrication is illustrated in Figure II. 23. All the devices realized in this thesis are on (100) single-crystalline silicon wafers of 200 mm in diameter and 550 µm in thickness. A 3-µm layer of SiO_2 insulator grown by thermal oxidation serves as a dielectric and support layer (Figure II. 23.a). There are two types of metals used for SynTs' electrodes, namely Platinum (Pt) and Titanium (Ti) (Figure II. 23.b). Pt is a noble transition metal that has remarkable resistance to corrosive reactions or thermal oxidation. To avoid platinum silicide compound (PtSi) formation at the interface [26], a thin TiO_2 is grown to play as a barrier and adhesion layer before coating Pt by sputtering. Ti metallization is realized by DC sputtering. The metal-coated wafers undergo common lithography patterning steps such as photoresist coating, insulation, development, and then metal etching. Pt metal is etched in a mixed solution of H_2O:HNO_3:HCl at 57 ˚C. The etching duration is controlled manually, and it depends on the etching conditions such as temperature, stirring, and also on material thickness. Ti metal and TiO_2 adhesion layers can be patterned with a solution of NH_3:H_2O_2:H_2O at room temperature (Figure II. 23.c). Another method to etch these two materials is by RIE with SF_6 gas. Subsequently, the channel layer is deposited using different methods (Figure II. 23.d).

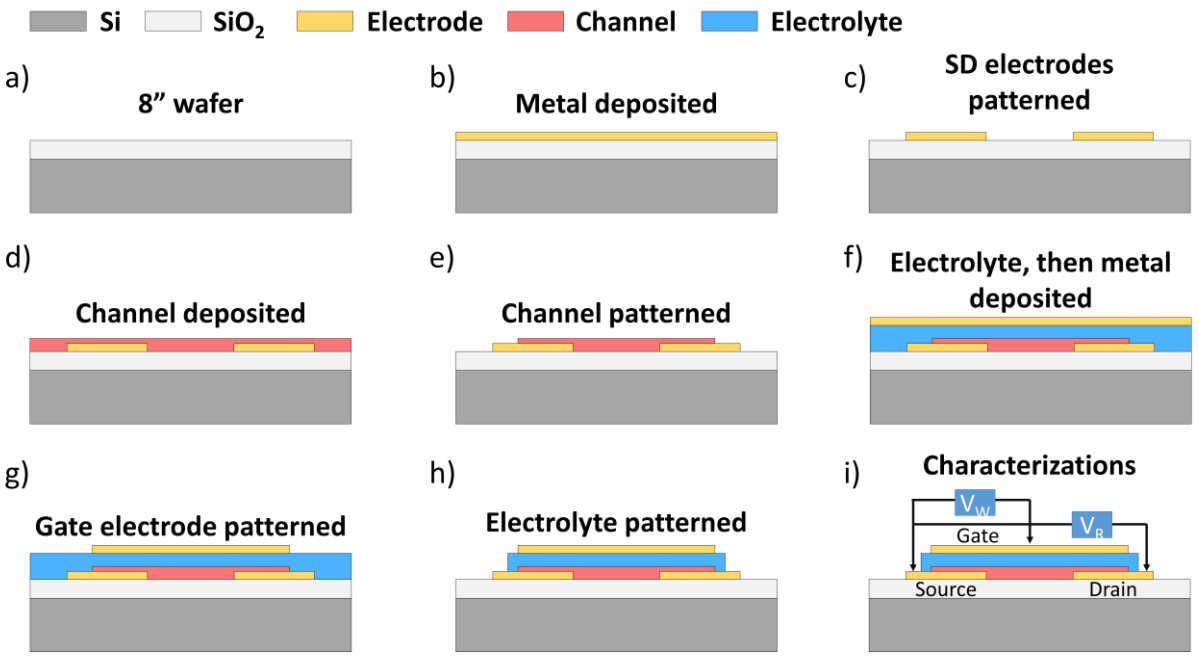

Figure II. 23: Process flow of SynTs.

There are two channel materials considered in this thesis work, TiO_2 deposited by ALD with TDMAT and water precursors, and $LiCoO_2$ deposited by sputtering. RIE recipes with SF_6 or CF_3/O_2 can be employed to pattern the TiO_2 layer, while wet etching is necessary

to handle the $LiCoO_2$ film (Figure II. 23.e). To etch $LiCoO_2$, the wafer is immersed in a bath with H_2SO_4:H_2O_2:H_2O solution at room temperature. Both channels have to undergo a thermal treatment at 400 ˚C and/or 700 ˚C to finalize the material preparation. With the channel ready, the LiPON electrolyte and Ti gate electrode are deposited subsequently (Figure II. 23.f). The LiPON film is grown on the substrate by RF sputtering using Li_3PO_4 target under the flow of N_2 gas; the Ti metal is coated by DC sputtering. We continue by patterning Ti top electrode (Figure II. 23.g). Unlike bottom Ti electrodes, RIE with SF_6 gas is the only option for gate patterning. Wet etching of Ti will not be possible because LiPON film is water sensitive. Finally, LiPON is patterned by the wet method using a TMAH solution (Figure II. 23.h), and the wafers with SynTs are ready for physical and electrical characterizations (Figure II. 23.i).

At the end of the fabrication process, we perform a FIB on our transistors to control the growth of the stacked layers. An example of the cross-sectional TEM image of an electrochemical synaptic transistor comprised of Ti metal electrodes (gate, source, and drain), TiO_2 channel layer, and LiPON solid-state electrolyte can be found in Figure II. 24. At 5k magnification, we can clearly observe the appearance of the films and verify the thicknesses of SiO_2 insulator layer, Ti electrodes, and LiPON electrolyte. Even though we can see a thin line representing the 10-nm TiO_2 layer at the interface of the LiPON and the bottom Ti films, we need to increase the magnification level of TEM to inspect this channel more clearly.

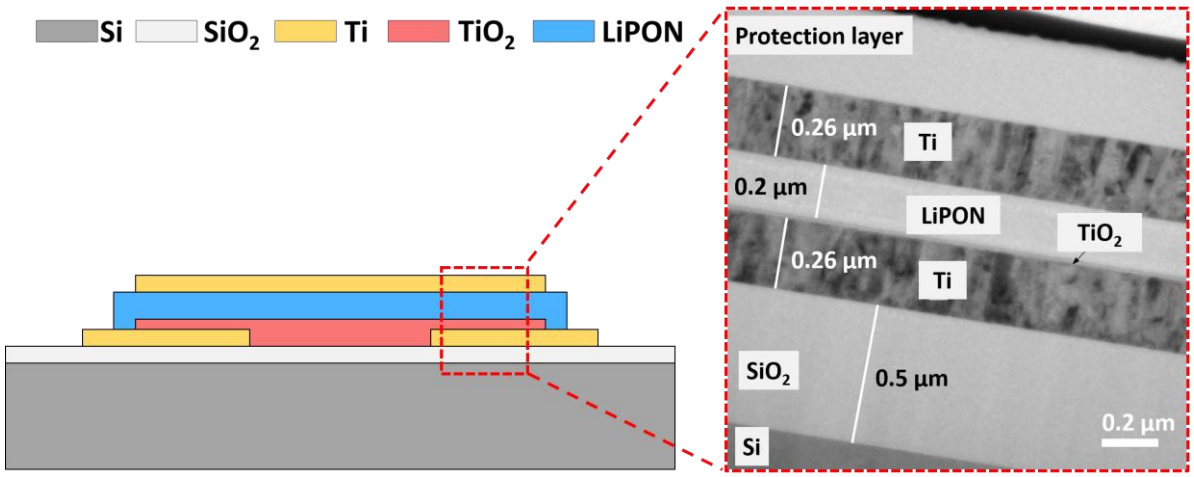

Figure II. 24: Cross-section TEM image of a SynT made of Ti electrodes, TiO_2 channel, and LiPON electrolyte.

3.2 Progressive optimization of microfabrication steps

3.2.1 SD patterning

The gap between the source and drain electrodes defines the length of the transistor channel. Therefore, precisely patterning these bottom electrodes is critical. For the first generation of SynTs, the gap between SD electrodes ranges from L = 1.5 µm (SynT-1) to L = 5 µm (SynT-5). However, it is not straightforward to pattern Pt metal electrodes with a few micrometer gaps because of two factors: the photoresist patterning and the etching method.

As described in Eq. 3, the resolvable dimension of photoresist can be affected by three elements, namely resist thickness, exposure gap, and the wavelength of the light source. The photoresist used for the protective mask in this process is positive Microposit SPR220 with 8-µm thickness after development. Even though the thickness of this photoresist is not ideal for open these gaps precisely (<5 µm), the available post-bake process with this resist is essential to protect the underlayer from corrosive wet etchants. The wafers are exposed using proximity (≈3 µm gap) mode and hard contact mode (no gap) with a Hg UV light. At the beginning of the thesis work, the photoresist layer is mainly patterned using proximity exposure. This exposure mode leads to unopened features for SynT-1 and SynT-2, whose gaps are smaller than the exposure gap (See Figure II. 25.a). These SD pitches are opened by employing the hard contact mode, eliminating the gap between the photomask and the substrate. The exposure time is also tuned around the theoretical dose of SPR220 at this thickness to avoid the under- and over-exposure effect that can change the shape of the developed resist. As a result, we can obtain the resist patterns with acceptable dimensions (See Figure II. 25.b).

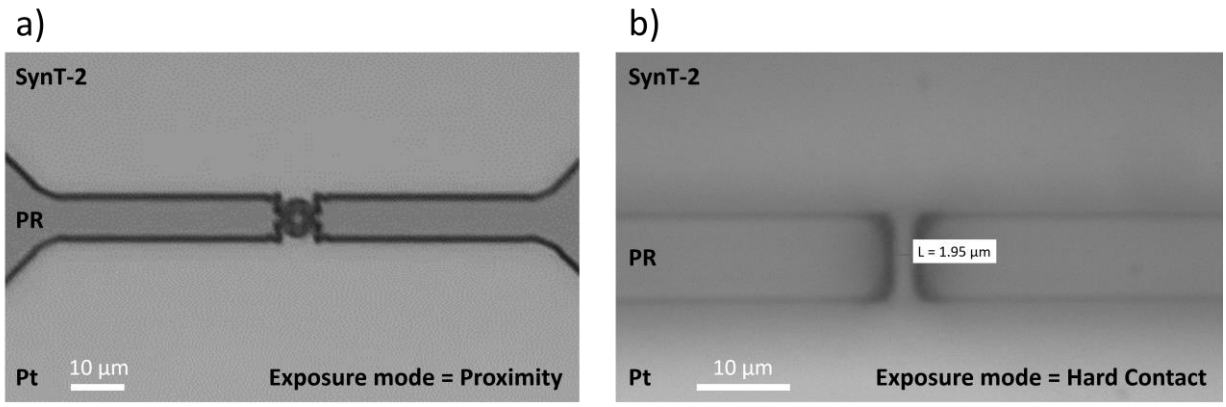

Figure II. 25: a) Optical microscopy image of connected patterns of photoresist under the effect of proximity exposure. b) The photoresist patterns are open and correspond well to the designed gap of SynT-2 by using hard contact exposure.

We have to take into account the isotropic profile of the wet etching method that contributes to the widening of the electrode gaps in the Pt patterning process. Thus, etching 250-nm thick layer of Pt results in at least a 0.5 µm gap widening due to over etch. Furthermore, the reactions happen inside a hot, corrosive mixture of concentrated nitric and hydrochloric acids. With this condition, it is hard to monitor the etching evolution closely. As a consequence, the etching profile is not uniform within the wafer, and the gaps of Pt electrodes are much wider than the designed ones (5 times wider in Figure II. 26.a) at the beginning of the thesis. To improve this overetching issue, several experiments with different etching duration are done to find out an optimized time for this etching is around 7 minutes and 30 seconds. Note that the etching substrate has to be rotated manually, and the solution bath has to be stirred constantly during the process so that the metal is removed equally at different positions within the wafer. Even though the etched patterns do not have the same dimensions as the designed ones, the SD gap is much improved (Figure II. 26.b).

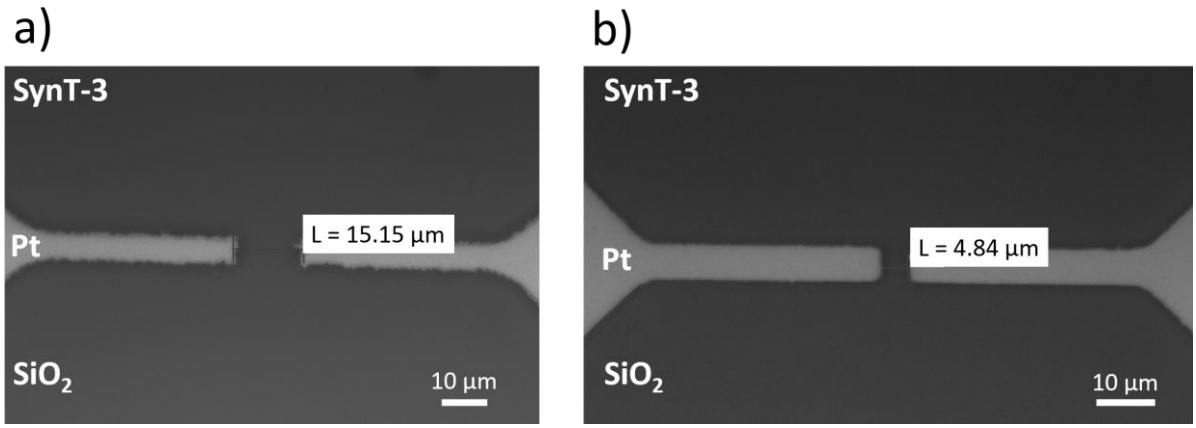

Figure II. 26: a) Severe overetching of Pt electrodes. b) An improvement of Pt electrode patterning by controlling the etching time and etching conditions.

SD electrodes made of Ti can have better resolutions (smallest gap of SynT-1 = 1.8 µm) due to the flexibility in the etching process. Bottom Ti metal can be etched with RIE etching with an anisotropic profile. In addition, with the wet etching process, Ti etching can be done with Microposit S1818 photoresist, which has four times thinner thickness compared to SPR220, resulting in a better definition of the features.

3.2.2 Channel processes

A channel is a patterned thin film connecting the source and the drain electrodes in a transistor. This layer can change its electrical properties under the field applied between the gate and the source electrodes. We consider ALD TiO_2 and PVD $LiCoO_2$ (LCO)

as the channels for the first generation of our electrochemical synaptic transistors. To prepare our channels, we first pattern and then anneal them using rapid thermal annealing (RTP tool). In the following parts, I will discuss the problems we faced and the ways to overcome both patterning and thermal treatment processes.

3.2.2.1 *Channel patterning*

With their large ratio of the area over thickness, our channels are supposed to be patterned easily. This is the case for the 10-nm TiO_2 channel. The film is etched with an SPR220 protective mask using CHF_3/O_2-based plasma etching. The patterned film has designed dimensions (Figure II. 27). The only issue observed in this process is the stripping of the photoresist after RIE. It is observed that after this etching, the photoresist structure becomes cross-linked and hardened because of the elevated temperature and UV radiation from the plasma inside the reactor. For this reason, resist remover 1112A at 50 ˚C is employed.

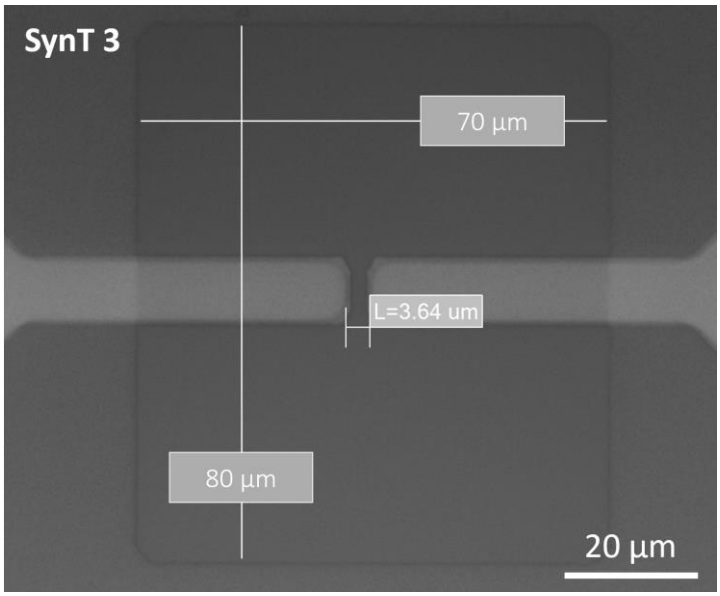

Figure II. 27: An image of the patterned TiO_2 channel with desired dimensions.

For thin-film $LiCoO_2$ with thickness from 50 nm to 100 nm, the etching happens rapidly under 10 s inside a mixed solution of H_2O:H_2O_2:H_2SO_4. At this time scale, greatly lateral overetching can be expected with a delay of a few seconds (Figure II. 28.a). To have better control of this etching process, I dilute the etching solution by increasing the H_2O:H_2O_2:H_2SO_4 ratio from 32:5:1 to 40:5:1 and perform the etching at room temperature instead of 35 ˚C as for micrometer-thick $LiCoO_2$ films. As a result, the etching process finishes around 30 s, and the control of over-etch is easier (Figure II. 28.b).

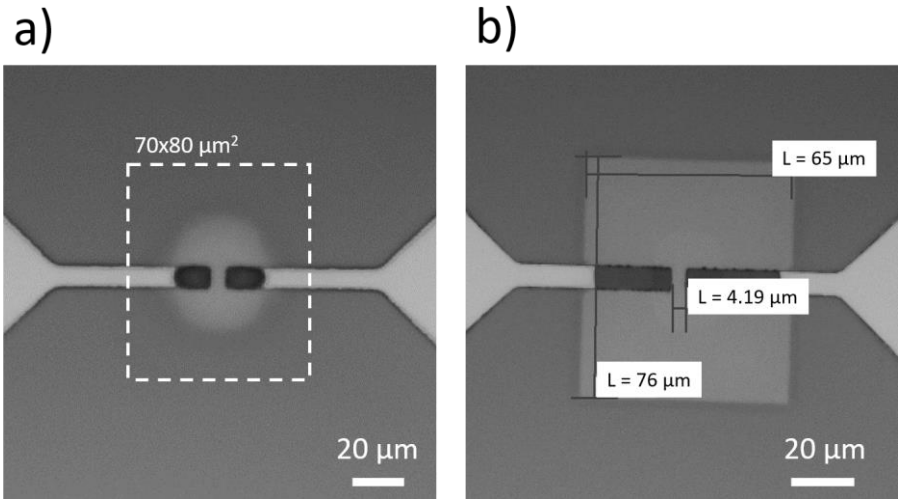

Figure II. 28: a) Lateral over-etching of LCO thin film with the dashed square being the patterned photoresist mask. b) An improved patterning of LCO channel with 5 µm over-etched on each side.

3.2.2.2 *Thermal processing*

Thermal annealing as a post process is important in thin-film microfabrication as it can alter the physical and chemical properties of the film, favor the formation of specific structural phases, and improve the surface roughness of inorganic films. For example, annealing the as-deposited LiCoO2 film at 400 ˚C and 700 ˚C can induce a low-temperature phase (LT-LCO) and a high temperature phase (HT-LCO), respectively [27]. LT-LCO has a spinel-type structure with space group Fd-3m. It is obtained at temperatures below 400 °C. HT-LCO has a hexagonal structure with space group R-3m. It is obtained at temperatures above 600 °C. With an ordered structure favoring the Li intercalation, HT-LCO is preferred as the cathode material in Li-ion battery application, and it is the target phase of our synaptic transistor channel. I will describe the process of development to obtain a qualified thin-film HT-LCO channel in the following part.

From the previous step, the as-patterned LCO channel will be annealed in the RTP tool at 700 ˚C to transform the LCO phase into HT-LCO. However, the annealed LCO layer does not inhibit an R-3m structure, but it shows a vibration type similar to the Fd-3m structure under the Raman spectroscopy technique (Process 1, Figure II. 29). This undesired phase transformation can be explained by the exposure of the thin-film LCO to the solution during the etching step. In an attempt to obtain the HT-LCO phase of the channel layer, we proceed with annealing after deposition of LCO before patterning (Process 2). With this modification of the step sequence, we successfully observed the R-3m structure signature with Raman spectroscopy. Nonetheless, the etching step is long (> 20 minutes) and more complex to completely remove LCO (residue all over the wafer). Finally, by inserting an intermediate annealing step at 400 °C between PVD deposition of LCO and patterning steps, we obtain the desired HT-LCO phase with a feasible etching step (process 3). The

intermediate annealing step is believed to transform the as-deposited LCO into spinel LT-LCO, which can protect the thin film from the effects of solution exposure. Process 3 is essential to obtain a functional LCO layer for both energy and synaptic transistor applications.

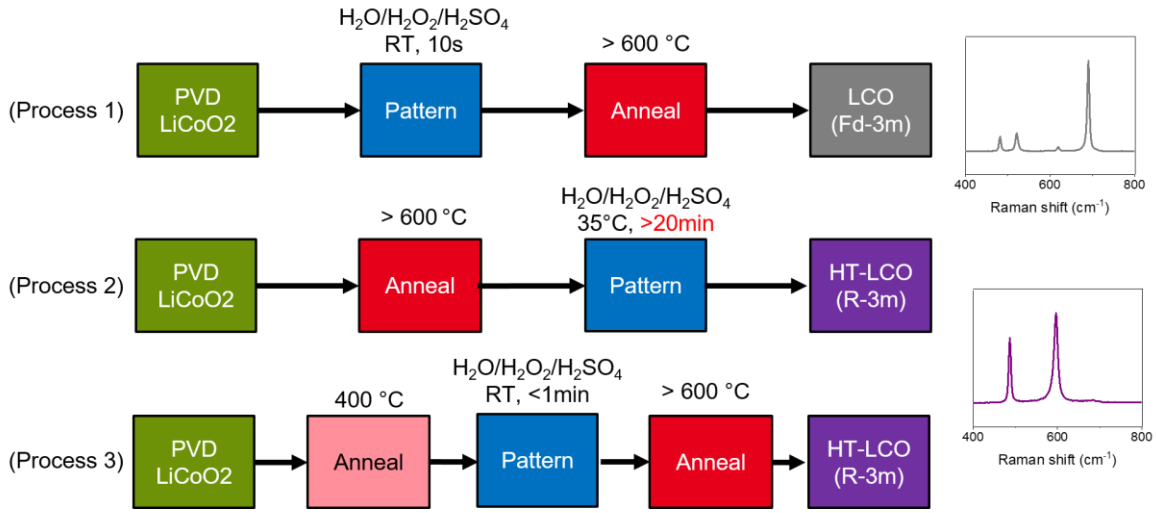

Figure II. 29: Evolution of processes to obtain HT-LCO channel layers.

3.2.3 Gate and electrolyte patterning

Now we have to pattern the Ti gate electrode and then the LiPON electrolyte. Unlike bottom Ti electrodes, RIE is the only option for gate patterning. In the first versions of SynT, the gate electrode is designed to be relatively small (12x14 μm^2). The size of this electrode makes it impossible to keep the square shape with proximity exposure due to sharp edge diffraction. As a result, the photoresist at the corners is over-exposed, creating a cross-shape gate electrode (Figure II. 30.a). As the gate is patterned, the wafer is coated with an S1818 positive photoresist (MICROPOSIT) immediately to reduce the air exposure time of the bare LiPON surface. LiPON film is patterned with a photomask with a designed area of 70x80 μm^2 marked with a dashed rectangle. However, the lateral over-etching rate of 100 – 200 nm LiPON film is extremely high inside a TMAH solution, creating a strange pattern as seen in Figure II. 30.a. An attempt to optimize the etching time has been done to pattern 100 nm LiPON on SiO_2 wafers. The immersing time inside the etching bath is 45 s for Figure II. 30.b and 32 s for Figure II. 30.c, and it is followed by 5 s of DI water rinsing. Although the overall shape of the pattern is monitored, the lateral over-etching may pose a risk of a short circuit between the gate and bottom electrodes, especially when adding the effect of bottom electrode topography.

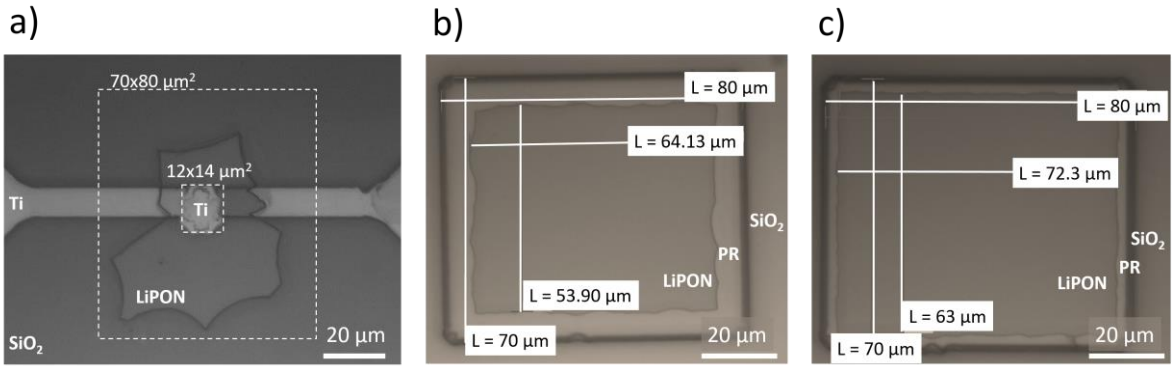

Figure II. 30: a) Result of gate electrode and electrolyte patterning at the beginning of the thesis. Wet etching of 100 nm LiPON layer on SiO_2 during b) 45 s + 5 s rinsing with DI water, and c) 32 s + 5 s rinsing with DI water.

To address the problems mentioned above, the designs of the gate electrode and the electrolyte film are extended in size (Figure II. 31.a). The gate electrode now has the same size as the channel layer (70x80 µm^2), creating a fully overlapped gate stack. With the new size of electrode, the impact of edge diffraction is minimized, and we can comfortably probe on the gate electrode without the need for a redistribution layer, therefore accelerating the development cycle of the device. The size of the LiPON pattern now is 470x425 µm^2, covering the active zone of the device completely. The size of the layer will eliminate the lateral over-etching risk from the wet process. The completed view of a SynT-1 device can be found in Figure II. 31.b, and this transistor is ready for further physical and electrical characterizations.

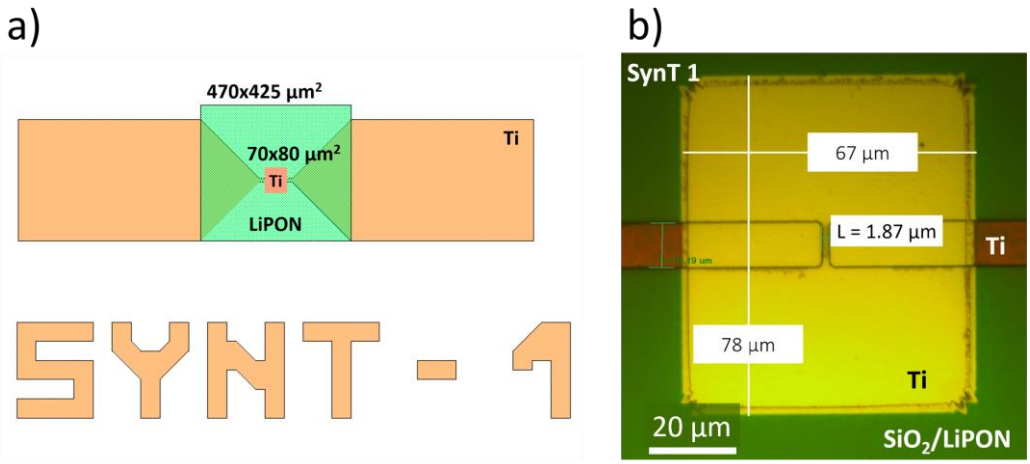

Figure II. 31: a) The layers' size adaptation of the SynT design to alleviate constrains from Ti gate electrode and LiPON electrolyte patterning. b) A top view of a complete synaptic transistor taken on SynT-1 device.

4 CONCLUSIONS

In this chapter, we have presented the microfabrication techniques used in my thesis. The deposition techniques are shown with a focus on the techniques that being used to elaborate our transistors: sputtering and atomic layer deposition. As examples, the deposition processes of the PVD LiPON and ALD TiO_2 thin films are described. Furthermore, we have discussed the methods of photoresist mask forming and material etching because they directly affect the transfer of designed layers into the deposited thin films. The pattern resolution of photoresist depends heavily on their thickness, the exposure gap, and the light source's wavelength. The final patterns of devices are dictated by the etching profiles of etching methods.

Physical characterizations such as SEM and TEM are often employed to inspect the microstructures, thin films' growth and their thicknesses, while EDS and Raman spectroscopy are particularly useful to obtain the chemical composition and the fingerprints of deposited layers. The working principles of these techniques and the examples from the literature and own work have been discussed in this chapter. These techniques are also useful for identifying possible defects and verifying the microfabrication steps. The physical understanding of the gate stacks allows us to interpret more precise the performance of the devices.

In the last section, we have shown the process flow to elaborate our electrochemical synaptic transistors with details on the steps, from material deposition to patterning. An example of cross-section view of a SynT is also illustrated with verified thicknesses of different components of the device. We encountered numerous problems while realizing this generation of device, and the progressive optimization process is mentioned. At each microfabrication steps, there are various challenges: unopen features, over-etching layers, and undesired channel structure. Resolving these difficulties required efforts from different approaches including adjusting exposure gap, monitoring patterning time and temperature control, and redesigning the photomask. The first functional SynTs have been elaborated using the described processes.

5 REFERENCES

[1] J. Bącela, M. B. Łabowska, J. Detyna, A. Zięty, and I. Michalak, "Functional Coatings for Orthodontic Archwires—A Review," *Materials*, vol. 13, no. 15, p. 3257, Jul. 2020, doi: 10.3390/ma13153257.

[2] S. Franssila, *Introduction to microfabrication*. Chichester, West Sussex, England ; Hoboken, NJ: J. Wiley, 2004.

[3] C. H. Choi, W. I. Cho, B. W. Cho, H. S. Kim, Y. S. Yoon, and Y. S. Tak, "Radio-Frequency Magnetron Sputtering Power Effect on the Ionic Conductivities of Lipon Films," *Electrochem. Solid-State Lett.*, vol. 5, no. 1, p. A14, 2002, doi: 10.1149/1.1420926.

[4] C. S. Nimisha, K. Y. Rao, G. Venkatesh, G. M. Rao, and N. Munichandraiah, "Sputter deposited LiPON thin films from powder target as electrolyte for thin film battery applications," *Thin Solid Films*, vol. 519, no. 10, pp. 3401–3406, Mar. 2011, doi: 10.1016/j.tsf.2011.01.087.

[5] B. Put, P. M. Vereecken, J. Meersschaut, A. Sepúlveda, and A. Stesmans, "Electrical Characterization of Ultrathin RF-Sputtered LiPON Layers for Nanoscale Batteries," *ACS Appl. Mater. Interfaces*, vol. 8, no. 11, pp. 7060–7069, Mar. 2016, doi: 10.1021/acsami.5b12500.

[6] H. Xia, H. L. Wang, W. Xiao, M. O. Lai, and L. Lu, "Thin film Li electrolytes for all-solid-state micro-batteries," *IJSURFSE*, vol. 3, no. 1/2, p. 23, 2009, doi: 10.1504/IJSURFSE.2009.024360.

[7] J. F. Ribeiro *et al.*, "Enhanced solid-state electrolytes made of lithium phosphorous oxynitride films," *Thin Solid Films*, vol. 522, pp. 85–89, Nov. 2012, doi: 10.1016/j.tsf.2012.09.007.

[8] T.-E. Lee, K. Toprasertpong, M. Takenaka, and S. Takagi, "Re-examination of effects of ALD high-k materials on defect reduction in SiGe metal–oxide–semiconductor interfaces," *AIP Advances*, vol. 11, no. 8, p. 085021, Aug. 2021, doi: 10.1063/5.0061573.

[9] Z. Li, J. Su, and X. Wang, "Atomic layer deposition in the development of supercapacitor and lithium-ion battery devices," *Carbon*, vol. 179, pp. 299–326, Jul. 2021, doi: 10.1016/j.carbon.2021.03.041.

[10] V. Sallaz, S. Oukassi, F. Voiron, R. Salot, and D. Berardan, "Assessing the potential of LiPON-based electrical double layer microsupercapacitors for on-chip power storage," *Journal of Power Sources*, vol. 451, p. 227786, Mar. 2020, doi: 10.1016/j.jpowsour.2020.227786.

[11] www.atomiclimits.com, "Atomic Layer Deposition Process Development." https://www.epfl.ch/research/facilities/cmi/wp-content/uploads/2020/07/ALD-Process-Development.pdf (accessed Jul. 28, 2022).

[12] D. Pan, L. Ma, Y. Xie, T. C. Jen, and C. Yuan, "On the physical and chemical details of alumina atomic layer deposition: A combined experimental and numerical approach," *Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films*, vol. 33, no. 2, p. 021511, Mar. 2015, doi: 10.1116/1.4905726.

[13] Alan Doolittle, "Lecture 7: Lithography and Pattern Transfer." Accessed: Aug. 10, 2022. [Online]. Available: https://alan.ece.gatech.edu/ECE6450/Lectures/ECE6450L7-Optical%20Lithography.pdf

[14] G. Roeder *et al.*, "Determination of the Dill parameters of thick positive resist for use in modeling applications," *Thin Solid Films*, vol. 519, no. 9, pp. 2978–2984, Feb. 2011, doi: 10.1016/j.tsf.2010.11.068.

[15] Minghao Qi, "Lecture 44: Etching," Aug. 10, 2022. [Online]. Available: https://nanohub.org/resources/25355/download/2016.04.13-ECE695Q-L44.pdf
[16] G. S. Oehrlein and J. F. Rembetski, "Plasma-based dry etching techniques in the silicon integrated circuit technology," *IBM J. Res. & Dev.*, vol. 36, no. 2, pp. 140–157, Mar. 1992, doi: 10.1147/rd.362.0140.
[17] J. Lee, H. W. Lee, and K.-H. Kwon, "Characteristics of etching residues on the upper sidewall after anisotropic plasma etching of silicon," *Applied Surface Science*, vol. 517, p. 146189, Jul. 2020, doi: 10.1016/j.apsusc.2020.146189.
[18] D. B. Williams and C. B. Carter, *Transmission electron microscopy: a textbook for materials science*, 2nd ed. New York: Springer, 2008.
[19] A. Hamilton and F. Quail, "Detailed State of the Art Review for the Different On-Line/In-Line Oil Analysis Techniques in Context of Wind Turbine Gearboxes," in *Volume 1: Aircraft Engine; Ceramics; Coal, Biomass and Alternative Fuels; Wind Turbine Technology*, Vancouver, British Columbia, Canada, Jan. 2011, pp. 971–988. doi: 10.1115/GT2011-46860.
[20] S. Oukassi, L. Baggetto, C. Dubarry, L. Le Van-Jodin, S. Poncet, and R. Salot, "Transparent Thin Film Solid-State Lithium Ion Batteries," *ACS Appl. Mater. Interfaces*, vol. 11, no. 1, pp. 683–690, Jan. 2019, doi: 10.1021/acsami.8b16364.
[21] J. Lee, R. D. Nikam, M. Kwak, and H. Hwang, "Strategies to Improve the Synaptic Characteristics of Oxygen-Based Electrochemical Random-Access Memory Based on Material Parameters Optimization," *ACS Appl. Mater. Interfaces*, vol. 14, no. 11, pp. 13450–13457, Mar. 2022, doi: 10.1021/acsami.1c21045.
[22] J. A. Carrero, G. Arana, and J. M. Madariaga, "Chapter 6. Use of Raman spectroscopy and scanning electron microscopy for the detection and analysis of road transport pollution," in *Spectroscopic Properties of Inorganic and Organometallic Compounds*, vol. 45, R. Douthwaite, S. Duckett, and J. Yarwood, Eds. Cambridge: Royal Society of Chemistry, 2014, pp. 178–210. doi: 10.1039/9781782621485-00178.
[23] T. Schmid and P. Dariz, "Raman Microspectroscopic Imaging of Binder Remnants in Historical Mortars Reveals Processing Conditions," *Heritage*, vol. 2, no. 2, pp. 1662–1683, Jun. 2019, doi: 10.3390/heritage2020102.
[24] O. Frank, M. Zukalova, B. Laskova, J. Kürti, J. Koltai, and L. Kavan, "Raman spectra of titanium dioxide (anatase, rutile) with identified oxygen isotopes (16, 17, 18)," *Phys. Chem. Chem. Phys.*, vol. 14, no. 42, p. 14567, 2012, doi: 10.1039/c2cp42763j.
[25] H. A. Alshamarti, J. J. Hassan, and H. L. Saadon, "Fabrication and Characterization of UV Photodetectors Based on Metal Doped and Undoped ZnO Nanorods," 2019, doi: 10.13140/RG.2.2.33555.43044.
[26] G. R. Fox, S. Trolier-McKinstry, S. B. Krupanidhi, and L. M. Casas, "Pt/Ti/SiO_2/Si substrates," *J. Mater. Res.*, vol. 10, no. 6, pp. 1508–1515, Jun. 1995, doi: 10.1557/JMR.1995.1508.
[27] H. Xia, Y. Wan, W. Assenmacher, W. Mader, G. Yuan, and L. Lu, "Facile synthesis of chain-like LiCoO2 nanowire arrays as three-dimensional cathode for microbatteries," *NPG Asia Mater*, vol. 6, no. 9, pp. e126–e126, Sep. 2014, doi: 10.1038/am.2014.72.

CHAPTER III

FIRST GENERATION OF ELECTROCHEMICAL SYNAPTIC TRANSISTORS

ABSTRACT

In the third chapter, we will introduce the first functional gate stacks of SynTs including the materials and the characterizations. The chapter is divided into three principle sections, the materials composing the transistors, the characterizations of SynTs with $LiCoO_2$/LiPON gate stack, and the characterizations of SynTs with Li_xTiO_2/LiPON gate stack.

Section 1 of this chapter is dedicated to introduce the properties of the materials used for the elaboration of the SynTs, including $LiCoO_2$ and Li_xTiO_2 for the channel, and LiPON for the solid-state electrolyte. For the channel materials, their crystal structures of different phases and their insulator metal transition phenomenon are mentioned. The advantages of LiPON solid electrolyte over other types of ionic conductors are presented to highlight the material choice of the first generation of SynTs.

Section 2 shows the development and a first study of SynTs composed of $LiCoO_2$/LiPON gate stack. With different test structures and electrical setup, the properties of this gate stack are revealed. The contribution of HT-LCO and LiPON electrolyte are verified with impedance spectroscopy and cyclic voltammetry characterizations. The resistive switching and analog programming desired in electrochemical synaptic transistors was demonstrated experimentally.

We will present an extended study of Li_xTiO_2/LiPON gate stack comprising of physical, electrical and electrochemical characterizations in the third section. SEM and EDS are used to study the films' thickness and composition. HRTEM affirms the quasiamorphous phase of Li_xTiO_2 channel material and the fact that there exists nanoinclusions inside the amorphous matrix. The good merits of this gate stack in transistor configuration such as low-conductance analog switching in the range of nano-Siemens and low writing energy ($fJ/\mu m^2$) are confirmed with electrical tests. Synaptic plasticity characteristics required for an artificial synaptic component are also demonstrated. The high operation speed is accounted for by the fast ion intercalation into pseudocapacitive behavior of the 10-nm thin TiO_2 channel, which is verified by the results of electrochemical tests on a two-terminal structure.

TABLE OF CONTENTS

1 MATERIALS 101

1.1 Lithium cobalt oxide ($LiCoO_2$) channel 101

1.2 Lithium titanium oxide (Li_xTiO_2) channel 103

1.3 Lithium phosphorous oxynitride (LiPON) electrolyte 105

2 PRELIMINARY STUDY OF SYNT WITH {$LiCoO_2$/LiPON} GATE STACK 107

2.1 Electrical measurement setup 107

2.2 Electrochemical characterization of the gate stack 109

2.3 Electrical characterization of the transistor 110

2.3.1 Resistive switching phenomenon 110

2.3.2 Analog state programming 111

2.4 Summary 112

3 EXTENDED STUDY OF SYNT WITH {Li_xTiO_2/LiPON} GATE STACK 114

3.1 Artificial electrochemical synapse 114

3.2 Physical characterizations 115

3.3 Electrical characterizations 116

3.4 Electrochemical characterizations 121

3.5 Summary 127

4 CONCLUSIONS 128

5 REFERENCES 130

INTRODUCTION: BACKGROUND RECALL

Among three-terminal artificial synapses, electrochemical synaptic transistors (SynTs) appear as suitable candidates due to their similarities with biological synapses. In a SynT, the gate dielectric is an electrolyte film, which conducts only ions. When an external voltage pulse (WRITE) is applied between the gate and source, mobile ions intercalate into the channel film through electrochemical reactions, which yields modulation of the channel's electronic conductivity. This similarity potentially places SynTs in a close position to its inspiration, the biological synapses, in terms of operation, energy, and speed.

However, elaboration of most electrochemical SynTs towards further scalability and CMOS compatibility suffers several kinds of difficulties. Liquid and solid polymer electrolytes (ionic liquids [1], ion gels [2], PEO-$LiClO_4$ [3]–[6]) show clear limitations towards wafer scale integration [7]. Furthermore, channels composed of mechanically exfoliated layers appear not suitable for future elaboration of networks composed of a large number of synaptic components [3], [5], [8]. For these reasons, fabrication of all-solid-state wafer-scale SynTs is highly desirable and is drawing considerable attention [9]–[15].

To overcome such problems, we investigated specific all-solid-state, wafer-scale fabricated electrochemical synaptic transistors: concerning the channel material, we first tried $LiCoO_2$ then Li_xTiO_2. Regarding the electrolyte, we used LiPON. In the following parts of the chapter, we will first present the properties of the constituting materials, followed by the characterizations of SynTs with the corresponding gate stacks.

1 MATERIALS

1.1 Lithium cobalt oxide ($LiCoO_2$) channel

In an electrochemical synaptic transistor, channel is a semiconductor material whose doping level can be altered by the injection of mobile ions under the field effect from the gate, thereby modulating the conductance states. Therefore, the channel material choice controls directly the working voltage window, dynamic range, and reading energy of the device.

$LiCoO_2$ is a common material in Li-based energy storage application. $LiCoO_2$ exhibits two crystal structures depending on the preparation and the synthesis temperature. While the high-temperature LiCoO2 (HT-LCO) has a hexagonal layered R-3m structure by synthesizing at T > 600 °C, the low-temperature $LiCoO_2$ (LT-LCO), synthesized at a lower temperature T < 400 °C, has a cubic spinel-related Fd-3m structure (see Figure III. 1) [16], [17]. For its superior electrochemical performance [18], only HT-LCO is considered in this thesis.

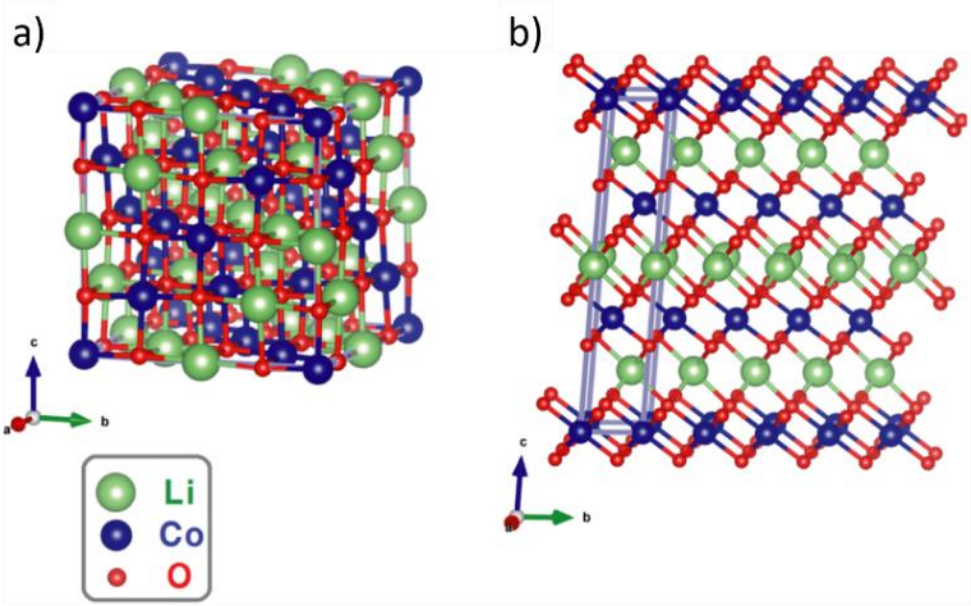

Figure III. 1: Crystal structures of (a) Fd-3m spinel $LiCoO_2$ (LT-LCO) and (b) R-3m layered $LiCoO_2$ (HT-LCO) [17].

HT-LCO is used as a cathode (positive electrode) in commercial Li-ion batteries for its high energy density and great cycling stability. Charge–discharge cycling with cell voltage in the range of 3.5 to 4.2 V corresponds to deintercalation/intercalation of about 0.5 Li per $LiCoO_2$ formula unit, giving a specific capacity of about 140 mAh/g [19]. Deeper lithium extraction causes structural instability of the Li_xCoO_2 cathode material and its

reaction with electrolyte. Stoichiometric Li_1CoO_2 crystalline film is reported to be an insulator [20]. As lithium ions are extracted from $LiCoO_2$ film, there exists an insulator-metal transition (IMT), and the nature of this phenomenon is thoroughly studied using different approaches in REFs [19], [21], [22]. The IMT of $LiCoO_2$ can be observed from its computed electronic structures for lithium content x = 1, 0.99 and 0.6 in Figure III. 2.

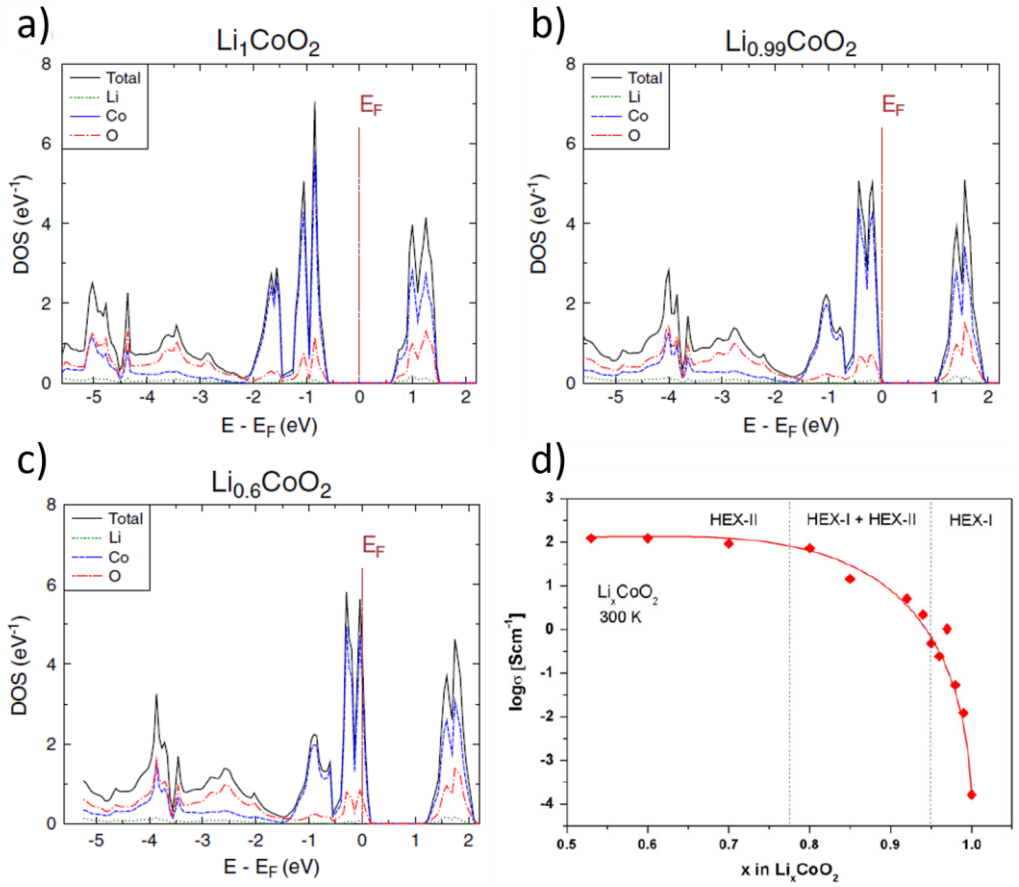

Figure III. 2: Total and site-decomposed DOS of Li_xCoO_2 oxides with different Li concentration, i.e. a) x=1, b) x= 0.99, and c) x = 0.6. d) Variation of the electrical conductivity of LixCoO2 at 300 K as a function of lithium concentration x. Adapted from [19].

With x = 1, the electronic structure of pure compound consists of a separation of valence and conduction bands by a gap of 1–1.2 eV with Fermi level E_F lying inside the gap, indicating a low conducting state of the film (Figure III. 2.a). Note that the position of the Fermi level relative to the band structure, the electronic conductivity of the electrode material can be determined, e.g., if a Fermi level crosses a band, the electrons can be excited to non-localized conduction states at the expense of very little energy. We can observe that with little deintercaltion of lithium from x= 1 to x = 0.99, the film becomes more conducting with E_F lying on top of the valence band (Figure III.1.b). With more lithium extraction at x = 0.6, Li_xCoO_2 film is an electronic conductor at room temperature (Figure III. 2.c). The electrical conductance profile of Li_xCoO_2 as a function of lithium concentration can be found in Figure III. 2.d. The total variation of conductivity under Li deintercalating

effect can be of six orders of magnitude, which is promising for electrochemical synaptic transistor applications [9], [11], [23]. Besides, $LiCoO_2$ had already been studied (in two-terminal configuration toward non-volatile memories [24], [25]) in a former collaboration between GeePs and CEA-LETI laboratories. For all these reasons, we select this material for the channel of our first SynTs.

1.2 Lithium titanium oxide (Li_xTiO_2) channel

Titanium oxide is an important metal oxide in surface science. This material is widely used for multiple applications, such as solar cells, gas sensors, optical waveguides, capacitors, and batteries because it has high chemical stability, good photoactivity, relatively low cost, and nontoxicity. Crystalline TiO_2 may exist in three polymorphs such as rutile, anatase, and brookite (see Figure III. 3).

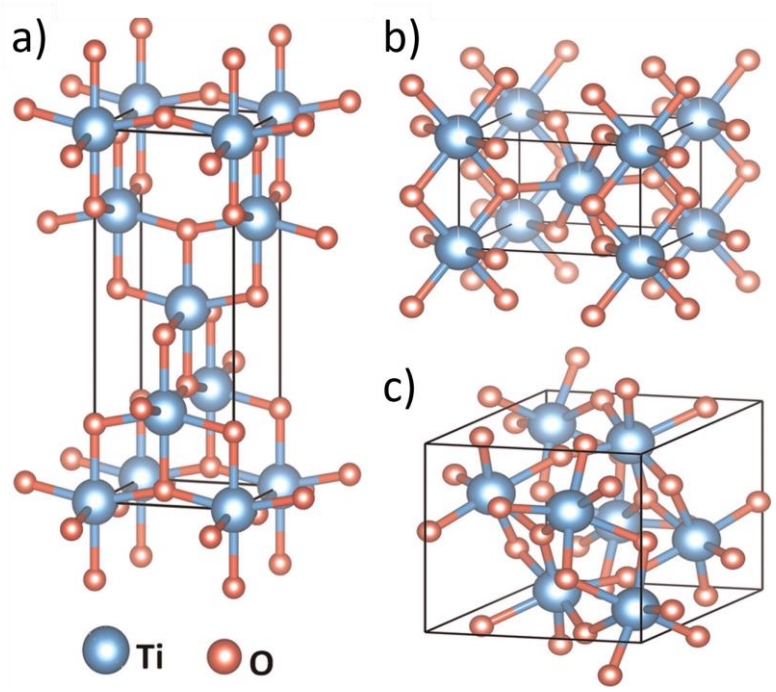

Figure III. 3: Crystal structures of TiO_2 (a) anatase (tetragonal), (b) rutile (tetragonal), and (c) brookite (orthorhombic) polymorphs [26].

Brookite is quite rare in nature. Rutile phase is stable at high temperature while it is not the case for anatase phase. However, the intercalation of lithium ion into rutile structure of TiO_2 is difficult due to its large distortion of the bulk lattice [27]. On the other hand, anatase TiO_2 is well known for its ability to accommodate charge in the form of interstitial lithium ions, thus more advantageous in applications require ion intercalation such as electrochemical energy storage and information storage [4], [28]–[31]. Upon lithium intercalation, anatase-based Li_xTiO_2 film observes an interesting change in electronic

conductivity, which was verified using experimental and theoretical approaches in REFs [32], [33], and is demonstrated with ab-initio calculation performed for TiO_2 and Li_1TiO_2 in Figure III. 4.

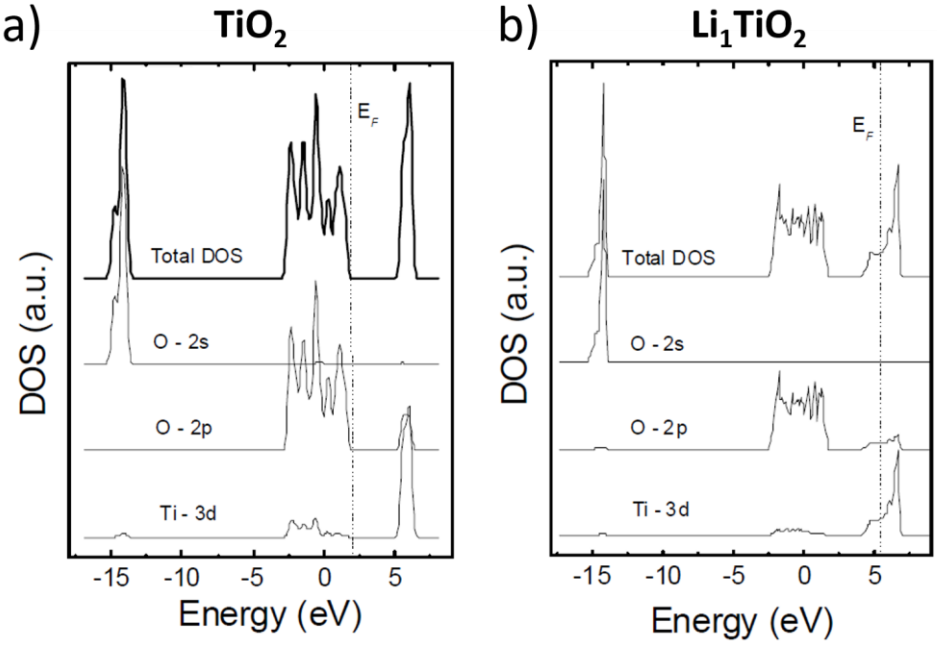

Figure III. 4: Total density of state of anatase-phase a) TiO_2 and b) $LiTiO_2$ anatase respectively related to the partial DOS of atomic orbitals of Ti and O.

From Figure III. 4, the material will be transformed from an insulator from TiO_2 to an electronic conductor upon the Li intercalation Li_1TiO_2. In the case of TiO_2 anatase, the Fermi energy E_F is located at the bottom of a ~3 eV bandgap, thus in this case, the anatase TiO_2 is referred to as an insulator at room temperature. On the other hand, in Li_1TiO_2 anatase, the Fermi energy crosses the conduction band, which is mainly contributed by the 3d-orbital of Ti. This implies the charge-compensating Li-2s electron to enter the material in a non-localized state (Li atoms are fully ionized and become Li+, resulting in an excess of electrons in the system filling partially Ti 3d and moving the Fermi level to the middle of the conduction band), transforming the pristine crystal from an insulator to an electronic conductor [33]–[35].

This IMT property was employed in an electrochemical synaptic transistor whose channel is made of anatase Li_xTiO_2 [4], see Figure III. 5. The memory cell, illustrated schematically in Figure III. 5.a, comprising of a redox transistor and a diffusive memristor selector. The transistor is made of a 30-nm Li_xTiO_2 channel and $LiClO_4$:PEO polymer electrolyte. A $Li_{0.7}FePO_4$ reference electrode controls the concentration of lithium inside Li_xTiO_2 channel during (de)intercalation, while its electronic conductivity modulation is sampled with a small bias (See Figure III. 5.b.). As Li ions incorporate into the material (V_{REF} < 1.6 V), the conductance rise significantly from nearly zero to 200 µS (red curve). This can

be explained by the total contribution of lithium doping as n-type donors. When more ions are inserted, the conductance drops rapidly, before rising again at V_{REF} = 1.9 V. This drop of conductance can be assigned to an anatase-to-Li-titanate phase transition studied in REF [32]. The same trend is observed in the reversed direction (blue curve) with a common voltage hysteresis in electrochemical systems. This study demonstrates that Li_xTiO_2 is an excellent channel material for synaptic transistor with many great merits. However, the conductance level of this channel is still high (µS range). Thinning down to 10-nm or less and working with a less conducting phase can further improve the energy performance of this channel. Furthermore, amorphous, ultrathin TiO_2 film exhibits extrinsic-pseudocapacitive behavior that endures rapid Li intercalation and tremendous cycling. For these reasons, we construct our first generation of SynTs with a channel made of thin and amorphous TiO_2 layer.

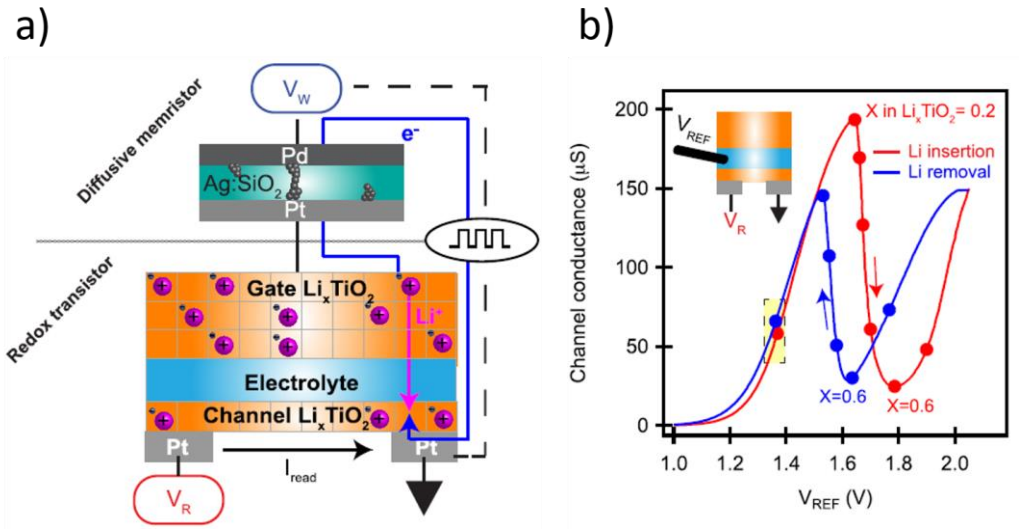

Figure III. 5: a) An electrochemical synaptic transistor made of an anatase-phase Li_xTiO_2 channel. b) The conductance variation of the channel under the change of Li content. Adapted from [4].

1.3 Lithium phosphorous oxynitride (LiPON) electrolyte

In electrochemical synaptic transistors, electrolytes serve as ion conducting media to modulate channel conductance via electrostatic mode (EDL formation) and intercalation mode (redox reactions). Thus, this layer contributes directly to the operation speed, energy, and state retention. LiPON is a common solid-state electrolyte for ion-based devices (battery, supercapacitor, electrochemical transistors) [9], [36], [37]. Solid-state electrolyte has advantages of operation temperature and stability over other types of electrolytes such as ionic liquids or polymer. LiPON can endure temperature up to 350 °C without degradation [38], much higher than polymer or organic-based electrolytes. In addition, they allow easy miniaturization and on-chip integration, which enables the elaboration of highly dense structures or for medical implant applications. LiPON thin film is selected to

be our electrolyte because of its known scalability (down to 20 nm), a large chemical stability window (0 – 5 V vs Li^+/Li), a high electrical resistivity (>10^{15} Ω.cm), and a reasonable ionic conductivity (10^{-6} S/cm). To verify the ionic conductivity of our LiPON layer, we perform the electrochemical impedance spectroscopy (EIS) experiment on MIM structure with different electrode areas (see Figure III. 6).

We propose a simple equivalent circuit as in Figure III. 6.a to fit the impedance response of the test structure. Here, the first part is a $\mathbf{R_{system}}$ represents the resistance contributed by the measurement system (wire and probe contact). The second part of the circuit is composed of two components in parallel: a resistance $\mathbf{R_L}$, which characterizes the ionic conductivity of the electrolyte, and a constant phase element (CPE) $\mathbf{Q_1}$, corresponding to the dielectric behaviors of LiPON layer. Finally, a second CPE $\mathbf{Q_2}$ emulates the double layer capacitance of this MIM capacitor. The impedance of the CPE is denoted by $Z_{CPE} = 1/Q(i\omega)^n$. With this circuit model, we can obtain the characteristic frequency of LiPON f_{c_LiPON} at the middle of the semicircle curve. Furthermore, the value of LiPON resistance (R_{LiPON}) can be extracted from the diameter of the semicircle. By measuring R_{LiPON} of MIM devices with 100-nm thick LiPON and the areas ranging from 19.6 mm^2 to 33.2 mm^2, we can calculate the ionic conductivity of LiPON (σ_{ion}) by following the relation:

$$R_{LiPON} = \left(\frac{1}{\sigma_{ion}}\right) \times \left(\frac{l}{S}\right), \qquad \text{Eq. 1}$$

where l is the thickness of the LiPON and S is the area (Figure III. 6.c).

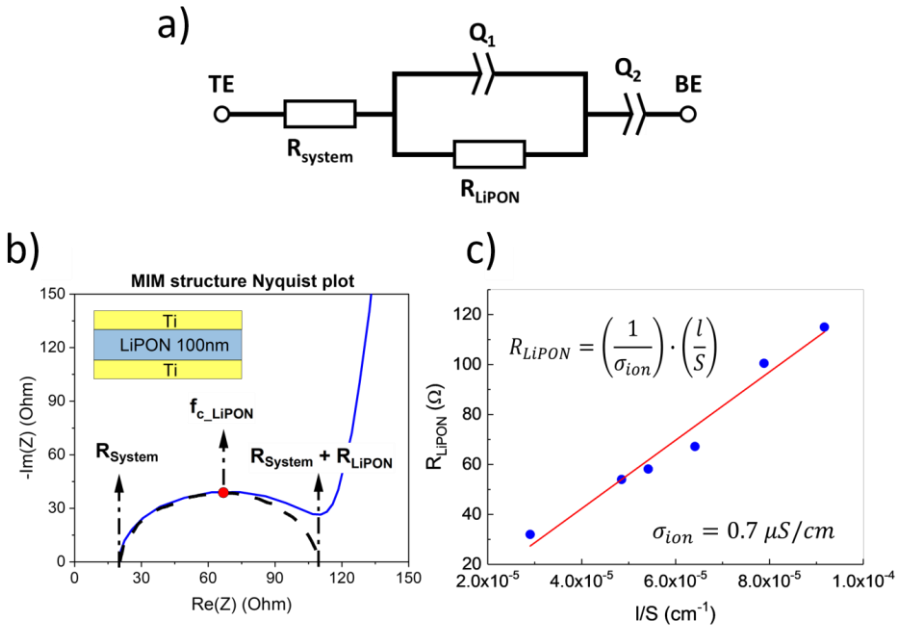

Figure III. 6: a) EIS spectra of a Ti/LiPON/Ti MIM structure. b) Deduced ionic conductivity of LiPON electrolyte from EIS spectra.

2 PRELIMINARY STUDY OF SYNT WITH {$LiCoO_2$/LiPON} GATE STACK

2.1 Electrical measurement setup

Before moving to the characterizations of the first realized synaptic transistors with $LiCoO_2$ channel and LiPON electrolyte, we will discuss the setup employed for both electrical and electrochemical measurements. The performance of the devices are analyzed using multi-channel VMP3 potentiostat (Bio-Logic) coupled with ultra-low current modules, allowing current in the range of nA to be measured with precision (See Figure III. 7.a). To prevent external electromagnetic noise sources, we employ also a Faraday box to shield our electrical measurements. The potentiostat is controlled by the EC-Lab software on an Ethernet-connected computer.

There are two types test devices available on the elaborated wafers, and they correspond to two types of connection on the VMP3. First, the battery-like structures refer to the cells with parallel electrodes, sandwiching an active material (cathode) and a layer of ionic conductor (electrolyte) (Figure III. 7.b). These devices allow us to closely study the electrochemical reactions and processes happen at the active material and the interfaces by performing techniques such as cyclic voltammetry, electrochemical impedance spectroscopy, Galvanostatic cycling, et cetera. To test these two-terminal structures, "working electrode" (denoted as the red probe and "+") is probed to the bottom side and "counter electrode" (denoted as the blue probe and "–") is probed to the top one. These electrodes are connected in the "Standard mode" where the working electrode is connected to CA2 and Ref1 (Ref for reference, and CA for current amplifier). The counter electrode to CA1, Ref2. Ref1, Ref2 are used to measure the voltage difference, e.g., the cell potential is measured by the potential difference between Ref1 and Ref2. CA2 and CA1 to apply the current in Galvanostatic mode.

Second, the principle devices are the transistors with Gate, Source and Drain electrodes. The active material now serves as the channel connecting the drain and source electrodes, and the electrolyte film lays between the gate electrode and the channel (See Figure III. 7.c). It is possible to use this potentiostat to work with multiple-terminal devices by switching the connection mode to "CE-to-Ground". In this mode, the counter electrodes, probing on Source electrode, are connected to the Ref2 and they are grounded. The other two working electrodes, one for Gate programming and one for Drain reading, are connected to the Ref3 and CA1 lines. With this connection, one can program the transistor by applying potential pulses on gate and source electrodes, and perform current read operations on drain and source electrodes without interfering the other. I will present the characterizations of the first gate stack in the next section.

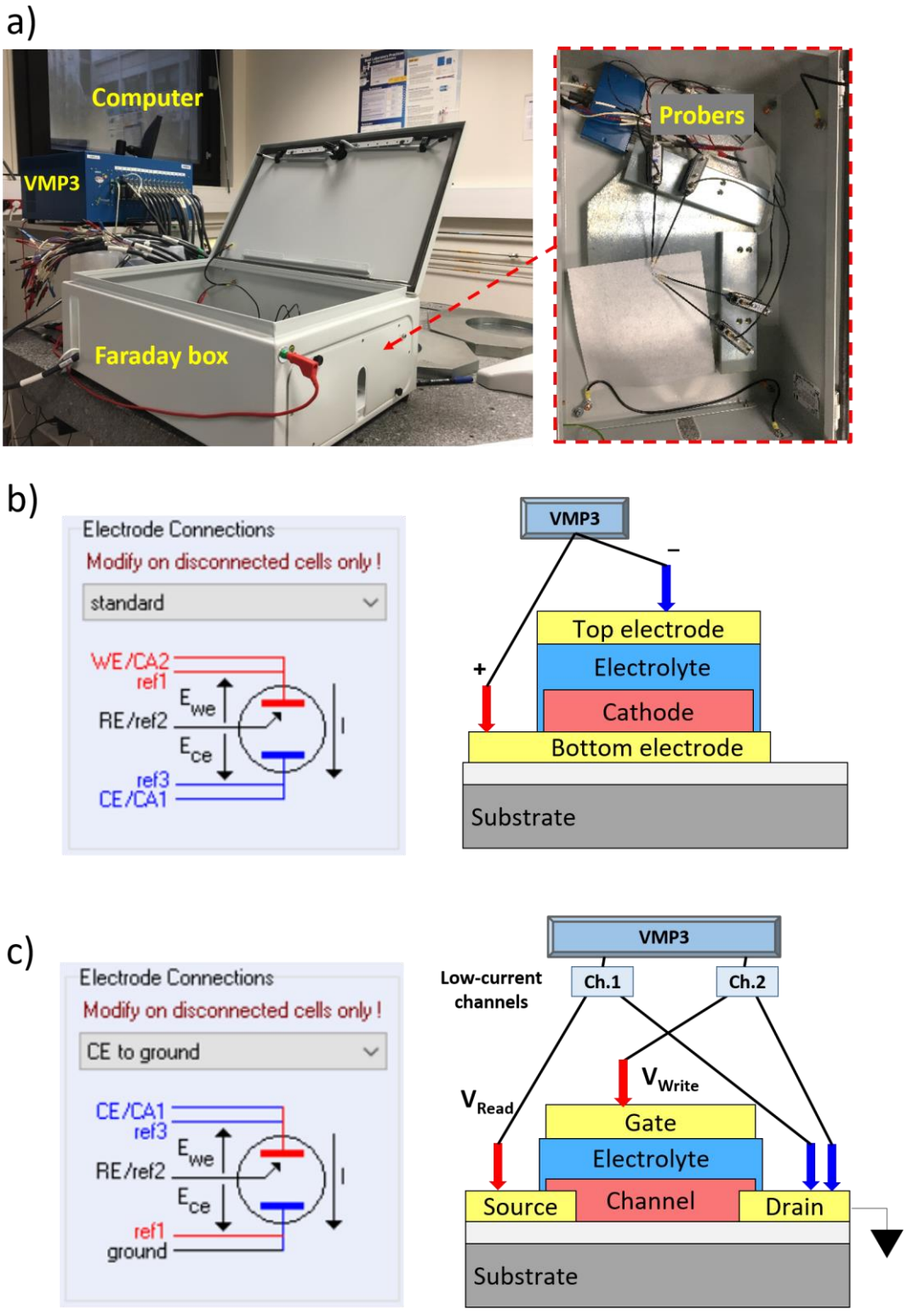

Figure III. 7: a) The electrical setup used to characterize electrochemical systems in this thesis. Schematic illustration of b) a two-terminal structure measurement and c) a transistor measurement.

2.2 Electrochemical characterization of the gate stack

We elaborate a wafer with devices comprising of 500 nm LCO and 500 nm LiPON electrolyte. An EIS test was done on a battery-like device with area $S = 1.13\ \mathrm{mm}^2$ to evaluate the electrochemical contribution of the vertical gate stack (Figure III. 8.a). The frequency was scanned from 1 Hz to 1 MHz and the complex impedance response was recorded and presented in Figure III. 8.b. The diameter of the semicircle impedance indicates the ionic resistance of the LiPON electrolyte, whose value is $R_{LiPON} = 5.2\ k\Omega$. The ionic conductivity is hence, $\sigma_{LiPON} = 0.85 \times 10^{-6}\ S/cm$, corresponding well to the reported value of this sputtered electrolyte. Figure III. 8.c shows the first cyclic voltammetry (CV) curve for the HT-$LiCoO_2$ active layer between 3.3 V and 4.2 V at a scan rate of 0.5mV/s. Li^+ ions can be intercalated into/extracted from Li_xCoO_2 active material following the redox reactions:

$$Li_1CoO_2 \leftrightarrow Li_{1-x}CoO_2 + xLi^+ + xe^- \qquad \text{Eq.2}$$

The CV curve for the HT-LCO exhibits one major cathodic peak and one anodic peak at 4.01 and 3.84 V respectively. There exists also two minor cathodic and anodic peaks at 4.09, 4.19 and 4.03, 4.15V, respectively. The pair of major redox peaks correspond to the first-order phase transition, while the two pairs of minor redox peaks correspond to the order-disorder phase transitions [39]. The inset illustrate the result of the whole scan from the initial OCV (0.43 V). We can observe that there is no activity recorded up to the reaction potentials around 4.0 V. With these results from electrochemical tests, we can verify the contribution from LiPON electrolyte and the HT-$LiCoO_2$ active material.

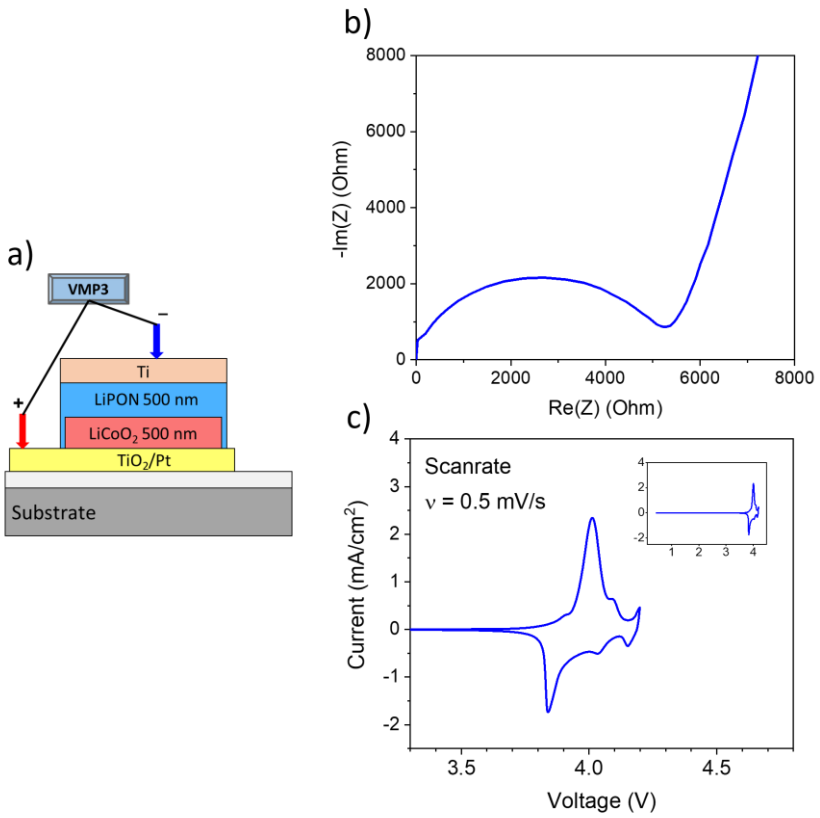

Figure III. 8: Electrochemical study of $LiCoO_2$/LiPON gate stack. a) The illustration of the battery-like structure used. b) EIS study of a test structure with an area of 1.13 mm^2. b) CV study of the same device showing the redox potentials.

2.3 Electrical characterization of the transistor

2.3.1 Resistive switching phenomenon

The first electrical characterizations of SynTs were dedicated to verify the IMT transition of LCO channel layer (see Figure III. 9.a). We extracted the Li ions by applying a constant potential of 4.2 V on the gate-source electrodes during 300 s. The change in conductivity of the channel was recorded with a constant potential of 0.1 V applied between the drain-source electrodes. Figure III. 9.b depicts the exponential current growth of over 5 decades in less than 50 s of ion extraction (similar to the charge process), and this agrees well with the reported IMT in the literature [19].

With this information of phase transformation, it is interesting to test if we can retrieve the insulator phase by intercalating Li ions into the channel. To do that, we performed a voltage scan on gate-source electrodes, with the voltage from the out-of-fab OCV of 0.7 V to 4.2 V. Then, the scan was reversed to –3.4 V and ended at 0 V. From Figure III. 9.c, the conductance does not rise until V_G reaches 4.1 V. At this potential, the

conductance of the channel rises significantly to 700 µS. These features agree well with the CV inspection of the gate stack. In the reversed direction, the ions are gradually intercalated into the channel, driving the conductance to decrease. The timescale of the reversed reaction is much slower, thus, leading to the difference between the rising and the falling slopes, which creates a hysteresis curve. This type of hysteresis is typical for non-volatile memory devices as it indicates the resistive switching phenomenon [8].

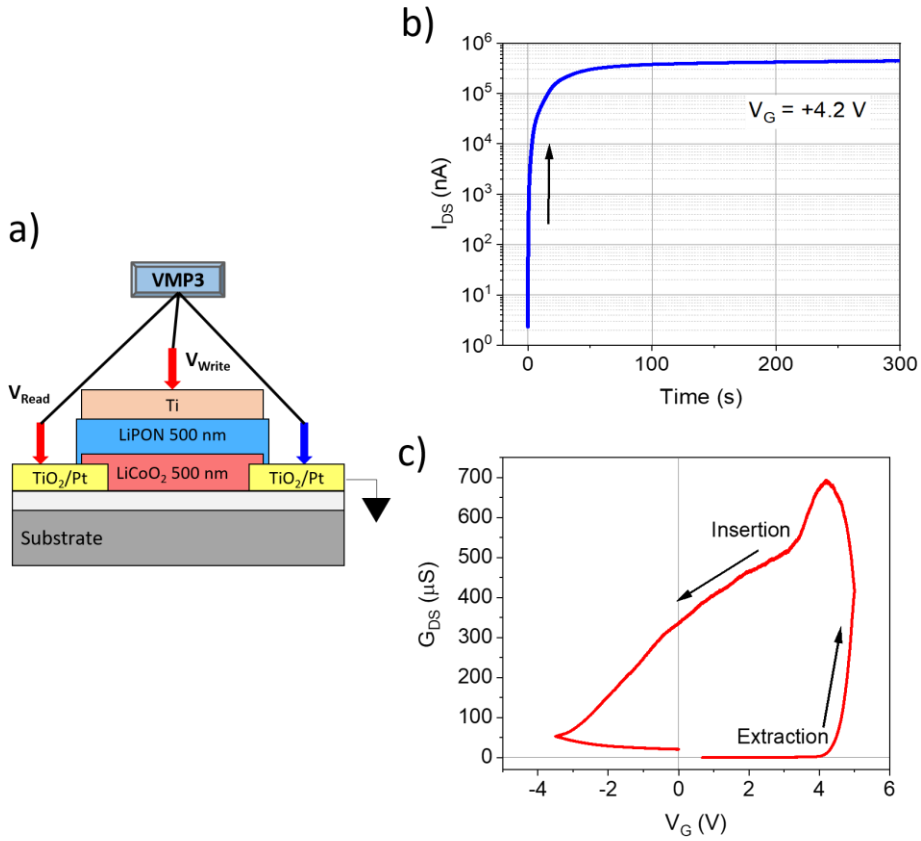

Figure III. 9: Electrical test of the synaptic transistor. a) The illustration of the SynT with a 500-nm $LiCoO_2$ channel and 500-nm LiPON electrolyte. b) The evolution of channel conductance under a constant extraction of Li ion at V_G = 4.2 V. c) Hysteresis conductance profile of the channel under Li intercalation.

2.3.2 Analog state programming

In this part, we study the analog programming of multiple states with this transistor using potential pulses. In Figure III. 10.a, we show the pulse scheme used for the writing and the recorded profile of drain-source conductance (G_{DS}). The pulses with V_G = 4.2 V are used to increase the channel conductance while the pulses with V_G = -4.0 V are applied to drive down the conductance states. After the application of pulses in 1 s, the gate voltage is switched OFF during 4 s (rest state). Here, the train of 5 up pulses results in a stepped increase in G_{DS} from 870 to 910 µS. During "write" operation, an increase in the G_{DS} is

observed due to Li leaving the channel, while during each "read" operation G_{DS} remains constant at a value corresponding to the new conductance state. This process is reversible, and a subsequent train of 5 down pulses returns the device to the initial conductance level. The consecutive programming was done with 40 pulses potentiation and depression per cycle with the same pulse scheme (+4.2 V up and – 4.0 V down, t_d = 1s). In this experiment, the rest state time was reduced to 1 s (See Figure III. 10). As a result, the channel conductance varies from around 1.3 mS to 1.5 mS. These tests confirm the concept of synaptic transistor of $LiCoO_2$/LiPON gate stack.

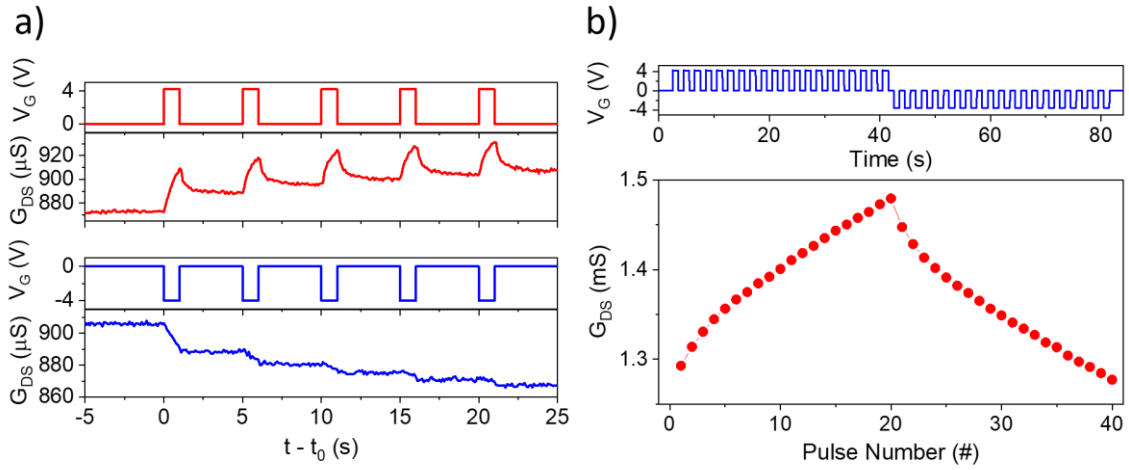

Figure III. 10: a) Analog conductance modulation demonstrated with five pulses of V_G = +4.2 and -4.0 V amplitude and 1 s duration. b) One cycle of conductance modulation with 40 pulses of potentiation and depression pulses.

2.4 Summary

In conclusion, we are able to realize the first functional wafer-scale and all-solid-state synaptic transistors made of $LiCoO_2$/LiPON gate stack. The battery-like structures allow us to decorrelate the contribution of the active material and the electrolyte. For the LiPON layer, the ionic conductivity and characteristic frequency are determined with EIS technique to be in line with the reported values of this material. With cyclic voltammetry, the redox potentials representing the phase transitions of HT-$LiCoO_2$ layer are clearly identified. Electrical tests focusing on the conductance modulation of the channel are done to confirm its reversible IMT under the intercalation of Li ion from the gate. With a train of programming pulses, the analog conductance state modulation is demonstrated with a cycle of 40 pulses and a channel conductance varying from 1.3 to 1.5 mS.

The realization and characterizations of the first devices allow great understanding of the microfabrication process flow and the operation principles of these electrochemical systems. However, the operation scheme and performance of the transistor can be further

improved. First, the working conductance of this device is in the range of mS, which can be costly to implement into a complex neural network comprising millions of transistors. The programming voltage is considerably high for this gate stack (~4 V). In addition, a long programming duration of 1 s of V_G pulses can contribute to the overall energy consumption of this transistor. These problems can be addressed by thinning down the channel material down to a few tens of nanometers, which can reduce the conductance and the programming time thanks to the reduced diffusion length of ions into the material. Furthermore, working with top electrode made of the same materials with the active material can reduce significantly the chemical potential difference, thus, minimizing the programming amplitudes. For these reasons, we explore the ultra-thin film Li_xTiO_2 for the channel material as a next step to optimize the performance of our first generation SynTs.

3 EXTENDED STUDY OF SYNT WITH {Li_xTiO_2/LiPON} GATE STACK

3.1 Artificial electrochemical synapse

Figure III. 11 shows an illustrative view of our micro-fabricated SynT device. The cell core (synaptic element) is a vertical stack consisting of an ultra-thin (10nm) titanium dioxide (TiO_2) channel, an electrolyte made of amorphous lithium phosphorus oxynitride (LiPON), and a top gate made of Ti. This vertical configuration was considered to allow for shorter diffusion path, thus increasing the operational speed [4], [40].

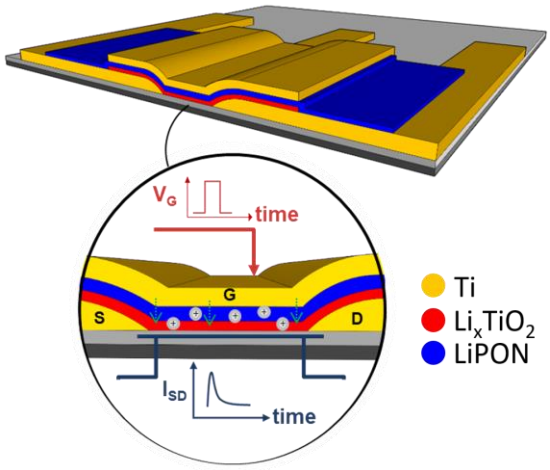

Figure III. 11: Schematic view of our SynT with Li_xTiO_2/LiPON gate stack.

Amorphous TiO_2 was selected as a channel material because of its well-studied intrinsic merits as suitable host for Li-ion intercalation in energy storage applications [41]–[43]. Additionally, TiO_2 in its amorphous form is known to exhibit pseudocapacitive characteristics which allow fast, reversible ion intercalation without phase transition [44], [45]. Furthermore, TiO_2 undergoes an insulator-to-metal transition upon ion intercalation, thus making it an appealing material for ion-based SynTs [32], [33]. LiPON has been chosen as solid-state electrolyte for its high chemical and electrochemical stability [46]–[49], and scalability [37], [50].

The inset of Figure III. 11 shows a schematic cross-section of the transistor. Square-shaped voltage pulses applied to the gate mimic the biological impulses of the pre-synaptic neurons. Under these spikes, Li^+ ions are inserted into the channel material (red layer), connecting the source and drain electrodes via the electrolyte (blue layer), hence creating a change in electrical conductance. This conductance change is captured by

electrically sampling with a small, constant potential bias (V_R) the current flowing between the source and drain.

3.2 Physical characterizations

The cross-section of a SynT is prepared using FIB technique for the SEM/EDS and TEM characterizations. From Figure III. 12.a (top panel), we can observe that the channel length (between source and drain electrodes) is about 3 µm, using Scanning Electron Microscopy (SEM). The LiPON layer thickness is 200 nm, as desired. The elemental mapping of the layers is observed with the help of energy dispersive X-ray spectrometry (EDS), where elements P and O represent the LiPON electrolyte. Furthermore, no interfacial inter-diffusion is discerned with the detectable elements.

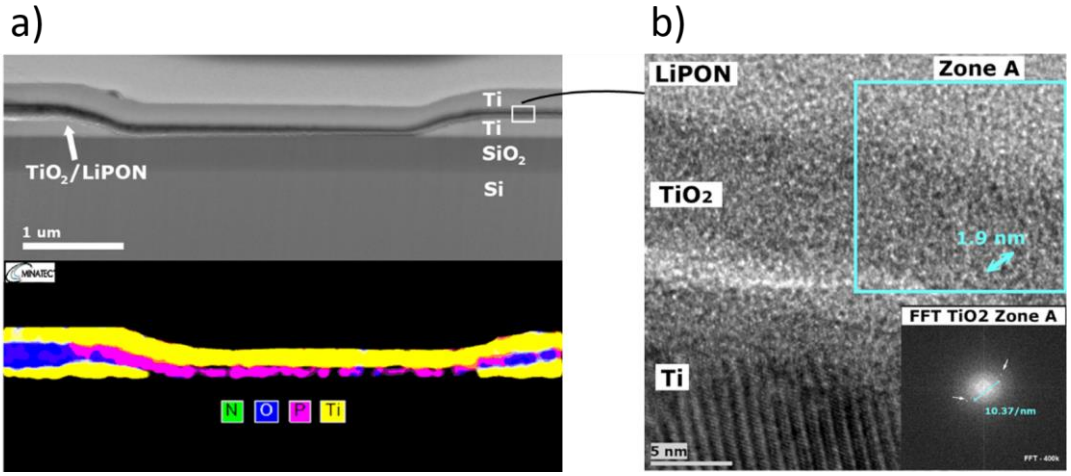

Figure III. 12: a) SEM and EDS characterizations of SynT's gate stack. b) HRTEM image with the focus on the neighborhood of TiO2 channel (inset: FFT on the selected Zone A)

High-resolution transmission electron microscopy (HRTEM) has been carried out to inspect the channel's structural properties (Figure III. 12.b). The stacked layers are differentiated based on the thicknesses and their contrasts. The 10nm thickness of the TiO_2 layer is clearly confirmed from the image. Besides, the HRTEM analysis gives valuable information about channel film properties, indicating the presence of 2 nm size nanocrystallites embedded in the amorphous matrix. The structure of TiO_2 channel layer appears to be an amorphous one. However, by doing local FFT, we observed some spots in the diffractogram. They are underlined by white arrows. The distance between spots are measured and indexed with reference to Si atomic planes. Si substrate images were acquired in a similar magnification as the TiO_2 images, and local FFT images were then done from comparable field of view (FOV), 13x13 nm^2 for 400k magnification. The known atomic plane distances of silicon served as reference for calculating a constant value

adapted to the magnification. The extracted constant value is 21.45 for x400k. Once the constant value was determined, we calculated the interatomic distance in real space d = 0.207 nm from the distance d* in the reciprocal space, d* = 10.37/nm in this case. The distance corresponded to the (210) crystal phases of the TiO_2 rutile (JCPDS no. 21-1276). Same experiment was done with 300k magnification on other nano-crystallites and the results showed that the major part of the crystal structure was TiO_2 rutile.

3.3 Electrical characterizations

The first electrical tests were done to confirm the conductance modulation phenomenon under the intercalation of Li ion. Figure III. 13.a depicts the evolution of the SD channel conductance G_{SD}, by application of a bidirectional sweeping gate voltage from -3.0 V to 3.0 V (at a rate of 50 mV/s). A small bias of 0.1 V was applied to read the change in Source-Drain current (I_{SD}). Initially, G_{SD} is very low ($G_{SD} \leq 20$nS at $V_G = 0$ V). Then G_{SD} increases up to a 100 times higher value, reaching 250 nS, due to intercalation of Li^+ ions into the TiO_2 quasi-amorphous channel. For the backward sweep (Li^+ extraction), G_{SD} decreased gradually back to its low conductance state, exhibiting a clear counter-clockwise hysteresis pattern. We can see that the highest slopes (which correspond to a more effective conductance variation) are located within the [-0.5 V, 1.5 V] V_G potential region, whereas other regions demonstrated slow, saturated modification of channel conductance. For this reason, we selected the [-0.5 V, 1.5 V] potential window (inset of Figure III. 13.a) to develop our SynTs. More details on the underlying electrochemistry of the Li^+ doping mechanism under different potential windows will be discussed in a following section.

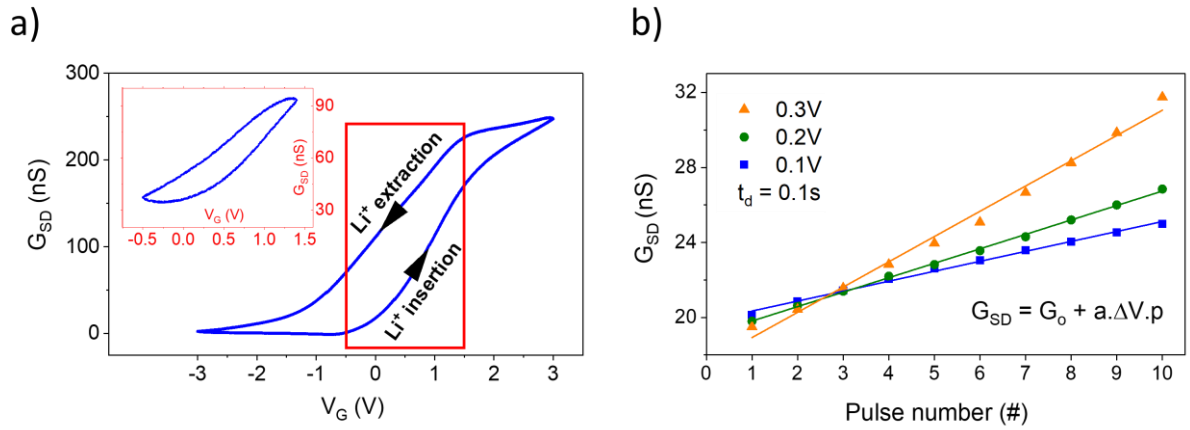

Figure III. 13: a) Charge transfer curve (channel conductance GSD as a function of gate voltage VG) with a gate sweep of 50 mV/s in the potential range [-3 V, 3 V] (inset: Focused working window of [-0.5 V, 1.5 V]). b) Conductance GSD change under incremental voltage amplitude pulses from 100 mV to 300 mV with a duration td of 0.1 s.

To demonstrate the ability to modify the analog states required for an artificial synapse, we programmed SynT with a train of 10 voltage pulses with different amplitudes relative to the open-circuit voltage (OCV) measured between the gate and source electrodes at rest. Note that in electrochemical devices, chemical potential gradients are generated by modifying the ionic content of one electrode, and this phenomenon is termed nanobattery effect.[51] Therefore, it is essential to program electrochemical synaptic transistors with a gate potential V_G whose amplitude takes into account the OCV values of the cell. After the application of each pulse, the gate terminal was switched OFF for 1 s, when the READ action occurs, to prevent the electron movement and perturbation of the programmed state. Figure III. 13.b illustrates the change of conductance states (ΔG) under the influence of WRITE pulses with the same duration of 0.1 s but different magnitudes (ΔV), from 100 mV to 300 mV. The pulses of higher amplitude result in a more pronounced change of conductance states. A simple linear fitting better shows the relation between the changes of SD conductance and the pulse amplitudes. Here, one can observe that by varying the pulse amplitude, we are driving the SynT channel conductance along the "Li^+ insertion" conductance curve from Figure III. 13.a but at a different V_G sweeping rate.

Maintaining the programmed states is of great importance to synaptic elements in an artificial neural network for a high learning precision [52]. In Figure III. 14.a, b, we demonstrate the retention of the SynTs with both potentiation and depression (increase and decrease of conductance level respectively). A series of pulses with ±200 mV magnitude and 0.5 s duration was used for programming the device. The pulses were followed by a resting time of 50 s while the G_{SD} was recorded.

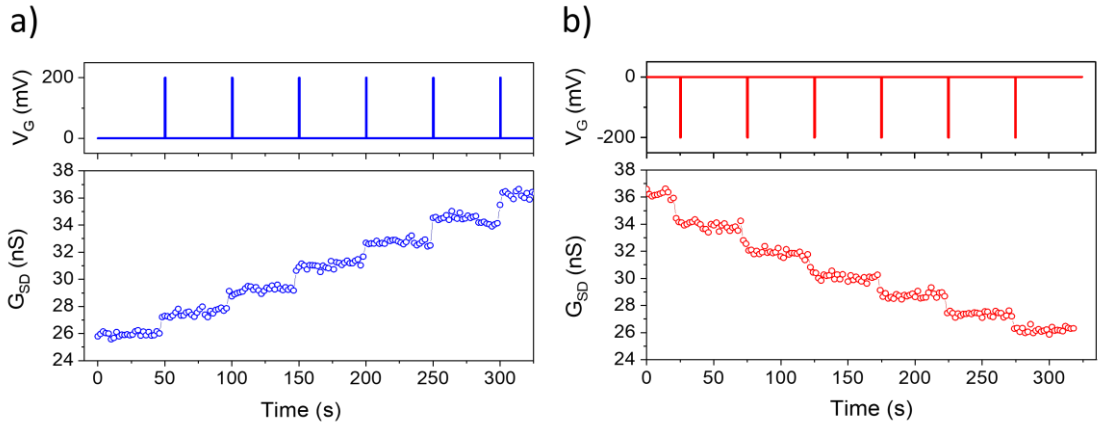

Figure III. 14: a) Retention (during 50s) of GSD after each potentiation pulse (200 mV, 0.5 s), and b) Retention of GSD after each depression pulse (-200 mV, 0.5 s). The two experiments are performed on the same device without any disruption.

We further checked the stability by applying a voltage pulse (500mV, 0.5s), then switched off the gate and measured the conductance evolution during over 4000 s (See

Figure III. 15). At 4000 s, the conductance decrease is lower than 15%. The initial small drop (~5%) could be explained by the discharge of the capacitor at the LiPON interface with a fast time constant. After this, the channel conductance decreases only very little, and at a much slower rate. Overall, we can see in this example that the programmed states appear quite stable. Such a stability may be due to ions intercalation (inside the channel) via Faradaic reactions.

To optimize even further the state retention of electrochemical synaptic transistors, several possibilities may be explored in the future. First, the chemical potential difference (while operating) can be diminished by designing the gate and the channel of the transistors to be the same materials (see for example REF [4]). Secondly, an additional electronic element (memristor/transistor) can be implemented on the gate electrode to act as a physical switch; this switch can serve as a selector for precise programming when the synaptic transistors are implemented in a crossbar-array structure (see REFs [4], [13], [53]).

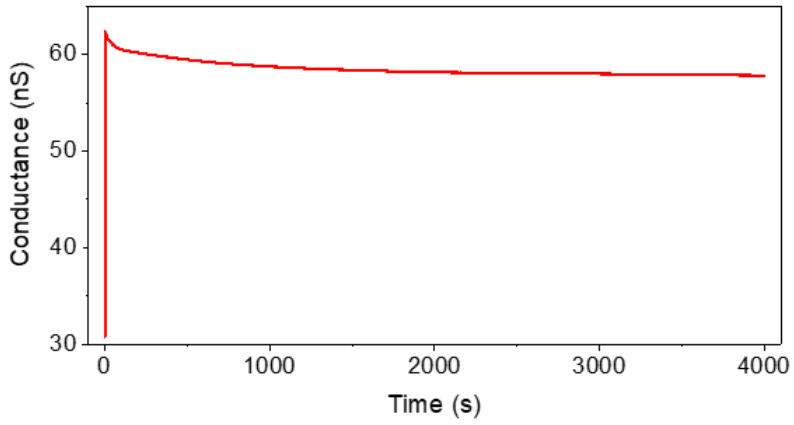

Figure III. 15: Retention profile of a conductance state over 4000 s.

Synaptic functions required for artificial synapses were demonstrated in this device. As shown in Figure III. 16.a, emulation of neuromorphic behavior, such as long-term potentiation (LTP) and long-term depression (LTD) was achieved by alternatively programming the SynT with 50 identical pulses (±100 mV, 0.1 s) and settling and reading time of 1 s. The conductance states are modified in an analog way from a low conductance level of 28 nS to a high conductance level of 74 nS, which corresponds roughly to a 1 nS increase from one state to the next one. The device-to-device variation was confirmed to be small across SynTs by conducting the same characterization in multiple devices (see Figure III. 16.b).

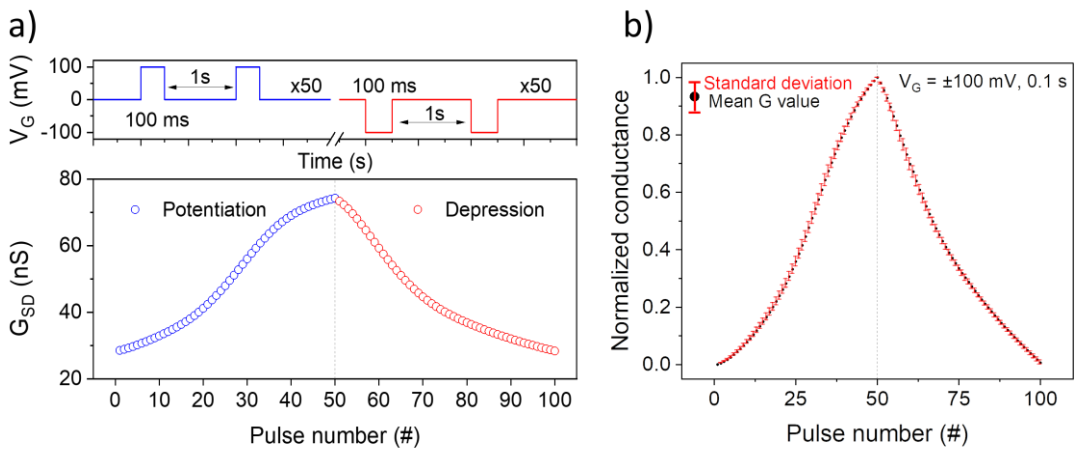

Figure III. 16: a) Long-Term Plasticity demonstration with 100 states of potentiation and depression. b) Conduction modulation with 100 states performed on 5 different SynTs.

The injected charge over 50 operations has been recorded in order to estimate the amount of energy to program a state to an adjacent one. The energy consumption for SynT devices is calculated by the expression:

$$E_w = \Delta Q \times V_w \qquad \text{Eq.3}$$

where ΔQ is the injected charge and V_w is the voltage used for programming. The average charge transferred for SynT was 90 pC, and the voltage used for writing is 0.1 V. Therefore, 9 pJ is spent for each writing operation. For our SynT with a TiO_2 area of 70x80 µm^2, we obtain the normalized energy consumption of 1.6 fJ/µm^2. Assuming the energy per write operation is directly proportional to the channel area, we obtain a projected programming energy of 16 aJ for a scaled 100x100 nm^2 device. Hence, together with the conductance levels in the range of nano-Siemens, our SynT can be considered as one of the most energy-efficient all-solid-state synaptic devices realized in both READ and WRITE operations.

The symmetry property of conductance modulation or the weight update between potentiation and depression processes is characterized by the asymmetric ratio (AR), defined as

$$AR = \left[\frac{\max|G_p(n) - G_d(n)|}{G_p(50) - G_d(50)}\right] \; for\; n = 1\; to\; 50, \qquad \text{Eq.4}$$

where $G_p(n)$ and $G_d(n)$ are channel conductance values at the nth state after the potentiation and depression pulses. For our SynT, the AR was calculated to be 0.31 (in Figure III. 16.a) for 100 pulses per cycle, indicating a good symmetry in comparison to results reported in literature [40], [54].

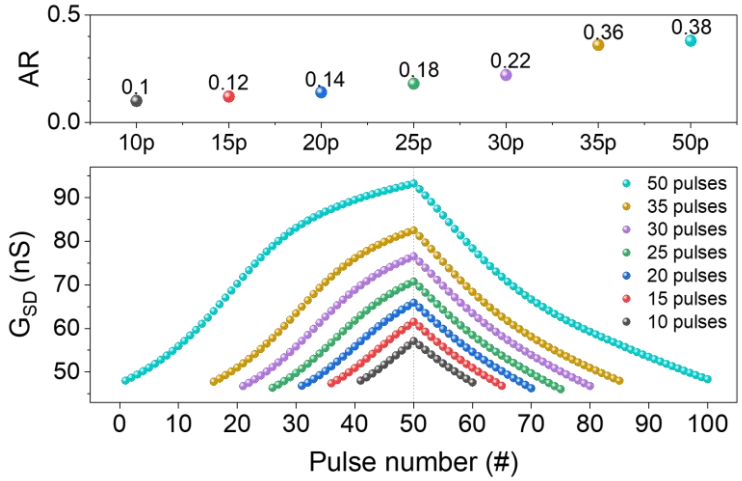

Figure III. 17: Correlation between the number of conductance states and the asymmetric ratio.

To analyze the relationship between the number of states per cycle and its AR, we performed a series of programming cycles on a device with a different number of intermediate states and analyzed the corresponding AR (see Figure III. 17). The number of gate pulses increases from 10 to 50 pulses, yielding an increase of AR from 0.1 for 10 pulses to 0.38 for 50 pulses. Thus, one needs to consider the compensation between the desired programming states and their AR for each application.

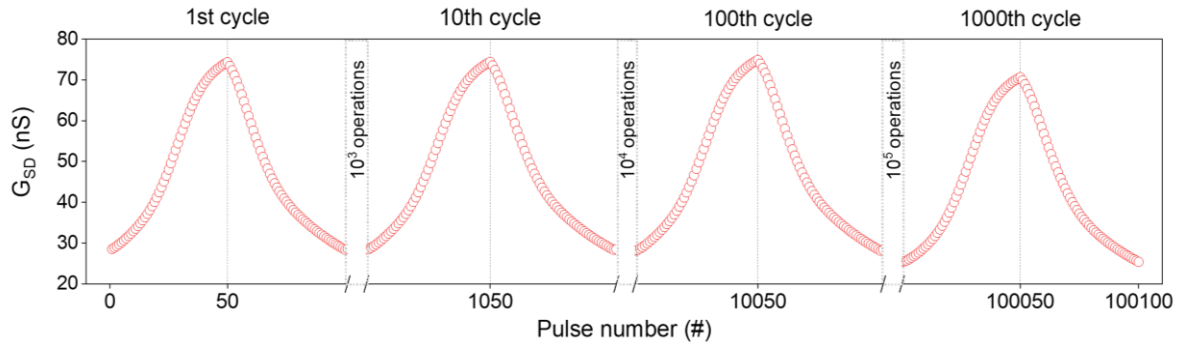

Figure III. 18: Endurance test with 1000 cycles and 10^5 operations shows no degradation.

Figure III. 18 demonstrates the endurance test of SynT after more than 1000 cycles, with 100 weight update operations in each cycle and with the same pulse scheme as in Figure III. 16.a. The endurance of this device is high since its small change of Max-Min conductance after 1000 cycles is calculated to be

$$\Delta G = (G_{max} - G_{min})_{1st} - (G_{max} - G_{min})_{1000th} = 0.5\,nS \qquad \text{Eq.5}$$

This endurance of 10^5 weight updates is considered largely sufficient for ANN pattern recognition training using artificial synaptic hardware, since not all of the synapses will be repeatedly updated in the course of the training epochs [55].

By scaling down the channel to 10 nm, our amorphous TiO_2 film exhibits extrinsic pseudocapacitive behavior, which allows simultaneously for ultra-low energy consumption and fast/reversible conductance modulation. Both features are indispensable for artificial synapse application. We can observe that low power dissipation is the highlight that this artificial synapse offers. The operational conductance (a few tens of nS) is comparable or lower than other types of three-terminal SynTs. In addition, by reaching an outstanding low energy consumption per spike (1.6 fJ/µm^2), close to the biological energy range of femtojoule, our present SynTs appear as an excellent candidate for large-scale energy-efficient neural networks. This high energy-efficient property stems from the choice of a resistive yet ion-intercalation-sensitive channel material TiO_2. Furthermore, with a 10 nm TiO_2, the intercalation process of Li ion happens at a fast pace without the solid-state diffusion limit. This has an important effect for SynTs because the conductance modulation can be stimulated by fast gate voltage pulses. This important phenomenon will be discussed thoroughly in the following section.

3.4 Electrochemical characterizations

In the specific context of electrochemical synaptic transistors, there is a correlation between conductance modulation and the various ion exchange reactions taking place in the electrodes' bulk and at the interfaces with the solid electrolyte [56]. In the following, the study of a vertical Li_xTiO_2/LiPON/Li structure has been carried out in an effort to investigate the channel Li_xTiO_2 electrochemical reactions and the relative uncorrelated effects on our SynT electrical response.

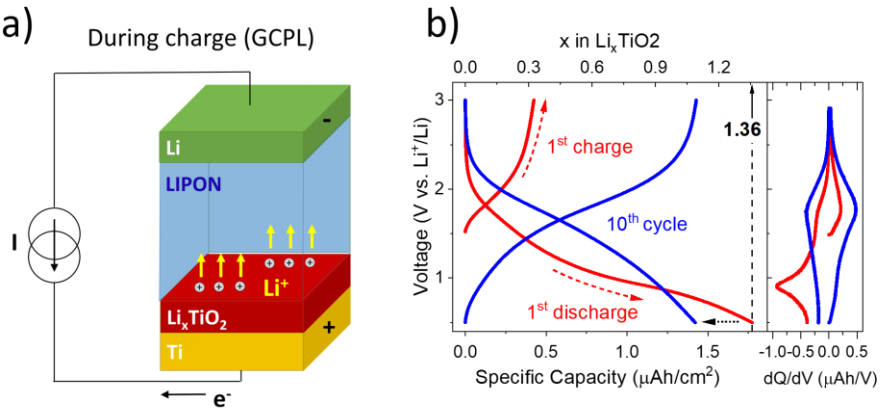

Figure III. 19: a) Schematic view of Ti/Li_xTiO_2/LiPON/Li vertical structure during GCPL charge process. b) Comparison of 1st and 10th cycle upon GCPL (The cell is charged to 3.0 V and discharged to 0.5 V with the current density of ±2.6 µA/cm^2).

The vertical structures (Figure III. 19.a) showed an average OCV around 1.5 V after fabrication, thus substantiating a diffusion of Li+ ions into TiO2 during LiPON deposition [47], [57], [58]. A first charge (delithiation) capacity corresponding to an initial Li0.32TiO2 stoichiometry (Figure III. 19.b) corroborated this fact. Furthermore, the following discharge (lithiation) exhibited an unexpected potential profile (1 V plateau) and a high (Li1.3TiO2) capacity; both characteristics have been already reported [44], [45] and attributed to an activation process [59]. Subsequent cycling curves have consistent features, revealing the highly reversible ion (de)intercalation reactions.

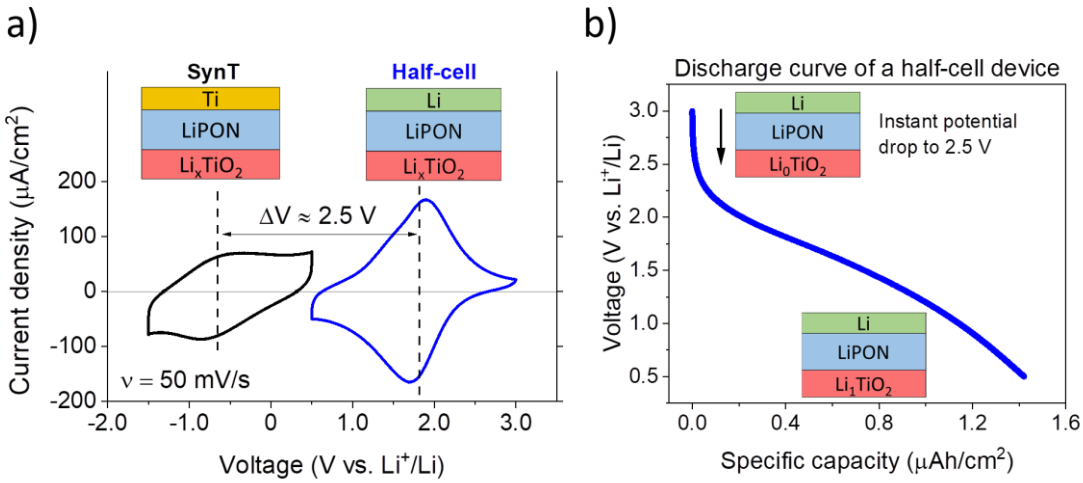

Figure III. 20: a) Voltammogram comparison of SynTs and vertical test structure. b) Discharge curve of the half-cell showing the cell potential drops from 3 V to 2.5 V when Li metal facing Li depleted TiO_2 (Li_0TiO_2).

Figure III. 20.a presents results of cyclic voltammetry analysis that has been carried out in order to compare the TiO_2 response upon ion (de)intercalation in two configurations: lithium excess (vertical structure) and deficiency (SynT). Li^+ ions can be intercalated into TiO_2 with the consecutive reduction of Ti(IV) to Ti(III) redox centers, given as:

$$TiO_2 + xLi^+ + xe^- \leftrightarrow Li_xTiO_2 \qquad \text{Eq.6}$$

with $0 \leq x \leq 1$. In a lithium excess configuration, TiO_2 films show broad anodic and cathodic peaks (at 1.88 and 1.66 V respectively) corresponding to Li^+ ion intercalation. However, in a lithium deficiency configuration, anodic and cathodic peaks reside at -0.5 and 0.8 V respectively. The peak shift of around 2.5 V corresponds to the difference of top electrodes (Li metal and Ti metal). We notice from our experiments that there is an oxide film formed on the interface between Ti top metal and LiPON. Thus, the potential difference is now between Li and TiO_2, which is illustrated in Figure III. 20.b. These redox peaks of SynT exhibit lower current densities due to insufficient ion quantity for a complete (de)intercalation reaction ($x_{available}$ = 0.32). Notwithstanding, the potential window of the redox peaks corresponds to the largest variation of electronic conductance, highlighting the correlation between both phenomena (See Figure III. 21). From the voltammogram of

the gate stack, we can see that the shortage of Li supply led to the wide shape of oxidation and reduction peaks, and the Li intercalation process will become capacitive instead. As a result, one may observe a slow change of SD conductance at the end of voltage sweeps.

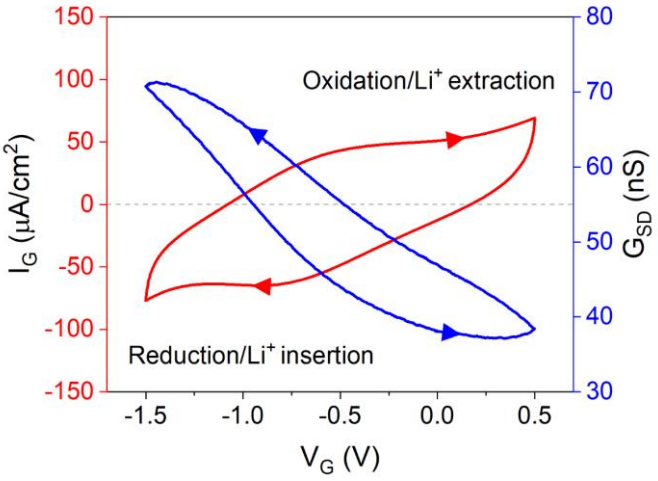

Figure III. 21: SynT channel conductance modulation (blue curve) and gate current response (red curve) as functions of the sweeping of gate voltage.

To further investigate the Li^+ ion (de)intercalation kinetics, additional CV experiments of vertical structures were performed at varying scan rates (Figure III. 22). Cyclic voltammetry was performed on a cell with an electrode area of 0.375 cm^2. The cell's potential was linearly swept from 0.5 V to 3.0 V vs. Li electrode at increasing speed from 1 mV/s to 1 V/s with 10 cycles for each rate. The current response at each scan rate is illustrated in Figure III. 22.a. The recorded current density increases with the increase of scan rate due to the reduction of diffusion layer thickness.

A closer examination of the CV scan rate dependence allows discriminating quantitatively the contributions of diffusion and surface controlled processes to the current response. At a fixed potential V, the current measured has a linear relation with the scan rate $i(v) \propto v \rightarrow i(v) = k_1 v^1$ for a capacitive process. In a Faradaic process, the relation of current and scan rate can be deduced from Cottrell equation: $i(v) \propto v^{1/2} \rightarrow i(v) = k_2 v^{1/2}$. Then, the current measured by CV can be decomposed as the following equation [28]:

$$i(v) = k_1 v + k_2 v^{1/2} \qquad \text{Eq.7}$$

where i, v, k denote the current measured, the scan rate applied and the process coefficient respectively. If we divide the Eq.7 by $v^{1/2}$, we will obtain:

$$\frac{i(v)}{v^{1/2}} = k_1 v^{1/2} + k_2 \qquad \text{Eq.8}$$

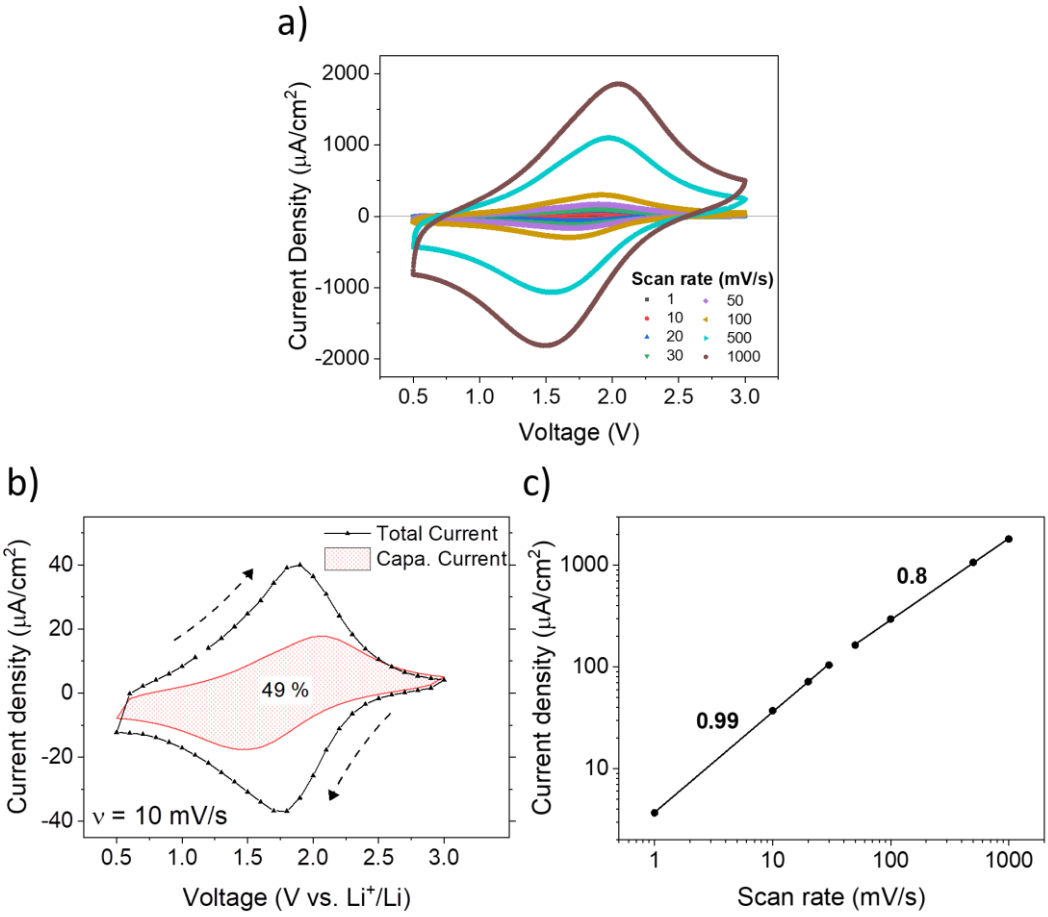

Figure III. 22: a) Cyclic voltammetry of TiO_2/LiPON/Li half-cell at increasing scan rates from 1 to 1000 mV/s. b) CV study at 10 mV/s and the calculated pseudocapacitive current contribution. c) Logarithmic relationship between cathodic peak current (Li^+ insertion) and scan rates between 1 mV/s and 1 V/s.

Hence, using Eq.8, we can fit the k_1 and k_2 coefficients as the slope and y-intercept from the data of measured current at different scan rates. These coefficients represent the contribution of Faradaic and non-Faradaic currents.

Interestingly, pseudocapacitive contribution dominates, overwhelmingly, the stored charge in TiO_2 over the entire potential window at scan rates between 10 mV/s and 1 V/s (51% and 90%, respectively). Consequently, a higher SynT performance is expected using such channel material insofar as diffusion-controlled processes are completely inefficient in these conditions.

The reaction kinetics is resolved by examining the variation of peak current (i_p) with scan rate (v) using the power-law relationship [60]:

$$i_p = av^b \qquad \text{Eq.9}$$

where the b-value of 0.5 indicates a diffusion-controlled process and the value of 1 suggests a surface-controlled or diffusion-irrelevant capacitive behavior.

Figure III. 22.c presents a plot of log (i) versus log (v) for the redox peaks of TiO_2. It is shown that the material exhibits fast surface driven intercalation (b = 0.99) up to 50 mV/s and remains the predominant contribution (b = 0.8) up to 1 V/s, which is consistent with the extrinsic pseudocapacitive intercalation in amorphous TiO_2 [44], [45], [61], [62]. In the literature, a mixed intercalation process was reported for this system. Ye et al. suggested that the separation of b values for amorphous TiO_2 is due to deeper sites in bulk being inaccessible for Li^+ ion intercalation on higher scan rates, thus decreasing the gravimetric current response with the increase of the thicknesses [45].

To test the effect of cycling on the contribution of the LiPON layer, we performed EIS before and after the CV cycling. The EIS plots remain unchanged over 100 cycles (Figure III. 23), proving consistency with a major and stable ion conductor contribution (ion conductivity $\sigma_{LiPON} = 0.5\ \mu S/\mathrm{cm}$, characteristic frequency $fc_{LiPON} = 38\ kHz$).

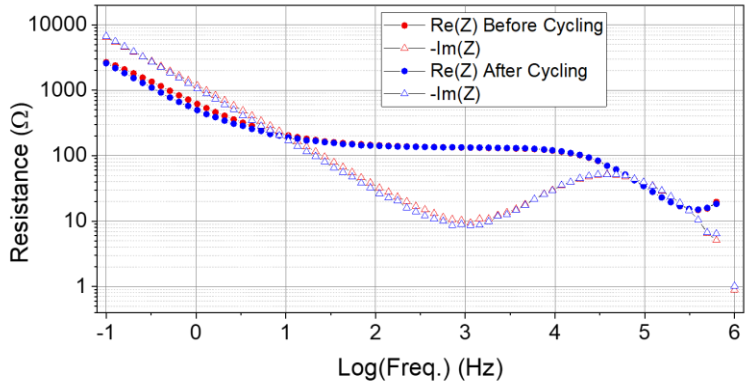

Figure III. 23: EIS spectra (at V = 1.2 V) before and after 100 CV cycles.

The specific capacity and Coulombic Efficiency (CE) per cycle are shown in Figure III. 24. CE is the ratio of total charge extracted out of active material to the total charge inserted into the active material over a cycle. Here we demonstrated a high rate capability of TiO_2 electrode with a capacity fading less than 50% for a 100 times current rate increase. As the current density was switched back to the low current rate, the capacity recovered its initial value. This recovery of the total capacity indicated that the capacity fading with incremental current rate was only related to kinetics limitation, and not material degradation or parasitic reactions. This high rate and reversibility were also confirmed by approximately 100% CE. These characteristics of TiO_2 confirmed a high-quality channel material for fast intercalation operations and high endurance.

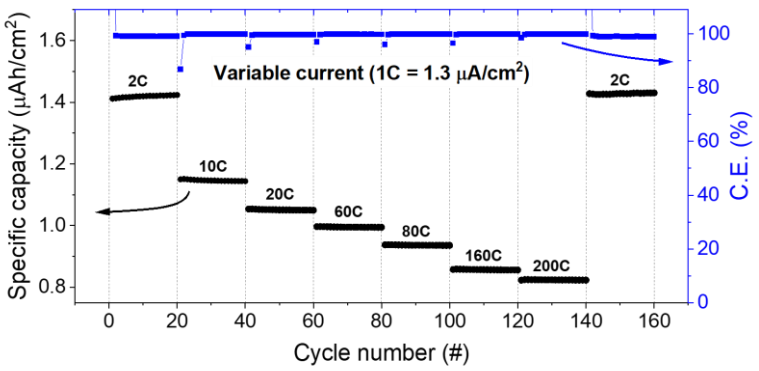

Figure III. 24: Specific capacity and Coulombic efficiency variation with current density and cycle number.

The high rate incorporation of Li into amorphous TiO_2 is a unique characteristic that allows fast conductance modification at the channel of SynT. Figure III. 24 depicts the recorded variation of SD conductance and its change of maximum conductance when experiencing increasing sweeping rates of V_G within the voltage range of V_G = [-0.5V, 1.5V]. For a full programming cycle (50 states of potentiation and 50 states of depression, in this voltage range), the average voltage difference between two adjacent states is calculated to be 40 mV. Thus, with such potential gap, the switching time for 50 mV/s sweeping rate is 0.8 s. Similarly, we have 4 ms switching operations at 1 V/s, while maintaining the "M-shaped" conductance modulation with only a 7% decrease of G_{max}.

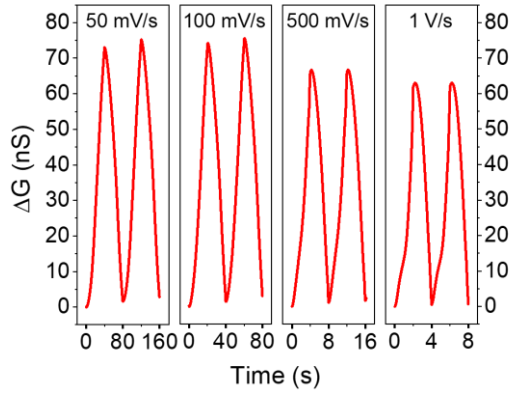

Figure III. 25: The channel conductance modulation under the increasing Gate voltage scan rates between 50 mV/s to 1 V/s.

This observation highlights the beneficial effect of TiO_2 pseudo-capacitive behavior to alleviate kinetic inhibition in all-solid state configuration. In addition to TiO_2, other intercalation materials serving as channels ($LiCoO_2$, WO_3, etc.) can be engineered to be "extrinsic pseudocapacitive" materials by thinning their film thickness to a few nanometers, thus significantly reducing the ion diffusion length and making the whole system more agile in terms of operation [63].

Overall, the electrochemical study allows us to prove the followings:

(i) TiO_2 was initially lithiated up to 0.32 Li due to the LiPON PVD process,
(ii) The initial lithiation was reversible,
(iii) The electronic conductance modulation was correlated to (de)intercalation reactions of Li^+ ions in TiO_2,
(iv) TiO_2 exhibited a pseudocapacitive behavior, with a fast and reversible (de)intercalation thus conferring to SynTs high performances in terms of response time and endurance [64].

3.5 Summary

In summary, we report a low energy consumption, all-solid-state electrochemical synaptic transistor prepared with wafer-scale microfabrication processes. The devices were assembled with an amorphous 10 nm thick TiO_2 channel and a LiPON electrolyte in a vertical configuration facilitating fast ions doping, nano-Siemens conducting level, and femtojoule writing energy. Synaptic plasticity characteristics required for an artificial synaptic component are also demonstrated. The stability and endurance of the transistors are confirmed by more than 1000 cycles and 10^5 reversible programming states in ambient conditions.

We proposed a systematic study of the vertical structures, involving several electrochemical characterizations. These investigations revealed valuable information on the electrochemical reactions which occur: (i) contribution of the channel bulk and its interfacial region, (ii) electrolyte contribution, and (iii) the reaction mechanism reversibility. Therefore, we can make a clear correlation between electrochemical reactions and the performance characteristics of our Li_xTiO_2-based three-terminal devices. The fast operation rate stemmed from the rapid Li diffusion into the pseudocapacitive amorphous TiO_2 layer, while high ionic activities around the potential 1.68 V vs. Li^+/Li suggested a highly efficient working voltage range. The EIS and rate capability tests further confirmed the TiO_2 thin film's resilience under different sweeping rates, thus making it an appropriate channel material for SynTs used for online training and high-speed, low-power neuromorphic systems.

In future work, the writing energy can be reduced by shrinking the dimensions of the devices. Miniaturization of the area for the gate stack has been reported to be practical to reduce the power spent on programming SynTs [10]. Similarly, by thinning the TiO_2 amorphous channel, the electrical conductivity is lower horizontally and faster vertically in terms of Li incorporation into ultrathin pseudocapacitive films. However, to assure the amount of mobile Li in the system, a stoichiometric $LiTiO_2$ is compulsory. Instead of relying on the passive Li diffusion after the PVD LiPON step, we will develop an ALD technique that allows depositing lithiated TiO_2 film.

4 CONCLUSIONS

In this chapter, we first presented the materials employed to build the electrochemical synaptic transistors, such as $LiCoO_2$ and Li_xTiO_2 channels, and LiPON electrolyte. The high-temperature phase $LiCoO_2$ exhibits a six-order of magnitude IMT upon Li extraction. Li_xTiO_2 film shows a reversible conductance modulation with the modification of Li content. The conductivity modulation of this layer can be accounted by the forming of EDL and the anatase-to-titanate phase transition. In addition, the solid-state LiPON electrolyte has electrochemical and temperature stability advantages over other phase of ionic conductors such as ionic liquid and polymer types, facilitating the microfabrication of wafer-scale, BeOL compatible synaptic transistors.

The preliminary results of SynTs with $LiCoO_2$/LiPON gate stack were presented as a proof-of-concept of the wafer-scale elaboration synaptic transistors. The electrochemical tests such as electrochemical impedance spectroscopy (EIS) and cyclic voltammetry (CV) on the battery-like structures help identifying the contribution of LiPON electrolyte (ionic conductivity and characteristic frequency) and $LiCoO_2$ active material (reaction potentials). The important IMT of the HT-$LiCoO_2$ channel was confirmed by measuring the channel current while extracting Li ions with a gate voltage of +4.2 V. Upon the re-intercalation of ions, the conductance decreased to the initial state at a different slope, creating the typical hysteresis of SynTs. A train of pulses allowed to modify the conductance in an analog manner for 40 times per cycle, proving the required functions for an artificial synaptic device. However, the programming voltage and the channel conductance of this gate stack were considerably high.

Finally, a comprehensive study on SynTs with the Li_xTiO_2/LiPON gate stack was conducted with the goal to improve the performance of the previous material composition. The cross-section of SynTs was prepared with FIB for SEM/EDS and TEM studies, providing useful information on the physical dimensions, presenting elements, and the quasi-amorphous channel material phase. With electrical tests, the SynT showed good merits of an electrochemical artificial synapse, such as fast programming, reversible conductance modulation, retention, linearity, endurance, and small device-to-device variation. This transistor was highly efficient in terms of energy consumption for both write ($fJ/\mu m^2$) and read (nS) operations. A systematic study using a two-terminal device representing the gate stack of SynT was done to decorrelate the electrochemical properties of the Li_xTiO_2 channel and its electrical performance. The test results highlighted the pseudocapacitive behavior of the ultra-thin Li_xTiO_2 film; making it an appropriate channel material for SynTs used for high-speed, low-power neuromorphic systems.

In Table 1, we present the materials and switching properties of the electrochemical synaptic transistor in this work and other reported SynTs. The purpose of this Table is to benchmark synaptic devices following wafer scale integration, and using microfabrication techniques and materials that are compatible with CMOS BEOL integration. A graphical

comparison among the technologies is presented in Figure III. 26. From this figure, we observe that our SynT has certain improvement in terms of performance, especially the energy consumption (write and read) aspects. However, the ON/OFF ratio of this device definitely require some improvement in the next generation to cover a wider dynamic range for different neural network applications.

Table 1: Summary on materials and switching properties of selected works on CMOS compatible, all-solid-state SynTs

Channel	Electrolyte	G_{SD} (nS)	Write duration (s)	G_{max}/G_{min}	# states	Prog. Energy (fJ/μm^2)	Endurance (Writes)	Reference
$LiCoO_2$	LiPON	290000	2	1.56	200	-	>8x10^3	[9]
WO_3	LiPON	24	100x10^{-9}	40	100	100	>10^5	[10]
$LiCoO_2$	Li_3POSe	40	1	19	80	-	>720	[11]
$WO_{2.7}$	Li_3PO_4	3500	1	6.4	60	1.4x10^6	>420	[12]
aNb_2O_5	$LiSiO_2$	100	0.01	10	100	20	>10^3	[13]
WO_x/Al_2O_3	Li_3PO_4	50	1	2.24	80	-	-	[15]
Li_xTiO_2	LiPON	75	0.1	2.6	100	1.6	>10^5	This work [65]

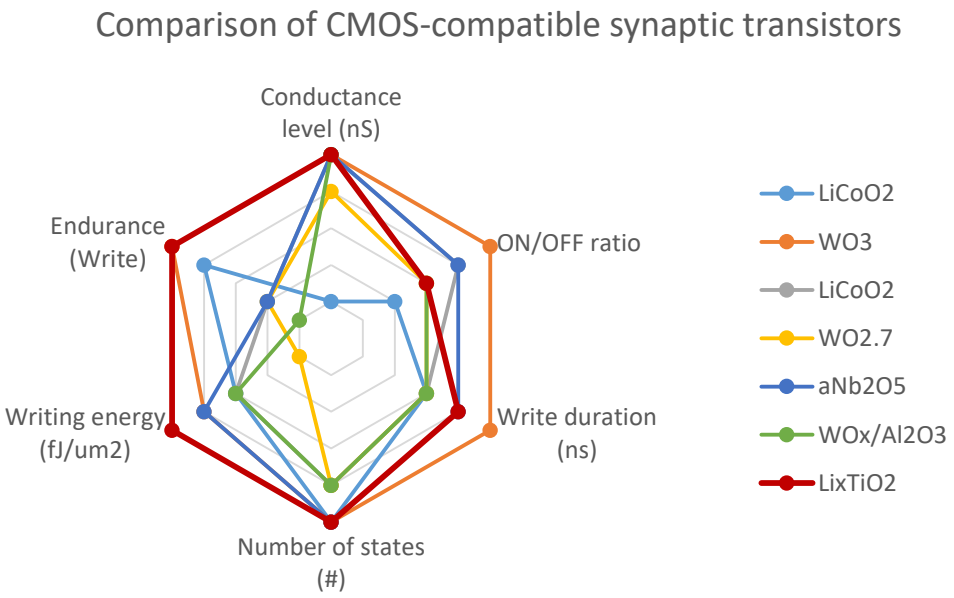

Figure III. 26: Graphical comparison of the reported CMOS-compatible synaptic transistors.

5 REFERENCES

[1] C. S. Yang *et al.*, "A Synaptic Transistor based on Quasi-2D Molybdenum Oxide," *Adv. Mater.*, vol. 29, no. 27, p. 1700906, Jul. 2017, doi: 10.1002/adma.201700906.

[2] J. Zhu *et al.*, "Ion Gated Synaptic Transistors Based on 2D van der Waals Crystals with Tunable Diffusive Dynamics," *Adv. Mater.*, vol. 30, no. 21, p. 1800195, May 2018, doi: 10.1002/adma.201800195.

[3] M. T. Sharbati, Y. Du, J. Torres, N. D. Ardolino, M. Yun, and F. Xiong, "Low-Power, Electrochemically Tunable Graphene Synapses for Neuromorphic Computing," *Adv. Mater.*, vol. 30, no. 36, p. 1802353, Sep. 2018, doi: 10.1002/adma.201802353.

[4] Y. Li *et al.*, "Low-Voltage, CMOS-Free Synaptic Memory Based on Li_xTiO_2 Redox Transistors," *ACS Appl. Mater. Interfaces*, vol. 11, no. 42, pp. 38982–38992, Oct. 2019, doi: 10.1021/acsami.9b14338.

[5] B. Yao *et al.*, "Non-Volatile Electrolyte-Gated Transistors Based on Graphdiyne/MoS_2 with Robust Stability for Low-Power Neuromorphic Computing and Logic-In-Memory," *Adv. Funct. Mater.*, vol. 31, no. 25, p. 2100069, Jun. 2021, doi: 10.1002/adfm.202100069.

[6] R. Tanaka, M. Sakurai, H. Sekiguchi, and M. Inoue, "Improvement of room-temperature conductivity and thermal stability of PEO–LiClO4 systems by addition of a small proportion of polyethylenimine," *Electrochimica Acta*, vol. 48, no. 14–16, pp. 2311–2316, Jun. 2003, doi: 10.1016/S0013-4686(03)00220-2.

[7] V. Raghunathan, T. Izuhara, J. Michel, and L. Kimerling, "Stability of polymer-dielectric bi-layers for athermal silicon photonics," *Opt. Express*, vol. 20, no. 14, p. 16059, Jul. 2012, doi: 10.1364/OE.20.016059.

[8] C. He *et al.*, "Artificial Synapse Based on van der Waals Heterostructures with Tunable Synaptic Functions for Neuromorphic Computing," *ACS Appl. Mater. Interfaces*, vol. 12, no. 10, pp. 11945–11954, Mar. 2020, doi: 10.1021/acsami.9b21747.

[9] E. J. Fuller *et al.*, "Li-Ion Synaptic Transistor for Low Power Analog Computing," *Adv. Mater.*, vol. 29, no. 4, p. 1604310, Jan. 2017, doi: 10.1002/adma.201604310.

[10] J. Tang *et al.*, "ECRAM as Scalable Synaptic Cell for High-Speed, Low-Power Neuromorphic Computing," in *2018 IEEE International Electron Devices Meeting (IEDM)*, San Francisco, CA, Dec. 2018, p. 13.1.1-13.1.4. doi: 10.1109/IEDM.2018.8614551.

[11] R. D. Nikam, M. Kwak, J. Lee, K. G. Rajput, W. Banerjee, and H. Hwang, "Near ideal synaptic functionalities in Li ion synaptic transistor using Li3POxSex electrolyte with high ionic conductivity," *Sci Rep*, vol. 9, no. 1, p. 18883, Dec. 2019, doi: 10.1038/s41598-019-55310-8.

[12] J. Lee, R. D. Nikam, S. Lim, M. Kwak, and H. Hwang, "Excellent synaptic behavior of lithium-based nano-ionic transistor based on optimal $WO_{2.7}$ stoichiometry with high ion diffusivity," *Nanotechnology*, vol. 31, no. 23, p. 235203, Mar. 2020, doi: 10.1088/1361-6528/ab793d.

[13] Y. Li *et al.*, "One Transistor One Electrolyte-Gated Transistor Based Spiking Neural Network for Power-Efficient Neuromorphic Computing System," *Adv. Funct. Mater.*, vol. 31, no. 26, p. 2100042, Jun. 2021, doi: 10.1002/adfm.202100042.

[14] J. Yu *et al.*, "All-Solid-State Ion Synaptic Transistor for Wafer-Scale Integration with Electrolyte of a Nanoscale Thickness," *Adv. Funct. Mater.*, vol. 31, no. 23, p. 2010971, Jun. 2021, doi: 10.1002/adfm.202010971.

[15] K. Lee *et al.*, "Improved synaptic functionalities of Li-based nano-ionic synaptic transistor with ultralow conductance enabled by Al_2O_3 barrier layer," *Nanotechnology*, vol. 32, no. 27, p. 275201, Jul. 2021, doi: 10.1088/1361-6528/abf071.
[16] K. Ariyoshi, K. Yuzawa, and Y. Yamada, "Reaction Mechanism and Kinetic Analysis of the Solid-State Reaction to Synthesize Single-Phase $Li_2Co_2O_4$ Spinel," *J. Phys. Chem. C*, vol. 124, no. 15, pp. 8170–8177, Apr. 2020, doi: 10.1021/acs.jpcc.0c01115.
[17] S. Kim *et al.*, "First-Principles Study of Lithium Cobalt Spinel Oxides: Correlating Structure and Electrochemistry," *ACS Appl. Mater. Interfaces*, vol. 10, no. 16, pp. 13479–13490, Apr. 2018, doi: 10.1021/acsami.8b00394.
[18] H. Xia, Y. Wan, W. Assenmacher, W. Mader, G. Yuan, and L. Lu, "Facile synthesis of chain-like LiCoO2 nanowire arrays as three-dimensional cathode for microbatteries," *NPG Asia Mater*, vol. 6, no. 9, pp. e126–e126, Sep. 2014, doi: 10.1038/am.2014.72.
[19] A. Milewska *et al.*, "The nature of the nonmetal–metal transition in LixCoO2 oxide," *Solid State Ionics*, vol. 263, pp. 110–118, Oct. 2014, doi: 10.1016/j.ssi.2014.05.011.
[20] M. Ménétrier, I. Saadoune, S. Levasseur, and C. Delmas, "The insulator-metal transition upon lithium deintercalation from LiCoO2: electronic properties and 7Li NMR study," *J. Mater. Chem.*, vol. 9, no. 5, pp. 1135–1140, 1999, doi: 10.1039/a900016j.
[21] C. A. Marianetti, G. Kotliar, and G. Ceder, "A first-order Mott transition in LixCoO2," *Nature Mater*, vol. 3, no. 9, pp. 627–631, Sep. 2004, doi: 10.1038/nmat1178.
[22] J. Molenda, "Modification in the electronic structure of cobalt bronze LixCoO2 and the resulting electrochemical properties," *Solid State Ionics*, vol. 36, no. 1–2, pp. 53–58, Oct. 1989, doi: 10.1016/0167-2738(89)90058-1.
[23] H. Yu *et al.*, "Nonvolatile multilevel switching in artificial synaptic transistors based on epitaxial LiCo O 2 thin films," *Phys. Rev. Materials*, vol. 5, no. 11, p. 115401, Nov. 2021, doi: 10.1103/PhysRevMaterials.5.115401.
[24] van Huy Mai. Étude de phénomènes de commutation de résistance de films minces de LixCoO2. Autre [cond-mat.other]. Université Paris Sud - Paris XI, 2014. Français. NNT : 2014PA112115. tel-01164971
[25] Van-Son Nguyen. LixCoO2-based thin films and devices for potential application to nonvolatile resistive memories. Micro and nanotechnologies/Microelectronics. Université Paris Saclay (COmUE), 2017. English. NNT : 2017SACLS344. tel-01900130
[26] H. Siddiqui, "Modification of Physical and Chemical Properties of Titanium Dioxide (TiO_2) by Ion Implantation for Dye Sensitized Solar Cells," in *Ion Beam Techniques and Applications*, I. Ahmad and T. Zhao, Eds. IntechOpen, 2020. doi: 10.5772/intechopen.83566.
[27] A. Stashans, S. Lunell, R. Bergström, A. Hagfeldt, and S.-E. Lindquist, "Theoretical study of lithium intercalation in rutile and anatase," *Phys. Rev. B*, vol. 53, no. 1, pp. 159–170, Jan. 1996, doi: 10.1103/PhysRevB.53.159.
[28] J. Wang, J. Polleux, J. Lim, and B. Dunn, "Pseudocapacitive Contributions to Electrochemical Energy Storage in TiO_2 (Anatase) Nanoparticles," *J. Phys. Chem. C*, vol. 111, no. 40, pp. 14925–14931, Oct. 2007, doi: 10.1021/jp074464w.
[29] S. S. El-Deen *et al.*, "Anatase TiO2 nanoparticles for lithium-ion batteries," *Ionics*, vol. 24, no. 10, pp. 2925–2934, Oct. 2018, doi: 10.1007/s11581-017-2425-y.
[30] X. Liu, P. Carvalho, M. N. Getz, T. Norby, and A. Chatzitakis, "Black Anatase TiO_2 Nanotubes with Tunable Orientation for High Performance Supercapacitors," *J. Phys. Chem. C*, vol. 123, no. 36, pp. 21931–21940, Sep. 2019, doi: 10.1021/acs.jpcc.9b05070.

[31] A. Zaffora, F. Di Franco, R. Macaluso, and M. Santamaria, "TiO2 in memristors and resistive random access memory devices," in *Titanium Dioxide (Tio_2) and Its Applications*, Elsevier, 2021, pp. 507–526. doi: 10.1016/B978-0-12-819960-2.00020-1.
[32] R. van de Krol, A. Goossens, and E. A. Meulenkamp, "Electrical and optical properties of TiO2 in accumulation and of lithium titanate Li0.5TiO2," *Journal of Applied Physics*, vol. 90, no. 5, pp. 2235–2242, Sep. 2001, doi: 10.1063/1.1388165.
[33] W. J. H. Borghols, D. Lützenkirchen-Hecht, U. Haake, E. R. H. van Eck, F. M. Mulder, and M. Wagemaker, "The electronic structure and ionic diffusion of nanoscale LiTiO2 anatase," *Phys. Chem. Chem. Phys.*, vol. 11, no. 27, p. 5742, 2009, doi: 10.1039/b823142g.
[34] C. Y. Ouyang, Z. Y. Zhong, and M. S. Lei, "Ab initio studies of structural and electronic properties of Li4Ti5O12 spinel," *Electrochemistry Communications*, vol. 9, no. 5, pp. 1107–1112, May 2007, doi: 10.1016/j.elecom.2007.01.013.
[35] S. Amaya-Roncancio *et al.*, "Ab initio calculations of lithium titanates related to anodes of lithium-ion batteries," *Journal of Physics and Chemistry of Solids*, vol. 141, p. 109405, Jun. 2020, doi: 10.1016/j.jpcs.2020.109405.
[36] S. Oukassi, L. Baggetto, C. Dubarry, L. Le Van-Jodin, S. Poncet, and R. Salot, "Transparent Thin Film Solid-State Lithium Ion Batteries," *ACS Appl. Mater. Interfaces*, vol. 11, no. 1, pp. 683–690, Jan. 2019, doi: 10.1021/acsami.8b16364.
[37] V. Sallaz, S. Oukassi, F. Voiron, R. Salot, and D. Berardan, "Assessing the potential of LiPON-based electrical double layer microsupercapacitors for on-chip power storage," *Journal of Power Sources*, vol. 451, p. 227786, Mar. 2020, doi: 10.1016/j.jpowsour.2020.227786.
[38] Y. Wu, S. Wang, H. Li, L. Chen, and F. Wu, "Progress in thermal stability of all-solid-state-Li-ion-batteries," *InfoMat*, vol. 3, no. 8, pp. 827–853, Aug. 2021, doi: 10.1002/inf2.12224.
[39] H. Xia, L. Lu, Y. S. Meng, and G. Ceder, "Phase Transitions and High-Voltage Electrochemical Behavior of LiCoO2 Thin Films Grown by Pulsed Laser Deposition," *Journal of The Electrochemical Society*, p. 6.
[40] C. Yang *et al.*, "All-Solid-State Synaptic Transistor with Ultralow Conductance for Neuromorphic Computing," *Adv. Funct. Mater.*, vol. 28, no. 42, p. 1804170, Oct. 2018, doi: 10.1002/adfm.201804170.
[41] G.-N. Zhu, Y.-G. Wang, and Y.-Y. Xia, "Ti-based compounds as anode materials for Li-ion batteries," *Energy Environ. Sci.*, vol. 5, no. 5, p. 6652, 2012, doi: 10.1039/c2ee03410g.
[42] W. J. H. Borghols *et al.*, "Lithium Storage in Amorphous TiO2 Nanoparticles," *Journal of The Electrochemical Society*, p. 8.
[43] J. Xu, C. Jia, B. Cao, and W. F. Zhang, "Electrochemical properties of anatase TiO2 nanotubes as an anode material for lithium-ion batteries," *Electrochimica Acta*, vol. 52, no. 28, pp. 8044–8047, Nov. 2007, doi: 10.1016/j.electacta.2007.06.077.
[44] S. Moitzheim, S. De Gendt, and P. M. Vereecken, "Investigation of the Li-Ion Insertion Mechanism for Amorphous and Anatase TiO_2 Thin-Films," *J. Electrochem. Soc.*, vol. 166, no. 2, pp. A1–A9, 2019, doi: 10.1149/2.1091816jes.
[45] J. Ye *et al.*, "Amorphization as a Pathway to Fast Charging Kinetics in Atomic Layer Deposition-Derived Titania Films for Lithium Ion Batteries," *Chem. Mater.*, vol. 30, no. 24, pp. 8871–8882, Dec. 2018, doi: 10.1021/acs.chemmater.8b04002.
[46] S. Oukassi *et al.*, "Millimeter scale thin film batteries for integrated high energy density storage," in *2019 IEEE International Electron Devices Meeting (IEDM)*, San Francisco, CA, USA, Dec. 2019, p. 26.1.1-26.1.4. doi: 10.1109/IEDM19573.2019.8993483.

[47] S. Oukassi, C. Giroud-Garampon, C. Dubarry, C. Ducros, and R. Salot, "All inorganic thin film electrochromic device using LiPON as the ion conductor," *Solar Energy Materials and Solar Cells*, vol. 145, pp. 2–7, Feb. 2016, doi: 10.1016/j.solmat.2015.06.052.
[48] L. Le Van-Jodin, F. Ducroquet, F. Sabary, and I. Chevalier, "Dielectric properties, conductivity and Li+ ion motion in LiPON thin films," *Solid State Ionics*, vol. 253, pp. 151–156, Dec. 2013, doi: 10.1016/j.ssi.2013.09.031.
[49] Y. Zhu, X. He, and Y. Mo, "Origin of Outstanding Stability in the Lithium Solid Electrolyte Materials: Insights from Thermodynamic Analyses Based on First-Principles Calculations," *ACS Appl. Mater. Interfaces*, vol. 7, no. 42, pp. 23685–23693, Oct. 2015, doi: 10.1021/acsami.5b07517.
[50] S. Nowak, F. Berkemeier, and G. Schmitz, "Ultra-thin LiPON films – Fundamental properties and application in solid state thin film model batteries," *Journal of Power Sources*, vol. 275, pp. 144–150, Feb. 2015, doi: 10.1016/j.jpowsour.2014.10.202.
[51] I. Valov *et al.*, "Nanobatteries in redox-based resistive switches require extension of memristor theory," *Nat Commun*, vol. 4, no. 1, p. 1771, Jun. 2013, doi: 10.1038/ncomms2784.
[52] S. Ambrogio *et al.*, "Equivalent-accuracy accelerated neural-network training using analogue memory," *Nature*, vol. 558, no. 7708, pp. 60–67, Jun. 2018, doi: 10.1038/s41586-018-0180-5.
[53] E. J. Fuller *et al.*, "Parallel programming of an ionic floating-gate memory array for scalable neuromorphic computing," p. 6, 2019, doi: 10.1126/science.aaw5581.
[54] R. D. Nikam, M. Kwak, J. Lee, K. G. Rajput, and H. Hwang, "Controlled Ionic Tunneling in Lithium Nanoionic Synaptic Transistor through Atomically Thin Graphene Layer for Neuromorphic Computing," *Adv. Electron. Mater.*, vol. 6, no. 2, p. 1901100, Feb. 2020, doi: 10.1002/aelm.201901100.
[55] S. Yu, "Neuro-Inspired Computing With Emerging Nonvolatile Memorys," *Proc. IEEE*, vol. 106, no. 2, pp. 260–285, Feb. 2018, doi: 10.1109/JPROC.2018.2790840.
[56] Q. Wan, M. T. Sharbati, J. R. Erickson, Y. Du, and F. Xiong, "Emerging Artificial Synaptic Devices for Neuromorphic Computing," *Adv. Mater. Technol.*, vol. 4, no. 4, p. 1900037, Apr. 2019, doi: 10.1002/admt.201900037.
[57] J. Rahn, E. Hüger, L. Dörrer, B. Ruprecht, P. Heitjans, and H. Schmidt, "Self-Diffusion of Lithium in Amorphous Lithium Niobate Layers," *Zeitschrift für Physikalische Chemie*, vol. 226, no. 5–6, pp. 439–448, Jun. 2012, doi: 10.1524/zpch.2012.0214.
[58] E. Hüger and H. Schmidt, "Lithium Permeability Increase in Nanosized Amorphous Silicon Layers," *J. Phys. Chem. C*, vol. 122, no. 50, pp. 28528–28536, Dec. 2018, doi: 10.1021/acs.jpcc.8b09719.
[59] H. Xiong *et al.*, "Self-Improving Anode for Lithium-Ion Batteries Based on Amorphous to Cubic Phase Transition in TiO_2 Nanotubes," *J. Phys. Chem. C*, vol. 116, no. 4, pp. 3181–3187, Feb. 2012, doi: 10.1021/jp210793u.
[60] S. Ardizzone, G. Fregonara, and S. Trasatti, "'Inner' and 'outer' active surface of RuO2 electrodes," *Electrochimica Acta*, vol. 35, no. 1, pp. 263–267, Jan. 1990, doi: 10.1016/0013-4686(90)85068-X.
[61] Y.-M. Lin *et al.*, "Morphology Dependence of the Lithium Storage Capability and Rate Performance of Amorphous TiO_2 Electrodes," *J. Phys. Chem. C*, vol. 115, no. 5, pp. 2585–2591, Feb. 2011, doi: 10.1021/jp110474y.

[62] X. Sun *et al.*, "Pseudocapacitance of Amorphous TiO_2 Thin Films Anchored to Graphene and Carbon Nanotubes Using Atomic Layer Deposition," *J. Phys. Chem. C*, vol. 117, no. 44, pp. 22497–22508, Nov. 2013, doi: 10.1021/jp4066955.
[63] C. Choi *et al.*, "Achieving high energy density and high power density with pseudocapacitive materials," *Nat Rev Mater*, vol. 5, no. 1, pp. 5–19, Jan. 2020, doi: 10.1038/s41578-019-0142-z.
[64] M. Mastragostino and F. Soavi, "Pseudocapacitive and Ion-Insertion Materials: A Bridge between Energy Storage, Electronics and Neuromorphic Computing," *ChemElectroChem*, vol. 8, no. 14, pp. 2630–2633, Jul. 2021, doi: 10.1002/celc.202100457.
[65] N. Nguyen *et al.*, "An Ultralow Power $Li_x TiO_2$ -Based Synaptic Transistor for Scalable Neuromorphic Computing," *Adv Elect Materials*, p. 2200607, Sep. 2022, doi: 10.1002/aelm.202200607.

CHAPTER IV

SIMULATIONS OF NEUROMORPHIC COMPUTING SYSTEMS COMPOSED OF OUR Li_xTiO_2-BASED SYNAPTIC TRANSISTORS

ABSTRACT

In chapter 4, we will introduce the neural networks that can make use of our electrochemical synaptic transistors (SynT), including artificial and spiking neural networks. These neural-inspired systems employ SynT in unique ways.

Section 1 of this chapter is dedicated to giving an overview on the concept of artificial neural network (ANN), working algorithms, and the associated hardware development to afford the ever-increasing size and complexity of these systems that handle data-intensive tasks. Among the approaches, the deep learning accelerators made from emerging nonvolatile memory technologies using the concept of in-memory computing raised significant interest. The crossbar architecture and its functions are described to highlight the need for electronic synaptic devices that can modify its conducting weight in a linear and controllable way. We used the CrossSim simulator to simulate, train, and test neural networks taking into account realistic, nonlinear features of SynTs on image pattern recognition tasks. The simulated results reveal the high performance of our SynTs as artificial synapses for deep learning accelerators.

Section 2 shows the potential applicability of SynTs toward spiking neural networks (SNNs) as artificial synapses. A neural network is designed using the simple "leaky integrate-and-fire" (LIF) neuron model, and our synaptic transistor as one of the synapses. With the experimentally verified synaptic plasticity, the SynTs will be an essential part of the training process of the circuit. The all-analog neural circuit is simulated using LT-SPICE software to demonstrate the associative memory behavior as in Pavlov's dog experiment. The simulation results reveal that the neural circuit can learn to link the stimuli from initially unconnected neurons and respond correspondingly, thus proving the applicability of our SynTs to the SNN. Conclusions on the applications of SynTs can be found at the end of the chapter.

TABLE OF CONTENTS

1 APPLICABILITY OF SYNTS TO ARTIFICIAL NEURAL NETWORKS 138
1.1 Background 138
1.2 Working principle of an Artificial Neural Network (ANN) 139
1.3 From near-memory to in-memory computing 142
1.3.1 Hardware approaches 142
1.3.2 The crossbar architecture as an analog in-memory accelerator 144
1.4 CrossSim crossbar simulator 146
1.5 Simulation of ANN with SynTs as synaptic elements 148
2 APPLICABILITY OF SYNTS TO SPIKING NEURON NETWORKS 152
2.1 Background and objective 152
2.2 Pavlovian conditioning experiment 153
2.3 Electronic versions of Pavlov's dog in literature 154
2.4 Design of an associative memory which involves our SynT 155
2.4.1 Neuron circuit model involving 1 synapse (resistor) 155
2.4.2 Neuron circuit model involving 2 synapses (resistors) 157
2.4.3 Neuron circuit model involving 2 synapses: 1 resistor and 1 SynT 158
2.5 Circuit simulation of the associative memory with SynT characteristics 161
3 CONCLUSIONS 164
4 REFERENCES 165

1 APPLICABILITY OF SYNTS TO ARTIFICIAL NEURAL NETWORKS

1.1 Background

The amount of data estimated to be generated in 2025 is 181 zettabytes, which is considered a substantial digital mine for technology development [1], [2]. However, it also poses a considerable challenge, requiring exhaustive computing resources to analyze and understand thoroughly. Artificial intelligence (AI) has been extensively used in the past decade to perform data-intensive tasks. This technology is becoming increasingly advanced and widespread in real-world applications, such as computer vision [3], natural language recognition [4], healthcare [5], and pattern classification [6]. Advances in AI technology have been achieved through the unprecedented success of deep-learning algorithms, coming along with the ever-increase sizes and complexities of neural networks and tasks [7]. Such developments outplayed the scaling trend of CMOS processors, which made the conventional digital computers based on von Neumann computing structures less appealing for jobs consisting of machine-learning operations [8]. It has been evaluated that, for many computing tasks, the majority of the energy and time is consumed in data movement process rather than computation [9]. This has necessitated the development of brain-inspired neuromorphic computing based on emerging non-volatile memory that can deliver efficient computing [10]–[12].

New and emerging non-volatile memory concepts have also been introduced into the traditional memory hierarchy to reduce the 'distance' between computing and the data [13], [14]. Instead of re-engineering conventional systems by individually improving the parallelism, memory bandwidth, or memory concept, in-memory computing aims to radically subvert the von Neumann architecture by carrying out calculations in situ at exactly the data location [15]. This approach is similar to the computing scheme in the human brain, where information is processed in sparse networks of neurons and synapses without any physical separation between computation and memory. In-memory computing offers a clear advantage by lowering the latency and energy burdens of the memory wall.

Emerging hardware constructed on the base of crossbar architecture and emerging non-volatile memory elements leads to a significant push in processing speed and energy efficiency by carrying out these vector-matrix multiplications in an analog fashion [16], [17]. In this section, we focus on the description of the structure and working mechanism of the novel hardware systems that allow the acceleration of deep learning algorithms. In this aspect, we simulate and train an analog-based ANN employing our SynTs as the artificial synapses.

1.2 Working principle of an Artificial Neural Network (ANN)

An ANN consists of neurons layers connected by "synaptic" weights. The first neuron layer is the input layer whose size is determined by the incoming data of interest: image pixels, reduced audio content, encoded words, etc. It is followed by a series of hidden layers. whose role is to perform nonlinear transformations of the input values entered into the network. If there is more than one hidden layer, an ANN is usually called DNN (Deep Neural Network).

Hidden layers vary depending on the function of the neural network, and similarly, the layers may vary depending on their associated weights. The output layer's size depends on the task that the DNN should accomplish, for example, classifying an image into a predefined set: a handwritten number or character, a type of flower, etc. There are two operation modes of DNN: Training and inference (illustrated in Figure IV. 1).

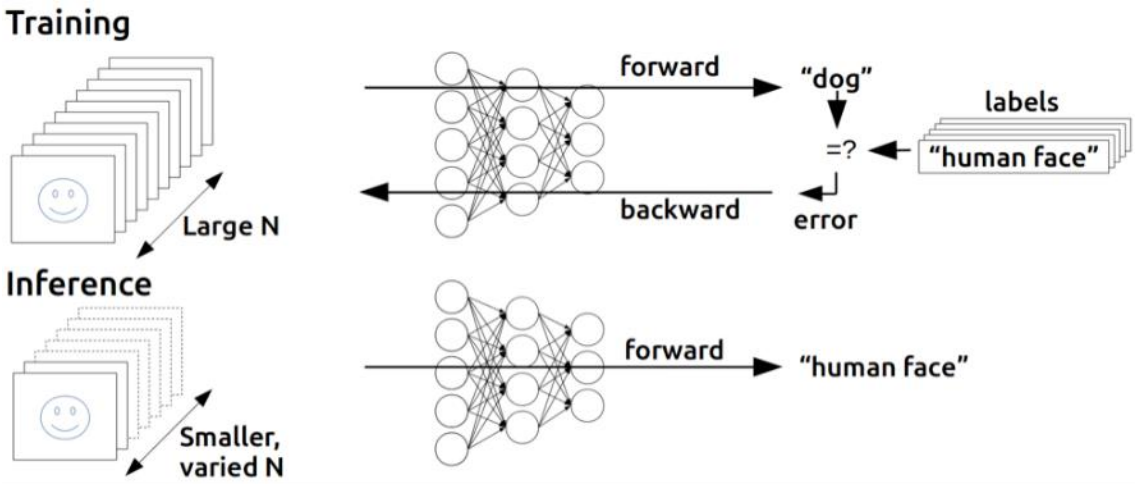

Figure IV. 1: Two operation modes of DNN: Inference and training [18].

The forward inference is the evaluation of a trained neural network with defined weights connection on one or more new vector inputs. The computational tasks involved in this process are light. One of the dominant computation tasks in this phase is vector-matrix multiplication (VMM) (Figure IV. 2). In this task, the input vector from the previous layer x_i must be multiplied by a matrix of weights w_{ij}, creating a new vector of neuron excitations for the next layer y_j. In another word, this operation can be broken down into a series of multiply and accumulate (MAC) operations, followed by a nonlinear squashing function $f(x)$.

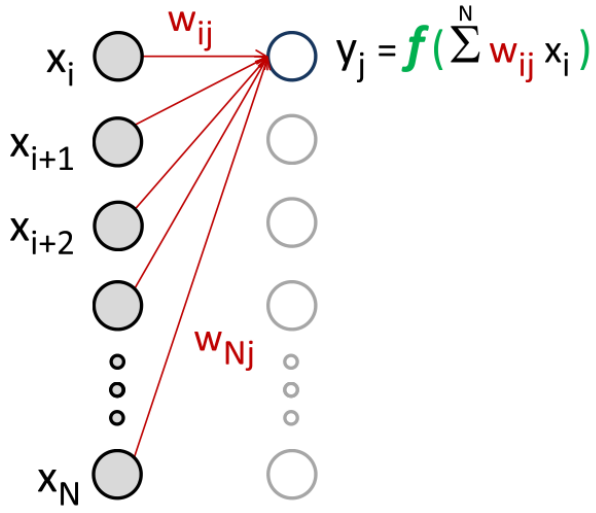

Figure IV. 2: Illustration of vector-matrix multiply operation during forward inference phase of DNN [11].

Such a nonlinear function is important from a neural network point of view because it prevents the forward-evaluate of a multi-layer network from collapsing into a single linear equation. Compared to the VMM of the previous step, the squashing function takes less computing effort. Typical nonlinear functions used for this purpose are ReLU or logistic sigmoid. ReLU is a piece-wise linear function with two segments: one along the x-axis, outputting zero for any input sum that is negative; and a second segment along the diagonal $f(x) = x$ directly passing any positive sum as the output. On the other hand, a sigmoid function is a mathematical function having a characteristic "S"-shaped curve, having the output from zero to one. These functions help to address the problems coming from saturating excitations and vanishing gradients.

In the forward path, the input data is prepared as a vector and fed into the network via the input layer, and then it propagates in series until it reaches the end of the network. At the output layer, a softmax operation is usually called. Here, each raw excitation y_j is put through an expanding nonlinearity and then normalized by the sum of all such intermediate results across the entire output layer. This operation guarantees that the produced outputs that fall between zero and one and to sum up to one as well. The output is now a probability vector, representing the guesses of the initially trained DNN on an example of data input.

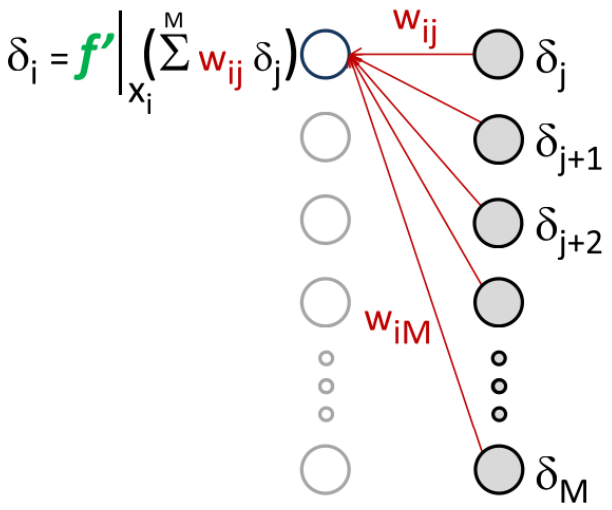

Figure IV. 3: The reverse propagation step in DNN training involves a VMM between a vector of errors (δ_j) and (the transpose of) the weight matrix (w_{ij}) [11].

An important step to improving the prediction accuracy of the neural network is training. For the training mode, the weights of the neural networks are adjusted using the supervised backpropagation algorithm. In this case, the input vector x_i is stored, and the output vector is compared with the label (expectation) to determine the error vector. The calculated vector reversely propagates through the network in a chain of VMMs from the end to the beginning (right to left), and the intermediate results are stored. Figure IV. 3 shows the reverse propagate step, as a vector of errors (δ_j) is multiplied by the transpose of the original weights w_{ij}. Instead of operating a nonlinear squashing function, the sum is multiplied by the derivative of the squashing function as evaluated at the original excitation, x_i, to compute the derivative of an 'energy function', E, for the overall DNN as a function of each individual weight, w_{ij}. Such energy functions can be minimized only when the guess of the DNN during inference matches with the data input's label. Backpropagation then allows the DNN to adjust each weight connection to improve the prediction accuracy of the network upon the next encounter. Weight update for each weight is then the product of the original upstream neuron excitation, x_i, and the downstream neuron's error, δ_j. Typically, this is scaled by a fairly small number, η, called the learning rate. The learning rate has to be chosen so that the learning is not too slow with small η, or the training cannot converge because of too big η. This process of training (inference, errors calculating, and weights updating) is iterated many times until the desired error rate is reached.

During the training phase, the complete dataset is presented to the network many times (epochs), which makes training a highly compute-intensive process, especially noting that modern datasets can have many millions of entries. In addition, while the forward propagation of the input vector and backward propagation of error processes are simple VMMs on stationary weights, the update operation manipulates the weights themselves.

From a computational complexity point of view, the update operation is much more complicated than other propagation operations as the system has to keep track of the weights, activations and errors at any time during the training process. Therefore, training is considerably heavier on memory requirements. The goal of computer architectures for machine learning with DNN is to make these operations as energy efficient as possible.

1.3 From near-memory to in-memory computing

1.3.1 Hardware approaches

To increase compute efficiency for machine learning tasks, there are different approaches that are often referred to as "compute-near-memory" [19] and "compute-in-memory" [20]. In principle, the idea is to perform the processing and the algorithms close to where memory resides, thus, minimizing the data shuttling efforts. Compute-near-memory (CNM) is currently at present the most popular approach in the field and is entirely carried out using conventional digital CMOS components and approaches (both inference and training of DNN). The common goals of this approach are bringing memory closer to the compute engine and deploying reduced precision representation of the digital content. A schematic view of CNM can be found in Figure IV. 4.

From Figure IV. 4, such a CNM structure allocates some small on-chip cache memory to each processor unit in order to mitigate the memory bottleneck. However, each processor in a GPU still needs to fetch data from the dedicated memory to the compute unit, execute the compute operation (such as MAC) and then write back to the local memory. Therefore, while the proximity of the memory units to the processors can reduce the effect of the Von Neumann bottleneck, the memory wall remains (due to the performance gap between the processor and the memory). While dynamic random-access-memories (DRAMs), with structures of 1-transistor-1-capacitor, can offer high memory density, one cannot afford them for MAC operations due to their energetical expense. On the other hand, the six-transistor static random access memory (SRAM) on-chip memories can provide high bandwidth, high reliability, and offer fast and energy-efficient read/write operations. However, the density of SRAM is significantly limited due to its large area footprint, hindering its use as an on-chip memory for MAC engines and other applications. This is a limitation of the current CMOS-based compute near-memory.

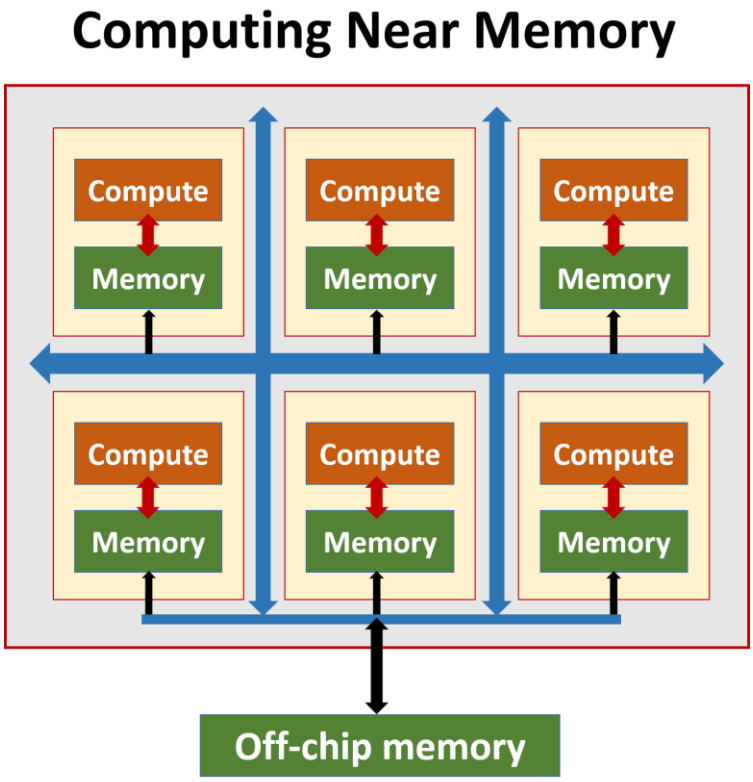

Figure IV. 4: Computing near memory structure [21].

In another approach, compute-in-memory (CIM) paradigm aims at reducing data movement completely by using a crossbar architecture and non-volatile memory components to accomplish the MAC operation in an analog fashion (Figure IV. 5). Within a CIM chip, there exists computational, on-chip compatible memory and compute units. The CIM cores contain the crossbar memory array combined with the peripheral circuitry (Analog-to-Digital converter – ADC, and control/communications digital circuitry). The tiles of CIM cores and the digital compute unit communicate (digitally) via a data bus. I/O circuitry provides the means to communicate off-chip. This approach can significantly increase parallelism and reduce data transport to and from off-chip memory. Memory elements storing the weights in one-bit or multi-bit capacity (digital memories) or analog values along a continuous scale between a high and low conductance value (emerging analog memories) are arranged at the nodes of a crossbar array of metal interconnects. We will mainly discuss the latter in this thesis. Here, the VMM operation during inference is carried out in an analog fashion using Kirchhoff's and Ohm's laws (which will be discussed in detail in the next section). This mode of operation also enables incremental changes of all the array elements in parallel at constant times, thus increasing the computing efficiency for inference and training of DNNs.

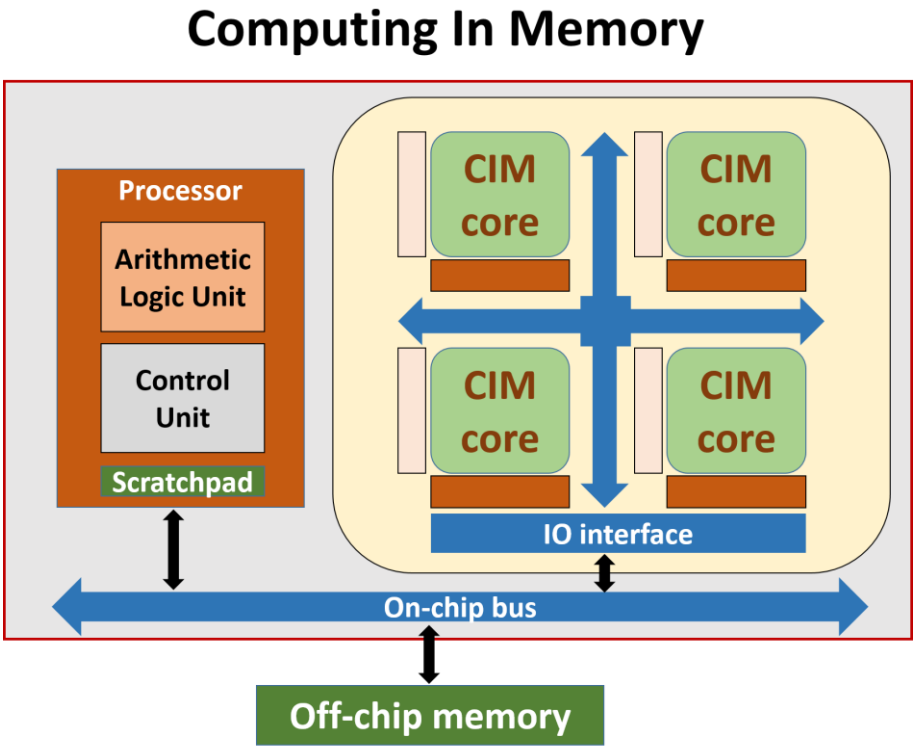

Figure IV. 5: Computing in memory structure [21].

In the next section, we will discuss how the emerging non-volatile memories can be implemented into crossbar architecture to accelerate the VMM operations.

1.3.2 The crossbar architecture as an analog in-memory accelerator

The vector-matrix multiplication operation, which dominates most DNN workloads, can now be executed in-situ with massive parallelism and order-of-magnitude energy reduction using crossbar architecture. Recall that during the inference phase and the training phase of a DNN, it is always involved in the propagation (forward/backward) of the neuron excitation vector or error vector. Figure IV. 6 illustrates such a concept in three forms: neural network, mathematical, and electrical representations. The propagation of excitation values X_i from Layer N to layer N+1 via weighted connection $W_{i,j}$ can be viewed as a multiplication of the vector $\vec{X}$ and the matrix $\boldsymbol{W}$: $\vec{X}\boldsymbol{W}$. Here, the vector X_i is transposed for the presentation purpose. Such mathematical expression can be reproduced by a crossbar array composed of analog electronics (artificial synapses) at the cross points.

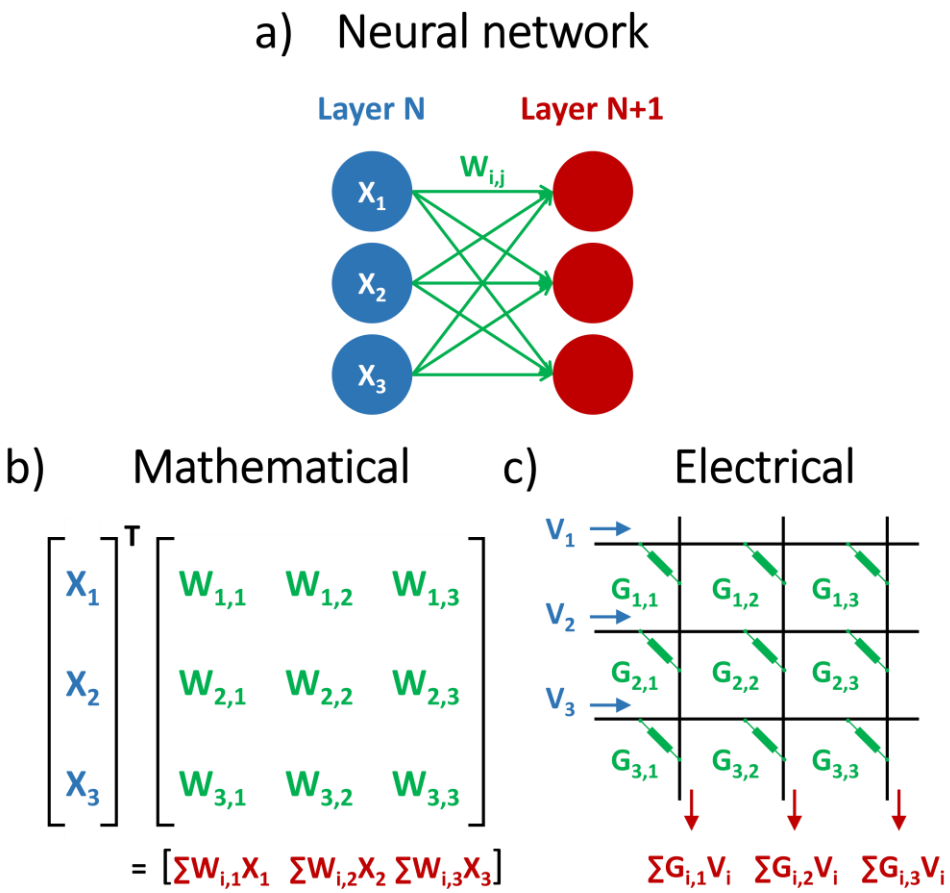

Figure IV. 6: Vector-matrix multiplication illustrated in a) neural network, b) mathematical, and c) electrical representations.

These synapses with conductance levels $G_{i,j}$ are set proportional to the value of the matrix $\boldsymbol{W}$, and a vector of input voltages $\vec{V}$, which is proportional to $\vec{X}$, is applied to the rows as read voltages. An analog multiplication is computed at each cell following the Ohm's law in which the current flowing out through the synapse is the product of its conductance $G_{i,j}$ and the applied voltage V_i. Along the columns, the accumulated current I_j will have the form of the dot product dictated by Kirchhoff's law. The analog dot products are then quantized using an analog-to-digital converter. The converted dot product can now be processed (by squashing with a nonlinear function or being multiplied with the derivative of the squashing function evaluated at the squashing function) and transmitted digitally to the next layer of the neural network.

During the training of DNN, another important operation accelerated using a crossbar array is parallel writing for the weight update. This operation is realized based on the rank one outer product update of the write duration and write amplitude (Figure IV. 7). Weight W_{ij} is updated by $x_i \times y_i$. In order to achieve a multiplicative effect, the x_i are encoded in time while the y_i are encoded in the height of a voltage pulse. The resistive

memory will only train when x_i is nonzero. The height of y_i determines the strength of training when x_i is nonzero.

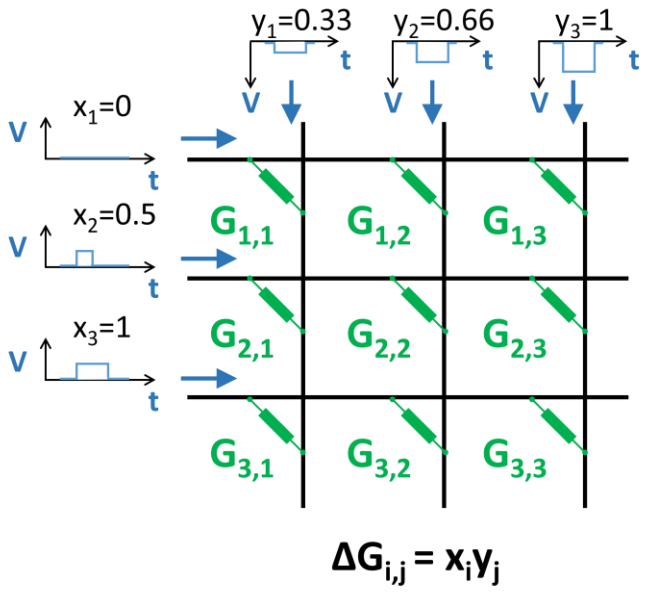

Figure IV. 7: Illustration of parallel weight update during training.

Such an analog approach for VMM has two fundamental advantages for energy efficiency: (i) A multi-bit multiplication is conducted with a single device, and summations of the current are simplified hardware-wise (just wire crossings), and (ii) the weight matrix does not need to be read out; only the inputs and outputs need to be communicated between processing cores. This dramatically reduces data movement energy. There exists several open-source simulation simulator package developed to simulate and benchmark analog-based deep learning (DL) accelerators taking into account the experimental characteristics of the artificial synapses in the literature, namely NeuroSim [22], AIHWKIT [23], and CrossSim [16]. In my thesis, CrossSim platform has been mainly used and it will be briefly described in the next part before introducing the simulation results of analog crossbar arrays made from SynTs.

1.4 CrossSim crossbar simulator

CrossSim, developed by Sandia National Laboratories, is a platform that provides a clean python application programming interface (API) allowing for the application of different algorithms built upon resistive memory crossbars while modeling realistic characteristics of devices [24]–[26].

Within the CIM model, a crossbar-based neural core can be used to perform the parallel vector matrix multiply and outer product update, while a more general purpose

digital core can be used to process the inputs and outputs of the crossbar (Figure IV. 8.a). The flexibility of the digital cores allow many different algorithms to be implemented, while still taking advantage of the neural cores to accelerate VMM operations. The digital cores can also use digital on-chip resistive memory instruction caches to store slowly changing data while reserving expensive SRAM caches only for the data being processed. The neural core consisting of a resistive memory crossbar and peripheral circuits (Digital-Analog converter, Analog-Digital converter, and op-amps) is illustrated in Figure IV. 8.b. The inputs are processed in digital domain and fed into the crossbar using DACs. Here, a bias row and column are added to the crossbar to allow for negative weights to mimic the inhibitory influence of neurons [27]. The rows and columns are driven by either variable length or variable height of potential pulses. The output currents are integrated and then converted to digital using an ADC.

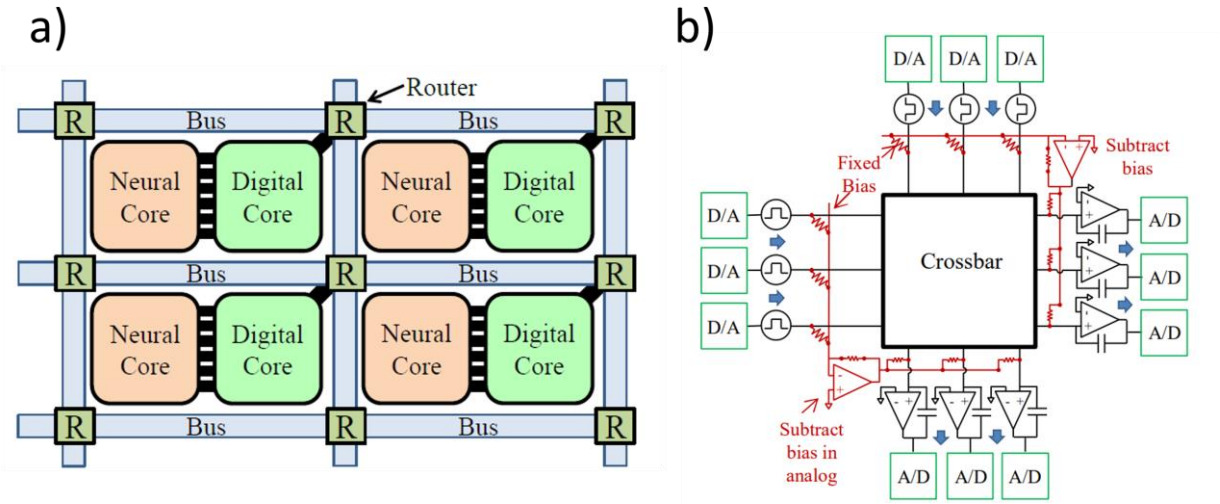

Figure IV. 8: a) Illustration of a neural architecture used for CrossSim simulation comprising of Digital cores dealing with inputs and outputs processing and Neuron cores for VMM acceleration. b) Illustration of a neuron core [24].

During inference, input vectors are converted to variable length pulses proportional to the magnitude of each element of the input vector. The resulting current is integrated at the output op-amps. This allows analog input vectors to be encoded and multiplied by the analog resistance values. The inputs along the rows come from a previous layer of sigmoid neurons which have outputs in a 0 to 1 range. The column current outputs feed into a sigmoid slope of 1. During backpropagation, the crossbar is updated by the outer product of two vectors. The columns are driven by output of the previous layers' neurons. This means the output will be in a 0 to 1 range. The rows will be driven by the product of the learning rate, η, the derivative of a sigmoid, and a backpropagated error. The simulator was designed considering this neuron model, but the generalized API can also be used with specialized hardware neurons.

1.5 Simulation of ANN with SynTs as synaptic elements

In the following, our objective is to simulate, using the CrossSim platform, an ANN built with realistic SynTs, then test such ANN for pattern recognition tasks, and benchmark its performance among several available technologies.

To do so, a three-layer neural network with one hidden layer is used, as shown in Figure IV. 9. In the neural core, the synaptic memory cell in the crossbar architecture comprises a SynT and a two-terminal access device. In fact, access devices (or selectors) are connected to the gate of the synaptic devices to increase the high OFF impedance of the gate stack after writing operations, thus, isolating the smallest potential perturbations from the gate to the channel and guaranteeing the long-term retention and programming accuracy required for DNN training. In addition, these selectors allow precisely parallel updating weights in large-scale systems [28]–[31]. Such a simulation of the access devices is to facilitate the benchmark afterward among other technologies using the same CrossSim platform as well.

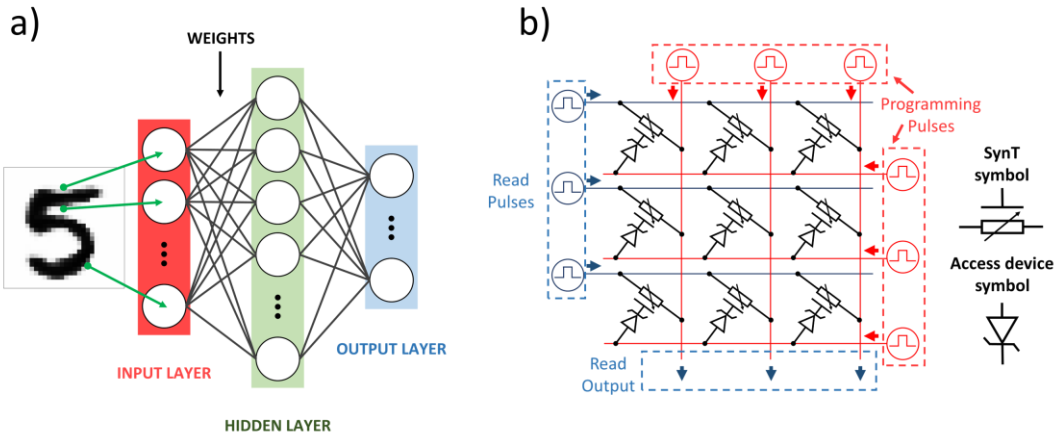

Figure IV. 9: a) Simulated three-layered ANN with one hidden layer. b) Crossbar array representation where SynTs serve as synaptic elements.

Here, each layer of the network was mapped onto a simulated crossbar, which carry out two main operations: vector matrix multiplication and parallel outer product weight update. The parallel weight updates are realized by varying both the length of the update on the rows and the voltage amplitude of the update on the columns with the assumption that the weight (conductance) change is linearly dependent on the amplitude and the duration of writing pulses. Concerning system and algorithm parameters, the neuron's nonlinear function is the Sigmoid function with a slope of 1 and a learning rate of $\eta = 0.1$. The conductance range of the cross-point memories and their initial weights are initialized by default following the scheme reported in [25].

To simulate a crossbar system with realistic synaptic devices, we simulate the nonlinearity and the write noise associated with the SynT operation (Figure IV. 10). We first start with a 100 cycles of conductance modulation from our SynT. The modulation profile exhibits the range from 25.8 to 73 nS with a clear saw-tooth shape for all the cycles considered. The step conductance ΔG from an initial conductance G_0 due to a write operation at each level is calculated by $\Delta G(G_0) = G - G_0$ (see insets). We then condense the cycling results into ΔG versus G_0 plots to analyze the write noise of the device.

The illustrations of the conductance step as a function of initial conductance show the characteristic nonlinearity of an electrochemical device. The step conductance ΔG is proportional to the amount of ions injected/extracted per pulse, thus depending on the operation gate-source potential (recall the I_GV_G graph of a SynT – Figure III. 20). During potentiation, conductance levels at around $50 - 60$ nS corresponding to the insertion potential of Li^+ ions into the channel, explaining the high values of ΔG compare to the other regions of the scan. The same trend is observed for the depression graph at around 60 nS.

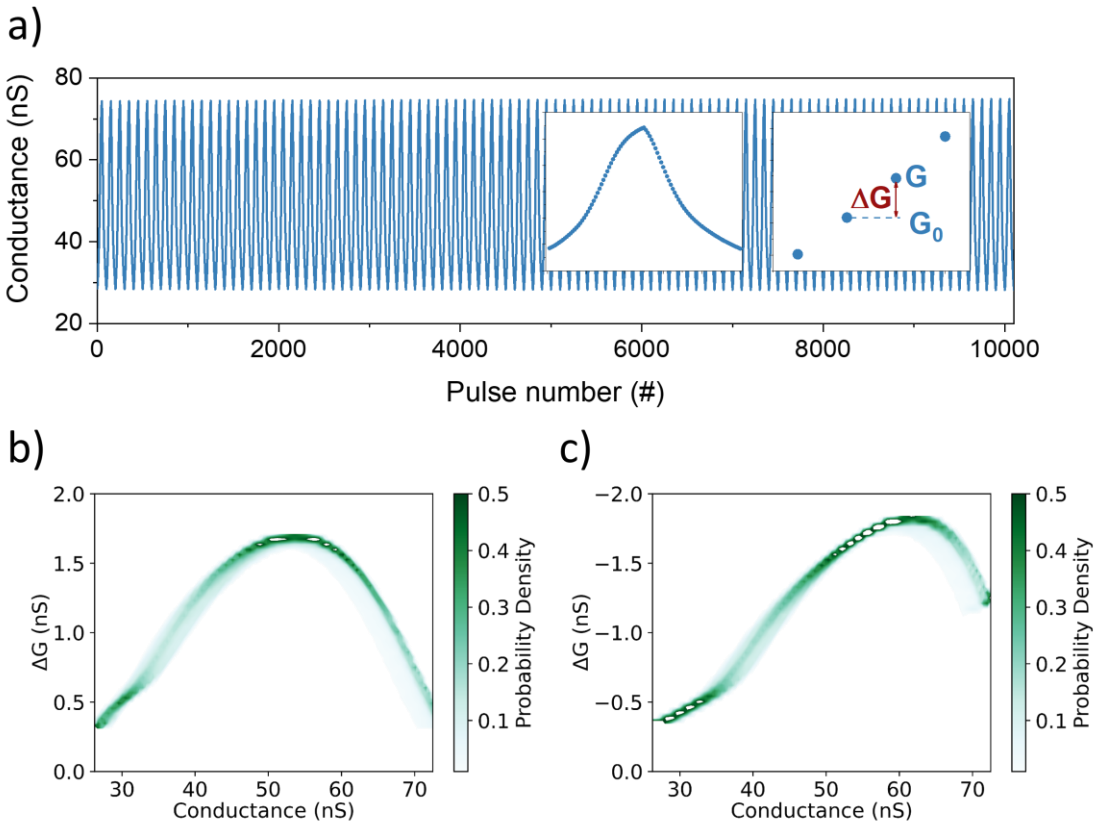

Figure IV. 10: a) 100 conductance modulation cycles are used as input to simulate the artificial synapses. The insets show the first cycle and a conductance step, respectively. b) and c): The probability function calculated for potentiation (b) and depression (c) operations.

The probability distribution function of the variable ΔG over conductance G induced during potentiation or depression programming operations are also plotted (Figure IV. 10.b, c). From these statistical plots, this SynT device exhibits an average standard deviation of $\sigma = 0.46 \pm 0.002$ nS and 0.52 ± 0.002 nS for potentiation and depression, respectively. By sampling from the probability distribution during weight updates to the crossbar, the device noise, nonlinearity, and asymmetry are taken into account. In order to find the update, first we determine the initial state of the device G_0, then we find the average weight update at G_0 and sample a noise value from the probability distribution at that level in Figure IV. 10.b, c. For instance, if $G_0 = 60$ nS, the mean update is 1.6 nS. Sampling from the probability distribution at that G_0 position, we might obtain an update of 2 nS, giving a noise of 0.46 nS added to the mean update. Note that the write accuracy of the device is calculated to be $(\Delta G/\sigma)^2 > 12$, which is higher than that observed on ReRAMs [11] and PCMs [32] ($(\Delta G/\sigma)^2 < 1$), which limits their application on ANN.

With the experimentally derived updates, we can now simulate how the ANN with SynTs as synaptic elements performs on different pattern recognition tasks using supervised backpropagation algorithm. The datasets and the associated networks' size can be found in Table 1.

Table 1: The information of the datasets and networks used for the simulation.

Dataset	Training examples	Test examples	Network size
UCI Small images (8x8 pixels)	3823	1797	64x36x10
MNIST large images (28x28 pixels)	60000	10000	784x300x10

The training and testing results of our SynT device are compared to an identical network constructed with ideal floating-point numeric precision, which represents the limit of neuromorphic algorithm and provides an important benchmark for the system.

For recognizing small, handwritten digits, the training accuracy of SynT reaches nearly the ideal numerical limit of 95.5% after 20 training epochs (Figure IV. 11.a). For the large digit dataset (Figure IV. 11.b), excellent accuracy is also obtained, reaching 95% compared to the ideal network of 98%. The decrease of recognition accuracy stems from the intrinsic operational nonlinearity of SynTs. However, with a low average write-noise level less than 1 nS, SynT performance on this pattern recognition test is relatively high compared to other available solutions.

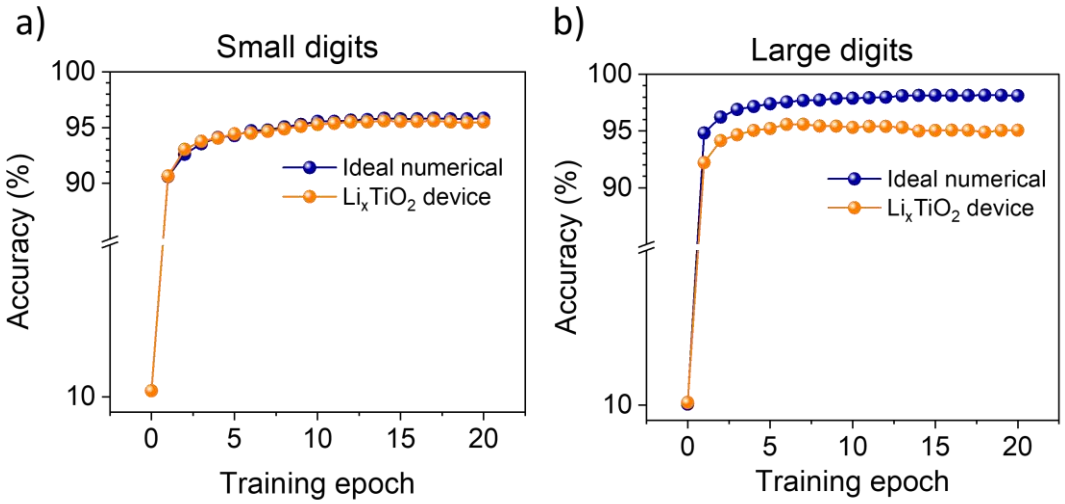

Figure IV. 11: Accuracy of the MNIST data recognition tests of a) small digits b) large digits.

A comparison based on the accuracy of the same recognition tests of different synaptic devices in the literatures can be found in Figure IV. 12. The devices are named based on their channel materials, namely α-MoO3 [33], PTIIG [34], SrFeO [35], IGZO [36], PEDOT:PSS/PEI [37], LiCoO2 [25]. From the results, we can clearly see that SynT performance is among the highest accuracy values of recently reported using the same simulation platform. Therefore, our CMOS-BEOL-compatible Li_xTiO_2-based SynTs, which exhibit flexible synaptic plasticity at low energy consumption, are demonstrated to be excellent artificial synapses for ANN application.

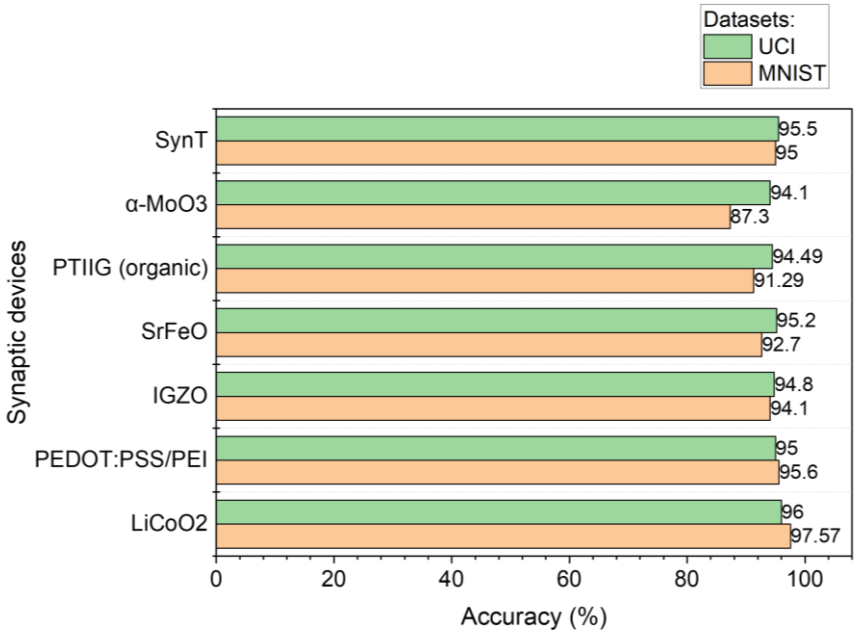

Figure IV. 12: Benchmarking SynT among the available synaptic transistors on the same image recognition tasks on the CrossSim platform.

2 APPLICABILITY OF SYNTS TO SPIKING NEURON NETWORKS

2.1 Background and objective

In the first part of chapter IV, the excellence of ANN in solving data-intensive, realistic problems such as image recognition, has been highlighted. The huge interest in this topic over the past decade elevates the neural network from a laboratory-level demonstration to a powerful tool for real world applications.

Nevertheless, the pursuit of greater accuracy in these networks have induced an unsustainable rate in the energy and processing demands for both training and deployment. Recently, there is a temptation to look for new approaches to build neural networks that are able to emulate biorealistic neuromorphic and cognitive properties of the brain, so called spiking neuron networks (SNNs). The SNN employs the concepts of spiking events: the information is encoded in the timing or frequency of the spikes. Many models have been proposed to account for the dynamics of a neuron. The leaky integrate-and-fire (LIF) model [38]–[40], which may be the simplest and most well-known one, is illustrated in Figure IV. 13. Action potentials (V1, V2, V3) from pre-neurons (travelling through axons) arrive at the synapses connecting a post-neuron, and are converted into currents which depend on the corresponding synaptic weights (w1, w2, w3) [41]. The sum of these currents (current I) progressively increases ("integrates") the membrane potential Vmem of a post-neuron (represented by a capacitance C). Since the membrane is not ideal, some leakage exists, represented by a resistance R. When Vmem reaches a constant threshold Vth, the post-neuron emits ("firing behavior") an output spike (Vspike), and the potential of Cmem is reset through a switch.

In the following, our main goal is to show the potential applicability of our SynTs to SNNs, by demonstrating the associative memory of a simple spiking neural network. To this end, we propose the design and the simulation (using a freeware circuit simulator "LT-SPICE") of an all-analog electronic circuit, composed of a LIF neuron and synapses, which reproduces very well the training of Pavlov's Dog (classical conditioning). Such a simple neural network, which involves the synaptic plasticity of our SynT, works in a "hardware" way, without any interface with software control machine. This interesting feature may give insights towards future embedded applications.

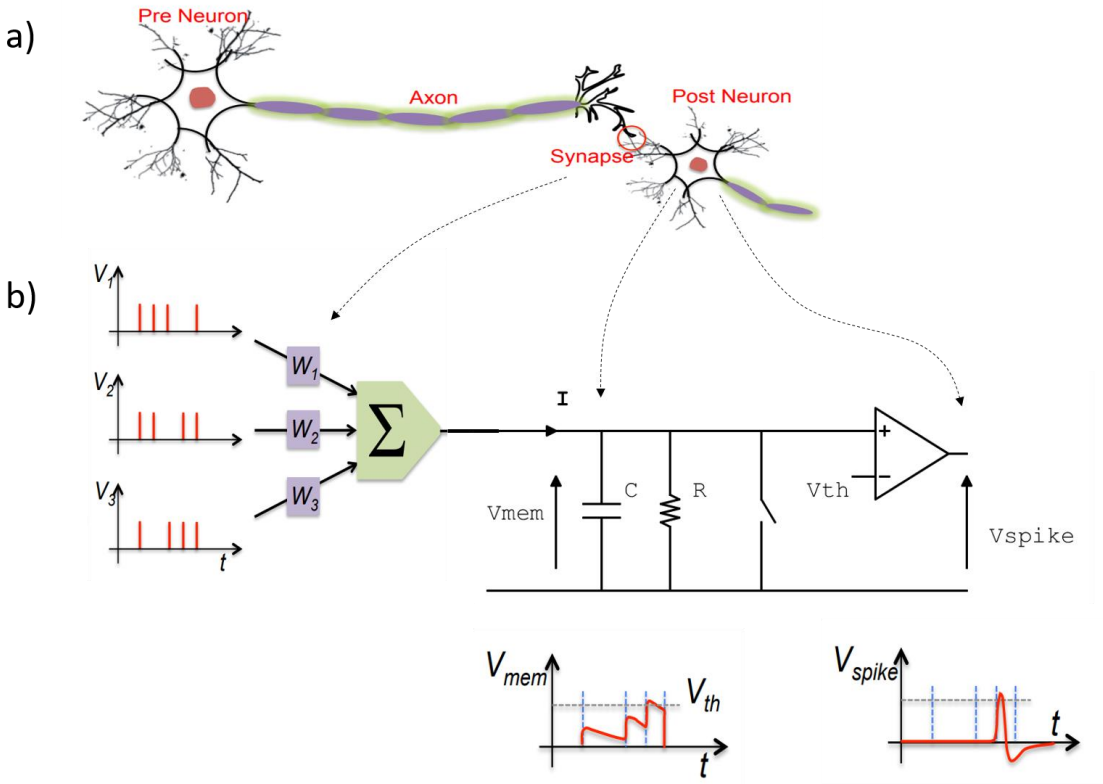

Figure IV. 13: a) A biological neuron with interconnecting synapses b) Representative LIF model for a biological neural network: at synaptic connections, the action potentials (V1, V2, V3) are converted into currents which depend on the synaptic weights (w1, w2, w3). The sum of these currents (current I) increases Vmem, until a threshold voltage (Vth) is reached. An output voltage spike (Vspike) is emitted and Vmem is reset through a switch. Extracted and modified from REF [41].

2.2 Pavlovian conditioning experiment

By getting a static shock when touching doorknobs several times in the winter, our brain is "trained" to stop touching these metal objects in dry ambience. This relates to the associative memory that we have, which dictates the ability to correlate different memories to the facts or events after a learning process [42]. This behavior has also been observed in many other animals. In 1897, Ivan Pavlov published his findings in associative learning (or classical/Pavlovian conditioning) on dogs [43]. The illustration of the experiment can be found in Figure IV. 14. There are four steps involved in the conditioning:

(I), (II) Before conditioning, salivation of the dog's mouth is set by the sight of food (unconditioned stimulus - US), and the dog does not salivate upon hearing the sound of the bell (neutral stimulus – NS).

(III) Then, during the conditioning period in which the sight of food is accompanied by a bell sound over a certain period, the dog learns gradually to associate/link the sound to the food.

(IV) After conditioning, the bell sound alone can trigger its salivation (conditioned stimulus – CS) without the intervention of vision.

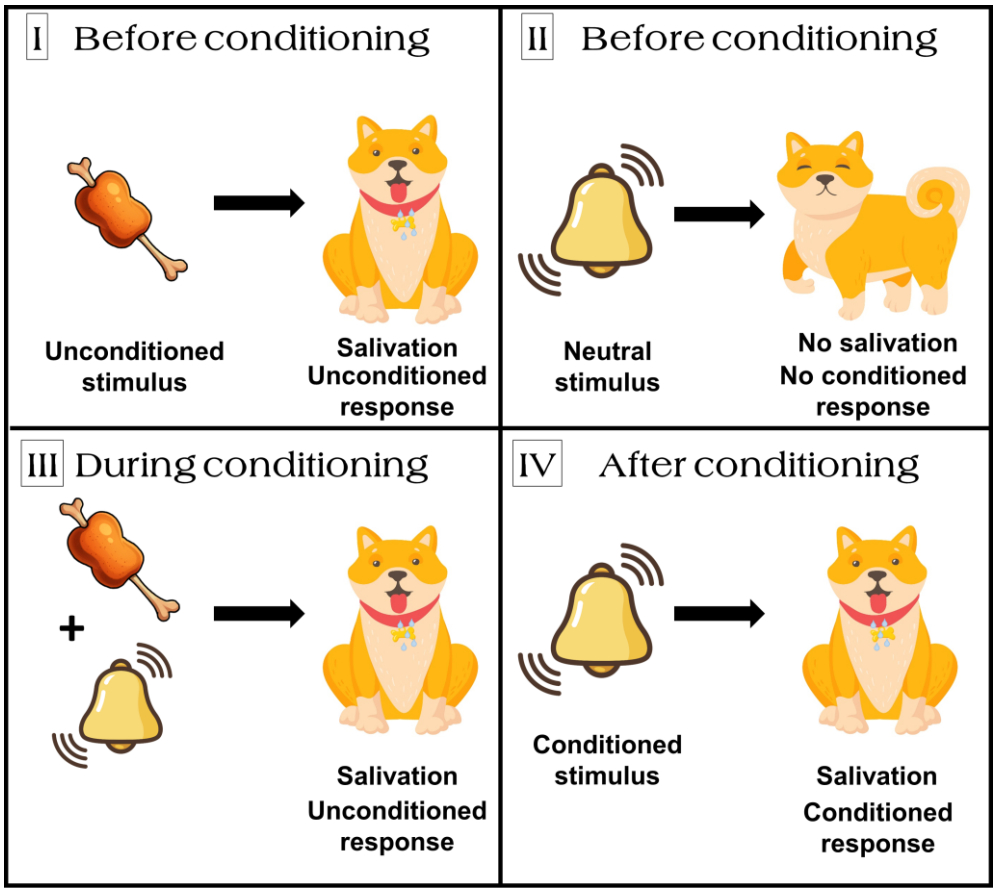

Figure IV. 14: Illustration of different steps in the Pavlov's dog experiment. Adapted from [44].

2.3 Electronic versions of Pavlov's dog in literature

This behavior has been intensively reproduced in artificial neural networks as a first important step in obtaining functionalities that resemble those of the human brain [45]–[49]. Pavlov's dog experiment can be modeled by a neural network circuit comprising two synapses and a neuron (Figure IV. 15). Here, the sight of food (US) and the sound of a bell (CS or NS), outputs of previous independent neurons, are modeled by electrical pulses sent to synapse 1 and 2 respectively. If the circuit is subjected to both input US and NS events, then after a sufficient number of occurrences, the output neuron starts to "salivate" upon the reception of CS only.

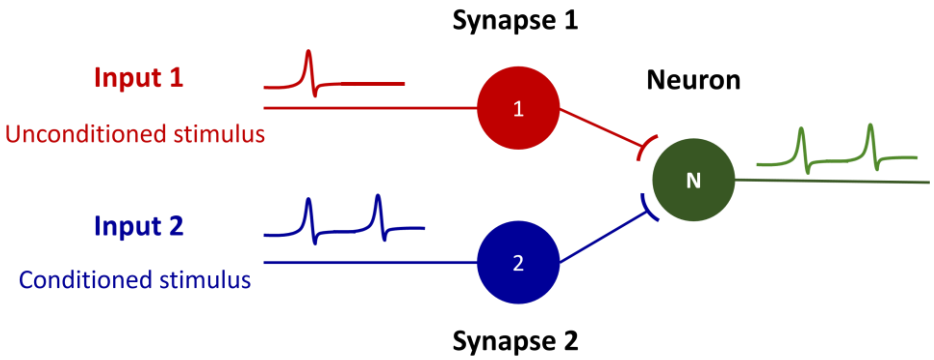

Figure IV. 15: Simple neural network model for Pavlov's dog experiment. Adapted from [45].

Electronic circuits which emulate the Pavlov Dog behavior have been reported previously in the literature. The Pavlovian conditioning was demonstrated mainly with two-terminal memristors: the proposed circuits include operational amplifiers [50]–[52], computing devices [53], or microcontrollers [45].

Associative learning was also demonstrated with three-terminal synaptic transistors [46]. Such configuration allows separating the read step from the write step. This characteristic yields several advantages. Specifically, the shape of the pulses applied during the write step can be conveniently chosen quite independently from the read pulses. The proposed circuit involves a computer interface to run the associative learning scheme.

Hereafter, the goal of our simulation task is to demonstrate the associative memory behavior, by using simple analog electronic elements, including our SynT with its experimental synaptic plasticity. Concerning the electronic neuron, we take advantage of a simple leaky integrate and fire (LIF) neuron circuit published recently [54], [55].

2.4 Design of an associative memory which involves our SynT

2.4.1 Neuron circuit model involving 1 synapse (resistor)

Toward the electronic implementation of the artificial neural network, we employ a LIF neuron model designed with analog components [54]. The circuit of the ultra-compact (UC) neuron and its simulated electronic response with a resistor as synapse, is illustrated in Figure IV. 16.

The neuron circuit is subdivided into two parts: "*leaky integrate-and-fire*" and "*axon signal transmission*" by the analogy to that of a biological neuron schematically drawn on the top panel. This analog LIF neuron circuit functions based on the I-V characteristic of

the silicon controlled rectifier (SCR) electronic component (marked as **U1** in the diagram). The key feature of the SCR is that it has a diode-like behavior with threshold and hysteresis that can be controlled by the SCR gate.

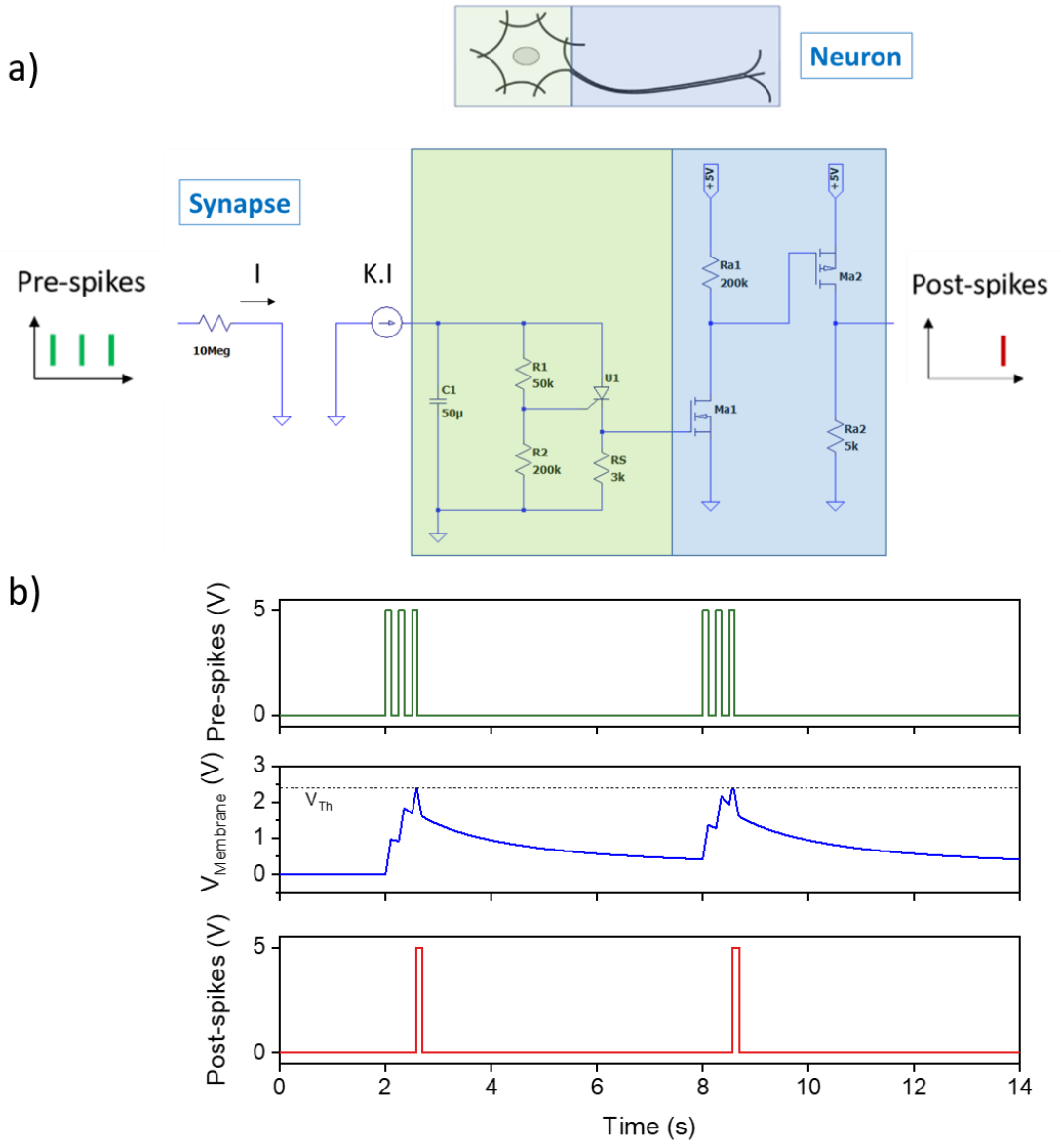

Figure IV. 16: a) Illustration of Leaky integrate-and-fire neuron model circuit and b) its electrical response to pre-spikes.

The *leaky integrate* feature is implemented by a RC pair (**C1**, **R1**, and **R2**). The capacitor acts as a membrane potential ($V_{Membrane}$), integrating the charge of incoming current spikes, which may leak out through the resistor (R = R1 + R2) during the time intervals between spikes. The key *fire* feature of the neuron is realized by the SCR's voltage threshold, which is set by its anode-cathode tension and is tuned by the gate through the resistors **R1** and **R2**. When the voltage threshold is attained, the SCR switches to the on-

state, and the capacitor quickly discharges through the small **RS**, generating a spike of current. The SCR remains in the on-state until the current decreases to the holding current value (I_{hold}) of the SCR, when the capacitor is almost fully discharged (then the SCR switches back to the off-state). In order for the spike to be able to drive a downstream neuron, the strength of the signal needs to be reinforced. As shown in Figure IV. 16.a, this is implemented by a pair of MOS transistors that play the role of the axon. Thus, the UC neuron is implemented with just one SCR and two transistors, plus one "membrane" capacitor and a few resistors.

Figure IV. 16.b shows an example of electrical response of an UC neuron circuit to which a synapse is connected (10MΩ, corresponding to a 100nS conductance). Here, we employ components with values **C1** = 50 µF, **R1** = 50 kΩ, **R2** = 200 kΩ, **RS** = 3 kΩ, **Ra1** = 200 kΩ, and **Ra2** = 5 kΩ to model the leaky integrate and then fire features upon the reception of a train of potential pre-spikes from the previous layer of neurons.

From the figure, the pre-spikes of voltage (5Volts) yield current spikes (I) which (after being amplified by a factor K ~ 1000) increase gradually the membrane potential $V_{Membrane}$ up to the threshold voltage before emitting an output spike and discharging the accumulated potential.

2.4.2 Neuron circuit model involving 2 synapses (resistors)

In the following proposed circuit, the signals of "food sight" (V_{FOOD}, connected to synapse 1) and "bell sound" (V_{BELL}, connected to synapse 2) neurons are simulated by voltage sources which provide voltage spikes (5V amplitude and 100ms duration).

Two synapses (synapse1: 10MΩ and synapse2: 30MΩ) are connected to the LIF neuron circuit, through a "current mirror" stage, which allows to sum up (and amplify: gain 25x35) the currents which flow through the two synapses. We selected the values (capacitor **C1** = 50 nF and resistors **R1** = 10 kΩ, **R2** = 200 kΩ, **RS** = 3 kΩ, **Ra1** = **Ra2** = 5 kΩ), so that one input spike coming from synapse1 allows one output spike, and so that synapse2 cannot yield any output spike. Figure IV. 19.b illustrates the obtained behavior.

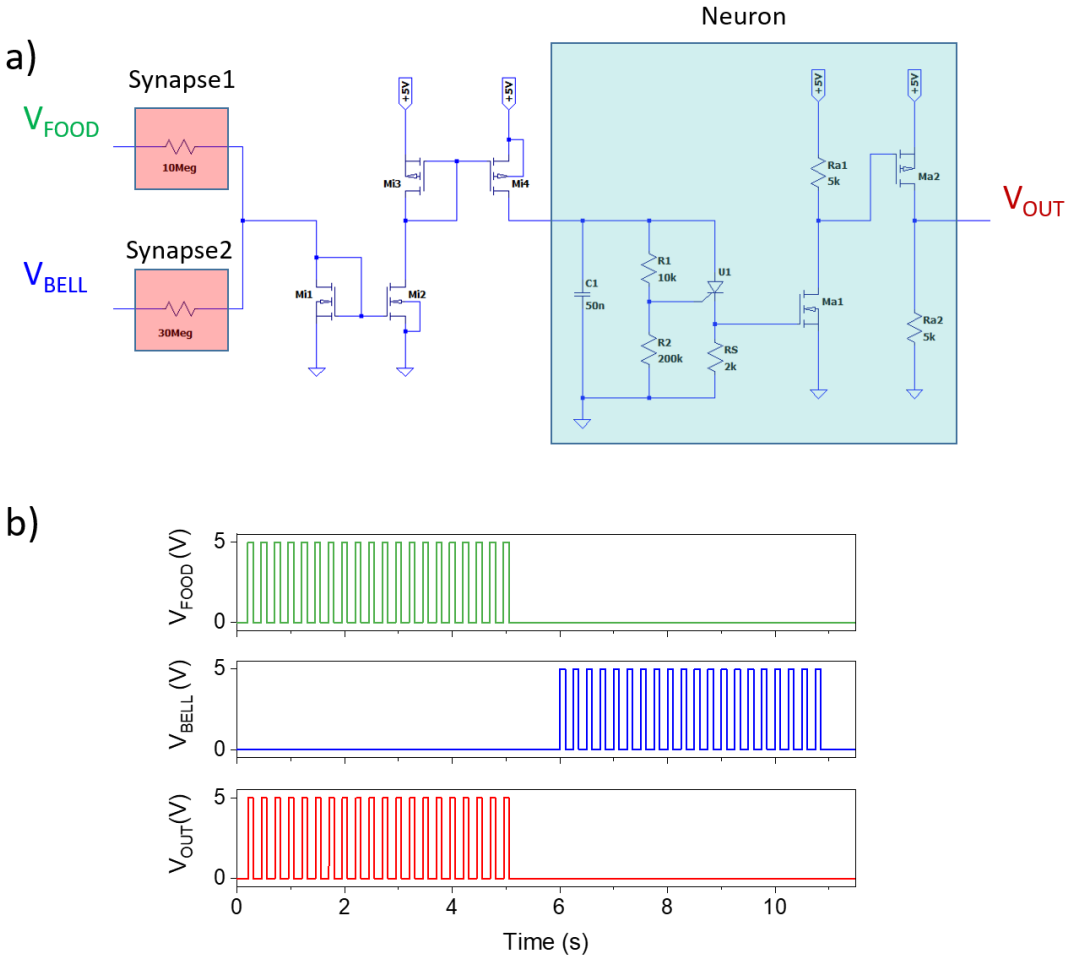

Figure IV. 17: a) Circuit with two synapses (resistors of 10 MΩ and 30 MΩ) in red, and the LIF neuron circuit in green b) Simulation results of the evolution of the V_{OUT} voltage, as a function of V_{FOOD} and V_{BELL}.

In the following, a circuit detecting the simultaneity of V_{FOOD} and V_{BELL} will be added to the present circuit, and synapse2 will be replaced by a model corresponding to the SynT behavior.

2.4.3 Neuron circuit model involving 2 synapses: 1 resistor and 1 SynT

The whole circuit is illustrated in Figure IV. 18 below. For Synapse1, we still use a resistor (10 MΩ). On the other hand, our SynT, with its synaptic plasticity, is adapted to become synapse2.

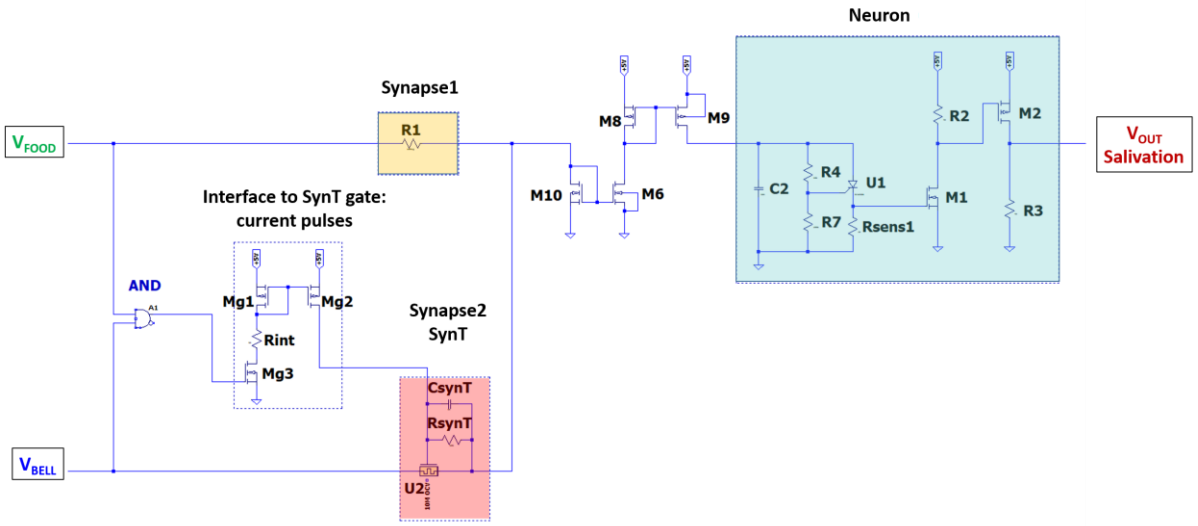

Figure IV. 18: Schematic of the whole circuit: when V_{FOOD} and V_{BELL} arrive simultaneously to the AND component, the latter delivers a 5V voltage output, which is converted to a (in the nA range) current arriving at the gate of synapse2 (SynT): this results in a gradual increase of the SynT conductance.

Both sources (V_{FOOD} and V_{BELL}) are connected to an AND logic component: when they arrive simultaneously to the AND component, the latter delivers a 5V output voltage, which is converted to a current (through an interface circuit). This current is injected into the gate of the SynT (synapse2) which allows the SynT conductance to increase gradually (during the conditioning period).

2.4.3.1 Interface to the SynT gate

Figure IV. 19 illustrates the function of this circuit. The pulses from **A1** will turn on the **Mg3** transistor, allowing a reference current to flow through the resistor **Rint**. By controlling the resistance value of **Rint**, the desired bias current for SynT programming can be tuned. A current mirror (**Mg1** and **Mg2** pair of pmos transistors) will "copy" such a current, which is then sent to the synaptic transistor gate.

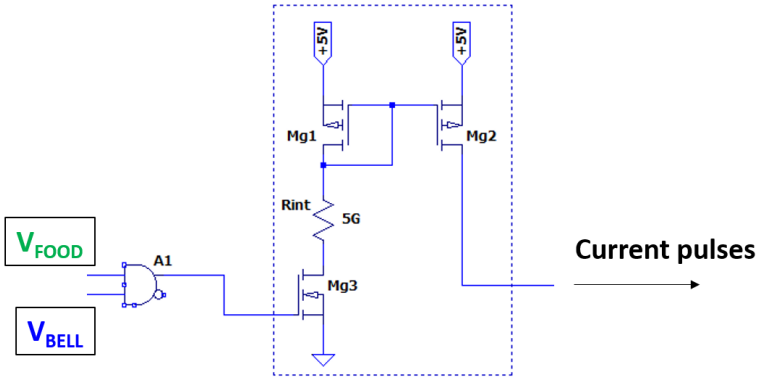

Figure IV. 19: Interface proposed to inject current to the SynT gate (current controlled mode).

2.4.3.2 Electrical model of the SynT device

We model our SynT behavior with a combination of three components (Figure IV.19): a small capacitor (**CsynT** = 1 nF) can be charged up by the incoming spikes, to simulate the behavior of SynT gate stack. This capacitance is estimated based on the measured capacitance of a MIM structure [56], taking into account the active area and the thickness of the transistor's gate stack. Besides, A resistor with a high value (**RsynT** = 10 GΩ) is used to mimic the high leakage resistance of the gate stack (an estimation of RsynT value awaits to be obtained in the future). Finally, we use a transistor **U2** in which the current I_{DS} can be controlled by the gate voltage V_G to model the conductance evolution: from the experimental measurement of the channel conductance as a function of gate voltage (Figure IV. 20.b), we fit the I_{DS} of the component **U2** with a linear function as follows:

$$I_{DS} = G_{DS} \times V_{DS} \qquad \text{with} \qquad G_{DS} = G_{Low} + \frac{(G_{High} - G_{Low})}{1.5V} \times V_C \qquad \text{Eq.1}$$

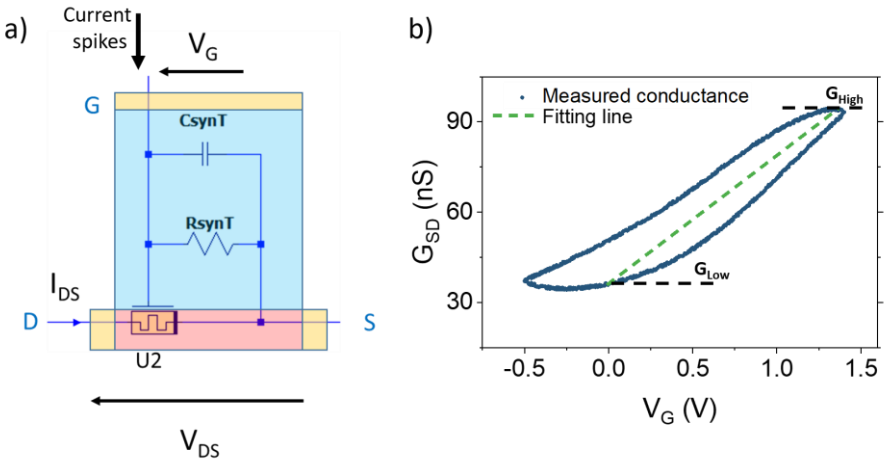

Figure IV. 20: a) Synaptic transistor compact model. b) Implementing the measured conductance as a function of the gate voltage.

The response of the synaptic transistor model to a train of current pulses is simulated in Figure IV. 21. Current pulses (I_G) of around 0.7 nA and 100 ms duration are injected into the gate (equivalent to 70 pC charge per pulse: this is approximately equivalent to the actual charge needed to switch experimentally between adjacent states of the electrochemical synaptic transistors). This leads to a gradual increase of the gate potential V_G (blue curve), thus to an increase of the G_{SD} conductance, and to an increasing magnitude of the current spikes I_{DS} (green curve) flowing across the SynT channel.

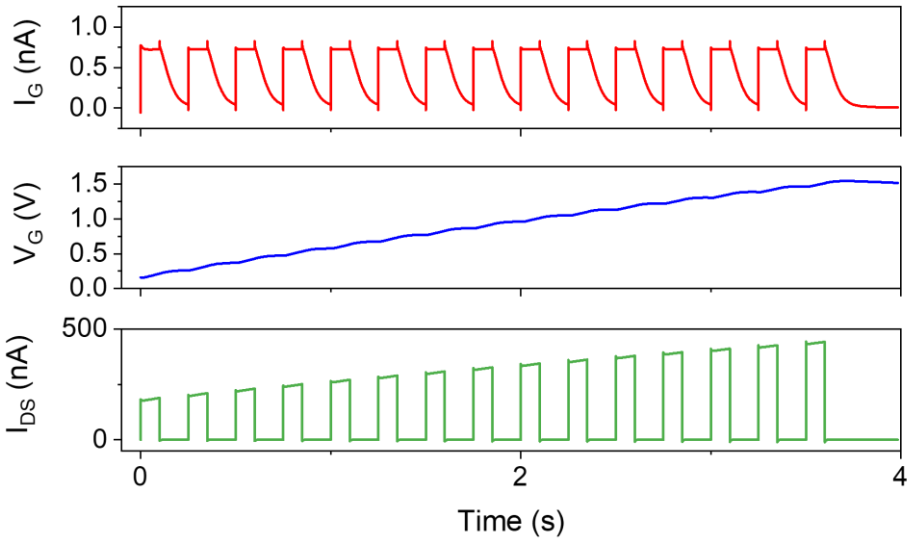

Figure IV. 21: Simulation of the SynT response (Top) The writing current pulses. (Middle) The increase of gate potential (simulated by the potential across C_{SynT}). (Bottom) The channel current I_{DS} of SynT.

Such evolution of G_{SD} will contribute to the associative memory behavior that will be simulated in the next step.

Note that we tried to program the synaptic transistor with potential pulses instead of current pulses. However, with voltage programming, the V_{CsynT} increased rapidly with a few first pulses. Thus, the conductance increase was quite nonlinear and not appropriate for the associative learning application.

2.5 Circuit simulation of the associative memory with SynT characteristics

With the components described above, we reproduce Pavlov's dog experiment with our neural network circuit. In the first step, applying "food sight" stimuli triggers the "salivation", whereas the "bell sound" stimuli do not induce any effect (step II). Electronically speaking, the output neuron fires because the currents through synapse 1 are high enough to get the neuron potential over the threshold (due to the resistance R1 = 10 MΩ, which

corresponds to a 100 nS conductance) while the current pulses sent to the neuron via synapse2 are too low (the initial conductance of the SynT is only 30 nS).

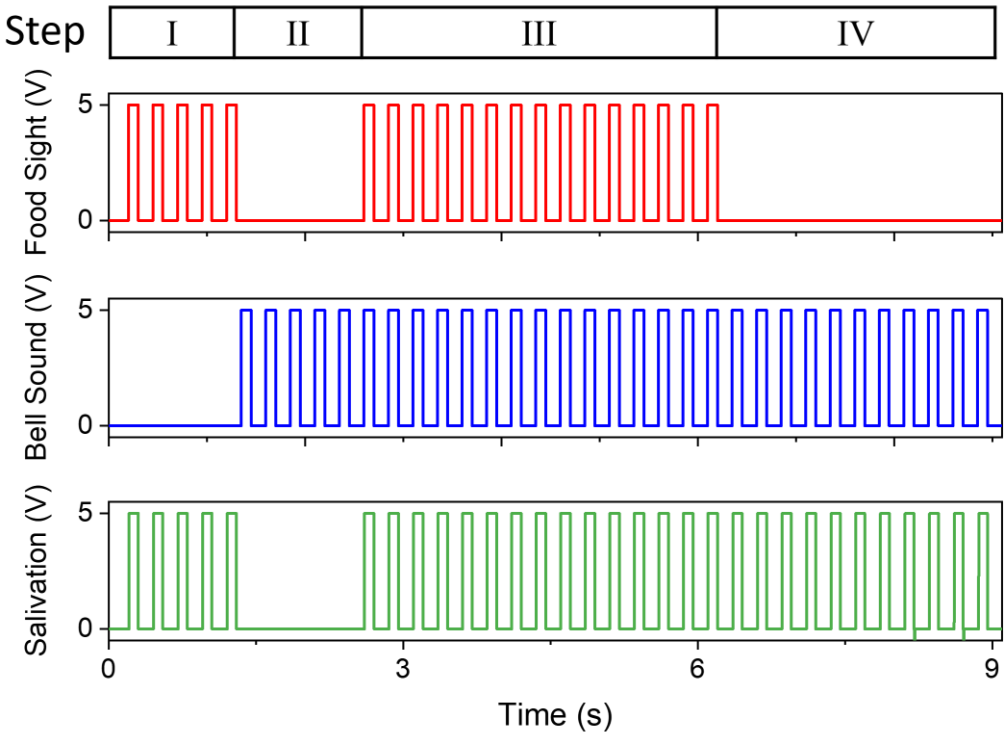

Figure IV. 22: Simulation of the development of the associative memory, with a neural network circuit composed of SynT as a synaptic element.

In the learning phase (step III), stimulus voltages are sent simultaneously to both synapses. As a direct result of "food sight," the "salivation" neuron fires correspondingly. At the same time, the overlap between two sources of stimulus pulses facilitates the potentiation of the synaptic transistor via a train of current pulses. Thus, conductance of synapse 2 increases gradually: an association between "food sight" and "bell sound" signals develops progressively.

Finally, in step IV, a train of signals of "bell sound" only is applied to the neuron: salivation still occurs, thereby proving the associative learning ability of the circuit.

We may mention that during the training of the circuit, the conductance of the SynTs could only increase (with positive current pulses). The realistic high leakage resistance guarantees the programmed states to be nonvolatile. In this consideration, the circuit can only demonstrate the most important feature, which is associative learning and not the extinction phase, in which the dog gradually forgets about the link between "bell sound" and "food sight".

With this circuit design featuring SynTs, we implemented efficiently an associative memory embedded network, which could be beneficial to real-world applications such as healthcare [57], [58], or other robotic sensing and reasoning by linking the realistic stimuli detected from sensory systems [59].

3 CONCLUSIONS

In this chapter, we investigated the use cases of the first generation of SynTs for different types of neural networks. First, we covered the working principle of the artificial neural network and then the related hardware engineering approaches to improve further the efficiency of the deep learning algorithms. The synaptic transistors could be used to construct computing-in-memory hardware to accelerate the dominant workload vector-matrix multiplication operations of deep neural networks. Their long-term plasticity could be employed to implement the weight matrix of the network, while the ability to switch the analog states independently by using the gate electrode can be helpful for the training phase of these systems.

Simulation of crossbar arrays with SynTs as cross-point memory devices was realized using the CrossSim simulator platform. With the experimental results from 100 conductance modulation cycles, the simulator could take into account the realistic nonlinearity of the SynT and benchmark its performance on handwritten datasets. Even though the switching nonlinearity of SynT can be observed on its probability distribution function graph of $\Delta G = f(G_0)$ due to the high activity of ion intercalation or extraction at around 1 V, its pattern recognition accuracy on given tasks remains high compared to other technologies employing the same simulation platform.

At the current stage, the access devices are mandatory for the parallel weight update operations of SynT crossbars. However, it would be beneficial from the energy and integration point of view if we could elaborate selector-free SynT arrays, which can be programmed in parallel (by stochastic update scheme for example). These SynTs can be integrated at BEOL, on top of other front-end-of-line circuit components (integrators and ADCs), which would be a major advantage in implementing area- and power-efficient neuromorphic processors. The approaches toward selector-free SynTs in the next generation will be proposed in the following chapter.

Our SynT was also used to demonstrate a bio-plausible learning rule – associative memory with an all-analog spiking neural network circuit. In this simulation using LT-SPICE, we proposed a design of a neural network consisting of two synapses, including a resistor and a compact model of SynT, and a leaky integrate-and-fire neuron. The training phase of Pavlov's dog experiment under the synchronous stimuli from "food sight" and "bell sound" was simulated by the potentiation process of the SynT. A train of current spikes of 0.7 nA was used to increase the current flowing through SynT's channel gradually. After the current is high enough to saturate the membrane potential in the neural, the circuit was successfully trained to react to "bell sound" stimuli by firing potential spikes out of the neuron. With the two preliminary examples, we showed that the SynTs can be employed in both types of neural networks.

4 REFERENCES

[1] "Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2020, with forecasts from 2021 to 2025." https://www.statista.com/statistics/871513/worldwide-data-created/ (accessed May 10, 2022).

[2] D. Reinsel, J. Gantz, and J. Rydning, "The Digitization of the World from Edge to Core," p. 28, 2018.

[3] S. Tulyakov *et al.*, "Time Lens: Event-based Video Frame Interpolation," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, TN, USA, Jun. 2021, pp. 16150–16159. doi: 10.1109/CVPR46437.2021.01589.

[4] J. Gu *et al.*, "Recent advances in convolutional neural networks," *Pattern Recognition*, vol. 77, pp. 354–377, May 2018, doi: 10.1016/j.patcog.2017.10.013.

[5] A. Rajkomar *et al.*, "Scalable and accurate deep learning with electronic health records," *npj Digital Med*, vol. 1, no. 1, p. 18, Dec. 2018, doi: 10.1038/s41746-018-0029-1.

[6] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90.

[7] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.

[8] J. Dean, "1.1 The Deep Learning Revolution and Its Implications for Computer Architecture and Chip Design," in *2020 IEEE International Solid- State Circuits Conference - (ISSCC)*, San Francisco, CA, USA, Feb. 2020, pp. 8–14. doi: 10.1109/ISSCC19947.2020.9063049.

[9] I. S. Choi and Y.-S. Kee, "Energy Efficient Scale-In Clusters with In-Storage Processing for Big-Data Analytics," in *Proceedings of the 2015 International Symposium on Memory Systems*, Washington DC DC USA, Oct. 2015, pp. 265–273. doi: 10.1145/2818950.2818983.

[10] H.-S. P. Wong and S. Salahuddin, "Memory leads the way to better computing," *Nature Nanotech*, vol. 10, no. 3, pp. 191–194, Mar. 2015, doi: 10.1038/nnano.2015.29.

[11] H. Tsai, S. Ambrogio, P. Narayanan, R. M. Shelby, and G. W. Burr, "Recent progress in analog memory-based accelerators for deep learning," *J. Phys. D: Appl. Phys.*, vol. 51, no. 28, p. 283001, Jul. 2018, doi: 10.1088/1361-6463/aac8a5.

[12] S. Dai *et al.*, "Recent Advances in Transistor-Based Artificial Synapses," *Adv. Funct. Mater.*, vol. 29, no. 42, p. 1903700, Oct. 2019, doi: 10.1002/adfm.201903700.

[13] H. Han, H. Yu, H. Wei, J. Gong, and W. Xu, "Recent Progress in Three-Terminal Artificial Synapses: From Device to System," *Small*, vol. 15, no. 32, p. 1900695, Aug. 2019, doi: 10.1002/smll.201900695.

[14] D. Ielmini and H.-S. P. Wong, "In-memory computing with resistive switching devices," *Nat Electron*, vol. 1, no. 6, pp. 333–343, Jun. 2018, doi: 10.1038/s41928-018-0092-2.

[15] A. Sebastian, M. Le Gallo, R. Khaddam-Aljameh, and E. Eleftheriou, "Memory devices and applications for in-memory computing," *Nat. Nanotechnol.*, vol. 15, no. 7, pp. 529–544, Jul. 2020, doi: 10.1038/s41565-020-0655-z.

[16] S. Agarwal *et al.*, "Resistive memory device requirements for a neural algorithm accelerator," in *2016 International Joint Conference on Neural Networks (IJCNN)*, Vancouver, BC, Canada, Jul. 2016, pp. 929–938. doi: 10.1109/IJCNN.2016.7727298.

[17] S. Ambrogio *et al.*, "Equivalent-accuracy accelerated neural-network training using analogue memory," *Nature*, vol. 558, no. 7708, pp. 60–67, Jun. 2018, doi: 10.1038/s41586-018-0180-5.
[18] "Inference: The Next Step in GPU-Accelerated Deep Learning." https://developer.nvidia.com/blog/inference-next-step-gpu-accelerated-deep-learning/ (accessed May 10, 2022).
[19] G. Singh *et al.*, "A Review of Near-Memory Computing Architectures: Opportunities and Challenges," in *2018 21st Euromicro Conference on Digital System Design (DSD)*, Prague, Aug. 2018, pp. 608–617. doi: 10.1109/DSD.2018.00106.
[20] S. Yu, H. Jiang, S. Huang, X. Peng, and A. Lu, "Compute-in-Memory Chips for Deep Learning: Recent Trends and Prospects," *IEEE Circuits Syst. Mag.*, vol. 21, no. 3, pp. 31–56, 2021, doi: 10.1109/MCAS.2021.3092533.
[21] W. Haensch *et al.*, "A Co-design view of Compute in-Memory with Non- Volatile Elements for Neural Networks," p. 56.
[22] X. Peng, S. Huang, Y. Luo, X. Sun, and S. Yu, "DNN+NeuroSim: An End-to-End Benchmarking Framework for Compute-in-Memory Accelerators with Versatile Device Technologies," in *2019 IEEE International Electron Devices Meeting (IEDM)*, San Francisco, CA, USA, Dec. 2019, p. 32.5.1-32.5.4. doi: 10.1109/IEDM19573.2019.8993491.
[23] M. J. Rasch *et al.*, "A Flexible and Fast PyTorch Toolkit for Simulating Training and Inference on Analog Crossbar Arrays," in *2021 IEEE 3rd International Conference on Artificial Intelligence Circuits and Systems (AICAS)*, Washington DC, DC, USA, Jun. 2021, pp. 1–4. doi: 10.1109/AICAS51828.2021.9458494.
[24] S. Agarwal *et al.*, "Energy Scaling Advantages of Resistive Memory Crossbar Based Computation and Its Application to Sparse Coding," *Front. Neurosci.*, vol. 9, Jan. 2016, doi: 10.3389/fnins.2015.00484.
[25] E. J. Fuller *et al.*, "Li-Ion Synaptic Transistor for Low Power Analog Computing," *Adv. Mater.*, vol. 29, no. 4, p. 1604310, Jan. 2017, doi: 10.1002/adma.201604310.
[26] "CrossSim: Crossbar Simulator." https://cross-sim.sandia.gov/ (accessed Jan. 10, 2020).
[27] P. Narayanan *et al.*, "Toward on-chip acceleration of the backpropagation algorithm using nonvolatile memory," *IBM J. Res. & Dev.*, vol. 61, no. 4/5, p. 11:1-11:11, Jul. 2017, doi: 10.1147/JRD.2017.2716579.
[28] G. W. Burr *et al.*, "Access devices for 3D crosspoint memory," *Journal of Vacuum Science & Technology B, Nanotechnology and Microelectronics: Materials, Processing, Measurement, and Phenomena*, vol. 32, no. 4, p. 040802, Jul. 2014, doi: 10.1116/1.4889999.
[29] Y. Li *et al.*, "Low-Voltage, CMOS-Free Synaptic Memory Based on Li_xTiO_2 Redox Transistors," *ACS Appl. Mater. Interfaces*, vol. 11, no. 42, pp. 38982–38992, Oct. 2019, doi: 10.1021/acsami.9b14338.
[30] E. J. Fuller *et al.*, "Parallel programming of an ionic floating-gate memory array for scalable neuromorphic computing," p. 6, 2019, doi: 10.1126/science.aaw5581.
[31] T. P. Xiao, C. H. Bennett, B. Feinberg, S. Agarwal, and M. J. Marinella, "Analog architectures for neural network acceleration based on non-volatile memory," *Applied Physics Reviews*, vol. 7, no. 3, p. 031301, Sep. 2020, doi: 10.1063/1.5143815.
[32] P. Hosseini, A. Sebastian, N. Papandreou, C. D. Wright, and H. Bhaskaran, "Accumulation-Based Computing Using Phase-Change Memories With FET Access Devices," *IEEE Electron Device Lett.*, vol. 36, no. 9, pp. 975–977, Sep. 2015, doi: 10.1109/LED.2015.2457243.

[33] C. Yang *et al.*, "All-Solid-State Synaptic Transistor with Ultralow Conductance for Neuromorphic Computing," *Adv. Funct. Mater.*, vol. 28, no. 42, p. 1804170, Oct. 2018, doi: 10.1002/adfm.201804170.
[34] D.-G. Seo *et al.*, "Versatile neuromorphic electronics by modulating synaptic decay of single organic synaptic transistor: From artificial neural networks to neuro-prosthetics," *Nano Energy*, vol. 65, p. 104035, Nov. 2019, doi: 10.1016/j.nanoen.2019.104035.
[35] P. Shi *et al.*, "Solid-state electrolyte gated synaptic transistor based on SrFeO2.5 film channel," *Materials & Design*, vol. 210, p. 110022, Nov. 2021, doi: 10.1016/j.matdes.2021.110022.
[36] Y. He *et al.*, "IGZO-based floating-gate synaptic transistors for neuromorphic computing," *J. Phys. D: Appl. Phys.*, vol. 53, no. 21, p. 215106, May 2020, doi: 10.1088/1361-6463/ab7bb4.
[37] Y. van de Burgt *et al.*, "A non-volatile organic electrochemical device as a low-voltage artificial synapse for neuromorphic computing," *Nature Mater*, vol. 16, no. 4, pp. 414–418, Apr. 2017, doi: 10.1038/nmat4856.
[38] P. Stoliar *et al.*, "A Leaky-Integrate-and-Fire Neuron Analog Realized with a Mott Insulator," *Adv. Funct. Mater.*, vol. 27, no. 11, p. 1604740, Mar. 2017, doi: 10.1002/adfm.201604740.
[39] H. Lim *et al.*, "Reliability of neuronal information conveyed by unreliable neuristor-based leaky integrate-and-fire neurons: a model study," *Sci Rep*, vol. 5, no. 1, p. 9776, Sep. 2015, doi: 10.1038/srep09776.
[40] H. Huang *et al.*, "Quasi-Hodgkin–Huxley Neurons with Leaky Integrate-and-Fire Functions Physically Realized with Memristive Devices," *Adv. Mater.*, vol. 31, no. 3, p. 1803849, Jan. 2019, doi: 10.1002/adma.201803849.
[41] S. Dutta, V. Kumar, A. Shukla, N. R. Mohapatra, and U. Ganguly, "Leaky Integrate and Fire Neuron by Charge-Discharge Dynamics in Floating-Body MOSFET," *Sci Rep*, vol. 7, no. 1, p. 8257, Dec. 2017, doi: 10.1038/s41598-017-07418-y.
[42] M. J. Kahana, "Associative symmetry and memory theory," *Memory & Cognition*, vol. 30, no. 6, pp. 823–840, Sep. 2002, doi: 10.3758/BF03195769.
[43] R. E. Clark, "The classical origins of Pavlov's conditioning," *Integr. psych. behav.*, vol. 39, no. 4, pp. 279–294, Oct. 2004, doi: 10.1007/BF02734167.
[44] M. Kumar, S. Abbas, J.-H. Lee, and J. Kim, "Controllable digital resistive switching for artificial synapses and pavlovian learning algorithm," *Nanoscale*, vol. 11, no. 33, pp. 15596–15604, 2019, doi: 10.1039/C9NR02027F.
[45] Y. V. Pershin and M. Di Ventra, "Experimental demonstration of associative memory with memristive neural networks," *Neural Networks*, vol. 23, no. 7, pp. 881–886, Sep. 2010, doi: 10.1016/j.neunet.2010.05.001.
[46] O. Bichler *et al.*, "Pavlov's Dog Associative Learning Demonstrated on Synaptic-Like Organic Transistors," *Neural Computation*, vol. 25, no. 2, pp. 549–566, Feb. 2013, doi: 10.1162/NECO_a_00377.
[47] A. Cisternas Ferri, A. Rapoport, P. I. Fierens, G. A. Patterson, E. Miranda, and J. Suñé, "On the Application of a Diffusive Memristor Compact Model to Neuromorphic Circuits," *Materials*, vol. 12, no. 14, p. 2260, Jul. 2019, doi: 10.3390/ma12142260.
[48] K. Moon *et al.*, "Hardware implementation of associative memory characteristics with analogue-type resistive-switching device," *Nanotechnology*, vol. 25, no. 49, p. 495204, Dec. 2014, doi: 10.1088/0957-4484/25/49/495204.
[49] S. G. Hu *et al.*, "Synaptic long-term potentiation realized in Pavlov's dog model based on a NiO_x-based memristor," *Journal of Applied Physics*, vol. 116, no. 21, p. 214502, Dec. 2014, doi: 10.1063/1.4902515.

[50] M. Ziegler *et al.*, "An Electronic Version of Pavlov's Dog," *Adv. Funct. Mater.*, vol. 22, no. 13, pp. 2744–2749, Jul. 2012, doi: 10.1002/adfm.201200244.
[51] L. Wang, H. Li, S. Duan, T. Huang, and H. Wang, "Pavlov associative memory in a memristive neural network and its circuit implementation," *Neurocomputing*, vol. 171, pp. 23–29, Jan. 2016, doi: 10.1016/j.neucom.2015.05.078.
[52] S. Du, Q. Deng, Q. Hong, and C. Wang, "A memristor-based circuit design of pavlov associative memory with secondary conditional reflex and its application," *Neurocomputing*, vol. 463, pp. 341–354, Nov. 2021, doi: 10.1016/j.neucom.2021.08.045.
[53] C. Sun, C. Wang, and C. Xu, "A full-function memristive pavlov associative memory circuit with inter-stimulus interval effect," *Neurocomputing*, vol. 506, pp. 68–83, Sep. 2022, doi: 10.1016/j.neucom.2022.07.044.
[54] M. J. Rozenberg, O. Schneegans, and P. Stoliar, "An ultra-compact leaky-integrate-and-fire model for building spiking neural networks," *Sci Rep*, vol. 9, no. 1, p. 11123, Dec. 2019, doi: 10.1038/s41598-019-47348-5.
[55] P. Stoliar, O. Schneegans, and M. J. Rozenberg, "Biologically Relevant Dynamical Behaviors Realized in an Ultra-Compact Neuron Model," *Front. Neurosci.*, vol. 14, p. 421, May 2020, doi: 10.3389/fnins.2020.00421.
[56] V. Sallaz, S. Oukassi, F. Voiron, R. Salot, and D. Berardan, "Assessing the potential of LiPON-based electrical double layer microsupercapacitors for on-chip power storage," *Journal of Power Sources*, vol. 451, p. 227786, Mar. 2020, doi: 10.1016/j.jpowsour.2020.227786.
[57] M. Aldape-Pérez, A. Alarcón-Paredes, C. Yáñez-Márquez, I. López-Yáñez, and O. Camacho-Nieto, "An Associative Memory Approach to Healthcare Monitoring and Decision Making," *Sensors*, vol. 18, no. 8, p. 2690, Aug. 2018, doi: 10.3390/s18082690.
[58] J. Wu, P. Huang, C. Lin, and C. Li, "Blood leakage detection during dialysis therapy based on fog computing with array photocell sensors and heteroassociative memory model," *Healthcare Technology Letters*, vol. 5, no. 1, pp. 38–44, Feb. 2018, doi: 10.1049/htl.2017.0091.
[59] M. Hampo *et al.*, "Associative Memory in Spiking Neural Network Form Implemented on Neuromorphic Hardware," in *International Conference on Neuromorphic Systems 2020*, Oak Ridge TN USA, Jul. 2020, pp. 1–8. doi: 10.1145/3407197.3407602.

CHAPTER V

OPTIMIZATION TOWARDS A SECOND GENERATION OF SYNAPTIC TRANSISTORS

ABSTRACT

In chapter 5, we will discuss the means to improve the performance of SynTs in the next generations using material engineering and device design approaches.

Section 1 of this chapter is dedicated to giving an overview of different metrics of the first generation (gen1) of SynTs and seek for improvement in various areas, especially the switching time, dynamic range, and endurance for neural network applications. To better build the next generation, we need to focus on searching for constituting materials exhibiting desired characteristics such as high ionic transport activity, low electronic conductance for electrolyte materials, and significant conductivity change upon Li intercalation for channel materials. The proposed functionalities can be obtained by elemental doping, phase engineering, or thickness downscaling. On the other hand, device design further optimizes the performance merits on top of the material properties. Some of the methods include shortening the migration pathway of ions by vertically stacking and maximizing area overlay. A passivation layer that protects the Li-based devices from environmental factors can enhance the retention and endurance of new SynTs.

Section 2 of the chapter describes some of the efforts that we have already initiated to advance the figures of merits in the second generation (gen2) SynTs. First, we show the new design of a set of masks with different test structures, which allows us to study the transistors systematically with the design parameters and the electrochemical properties of the gate stack. In this design, the transistors have channel gaps varying from 200 nm to 1.5 µm. In addition, we have started the study of ultrathin films of LiPON (prepared by atomic layer deposition – ALD) and $LiNbO_3$ (prepared by pulsed laser deposition – PLD) as the electrolyte and channel layers, respectively. The preliminary results reveal the potential of these materials for synaptic transistor applications. The optimization work for these exercises remains to be done. Some conclusions in this aspect will be drawn in the last section.

TABLE OF CONTENTS

1 IDEAS TO IMPROVE THE PERFORMANCE OF SYNTS 172
1.1 General considerations 172
1.2 Materials engineering 173
1.2.1 Channel materials 173
1.2.2 Electrolyte materials 174
1.3 Device design 176
1.3.1 Thin-film stacking and overlay 176
1.3.2 Dimension scaling 176
1.3.3 Gate stack design 177
1.4 Summary on the approaches 177
2 CURRENT WORK TOWARDS A SECOND GENERATION OF SYNTS 179
2.1 New design 179
2.2 Materials 181
3 CONCLUSIONS 183
4 REFERENCES 184

1 IDEAS TO IMPROVE THE PERFORMANCE OF SYNTS

1.1 General considerations

In my thesis, the first functional generation of the electrochemical synaptic transistor, the gen1 SynT, has been developed. While excellent synaptic functionalities have been demonstrated on these devices, such as ultralow operation power, linearity analog switching, good write endurance, etc., they still inherit the common characteristics of electrochemical systems, namely sluggish ionic kinetics, limited state retention, and small dynamic range with the modest amount of available ions. For these reasons, the goals of this chapter are to reconsider two main approaches: materials engineering and device design to boost the performance of the gen2 SynTs (see Figure V. 1).

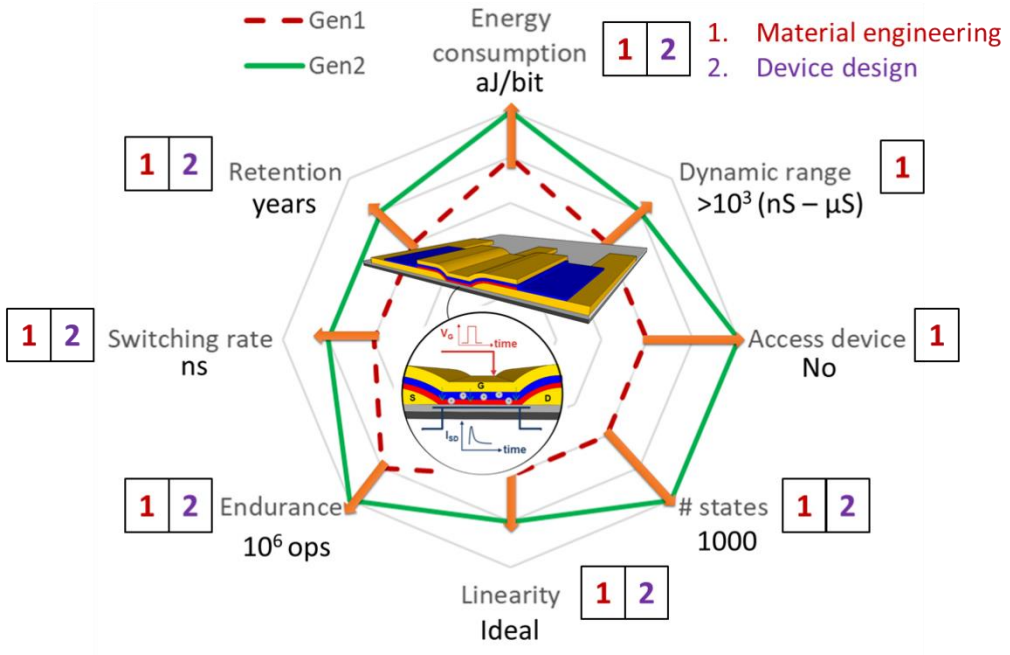

Figure V. 1: The estimated improvement of performance metrics for the second generation SynT compared to that of the first generation.

It is noteworthy that some figures of merits are heavily correlated in the sense that the improvement of one metric may result in the diminution of another one. For example, while thinning down the channel layer can reduce the write time and energy, it has a counter effect on the endurance property of the film under ionic intercalation reactions. Therefore, the essential work is to consider the influence of different solutions on the device's performance and then find an optimized trade-off among them.

1.2 Materials engineering

A wide range of materials can be used to construct electrochemical transistors whose functionalities are directly dictated by the redox reactions, e.g., organic, polymer, two-dimensional (2D), and metal-oxide materials. Polymer and organic materials certainly have limitations regarding on-chip integration and operation. Therefore, to align ourselves with the roadmap of the neuromorphic chip industry, we will narrow our search to metal oxide materials, which are more adapted to the current CMOS technologies.

1.2.1 Channel materials

The channel layer is one of the critical blocks of SynTs, which governs the performance of the devices in the areas of switching rate, linearity, dynamic range, and the number of states. For this reason, the materials have to be carefully chosen with the suggested properties:

(i) Exhibiting insulator-metal-transition induced by redox reactions. This phenomenon is often related to a structural phase transition. Many Li-based materials such as Li_xWO_3 [1], $LiCoO_2$ [2], $LiNbO_x$ [3] are demonstrated to have 3 – 5 orders of magnitude change in terms of resistivity upon controllable ionic intercalation, which are promising for multiple-state, high dynamic range transistors. A sufficiently high dynamic range is desired for the operation of the crosspoint memory to accommodate multiple states while tolerating write noise. For power consumption purposes, the conductance range should reside within nS to µS.

(ii) Illustrating high ionic diffusivity. This has two consequences, the first of which allows ions to be rapidly doped "vertically" into the channel via the electrolyte layer via gate pulses. Thus, the accumulation of ions at the interface, which results in an asymmetric change of the channel conductance, can be avoided. Secondly, the high ionic activity of the active material facilitates the horizontal redistribution of doped ions inside the channel matrix, which will significantly reduce the time required for state rest-and-read duration after writing pulses.

(iii) Displaying a nonlinear (ideally exponential) conductance modulation response to the gate sweep. Such characteristics will allow us to integrate these SynTs into crossbars in a selector-free manner. The gate terminals in the same row may also be directly connected to each other and used for update operations without cross-talk. By applying voltage pulses with opposite polarities at matching gate lines and columns using the half-voltage selection scheme, parallel and sequential programming are facilitated with minimal disturbance at the unselected device using the half-select update scheme [4].

With the given suggestions, ultrathin crystalline-phase channel materials are more favorable than the amorphous phase with demonstrated phase transitions. Furthermore, the redox reactions of crystalline phase intercalation materials are triggered at a specific value of gate potential, which means below that potential, the ionic exchange is negligible. Knowing that the conductance change is proportional to the intercalation charge, crystalline materials support point (iii). On point (ii), the crystalline materials can be engineered to have a sub-10nm thickness, which activates the extrinsic pseudocapacitive behaviors and allows fast ion intercalation time by reducing the diffusion length [5]–[8]. However, significant decreases (to below 5 nm) would increase the effects of the electrode/electrolyte interface, requiring tighter control of parasitic interfacial chemical and electrochemical reactions to avoid detrimental device performance. Therefore, in the next generations, we need to employ 5 – 10 nm channel layers made of prelithiated materials (Li stoichiometry is intrinsic to deposition process) such as Li_xWO_3, $LiCoO_2$, and $LiNbO_x$.

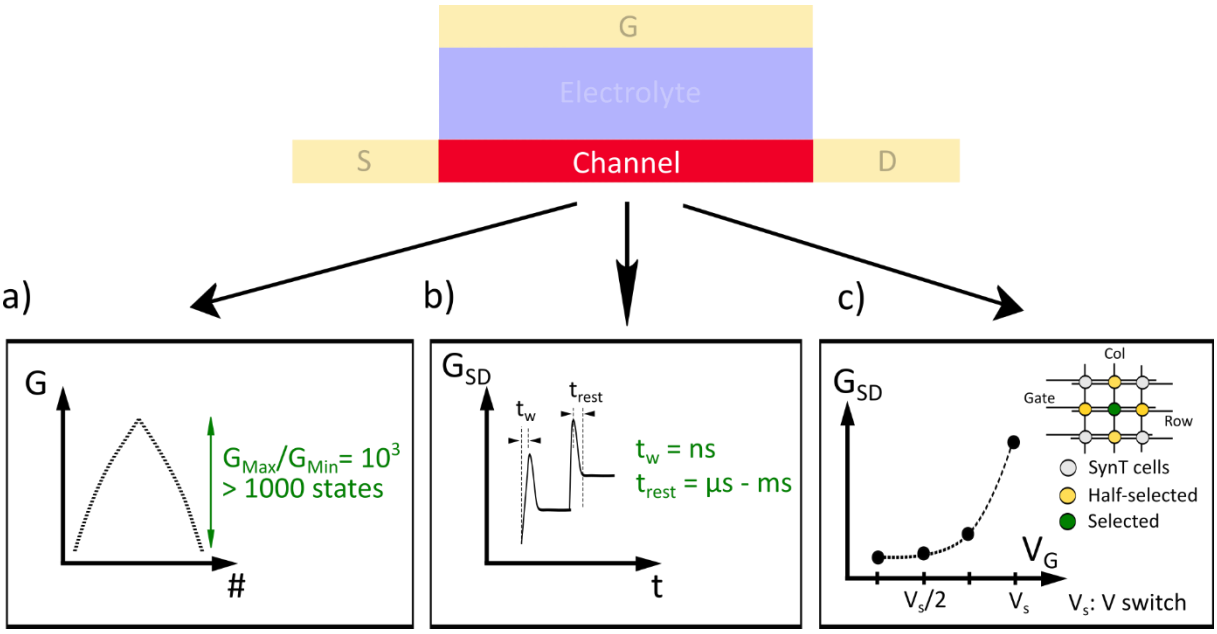

Figure V. 2: The choice of channel materials can affect a) The dynamic range and number of conductance states, b) The switching rate including writing and resting time, and c) The potential to obtain selector-less SynTs.

1.2.2 Electrolyte materials

The solid electrolyte layer is responsible for the conduction of mobile ions to (or from) the channel and not electrons. For this reason, it is desirable for this layer to have a good ionic conductivity and a high electronic resistivity. While high ionic conductivity of

electrolytes can contribute directly to the write time and energy, high electronic resistivity allows SynTs to maintain the programmed states with minimal leakage current.

There are different methods to obtain better ionic conductivity in solid-state electrolytes. First, we can dope the metal oxide electrolyte and generate extrinsic mobile Li ions so that the activation is determined by the migration energy alone (i.e., $E_A = E_m$) [9]. The ionic conductivity is related to the activation energy by the equation:

$$\sigma(T) = A.exp[-E_a/(kT)] \quad (1)$$

where T is temperature, A is a pre-exponential factor, E_a is the activation energy, and k is the Boltzmann constant.

Thus, by decreasing the activation energy, we increase the ionic conductivity of the layer. Furthermore, elemental doping can cause disordered Li-ion sublattice that provides more Li interstitial sites for easy Li ion migration [9]. With determined conductivity, a common method to obtain better ionic conductance is to scale down the layer's thickness. The ionic conductance is calculated following the equation $G_{ion} = (\sigma \times S)/d$, where G_{ion} is the ionic conductance, d is electrolyte thickness, and S is the device area. With this equation, the ionic conductance of an electrolyte layer will increase linearly with the decrease in thickness. This miniaturization is accelerated by the development of advanced deposition techniques, such as pulsed laser deposition (PLD) and atomic layer deposition (ALD), allowing the deposition of ultrathin electrolytes with a thickness less than 20 nm.

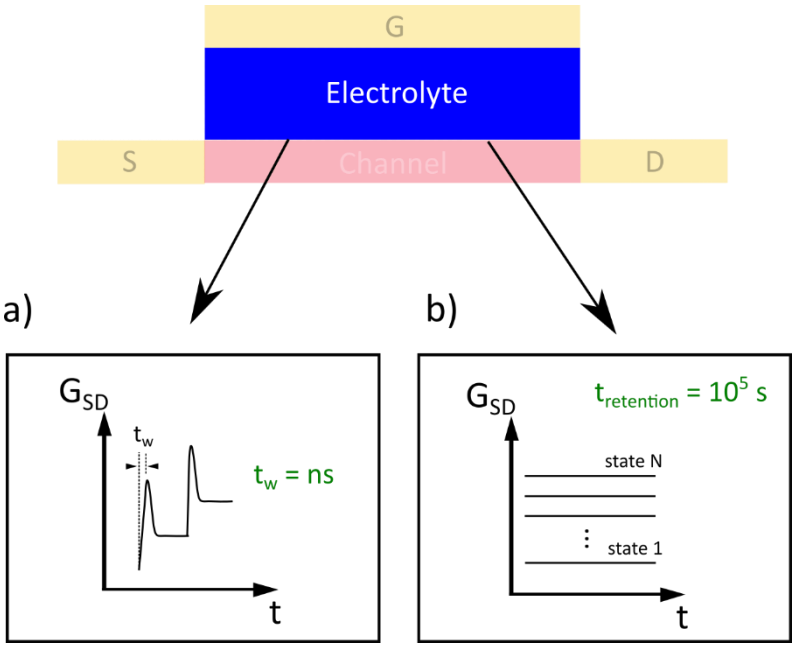

Figure V. 3: The electrolyte can have direct impacts on a) Write duration with their ionic conductance and b) State retention with their electronic resistance.

However, we cannot thin down the layer too much, due to constraints from electronic conductivity, which also has a linear dependence on the thickness of the layer, and the mechanical and electrochemical stability. A wide stability window allows SynTs to be operated in a wide range of gate voltage, duration, or in current-controlled mode (voltage is blindly applied). Therefore, a study to optimize the thickness of the electrolyte that can maintain the integrity of the layer while maximizing the ionic activity is remained to be done.

1.3 Device design

Apart from the intrinsic properties of the gate-stack materials, there are several performance merits of SynTs that can be improved from the device design point of view, for example, switching rate and energy, operation voltage, endurance, and retention.

1.3.1 Thin-film stacking and overlay

Thin-film sandwiching is an effective method to decrease the migration path of the ions vertically to/from the channel, which directly affects the write speed. Indeed, by designing SynTs with different thin-film layers, the migration and diffusion length required for the ions are in the order of the film thicknesses (~nm), which is normally several orders of magnitude higher than that of the coplanar devices, which is limited by the lithography techniques (~µm). Another point on the design that can help to improve the switching rate is to maximizing the overlaying area of the gate electrode and the channel gap. A device with this configuration allows faster measurement of the conductance change by reducing the time required for the ions to diffuse horizontally within the channel matrix (corresponding to the rest time after each write pulse). Otherwise, in the designs where the gate area covers partly the channel gap, the application of write voltage induces ionic intercalation locally on a part of the channel, which requires a period for the ions or holes propagate to the other part before stable states could be measured.

1.3.2 Dimension scaling

Device downscaling is referred to as a key solution to improve the performance of the electrochemical transistors in terms of energy and time efficiency. As previously mentioned, the thickness of the active channel and the electrolyte can be thinned down to facilitate fast ion migration and diffusion. On the other hand, scaling down the device area

(the vertically overlapped area between the gate electrode and the channel) can help to decrease the energy consumed per write operation. Indeed, the charge involved in the conductance switching (ΔQ) has been shown to scale approximately with the volume of the active material channel in SynTs [10]–[12]. For this reason, an energy gain of two orders is achievable for write operations by reducing the gate area to less than 1 μm^2 and the channel thickness to 5 – 10 nm. The gap between the source and drain electrode is also worth downscaling to 200 – 500 nm as it would allow the measurement of highly insulating channel materials under the effect of ionic intercalation. Nevertheless, the process to open a 200 nm gap could be time-consuming to realize with common photolithography.

1.3.3 Gate stack design

From the first generation of SynT, we observe that the working voltage range depends heavily on the potential difference between the gate and the channel electrodes. Such a potential difference comes from the Li stoichiometry of the gate stack. In the next steps of SynT development, we suggest creating a symmetrical gate stack by adding a reservoir layer of the same material family as the channel between the electrolyte and the gate electrode. This layer is first beneficial to reduce the initial potential difference of the cell, thus, reducing the energy spent for each writing operation. Furthermore, this film may allow us to introduce a supplementary amount of mobile ions to the system, which potentially affects the linearity and the total number of programmable states of the channel [12]. However, it is apparent that introducing another layer to the process flow requires considerable effort in terms of integration.

Another technological brick that we did not spend time developing in the first functional devices is the passivation step. Passivation or encapsulation is usually a shielding layer made of metal oxide or polymer passivation placed at the end of the process to the device from environmental factors such as temperature, humidity, and air exposure [13]–[15]. This step is considered to be critical for some technologies, especially those that involve air-sensitive Li ions and Li-based compounds [16]–[18]. For the gen1 SynTs, the lack of passivation led to the degradation of the LiPON electrolyte and Li ion oxidation upon air contact. With a new protecting layer, gen2 SynTs will potentially improve their state retention and endurance.

1.4 Summary on the approaches

In general, different figures of merit could be enhanced using mainly two approaches: materials engineering by tuning the thickness and specific associated properties and device design with extra layers and structure (see Figure V. 4). The

suggestions on different levels can be used as a guideline to study and develop next generations of SynTs.

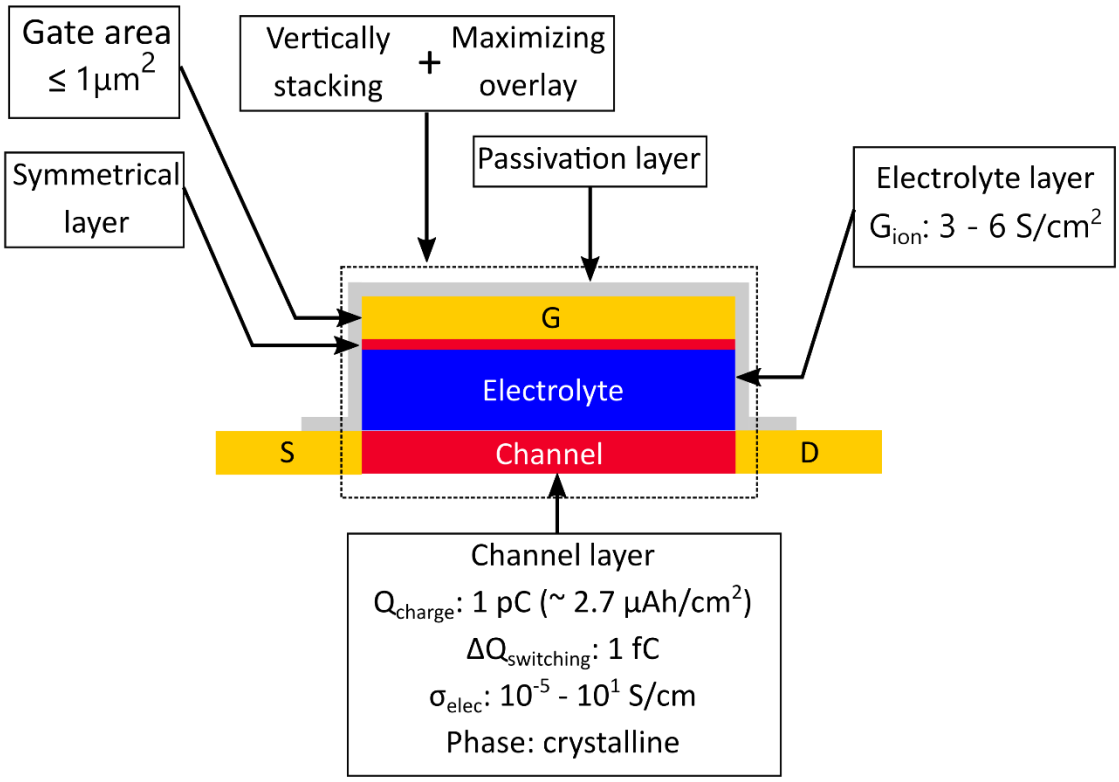

Figure V. 4: Suggestions on the properties of the thin films and the structure of the next SynTs.

2 CURRENT WORK TOWARDS A SECOND GENERATION OF SYNTS

We have initiated several works towards a new generation of SynT (gen2). These works include the design of a new set of masks with multiple transistors and test structures, and efforts from the lithography side to realize a 200 nm channel gap. In addition, we have conducted some material developments to this purpose. A brief summary of these exercises will be included in this section.

2.1 New design

We have designed and fabricated a set of 9" chrome masks (Figure V. 5.a) with five main layers: bottom electrodes, active channel, electrolyte, top electrode, and passivation layer. Following the guideline in the previous section, the new SynTs are designed to have small gate areas (varying with A = 100 to 400 μm^2) and maximized gate and channel overlap (marked as (1) in Figure V. 5.a). A closer view of a SynT is shown in Figure V. 5.b. The new SynTs come in variable dimensions (device area and channel gap), which allow us to study the correlation between transistor size and performance merits statistically. An encapsulation layer is designed to cover sufficiently the region of interest near the gate stack. Such a passivation does not require additional redistribution layer of metal, reducing the lithography steps to finalize the functional devices. Besides, in each die, we also place different test structures, such as one-gate-multiple-channel transistors (2) and battery-like structure (3). In general, these devices allow us to elucidate the electrical contribution from the channel geometry (length and width) and the electrochemical properties of the electrolyte.

One of the interesting features of the gen2 SynT is that they might have a much smaller channel gap (smallest designed L = 200 nm) compared to the gen1, which could be defined by photolithography with an ASML ASM300 DUV (deep ultra-violet) stepper. In the first attempts, we were able to realize bottom electrodes with open gaps from 300 nm to 5 µm, which are sufficiently good compared to the reported SynTs. Even though the lithography of the gaps of 200 nm is theoretically possible, etching and opening the gaps completely without metal residues require extra engineering efforts (see Figure V. 6).

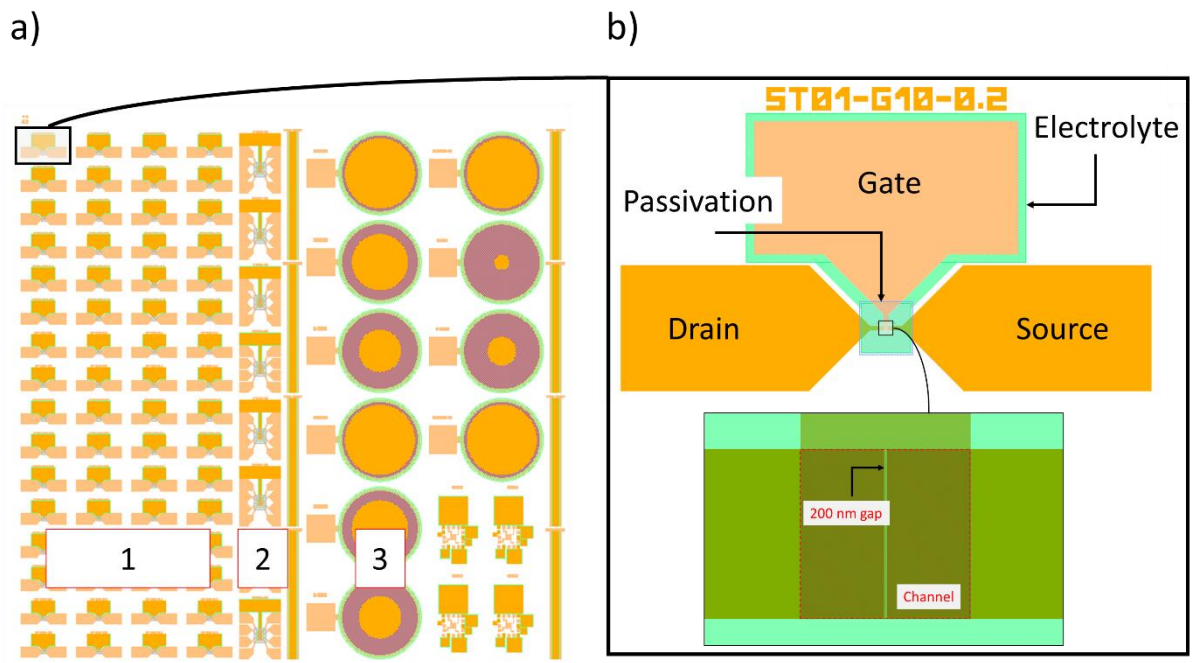

Figure V. 5: a) A die containing different test structures: SynTs (1), (2), and battery-like structures (3). b) A zoom into a SynT device.

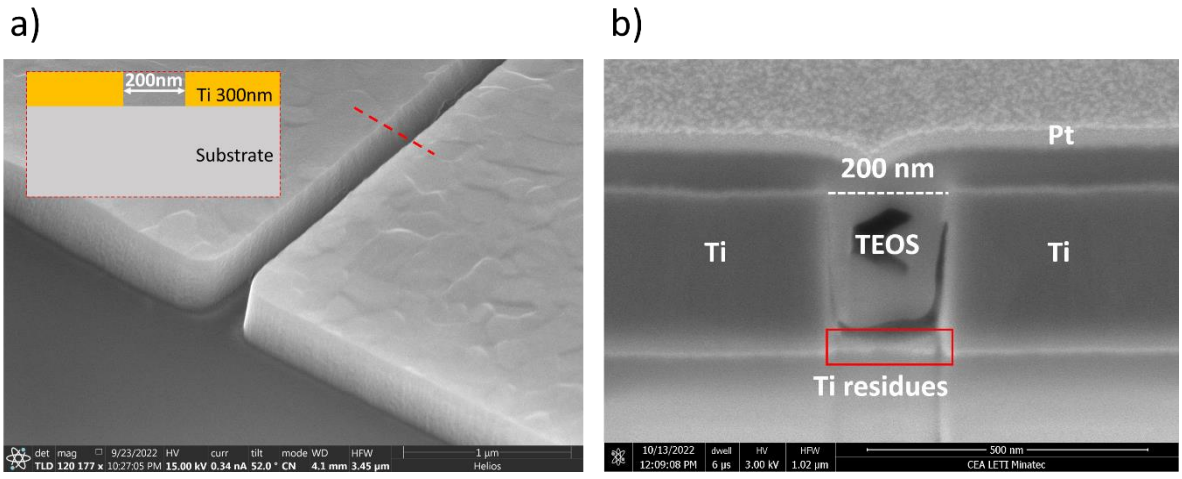

Figure V. 6: a) SEM observation of bottom electrodes belonging to a SynT device with the channel gap of 200 nm after patterning (inset: schematic illustration of the cross-section at the dashed line). b) FIB/SEM observation with the focus on the gap of the electrodes. The red rectangle highlight the Ti residue after etching indicating a failure to open this device. (TEOS - Tetraethyl orthosilicate and Pt films were coated to observe the gap)

2.2 Materials

In parallel with lithography flow development, we have been seeking new materials that are compatible with our current process and have interesting ionic and electronic properties. The first one we need to mention is the ultrathin LiPON with a thickness of around 20 nm, which has been deposited by ALD with LiHMDS and DEPA precursors using Picosun R200. A TEM observation of a MIM stack composed of TiN/LiPON 20nm/Ti is shown in Figure V. 7.a. With the electrochemical impedance spectroscopy (EIS) measurement performed on the MIM structure, we can confirm the ionic conductivity is σ = 0.6 µS/cm (Figure V. 7.b), similar to that of sputtering LiPON. Such an ultrathin layer of electrolyte will enhance the write speed with its high ionic conductance 10 times compared to the gen1 SynT. However, more work needs to be done to master the deposition process and optimize the ionic properties of the ALD LiPON.

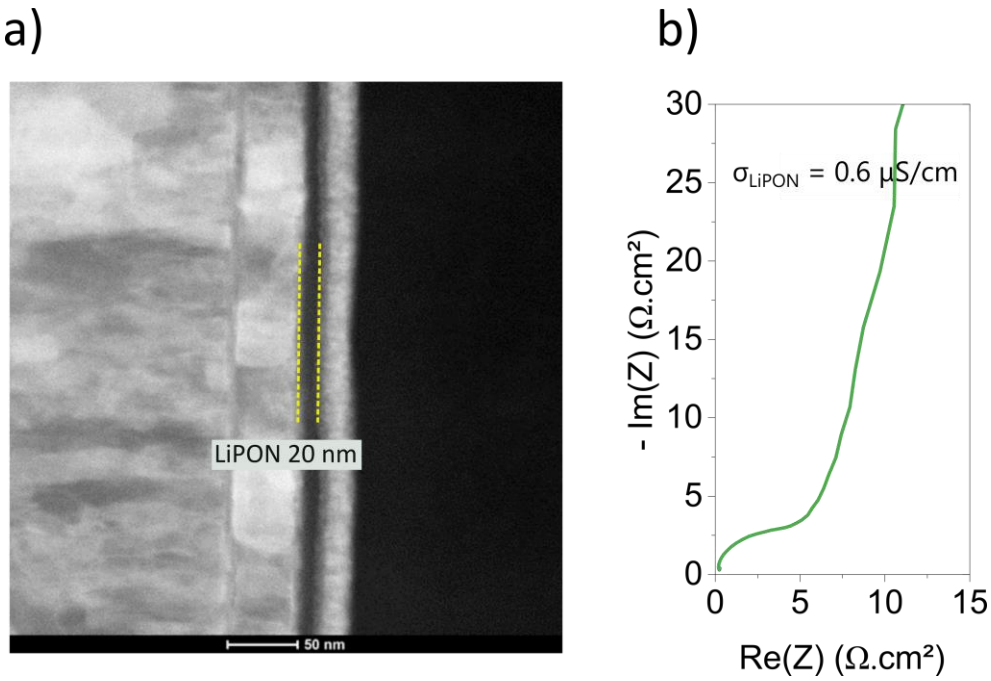

Figure V. 7: a) TEM observation of a MIM structure composed of TiN/LiPON 20nm/Ti with the LiPON film deposited by ALD. b) The EIS measurement on the MIM structure indicates the ionic conductivity of the ALD LiPON to be σ = 0.6 µS/cm.

Apart from the LiPON electrolyte, we have initiated the study of ultrathin lithium niobium oxide ($LiNbO_3$) film prepared by PLD as an active material for the synaptic transistor. In the literature, the Li-Nb-O material family is reported to have conducting, semiconducting, and insulating properties across a wide resistivity and bandgap range, depending on the oxidation state of Nb (from metal Nb to $LiNbO_3$ insulator) [3]. The goal of the exercise is to induce the reduction of Nb with Li ions and make use of the IMT of

the channel material. In the first experiments, we successfully integrated the $LiNbO_3$ layer as a channel into the current process flow. The patterning step was realized using plasma etching, similar to that previously developed etching recipe of Li_xTiO_2. We were able to characterize the SynT gate stack ($LiNbO_3$ 10 nm/LiPON 200 nm) using the battery-like test structure (see Figure V. 8). The CV performed on such test devices allowed us to locate the redox peak potentials (at 2.0, 2.2 V for oxidation and 1.4, 1.85 V for reduction). By integrating the current density over the time in the first Li extraction, we obtain the specific charge capacity extracted from the pristine active layer to be ΔQ_{LiNbO3} = 0.87 µAh/cm^2. This initial amount of charge is twice higher than of Li_xTiO_2 layer in the gen1 SynTs ($\Delta Q_{LixTiO2}$ = 0.42 µAh/cm^2). The first successful electrochemical characterizations of the battery-like structures verify the functionality of the ionic activity within the gate stack. Nevertheless, we did not observe any controllable conductance modulation using the transistor configuration. We hypothesize that the conductance of $LiNbO_3$ with the oxidation state of 5 is too low, leading to undetectable change of channel current. Therefore, more work will be done to optimize the devices material-wise.

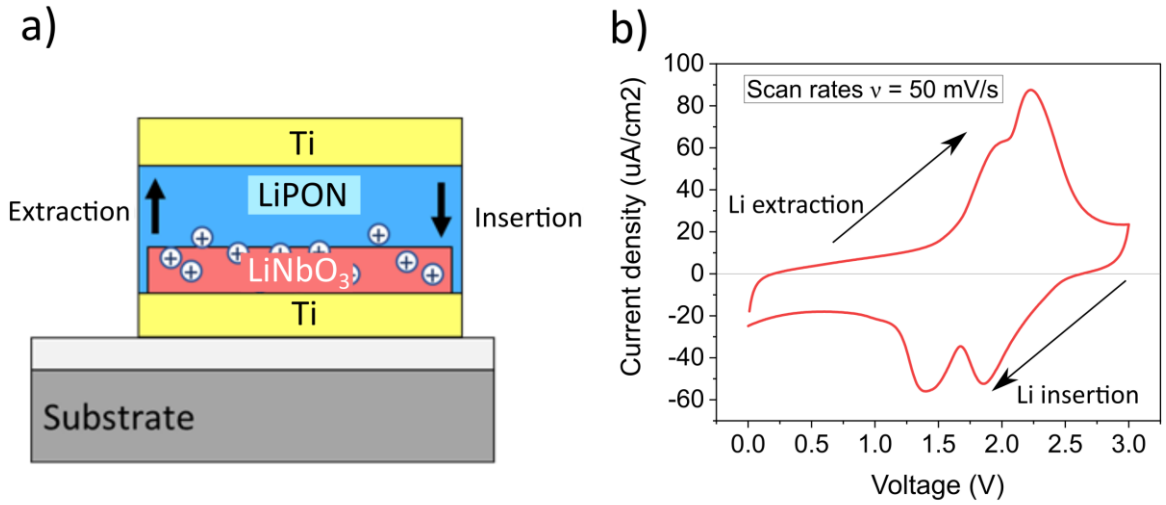

Figure V. 8: Preliminary study of $LiNbO_3$ as an active channel material. a) The battery-like test structure composed of Ti/$LiNbO_3$/LiPON/Ti. b) The voltammogram of the test device confirms the functionality of the gate stack.

3 CONCLUSIONS

In this chapter, we suggested a few ideas to improve the performance of the new generation of SynTs based on our first generation. We may consider two main approaches at the device level: material engineering and device design. While the former focus on selecting and optimizing the ionic/electronic conductivity of the constituting films, the latter aims attention to improving the arrangement of different layers to boost the device performance on top of the intrinsic material properties of the gate stack. To summarize, the switching rate and linearity of the electrochemical transistors depend on the ionic transport activity of the electrolyte and active channel, which can be enhanced by minimizing the migration path of the ions. The dynamic range and the total number of conductance states depend heavily on the amount of available mobile ions, the IMT, and the charge capacity of the channel material. A passivation layer could improve the endurance and retention merits of SynT. Even though downscaling dimensions could be viewed as a general trend to advance the new SynTs, we have to study and quantify the inter-correlation effect among the merits and find a reasonable trade-off.

In the second part of the chapter, we covered some of the optimization work and experiments for the gen2 SynTs, including new mask design, process optimization, and new materials research. The new set of masks consists of multiple SynTs with different parameters, allowing us to systematically study the performance of SynTs with different device parameters, such as gate area and channel gap. Furthermore, other test structures facilitating the decorrelation of the channel and electrolyte contribution are included in each dice. We could realize bottom electrodes made of Ti with channel gaps greater than 300 nm. Concerning the material engineering aspect, we managed to realize 20 nm LiPON with the ALD technique and confirmed its ionic conductivity to be around 0.6 µS/cm. $LiNbO_3$ channel layer was successfully integrated into the current process flow and proven its functionality as an electrochemically active material with cyclic voltammetry technique. However, these layers require further improvement in future experiments and processes.

4 REFERENCES

[1] K. Yoshimatsu, T. Soma, and A. Ohtomo, "Insulator-to-metal transition of WO_3 epitaxial films induced by electrochemical Li-ion intercalation," *Appl. Phys. Express*, vol. 9, no. 7, p. 075802, Jul. 2016, doi: 10.7567/APEX.9.075802.

[2] C. A. Marianetti, G. Kotliar, and G. Ceder, "A first-order Mott transition in LixCoO2," *Nature Mater*, vol. 3, no. 9, pp. 627–631, Sep. 2004, doi: 10.1038/nmat1178.

[3] M. B. Tellekamp, J. C. Shank, and W. A. Doolittle, "Molecular Beam Epitaxy of lithium niobium oxide multifunctional materials," *Journal of Crystal Growth*, vol. 463, pp. 156–161, Apr. 2017, doi: 10.1016/j.jcrysgro.2017.02.020.

[4] T. Gokmen and Y. Vlasov, "Acceleration of Deep Neural Network Training with Resistive Cross-Point Devices: Design Considerations," *Front. Neurosci.*, vol. 10, Jul. 2016, doi: 10.3389/fnins.2016.00333.

[5] J. Ye *et al.*, "Amorphization as a Pathway to Fast Charging Kinetics in Atomic Layer Deposition-Derived Titania Films for Lithium Ion Batteries," *Chem. Mater.*, vol. 30, no. 24, pp. 8871–8882, Dec. 2018, doi: 10.1021/acs.chemmater.8b04002.

[6] S. Moitzheim, S. De Gendt, and P. M. Vereecken, "Investigation of the Li-Ion Insertion Mechanism for Amorphous and Anatase TiO_2 Thin-Films," *J. Electrochem. Soc.*, vol. 166, no. 2, pp. A1–A9, 2019, doi: 10.1149/2.1091816jes.

[7] C. Choi *et al.*, "Achieving high energy density and high power density with pseudocapacitive materials," *Nat Rev Mater*, vol. 5, no. 1, pp. 5–19, Jan. 2020, doi: 10.1038/s41578-019-0142-z.

[8] Y. Liu, S. P. Jiang, and Z. Shao, "Intercalation pseudocapacitance in electrochemical energy storage: recent advances in fundamental understanding and materials development," *Materials Today Advances*, vol. 7, p. 100072, Sep. 2020, doi: 10.1016/j.mtadv.2020.100072.

[9] S. Muy *et al.*, "Tuning mobility and stability of lithium ion conductors based on lattice dynamics," *Energy Environ. Sci.*, vol. 11, no. 4, pp. 850–859, 2018, doi: 10.1039/C7EE03364H.

[10] Y. van de Burgt *et al.*, "A non-volatile organic electrochemical device as a low-voltage artificial synapse for neuromorphic computing," *Nature Mater*, vol. 16, no. 4, pp. 414–418, Apr. 2017, doi: 10.1038/nmat4856.

[11] M. T. Sharbati, Y. Du, J. Torres, N. D. Ardolino, M. Yun, and F. Xiong, "Low-Power, Electrochemically Tunable Graphene Synapses for Neuromorphic Computing," *Adv. Mater.*, vol. 30, no. 36, p. 1802353, Sep. 2018, doi: 10.1002/adma.201802353.

[12] J. Tang *et al.*, "ECRAM as Scalable Synaptic Cell for High-Speed, Low-Power Neuromorphic Computing," in *2018 IEEE International Electron Devices Meeting (IEDM)*, San Francisco, CA, Dec. 2018, p. 13.1.1-13.1.4. doi: 10.1109/IEDM.2018.8614551.

[13] C. Yang *et al.*, "All-Solid-State Synaptic Transistor with Ultralow Conductance for Neuromorphic Computing," *Adv. Funct. Mater.*, vol. 28, no. 42, p. 1804170, Oct. 2018, doi: 10.1002/adfm.201804170.

[14] Z. Xie *et al.*, "All-Solid-State Vertical Three-Terminal N-Type Organic Synaptic Devices for Neuromorphic Computing," *Adv Funct Materials*, vol. 32, no. 21, p. 2107314, May 2022, doi: 10.1002/adfm.202107314.

[15] A. Melianas *et al.*, "High-Speed Ionic Synaptic Memory Based on 2D Titanium Carbide MXene," *Adv Funct Materials*, vol. 32, no. 12, p. 2109970, Mar. 2022, doi: 10.1002/adfm.202109970.

[16] D.-J. Yun *et al.*, "Chemical/morphological transition behavior of lithium phosphorus oxynitride solid-electrolyte in air: An analytical approach based on X-ray photoelectron spectroscopy and atomic force microscopy," *Journal of Power Sources*, vol. 399, pp. 231–237, Sep. 2018, doi: 10.1016/j.jpowsour.2018.07.081.
[17] Y. Wang *et al.*, "Inhibitory property of lithium phosphorus oxynitride surface grown by atomic layer deposition," *Surfaces and Interfaces*, vol. 33, p. 102280, Oct. 2022, doi: 10.1016/j.surfin.2022.102280.
[18] C. Busà, M. Belekoukia, and M. J. Loveridge, "The effects of ambient storage conditions on the structural and electrochemical properties of NMC-811 cathodes for Li-ion batteries," *Electrochimica Acta*, vol. 366, p. 137358, Jan. 2021, doi: 10.1016/j.electacta.2020.137358.
[19] Y. Zhang, L. Yang, W. Wang, and G. Wang, "Nitrogen-doped carbon nanotube in-situ loaded LiNbO3 anode with high capacitance contribution for lithium-ion capacitors," *Electrochimica Acta*, vol. 435, p. 141354, Dec. 2022, doi: 10.1016/j.electacta.2022.141354.

CONCLUSIONS AND PERSPECTIVES

Neuromorphic computing is one of the solutions to address the computing bottleneck (e.g., shared data bus between memory and processing units) created by the conventional Von Neumann architecture, paving the way to handling data-intensive jobs in an energy- and time-efficient manner. Many technologies have been suggested to construct the synaptic hardware for such brain-inspired systems, namely two-terminal devices (ReRAM, PCM, and MRAM) and three-terminal devices (FeFET and SynT). Each has both advantages and drawbacks when implemented as artificial synapses. Researchers have demonstrated nanoscale SynT devices to have an excellent overall performance with high programming precision, ultralow operation power, linear analog-state modulation, and potentially low device variations with a robust working mechanism. However, research efforts were essential to improve several performance merits of SynTs (i.e., programming speed, state retention, and ON/OFF ratio) to suit the neural network applications better. The state-of-the-art SynTs and the approaches to go beyond were reported in chapter 1.

To advance the performance of SynTs, the goals of this thesis were: (i) Elaborating and characterizing solid-state Li-based electrochemical synaptic transistors with CMOS-compatible processes. (ii) Demonstrating their required synaptic functionalities owing to Li-ion intercalation leading to rapid and ultralow power analog switching, showing possible applicability in neural network computing systems based on the performance of functional devices. (iii) Proposing innovative solid-state gate stacks and designs to optimize the overall performance and further upscaling.

As a first step, we showed the possibility of elaborating nanoscale synaptic transistors using CMOS-compatible microfabrication processes such as material deposition (sputtering and atomic layer deposition), photolithography with photomask, and thin-film etching using either reactive ions or chemical etchants. Subsequently, different physical characterizing methods, including SEM, TEM, EDS, and Raman spectroscopy, were covered as beneficial ways to obtain the material properties of deposited layers. The physical understanding of the gate stacks allows us to interpret the performance of the devices more precisely. The process flow to fabricate our electrochemical synaptic transistors with details on the steps was described. We revealed especially the encountered problems while realizing this generation of devices and the progressive optimization process. The first functional SynTs have been elaborated using the described microfabrication steps.

Afterward, we presented the interesting electrical and electrochemical properties of building materials of the first generation of the electrochemical synaptic transistor ($LiCoO_2$ and Li_xTiO_2 channels and LiPON electrolyte) in chapter 3. The stability of these materials facilitates the microfabrication of wafer-scale, BeOL-compatible synaptic transistors. The preliminary results of SynTs with $LiCoO_2$/LiPON gate stack were presented as a proof-of-concept of the wafer-scale elaboration synaptic transistors. The electrochemical tests such

as electrochemical impedance spectroscopy (EIS) and cyclic voltammetry (CV) on the battery-like structures allowed studying the contribution of LiPON electrolyte and $LiCoO_2$ active material. The important insulator-metal-transition of the HT-$LiCoO_2$ channel was confirmed by measuring the channel current while extracting Li ions with a gate voltage. Upon the re-intercalation of ions, the conductance decreased to the initial state. A programming cycle with 40 conductance states was demonstrated, proving the required functions for an artificial synaptic device.

Then, a comprehensive study on SynTs with the Li_xTiO_2/LiPON gate stack was conducted to improve the performance of the previous material composition. The cross-section of SynTs was studied with SEM/EDS and TEM, providing information on the physical and chemical properties of the gate stack. In addition, the SynT exhibited good merits of an artificial electrochemical synapse, such as fast programming, reversible conductance modulation, retention, linearity, endurance, and slight device-to-device variation. This transistor was highly efficient in energy consumption for both write ($fJ/\mu m^2$) and read (nS) operations. A systematic study using a two-terminal device was further performed to highlighted the pseudocapacitive behavior of the ultra-thin Li_xTiO_2 film; making it an appropriate channel material for SynTs used for high-speed, low-power synapses.

We then investigated the use cases of the first generation of SynTs for different types of neural networks: ANN and SNN, in chapter 4. First, we use the experimental results of SynTs to simulate computing-in-memory hardware to accelerate the vector-matrix multiplication operations of ANNs. With the CrossSim simulator platform, the realistic nonlinearity of the SynT was taken into account to practically benchmark its performance on handwritten datasets with other available technologies. Our SynT was also used to simulate the associative memory with an all-analog spiking neural network circuit using LT-Spice. We proposed a design of a neural network consisting of two synapses, including a resistor and a compact model of SynT, and a leaky integrate-and-fire neuron. We successfully trained the circuit to react to neutral stimuli by firing potential spikes out of the neuron. With the two preliminary examples, we showed that the SynTs could be employed in both types of neural networks.

In the future phases of this work, we intend to focus on optimizing further the performance of this SynT (i.e., the programming rate, the dynamic range, the retention and number of states, etc.) based on two main approaches: (i) Materials engineering, and (II) Device design (some ideas have already been proposed in chapter5). Concerning the materials, we will improve specific properties of the existing layers, such as ionic conductivity and voltage range stability of the electrolyte, ionic diffusivity, and the available mobile ions of the channel materials by thermal or deposition processes. Significantly, tuning the phases of channel material affects the programming voltage and the switching linearity, potentially allowing an area- and energy-efficient selector-free artificial synapse for large crossbar arrays. In the design of SynTs, we aim to reduce the programming energy per write operation by reducing the active area size and increasing the overlap between the gate and the bottom electrodes. A passivation layer on top of the gate can be designed

to prevent ion loss from oxidation upon air exposure, thus increasing both the retention and endurance of the devices.

RÉSUMÉ EN FRANÇAIS

Introduction

L'architecture classique de Von Neumann joue un rôle essentiel dans la résolution de problèmes de différents niveaux de complexité dans presque tous les domaines de la vie. Toutefois, cette architecture informatique présente un goulot d'étranglement : une quantité importante de temps et d'énergie est nécessaire pour transmettre les données entre les processeurs et la mémoire. Cet obstacle limite inévitablement l'efficacité des calculs, en particulier pour la résolution de tâches complexes telles que la reconnaissance des formes ou de sons, par exemple. S'inspirant du cerveau humain, les systèmes informatiques neuromorphiques devraient pouvoir surmonter cette limite en effectuant les calculs de manière massivement parallèle.

Le développement de tels systèmes neuromorphiques nécessite cependant des synapses artificielles qui imitent le comportement de leurs homologues biologiques. Des efforts ont été réalisés pour simuler les fonctions synaptiques avec des circuits analogiques CMOS. Ils se heurtent cependant à un problème majeur de miniaturisation (des dizaines de composants sont nécessaires pour imiter une seule synapse). De ce fait, de nombreuses recherches se concentrent actuellement sur des dispositifs spécifiques (memristors), facilement miniaturisables, dont la conductance électrique peut être modulée, pour émuler l'évolution des connexions synaptiques biologiques. Ces composants peuvent être subdivisés en dispositifs à 2 terminaux et dispositifs à 3 terminaux. Chaque configuration a ses propres forces et faiblesses. En particulier, dans la configuration à 3 terminaux, l'opération d'écriture (modulation du poids synaptique) est découplée de l'opération de lecture, ce qui permet un meilleur contrôle de la conductance.

Parmi les synapses artificielles à 3 terminaux, les transistors ioniques apparaissent comme de bons candidats potentiels. Leur fonctionnement repose sur un empilement {canal/conducteur ionique} qui permet d'injecter/extraire des ions (via le conducteur ionique) dans la partie active du transistor (le canal), et de moduler ainsi finement la conductance électrique du composant.

L'intégration à grande échelle de tels dispositifs synaptiques est indispensable pour développer l'architecture l'informatique neuromorphique. Des dispositifs composés d'électrolytes liquides et polymères, et de canaux exfoliés manuellement, ne permettent malheureusement pas leur intégration ultérieure dans des puces informatiques de grande densité. Pour surmonter ces limitations, les objectifs de ma thèse sont (i) de proposer des transistors synaptiques innovants qui permettent d'optimiser leurs performances globales (ii) d'élaborer ces transistors synaptiques électrochimiques avec des procédés compatibles CMOS, et (iii) de démontrer leurs fonctionnalités synaptiques. Les travaux de cette thèse ont été réalisés grâce à une collaboration entre le Laboratoire Composants pour la RF et

l'Energie (LCRE) au CEA-LETI àGrenoble, et le Laboratoire Génie Électrique et Électronique de Paris (GeePs), à Gif sur Yvette.

Chapitre 1

Dans le chapitre 1, nous avons souligné la nécessité de développer un nouveau paradigme informatique pour remédier au problème posé par l'architecture informatique conventionnelle de Von Neumann. S'inspirant du cerveau humain, les systèmes informatiques neuromorphiques ont un grand potentiel pour réaliser des tâches complexes de manière plus efficace en termes d'énergie et de temps.

Plusieurs principes physiques et technologies différents peuvent permettre de réaliser des dispositifs électroniques qui émulent des synapses artificielles pour l'informatique neuromorphique. Plusieurs technologies sont décrites dans ce chapitre, notamment les dispositifs à deux terminaux (ReRAM, PCM et MRAM) et les dispositifs à trois terminaux (FeFET et SynT). Chacune de ces technologies présente des avantages et des inconvénients. Les dispositifs à deux terminaux sont en général caractérisés par une grande rapidité, une endurance élevée, une grande gamme dynamique et une bonne stabilité. Cependant, ils présentent également un comportement non linéaire et une consommation par opération (Ecriture/Lecture) élevée, qui peut limiter la précision (dans la phase d'apprentissage) et l'efficacité énergétique dans les réseaux neuronaux à grande échelle.

Parmi les dispositifs à trois terminaux, les synapses de type FeFET permettent des opérations de programmation rapides, et possèdent une grande stabilité. Néanmoins, ce type de dispositif souffre des mêmes problèmes de miniaturisation que la DRAM et les mémoires à grille flottante (car il s'agit essentiellement de mémoires à base de charges). En outre, certaines fonctions synaptiques sont difficilement réalisables sur de tels composants. Les dispositifs de type SynT (transistors ioniques), eux, ont une consommation énergétique très faible, permettent une modulation précise et linéaire des états de conductance, et des variations potentiellement faibles d'un composant à l'autre. Néanmoins, la faible vitesse de programmation, la stabilité médiocre des états et le rapport ON/OFF faible sont des caractéristiques à améliorer pour que les SynTs deviennent plus attrayants pour des applications aux réseaux neuronaux.

Un état de l'art sur les technologies de réalisation des SynTs est dressé, en mettant l'accent sur plusieurs paramètres importants (figures de mérite) tels que la modulation de conductance (intervalle et nombre d'états possible sur cet intervalle), la linéarité, la stabilité, la consommation d'énergie, l'endurance et la dynamique temporelle. Afin d'optimiser davantage les performances des SynTs, plusieurs approches ont été suggérées dans la littérature, notamment la conception de nouveaux matériaux, la diminution des dimensions, et la mise en œuvre de processus d'élaboration qui soient compatibles CMOS.

Chapitre 2

Dans ce chapitre, nous avons présenté diverses techniques de microfabrication, notamment celles utilisées dans ma thèse pour élaborer nos transistors (pulvérisation cathodique et ALD). En particulier, les processus de dépôt des couches minces de LiPON par PVD et de TiO_2 par ALD sont détaillés. Nous avons également abordé les méthodes de photolithographie et de gravure des matériaux. La résolution des motifs réalisés sur la résine photosensible dépend notamment de l'épaisseur de la résine, du temps d'exposition, et de la longueur d'onde de la source lumineuse. Les motifs finaux des dispositifs sont dictés par les méthodes de gravure utilisées.

Des techniques de caractérisations physiques (MEB : microscope électronique à balayage, et MET : microscope électronique à transmission) ont été utilisées dans ma thèse pour contrôler la microstructure, la croissance des films minces et leurs épaisseurs, tandis que la méthode EDS-EDX et la spectroscopie Raman ont été particulièrement utiles pour obtenir la composition chimique des couches déposées. Le principe de fonctionnement de ces techniques a été illustré au travers d'exemples de la littérature, et de mon propre travail. Ces techniques ont permis d'identifier les défauts et vérifier toutes les étapes de microfabrication.

Nous avons rencontré de nombreux problèmes lors de la réalisation de la première génération de dispositifs SynTs. À chaque étape de la microfabrication, divers problèmes se sont posés (taux de développement trop faible de la résine, couches sur-gravées, résolution latérale du canal insuffisante, …). Résoudre ces difficultés a nécessité des efforts dans plusieurs directions, notamment l'ajustement de l'écart d'exposition, la surveillance du temps de formation du motif et le contrôle de la température, ainsi qu'une nouvelle conception du masque photolithographique. L'optimisation progressive de tous ces paramètres a permis d'obtenir les premiers SynTs fonctionnels.

Chapitre 3

Dans ce chapitre, nous avons d'abord détaillé les matériaux utilisés pour l'élaboration des transistors synaptiques électrochimiques (canal en $LiCoO_2$ ou Li_xTiO_2, et électrolyte en LiPON). L'obtention de dispositifs SynTs avec un empilement {$LiCoO_2$/LiPON} a d'abord fourni une preuve de concept de l'élaboration de transistors synaptiques à l'échelle du wafer. Des tests électrochimiques tels que la spectroscopie d'impédance électrochimique (EIS) et la voltampérométrie cyclique (CV) sur les structures ont permis d'identifier la contribution de l'électrolyte LiPON (conductivité ionique et fréquence caractéristique) et du matériau actif (HT-$LiCoO_2$). L'importante transition métal-isolant du canal HT-$LiCoO_2$ a été confirmée en mesurant le courant du canal lors de l'extraction d'ions Li avec une tension de grille de +4,2 V. Lors de la ré-intercalation des ions, la conductance a diminué jusqu'à l'état initial avec une pente différente, créant un hystérésis typique des SynTs. Un train d'impulsions a permis de modifier la conductance de manière fine (40 états

de conductance par cycle), démontrant les fonctions requises pour un dispositif synaptique artificiel. Cependant, la tension de programmation et la conductance du canal étaient considérablement élevées.

Une étude complète sur des SynTs avec empilement {Li_xTiO_2/LiPON} a ensuite été menée afin d'améliorer les performances obtenues précédemment. La section transversale de ces SynTs a examinée par MEB, EDS-EDX et MET, fournissant des informations utiles sur les dimensions physiques, les éléments, et la phase (quasi-amorphe) du matériau constituant le canal. Lors des tests électriques, ces SynTs ont permis d'obtenir de très bonnes figures de mérite (programmation rapide, modulation réversible de la conductance, stabilité, linéarité, endurance et faible variation d'un dispositif à l'autre). Ces transistors se sont notamment montrés très efficaces en termes de consommation énergétique (fJ/µm2) pour les opérations d'écriture/lecture. Une étude électrochimique systématique (utilisant un dispositif à deux terminaux représentant l'empilement de grille du SynT) a été réalisée pour pouvoir se focaliser uniquement sur les propriétés électrochimiques du canal Li_xTiO_2 et ses performances électriques. Cette étude a mis en évidence le comportement pseudo-capacitif du film ultra-mince de Li_xTiO_2, ce qui en fait un matériau très utile pour des systèmes neuromorphiques très rapides et à faible consommation énergétique.

Chapitre 4

Dans ce chapitre, nous avons étudié l'applicabilité potentielle de la première génération de SynTs à différents types de réseaux neuronaux (ANNs et SNNs). En ce qui concerne les réseaux de neurones artificiels (ANNs), la simulation de matrices de transistors synaptiques a été réalisée à l'aide de la plateforme de simulation «CrossSim». A partir des résultats expérimentaux (100 cycles de modulation de conductance), le simulateur a pu prendre en compte une non-linéarité réaliste du SynT et évaluer ses performances sur des bases de données manuscrites (MNIST). Malgré la non-linéarité de commutation du SynT (due notamment à la forte activité d'intercalation ou d'extraction d'ions à environ 1 V), la précision de reconnaissance de formes sur ces tâches reste élevée ($\geq$ 95%), comparée à d'autres technologies employant la même plateforme de simulation.

Nous avons également montré l'applicabilité potentielle de nos dispositifs SynTs aux réseaux de neurones à impulsions (SNNs). Pour cela, nous avons étudié une règle d'apprentissage - mémoire associative – en proposant le design d'un circuit entièrement analogique. Dans cette simulation (utilisant LT-SPICE), le réseau neuronal est composé d'un neurone LIF (Integrate-and-Fire) et de deux synapses : l'une de ces synapses est représentée par un modèle compact de nos dispositifs (SynT). La phase d'entrainement de l'expérience du Chien de Pavlov, via les stimuli synchrones "vue de la nourriture" et "son de cloche" fait intervenir l'augmentation progressive de la conductance du SynT (au travers

d'un train d'impulsions de courant d'environ 1nA). A partir de là, le comportement de mémoire associative a été clairement mis en évidence.

Chapitre 5

Dans ce chapitre, nous avons abordé différentes pistes (choix des matériaux et configuration des dispositifs) qui permettraient d'améliorer encore les performances des transistors synaptiques. Les matériaux à rechercher doivent pouvoir répondre à des critères spécifiques (bon transport ionique et faible conductance électronique pour les matériaux d'électrolyte, et changement de conductivité significatif lors de l'intercalation de lithium pour les matériaux de canal). Concernant la configuration des dispositifs, il s'agit notamment de diminuer encore les distances de migration des ions.

Quelques travaux préliminaires ont déjà été réalisés dans ce sens. Nous avons conçu et fabriqué de nouveaux types de masques, dans lesquels la distance source-drain des transistors varie dans l'intervalle [200 nm - 1,5 µm]. De plus, nous avons commencé à étudier des films ultra-minces de LiPON (électrolyte) et de $LiNbO_3$ (canal). Les résultats préliminaires sont prometteurs : ce travail d'optimisation sera à poursuivre dans le futur.

Conclusions et perspectives

Dans cette thèse, nous avons exploré de nouveaux types de transistors nano-ioniques pour la réalisation de synapses artificielles. Nous avons d'abord élaboré des transistors synaptiques tout-solide à l'échelle d'un wafer en utilisant des techniques de microfabrication compatibles CMOS : une première génération de composants (deux types d'empilements possibles : $LiCoO_2$/LiPON, Li_xTiO_2/LiPON) a été réalisée. Les propriétés physiques et structurales de tels transistors ont été caractérisées par différentes techniques de microscopie et de spectroscopie (MEB, MET, spectroscopie Raman). Leurs performances en termes de comportement synaptique (modulation de la conductance, stabilité des états, non-linéarité, consommation d'énergie et endurance) ont été clairement démontrées. Une étude électrochimique systématique (focalisée sur le matériau constituant le canal du transistor) a permis d'expliquer l'origine des performances de ces composants (comportement pseudo-capacitif de Li_xTiO_2). À partir des résultats expérimentaux, des réseaux de calcul neuromorphique (ANNs et SNNs) ont été simulés. En particulier, un réseau de neurones artificiels (ANN) composés de matrices de transistors synaptiques a été simulé et testé sur différentes tâches de reconnaissance de formes. Le comportement cognitif de conditionnement classique (expérience de Pavlov) a également été simulé, montrant l'applicabilité potentielle de nos transistors synaptiques aux réseaux de neurones à impulsions (SNNs).

Dans le futur, nous nous concentrerons sur l'optimisation des performances de ces dispositifs (nombres d'états, gamme de conductance, stabilité, etc.) en nous basant sur

deux approches principales : (i) l'ingénierie des matériaux, et (II) la configuration des dispositifs. En ce qui concerne les matériaux, nous chercherons à améliorer les propriétés spécifiques des couches existantes, telles que la conductivité ionique et la stabilité de la gamme de tension de l'électrolyte, la diffusivité ionique et les ions mobiles disponibles des matériaux du canal (via des processus thermiques ou de dépôt). Pouvoir contrôler la nature des phases du matériau constituant le canal affecte la tension de programmation et la linéarité de la commutation, pourrait potentiellement permettre d'obtenir une synapse artificielle sans sélecteur (gain en termes de surface et d'énergie). En ce qui concerne la configuration des dispositifs, nous visons à réduire l'énergie de programmation, en réduisant la taille de la zone active et en augmentant le chevauchement entre la grille et les électrodes inférieures. Une couche de passivation sur le dessus de la grille pourrait également être conçue pour empêcher la perte d'ions par oxydation lors de l'exposition à l'air, augmentant ainsi la stabilité et l'endurance des dispositifs.

PUBLICATIONS DURING THE THESIS

Research article :

- Ngoc-Anh Nguyen, Olivier Schneegans, Raphaël Salot, Yann Lamy, John Giapintzakis, Van Huy Mai, Sami Oukassi, "**An Ultra-Low Power Li_xTiO_2-based Synaptic Transistor for Scalable Neuromorphic Computing**,"
  Advanced Electronic Materials, 8, 2200607 (2022)
  https://doi.org/10.1002/aelm.202200607

Conferences :

- Ngoc-Anh Nguyen, Olivier Schneegans, Jouhaiz Rouchou, Yann Lamy, Jean-Marc Boissel, Marjolaine Allain, Sylvain Poulet, Sami Oukassi, "**Elaboration and Characterization of CMOS Compatible, Pico-Joule Energy Consumption, Electrochemical Synaptic Transistors for Neuromorphic Computing.**"
  241st Electrochemical Society Meeting, Vancouver, Canada, June 2022.
  https://ecs.confex.com/ecs/241/meetingapp.cgi/Paper/155692

- Ngoc-Anh Nguyen, Sami Oukassi, Marjolaine Allain, Clémence Hellion, Yann Lamy, Cécilia Dupré, Marcelo Rozenberg, Kang Wang, Pascale Senzier, Claude Pasquier, John Giapintzakis, Van Huy Mai, Olivier Schneegans, "**Associative Memory Demonstrated By a Simple Design of Spiking Neural Network with an Ionic Synaptic Transistor**"
  Submitted to IEEE International Symposium on Circuits and Systems 2023 (ISCAS 2023)