Two Dimensional Dynamic Synapse With Programmable Spatio-Temporal Dynamics For Neuromorphic Computing

by

Mohammad Taghi Sharbati

B.S. in Electrical Engineering, Mazandaran University, Iran, 2006

M.S. in Electrical and Computer Engineering, Shiraz University of Technology, Iran, 2009

M.S. in Electrical and Computer Engineering, University of Pittsburgh, US, 2016

Submitted to the Graduate Faculty of

Swanson School of Engineering in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2021

UNIVERSITY OF PITTSBURGH

SWANSON SCHOOL OF ENGINEERING

This dissertation was presented

by

Mohammad Taghi Sharbati

It was defended on

September 9, 2021

and approved by

Mahmoud El Nokali, Ph.D., Associate Professor, Department of Electrical and Computer Engineering

Kevin P. Chen, Ph.D., Professor, Department of Electrical and Computer Engineering

Samuel Dickerson, Ph.D., Assistant Professor, Department of Electrical and Computer Engineering

Ryad Benosman, Ph.D., Professor, Department of Ophthalmology, School of Medicine

Dissertation Director: Feng Xiong, Ph.D., Assistant Professor, Department of Electrical and Computer Engineering

Two Dimensional Dynamic Synapse With Programmable Spatio-Temporal Dynamics For Neuromorphic Computing

Mohammad Taghi Sharbati, Ph.D.

University of Pittsburgh, 2021

In today's era of big-data, a new computing paradigm beyond today's von-Neumann architecture is needed to process large-scale datasets efficiently. In response to this need, the field of neuromorphic computing has recently emerged. Inspired by the brain, neuromorphic approaches are better at complex tasks than even supercomputers and show much better efficiency. This is because, unlike modern computers that use digital '0' and '1' for computation, biological neural networks exhibit analog changes in synaptic connections during the decision-making and learning processes. However, the existing approaches of using digital complementary metal-oxide-semiconductor (CMOS) devices to emulate gradual/analog behaviors in the neural network are energy intensive and unsustainable; furthermore, emerging memristor devices still face challenges such as non-linearities and large write noise. Here, we propose a novel artificial synaptic device use of an electrochemical dynamic synapse based on two-dimensional (2D) materials. The synaptic weight (channel conductance) of these dynamic synapses can be tuned via both a long-term doping effect from electrochemical intercalation and a short-term doping effect from ionic gating, thereby demonstrating programmable spatio-temporal dynamics, an essential feature for implementing spiking neural networks (SNNs). The electrical conductance of the channel is reversibly modulated by a concentration of Li ions between the layers of the 2D materials. This fundamentally different mechanism allows us to achieve a good energy efficiency (< 700 aJ per switching event), analog tunability (>5000 non-volatile states), good endurance and retention performances, and a linear

and symmetric resistance response. We demonstrate essential neuronal functions such as excitatory and inhibitory synapses, short term and long term plasticity, paired pulse facilitation (PPF), spike timing dependent plasticity (STDP), and spike rating dependent plasticity (SRDP), with good repeatability. Our scaling study suggests that this simple, two-dimensional (2D) synapse is scalable in terms of switching energy and speed. This work can lead to the low-power hardware implementation of neural networks for neuromorphic computing.

Table of Contents

List of Tables

List of Figures

Preface

I would like to express my deepest appreciation to my advisor, Professor Feng Xiong, who has continually and convincingly conveyed a spirit of adventure in regard of research. Without his guidance and persistent help this research would not have been possible.

I'm deeply indebted to labmates , Yanhao Du, Qingzhou Wan, and John Erickson, for the stimulus discussions, for the days we were working together and for all the funs we have had in the past five years.

I would like to extend my sincere thanks to my committee members Professor Mahmoud El Nokali, Professor Ryad Benosman, Professor Kevin Chen, and Professor Samuel Dickerson for serving as my committee members and for all your suggestions.

I would also like to extend my deepest gratitude to my parents Abbas Ali and Ommolbanin Sharbati for their wise counsel, supports and all the scarifies that you have made. I cannot begin to express my thanks to my beloved wife, Samane Khademi, who my success would not have been possible without her supports and encouragement to keeping continue and overcoming the challenges. I would like to express my thanks to my beloved son Ali Sharbati for being such a good boy always chirring me up.

1.0 Introduction

1.1. Motivation

1.1.1 Neuromorphic Computing and Artificial Neural Networks (ANNs)

Necessity for the next generation electronic devices with improved performances in terms of energy and speed is increasing speedily as transistor technology is closing to its Moore's law. The human brain as the most efficient computational system has attracted a lot of attention between researchers to develop highly efficient device by emulating brain behaviors. The human brain only consumes 20 W to do very complicated tasks such as speech and pattern recognition and all daily calculations required for our body to be functional daily life.[1] In contrary, the most famous super computer introduced by Oak Ridge National Laboratory uses significant energy 9.78 MW to do a complicated task.[2] The major challenges for the traditional computers that make them inefficient in energy consumption and speed is Von Neumann bottleneck due to the physical separation of the memory and central processing unit (CPU) and the information needs to be transferred between these two location due to the sequential execution of the operations.[3,4]

Contrarily, biological systems able to do both processing and storing at the same location that enable to execute computations in massively parallel networks, reducing the energy cost per operation.[5] Consequently, researchers are inpired to develop neuromorphic computing systems that can emulate brain functionalities to acheive or even exceed the brain efficiency. To fulfill this, scientists have successfully developed and applied ANNs in numerous fields including: image and pattern recognition,[6] speech recognition,[7] machine translation,[8] and beating humans at

chess and recently, Go.[9] The hardware implementation of the ANNs are hindered by digital transistor as building block of current artificial synapses as thay cannot trully imitate the analoge bahavior of the biological synapses as building block of the biological neural network. The basic neural network operations such as neurpmorphic computations, Deep Neural Networks (DNNs) and Spiking Neural Networks (SNNs), will be reviewed in the folloaing sections. In addition, variouse of current approach that are under investigation to improve the synaptic devices performances with regard to the hardware acceleration of ANNs. Furthermore, different number of emerging arificial synaptic devices including phase change memory (PCM) based synapses, resistive random-access memory (RRAM) based synapses, Electrochemical synapses and two-dimensional (2D) materials based synapses will be introduced and reviewed.

1.2 Background

1.2.1 Neural Network Basics: Synapse and Network Operation

The human brain, with its dense neural networks and adaptable memory elements (synapses), presents a fascinating computing model with extraordinarily low power consumption.[10] The human brain has almost ~10^{11} neurons, and each neuron is connected to another neuron through ~10^{4} synapses, for a total of ~10^{15} total synapses.[10,11] As illustrated in **Figure 1.1**, each neuron is constituted of many different parts, including the soma that forms the body and is connected to the network via dendrites and an axon. Information is received from other neurons (inputs) via dendrites, and the terminal branches in the axon transmit the information on (output). The small gaps (20-40 nm) between the axon end of the presynaptic neuron and the dendrites of the postsynaptic neurons are known as synapses. The connection strength between neurons, known as synaptic weight, can be strengthened (potentiation) or weakened (depression) during a procedure called synaptic plasticity, which is widely believed to be of major importance in the process of learning and memory making in the brain.

After firing a neuron, a signal is realized and travels down the axon into the dendrites of the next neuron through synapses. The size of this signal is dependent on the neural connection strength or synaptic weight. Many artificial synaptic systems adopt a crossbar synaptic array structure that can be scaled up for an (Artificial Neural Network) ANN implementation, as depicted in **Figure 1.1**. Each input line through programmable resistors is connected to the output lines and each input influences the total output of each line. These resistors as synaptic devices determine how much of the input signal goes to each output signal. Various selection devices in modern arrays impede errors arising from so called sneak currents falsely influencing the

output.[12] Comprehensive research has been done to optimize these synaptic devices and their performance and establish the best possible neuromorphic system. Here besides reviewing the working principles behind ANNs, the device characteristics and consequent impact on these networks' performances has been discussed.

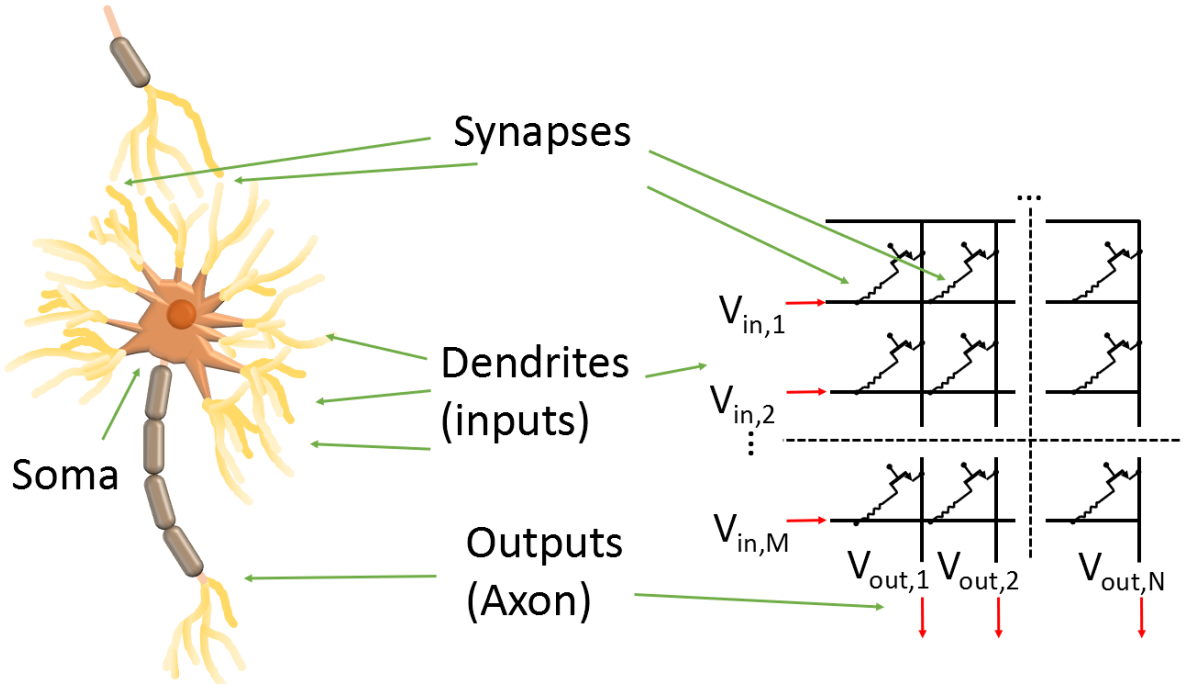

Figure 1.1. Comparison of biological and artificial neurons. Schematics of a biological neuron (left) vs. an artificial neural network (right).

1.2.1.1 Neuromorphic Computation

To show how ANNs execute their computation, we use a simple example of the recognition of handwritten digits across a feed-forward network, as illustrated in **Figure 1.2**. Initially the input image is separated into N binary iputes in black and white and each input is representing a specific

area of the initial image. The inputs then feed to a one hidden layer that is representing M hidden neurons for computations after receiving N inputs. Each neuron of an M hidden neuron in turn is connected to each 10 output neurons. A SET threshold level has been defined for the hidden layer, and if the summation of each input comming from the firat layer passes this level, the corresponding signal will be propagated into the next layer. This process occurs in all network layers untill it hits 10 output neurons (corresponding to 0-9 digits) in the output layer. The system then can realize the initial digit image, depending on the firing of the final neuron. Two important parameters that need to be considered are determining the threshold level and synaptic weight (connection strength) between the neurons. The synaptic weight is determining the weight of the distribution of the each input signal to each neuron of the next layer (M hidden layer). The synaptic weight can be trained using various algorithms and different methods. The main two methods are: offline training and online training. In offline training, a huge series of known data is fed into the system and the relevant synaptic weight is adjusted until the output is the known correct output. However, this method has some disadvantages such as the requirement of a large and established data SET to feed into the system and extensive trainingg time. For online training network training happens at the same time the data is entered. However, although this methods allows processing of dynamic data sets, sizable peripheral circuits are required for performing a vast number of weight updates in real time, as well as more on-chip memory for storage of the new weight values.[13] In online training the synaptic devices are required to show high endurance as the synapses are usually more on-the-fly. Furthermore, a large number of conductance states and a linear synaptic weight change are required to make it easier for the system to converge to an error minima. After completing the training, the self operation of the network, depending on the efficiency of the training scheme, will show different degrees of success. Two important

parameters to determine the choice of type of training, either offline or online, are the task that needs to be completed and the network type.[14] DNNs and SNNs are the two most well-known network types, where the former is useful for applications such as speech recognition and image classifications while the latter are used to allow ANNs to reach to maximum potential in artificial networks similar to the biological neural systems.

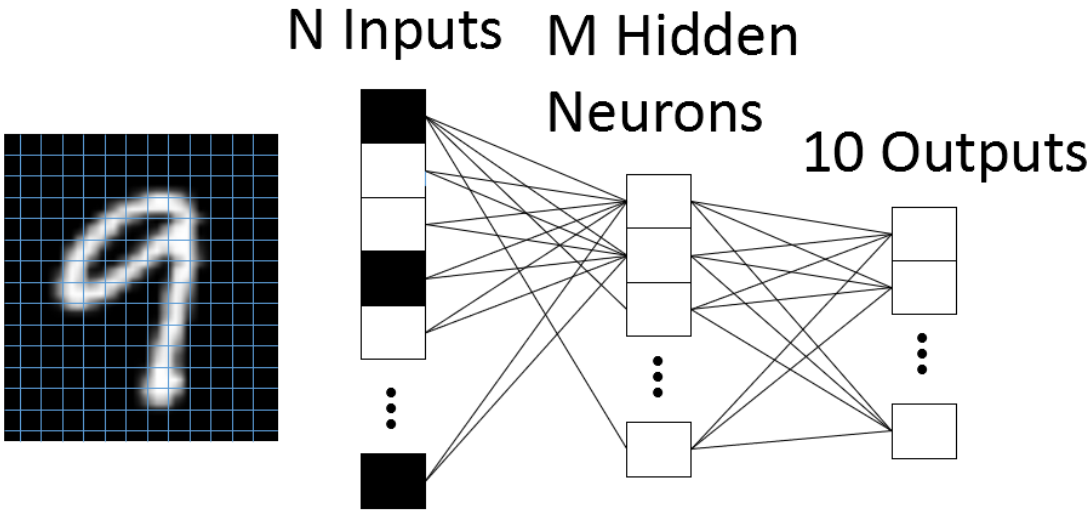

Figure 1.2. Simple pattern recognition example. 2-Layer multilayer perceptron example of a simple network structure for handwritten digits recognition. The image is broken into a grid of N total areas, (the N input elements) those elements are all connected into a Hidden Neuron layer of size M, and those Hidden Neurons are then connected to the 10 Output Neurons, representing the digits 0-9. Reproduced with permission.[15] Copyright 2017, IEEE.

1.2.1.2 Deep Neural Networks (DNNs)

In 1943 McCulloch and Pitts developed a neuoal network for the first time, with the first perception of it implemented in 1958.[16] Because of computing power deficiency, the first devices

were unremembered until the 1980s. However, with the discovery of the error backpropagation techniques and chain rule in the late 20th century, we begin to observe a continiuous increase in accuracy and sophistication in neural network systems .[16] DNNs (Multilevel Perceptrons, Deep Belief Networks, Convolutional Neural Networks etc.), as motivation behind modern day neural network advances, experienced a speedily explore by emerging modern computational power in the early 21st century. DNNs, with their vast complexity and many layers, were capable of unsupervised, semi-supervised, and supervised learning and exceeded in tasks that had extensive training data.[16] Moreover, the increase in commecial availability of graphic processor units (GPUs) and field programmable gate arrays (FPGAs) as powerful parallel computation devices allowed development of new algorithms and network structures, resulting in further significant progress in DNN development .[17] Nevertheless, along with the continiuous progress in DNN devleopment came increases in complexity that requireda vast number of computational resources for their training. Consequently, substantial research efforts have been devoted o the maturing of DNN accelerators using custom designs. To fabricate a DNN accelerator, current CMOS devices are utilized with the help of on-chip buffers.[18] The major drawback for theses solutions, however, is scalability, as the large number of memory devices, usually static random access memories (SRAMs), are required. In addition, the size SRAM cells (each SRAM cell includes eight transistors) results in low power efficiency caused by leakage current, creating another challenge for these memory devices.[18] Advances in artificial synaptic devices that can emulate human cognitive capabilities are indispensable to improving existing DNN configurations to acheive dense and energy efficient devices. Thus researchers began to develop Spiking Neural Networks (SNNs).

1.2.1.3 Spiking Neural Networks (SNNs)

Due to their close resemblance to biological systems, SNNs have attracted a lot of attention. It is believed that the major reason that capable brain in sequence recognition and memory is its potential to process spikes.[19,20] To make SNNs a reality, a technique to govern spike-timing-dependent plasticity (STDP) is required. This technique is hypothesized to be what dictates causality in human brain. If one stimuli correlates/causes another one several times, this correlation/causality is strengthened via synaptic connection change in the brain.[21] As summarized by Löwel and Singer "neurons that fire together, wire together."[22] As illustrated in **Figure 1.3a,** for the single synapse structure with pre- and postsynaptic neuron, by continuous firing of presynaptic neuron without considering firing of post-synaptic neuron, it would seem that the pre-synaptic neuron has a direct effect on post synaptic neuron whether or not the post-synaptic neuron fires. As a result, the one observes a strengthened (excitatory) or weakened (inhibitory) in connection between these two synapses depending on the synapse's type. On the other hand, with continuous firing of the post-synaptic neuron before firing in pre-synaptic neuron, it would seem that the pre-synaptic neuron has no effect on the relevant post-synaptic neuron and one observes weakened (excitatory) or strengthened (inhibitory) in connection area between these two neurons. There are four forms of STDP and a related synaptic weight change in the biological brain as illustrated in **Figure 1.3b**. Besides STDP, other parameters like dendritic locations, spiking orders, and firing rate have displayed to influence synaptic plasticity in the brain as well.[23]

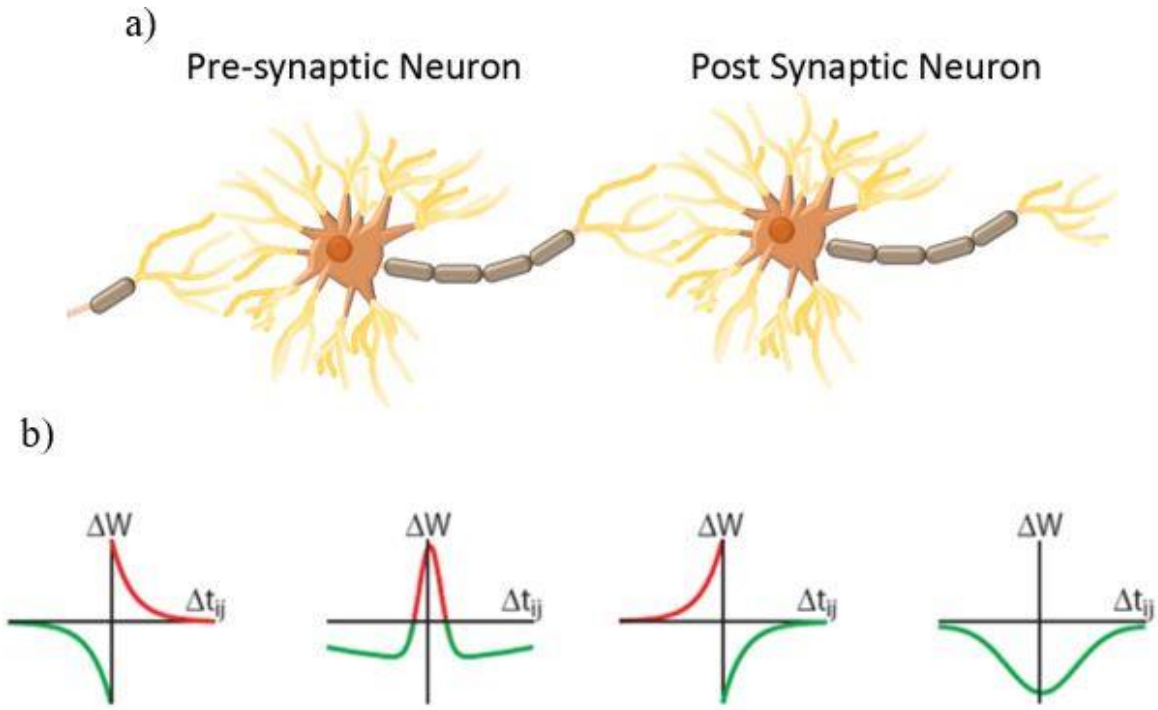

Figure 1.3. Spike-timing-dependent plasticity (STDP) models. a) Example of a neuron (Neuron A) with its prior neuron layer (Neuron B). b) Examples of STDP desired weight change based on the time difference between Neuron A and Neuron B firing. Reproduced with permission.[24] Copyright 2018, John Wiley and Sons.

Similar to DNNs, which can be built using many different approaches,, many techniques have been developed to build SNNs as well, depending on the required applications and the involved artificial synaptic device.[25] However, all the approaches need stringent timing mechanisms in addition to external secondary circuity to accomplish. In addition to the large physical chip area that is required for these incidental circuitries, a huge amount of memory is essential, resulting in unrealistic scalability of these synaptic devices.[25] As a result, to implement the next generation of SNNs, a synaptic device that is capable of internally incorporating STDP behaviors is required.

1.2.1.4 Existing Approaches and Synaptic Device Requriements

A few neuromorphic computing systems are available to be used for hardware implementations of ANNs. The popular DNN system called TrueNorth uses almost 4096 CMOS based neuromorphic cores consisting of 10^6 neurons and 246×10^6 synapses with a capability of conducting 58 G-synaptic operations per second (GSOPS). Despite its efficcient power consumption per operation (~ 2.5 pJ/op), compared to the human brain with a significant energy efficiency of per operation (~2 fJ/op), a lot of progress needs to be made and challemges addressed.[26, 27] Another well-known neuromorphic system based on SNNs is SpiNNaker.[28] Each of its 18 general purpose CPUs can model a few hundred neurons. Each neuron with its 103 synapses along with the software considerations execute the required computations. Although SpiNNaker consumes more power compared to other existing computing systems, it surpasses others at training speed. While other computing systems uses offline training that might take hours, SpiNNaker uses online training in real time for large scale SNNs simulations, which provides a highly efficient spiking system.[29] Despite these promissing aspects, these systems suffer from scalability and high power consumption as they are utilizing traditional digital CMOS technology in their artchitecture. Therefore exploring a new synaptic device for hardware implementation and ANN accelerations is critical, and researchers are working on various types of artificial synapses to address the aforementioned challenges.

In recent years various synaptic devices have been proposed. On the face of it, synaptic devices and emerging non-volatile memories (NVMs) have a lot of properties in common such as easy programming, easy reading, and high data retention. Accordingly, emerging NVM devices such as RRAM, spin-torque transfer random access memory (STT-RAM) and PCM have been adopted for developing synaptic devices due to their fast switching speed, low programing power,

good retention and scalability, all characteristics which have a vital role in emerging neuromorphic computing devices.[15]

Despite their similarities, NVMs and synaptic devices have different applications and so their device metrics can be different. The most importan difference between NVMs and artificial synapses is the number of states available. NVMs are usually binary or at most multi-level cells (MLC) with maximum 8 states per device; however, for the purpose of learning accuracy and device density, artificial synaptic devices require a high number of distinct states. Moreover, while a high on/off ratio is prefered for NVMs for signal-to-noise ratio improvement, in synaptic devices that can result in low sensivitivity to individual synapses in larger arrays. Furthermore, some properties that are not important in NVMs such as symmetry and linearirty of the current-voltage govern DNN accuracy and play a key role in synaptic devices for ANN applications. The endurance and retention that are required in NVM applications, are important in synaptic devices as well depends on the required application. For example, a good endurance is required for a neural network that performs weight updates using online training. However, retention performance that does not play vital role in these systems are important in offline neural network with presynaptic weight.

Now, the different types of artificial synaptic devices that have been developed so far will be appraised. There are four major types of synaptic devices that we will discuss including PCMs, RRAMs (including conductive-bridging, filamentary, and interfacial), electrochemical intercalation based synapses, and 2D material based synapses. We will review the working prenciples and device metrics for each device, evaulating their potential for implementation of the current ANN frameworks in terms of operating speed and energy, number of states (precision), the

degree of symmetry and linearity of the state switching, reliability (device variations, endurance, and retention), and, finally, their potential in spatio-temporal dynamics for SNNs.

2.0 Emerging Artificial Electronic Synapses

2.1 Phase Change Memory (PCM)

Recently a promising NVM technology, PCM, has emerged with good reliability, fast programming speed, scalability, low device-to-device variations, good endurance, multiple programming resistance levels, and high packing density.[30-35] In **Figure 2.1**a, a commonly used PCM cell structure, mushroom or T-cell, is illustrated in which the nanoscale PCM material is sandwiched between two bottom and top electrodes, where the phase change material can be switched between a crystalline phase in SET process (i.e. low resistance state or LRS) and an amorphous phase in RESET process (i.e. high resistance state or HRS) via Joule heating.[36] The concept of using phase change materials for memory application was originally proposed by Ovshinsky in 1969,[37] but it was the discovery of a family of fast switching (<100 ns) chalcogenides (Ge, Sb, and Te alloys) by Yamada *et al.* (**Figure 2.1b**),[38] which prompted the commercialization of PCM based optical data storage (DVDs and blu-rays) in the 1990s and, subsequently, the development of electrically programmable phase change random access memory.[39]

In SET process, as shown in **Figure 2.1c**, [40] upon application of an electrical field (a series of electrical pulses with low amplitude) and consequence threshold switching, Joule heating is induced, causing the phase change material to heat up to its crystallization temperature (typically ~150 °C) .[41] Upon crystallization, the material switched from the amorphous phase, where it is in a high resistance state (HRS), to a crystalline phase with a low resistance state (LRS). However, in the RESET process (the phase transition from crystalline to amorphous state) with application

of an electrical voltage pulse with high amplitude (to heat up the cell to above its melting temperature, usually >650 °C[42]), the programming region will melt due to the Joule heating induced by the current moving through the cell and quench rapidly due to the glass transition when the reset pulse is cut off. The SET process is the rate limiting step in PCM since the crystallization process involves atomic movement whereas RESET is the power limiting step as the cell needs to be heated up to its melting temperature.

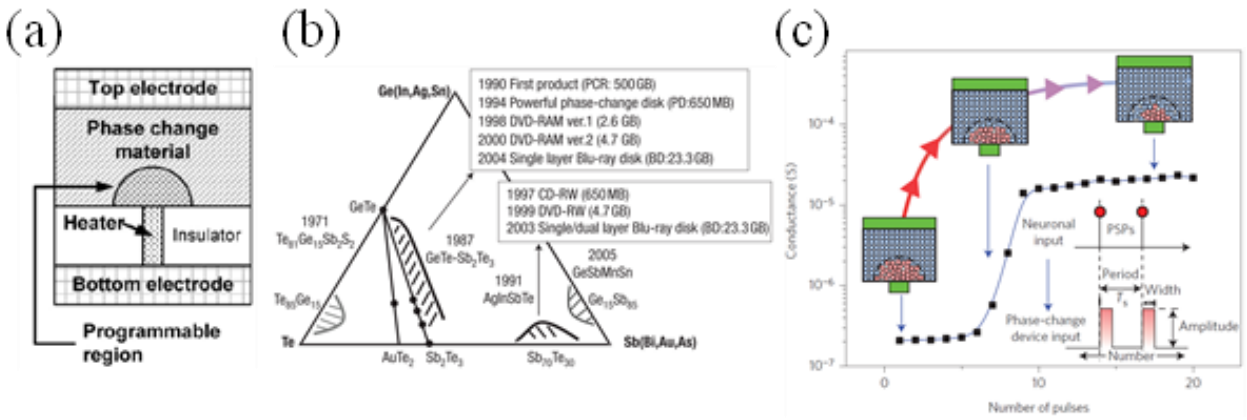

Figure 2.1. PCM synapse. a) Cross-section schematic of the mushroom PCM device. b) Family tree of fast switching chalcogenides. c) Schematic for implementing synaptic plasticity in PCM synapses. a) Reproduced with permission.[36] Copyright 2010, IEEE. b) Reproduced with permission.[38] Copyright 2007, AIP. c) Reproduced with permission.[40] Copyright 2016, Nature.

Unlike with binary memory applications, which switch immediately from one state to the other, the gradual phase transition in PCM between crystalline and amorphous states enables PCM materials to be useful for neuromorphic applications and ANN hardware accelerations due to its good scalability, fast speed, and its potential for analog switching. Various pulsing schemes, as shown in **Figure 2.2a**, have been explored to program PCM devices for synaptic plasticity

implementation. The PCM conductance can be tuned by modulating the amorphous to crystalline resistance ratio .[40] In **Figure 2.2b**, the cross-section transmission electron microscopy (TEM) images of a gradual programming process with fine control of resistance of a GST-PCM device is shown for both the SET and RESET processes through 100 steps.[11, 43] This working mechanism and multilevel cell operation has been employed for array density improvement in NVM applications. However, for neuromorphic applications a synaptic device with higher level of precision (# of available distinct states or precision level per device) is indispensable to enable it to be used for ANN implementation. A $Ge_2Sb_2Te_5$ (GST)-based PCM synaptic device reported by Wright et al. displayed 10× precision with a dynamic range of 50× as illustrated in **Figure 2.3a**.[44] The GST PCM studied by Zhong *et al.* showed 10× precision and 5× dynamic range per device (**Figure 2.3b**).[45] Kuzum et al., in 2012, were able to demonstrate a PCM device with high precision (120 states per device) and reasonable dynamic range (~50×) with the help of a train of pulses with increasing amplitude (**Figure 2.3c**).[43] This considerable improvement in precision level in PCM artificial synapses promoted utilizing these devices for the hardware implementation of neural networks. Despite the promising aspects of???, such as good symmetry observed in the conductance response, the large cycle-to-cycle and device-to-device variation seen and high nonlinear behavior due to the stochastic nature of the phase transition processes negatively affects the learning performance in PCM synaptic arrays.[46-49]

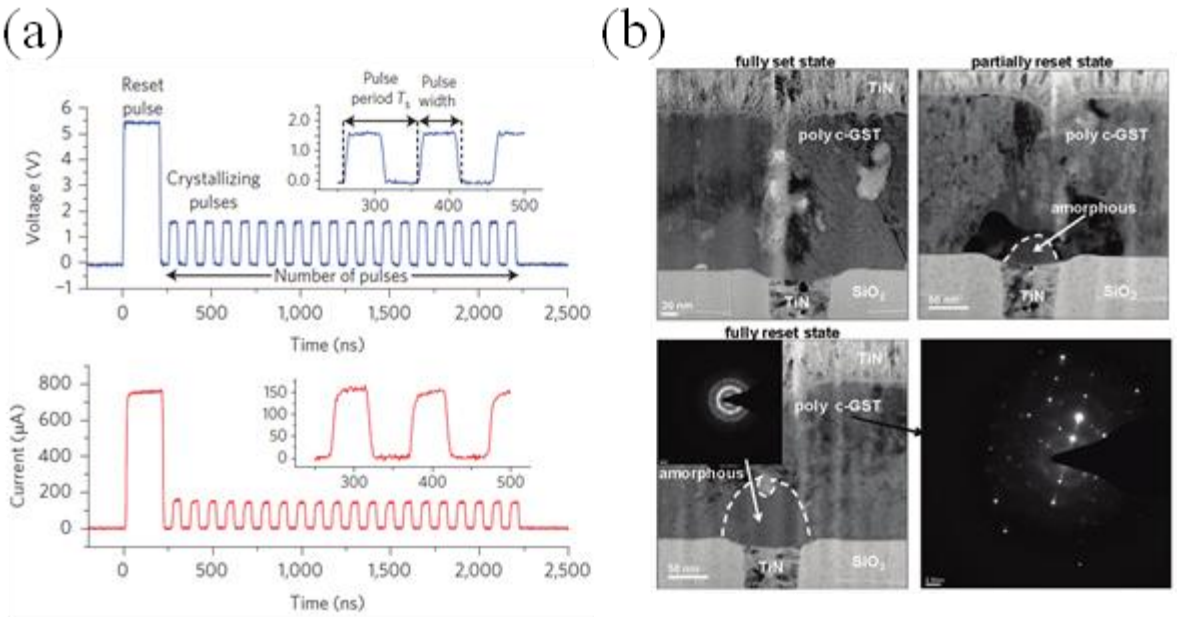

Figure 2.2. SET and RESET in PCMs. a) A series of set pulses with low amplitude for set cycle and single high amplitude pulses required for reset cycle in blue (top), and corresponding set and reset current in red (bottom). b) TEM images illustrating the phase transition from fully set state to partially reset state with small amorphous region shown by small mushroom and fully reset state. a) Reproduced with permission.[40] Copyright 2016, Nature. b) Reproduced with permission.[43] Copyright 2012, ACS.

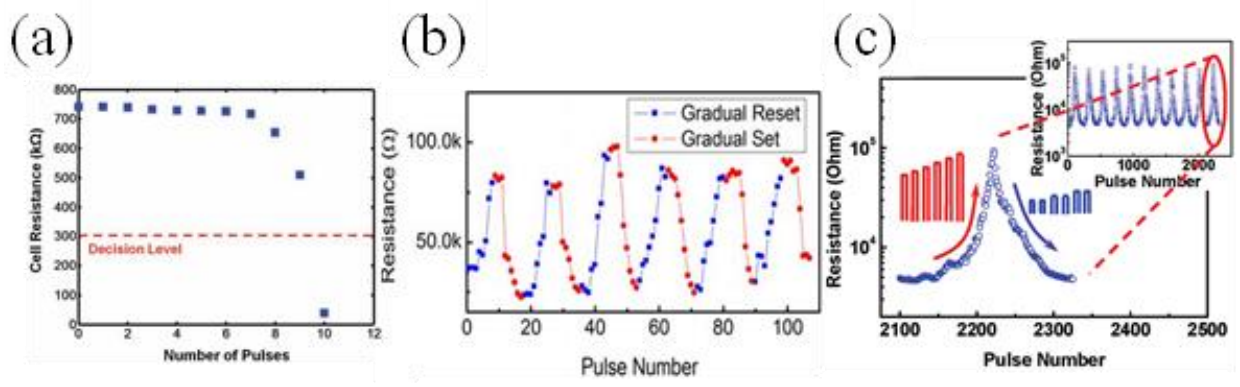

Figure 2.3. Synaptic plasticity in PCMs. a) Resistance changes of the PCM device upon the application of 10 pulses (~1 V, 60 s). b) Resistance change using simple square pulse for the both set and reset cycles. c) Gradual reset and set using pulses with increasing amplitude. a) Reproduced with permission.[44] Copyright 2013, John Wiley and Sons. b) Reproduced with permission.[45] Copyright 2015, John Wiley and Sons. c) Reproduced with permission.[43] Copyright 2012, ACS.

2.2 Resistive Random Access Memory (RRAM)

RRAM devices are one of the most attractive emerging NVM technologies with their programmable resistive level for information encoding. The promising properties of these devices, such as a simple two terminal structure, low energy switching, fast switching speed, scalability and compatibility with established CMOS technology, have encouraged a lot of resarchers to work on developing these devices. In addition to their usefulness for memory applications, their non-volatility and minimal standby leakage power compared to SRAM devices mean that RRAM devices are also being explored for the hardware hastening of artificial neural networks. Moreover, RRAM devices with a crossbar array structure have the potential for large scale integration, a quality indispensable for ANNs. RRAM can be categorized into the three major types based on the switching mechanisms: CBRAM, filamentary RRAM, and interfacial RRAM (**Figure 2.4**).

The working mechanism and suitability of each of above-mentioned RRAM devices as artificial synaptic device are investigated in the following paragraphs.

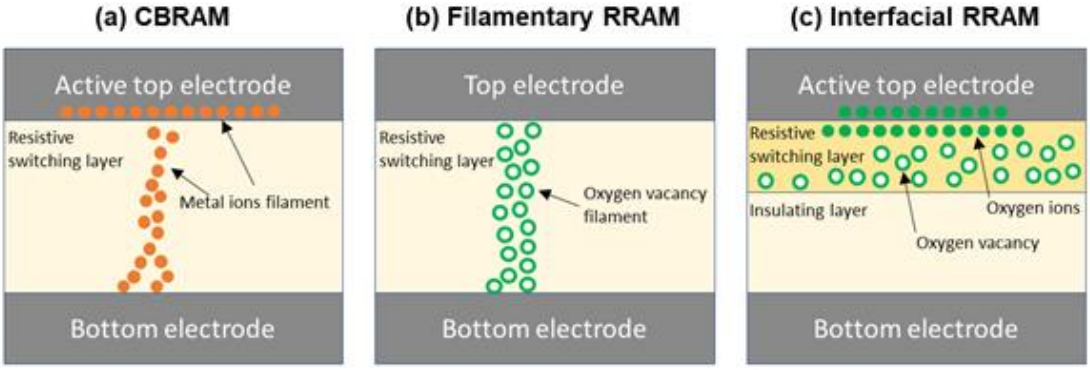

Figure 2.4. Device structure and working principle of CBRAM, filamentary RRAM, and interfacial RRAM. a) CBRAM is based on the conductive metal-ions filament formed electrochemically between an active top electrode and resistive switching layer. b) Filamentary RRAM is based on an oxygen vacancy filament through the resistive switching layer that is sandwiched between two inert metal electrodes. c) Interfacial-type RRAM is based on the barrier modulation effect at the metal /switching material interface layer, where the migration of oxygen ions towards the electrode reduces the effective barrier width of electron tunneling and immobile oxygen vacancies randomly distribute in the resistive switching layer.

As illustrated in **Figure 2.4a**, the CBRAM operates via formation and dissolution of a conductive bridge pathway during the SET and RESET processes and switches between the resulting LRS and HRS states. The bridge is comprised of active metal cations such as Ag or Cu that are formed in the insulating switching layer such as metal oxides,[50] amorphous silicon,[51] or solid electrolyte.[52] Upon applying an electrical field in the SET process, the cations from the top

electrodes are injected into the switching layers and the metallic conductive pathway bridges between the top and bottom electrode, and the device resistance switches to an LRS state. On the other hand, upon applying an electrical field with opposite polarity in the RESET process, this conductive bridge dissolves and device resistance transits to the HRS state. Characteristics such as simple structure and scalability, good retention, fast switching speed, and low switching voltage make CBRAM a promising candidate for synaptic device development. However, some drawbacks such as non-linear and asymmetrical conductive responses, large cycle-to-cycle and device-to-device variation due to the stochastic nature of the conductive filament formation and dissolution have a negative effect on ANN performance and need to be addressed. Moreover, high conductivity in the LRS state (typically ranging from 300 Ω to 1 kΩ)[53] may result in huge leakage current in the crossbar structure, which is another issue for CBRAM devices implementation in hardware.

In filamentary RRAM, similar to in CBRAM, the conductance change is based on a filamentary bridge forming and rupturing, only in this case with the help of available oxygen vacancies instead of metal cations, in the insulating material sandwiched between two electrodes. In **Figure 2.4b**, a typical metal-insulator-metal (MIM) vertical and two terminal structure of the filamentary-based RRAM is illustrated, where the insulating layer (switching layer) is made of metal oxides such as TiO_x,[54, 55] HfO_x,[56-59] AlO_x,[54, 56] WO_x,[60, 61] and TaO_x.[62, 63] Upon application of an electric field during the SET process, a conductive pathway consisting of oxygen vacancies is formed, inducing a soft breakdown in the oxide material (with limited current compliance to control the process) and device transit from HRS to LRS state. During RESET process, on the other hand, the filament cracks either by recombination of oxygen vacancies with oxygen ions upon application of an electric field with opposite polarity (bipolar RRAM) or through

Joule heat under a larger electric field (unipolar RRAM), causing transition from LRS to HRS.[64] Despite the promising aspects of filamentary RRAM devices such as their simple structure, low cost, good scalability, low switching voltage, and highly accessible switching materials compatible with established CMOS technology and fabrication facilities, they suffer from some drawbacks, including, like CBRAM devices, the stochastic nature of the switching mechanism. This property due leads to difficulty controlling filament formation and dissolution; this lack of control induces large cycle-to-cycle as well as device variation, non-linear conductance response, and limited analog states, thereby limiting their performance as artificial synaptic devices.[62] The CBRAMs have some advantages over filamentary RRAMs including better reliability (retention and endurance) performance because of the greater stability of oxygen based filaments in comparison to metal cation based filaments. Furthermore, the RESET mechanism in CBRAM has smaller sensitivity to the electric field induced drift effect. However, some issues like voltage-time dilemma[65, 66] in filamentary RRAM, in particular in metal oxide-based devices, has been reported; this limits its performance because there is a tradeoff between long retention time and fast switching speed under the reading voltage stress test. A device with a TiN/HfO_x/AlO_x/Pt structure reported by Yu et al. exhibited good retention at high temperature (7200 s at 100 °C) and good cycling endurance (>10^5) upon application of short pulse (50 ns) indicating fast programming speed.[56] Additionally, Prezioso et al. showed that their Pt/TiO_{2-x}/Al_2O_3/Pt device exhibited significant retention (10 years) operating at room temperature with life cycle > 5000 at relatively slow switching speed of 500 µs.[54] However, filamentary RRAM based artificial synapses are suffering from large cycle-to-cycle and device-to-device variation because of the stochastic nature of the switching mechanism involved in the random distribution of oxygen defects.

The third type of RRAM, non-filamentary or interfacial RRAM, works based on the metal/switching material tunnel barrier modulation effect via oxygen ion migration. In this type of RRAM, as illustrated in **Figure 2.4c**, unlike in filamentary RRAM an extra oxide layer (insulating layer) is used to serve as a load resistor for thermal runaway prevention and elimination of the DC current compliance requirement. An exotic oxide like TaO_x and $Pr_{1-x}Ca_xMnO_3$ (PCMO) is incorporated in a typical resistive switching layer.[67-69] The SET process occurs whenoxygen ion migrate from the resistive switching layer in the direction of the metal/switching material interface upon application of an electric field. The electrons become capable of tunneling from the metal electrode to the switching layer due to barrier height reduction and cause transition from an HRS to an LRS state. On the other hand, in the RESET process, an electrical field with opposite polarity induces reverse migration of oxygen ions into the switching layer, resulting in strengthening of the tunnel barrier height and transition from the LRS to the HRS state. Interfacial RRAM is a promising candidate for synaptic applications due to its gradual switching process and forming of a free mechanism. However, despite the unique features presented by interfacial RRAM, including self-compliance, forming-free, and gradual switching with good analog precision, its relatively low programming speed cannot provide sufficient energy to build a high energy barrier to reach good retention.[10, 65] Furthermore, usage of exotic switching materials make the fabrication more complicated and expensive.

2.3 Electrochemical Devices

Apart from emerging technologies such as PCM and RRAM, which were originally developed with a goal towards enabling their use in NVM applications, scientists are demonstrating new devices created specifically for use in synaptic applications. For example,

recently researchers have developed electrochemical based artificial synapses with high precision levels, low programming energy, high scalability, a built-in timing mechanism and linear and symmetrical conductance responses appropriate for SNN applications.[24, 70-72] In this type of device, the channel conductance (the synaptic weight) can be modulated through gate terminals with precise control by controlling the ionic concentration in the channel (synapse). The ionic exchange is facilitated via electrochemical reactions in a gel electrolyte.

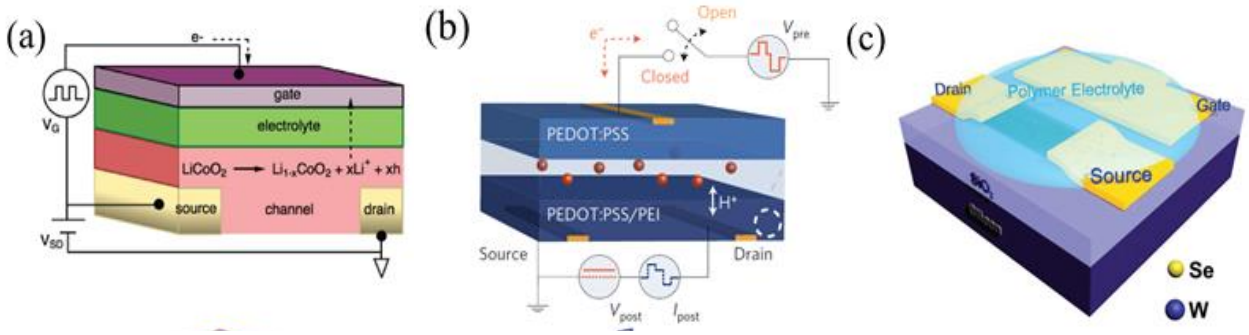

Figure 2.5. Electrical chemical synapses. a) A schematic of synaptic transistor with LiCoO2 as active channel. b) An organic polymer synapse consisting of liquid electrolyte (NaCl or KCl) sandwiched by PEDOT:PSS and PEI treated PEDOT:PSS layers. c) An ionic-gated synaptic transistor based on metal dichalcogenide crystal (WSe2) and phosphorus trichalcogenide (NiPS3 and FePSe3). a) Reproduced with permission.[70] Copyright 2017, John Wiley and Sons. b) Reproduced with permission.[71] Copyright 2017, Nature. c) Reproduced with permission.[72] Copyright 2018, John Wiley and Sons.

Three pioneer works that utilize an electrochemical intercalation technique are shown in **Figure 2.5a-c**, each of which uses different channel materials: lithium cobalt oxide ($LiCoO_2$),[70] organic polymers (poly(3,4-ethylenedioxythiophene) including polystyrene sulfonate or PEDOT:PSS),[71] and tungsten diselenide (WSe_2),[72] respectively. The key features necessary for

use in synaptic devices that these devices offer, such as good precision, good scalability, highly linear and symmetric response, and low power consumption, make them good candidates in terms of artificial synapse development. Despite having three terminal structures, low switching energy as well as good retention properties have also been achievable in these devices. This is because unlike in two-terminal devices such as RRAM[62] and PCM, [73] where there is a tradeoff between switching energy and retention due to the same barrier being used for programming and retention, in the three terminal structure introduced in electrochemical synapses, low switching energy and good retention characteristics are achievable as programming and retention mechanisms are decoupled.

A new approach for synaptic devices using intercalating and deintercalating mobile ions in the active channel as synapse was developed by Fuller et al.[70] in 2017 (**Figure 2.5a**). In this device, $LiCoO_2$, a common cathode material used in Li ion batteries (LIBs), and lithium phosphorus oxynitride (LiPON), havinggood chemical stability, are employed as an active channel and solid electrolyte, respectively.[74] They chose $LiCoO_2$ due to its well-characterized electrochemical properties and their good endurance, as demonstrated in LIBs.[75] The authors reported a precise, reversible, and controllable linear channel conductance change (synaptic weight) with a dynamic range of 100 μS and 200 distinct states, as illustrated in **Figure 2.6a**.

The same year, Burgt *et al*. .[71] revealed their nonvolatile organic device utilizing reversible electrochemical intercalation for neuromorphic computing applications (**Figure 2.5b**). They introduced PEDOT:PSS for use as the active channel where insertion and extraction of protons enable linear conductance change in the organic synapse. They reported good precision of 400 states with a dynamic range of 300 μS in a linear, repeatable, and symmetric conductance response, as depicted in **Figure 2.6b.**

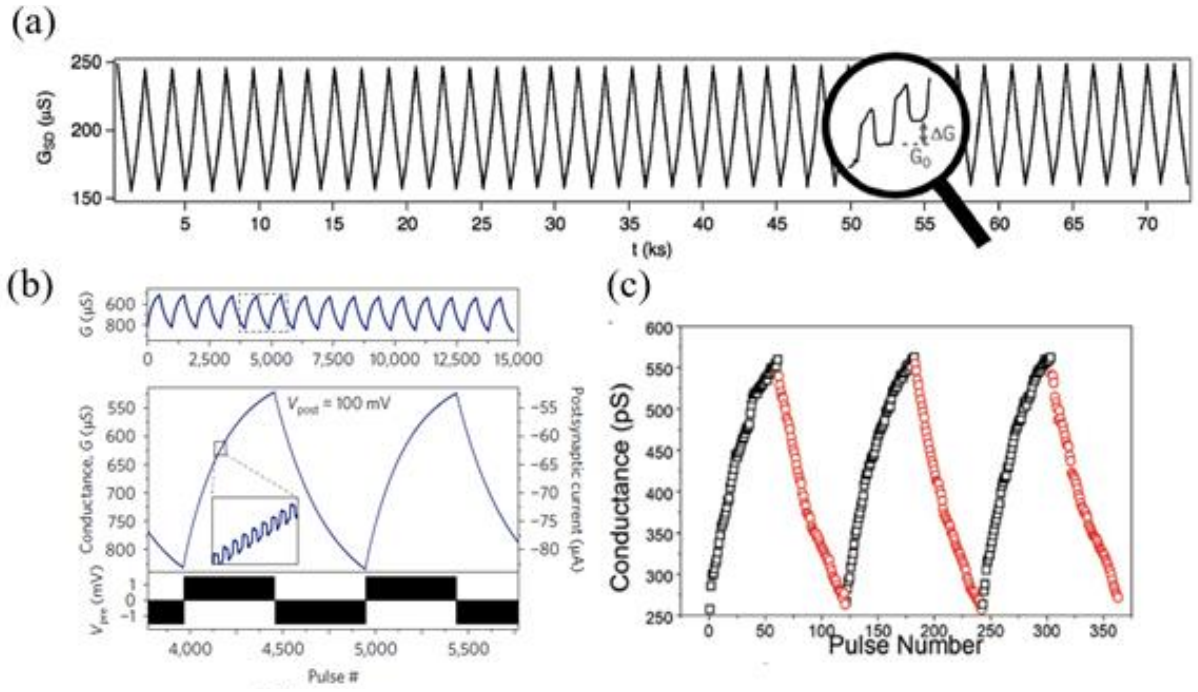

Figure 2.6. Long term potentiation (LTP) and depression (LTD) in electrochemial synapses. a) the $LiCoO_2$ synaptic transistor by Fuller *et. al*. b) the polymer synapse by Burgt *et al.* c) the WSe_2 synapse by Zhu *et al.* a) Reproduced with permission.[70] Copyright 2017, John Wiley and Sons. b) Reproduced with permission.[71] Copyright 2017, Nature. c) Reproduced with permission.[72] Copyright 2018, John Wiley and Sons.

In 2018, Zhu et al.[72] reported on their 2D materials-based electrochemical synapse. 2D materials are attractive materials for the interlayer gaps as they provide good sites for mobile ion accommodation through electrochemical reactions (**Figure 2.5c**). In the reported work, WSe_2 and phosphorus trichalcogenide were employed as the artificial synapse, and a gel electrolyte was used for ionic exchange between the channel and the metal electrode. They reported 60 distinct states with a 307 pS dynamic range for this device, as displayed in **Figure 2.6c.** In general, the artificial synapses developed based on electrochemical reactions showed higher precision levels in comparison with other technologies that were mainly designed for MLC NVMs; as the channel

conductance can be precisely tuned through gate terminals and mobile ion concentration within the active channel, better linear conductance response is achievable compared to PCM and RRAM synaptic devices.

2.4 Two-Dimensional (2D) Devices

Unique and remarkable optical,[76] electrical,[77] and thermal features[78] have been observed in 2D materials such as graphene (**Figure 2.7a**), hexagonal boron nitride (h-BN) (**Figure 2.7b**), transition metal dichalcogenides (TMDs) with the form of MX_2 (where M = transition metal and X = chalcogen, such as MoS_2 in **Figure 2.7c**), and black phosphorus (BP) (**Figure 2.7d**). Furthermore, 2D materials offer some special properties including sub-nanometer thickness without dangling bonds, indirect (bulk) to direct (monolayer) band gaps transistion when thinned down,[79] highly in-plane anisotropy (BP)[80] and anisotropic thermal transport (high in the in-plane direction[81] and low in the cross-plane direction[82]). In addition, 2D materials with qualities such as diverse band structures (semi-metals such as graphene, WTe_2),[83] small band gaps for low-power transistor applications ($MoTe_2$, $HfSe_2$),[84] ability to semiconduct (MoS_2, WS_2, BP),[85] and large band gap or insulating h-BN, have a strong potential for implementation in next generation computing systems. The artificial synaptic devices based on 2D materials, as shown in **Figure 2.7e-j**, are under demonstrations with the help of physical mechanisms displayed by them like charge trapping, resistive switching, Joule heating etc. One of the first 2D based artificial synapses using twisted bilayer graphene was reported by Tian et al. **(Figure 2.7e)**.[86].

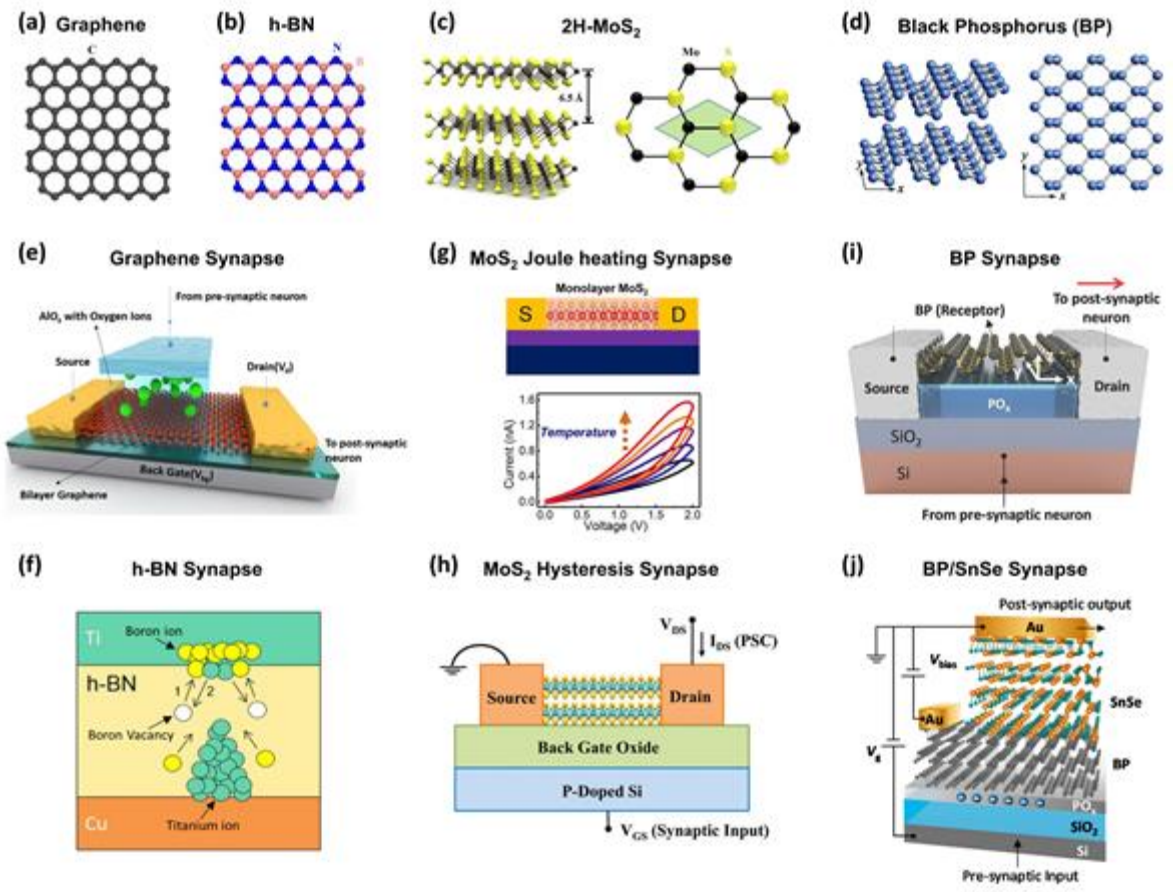

Figure 2.7. 2D synapses. a, b) Lattice structures of graphene and h-BN. Reproduced with permission.[88] 2017, Springer Nature. c) Lattice structure of MoS_2. Reproduced with permission.[89] 2013, American Chemical Society. d) Lattice structure of BP. Reproduced with permission.[90] 2014, American Physical Society. e) A schematic of a graphene dynamic synapse. Reproduced with permission.[86] 2015, American Chemical Society. f) A schematic of a h-BN synapse. Adapted with permission.[87] 2018, Springer Nature. g) A schematic of a MoS_2 synapses based on Joule heating. h) A schematic of a back-gated MoS_2 hysteresis synapse. i) A schematic of a BP synaptic device. j) A schematic of the BP/SnSe heterojunction synaptic device. a-g) Adapted with permission.[91] 2018, American Chemical Society. h) Reproduced with permission.[92] 2017, American Chemical Society. i) Reproduced with permission.[93] 2016, Wiley. j) Reproduced with permission.[94] 2017, American Chemical Society.

They report that graphene conductance (synaptic weight) can be tuned by trapping charge concentration at the AlO_x defect sites. Shi et al. reported a multilayer h-BN artificial synapse shown in **Figure 2.7f**, where the boron vacancies and metallic ions conductive pathway are responsible for modulating the device conductance.[87] The switching mechanism in metal/h-BN/metal synapses and CBRAM are similar, except the cation pathways formation in h-BN synapses is performed by the generated boron vacancies.[87]

Sun et al.[91] reported on a mono layer MoS_2 artificial synapse where synaptic weight modulation is based on the Joule heating effect (**Figure 2.7g**).[91] The conductance modulation in this device is dependent on the residual temperature increase caused by the resistive heating in MoS_2. Arnold et al. [92] developed a MoS_2 synapse where the adsorbed gas molecules and/or by MoS_2/SiO_2 interface and/or by defects in MoS_2 led charge trapping shown in hysteresis behavior in its I-V characteristics. Tian et al. demonstrated synaptic plasticity through charge transfer between the BP channel and the native PO_x functional layer in a BP based artificial synapse (**Figure 2.10i**).[93]. Finally, a BP/SnSe junction-based synaptic device was reported by Tian et al. as well that uses emulation of synaptic plasticity bt taking advantage of the tunable electronic properties of the BP and SnSe heterojunction(**Figure 2.7j**).[94]

2.5 Device Metrics Comparison

The development of synaptic electronics for implementations in the hardware of ANNs has progressed rapidly in the last few years. Because of their similar requirements in terms of programming and retention, emerging NVM technologies such as PCM and RRAMs have attracted much attention as possible candidates for building large-scale artificial neural networks. It should also be noted that an ideal synaptic device also possesses additional properties that are traditionally

less important (even irrelevant) in NVM applications such as good precision levels, high linearity, and symmetrical conductance responses. In addition, the specific requirements of synaptic devices are likely application-dependent, e.g., good retention is more important for a neural network that is trained offline, where online training requires better endurance and more linear behaviors due to the frequency of weight updates. Hence, emerging devices based on electrochemical reactions or 2D materials, which are not originally geared towards NVM applications, are also being extensively investigated by researchers and reviewed here.

Figure 2.8 summarizes the performances of a number of synaptic devices based on PCM, the three types of RRAMs (CBRAM, filamentary, and interfacial), electrochemical devices, and 2D materials in terms of crucial device metrics such as precision (# of states), switching energy, operating speed, the degree of linearity and symmetry of the conductance response, reliability (device variations, endurance, and retention), and the potential in spatio-temporal dynamics for SNNs. PCM synapses have fast speed, decent precision and adequate endurances; but their non-linear behavior, high RESET power, and large variation from the melt-quench process may hamper their performance. RRAMs offer CMOS compatibility, simple structure, and good scalability; however large device variation from the stochastic nature of the switching mechanisms can limit learning accuracy in large-scale neural networks. In addition, the abrupt SET process in CBRAM and filamentary RRAM can result in non-linear and asymmetrical response and also limit the device precision. Interfacial RRAMs have shown more gradual switching with better precision, though this is typically accompanied by a smaller dynamic range and a slower switching speed compared to other RRAM devices.

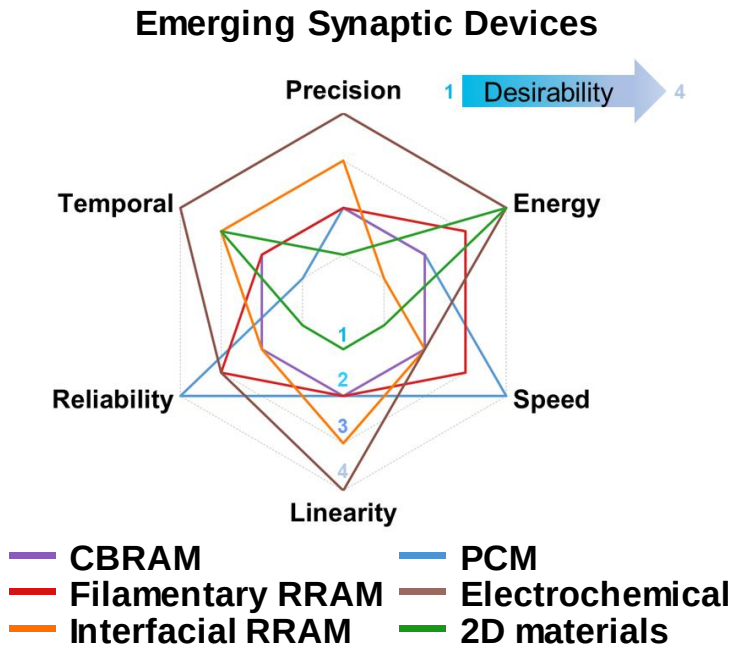

Figure 2.8. Radar graph comparing the device metrics among PCM (blue lines), CBRAM (purple lines), Filamentary RRAM (red lines), Interfacial RRAM (Orange lines), Electrochemical synaptic devices (brown lines) and 2D Materials-based synaptic devices (green lines), in terms of precision, energy, speed, linearity, reliability, temporal characteristics. 1 to 4 represents the four different degrees of desirability, respectively.

Nanoscale devices based on electrochemical reactions offer good precision and linearity, as well as potentially low device variations. The low operating speed due to the slow ionic movement may be mitigated by device scaling. But reactive cations (e.g. Li^+) can cause contamination issues and may be incompatible with CMOS devices. 2D materials offer good scalability and the potential for in-memory computing as 2D devices are considered promising candidates for next-generation logic devices. However, researchers are still looking for an ideal

switching mechanism in 2D devices as current 2D synapses based on charge trapping offer limited precision, non-linear and asymmetrical responses, and poor reliability.

To implement timing-based plasticity for SNNs, many of the existing approaches employ complex timing circuitry and are thus not scalable. Techniques that utilize secondary effects with built-in temporal components such as electrostatic gating, Joule heating, or charge trapping are preferred.

While most existing synaptic devices cannot fulfill the requirement of an ideal electronic synapse at the moment, there is an enormous opportunity to achieve orders of magnitude improvement in computation capability and energy efficiency through the hardware acceleration of ANNs. By combining the electrochemical reaction effect with 2D materials, an artificial electronic synapse optimized for the hardware acceleration of ANNs can be achieved. In the following chapters, an electrochemical synapse that utilizes a topological insulator $(Bi_xSb_{1-x})_2Te_3$ film will be presented that exhibits programmable spatiotemporal dynamics, high precision, linear and symmetric weight response, and impressive scalability in terms of energy and speed, thus demonstrating a promising potential to lead to the hardware acceleration of truly neurorealistic ANNs with superior cognitive capabilities and excellent energy efficiency.

3.0 Two Dimensional Electrochemical Graphene Synapse

3.1 Synaptic Device Fabrication

We choose to study graphene because graphite is a common anode material for Li ion batteries (LIBs), confirming graphene's compatibility for Li intercalation [95]. We fabricated the graphene device through mechanical exfoliation (Kish graphite from HQ Graphene) and Graphene flakes were transferred onto SiO_2 (285 nm) on Si substrate. The metal contacts (80 nm of Cu) were defined by electron-beam lithography and deposited through electron-beam evaporation. All devices were annealed under forming gas flow (Ar/N_2) at 300 °C for an hour to improve contact resistance and remove tape residues and other surface adsorbents. Our electrochemical cell (**Figure 3.1**) is similar to a planar "nano-battery" with the graphene and lithium iron phosphate (LFP) serve as the working and reference electrodes, respectively. The electrolyte ($LiClO_4$ in PEO) was prepared following a recipe similar in Ref 1. We mixed $LiClO_4$ (0.3 g) and PEO (1.0 g) powders with 15 ml of anhydrous methanol. We then stirred the mixture overnight at 50 °C. After applying the PEO electrolyte, the devices were annealed at 370 K in a vacuum probe station (~1.0×10^{-5} mbar) for two hours to eliminate residual moisture before electrical measurements.

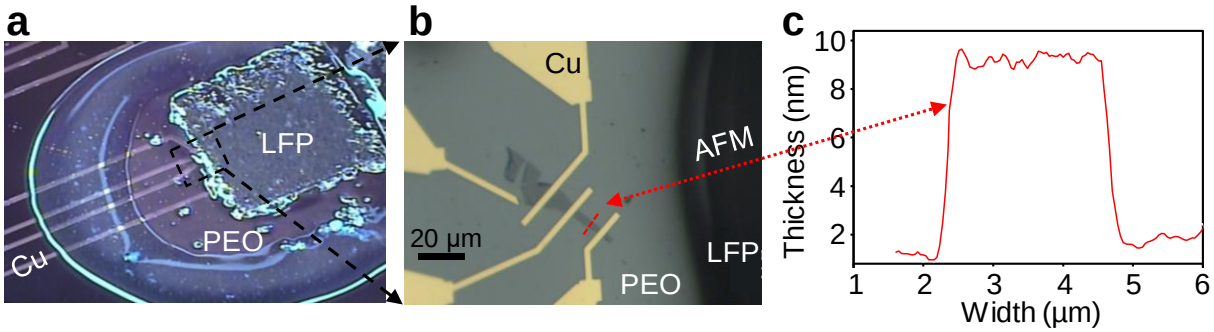

Figure 3.1 | Electrochemical graphene synaptic device structure. a, An optical image of an as-fabricated electrochemical graphene synapse comprising of graphene, LFP, electrolyte, and metal contacts (80 nm of Cu). A zoomed-in optical image (**b**) and the atomic force microscopy (AFM) measurement (**c**) of the graphene synapse. The scale bar is 20 μm.

3.2 Electrical and Electrochemical Characterization

Electrochemical characterizations were carried out with a SP-200 BioLogic workstation. During galvanostatic discharge measurements, a constant 100 pA charging/discharging current was applied with the graphene connected to the working electrode and the LFP connected to the counter/reference electrodes. Electrical and pulse measurements were performed with a Keithley Semiconductor Parameter Analyzer (4200-SCS) with pulse measuring units (PMUs). All electrical measurements were performed in a vacuum probe station (~10^{-5} Torr).

3.3 Advantages of Using Graphene

The 2D nature of graphene makes it an excellent host for accommodating guest ions between its layers.[95] We choose to intercalate Li ions in few-layer graphene in this work because (1) 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 (LIBs)[95-97] and (2) Li ions have been reported to have an unusually high diffusion coefficient (7×10-5 cm2s-1) in bilayer graphene at room temperature,[98] suggesting that a fast switching speed is possible. While Li ions can potentially be a source of contamination for Si devices[99] if not handled and contained properly, it is a good material to start with for our proof-of-concept demonstration because of its well-known electrochemical behaviors. In addition to graphene's compatibility with the CMOS technology (good thermal stability and 2D nature)[100] and its potential for application in flexible electronics (excellent mechanical properties and ultrathin layers),[101] graphene can be easily integrated with emerging 2D electronics to form a densely connected logic and memory network, enabling a new paradigm in computing architectures and breaking the existing "memory bottleneck".

3.4 Advantages of Our Electrochemical Based Synapse

Schematics of a biological synapse and our electrochemical graphene synapse are illustrated in **Figure 3.2a-b**, where the graphene channel resistance represents the synaptic weight, and the plasticity of the graphene synapse can be modulated by the signal from the reference electrode (lithium iron phosphate, or LFP), representing signals from presynaptic neurons. The solid electrolyte ($LiClO_4$ in poly (ethylene oxide) (PEO)) allows ionic exchange to occur between

graphene and LFP. A major advantage of our approach is that we can achieve the reversible and precise tuning of the graphene device conductance by controlling the concentration of Li ions in graphene[95, 96] to mimic synaptic plasticity. In addition, unlike conventional memory technologies, the programming and storage mechanisms in the electrochemical synapse are decoupled,[71] allowing us to achieve both low-power switching and good retention properties. **Figure 3.2c** shows a schematic of the energy barriers for typical thermodynamic memory (ferroelectric and resistive random-access)[102] and kinetic memory (flash and PCM)[103] devices. The barrier heights for these memory devices are the same during the write and read operations, imposing a tradeoff between a low programming power (write) and good retention (read) in conventional memory technologies. In our electrochemical device, the programming (write) and retention (read) barriers can be decoupled by a circuit switch between graphene and LFP (**Figure 3.2d**). During the write operation, the circuit is closed, which allows both electrons (through the external circuit) and ions (through the electrolyte) to be exchanged between graphene and the reference electrode under a small bias. During the read operation, where the graphene resistance is sensed by a low-voltage pulse across the graphene channel with the circuit between graphene and LFP open, the electron-blocking electrolyte prevents any electrochemical reactions and therefore maintains the state of the graphene synapse.

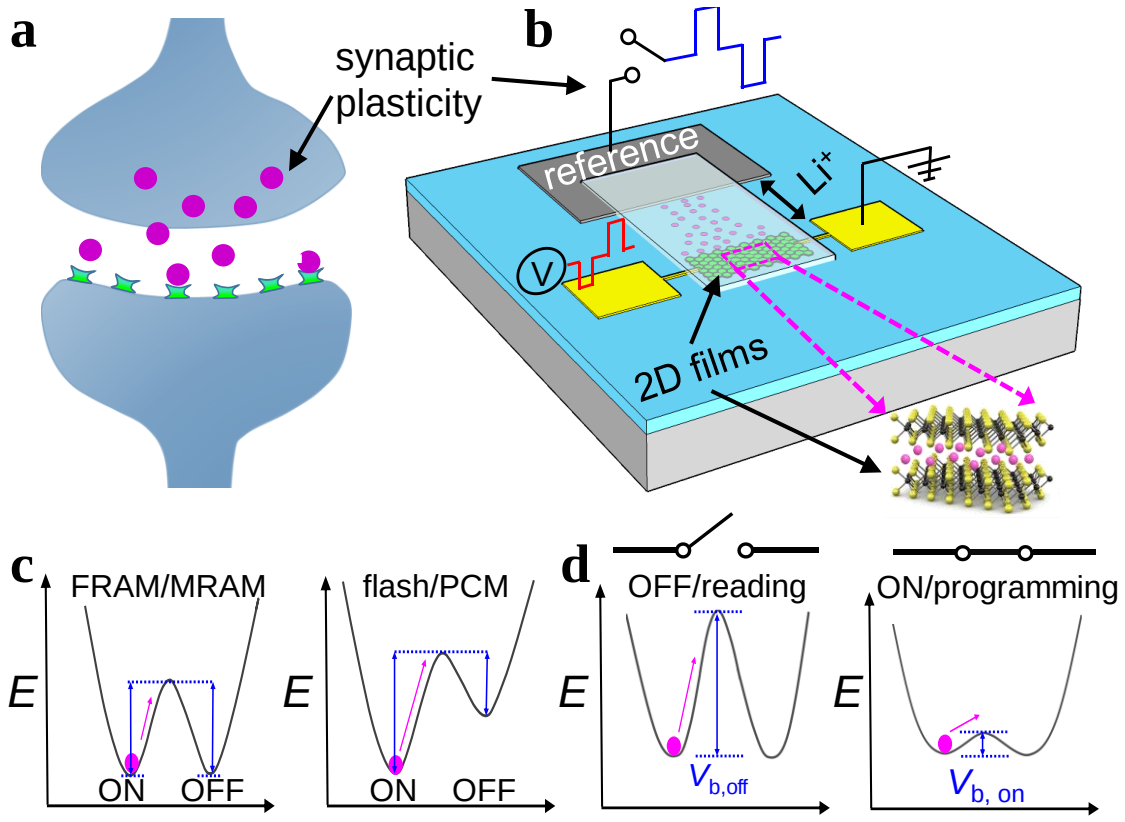

Figure 3.2. Electrochemical graphene synapse. a) Schematic of a biological synapse. b) Schematic of our graphene synapse. The synaptic weight is encoded into the channel conductance of the graphene device. We can modulate the synaptic plasticity (channel device) in our device by adjusting the Li ion concentration between graphene layers via electrochemical intercalation. c) Schematic of conventional memory technologies, where the energy barriers for programming and retention are the same, thereby creating a tradeoff between low-power writing and good retention. d) Schematics of the energy barriers for our electrochemical synapse during retention (read) and programming (write). The retention and programming mechanisms are decoupled through a circuit switch. The high $V_{b,\,off}$ barrier during the open circuit read operation ensures good retention, while the low $V_{b,\,on}$ barrier during the closed circuit write operation allows low-power programming.

3.5 Electrochemical Intercalation's Working Mechanism

Electrochemical intercalation has been the cornerstone of the LIB industry for decades, but it has only recently been employed as a technique for engineering the properties of 2D materials through charge transfer, impurity scattering, and band gap engineering.[104] Similar to a Li ion battery, our device has a working electrode (2D materials), a reference electrode (Li metal), and a liquid electrolyte to facilitate ionic exchange between the working and reference electrodes. The direction of the ionic movement in the system is dictated by the direction of current flow in the external circuit, as indicated by the red and blue arrows in **Figure 3.3**. To lower the resistance of the graphene synapse through intercalation, we used an external current to drive both negatively charged electrons and positively charged Li^+ ions from the reference electrode (LFP) to graphene, indicated by the pink arrows. When we reversed the polarity of the external current, Li ions was driven out of the graphene (de-intercalation), indicated by the blue arrows, and the graphene resistance was increased. Governed by charge neutrality, for every electron, e^-, flowing into/out of graphene in the external circuit, a corresponding Li^+ ion enters/leaves graphene through the electrolyte. This process allows us to precisely modulate the Li ion concentration in our graphene synapse (thus, its synaptic weight) by controlling the total amount of charge (= current × time) transferred into/out of the graphene.

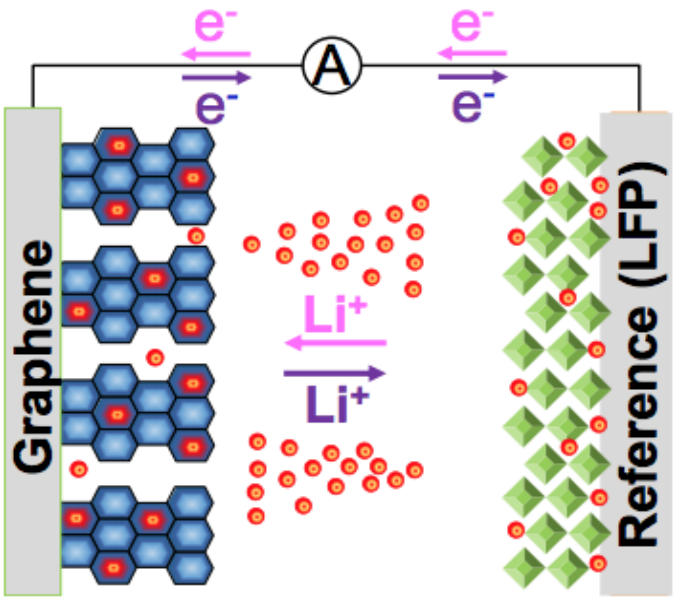

Figure 3.3. Schematic of the reversible electrochemical Li intercalation process. The pink (purple) arrows indicate the intercalation (de-intercalation) process, where both electrons, e^-, and Li^+ ions flow into (out of) the graphene working electrode.

3.6 Galvanostatic Discharge of Graphene

We performed galvanostatic charge/discharge measurement on our graphene sample with an electrochemical workstation (BioLogic SP-200). The constant charging/discharging current was 100 pA. As the Li concentration increases upon intercalation, the electrochemical potential of graphene vs. Li/Li^+ decreases. We also observed various voltage plateaus as a result of the staging phenomena, consistent with previous reports on Li intercalation in graphite [95.96,105]. Even though intercalation is initiated from the edges of the 2D crystal [106], previous studies [96,107] have shown that all of the intermediate intercalated compounds (e.g. LiC_{72}) have an approximately uniform distribution of intercalant (Li). This has been attributed to the formation and rapid growth of the

intercalate nucleus [96,107]. This is advantageous for the development of graphene synapse since it ensures that graphene devices with similar dimensions will have the same electrochemical response when they are intercalated to the same stage.

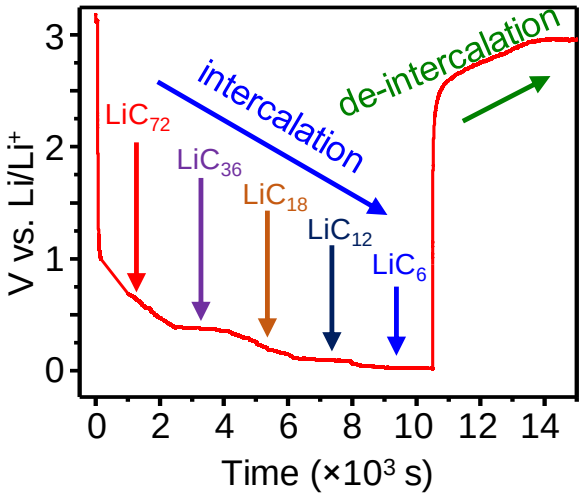

Figure 3.4. The galvanostatic discharge process of graphene. The voltage plateau indicates the formation of different intermediate intercalated compounds: LiC_{72}, LiC_{36}, LiC_{18}, LiC_{12}, and the stoichiometric limit LiC_6. The blue and green arrows indicate the direction of intercalation and de-intercalation, respectively.

Figure 3.4 shows the galvanostatic discharge measurement of a 20-nm graphite (see **Figure 3.6b** inset) thin film during Li intercalation. The intercalation process was controlled by a BioLogic SP-200 electrochemical workstation, and the discharge current was kept at 100 pA. We monitored the electrochemical potential of graphene by measuring the voltage difference between graphene and the reference LFP electrode. We observed multiple plateaus in our measurement, which corresponded to the formation of the intermediate compounds LiC_{72}, LiC_{36}, LiC_{18}, LiC_{12}, and, lastly, the stoichiometric limit LiC_6, consistent with similar measurements reported in the

literature.[95, 96, 108] Even though intercalation is initiated from the edges of the 2D crystal,[109] previous kinetics studies[96, 109] have shown that all of the intermediate intercalated compounds (e.g. LiC_{72}) have an approximately uniform distribution of intercalant (Li). This has been attributed to the formation and rapid growth of the intercalate nucleus.[96, 109] This is advantageous for the development of graphene synapse since it ensures that graphene devices with similar dimensions will have the same electrochemical response when they are intercalated to the same stage.

3.7 In-situ Raman Spectra of Graphene at Different Intercalated States

We also performed in situ Raman measurements to examine the different stages of graphene intercalation compounds, as illustrated in **Figure 3.5**. The Raman spectroscopy was performed using a Horiba LabRam instrument with a 532 nm laser. With a 100× objective, the laser spot radius was ~300 nm. The absorbed laser power was kept low (<20 μW) to avoid laser heating. Our pristine graphene displayed good film quality with no D band. Upon Li intercalation, we first noticed a slight upshift of the G band from 1580 cm^{-1} to 1605 cm^{-1}, which was likely due to the doping effect of Li intercalation.[106, 108, 109] Subsequently, the G band split into two peaks, which was possibly due to a weakening of the C-C bond upon the influx of Li ions.[108, 109] Eventually, the G band vanished into the background noise at the LiC_6 stage. The G band reappeared upon de-intercalation, thereby confirming the reversible nature of the electrochemical process.

We used Raman spectroscopy to probe the change in bonding, strain, and doping during the intercalation process. **Figure 3.5a** shows the in-situ Raman measurement setup, where we connected the portable electrochemical workstation to our graphene synapse in the Raman stage.

Before Li intercalation, our graphene sample shows good film quality as we observe no D band in our Raman spectrum **(Figure 3.5b)**. In the dilute stage (i.e. the initial intercalation stage before the formation of the first staging at LiC_{72}), we observed a significant upshift (from 1580 cm^{-1} to 1605 cm^{-1}) in the G band, suggesting doping from the charge transfer process [95, 105, 106]. We also noticed a small downshift of the 2D band, likely due to the increased tensile strain [105, 106]. Upon further Li intercalation, the G band split into two peaks in the LiC_{36} and LiC_{12} stages. This has been attributed to the difference between the graphene layers adjacent to Li layer to those next to the empty galleries [105]. The 2D band disappeared into the background level consistent with previous reports, which indicated that electron doping can cause a rapid decrease in intensity in 2D band [105, 110]. Upon Li de-intercalation, the G band re-appeared at 1580 cm^{-1}.

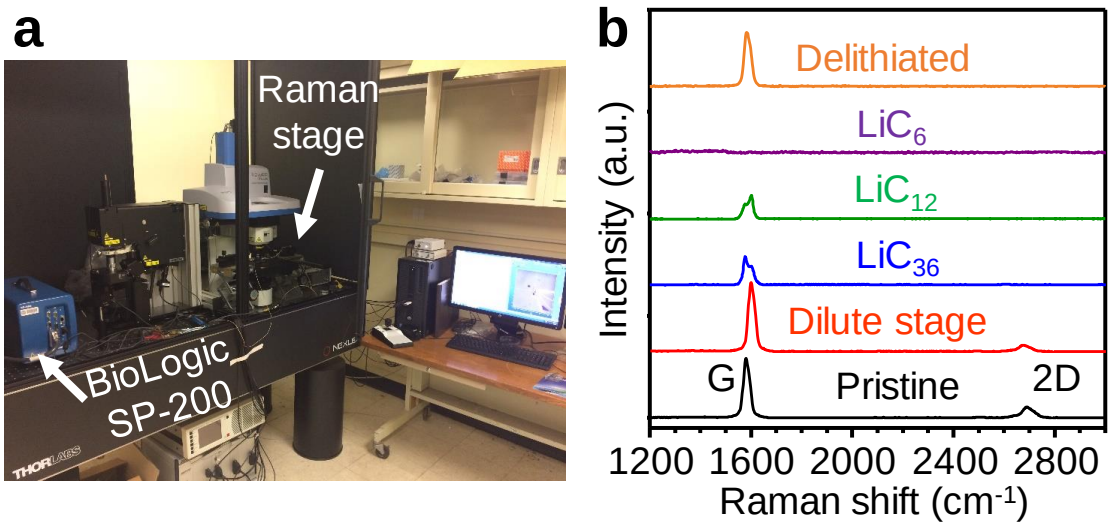

Figure 3.5. In-situ Raman measurement. a) Experimental setup for in-situ Raman measurement. b) Raman spectra of the graphene synapse at different intercalation stages.

3.8 Electrical Conductance Modulation

We performed direct current (DC) measurements to measure the electrical conductance of graphene as a function of its Li content, i.e., electrochemical potential. The electrochemical potential of graphene vs. Li/Li+ indicates its Li content. As shown in **Figure 3.6a**, in its pristine state, graphene has an electrochemical potential of ~3.0 vs. Li/Li^+. As we intercalate more Li+ into graphene, its potential decreases until it becomes fully lithiated (LiC_6) at 0.1 vs. Li/Li^+. Thus, when we varied its electrochemical potential by changing the voltage difference between graphene and the Li metal in **Figure 3.6b**, we observed an ~700% modulation in the synaptic weight (channel resistance) in our graphene synapse, which is significantly more than the ~150% change that occurs in biological synapses.[111] This large modulation in electrical conductance suggests that the electrical characteristics of the graphene electrochemical synapse is mostly dominated by the electrochemical state (i.e. Li concentration) of the device, hence making our electrochemical synapse more immune to the variabilities of graphene devices from the growth and/or fabrication processes and suitable for scaling up to dense neural networks. Looking into the future, our electrochemical synapse's tolerance for device-to-device variation in graphene will allow us to adopt the high-throughput, large-area chemical vapor deposition (CVD) growth of graphene thin films[113] for array-level integrations.

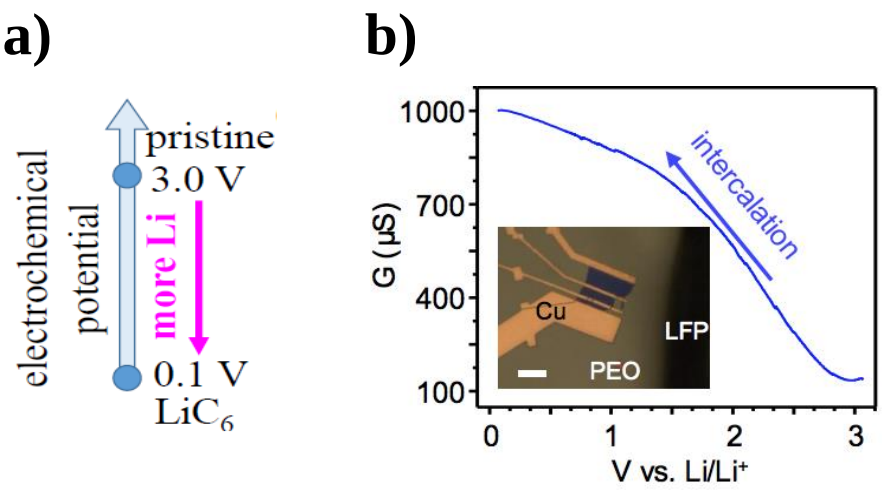

Figure 3.6. Electrochemical modulation of electrical conductance of graphene device. a, The electrochemical potential of graphene vs. Li/Li+ indicates its Li content. b, Modulation of the graphene conductance (synaptic weight) as a function of its electrochemical potential (Li concentration). The inset shows an optical image of the graphene synapse. The scale bar is 20 μm.

3.9 Neuronal Functionality of Graphene Synapse

3.9.1 Excitatory and Inhibitory Synaptic Response

We could mimic both excitatory and inhibitory synaptic responses with our 2D electrochemical synapse, as illustrated in **Figure 3.7a-b**, respectively. When we sent a single intercalation current pulse (50 pA, 10 ms) to a 8-nm thick graphene synapse (**see Figure 3.1b**) to emulate a presynaptic input, we observed an ~30 Ω (0. 67%) decrease in the channel resistance *R*, which mimics an increase in the synaptic weight of an excitatory synapse upon stimulation[111, 113], as shown in **Figure 3.7a**. The Δ*R* value gradually diminished to a stable value of Δ*R*' = -10 Ω (-0.22%). This result is similar to the short-term potentiation effect in a neural network, where the synaptic strength changes upon initial stimulation. This learning effect (increase in synaptic weight) gradually fades over time but still results in a permanent, smaller change in the synaptic strength.[114]

Physically, the initial large change in the channel conductance of graphene likely results from a combination of the doping effect of Li intercalation[106] and the electric double layer (EDL) gating effect[115] from the intercalation pulse. The EDL gating effect, which originates from the introduction of a large amount of charge at the graphene/electrolyte interface by the applied voltage, is capacitive in nature and thus dissipates[115] when the applied pulse is off. The doping effect of Li intercalation is non-volatile[106] and contributes to the lasting, smaller increase in the channel conductance. **Figure 3.7b** shows the inhibitory synaptic behavior, where the synaptic weight (channel conductance) decreases upon presynaptic stimulation (de-intercalation pulse). Similarly, we observed a large initial increase (Δ*R* = + 30 Ω) in the channel resistance, which

settled down to a stable value at $\Delta R' = +10\ \Omega$, representing the non-volatile effect from Li deintercalation.

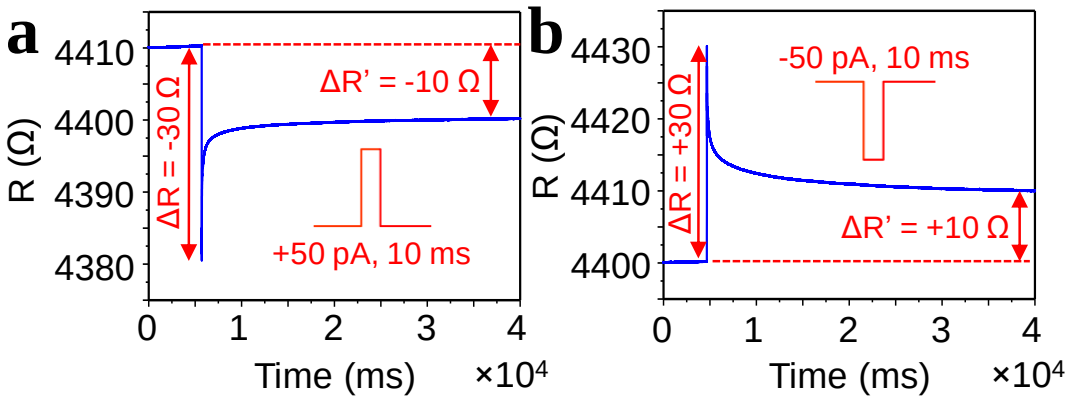

Figure 3.7. Synaptic plasticity. Excitatory a) and inhibitory b) synaptic behaviors of an electrochemical graphene synapse. The decrease (increase) in the channel resistance of graphene upon an intercalation (de-intercalation) pulse (50 pA, 10 ms) mimics an excitatory (inhibitory) synaptic response.

3.9.2 Programmable Long-Term Potentiation (LTP) and Long-Term Depression (LTD)

Long-term potentiation (LTP) and depression (LTD), where the synaptic strength can be permanently improved and diminished through repeated stimulation, are two essential neuronal functions in learning.[116] In **Figure 3.8**, we demonstrated the occurrence of LTP and LTD in our 2D electrochemical synapse through the application of a series of intercalation and de-intercalation pulses (50 pA, 10 ms). We were able to program over 400 distinct states into our device, suggesting that an extremely large number of non-volatile states are available for computations in our synapse. The synaptic response was linear and reproducible, which is desirable for system-level integration.[117] The R-squared value of the linear fit of the conductance response between 10 to

90% lithiation stage is 0.994, indicating a good linearity of the conductance response. This is expected because the increase in conductance depends on the amount of charge transferred into (out of) the graphene through the intercalation (de-intercalation) pulse.[71] This relationship is evident in our results, where ΔG is linearly dependent on the amplitude of the applied pulse. We studied the effect of the amplitude of the programming pulse on the change in synaptic weight (graphene channel conductance) on our electrochemical synapse by applying a series of 10-ms intercalating pulses with different amplitude (10 to 3000 pA). After each pulse, we applied a de-intercalating pulse to switch the graphene synapse back to its original state before the next measurement. We observed a linear relationship between the change in synaptic weight ($\Delta G/G$) with the pulse amplitude as shown in **Figure 3.9**. This is likely due to the doping effect from intercalation [95, 106]. From charge neutrality, the amount of intercalant (Li ion) and the associated charge transfer to the graphene through the electrolyte should equal to the same amount of electrical charge passed to the graphene in the external circuit (pulse amplitude × duration). This result suggests that we can further refine the step size (ΔG) and increase the memory density in our electrochemical synapse by modulating the amplitude of the applied pulses and our electrochemical synapse is highly scalable with excellent tunability.

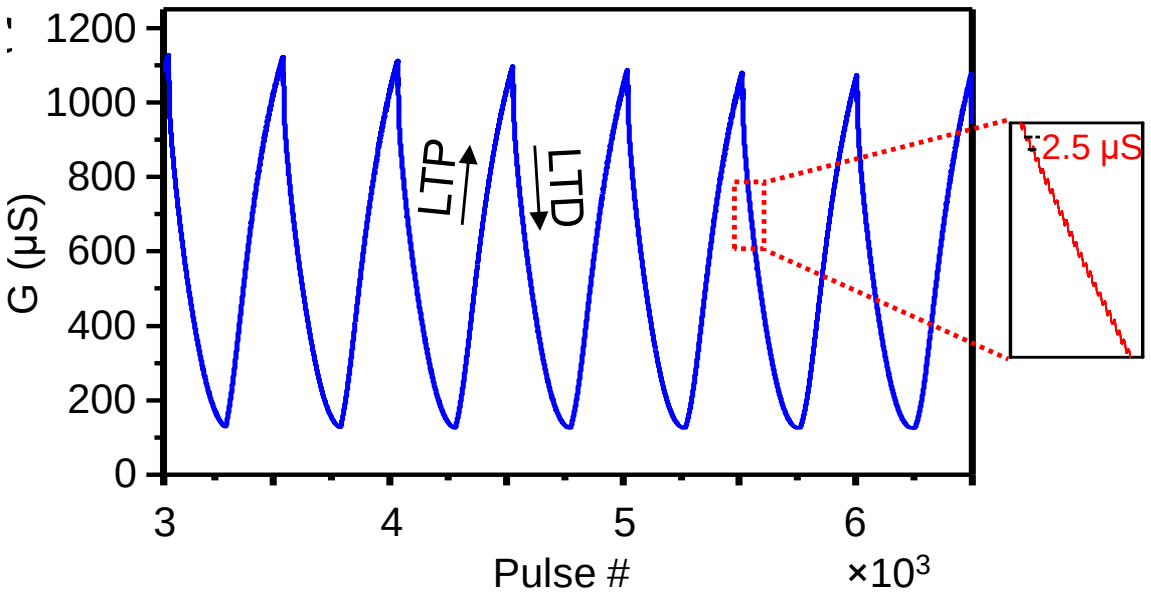

Figure 3.8. Synaptic plasticity. Demonstration of long-term potentiation (LTP) and depression (LTD) in our 2D synapse through controlled intercalation. We programmed >400 distinct states into our electrochemical synapse through a series of intercalation/de-intercalation pulses. The synaptic response is symmetric, linear, and repeatable.

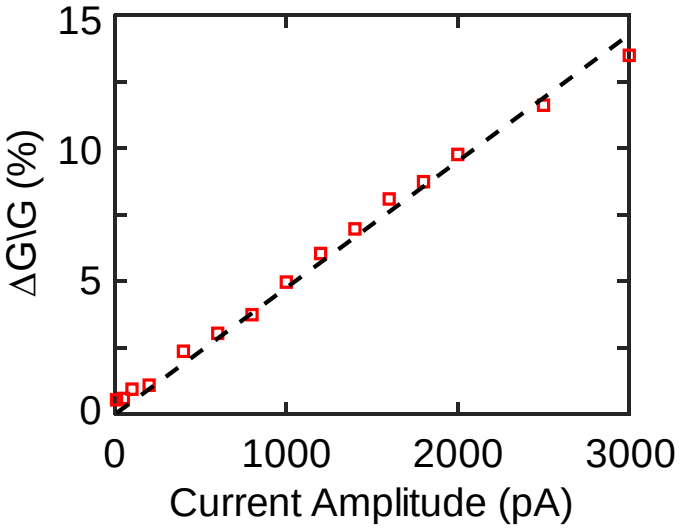

Figure 3.9. Pulse measurement. The change in synaptic weight (ΔG/G) as a function of the applied pre-synaptic pulse.

3.9.3 Spike Timing Dependent Plasticity (STDP) of Graphene Synapse

Spike timing dependent plasticity (STDP) plays an important role in learning and memory in biological neural networks.[114] As illustrated in different forms of STDP shown in **Figure 3.10a**, an STDP scheme governs how the spike (presynaptic signal) timing difference (Δt) determines the amplitude of the change in the synaptic weight (Δw).[11] We demonstrated the STDP capability in our electrochemical synapse through a two-pulse programming scheme, where the change in graphene conductance (ΔG) depends on the timing difference between the two pulses (Δt). As shown in **Figure 3.10b**, when the timing difference between the two intercalation pulses was small, we observed a large increase in the channel conductance (ΔG), similar to the strong memory effect observed for two temporally close events. When the timing difference was large, the change in channel conductance was subsequently smaller, analogous to the scenario in which the memory effect is weak when two events are temporally far apart.

Interestingly, the conductance change, ΔG, followed an exponentially decaying trend with three characteristic time constants, τ_1 = 22 ms, τ_2 = 315 ms and τ_3 = 19 s, similar to the timescale of biological synapses.[113] We speculate that the smaller time constants (τ_1 and τ_2) are related to the diffusion process during intercalation and thus should depend on the dimension (width and length) of the graphene synapse. The third (and much larger) time constant (τ_3), which may rise from the diffusion process from LFP to graphene, is needed to better fit the tail end of the decay, as we illustrated in **Figure 3.11**. We observed a power law dependence of τ_1 and τ_2 on device dimensions, as predicted by the Einstein diffusion equation $\tau_D \approx L^2/(2D)$. Using a diffusion coefficient, D, of ~ 4×10^{-6} cm^2/s for Li ions in graphene[99] and lateral dimensions of 4 μm (width) and 15 μm (length) for the graphene device, we estimated the average diffusion time ($\tau_D \approx L^2/2D$) for our electrochemical synapse to be ~20 ms and 280 ms in the width and length directions,

respectively, consistent with our measured time constants (22 ms and 315 ms). These result are consistent with switching times reported by van de Burgt *et al.* for the organic electrochemical synapses[71] and suggest that we can engineer and optimize the timing of the STDP for our 2D electrochemical synapse by altering the device geometry.

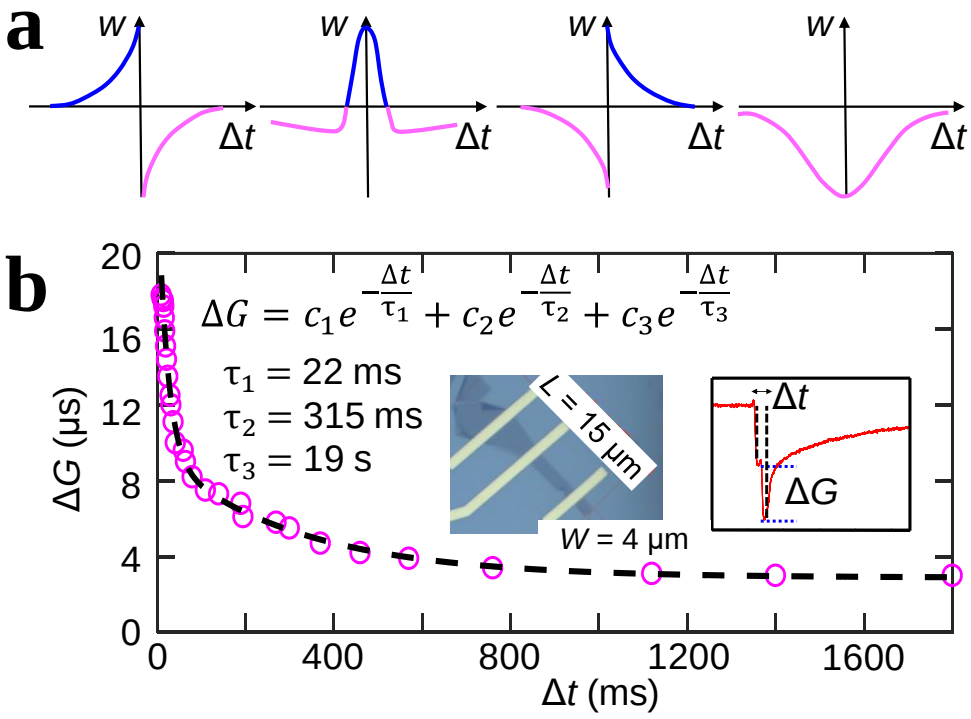

Figure 3.10. Spike timing dependent plasticity (STDP). a) Different forms of STDP in biological synapses, where the change in synaptic weight (w) depends on the timing difference (Δt) of synaptic events. b) In our graphene synapse, the change in resistance (synaptic weight) depends on the timing difference between programming pulses, mimicking the STDP scheme. The time constants τ_1, τ_2, and τ_3 are ~22 ms, 315 ms, and 19 s, similar to biological synapses. The ratio of the time constants $\tau_2/\tau_1 = 14.5$ is similar to the ratio of length/width for graphene $(L/W)^2 = 14.1$, as predicted by the Einstein diffusion equation $\tau_D \approx L^2/(2D)$.

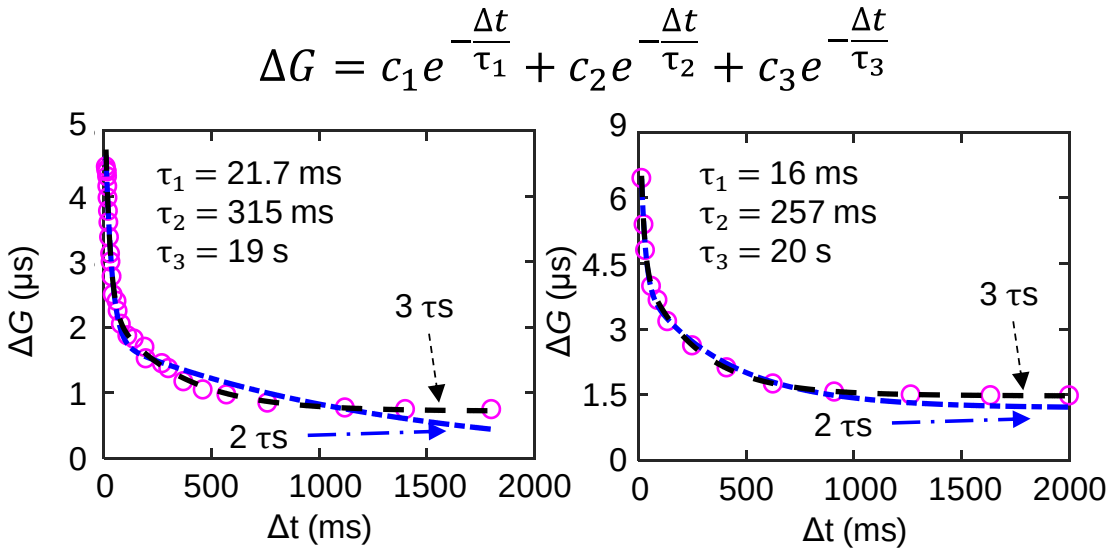

Figure 3.11 Spike timing dependent plasticity with one and two time scales. The change in synaptic weight (ΔG) as a function of the applied pre-synaptic pulse. The blue dashed time is an exponential decay fit with two time constants while the black dashed fit has three time constants, which fits the tail end of the decay better. We speculate that the two smaller time constants are related to the device dimension while the third, larger time constant depend on the distance between the LFP and the graphene.

3.9.4 Scaling Performance of Graphene Synapses

We studied the scalability of our 2D electrochemical synapse over a wide range of dimensions. The size and geometry of the graphene synapse determine the density of our analog memory as well as the programming speed and energy. As illustrated in **Figure 3.12a**, the time constants for STDP in our graphene synapse exhibited a power law dependence ($y = cx^{\alpha}$ with $\alpha = 1.9$) over the device dimensions, as governed by Einstein's diffusion equation, $\tau_D \approx L^2/2D$. We project that the diffusion time for a graphene flake with 30 nm × 30 nm lateral dimensions can be

as fast as ~100 ns (10 MHz), which is comparable to that of dynamic random-access memory (DRAM) devices.[121]

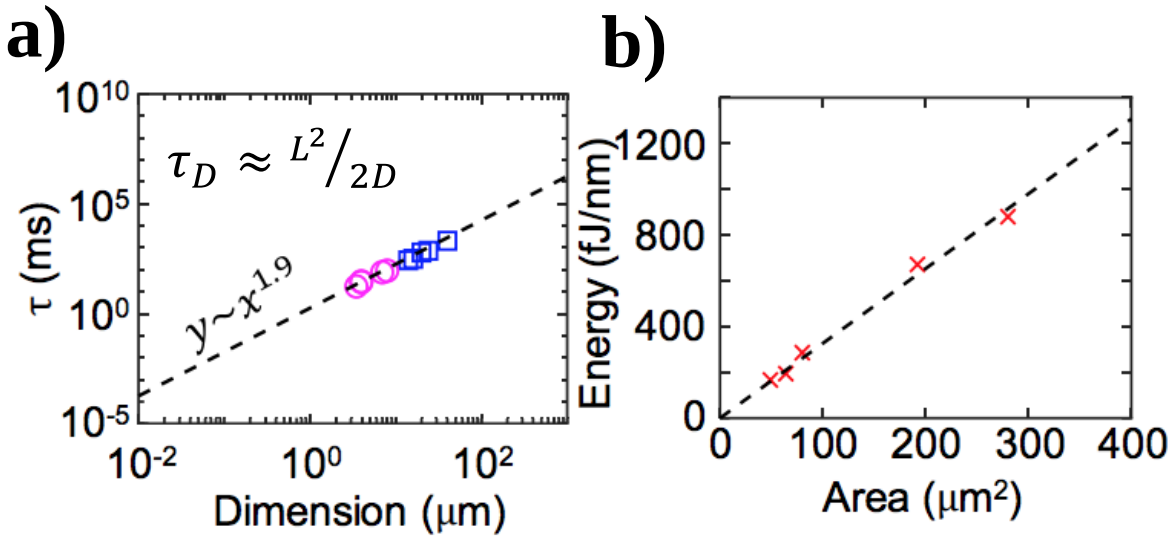

Figure 3.12. Scaling performance. a) Scaling trend of STDP time constants τ_1 (pink circle) and τ_2 (blue square) as a function of device dimension (W or L). b) Scaling trend of switching energy (normalized by thickness) as a function of flake area.

We also plotted the switching energy (normalized to the thickness) needed to induce a 0.25% change in the graphene synapse as a function of the area of the flake in **Figure 3.12b**. The switching energy ($E = I \times V \times t$) depends on the programming voltage (V) and the amount of charge ($I \times t$) transferred. The programming voltage is determined by the electrochemical potential of graphene and remains approximately constant. The total charge should scale linearly with the area of the flake (assuming uniform thickness). The linear dependence of E on the area confirms our analysis and suggests an excellent scaling potential of the electrochemical synapse. The switching energy per synaptic event for our smallest device (3 μm × 12 μm × 3 nm) is 500 fJ,

comparable to the switching energy of a biological synapse. We project that the switching energy of a submicron device (30 nm × 30 nm × 1 nm) can be as low as ~4 aJ, which is several orders of magnitude lower than the switching energy of biological synapses[111] and will significantly improve the energy efficiency of artificial neural networks.

3.9.5 Reliability Performance of Graphene Synapses

The graphene synapse displays good retention and endurance behaviors. The device conductance (synaptic weight) was stable under a 0.1 V reading bias (across the graphene channel) for more than 13 hours after a synaptic switching event, as shown in **Figure 3.13a**. The graphene device was disconnected from the reference electrode (LFP). We measured a 3.2% decrease in conductance at the end of the 13-hour stress test, consistent with the self-discharge effect observed in LIBs, where the conductance of the fully charged batteries typically decreases by ~3-5% in the first 24 hours due to impurities in the electrolyte and electrodes.[118] While this conductance drift is less than ideal for long term retention, it may be mitigated by adding a thin passivation layer on graphene, as it has been employed to reduce the self-discharge rate in LIBs.[119] Furthermore, this self-discharge rate in LIBs drastically slows down to approximately 1-2% per month thereafter,[121] projecting potentially good long-term retention properties for the graphene synapse.

To study the endurance performance of our graphene electrochemical synapse, we cycled our device between its two intermediate states by a train of 10 ms pulses with alternating amplitude (±50 pA). We observed two distinct states with no signs of degradation for 500 cycles (**Figure 3.13b**). The standard deviation for the two states are 0.047% (blue) and 0.045% (magenta), respectively.

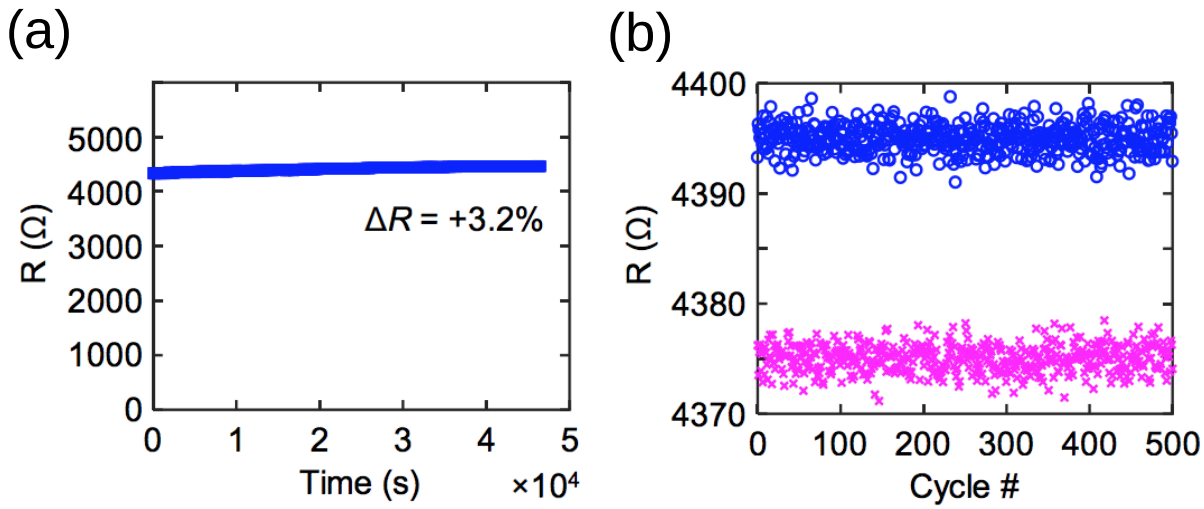

Figure 3.13. Reliability Performance. a) Retention. The graphene resistance (synaptic weight) as a function of time under a 100 mV bias. The graphene synapse was disconnected with the reference electrode. b) Device endurance. We cycled our graphene synapse for 500 times through pulse measurements with no signs of degradation.

4.0 Two Dimensional Electrochemical MoS_2 Synapse With Ultra-Fast Switching Speed And Low Programming Energy

4.1 Introduction

Neuromorphic computing has attracted remarkable attention as an emerging computational method for next generation computers. By designing new, and adapting existing, electronic devices the brain's neural functionalities can be mimicked in several ways through use of algorithms [21], architecture [28] and hardware [117]. The resulting neuromorphic system would offer huge improvements in computational performance as well as energy efficiency in deep learning algorithms when implemented in multi layered artificial neural networks (ANNs) as compared to traditional von Neuman machines [122]. However, despite the massive advancement of the field in recent years, artificial neural networks (ANNs) still pale in comparison to the human brain in terms of energy efficiency and scalability [123].

Traditional ANNs operate through a large number of conventional computing systems working in parallel. Ideally, high-level computations would be performed at an element (artificial synapse) which offers both data processing and storge at a single location, but von Neumann based ANNs still need to move data between processing and storage units [124]. As the field has matured, however, two main categories of networks have emerged: deep neural networks (DNNs) and spiking neural networks (SNNs).

Deep neural networks (DNNs) have pioneered significant progress in various fields including computer vision problems [125,126], speech recognition [127-132], and natural language processing [133]. However, training DNNs typically requires an enormous amount of

computational power due to the vast number of parameters that need to be tuned, as well as the massive datasets required to do so.

Spiking neural networks (SNNs) use signal spikes to process time-dependent information, a trait that plays a critical role in how the human brain operates [134]. These techniques, such as basic encoding of data as spikes, event driven computation, and temporal integration of signals have been mostly ignored in the development in DNNs. It would then seem that to have an efficient learning system that rivals the brain, demonstration and optimization of SNNs is of utmost importance, particularly for embedded applications with strict memory and energy constraints [135-141].

Application specific integrated circuit (ASIC) designers at IBM, Intel, and Manchester University developed TrueNorth [142], Loihi [143], and SpiNNaker [144], respectively, by implementing SNN techniques. Despite energy efficiency improvements of up to three orders of magnitude, these fully digital systems based on complementary metal-oxide-semiconductor (CMOS) transistors are not able to truly simulate the analog behaviors of biological synapses. Furthermore, if these systems were scaled to the same complexity level as the human brain, which consists of $\sim 10^{11}$ neurons and $\sim 10^{15}$ synapses [145], the huge amount of energy consumption required for these devices would inhibit any practical applications, indicating poor scalability.

Three emerging synaptic devices - resistive random-access memories (ReRAM) [56, 146-148], phase change memory (PCM) [40,149,150], and electrochemical metallization memory (ECM) [151,152] - have been explored as synaptic computing units to overcome this scaling challenge because of their analog response, fast operating speed, small footprint, and ability to be easily integrated into existing semiconductor fabrication processes. Despite these promising aspects, these technologies are hindered by basic limitations: poor reliability and high

programming current in PCM [153,154] and variabilities in both temporal and spatial properties, as well as the inherently nonlinear operating process of both ECM and ReRAM [145,152,155,156].

To address these issues, electrochemical three-terminal synaptic transistors have recently arisen as a promising substitute [24,70-72,157,158-161]. Their decoupled programing and storage mechanisms, as well as the precise modulation of channel conductance via ion insertion/extraction, results in linear and symmetrical switching, low programming power, low operating noise, high precision, and high endurances. Although polymer and titanium carbide MXene based devices have shown promising switching speeds down to sub microseconds, they suffer from challenges (e.g. low working temperatures, contaminations) that limit their potential to be implemented in ANNs [162]. These devices can also suffer from poor dynamic range (0.7 μS for 200 ns write pulse) and poor linearity for even 50 events, which limits their capability for ANN applications. Therefore, in order to fabricate a synaptic device which satisfies all the desired requirements for neuromorphic computing, different material systems need to be explored.

Two-dimensional (2D) van der Waal (vdW) materials have attracted a lot of attention in low dimensional electronic, optoelectronic, and piezoelectric applications due to their special features and tunable band structures. Among these materials, semiconducting transition metal dichalcogenides (TMDCs) have become a strong prospect for novel electronic applications because of their resistance to short-channel effects, remarkable gate controllability at nanoscale channel lengths, as well as superb electrical characteristics [163-168]. Molybdenum disulfide (MoS_2) is one of the most popular TMDC materials and its wide interlayer gap and scalable device size give it the potential to accommodate alkali ions like Na^+, K^+, and Li^+ [169,170]. With this in mind, MoS_2 can then be reversibly switched between a semiconducting (2H) and a metallic (1T) phase via intercalation of these alkali ions [106,171-173]. This phase transition, in addition to the

controllability of the intercalation process [174], makes MoS_2 a strong candidate for neuromorphic computing applications by providing another route to mimic analog synaptic activities.

Here, we present an artificial synaptic device with fast speeds (< 100 ns) and low energy consumption per event (< 4 fJ) based on MoS_2 thin films. Three terminal solid state synaptic transistors with two gates using 2D-MoS_2 have been investigated to determine their application for event based high performance learning systems. The short-term ionic gating effect in electrochemical transistors has been utilized in our recent work [175] to exhibit time dependent spatiotemporal dynamics for spike-based computations. The resulting SNNs, with their event driven nature and integration of temporal components, promise an exciting substitute for DNNs, with an outstanding energy efficiency and bandwidth, as well as a capability for cognitive computing by virtue of logical interface implementation to the existing hardware.

This MoS_2 based synapses utilize two different floating gates: Au for ionic gating, and LFP/Au for combined ionic gating and electrochemical intercalation processes. This enables the device to have both short- and long-term plasticity (STP and LTP) characteristics, which are used to demonstrate complex brain functionalities such as paired pulse facilitation (PPF), STP, LTP, and dynamic filtering. In addition, the conductance response is excellent in terms of linearity, symmetricity, and dynamic range. These controllable responses, as well as the flexibility offered by two different stimulus sources, make these synaptic devices superb candidates for future neuromorphic applications.

4.2 MoS_2 Device Fabrication and Working Mechanism

Schematics of a biological synapse and our electrochemical MoS_2 synapse with two side gates are shown in **Figure 4.1a,b** where the MoS_2 channel conductance constitutes the synaptic weight of the device. Lithium-ion phosphate (LFP), a Li^+ ion reservoir, and Au have been employed for presynaptic electrodes to modulate Li concentration in the MoS_2 channel via presynaptic electrical signals. It is worth mentioning that LFP has recently emerged as a cathode material in Li ion batteries (LIBs) because of its extended lifetime and large power density, as well as the fact that it is air-stable [176,177]. The reversible ionic exchange through a solid electrolyte $LiClO_4$ in poly (ethylene oxide) [24] between MoS_2 and the two side gates (LFP/Au and Au) leads to synaptic plasticity behaviors mimicking biological neurons.

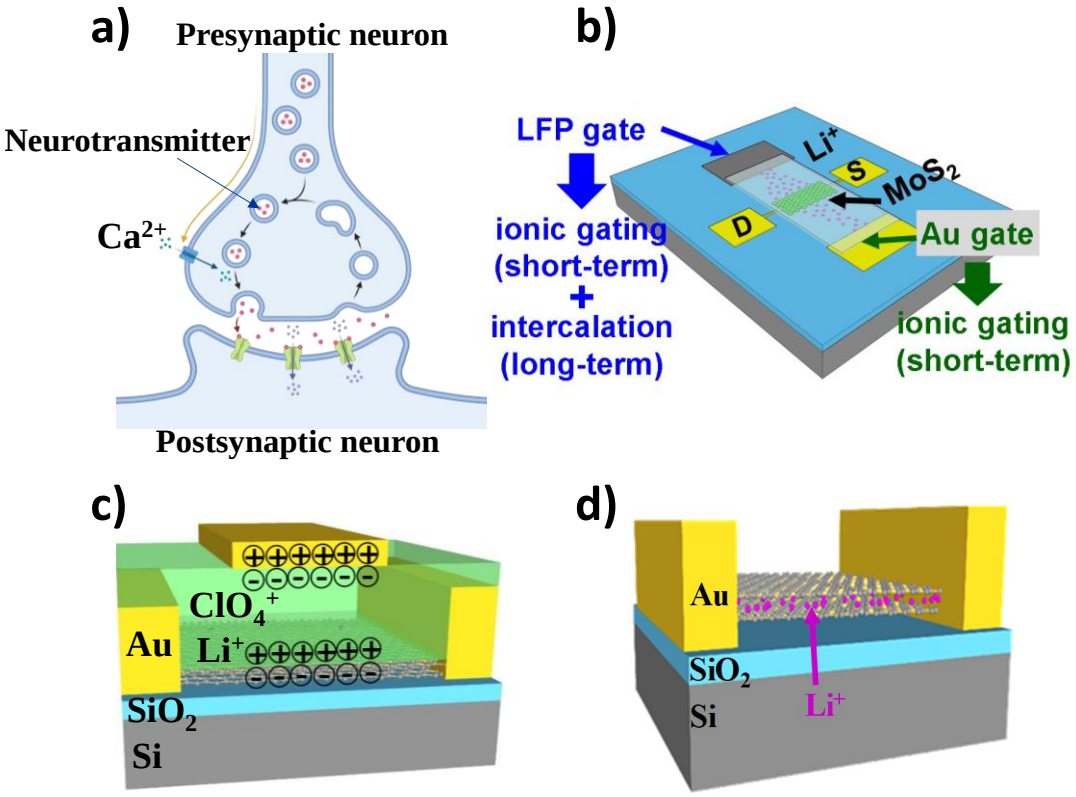

Figure 4.1. Electrochemical MoS_2 synapse. a) Schematic illustration of biological synapse. Created with BioRender.com. b) Schematic diagram of three terminal MoS_2 artificial synapse in which the two Au and LFP gate terminals represent postsynaptic terminals and drain and source, emulating postsynaptic electrodes. The channel conductance (synaptic weight) of the MoS_2 synapse can be modulated by Li^+ ion migration to the channel upon application of electrical signal to the gates. c) Short-term doping due to accommodation of positive Li^+ ions at the interface of the electrolyte and MoS_2 and formation of electrical double layer gating effect by applying voltage pulse to the Au gate d) Long-term doping due to Li ion electrochemical intercalation and permanent accommodation into the MoS2 layered structured.

As illustrated in **Figure 4.1c,d**, the plasticity of the MoS_2 channel is driven by ionic gating and electrochemical intercalations processes caused by the Au and LFP gates, respectively. We

demonstrated that plasticity due to ionic gating (Au) is short term, regardless of the gate stimulation condition, such as duration, magnitude, frequency, and number. However, the LFP gate can cause short- or long-term plasticity in accordance with the presynaptic stimuli. Physically, these behaviors are due to the combined electrostatic (from ionic gating, **Figure 4.1c**) and electrochemical (from Li intercalation, **Figure 4.1d**) effects. When a potentiation pulse is applied, mobile Li^+ ions are continuously inserted into the MoS_2 lattice (electrochemical intercalation) and at the same time, the positive Li^+ ions also accumulate at the interface of the electrolyte and MoS_2, inducing more electrons in MoS_2 through the field effect (electrostatic ionic gating). The electrochemical MoS_2 synapse allows a precise change in channel conductance by controlling the Li^+ ion content in the channel either by electrostatic effect or reversible Li^+ intercalation/de-intercalation. Such decoupled programming and storage mechanisms in this device have led to ultra-low energy consumption per switching and good retention properties.

4.3 Electrochemical and Electrical Characterization

The galvanostatic discharge measurement of a 10-nm thick MoS_2 flake was performed using a BioLogic SP-200 electrochemical workstation with a constant discharge current 100 pA. The measured voltage difference between the LFP and MoS_2 shows the electrochemical potential of MoS_2 during the Li intercalation process. As shown in **Figure 4.2a**, when Li ions were intercalated into the MoS_2, the electrochemical potential decreased, leading to high controllability of the channel synaptic weight and thereby allowing for analog synaptic behavior for ANN applications. We observed a plateau during the galvanostatic discharge at 1.1 V vs. Li/Li^+. This plateau indicates a semiconducting (2H) to metallic (1T) phase transformation which is consistent with previous studies [178,179].

The electrical characteristics (Id-Vg) of our MoS_2 synapse was measured for pre- and post-lithiation processes under V_{DS}= 0.5 V (**Figure 4.2b)**. After lithiation, we observed that the MoS_2 current level experienced an increase and the on/off ratio dropped significantly from 75,000 to 2.8, indicating successful lithiation, and thus successful phase transition, from semiconducting (2H) to metallic (1T).

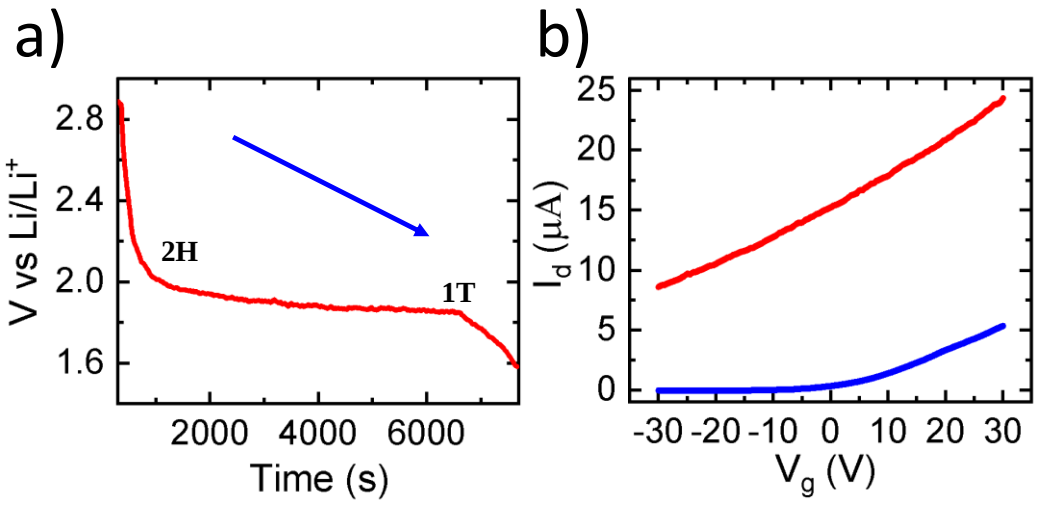

Figure 4.2. Electrochemical Li intercalation in MoS_2 synapse. a) The galvanostatic discharge curve in MoS_2 vs. Li/Li^+ with LFP as the reference electrode. The plateau indicates phase transition from semiconducting (1H) to metallic (1T) during intercalation process. b) The electrical I-V curve measurement before and after intercalation. Substantial decrease in ON/OFF ratio after intercalation process indicates permanent Li accommodation in MoS_2 layers.

4.4 Fast Speed Measurement Setup

To characterize these synaptic devices using ultra short pulse stimuli (pulse width < 200 ns) and observing how fast the Li^+ ions can travel, we developed a high-speed measurement setup to operate fast programming and the storage process, as shown in **Figure 4.3**. In this setup we used

an oscilloscope (OSC) with a fast-reading rate for reading and a current amplifier in series with the OSC to amplify the signal coming from the device under test (DUT).

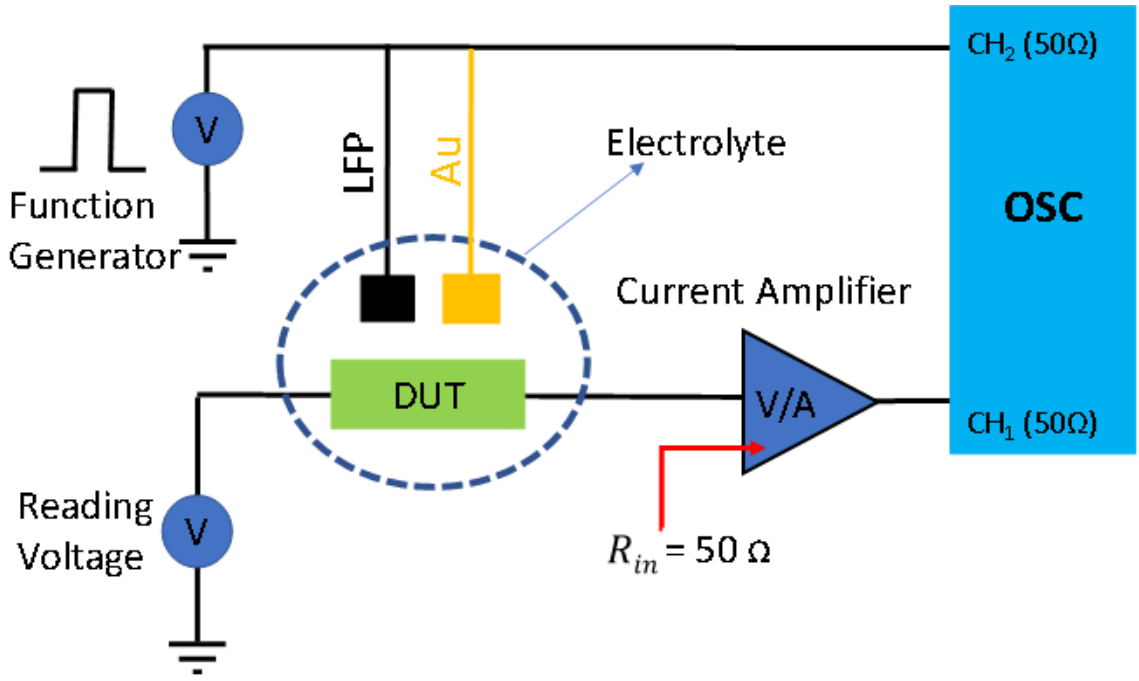

Figure 4.3. Fast speed measurement setup. We used 4200 SCS for applying a 100 mV DC bias signal to the drain terminal and reading the output signal via oscilloscope (OSC) with high horizontal resolution in the range of nanoseconds desirable for fast speed reading.

4.5 Short-Term Potentiation and Depression Induced by Ionic Gating and Electrochemical Intercalation Mechanisms Using Fast Single Stimuli

In the brain, the activity dependent synaptic plasticity in the form of STP and LTP plays a crucial role in brain functionalities in terms of computation and memory [180, 181]. The influx of Ca^{2+} to the presynaptic membrane triggered by arriving stimuli causes neurotransmitter liberation and enhancement of the synaptic weight, which leads to STP. However, this temporal synaptic

enhancement will be diminished in tens of milliseconds due to the spontaneous extraction of Ca^{2+} from the postsynaptic membrane. By applying a positive presynaptic voltage pulse (+1V, 100 ns) to the Au gate, the Li^{+} ions are driven to the channel surface through the ion gel electrolyte. This increase in ion concentration in the channel surface introduces more electrons in the MoS_2 (**Figure 4.4a, b**), increasing the channel conductance. When the pulse is removed, the ions retreat back out of the channel and into the electrolyte (**Figure 4.4c**) and the conductance returns to its original state, demonstrating short term potentiation (STP), as shown in **Figure 4.5 (a)**. This short-term effect is due to the Li^{+} ions donating excess electrons and inducing an electric double layer (EDL) gating effect [72,182]. By changing the voltage pulse polarity and driving ClO_4^{-} to the channel surface **(Figure 4.4d)**, the conductance will be decreased due to the induced holes in the channel, demonstrating short-term depression (STD), as depicted in **Figure 4.5 (b)**.

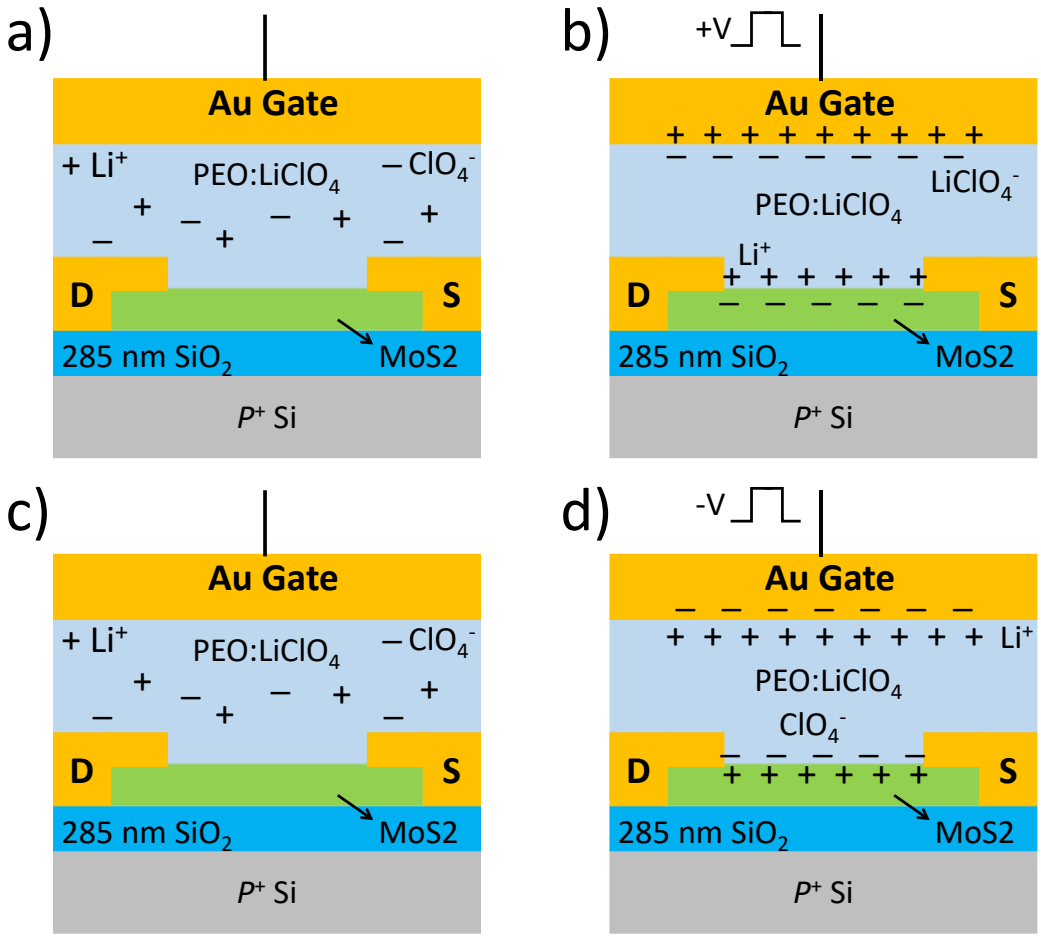

Figure 4.4. Schematics of ion gated MoS2 synapse with three terminal transistor structures where the voltage pulse is injected to the side Au gate through the electrolyte. a) At Vg open, positive (Li^+) and negative ($LiClO_4^-$) ions are randomly distributed and the MoS2 channel is in its pristine state. b) At Vg > 0, the positive Li^+ ions induce n-type doping due to EDl and improve the channel conductance c) after removing the pulse, the transient formed EDL will be disappeared and the MoS2 returns to its pristine state d) At Vg < 0, the negative $LiClO_4^-$ ions induce p-type doping, reducing channel conductance.

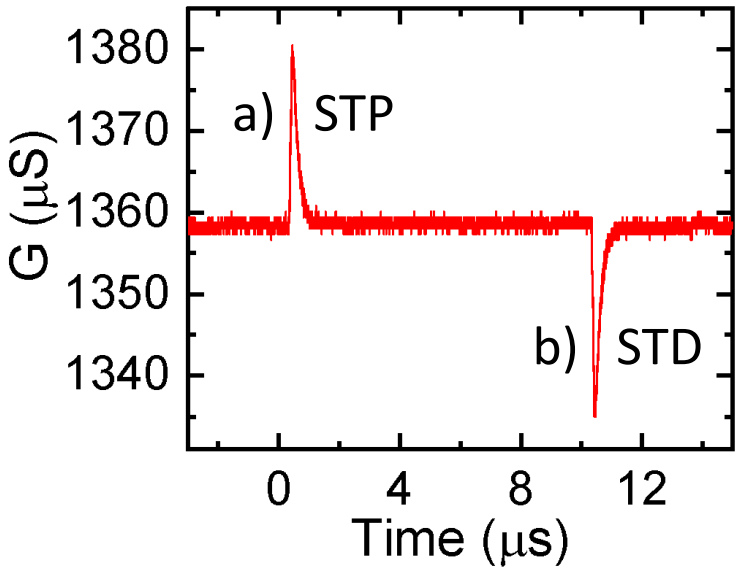

Figure 4.5. Short term plasticity. (a) short term potentiation due to EDL by applying a (1V, 100 ns) potentiation voltage pulse to the Au gate terminal. (b) Short term depression due to EDL by sending a (-1V, 100ns) depression voltage pulse to the Au gate terminal.

4.6 Electrochemical Intercalation Process with Long-Term Doping in MoS2 Synapse

In the electrochemical intercalation (E.I.) process **(**Equation 4.1**)**, Li^+ ions are driven into the vdW interlayers, not just the channel surface. Once the pulse is removed, they are expected to not retreat back into the ion gel spontaneously and to exhibit long-term plasticity effects. However, short term plasticity can be observed following the positive programming short pulse (1V, 100ns) to the LFP/Au gate. This implies that short pulse widths do not give the intercalated ions enough kinetic energy to diffuse deeper into the bottom layers of the MoS_2 vdW material. However, as shown in **Figure. 4.6a**, when the response behaviors due to EDL and E.I. gates are compared after removal of the input voltage pulse, the time constant of the E.I. is larger than what is seen from the EDL effect, confirming that there are a limited number of Li^+ ions intercalated into the MoS_2

layers. The Li ion diffusion time measured is 10 ns **(Figure 4.6b)**, showing the electrolyte used here is a good interface material for Li^+ ion exchanges with high travel speed. Furthermore, the larger conductance change in E.I. confirms additional contributions from EDL mechanisms; however, the shallowness of the ions still allows the ions to readily diffuse back into the electrolyte on the pulse end, making the effect short term.

$$LiFePO_4 \rightarrow xLi^+ + xe^- + FePO_4^- \quad (4.1)$$

$$MoS_2 + xLi^+ + xe^- \rightarrow Li_xMoS_2$$

However, for much longer injected pulses (~900 μs) long term potentiation (LTP) effects were observed from the LFP/ Au gate, as shown in **Figure 4.6c**. For these pulses the intercalated Li^+ ions are no longer able to be naturally released from the vdW layers and so cause permanent change in synaptic weight. The same pulses at the Au gate still demonstrate short term plasticity, showing that the Li^+ ions retreat into the ion gel when the gate bias is removed. We were able to fit the exponential decay for the both I.G and E.I responses using one and two term time scale equations, respectively. The results indicate an ionic gating effect for the Au gate and combined ionic gating and electrochemical intercalation for the LFP/Au gate due to contribution to the lateral geometry of the device in the process in addition to the channel surface.

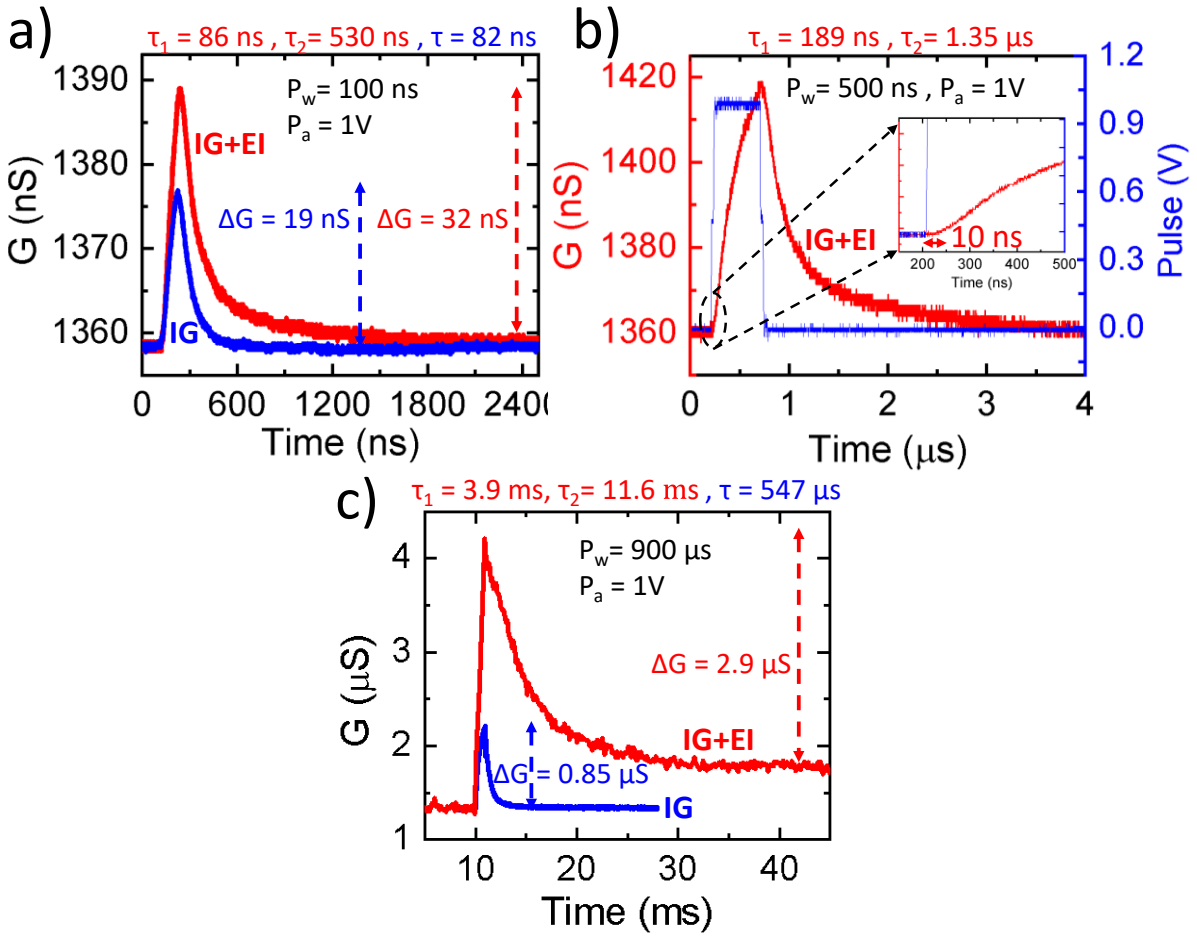

Figure 4.6. Synaptic plasticity. a) Upon application of short stimulation (1V, 100 ns) to the side gates, we observed short term potentiation due to EDL for the Au gate and combined EDL and electrochemical intercalation for the LFP/Au gate. The larger conductance and two term time constants after fitting indicate an E.I. effect. b) Mobile Li^+ ions diffusion time from electrolyte toward MoS_2 channel c) By applying a longer pulse (900 µs) to the side gates, the Au gate resulted in short-term doping independence of the pulse condition, and the LFP/Au gate caused long-term doping due to the permanent Li^+ ions accommodation in the MoS_2 layers.

4.7 Synaptic Weight Modulation Through Short Pulse Width/Amplitude Variations

The channel weight change (ΔG) is tunable for these synaptic devices by changing the presynaptic signals, similar to the process in biological synapses. This behavior is demonstrated for the LFP/Au gate via changing the pulse width and amplitude. **Figure 4.7a, b** shows ΔG as a function of pulse width up to 400 ns (fixed pulse amplitude 1 V) and pulse amplitude up to 2.5 V (fixed pulse width 100 ns). Both gates show close to linear relations between ΔG and pulse width/amplitude. The corresponding energy consumption per spike as a function of pulse width and amplitudes was also measured and shown in **Figure 4.7c, f**. This ultralow energy consumption, down to ~ 700 aJ, is much smaller than the amount used in biological synapses. In addition, this energy consumption can be further lowered by reducing the MoS_2 geometry, as we have shown for our graphene synapse.

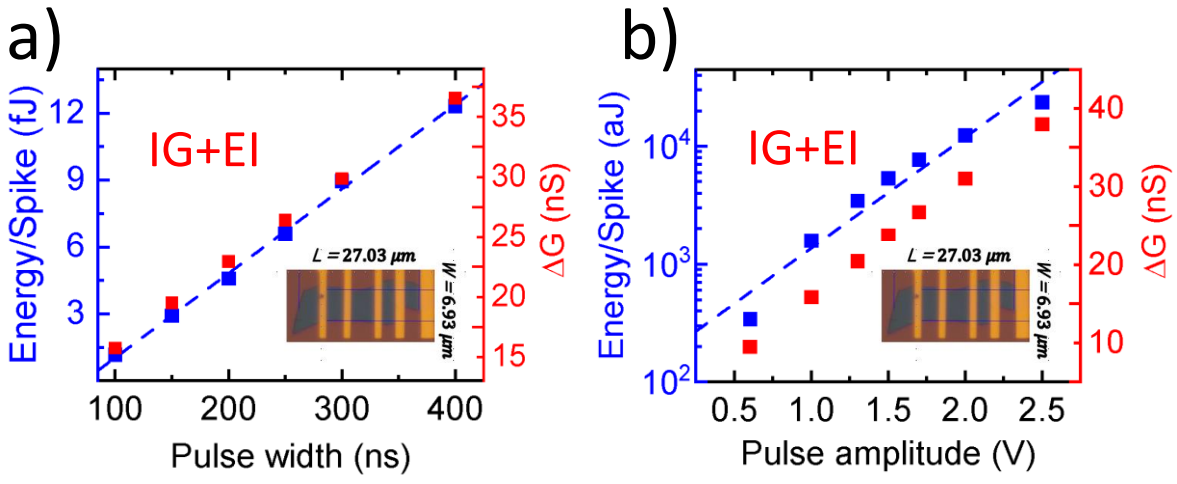

Figure 4.7. Synaptic weight modulation for MoS_2 Synapse. a) Synaptic weight change and corresponding energy consumption per spike by changing the pulse width and b) Amplitude showed our device synaptic weight mouldability desirable for DNN applications.

4.8 Long-Term Potentiation and Depression Emulation Using A Series of Potentiation And Depression Pulses

Two forms of synaptic plasticity in biological neurons that are essential for mimicking natural learning processes are long-term potentiation (LTP) and long-term depression (LTD). These persistent activity-dependent changes in the synaptic strength play a critical role in learning and successive memory formation by emulating the behavioral properties of associative learning and spatial/temporal encoding of learned events in the brain.

To mimic these brain functionalities in our MoS_2 synapse, a train of consecutive identically programmed positive and negative pulses ($|1V|$, 100ns) was applied to both floating gates, and the reversible LTP and LTD behaviors required for learning algorithms was demonstrated, as illustrated in **Figure 4.8a, b.** This reversible process for the Au gate is purely due to the electrostatic charge effect, indicating EDL for the Au gate independent of the pulse condition (**Figure 4.8a**). Similar to IG gate (Au) we observed short term effects for short pulse (100ns) in EI gate (LFP/Au), however, the response showed higher linearity and symmetricity, two important benchmarks for implementation of synaptic hardware into neural networks (**Figure 4.8b**). In addition, the dynamic range and number of distinct states are critical for DNN's recognition accuracy using artificial synapses. Here, EI process fulfilled these requirements by expressing 100 states and highly linear behavior compared to the IG procedure. To characterize our pulses, each potentiation/depression cycle was normalized and fit to the usual equations (Equation 4.2 and Equation 4.3):

$$G = G_{min} + G_o(1 - e^{-\nu_{pot} \cdot p}) \quad (4.2)$$

$$G = G_{max} - G_o(1 - e^{\nu_{dep}(p-1)}) \quad (4.3)$$

where p is the normalized pulse number and ν_{pot} and ν_{dep} are the nonlinearity fitting coefficients for potentiation and depression, respectively. For an ideal conductance response, both nonlinearity fitting values should be 0, indicating linear changes in conductance.

To show the superior performance of our artificial synapse, the device was programmed with 5000 up/down pulses ($|1V|$, 1 µs) over > 12 x 10^3 cycles, as demonstrated in **Figure 4.8c**. The average linearity values of 0.440 +/- 0.112 and 0.426 +/- 0.107 for potentiation and depression, respectively, high symmetricity of the switching and large number of possible states make our MoS_2 synapse an excellent candidate for neuromorphic computing. The same device was then characterized under the same conditions except with a longer (10 µs) pulse (**Figure 4.8d**). Results here showed that there is a trade-off between dynamic range and linearity, with the 10 µs pulse resulting in a larger dynamic range (~ 68 µS) and increased nonlinearity to 3.224 +/- 0.154 and 3.530 +/- 0.105 for potentiation and depression, respectively.

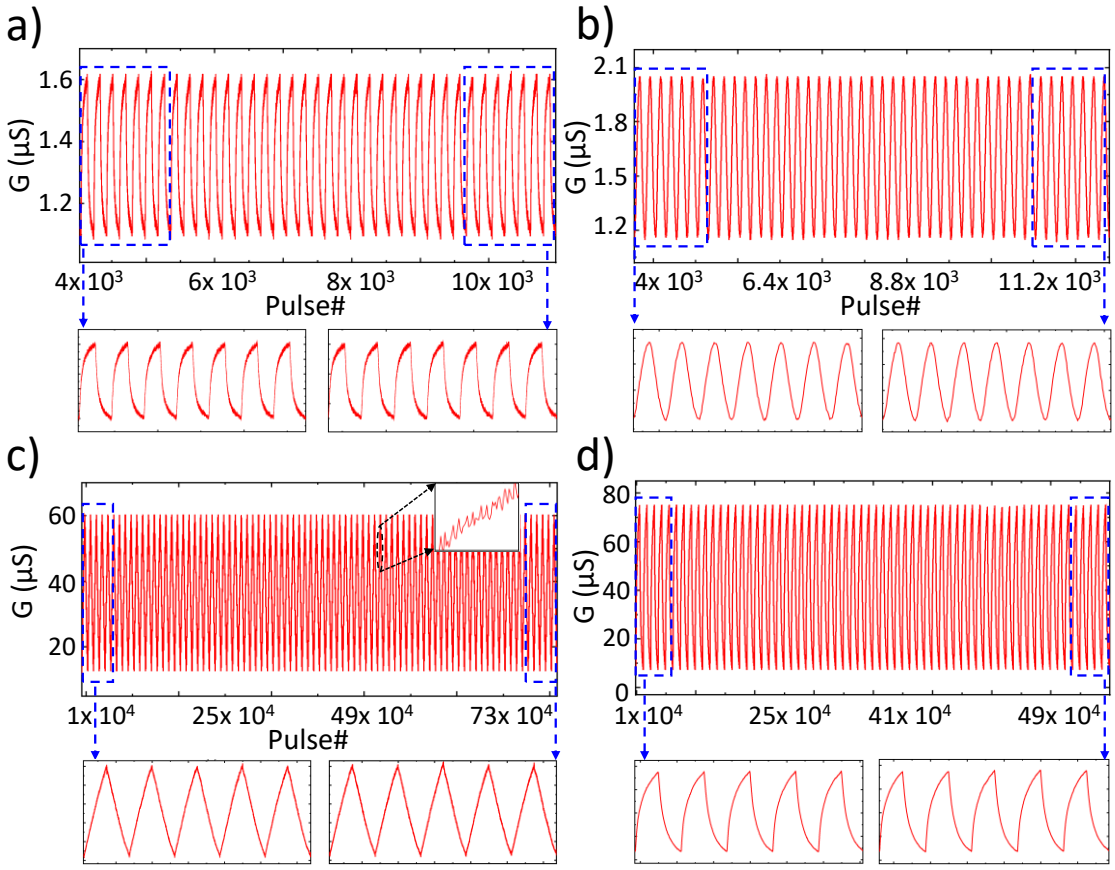

Figure 4.8. Long term synaptic plasticity. a) LTP and LTD emulation of our MoS_2 synapse due to EDL by applying a 100 of series pulses (1V, 100ns) to the Au gate. b) LTP and LTD due to combined EDL and EI effects by applying 100 consecutive pulses (100 ns) to the LFP/Au gate. Results showed a symmetric, reversible, and high linearity (compared to I.G.), which is preferable for DNN applications. c,d) Emulation of LTP and LTD introducing high precision level desirable for ANNs upon application of 1μs and 10 μs pulse widths.

When we plotted the energy per spike with respect to the pulse width (100 ns, 1 μs, 5 μs, and 10 μs) for EI responses, as depicted in **Figure 4.9**, a linear relationship between switching energy and pulse width was found. The projected ultra-low energy scale down to ~ 7 aJ for 1 ns pulse width reaffirms a super energy efficient MoS_2 synapse for neuromorphic applications.

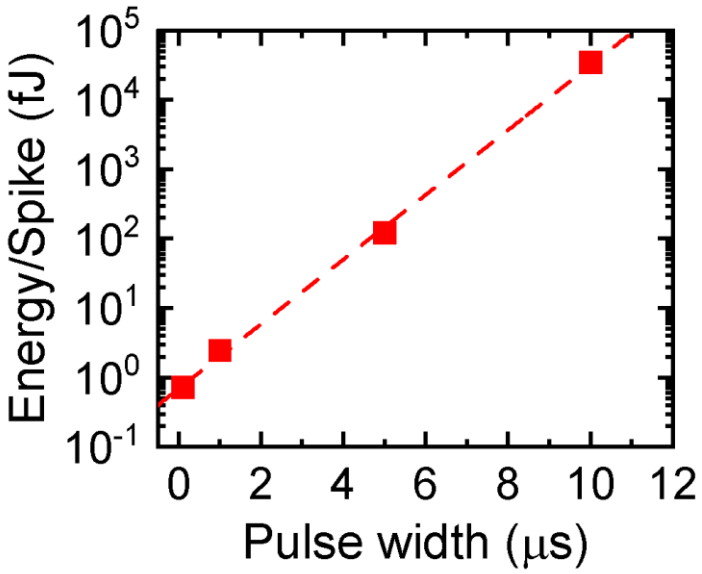

Figure 4.9. Programming energy per spike as a function of pulse width for electrochemical intercalation responses predicts low switching energy (~ 700 aJ) for MoS_2 artificial synapse.

The endurance characteristic of these synapses was also investigated for both the floating gates. As illustrated in **Figure 4.10a, b**, we observed smaller variations in conductance spreads for 100 ns pulse trains for the E.I. gate (3.421 +/- 0.0182 and 0.567 +/- 0.022 μS for G_{max} and G_{min}, respectively) compared to I.G. (1.6179 +/- 0.003 and 3.515 +/- 0.016 μS for G_{max} and G_{min}, respectively). The longer pulse widths 1 μs (61.109 +/- 1.664 and 12.104 +/- 0.255 for G_{max} and G_{min}, respectively) (**Figure 4.10c)**, 10 μs (75.991+/-0.714 and 8.203+/-0.314 for G_{max} and G_{min}, respectively) (**Figure 4.10d**), and 900 μs (64.386 +/- 0.298 and 14.545 +/- 0.234 for G_{max} and G_{min}, respectively) for the E.I gate exhibit low variations. The full characterization of each programming method is summarized in **Table 4.1** for clarity. The significantly low variation for the E.I. gate, even for large number of states (5000) and long pulses (900 μs), indicate a stable and controllable

electrochemical intercalation process as well as Li insertion/removal in MoS_2 vdW layers with high yield.

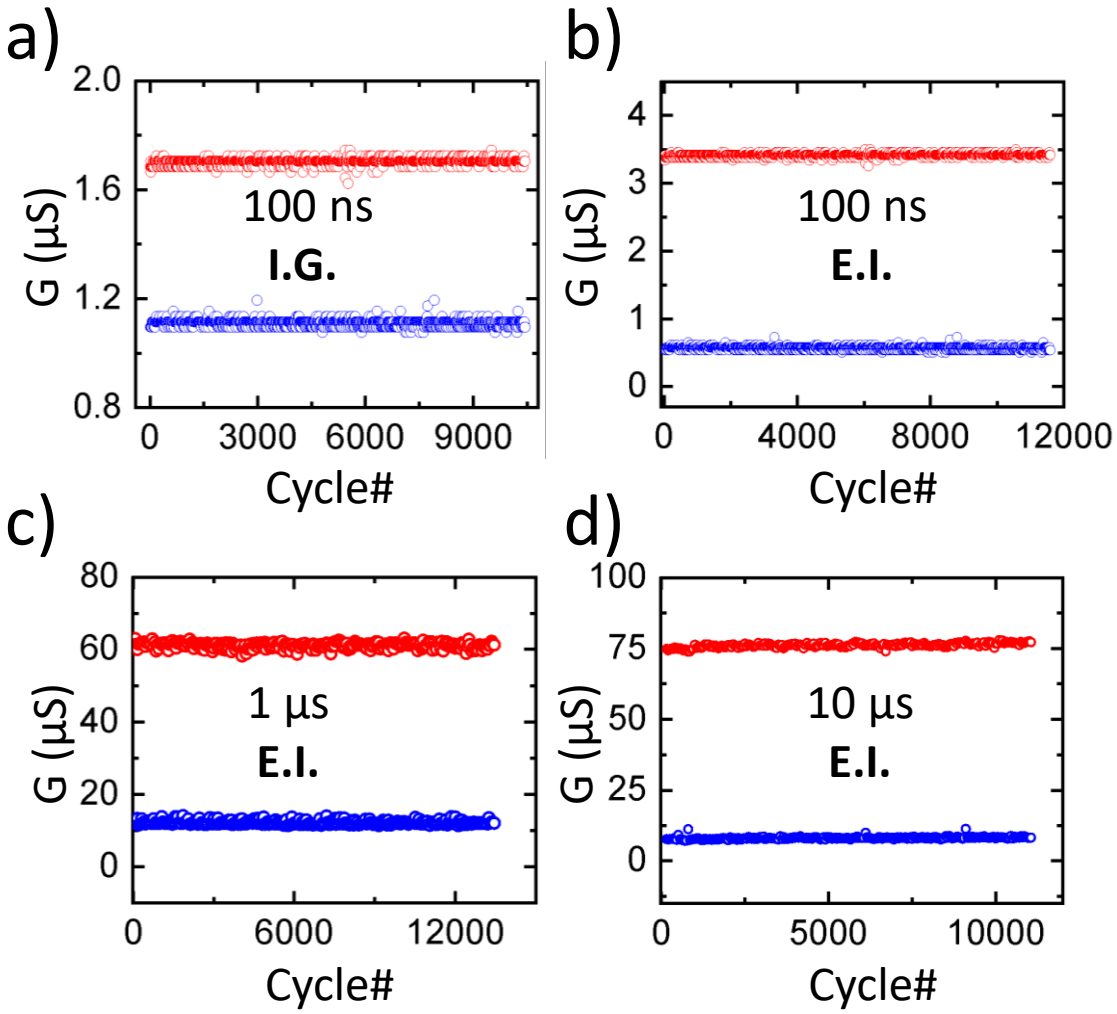

Figure 4.10. The endurance performance for both I.G. and EI a) Endurance for IG with 100 potentiation and depression pulses (1V, 100 ns). b) Endurance for E.I with 100 potentiation and depression pulses (1V, 100 ns). c) Endurance for E.I with 5000 potentiation and depression pulses (1V, 1 μs) d) Endurance for E.I with 5000 potentiation and depression pulses (1V, 10 μ s).

Table 4.1. Conductance response characterization for the both IG and EI gate terminals for different pulse widths.

	G_{max}	G_{min}	ν_{pot}	ν_{dep}
1 µs (E.I.)	61.108+/-1.664	12.104+/-0.255	0.441+/-0.056	0.426+/-0.0536
10 µs (E.I.)	75.991+/-0.714	8.203+/-0.314	3.224+/-0.077	3.530+/-0.0524
900 µs (E.I.)	64.386+/-0.298	14.545+/-0.234	2.502+/-0.049	2.047+/-0.047
100 ns (E.I.)	3.421+/-0.018	0.567+/-0.022	0.447+/-0.107	0.500+/-0.030
100 ns (I.G.)	1.618+/-0.003	1.093+/-0.016	4.211+/-0.718	4.512+/-0.822

4.9 Pair Pulse Facilitation (PPF) Emulation in MoS_2 Synapse

In neuroscience, neural facilitation, also known as paired-pulse facilitation (PPF), refers to an occurrence in which postsynaptic neuron potential invoked by two successive stimuli received by presynaptic neurons that cause an increase in presynaptic Ca^{+2} concentration. PPF as a form of short-term synaptic plasticity may be entangled in some neural tasks such as simple learning, information processing, and sound-source localization.

The success of PPF in our electrochemical MoS_2 synapse was demonstrated by applying consecutive paired pulses (100ns, 1V) to both side gates. As illustrated in **Figure 4.11a,** for IG, the channel conductance consistently decreases as the time interval of the two applied pulses increases, analogous to the manner of biological neurons. The synaptic weight change (ΔG) exponentially decayed with one time constant (τ_1= 102 ns) much faster than in biological synapses, which makes our artificial synapse unique in terms of programming speed in

neuromorphic computation applications. We speculate that the single time constant is related to the Li^+ ion interactions with the MoS_2 surface, resulting in an ionic gating effect. We repeated the PPF measurement with the same pulse parameter (100ns, 1V) on the EI gate and the result, as shown in **Figure 4.11b,** followed a double exponential decay function, similar to biological neurons, with two-time constants τ_1= 9 ns and τ_2= 131 ns. As we showed in our previous report [24] on graphene synapses, the two-time constants here are related to the MoS_2 device dimension, with the smaller and larger time constants resulting from the diffusion process during intercalation in width and length direction, respectively.

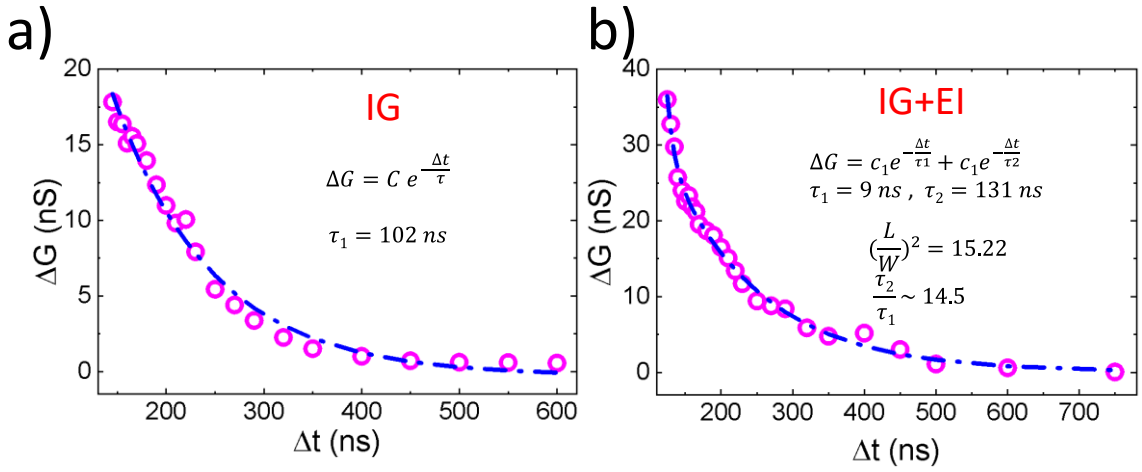

Figure 4.11. Paired pulse facilitation (PPF) where the synaptic weight can be modulated by two consecutive events at different time intervals. a) The PPF induced by EDL for the Au gate showed an exponential decrease in synaptic weight by increasing the time interval between two pulses (1V, 100ns). The time constant obtained from fitting shows a speed response much higher than that of biological neurons. b) The PPF response as a result of IG and EI effect by sending two consecutive pulses (1V, 100 ns) to the LFP/Au side gate. We observed two-time constants, 9 ns and 131 ns, which indicate ultra-speed response on the one hand and the potential of the speed to be improved by changing the device geometry on the other hand.

4.10 Pulse Number Effect in Synaptic Weight Modulation and Short-Term to Long-Term Transition

In the brain, synaptic activity is an activity specific modification of the strength of the synaptic connection (synaptic weight). The connection can be categorized as either STP or LTP, depending on its retention time. Enhanced synaptic plasticity, which is dependent on the synaptic connection, is caused by Ca^{2+} influx into the postsynaptic membrane as a result of the stimulus received in presynaptic membrane. This plasticity (STP) will last for only tens of milliseconds due to spontaneous liberation of Ca^{2+} from the postsynaptic membrane. However, this can be transformed to a comparatively permanent change (LTP) via repeated stimuli, depending on the structural change in the dendrite and synaptic connections caused by Ca^{2+} concentrations [183-185].

To emulate this phenomenon in our MoS_2 synapses, different numbers of identical voltage pulses were applied to the presynaptic side gates while other pulse parameters were held fixed (0.5V, 1µs). For the Au gate, although the channel conductance experienced enhancement upon stimulation, the retention time and consequently the transition from STP to LTP was observed even for a large number of pulses, as illustrated in **Figure 4.12a**. We observed a linear increase in ΔG as we increase the pulse number to 7×10^4, as shown in **Figure 4.12b**, and a small drop at pulse number 10^5, which can be attributed to the induced charge accumulation and saturation at the interface of the electrolyte and MoS_2 surface. Then we applied the same pulses (1V, 1µs) in differing amounts ($300 < N < 20000$) to the LFP/Au gate of the same MoS_2 device. For total pulse numbers below 300, the plasticity observed was STP, but as the pulse number increased, not only did the magnitude of the channel conductance increase ($\Delta G_{S.T+L.T}$), but also the retention time underwent a significant increase, yielding LTP behavior, as shown in **Figure 4.12c**.

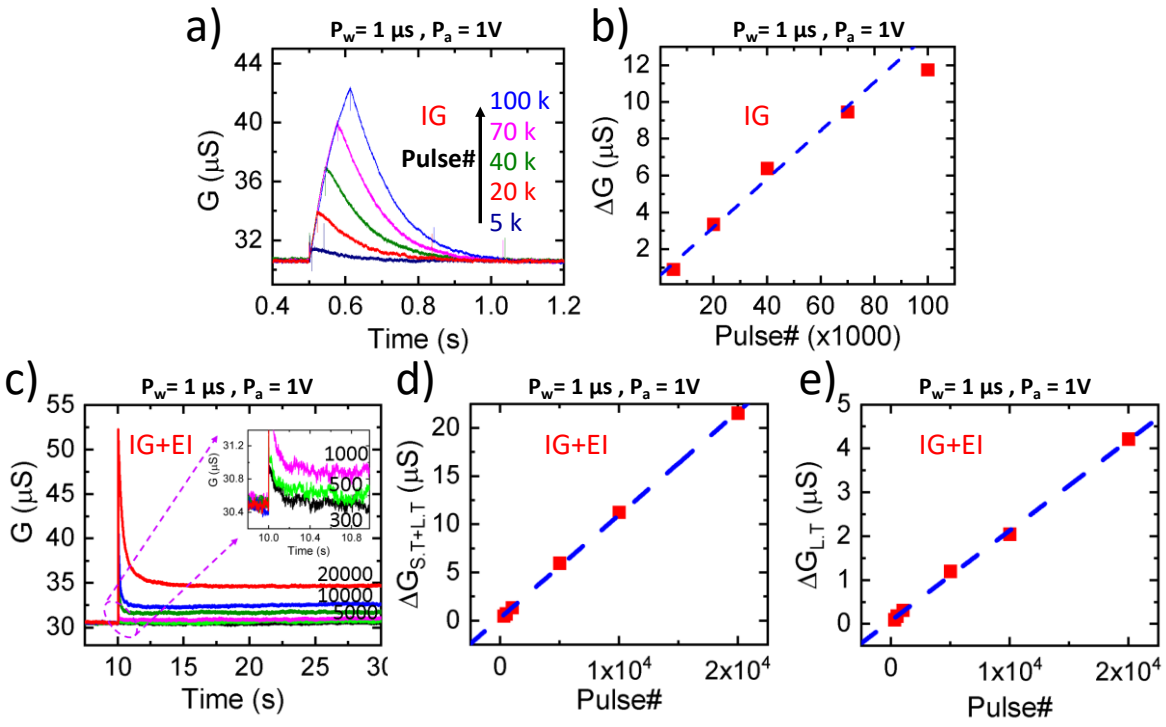

Figure 4.12. Synaptic weight modulation using pulse number. a) MoS_2 conductance amplitude upon application of a large number of a series short pulses 1v, 1µ (5k to 100k) to the Au gate and monitoring the conductance response, the short-term potentiation due to fully EDL effect observed independence of number of stimulations, which is desirable for SNN applications. b) The linear relationship between conductance change amplitude and pulse number. The conductance drop at 100k indicates saturation of the induced charge due to EDL. c) The conductance amplitude for different pulse numbers applied to the LFP/Au side gate. The short-term to long-term transition observed even with smaller pulse number compared to I.G. case indicates Li^+ ions accommodation in MoS2 layers due to E.I. effect. d,e) The total synaptic change ($\Delta G_{S.T+L.T}$), and permanent synaptic change ($\Delta G_{L.T}$), showing highly linear relationship with pulse number without any degradation, indicating denser available distinct states for the E.I case compared to I.G.

Overall, the total and permanent conductance change, $\Delta G_{S.T+L.T}$, and $\Delta G_{S.T+L.T}$, respectively, were shown to be highly dependent on the pulse numbers and experienced linear changes indicating a spacious layered structure for Li^+ ion accommodation to achieve high distinct states in our ECRAM, as shown in **Figure 4.12d, e.** To understand the transition behavior, the decayed curves of the device for each number of pulses were fit to an exponential decay curve using a modified "Kohlrausch" equation (Equation 4.4):

$$G_t = G_0 + A \exp(-\frac{t}{\tau}) \tag{4.4}$$

where G_0 is the channel conductance in the permanent state, A is a perfector, and τ is a relaxation time constant indicating forgetting rate. This type of curve is frequently used in psychology to model forgetful behaviors [186]. For the I.G. process, independent of the pulse number we observed the same forgetting rate, ~111 ms, expressing the short-term memory behavior in our brain that is desirable for SNN applications (**Figure 4.13**).

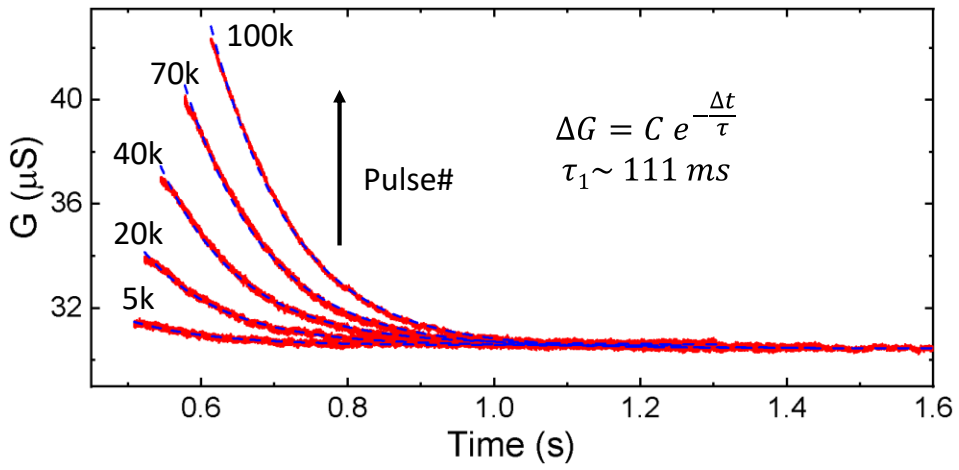

Figure 4.13. Short term plasticity and relaxation time for various pulse numbers in the IG process.

For the E.I. response, not only did the device conductance improve, even at a small number of consecutive pulses compared to the I.G. case, but the relaxation time also improved considerably by increasing the pulse number. To verify this result, we used a modified version of Eq. 4.2, adding a time constant term. The results of this fitting are shown in **Figure 4.14a-e**. Results indicate that upon increasing the pulse number, the retention time and synaptic weight rose. The plotted retention time (τ) as a function of pulse number in **Figure 4.14f** shows a clear improvement of τ by increasing the number of stimulations, indicating short term to long term potentiation transition.

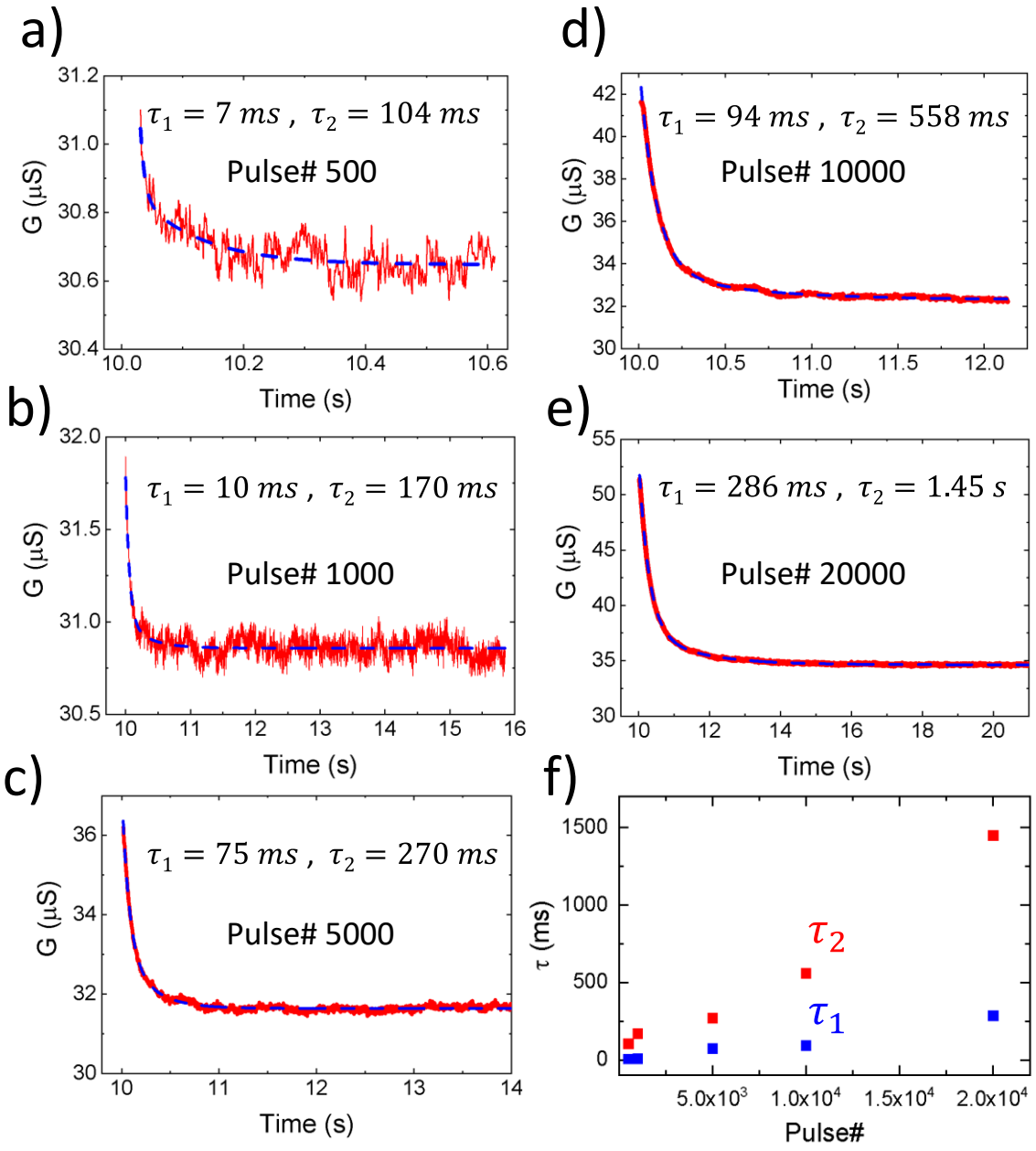

Figure 4.14. Short-term to long-term transition in the E.I. process. a-c) The retention for various pulse numbers shows two different plasticity from short to long term memory. f) The visualization of relaxation times (τ_1, τ_2) versus pulse number indicate by increasing the pulse number we can improve the long-term memory effect in our MoS_2 synapse.

4.11 Temporal Dynamic Filtering and Synaptic Weight Modulation Using Pulse Frequency Variation

The short-term synaptic depression and facilitation features in the human nervous system enable synapses to act as dynamic low-pass and high-pass filters, respectively. The specific behavior is controlled by stimulus frequencies, which are used for information transmission. For our synapse we were able to demonstrate high-pass temporal filtering, which is a desirable feature for SNN applications, by emulating PPF as discussed previously, where the higher conductance change (synaptic weight) appears for shorter time periods, Δt. Specifically, with the application of 40,000 successive pulses (0.5V, 1μs) at different frequencies ranging from 10 kHz to 800 kHz to the Au gate, we observed an increase in conductance amplitude as the frequency increased, leading to a high-pass dynamic filter, as illustrated in **Figure 4.15a**. By varying the signal frequency and thus changing the pulse intervals, we can regulate Li^+ diffusion and relaxation processes and are then able to modulate the synaptic weight magnitude (**Figure 4.15b**). Using this we are then able to emulate the spiking rate dependent plasticity (SRDP) observed in biological neurons. Similarly, we are able to imitate SRDP for the LFP/Au side gate by sending stimulations (20,000) with the same pulse width and amplitude (0.5V, 1μs), as illustrated in **Figure 4.15c**. As expected, by increasing the frequency (10 kHz to 800 kHz) and thus tuning the Li^+ ions diffusive dynamic, not only on the device surface but also within the MoS_2 vdW interlayers we were able to increase channel conductance in larger magnitudes compared to the those yielded by the Au gate and EDL formation (**Figure 4.15d**). Furthermore, the short term to long term plasticity transition was emulated for frequencies larger than 400 kHz due to the Li^+ ion intercalation in the vdW layers and the LTP level was able to be increased by increasing the frequencies (**Figure 4.15c**).

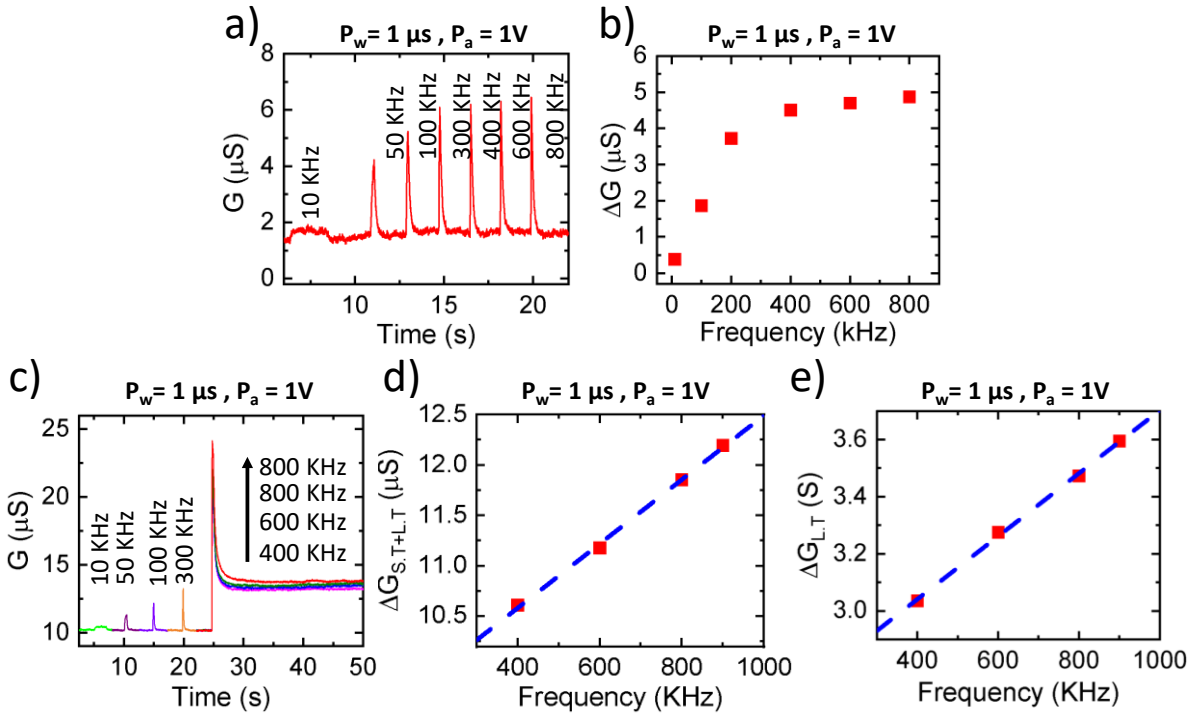

Figure 4.15. Dynamic filtering. a) High-pass temporal filtering characteristics upon application of 40000 consecutive potentiation pulses (1V, 1µs) with different frequencies (10k, 100k, 200k, 400k, 600k, and 800 kHz) to the Au gate. b) The pseudolinear increase in synaptic weight by increasing the frequencies. a) The frequency dependent synaptic weight via applying 20000 pulses with various frequencies (10 k, 50 k, 100 k, 200 k, 400 k, 600 k, and 800kHz) to the LFP/Au side gate. For frequencies below 400 kHz, we observed temporal high-pass dynamic filtering; for higher frequencies short-term to long-term transition occurs due to E.I. effect with d) linear change in total synaptic weight and e) permanent change in channel conductance.

4.12 Device Network Demonstration for ANN Implementation

To show the viability of using these synapses to form hardware for native ANN implementation, different network simulations were performed using MLP+ Neurosim [187]. This was used for devices characterized for both E.I. (100 ns, 10 us pulses) as well as Li Ion intercalation

(100 ns, 10 us, 100 us). The characterized synaptic devices were used to simulate a basic 2-layer perceptron for image recognition using the MNSIT standard number set [187] as well as ADAM network training methods [187]. The network was trained using 60,000 training images as well as 10,000 test images per epoch, for a total of 125 training epochs. The hardware was simulated as a standard 1T1R crossbar layout, which uses the standard circuit peripheries described by MLP+ Neurosim [187].

To determine the learning potential of these networks, the accuracy vs training epoch data was fitted to a saturating equation of the form (Equation 4.5):

$$A * (1 - e^{-b*x}) \tag{4.5}$$

Where A is the saturation value of the network accuracy and b is an empirical fitting value. The accuracies of these networks are shown below in **Figure 4.16**, with their saturation accuracy shown in the insets. As can be seen, the more linear the devices' switching property was, the better the accuracy. Specifically, both 100 ns pulses for E.I. and ion intercalation vastly outperformed the others when energy efficiency is taken into account. It is worth noting the initialization issues that occurred with the 10 μs data; the network seemed to have issues with settling into a stable initial solution for a few epochs before getting back out of it, so the final saturation number could be different in a real network of this architecture.

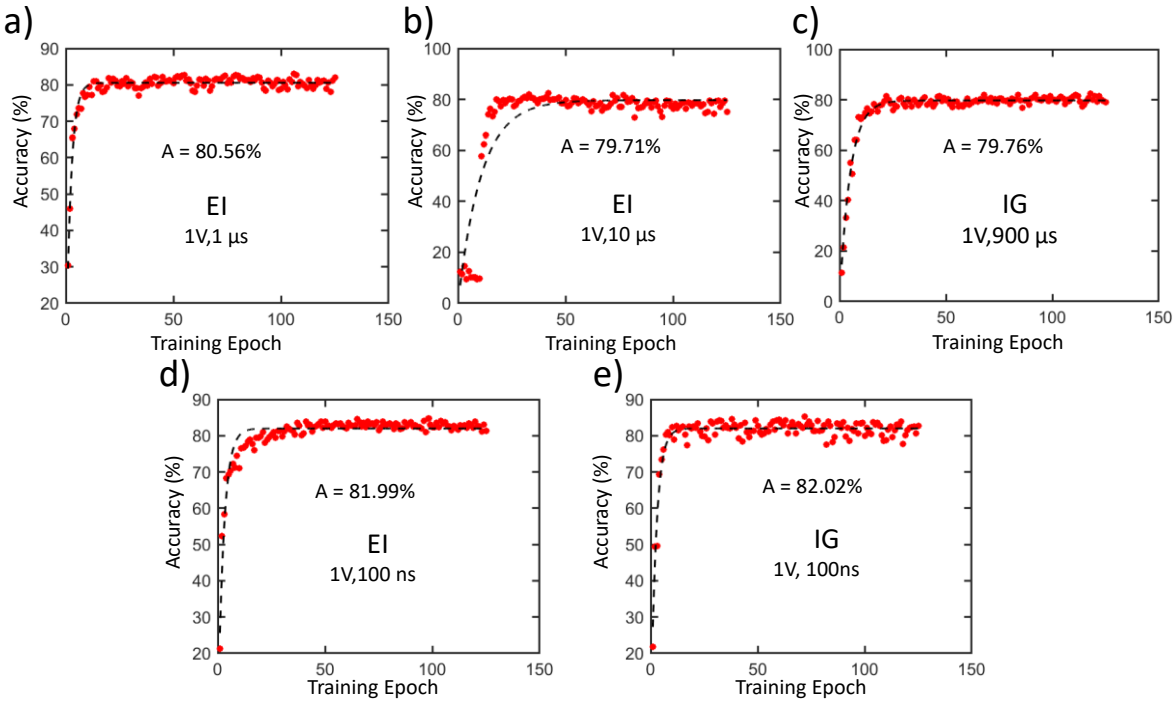

Figure 4.16. Network accuracy vs Training Epoch simulation results of training a 2-layer perceptron from MLP+ Neurosim for 125 training epochs. a-c) The results from 1 µs, 10 µs and 900 µs of E.I. programming methods respectively. d,e) Comparisons of 100 ns pulses for E.I. and I.G. training methods, respectively.

4.13 Event-Based Spatio Temporal Pattern Recognition

The use of electrochemical memristors for spatio-temporal pattern recognition has recently been demonstrated successfully [188]. This type of approach makes use of the transitive conductance response of these devices to single pulse stimuli, which can be used to model a synapse in a neuromorphic network. Such networks allow for power efficient classification of temporal and spatio-temporal patterns of activation. This is an important step in large neuromorphic systems, where information is often represented as asynchronous patterns of spikes

or events in time. Since all analog devices are intrinsically stochastic (leading to noise, cycle-to-cycle and device-to-device variation), our previous studies [188] aimed to evaluate the effect of such variability on the accuracy of pattern recognition. Here, we test the proposed device with both methodologies in order to evaluate performance for both gates in presence of noise and cycle-to-cycle variation. We built a stochastic model that could simulate our device's response to arbitrary spike trains. To achieve this, we defined a descriptive model that was used to fit the device response to single pulses and extracted the distribution of parameters that would be used for its stochastic counterpart, as described in Equation 4.6.

$$G(t) = \begin{cases} G_{rise}(t, t_0) \;\; when \;\; (t - t_0) < w | t > t_0 \\ G_{decay}(t, t_0) \;\; when \;\; (t - t_0) \geq w | t > t_0 \\ 0 \;\; otherwise \end{cases} \tag{4.6}$$

$$G_{rise}(t, t_0) = (A_1 + A_2)\frac{t - t_0}{w}$$

$$G_{decay}(t, t_0) = A_1 e^{-(\frac{t - t_0 - w}{\tau_1})} + A_2 e^{-(\frac{t - t_0 - w}{\tau_2})}$$

Where t_0 is the timestamp of the pulse application and w is its duration. The total conductance response G(t) is composed of a linear rise function G_{rise}, which models the conductance increase for the duration of the applied input pulse and a double decay function G_{decay}, which describes the subsequent conductance decay after the pulse is removed.

Fitting the device response to multiple iterations of the same pulse allowed us to determine sets of values for linear coefficients A_1, A_2 and temporal coefficients τ_1, τ_2. We then extracted parameters for the ionic gating (IG) and the electrochemical intercalation (EI) gates for (3V, 800 μs) pulses, which allowed generation of conductive temporal decays in the range of several milliseconds.

Table 4.2. Model parameters obtained with a (3 V, 800 µs) pulse. Each parameter is presented with its mean value ± Standard Deviation. The root mean square error (RMSE) serves as a measure of the fitting precision.

Gate	$\overline{A_1}$	$\overline{\tau_1}$	$\overline{A_2}$	$\overline{\tau_2}$	RMSE
IG	1.114±.05	15.58ms±.47ms	N/A	N/A	0.04
EI	0.52±.05	24.00ms±3.49ms	0.48±.05	120.53ms±1.83ms	0.01

Before fitting the data was normalized so that the mean maxima of the conductance was 1 and the baseline conductance was zero. This allowed us to use the same stochastic model as that presented in our previous work [188] to simulate the device response to arbitrary input spike trains. Similar to in our previous work, stochastic parameters were modeled as Gaussian distributions, with mean and standard deviation obtained by fitting the experimental data, as shown in **Table 4.2**.

$$\begin{cases} \hat{A}_1 = \mathcal{N}(\overline{A}_1, \sigma_{A_1}) \\ \hat{A}_2 = \mathcal{N}(\overline{A}_2, \sigma_{A_1}) \\ \hat{\tau}_1 = \mathcal{N}(\overline{\tau}_1, \sigma_{\tau_1}) \\ \hat{\tau}_2 = \mathcal{N}(\overline{\tau}_2, \sigma_{\tau_2}) \\ \hat{\eta} = \mathcal{N}(0, \sigma_\eta) \end{cases} \tag{4.7}$$

Where $\hat{\eta}$ is a normal random noise parameter used as additive noise induced in the device conductance. Its standard deviation, σ_η, is obtained from the RMSE in **Table 2.** For every individual input pulse sent to the memristive device, a set of parameters is drawn from the Gaussian distributions A_1, A_2, τ_1, τ_2 to model cycle-to-cycle variation (Equation 4.7). Additionally, additive

noise samples are drawn at a fixed sampling interval of 0.1 ms to add temporal white noise. The resulting stochastic model is described by the Equation 4.8-13.

$$G(t) = \sum_{t_i} G_{rise}(t, t_i) + G_{decay}(t, t_i) + \hat{\eta} \tag{4.8}$$

$$G_{rise}(t, t_i) = L_1(t, t_i) + L_2(t, t_i) \tag{4.9}$$

$$G_{decay}(t, t_i) = E_1(t, t_i) + E_2(t, t_i) \tag{4.10}$$

$$\begin{cases} L_1(t, t_i) = \hat{A}_1 \frac{t - t_i}{w} + B_1(t_i) \;\; t \in S_{rise} \\ L_2(t, t_i) = \hat{A}_2 \frac{t - t_i}{w} + B_2(t_i) \;\; t \in S_{rise} \\ L_1(t, t_i), L_2(t, t_i) = 0 \;\; t \notin S_{rise} \\ S_{rise} = \{(t - t_i) < w \mid t_{i-1} > t > t_i\} \end{cases} \tag{4.11}$$

$$\begin{cases} E_1(t, t_i) = (\hat{A}_1 + B_1(t_i)) e^{-(\frac{t - t_i - w}{\hat{\tau}_1})} \;\; t \in S_{decay} \\ E_2(t, t_i) = (\hat{A}_2 + B_2(t_i))) e^{-(\frac{t - t_i - w}{\hat{\tau}_2})} \;\; t \in S_{decay} \\ E_1(t, t_i), E_2(t, t_i) = 0 \;\; t \notin S_{decay} \\ S_{decay} = \{(t - t_i) \geq w \mid t_{i-1} > t > t_i\} \end{cases} \tag{4.12}$$

$$\begin{cases} B_1(t_i) = L_1(t_i, t_{i-1}) + E_1(t_i, t_{i-1}) \\ B_2(t_i) = L_2(t_i, t_{i-1}) + E_2(t_i, t_{i-1}) \end{cases} \tag{4.13}$$

B_1 and B_2 model the baseline conductance reached prior to each incoming spike t_i. B_1 models the population of charges driven by A_1 and τ_1 while B_2 models the charges driven by A_1 and τ_1. Since we want to evaluate how the device's stochastic behavior can affect computation, we compare the full stochastic model driven by the distribution of parameters in Equation 4.8 (called "Noisy" hereafter), against a fully deterministic "Ideal" model with fixed parameters (Equation 4.14).

$$\begin{cases} \hat{A_1} = \overline{A}_1 \\ \hat{A_2} = \overline{A}_2 \\ \hat{\tau_1} = \overline{\tau}_1 \\ \hat{\tau_2} = \overline{\tau}_2 \\ \hat{\eta} = 0 \end{cases} \tag{4.14}$$

The proposed device was tested using a discrimination task proposed in Wan et al. [175], which allows analysis of the memristive synapse's ability to enhance separation between different classes of spike trains by integrating temporal information. **Figure 4.17** shows the overall scheme and results for this task. A simple one-to-one connection network is used with a pre-synaptic neuron connected to a post-synaptic neuron through the memristive synaptic device (**Figure 4.17a)**. The pre-synaptic neuron elicits spike trains modelled coming from a Poisson generator **(Figure 4.17b-top)**. The neuron can generate different types of spike trains. These are treated as different classes for a discrimination task. The pre-synaptic neuron is connected through a "synaptic device" to a Leaky Integrator that acts as a post-synaptic neuron. The synaptic device can be either a "No-STP" simple memristor, modelled as a static multiplicative weight w, our "Noisy" memristor, or its "Ideal" counterpart. The principle behind this experiment is that in a Spiking Neural Network (SNN), the neuron membrane potential encodes and stores information before being transmitted as spiking activity to the rest of the network. Therefore, if a neuron is sensitive to a spike train temporal pattern, its membrane potential to different input inter-spike intervals should change significantly while responses to the same input spike train should be identical.

In this process, different spike trains (or classes) generated randomly are sent to the same synaptic device/post-synaptic neuron. The post-synaptic membrane is then read at a given time tread after a "Sequence End" spike occurring at the time end. The individual membrane values are

compared calculating the mean Intra-class distance ($\overline{IntraD}$, or the mean absolute distance between all the combinations of values of the same class) and the mean Inter-class distance ($\overline{InterD}$, or the mean absolute distance between all the combinations of values of different classes). Separability of classes is then calculated as $Separability = \overline{InterD}$ - $\overline{IntraD}$. In **Figure 4.17d** we test separability for the two gates (stimulated with a 3V, 800 µs pulse, parameters in **Table 2**) to compare the different synaptic devices. The Poisson generator frequency was set to 10Hz. We performed our test with a slow post-synaptic neuron capable of integrating the spike train's temporal information (τ= 100 ms), and a fast post-synaptic neuron that could not integrate temporal information (τ = 10ms). The weight w for the No-STP was set so the maximum post-synaptic membrane potential matched that of the ideal memristor (IG slow w = 14.33, EI slow w = 40.61, I.G. fast w = 7.40, E.I. fast w = 13.30). The times for the end of the spike train and read were set as t_{end} = 970 ms and t_{read} = 1000 ms, respectively. A total of 50 classes were generated, with 10 samples in each class. Unsurprisingly, for fast post-synaptic neurons, the proposed device natural STP offered significant improvements in class separation for temporal based tasks as the one proposed, as shown in **Figure 4.17d** (fast neuron). The EI gate offered longer decays (**Table 2**) for the same pulse characteristic, which made it perform better for both slow and fast neurons in this task. It is worth noting that even for the fast Ionic Gate and slow neuron, the Ideal model performed better than the No-STP synapses (up to 1.47x better separation). As expected for this highly time sensitive scenario, stochasticity can hurt separation by increasing intra-class distance, although in all tested scenarios the device still performed better than "No-STP" standard memristors. This test can be useful to highlight the importance and possible use of this type of novel dynamical memristor, but it is only indicative of performance in real world scenarios.

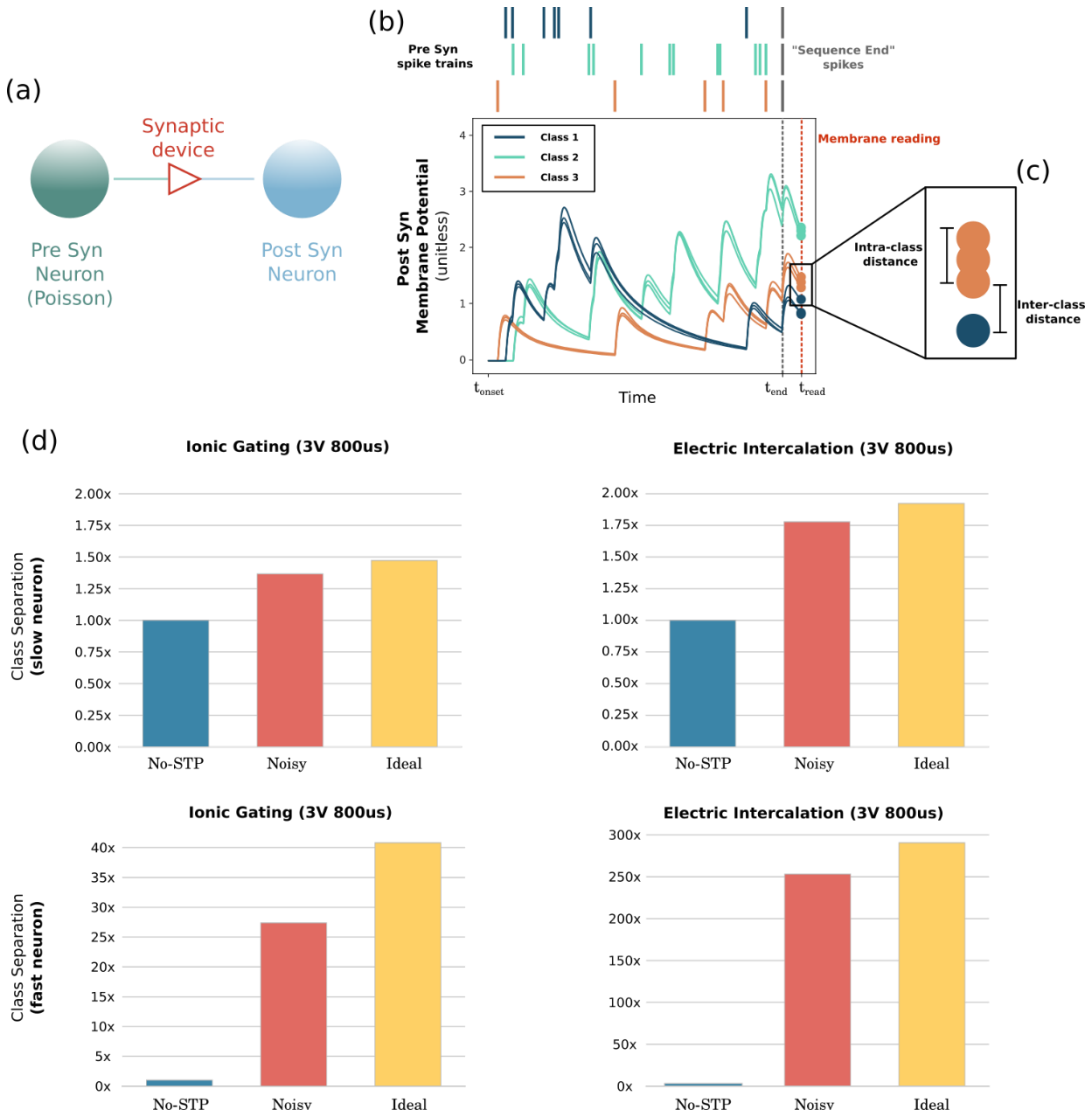

Figure 4.17. SNNs computation-based STP of our MoS2 synapse a) The diagram of our network, a Poisson pre-synaptic neuron connected to a leaky integrating post-synaptic neuron through a synaptic device (No-STP, Noisy or Ideal memristor). b) A simplified scheme of the benchmark. Different spike trains of fixed duration are generated by the pre-synaptic neuron, with a "Sequence end" spike added at tend to indicate the end of the stimuli. Each spike train is a class of a classification problem, and it is sent to the different synaptic devices and post-synaptic neurons multiple times to obtain multiple responses to the same class. The output of the network is represented by the post-synaptic neuron membrane potential

at the time tread. c) We calculate the mean Intra-class distance ($\overline{IntraD}$) and Inter-class distance ($\overline{InterD}$) between the membrane reads to obtain the separability metric: $Separability = \overline{InterD}$ - $\overline{IntraD}$. d) We test separabilty for the two gates (stimulated with a 3V, 800 µs pulse) both as an Ideal case (yellow) and stochastic Noisy device (red). We compare against a No-STP model (blue) which represents classical memristors with "static weights". We performed our test with a slow post-synaptic neuron capable of integrating the spike train's temporal information (τ= 100 ms), and a fast post-synaptic neuron that could not integrate temporal information (τ= 10 ms).

Moreover, the effect of noise and stochasticity on computation can be reduced by learning algorithms and highly redundant architectures [189] at a point where it can even improve computation [190-193]. To see how the device behaves on a real-world dataset, we tested it with the algorithm proposed by Rasetto et al. [188] on the N-MNIST dataset [194]. The network makes use of descriptors called Time Surfaces, which are spatial maps obtained by interpolating events with temporal kernels, usually exponential decays. In the case of memristors and image-based datasets like N-MNIST, each pixel produces an individual spike train based on a locally observed change of light, as illustrated in **Figure 4.18a, b.** The spikes then trigger write pulses for the memristor of the given pixel, and the conductance of spatial patches of memristor are sampled around the position of every new incoming spike, producing Time Surfaces (**Figure 4.18 d, e).** These descriptors can carry important spatiotemporal information about the input and can be clustered through unsupervised algorithms such as k-means to extract sets of features for classification or tracking.

Here we test the same architecture proposed by Rasetto *et al.* [188] shown in **Figure 4.18f**. We test again I.G. and E.I. gates in both the "Ideal" and the "Noisy" case. Differently from the single neuron SNN test, we show that a full architecture can accommodate errors introduced by

the stochastic behavior of the device, resulting in no significant difference between the Ideal and Noisy case for both gates (**Figure 4.18g**). The pulse response of the EI gate seems to be better suited for N-MNIST classification with a 5% increase in accuracy. It could be due to the longer decay of the EI gate versus the IG or the presence of two different time constants in the EI gate, which could allow integration at two different time scales.

In both tests we presented (the SNN and the full architecture), we can see that IG and EI gates differ in performance. The reason for this discrepancy has to be identified in the different temporal coefficients, which allow for integration at specific time scales. Ideally these temporal constants would have to be set during the design phase or by controlling the write pulse characteristics, hence our effort to determine the complex relation between device physics and temporal properties.

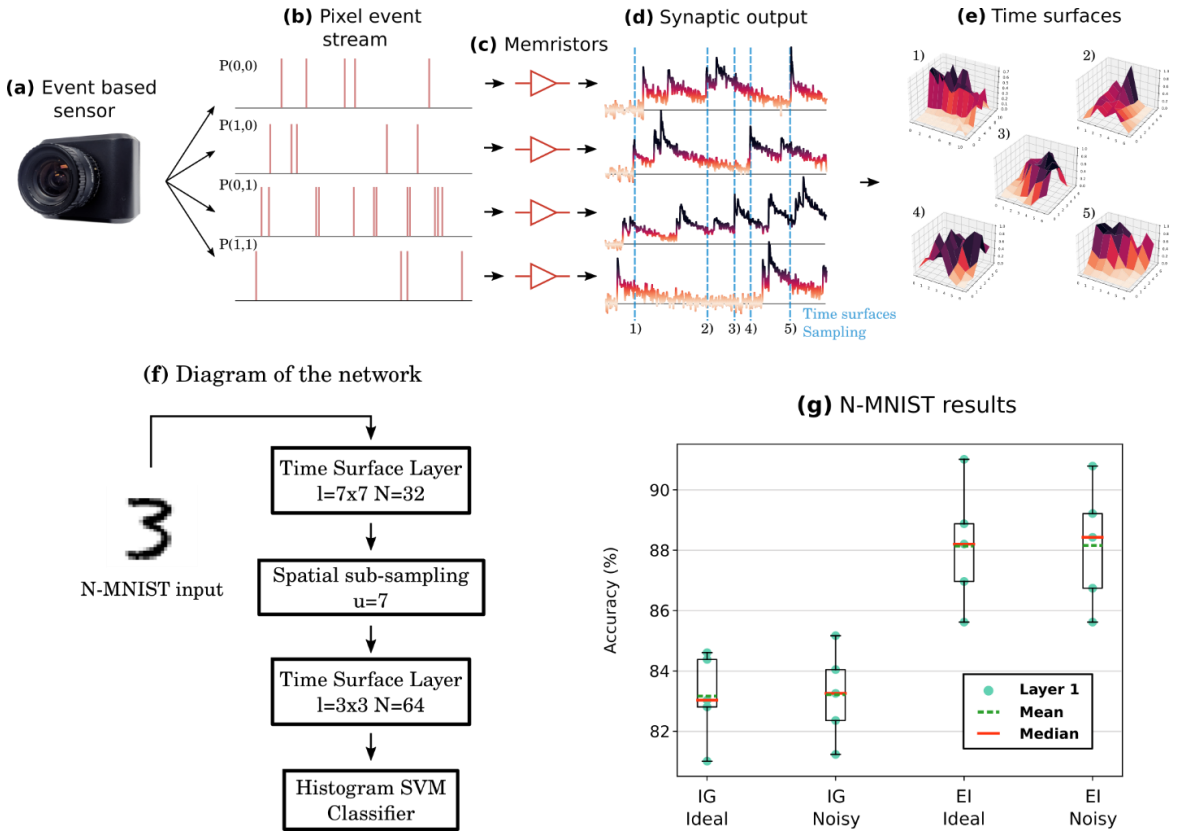

Figure 4.18. The process behind time surface generation. a-e) A diagram of the use of memristors for generating time surfaces, a descriptor for event-based datasets. a) An event-based camera emits events for different pixels P(Y,X). b). Each event is a settable writing pulse for different memristors connected individually to pixels. c) MoS2 artificial synapse. d) The conductance of the memristors can be sampled across different devices at each new event to create 2D maps called Time Surfaces (e). f) The scheme of the network we tested, black arrows indicate event streams; Time Surface Layers are determined by the spatial dimensions of the surfaces l and the number of clusters or features N obtained using k-means. Spatial sub-sampling was performed by reducing spatial coordinates of the input data by a parameter *u*. The number of cluster activations per recording were used to generate a histogram which was then fed to a Support Vector Machine allowing classification. g) The results on 10% of the M-MNIST dataset for 5 consecutive runs, showing virtually no difference between the Ideal and Noisy cases for both I.G. and E.I. (both gates were stimulated with 3V, 800 µs pulses).

4.14 Conclusion

In conclusion, we developed a programmable three terminal MoS2-based synaptic transistor using two ionic and electrochemical side gates. The coupling of ionic and electrochemical modulation in our MoS2 dynamic synapse initiate a major breakthrough in neuromorphic as well as energy technologies. In our MoS2 synapse, we were able to mimic basic brain functionalities such as STP, LTP, PPF, SRDP, and temporal filtering with ultra-fast switching speed (100 ns), ultra-low energy consumption (< 700 aJ), and extremely high precision (5000) with extraordinary linearity and symmetry, and good endurance performance. Moreover, biorealistic short term to long term transition was successfully emulated by varying pulse number and filtering frequency. We tested the STP effect with two previously proposed benchmarks for neuromorphic pattern recognition. We demonstrated a large boost in performance (up to 250x) over standard, non STP capable memristors in highly time-dependent scenarios, and we showed that device stochasticity, while affecting single neuron networks, might have virtually no effect on recognition rates for larger 2-layer architectures, all of which makes this device a perfect candidate for neuromorphic architectures. As far as we are aware, this artificial synapse exhibits the fastest ionic and charge transfer doping synaptic device with all the desired requirements for neuromorphic computing applications.

5.0 Programming MoS2 Dynamic Synapse With Long Voltage Pulse

5.1 Introduction

Despite the excellent performance of Von Neumann architecture-based computers in pure numerical calculations, their massive power consumption and slow speed make them incapable of efficiently handling modern intelligent tasks such as visual and speech recognition and real time analysis of large data sets [61]. On the other hand, the human brain system excels in terms of energy and speed at such complicated tasks, since the storage and processing of information occurs at the same locations in massively parallel and fault-tolerant neural networks [195-199]. With this in mind, new artificial neuromorphic computers inspired by the brain have attracted a great deal of attention from researchers as a new computing prototype [60,61,200-202].

Researchers have first looked to how neurons work. In the brain, neurons are connected via synapses, which can be considered a fundamental computing element. The connection strength between the neurons in the brain, synaptic weight, is precisely modified by Ca^{+2} concentration [184,195,203]. As a result, changing the synaptic weight (synaptic plasticity) is a key process for both learning and memory in biological neural systems [204]. So far, different methods of synaptic plasticity seen in biological synapses, such as short-term and long-term plasticity (STP and LTP) [24,72,145, 175,156, 205-207], pair pulse facilitation (PPF) [24,175,208,209], and spike timing dependent plasticity (STDP) [43,210,211], have been successfully demonstrated by artificial synapses. However, artificial neural networks have been traditionally implemented using existing CMOS transistors which are not ideal for this task due to their large power consumption, non-scalability, and most importantly, inability to mimic the fully analog changes found in biological

synapses [117, 144]. In recent years, some emerging two-terminal non-volatile memories such as resistive random-access memory (ReRAM) [56,140,146,147], phase change memory (PCM) [40,149,150], ferroelectric random-access memory (FeRAM) [212,213], and electrochemical metallization memory (ECM) [151,152] have been widely investigated as artificial synapses. Despite some promising features like analog programming, high programming speed, and simple two terminal structures, there are some outstanding challenges that need to be addressed (high programming power and poor reliability performance in PCM and large variations in ReRAM and ECM devices) in order for these for neuromorphic applications and hardware implementations. Furthermore, these devices have limited applications for next generation spiking neural networks (SNN), which process information through generating time-encoded spiked signals.

To overcome the limitations of existing device architectures, three-terminal electrolyte gated transistors based on an electrochemical redox mechanism have been proposed as artificial synaptic devices [24,71,72,175,158]. In these three-terminal devices, the channel conductance, the device's synaptic weight, can be precisely controlled by applying an electrical signal to the gate electrode. Basic synaptic functionalities (high dynamic range and precision, symmetric and linear response, good reliability) have been demonstrated with low power consumption. Moreover, due to these devices having electrochemical reactions and electric double layer gating effects, both long term and short-term plasticity have been reported in the same device, a key property that plays a vital role in event driven computation in artificial neural networks [72, 175].

Here we present an electrochemical three terminal transistor using MoS_2 films. MoS_2 is an exciting material in the transition-metal dichalcogenides (TMDs) family that has garnered a lot of attention in recent years in a multitude of applications. Here, its layered structure and anisotropic ionic transport and electrical properties, both in-plane and out-of-plane, as well as its ability to

modulate its physical characteristics by inducing a reversible phase change between the semiconducting (2H) and metallic (1T) phases allow for a robust device [214,215-217]. We demonstrate phase modulation in MoS_2 thin films through field induced Li ion migration in the Van der Wal layers of the MoS_2 channel, where low and high concentration of Li^+ ions indicate 2H and 1T phases, respectively. According to the literature [218- 220], the in-plane Li^+ ion diffusivity in MoS_2 films is considerable and allows the reversible phase transition via both an electrical field induced Li^+ ion migration and electrochemical intercalation (E.I.) mechanism. The accommodation of Li ions in 2H multilayer MoS_2 thin films (interlayer spacing of 0.62 nm) drives the phase transition ($2H \rightarrow 1T$) by destabilizing the 2H-MoS_2 crystal [218]. Interestingly, this typical interlayer gap size has the potential to be expanded in the future and enable MoS_2 to be a good host for other alkali metal ions accommodation as well [221-223]. Their capability for reversible ionic exchange and the resulting analogue conductance change make these MoS_2 based devices a strong candidate for neuromorphic computing applications. Although promising MoS_2 artificial synapses have been demonstrated, their high programming voltage (3V), poor precision (100) and low speed (1ms) [219] need to be addressed for any hardware implementation in artificial neural networks. Here we present a three-terminal planar MoS_2 transistor that displayed an improvement in programming voltage (1 V), and energy (4.1 pJ), as well as the high precision (5000 states) required for DNN applications. Furthermore, we explore the device potential for SNN applications by demonstrating time dependent synaptic features such as paired pulse facilitation (PPF), spike timing dependent plasticity (STDP), and temporal dynamic filtering.

5.2 Device Structure and Working Principles

The schematic of a biological synapse and the proposed three terminal synaptic device are shown in **Figure 5.1a,b**, respectively. In the biological synapse (**Figure 5.1a**), the voltage-gated calcium channel is activated by an arriving action potential at the pre-synaptic cell (axon terminal) which allows an influx of Ca^{+2} ions into the axon terminal and triggers the synaptic vesicles to release their neurotransmitters into the synaptic cleft (20-40 nm gap between pre and post synaptic neuron). The neurotransmitters bind to receptors and cause ion channels to open (or close) in the cell membrane generating a connection strength (weight) change in membrane potential. The post synaptic potential can therefore be strengthened (potentiation) or weakened (depression) by tuning the Ca^{2+} concentration (synaptic plasticity). Here, we develop artificial synapse devices, as illustrated in **Figure 5.1b,** to demonstrate biorealistic synaptic functions. The device consists of drain and source electrodes (Au [50 nm] / Ti [2 nm]) as well as two side gates (LFP/Au and Au) with a 10 μm physical distance from the MoS2 channel, as illustrated in **Figure 5.1c,d**. The drain/source electrodes read the conductance change (synaptic plasticity) of the device via a read-out voltage (100 mV) while the two side gates control the electrolyte, which serves as the source for exchanging the Li^{+} ions between MoS_2 and gate electrodes.

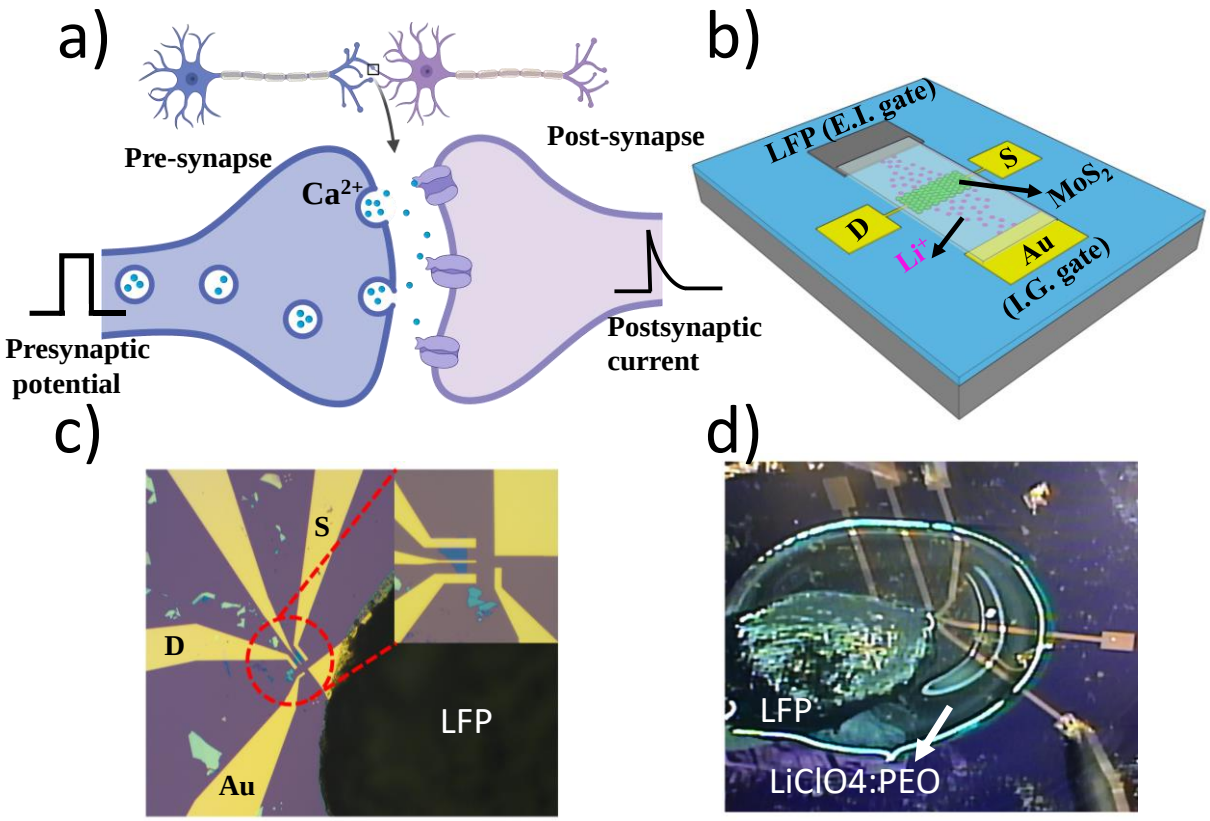

Figure 5.1. MoS2 synapse. a) schematic of biological synapse b) schematic of our MoS_2 synapse. c,d) optical images of our MoS_2 synapse.

5.3 Synaptic Plasticity Role in Human Brain

Synaptic plasticity is one of the most important characteristics of the human brain. It's modification on the neural circuits is thought to be how behavior, feelings and general learning processes are formed. This event-based plasticity clearly has a key role in different aspects of the fundamental brain functionalities, and there is major ongoing research in order to understand its role in the formation of neural circuitry.

5.3.1 Mimic of Excitatory and Inhibitory Synapses Using Long Voltage Pulse

In our brain, chemical synapses are the building block in neural circuits and play a basic role in memory, learning, and information processing. The most abundant and basic synaptic actions are classified as excitatory and inhibitory, where the synaptic weight increases or decreases, respectively, upon the action potential arriving. Upon application of a short stimulus to the presynaptic synapse and resulting temporal accumulation (removal) of Ca^{+} ions in presynaptic synapse, short-term plasticity occurs due to an enhancement (reduction) in synaptic weight. In electronic synaptic devices based on CMOS transistors and memristors, this key property is missing. Being able to fabricate devices that can emulate this behavior will strengthen neuromorphic computation functionalities, enabling them to better implement complicated tasks such as pattern recognition, speech recognition, cognition, learning and decision making.

In our 2D ECRAM device we are able to mimic both excitatory and inhibitory synapses, respectively, simply by applying a single intercalation or deintercalation voltage pulse (|1 V|, 20 ms) to the side gate, respectively, as shown in **Figure 5.2a,b**. The resulting potentiation and depression for the Au gate presynaptic electrode is fully transient due to the electric double layer (EDL) gating effect involved (**Figure 5.2a)**. We observed an abrupt increase (decrease) ΔG_{tot}= 0.65 µS in channel conductance due to Li^{+} ions adsorption onto the MoS2 surface, which mimics an increase (decrease) in the synaptic weight of a biological excitatory (inhibitory) synapse. The value of ΔG is gradually diminished back to its original state (G = 49.1 µS) after removal of the applied pulse, and the weakly attached Li^{+} ions are liberated from the surface. On the other hand, upon applying the same voltage pulse to the LFP/Au gate, we observed more conductance change ($\Delta G_{L.T}$= 2.1 µS) compared to that at the pure Au gate, as illustrated in **Figure 5.2b**. This can be

attributed to the electrochemical intercalation (deintercalation) process that happens upon stimulation.

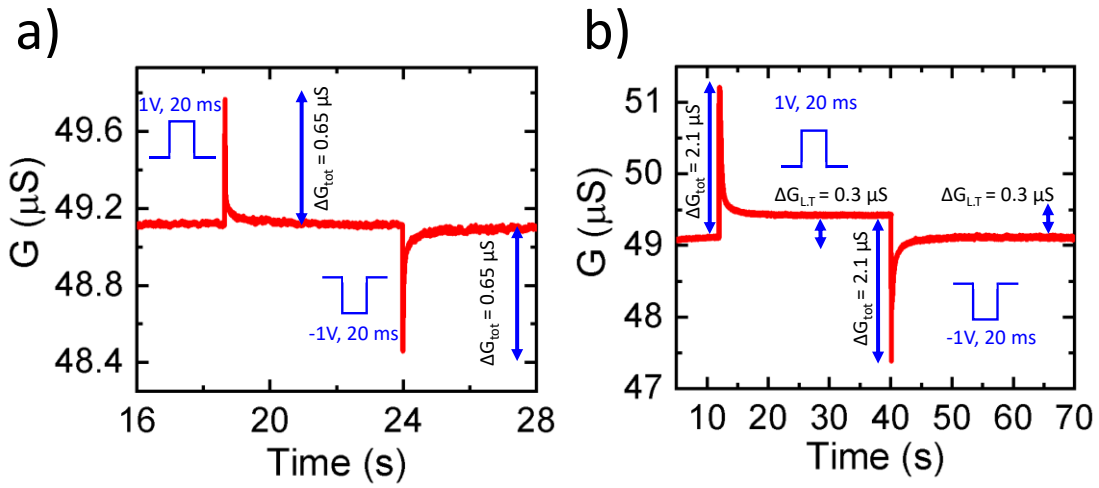

Figure 5.2. Excitatory and inhibitory synapses. a) the emulation of short-term potentiation and depression of biological synapse through EDL effect upon application of positive and negative pulse signals to the Au ionic gate b) mimicking of potentiation and depression effect upon application of intercalating and deintercalating pulse signals and resulting combined EDL and electrochemical reaction phenomena.

Moreover, the value of ΔG gradually decreased (increased) over time after experiencing a sharp change and reached a permanent and stable value at G = 49.4 µS. As we reported for our graphene synapse [24], the initial sudden transient change in channel conductance is coming from both short-term doping due to EDL and a long-term doping effect due to Li^+ intercalation. However, doping of Li^+ ions in vdW layers of MoS_2 through electrochemical intercalation mechanism is permanent and the Li^+ ions cannot diffuse back into the ion gel electrolyte without injecting a negative presynaptic voltage pulse. As a result, we observed nonvolatile and permanent effects which emulates of long-term potation effects in biological neural networks. As Shown in

Figure 5.2b, the permanent effect (ΔG= 2.1 μS) due to the positive intercalating pulse (+1V, 20 ms) can be eliminated and the original conductance (49.1 μS) recovered upon application of deintercalating pulse (-1V, 20ms).

5.3.2 Relaxation Process and Dynamic Behavior for The Response of The Both Ionic Gating and Electrochemical Intercalation Mechanisms

The fitting parameters and calculated time scales for the responses of both side gates indicate their dynamic behavior after removal of the electrical field. For the Au gate, we observed a single time scale for the relaxation curve for both potentiation **(Figure 5.3a)** and depression **(Figure 5.3b)** actions as the adsorbed mobile ions on the surface can easily diffuse back into the gel electrolyte. However, for LFP/Au gate, the mobile ions can also intercalate into the inter layers of the MoS_2. Once the pulse is removed, however, the backward diffusion is slower and accordingly we observed two-time scales due to the combined effects of surface adsorption and intercalation **(Figure 5.3c.d)**.

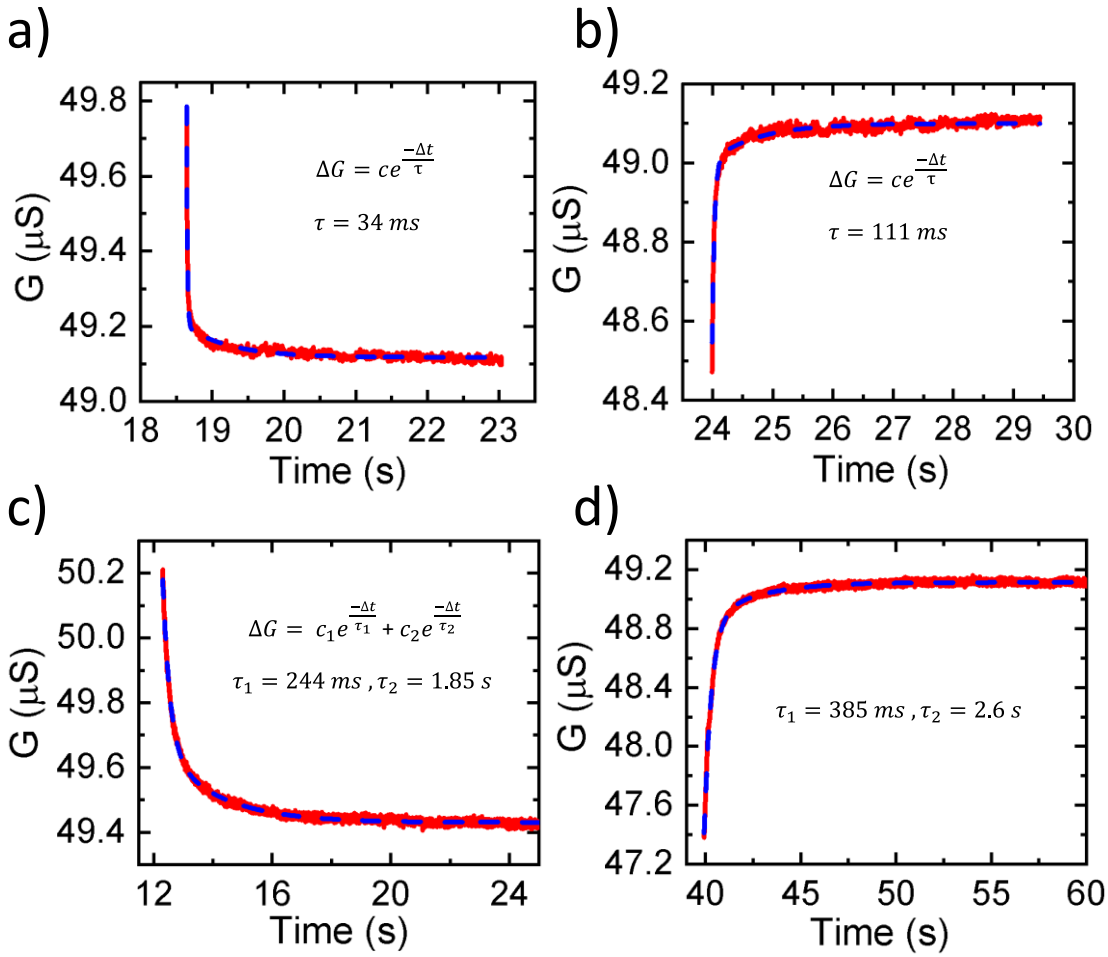

Figure 5.3. Relaxation process of the injected mobile ions. a,b) the relaxation time versus time and fitting parameters for the both potentiation and depression mechanism after removal applied positive (+1V, 20ms) and negative (-1V, 20ms) pulses to the EDL gate (Au). c,d) relaxation process with respect to time after injecting positive (+1V, 20ms) and negative (-1V, 20ms) pulse signals to the LFP/Au gate, resulting in two time constants after pulse removal.

5.3.3 Energy Consumption as A function of MoS_2 Synapse Thickness

We have also investigated the dependency of the long-term synaptic weight on the vdW layer number by applying a single positive voltage pulse (+1V, 20 ms) to the E.I. gate. As seen in **Figure 5.4**, as MoS_2 thickness increases the value of the permanent change increases, indicating more site availabilities for Li^+ ions to be permanently accommodated. However, we have not seen any meaningful trend for the I.G case as all the Li+ ions diffused back due to the EDL and no intercalation occurred.

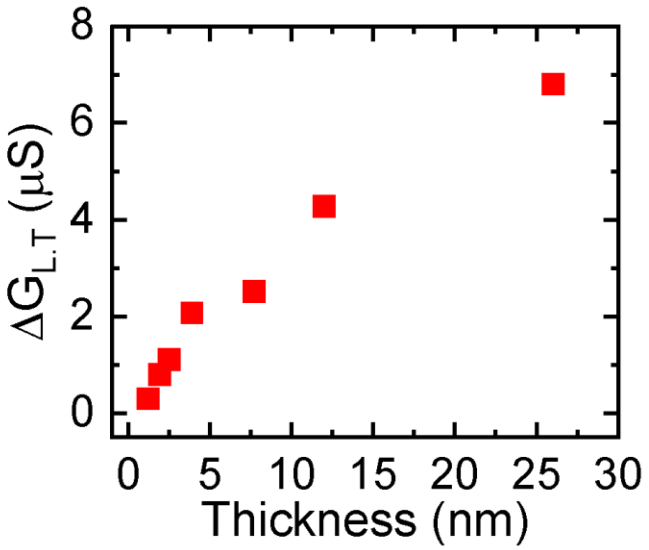

Figure 5.4. Long-term potentiation versus device thickness. Long term doping with respect to MoS_2 thickness indicates more space for mobile ion accommodation during intercalation process.

5.4 Emulation of Long-Term Potentiation and Depression Via Consecutive Potentiation and Depression Pulses For The Both Ionic Gating and Electrochemical Intercalation Mechanism

Information storage is the main task of the brain, and it is widely believed that the majority of data and information is stored at synapses in the shape of changes in synaptic plasticity. Long-term potentiation (LTP) and long-term depression (LTD), which result in permanent improvement/diminution in synaptic weight, have been widely studied as two major forms of synaptic plasticity in biological neurons and are believed to also play a vital role in learning and memory [224,225]. Both LTP and LTD have the potential to occur in a single synapse in response to various forms of activation of N-methyl-D-aspartate receptors [224]. We demonstrate LTP and LTD in our MoS_2 synapse through the application of a train of pulses with opposite polarities (|1V|, 20ms) to either side gates. By sending consecutive voltage pulses to the controlling electrode, we could mimic LTP by increasing the synaptic plasticity. Conversely, to avert new information encoding at the same synapse we selectively weaken the synaptic plasticity by sending negative voltage pulses to the controlling electrode. We programmed the MoS_2 synapses via both side gates separately and recorded the synaptic weight modification for comparison. Although, as shown in **Figure 5.5a,b**, the synaptic weights experienced a large enhancement via both gate terminals, the capacity of the conductance change modifications due to the intercalation process obtained in LFP/Au gates is larger and with the same number of pulses we achieved a highly linear, symmetric, and reproducible response. However, for the Au gate, as illustrated in **Figure 5.5a**, despite the high linearity observed at the initial stage of injected pulses, a gentle nonlinearity is detected at the end, with an even smaller pulse number (20 compared to LFP/Au gate's 125). This indicates charge accumulation and saturation on the surface of the MoS_2 due to the EDL effect. Moreover, the high

ON/OFF ratio for intercalation response (~15) compared to the EDL response (~1.5) indicates large dynamic range through the LFP/Au gate terminal, demonstrating its potential for DNN applications and computations in neural network systems. The good cycling endurance (> 10^4) for both Au **(Figure 5.5c)** and LFP/Au **(Figure 5.5d)** gate terminals as a result of EDL and combined EDL and intercalation effects also demonstrate its capability for hardware implementation.

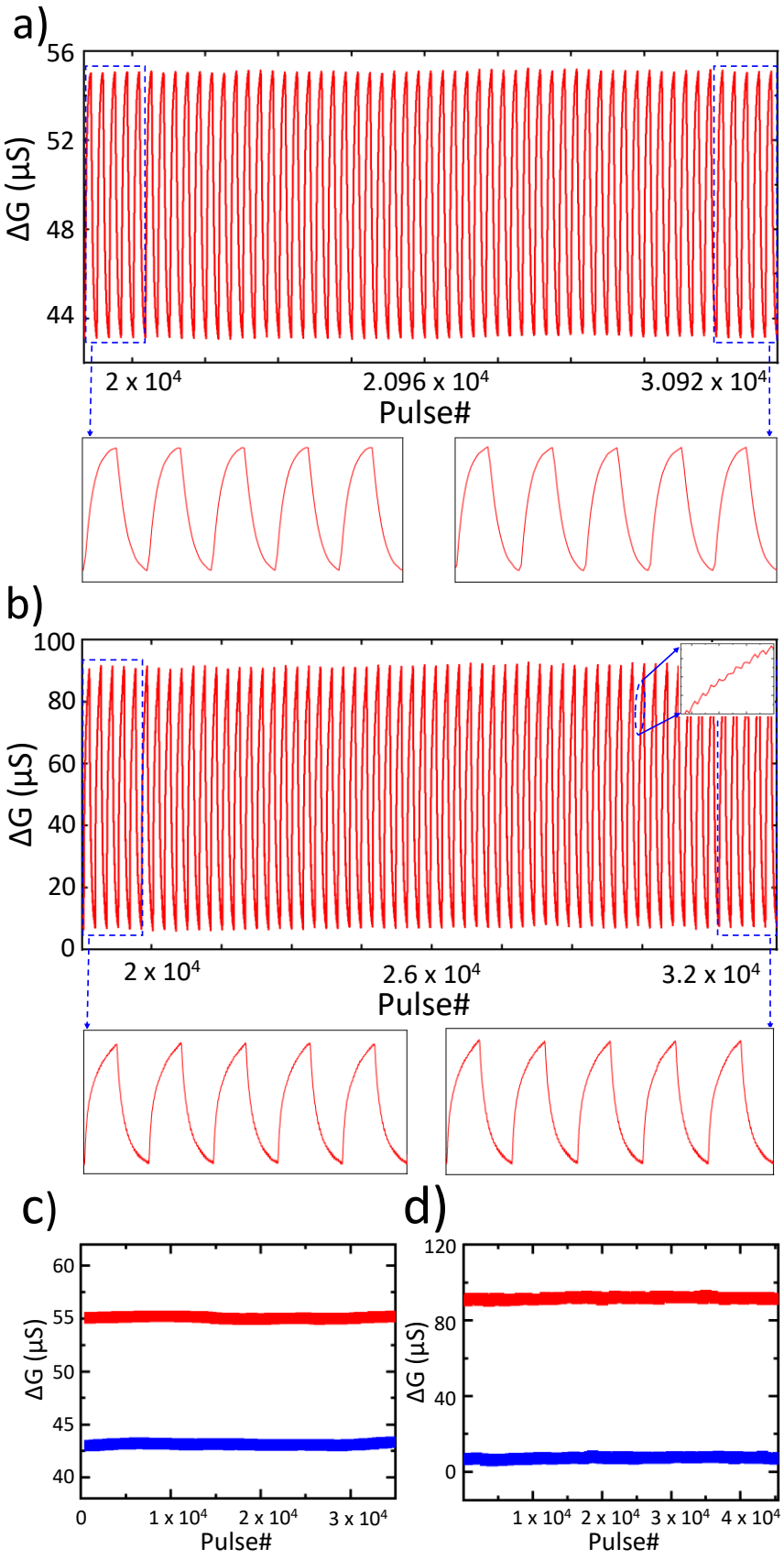
a)
56
52
48
44
ΔG (μS)
2 x 10^4
2.096 x 10^4
3.092 x 10^4
Pulse#
b)
100
80
60
40
20
0
ΔG (μS)
2 x 10^4
2.6 x 10^4
3.2 x 10^4
Pulse#
c)
60
55
50
45
40
ΔG (μS)
0
1 x 10^4
2 x 10^4
3 x 10^4
Pulse#
d)
120
80
40
0
ΔG (μS)
1 x 10^4
2 x 10^4
3 x 10^4
4 x 10^4
Pulse#

Figure 5.5. LTP and LTD. a) emulation of LTP and LTD due to EDL effect upon application of 20 consecutive positive and negative pulses to the Au gate terminal. b) emulation of LTP and LTD due to the combined EDL and intercalation/deintercalation effect with the application of a train of pulses (#125) to the LFP/Au gate terminal. c,d) demonstrating cycling endurance with good performances ($> 10^4$) for the both responses induced by EDL and combined EDL and electrochemical process.

5.5 Synaptic Weight Modulation and Sort-Term to Long-Term Transition Using Multiple Potentiation Pulses and Retention Behavior

Another important feature of the biological synapse that we were able to mimic in our artificial MoS_2 synapse is STP to LTP transition upon applying a series of voltage pulses to the gate terminal. Similar to the biological synapse where the Ca^{2+} ions concentration in the cell plays a critical role in neural functions, Li^+ incorporation and extrusion result in a dynamic balance in the artificial synapstic device, which heavily influences its learning and memory. As depicted in **Figure 5.6a**, no transition has been seen for the Au gate by increasing the number of applied pulses (1V, 50ms), indicating that all the Li^+ ions that migrated toward the MoS_2 surface due to the EDL effect diffused back into the electrolyte even after injecting 50 pulses. This transient nature of the synaptic plasticity in respect to the functionality and internal dynamics for the presented artificial synapse make it a strong candidate for SNN applications in neuromorphic computing systems. However, an obvious LTP effect has been observed upon applying successive voltage pulses to the LFP/Au gate, as illustrated in **Figure 5.6b**. Although we observed a short relaxation with time due to combination of the transient EDL effect and a spontaneous Li^+ deintercalation process, the device conductance could not reach to its original state over time, implicating transition to a permanent level of device conductance due to enduring Li^+ accommodation into the channel. The

retention measurement for 50 pulses confirms the LTP emergence and stability for a long period of time, > $2x10^5$ s (**Figure 5.6c**). Physically, due to the subsequent stimuli, the appeared potentiation caused by the former pulse has no chance to be fully relaxed and the accumulated potentiation over the process results in conversion to LTP.

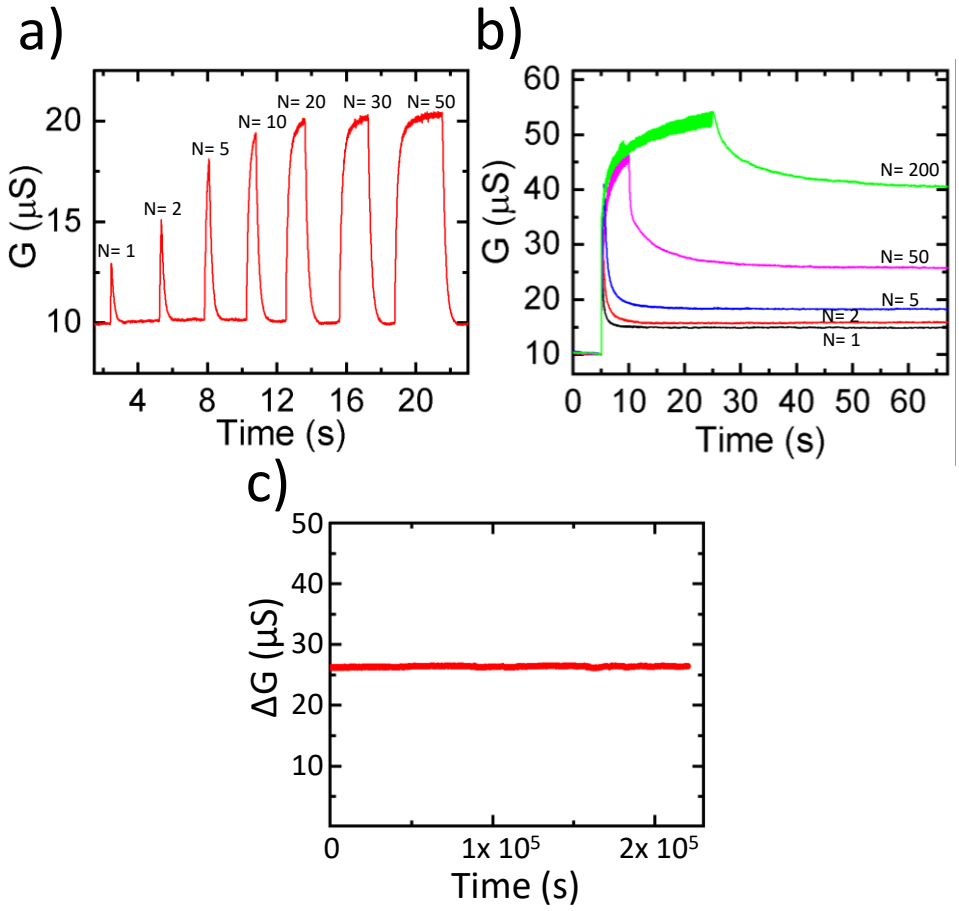

Figure 5.6. Short-term and long-term transition emulation using pulse number (N). a) conductance change (synaptic weight) in MoS_2 channel upon application of consecutive potentiation pulses to Au gate terminal. The response shows short-term plasticity independent of the pulse number due to the EDL effect. b) the long-term potentiation effect upon application of successive positive potentiation pulses with different number. The LTP level increases by increasing the pulse number, indicating more Li^+ ion concentration in the MoS_2 layers due to electrochemical processes. c) retention behavior with more than $2x10^5$ s for N= 50.

5.6 Relaxation Behavior of The Synaptic Response Upon Application of Consecutive Potentiation Pulses to The Both Ionic Gating and Electrochemical Intercalation

To understand the transition behavior, the decayed curve of the device numbers for each pulse number were fit to an exponential decay curve using a modified "Kohlrausch" equation (Equation 5.1):

$$G_t = G_0 + A\exp(-\frac{t}{\tau}) \tag{5.1}$$

where G_0 is the channel conductance in the permanent state, A is a perfector, and τ is a relaxation time constant indicating forgetting rate. This type of curve is frequently used in psychology to model forgetful behaviors [186]. A train of pulse with 1V amplitude and 100 μs width was applied to both programming gates to demonstrate this feature. For the I.G. process, we observed the same forgetting rate ~125 ms (**Figure 5.7**) independent of pulse number, exhibiting the short-term memory behavior that is desirable for SNN applications. For the E.I. gate and consequent electrochemical intercalation process, however, the retention time and synaptic weight was raised by increasing the applied pulse number, as shown in **Figure 5.8**.

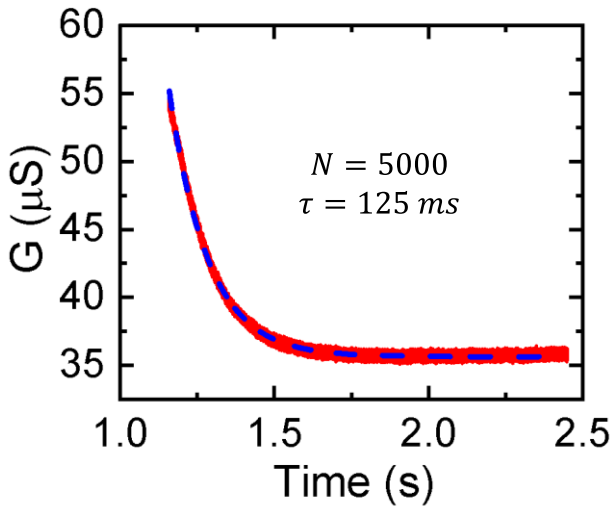

Figure 5.7. Relaxation time and fitting curve of EDL response after injecting 5000 potentiation voltage pulses (1V, 100 μs) shows a short-term effect with one time constant.

For E.I. gate and consequent electrochemical intercalation process, however, the retention time and synaptic weight raised by increasing the applied pulse number, as shown in **Figure 5.8**.

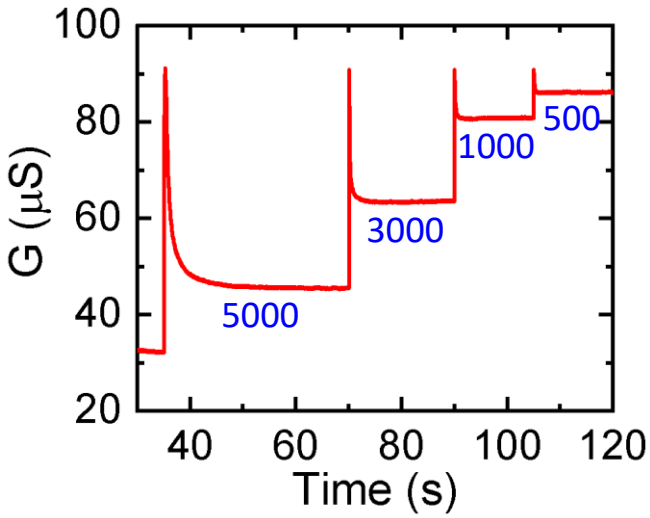

Figure 5.8. Conductance change (synaptic weight) for consecutive potential voltage pulse (+1V, 100 μs) with various number.

Physically, for the highest pulse number (N=5000) a large number of Li^+ ions were intercalated in the MoS_2 layers and due to their permanent accommodation, long term effects appeared. For the subsequent smaller pulse numbers, more Li^+ ions had the chance to be doped permanently in the vdW layers and cause enhancement in the long-term level. We plotted the fitted relaxation curve with respect to time for different pulse numbers (**Figure 5.9a-d)**. As seen, the two obtained time constants are clearly increasing as the pulse number increases. As depicted in **Figure 5.9c**, we observe a smaller time constant as the pulse number is decreasing because of the free Li^+ ions in the vdW layers that can diffuse back into the electrolyte.

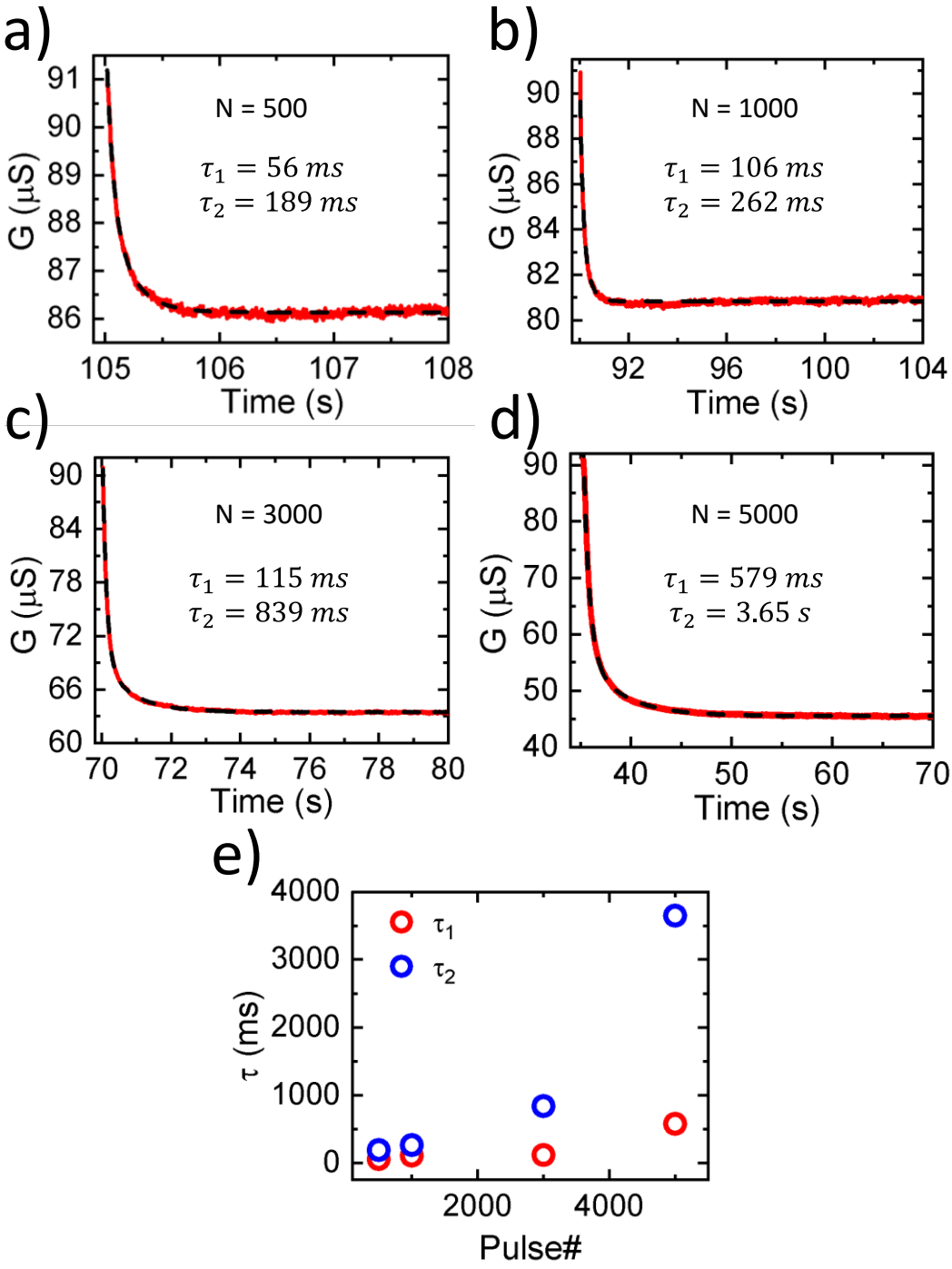

Figure 5.9. (a-d) Relaxation time and fitted parameters for various pulse number. e) Two obtained time constants versus pulse number.

5.6.1 Emulation of Learning, Forgetting, and Relearning Processes for Electrochemical Intercalation Process

Using the results shown in **Figures 5.8** and **5.9** we were able to demonstrate learning, forgetting, and relearning operations for our artificial synapse. As seen in **Figure 5.10**, following the synaptic weight improvement and relaxation decay for N = 5000, the synaptic weight reached its maximum weight after only N = 3000 stimulations, indicating a successful relearning process, similar to how our brain's learning of the forgotten information for the second time is faster than the first time [226,227].

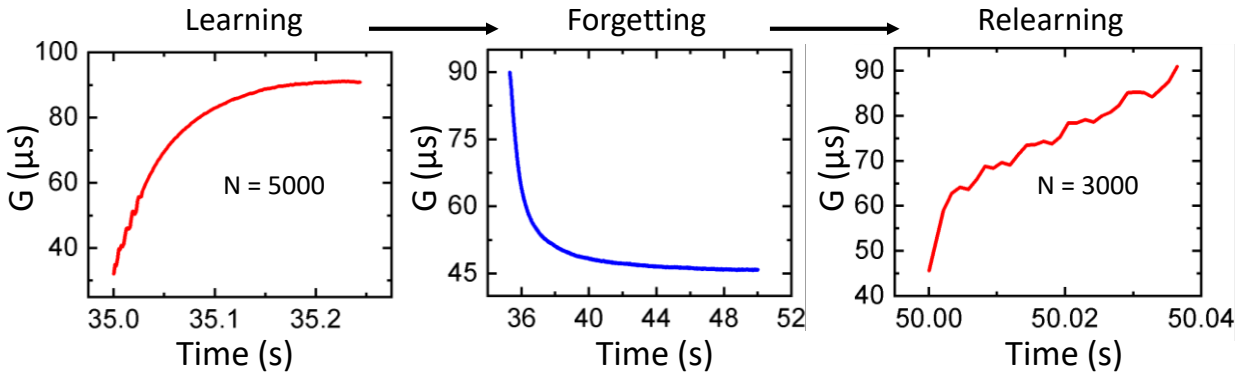

Figure 5.10. Emulation of Learning, forgetting, and relearning process of the human brain using MoS_2 artificial synapse.

5.7 Synaptic Weight Modulation by Changing Pulse Width and Amplitude

In a biological neuron, the pattern of activity and Ca^{+2} influx through receptors can change the synaptic plasticity at an excitatory synapse and the new neural network formation required for learning and memory [228]. This feature is investigated in our artificial synapse, where the channel conductance change (ΔG_{SD}) represents synaptic plasticity by varying the programming pulse width and amplitude applied to both side gates. As illustrated in **Figure 5.11** and **Figure 5.12**, we observed pseudo-linear and linear relationships between conductance change (ΔG_{SD}) with regard to pulse width (from 10 ms to 2 s, 1V) and pulse amplitude (from 0.5 V to 3 V, 20 ms), respectively, for both pure EDL (Au gate, **Figure 5.11a-d**) and combined EDL and intercalation (LFP/Au, **Figure 5.12a,b**). As expected, the amount of conductance change through the electrochemical intercalation effect is much higher than the sole EDL mechanism. The observed saturation for EDL in wider pulse widths (500 ms-2 s) as shown for a 2 s pulse (**Figure 5.11b**) is due to charge accumulation on the MoS_2 surface, as mentioned above for a large number of applied pulses. Likewise, the saturation observed for larger pulse widths during intercalation (**Figure 5.12a**) is attributed to the high concentration of intercalated Li^+ ions in the vdW layers of the MoS_2 channel and resulting lack of space for newly arriving mobile Li^+ ions. It is worth mentioning that the linear relationship between the channel conductance and pulse width amplitude indicates that we can further decrease the ΔG_{SD} step size and increase the number of nonvolatile distinct states in our ECRAM during the LTP and LTD process by simply tuning the pulse width or amplitude.

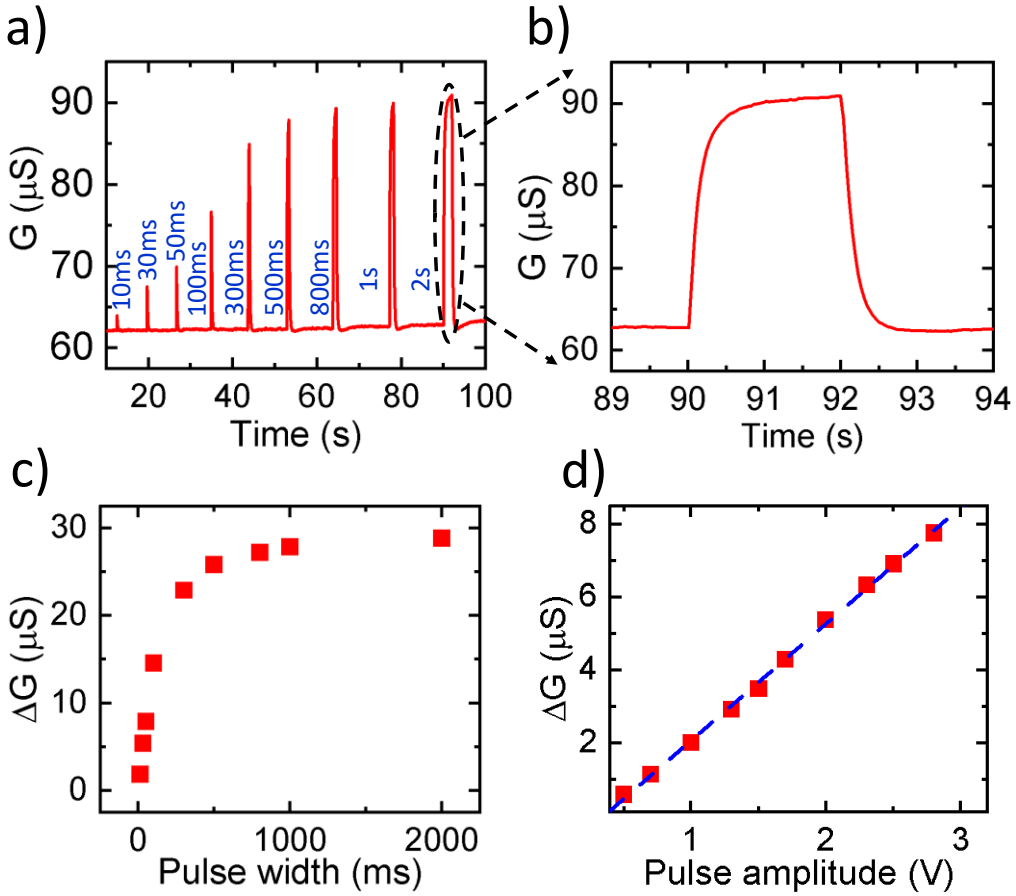

Figure 5.11. Synaptic weight (channel conductance) tuning in EDL based MoS2 synapse via pulse width/amplitude variation. a) MoS2 channel conductance for different pulse widths b) conductance saturation for 2s pulse width due to charge accumulation on MoS_2 surface c) synaptic weight change as a function of pulse width d) synaptic weight change with respect to different pulse amplitudes

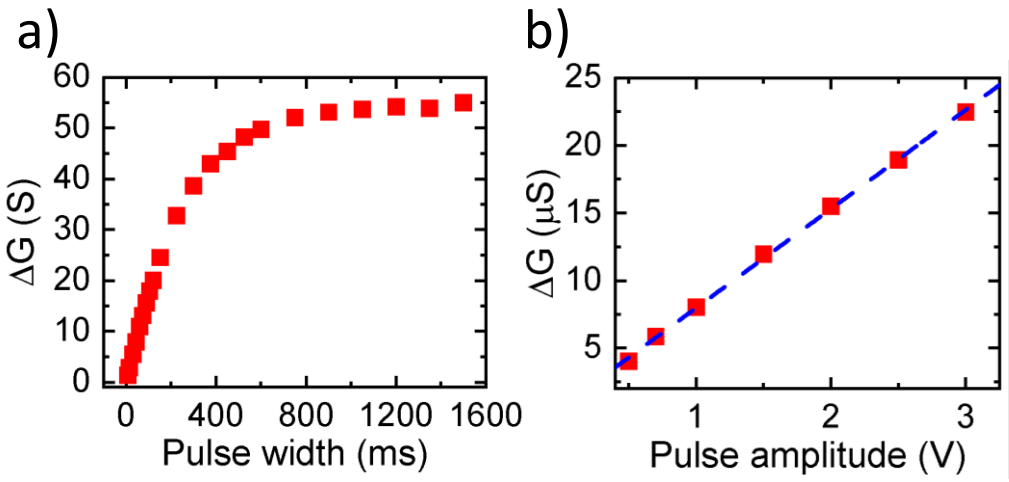

Figure 5.12. Synaptic weight modulation in electrochemical based MoS2 synapse via pulse amplitude variation. a) synaptic weight change as a function of pulse width. b) synaptic weight change with regard to transferred charge due to intercalation effect. c) synaptic weight change as a function of pulse amplitude.

5.8 Neural Facilitation Emulation and Time Constants Dependency to Pulse Width for SNN Applications

Paired-pulse facilitation (PPF), also known as neural facilitation, is a type of short-term synaptic plasticity in biological neurons that plays a key role in SNN applications such as learning, information processing, and sound-source localization [229,230]. The PPF exhibited in a biological synapse is based on two events within a short interval where the response to the second event intensifies with respect to the response to the first event. The reason for this enhancement is due to the liberation of Ca^{+2} caused by the second action potential that is added to those ions left over from the first stimulus [231,232]. This scenario was emulated for our artificial synapses for both side gates. We sent two consecutive voltage pulses (1V, 20 ms) with different time intervals

to the presynaptic electrode and monitored the channel conductance change (ΔG). As shown in **Figure 5.13a,b**, the synaptic weight (ΔG) was determined by the time interval (Δt) between the two pulses, and as Δt increased, the observed ΔG decreased, emulating the dynamic short-term phenomenon in biological neural networks [232]. This mimics the learning/memory mechanism in our brain where the reinforcing learning/memory effect is strong when two incident stimuluses are temporarily close. For the Au gate, and thus channel conductance induced by EDL as shown in **Figure 5.13a**, the conductance change (ΔG) followed an exponential decay ($\Delta G = Ce^{\frac{-\Delta t}{\tau}}$) and the time constant extracted was similar to the time scale in biological synapses (τ= 21 ms) [232]. In addition, as shown in **Figure 5.13c**, we showed that by increasing the pulse width in the PPF process, the time constant increases, which is due to more Li^+ ions accumulated at the interface and incomplete recovery before the second pulse.

To investigate the electrochemical intercalation effect in PPF, paired pulses with varied time intervals were sent to the LFP/Au electrode (E.I.) and similar to with the I.G. process, the conductance change followed an exponential decaying trend ($\Delta G = C_1\, e^{-\frac{\Delta t}{\tau_1}} + C_2\, e^{-\frac{\Delta t}{\tau_2}}$), as illustrated in **Figure 5.13b**; however this time there were two characteristic timescales $\tau_1 = 3\, ms$ and $\tau_2 = 44.9\, ms$. These values are analogous to those in biological synapses [113] and previous artificial synapses reported in the literature [24, 161,175,208]. These two-time scales, τ_1 and τ_2, are related to the diffusion process in the MoS_2 vdW layers and thus should be attributed to the device lateral dimension, width and length, respectively, as established in prior reports [24,175].

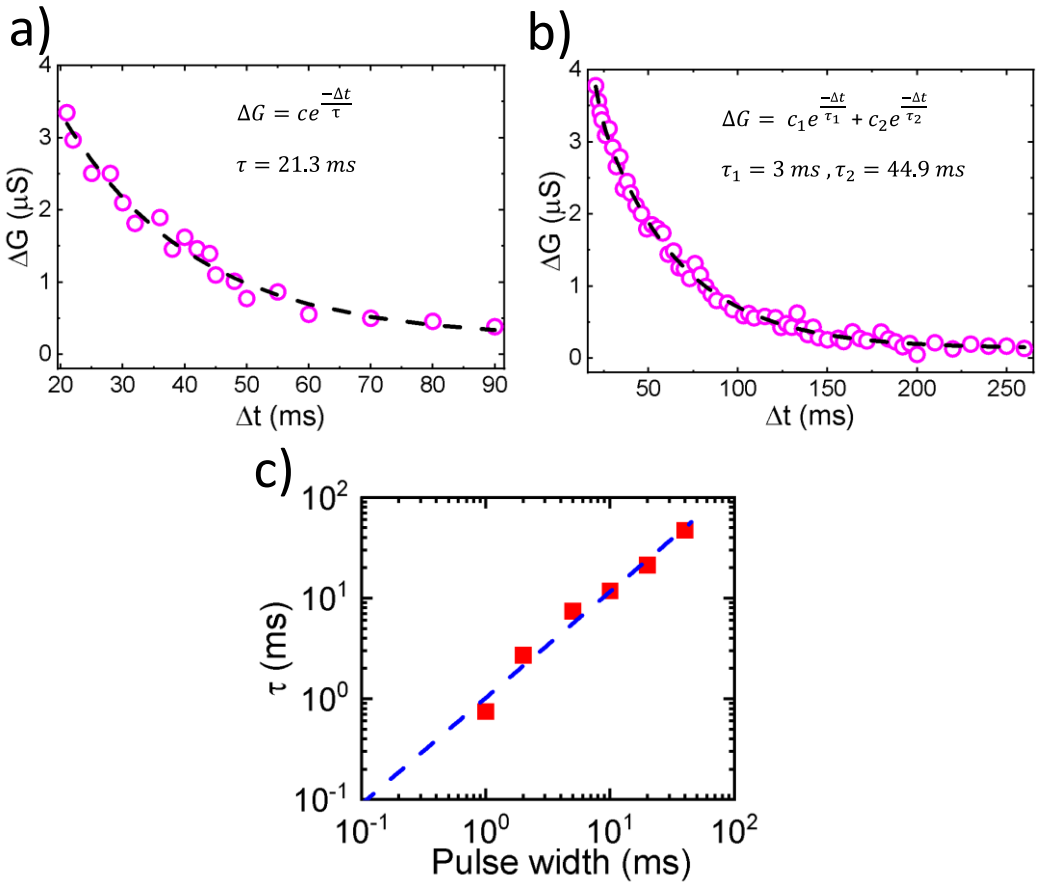

Figure 5.13. Time dependent synaptic facilitation for SNN applications and learning process by sending two consecutive stimuli. a) Paired pulse facilitation emulation in our MoS_2 synapse with one time scale fitting via ionic gating terminal. b) Pulse pair facilitation with two-time scales exponential decay fitting via E.I. gate terminal. c) Time scales obtained from PPF process in I.G. gate terminal with respect to the pulse widths

5.8.1 Programming Speed and Energy Scalability by Engineering the Device Dimension

We were able to engineer the switching speed of our MoS_2 artificial synapse by engineering the device size, as displayed in **Figure 5.14a,** for a wide range of dimensions. The power law

dependency of the time constants over the device dimensions confirms their relationships with the device size and geometry [24]. We project that by scaling the device's lateral dimension down to 10 nm × 10 nm, the programming speed can be scaled down to ≈ 1 GHz, comparable to dynamic random-access memory (DRAM) devices [121].

The switching energy for the smallest flake size (26 μm^2) is 4.5 pJ for the programming pulse with 1V amplitude and 20 ms width. To explore the scalability of the writing energy of the present ECRAM device, we monitored the switching energy ($E = I \cdot V \cdot t_w$) for different devices with different areas, as shown in **Figure 5.14b**. The projection reveals that the programming energy could be scaled down to 0.1 aJ for a device of 10 nm × 10 nm. This trend indicates that our artificial synapse is an extremely promising candidate for artificial neural networks applications. To study the temperature effect on the switching speed in the E.I. process, we performed PPF measurements at different temperatures. As shown in **Figure 5.14c**, by increasing the working temperature in a vacuum, we observed smaller time constants, indicating faster switching speed for our artificial synapse as the programming speed is highly dependent on the intrinsic electrolyte transport and ion diffusivity [26].

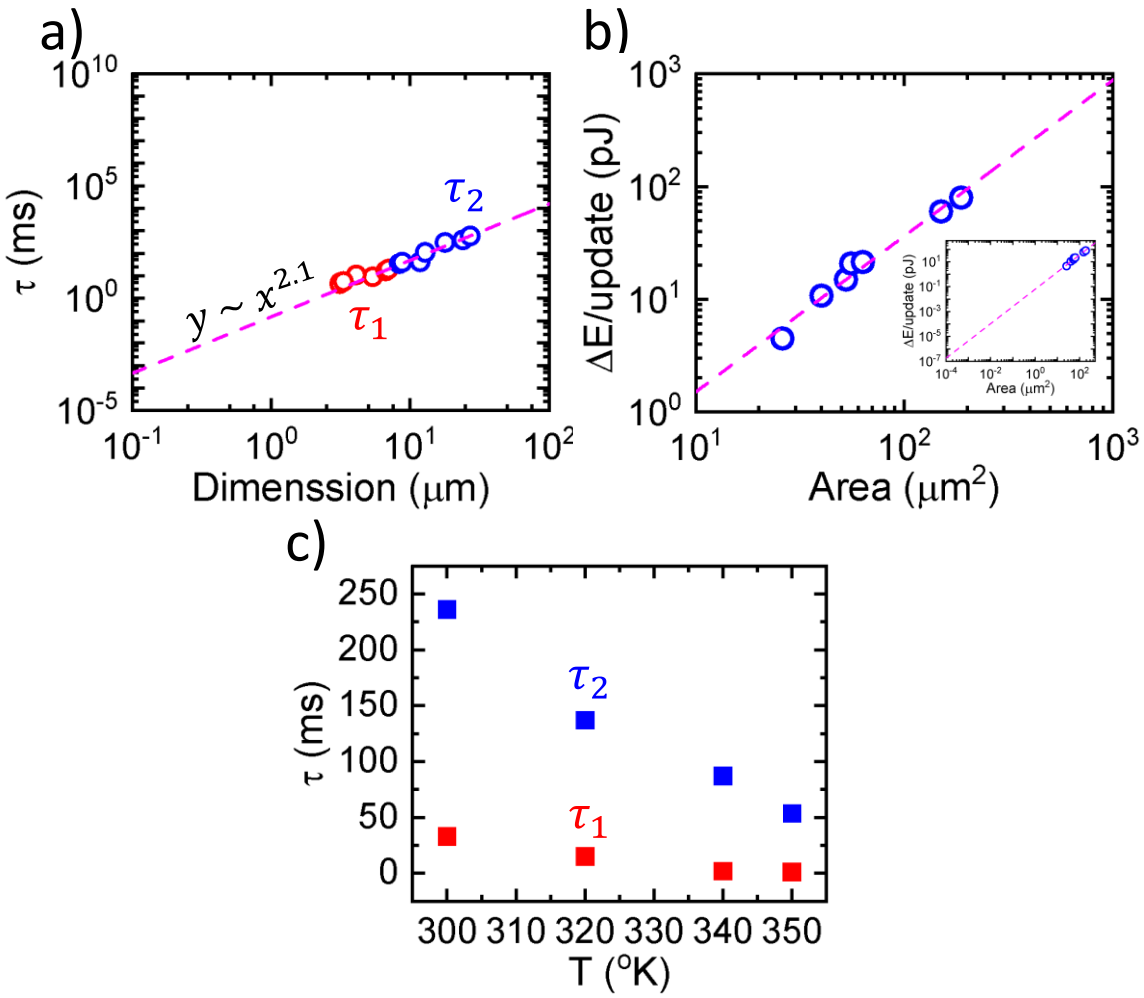

Figure 5.14. Scalable programming speed and energy. a) switching energy as a function of device lateral dimension. 1 GHz speed is achievable for device scaled down to 10 nm x 10 nm. b) Programming energy per event as a function of device area, indicating ultra-low switching energy (~ 0.1 aJ/update) for a device with 10 nm x 10 nm dimension. c) Time constant as a function of working temperature.

5.9 Spike Timing Dependent Plasticity for Electrochemical Intercalation Process

In addition to the PPF, we demonstrated spike timing dependent plasticity (STDP) as a Hebbian synaptic rule foundation for our MoS_2 artificial synapse. Of the four types of STDP in a

biological neuron [23], we exhibited the asymmetric STDP as the governing type in many systems [233-236]. To implement STDP in our artificial synapses, we sent different spikes as pre and post synaptic events ($|V_{pre}|$= 2V and $|V_{post}|$=1V, 20 ms) with various time differences to the presynaptic E.I. electrode using a commercially available multiplexer and monitored the synaptic change in response to the incident signals. We selected a saw-edged single pulse as V_{pre} and a square pulse as V_{post}, where the temporal correlations and timing between these two spikes ($\Delta t = V_{pre} - V_{post}$) is sent to the presynaptic gate and tunes the synaptic weight as shown in **Figure 5.15a.** When the $\Delta t > 0$ and V_{pre} precedes the V_{post}, long term potentiation occurs and synaptic weight increases (**Figure 5.15b**, bottom left). Conversely, when V_{pre} succeeds the V_{post} ($\Delta t < 0$) long-term depression ensues (**Figure 5.15b**, top-right). We then fit each potentiation using a two-term exponential function, and the characteristic time constants ($\tau_1 = 12\,ms$, $\tau_2 = 50.8\,ms$ for potentiation, $|\tau_1| = 11.6\,ms$, $|\tau_2| = 50.1\,ms$ for depression) are compatible with those of biological synapses. Moreover, the time constant scale with the device lateral dimension in each part indicates their scalability.

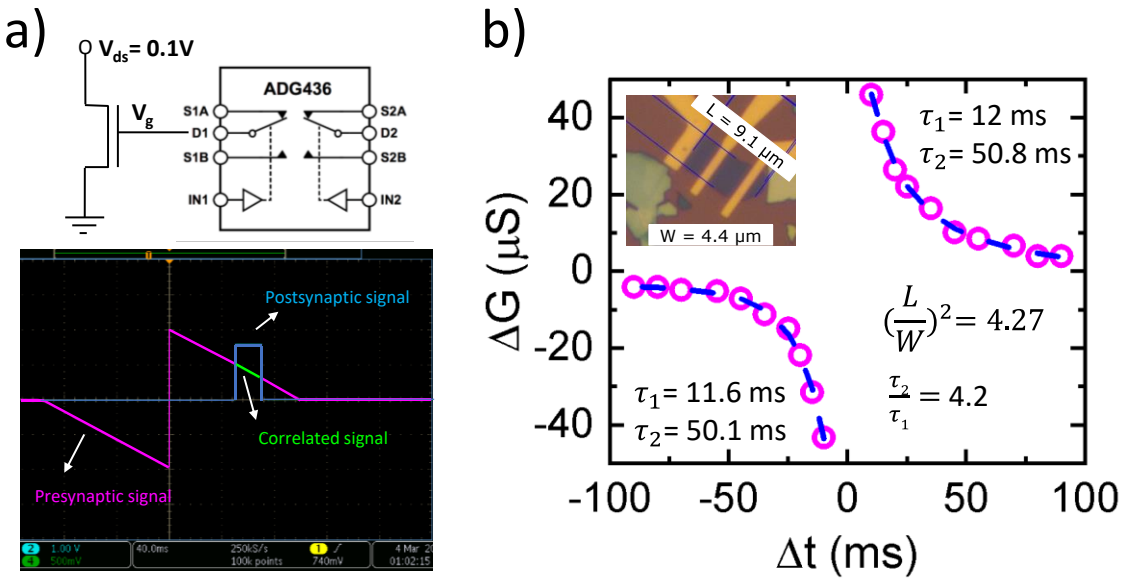

Figure 5.15. Spike timing dependent plasticity (STDP). a) Presynaptic and postsynaptic signals correlation to be sent to E.I. gate terminal using a multiplexer. b) Synaptic weight (channel conductance) change as a function of time interval between presynaptic and postsynaptic signals. In the asymmetric STDP, the obtained two-time constants in both sections of plasticity using two terms exponential decay equation are well matched with biological response. The ratio of the time constants $\frac{\tau 2}{\tau 1} = 4.2$ and the ratio of the lateral dimension of the device $(\frac{L}{W})^2 = 4.27$ indicates scalability of the programming speed and power.

5.10 High-Pass Dynamic Filtering Emulation and Short-Term to Long-Term Transition Investigation For The Both Ionic Gating and Electrochemical Intercalation Processes

In addition to the investigations on pulse number, width, and amplitude, we studied diffusive dynamic of the Li^+ ions as a result of stimulus frequency in both the EDL and

intercalation phenomena. A fixed number of consecutive voltage pulses (20 pulses, 1V, 20 ms) with various duty cycles were applied to the both presynaptic terminals. As illustrated in **Figure 5.16a**, for different frequencies (2 Hz < f < 200 Hz) applied to the I.G. gate we could not see long term effects, indicating short term facilitation and high pass dynamic filtering. This behavior demonstrates biological synaptic actions, which is desirable for information processing in SNN applications. Physically, the field induced ion migration to the MoS_2 channel and Li^+ ions absorbed to the surface temporally, increasing the synaptic weight. After removal of the driving force, all the mobile ions diffused back into the electrolyte, with a one time scale relaxation decay, and short-term potentiation happened independent of the pulse frequency.

However, the same scenario for the E.I. gate induced long term effects, and upon varying the duty cycle of the pulses (decreasing the time intervals) we detected monotonous enhancement in the synaptic weight change, much more than in the I.G. case. In addition, the long-term potentiation caused by Li^+ ion liberation and relaxation reduction indicates short term to long term transition, as illustrated in **Figure 5.16b**. For frequency range 2 Hz < f < 20 Hz no transition happened, and we observed short term plasticity independent of the pulse frequency useful for high pass filtering emulation. However, at higher frequency (> 30 Hz), it transitions from short term to long term effects, an important feature for learning processes and memory in an artificial synaptic device. The synaptic weight change for the EDL gate, as shown in **Figure 15c**, has a pseudo-linear relationship with frequency. For lower frequencies, a highly linear trend was observed. However, upon increasing the frequency (> 50 Hz), the higher concentration of mobile Li^+ ions at the MoS_2 surface caused a charge accumulation and the device's conductance change no longer followed a linear increase. Likewise, for the E.I. gate we observed a linear relationship between conductance change and frequencies lower than 20 Hz with a larger weight compared to EDL gate. However,

for frequencies larger than 20 Hz, a sharp increase was observed due to the electrochemical intercalation process and site availability for high concentration of Li^+ ions accommodation.

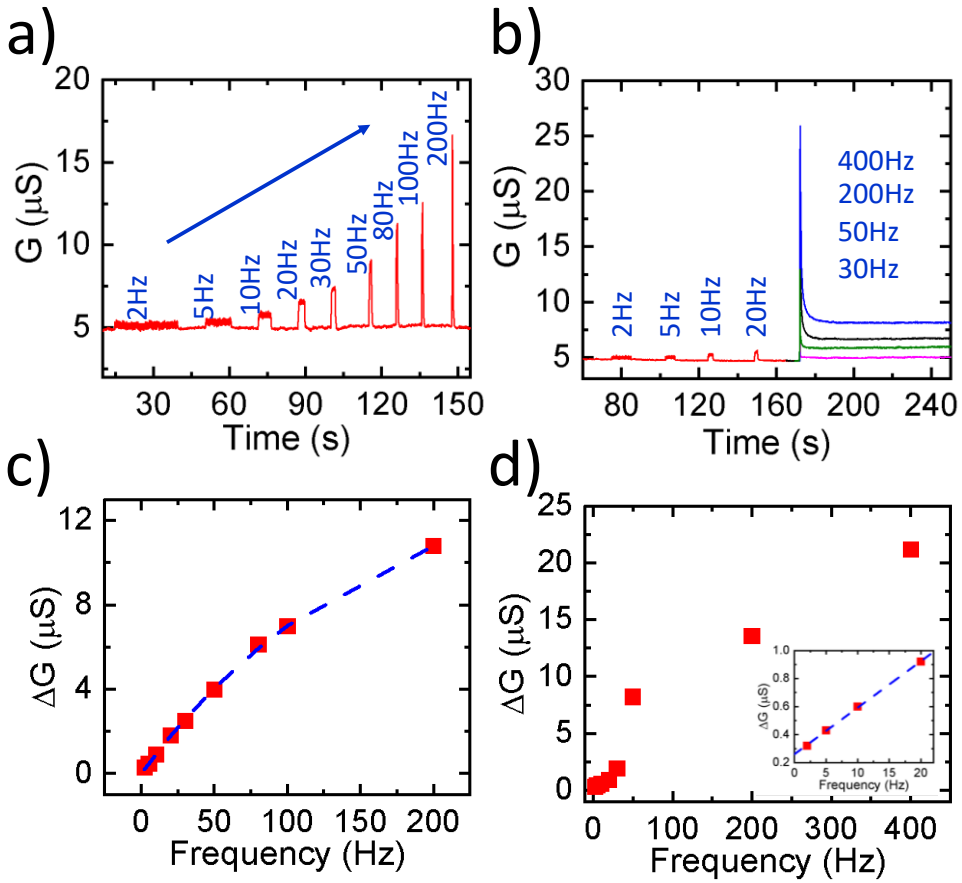

Figure 5.16. Frequency dependent synaptic weight for dynamic filtering emulation. a) Channel conductance as a function of frequency upon application of 20 potentiation pulses (1V, 20 ms) to EDL gate terminal. The short-term plasticity independent of the pulse frequency shows High-past the temporal filtering characteristics of our MoS_2 synapse. b) Channel conductance as a function of frequency demonstrates two regions, including short term plasticity as dynamic filtering emulation for SNN applications and long-term plasticity for learning and memory applications. c) Synaptic weight versus frequency originated from EDL gate. d) Synaptic weight with regards to frequency caused by E.I. gate electrode.

5.11 Conclusion

In summary, we report an artificial MoS_2 synapse with reversible modulation of synaptic weight through Li^+ ion electrochemical intercalation and EDL phenomena. Our MoS_2 synapse exhibited switching energy as efficient as 4.5 pJ (for smallest device, 26 μm^2) for a 1V, 10 ms pulse, and high precision (5000 states) upon administration of a100 µs pulse. In addition, our device performs with highly linear and symmetric responses, low variations, and good reliability ($> 2 \times 10^5$ s retention and $> 4 \times 10^4$ cycling endurance), making it a promising candidate for DNN applications. Moreover, the device displayed both STP and LTP caused by ionic gating and charge transfer doping, which is preferable for SNN applications. We demonstrate its capability for further biological synaptic functions, such as PPF, STDP as well as temporal filtering. The MoS_2 synaptic device has the potential to further increase to faster switching speeds (1 GHz) and lower energy consumption (0.1 aJ) per update by scaling down the device dimension to 10 nm × 10 nm.

6.0 Summary

In conclusion, we developed novel two-dimensional electrochemical devices as an artificial synapse using two dimensional dichalcogenides materials, including graphene and molybdenum disulfide (MoS_2). Our graphene synapse displayed an energy efficiency ~500 fJ/switching, analog tunability (>400 distinct states), a linear and symmetrical response, and good retention and endurance performances. We demonstrated their ability to perform basic neuronal functions, such as short-term and long-term potentiation/depression as well as STDP. Our graphene synapse also exhibited potential for scaling, where we project a nanoscale electrochemical synapse can operate with as low of a switching energy as 4 aJ per switching event at 10 MHz.

We also developed electrochemical synapses using MoS2, which is a transition metal dichalcogenide (TMD) with a bandgap, as another 2D material. MoS2 can be reversibly switched between a semiconducting (2H) and a metallic (1T) phase via intercalation. This phase transition provides another route to mimic synaptic activities in addition to the analog tuning. To have a comparable device in terms of switching speed with other non-volatile memories, we developed a fast speed measurement setup to improve the switching speed from milliseconds to nanoseconds. In MoS_2 electrochemical synapses, we designed three terminal synaptic transistors with two floating gates for ionic gating and electrochemical investigations. Our MoS2 artificial synapse could mimic basic functionalities of the human brain such as STP, LTP, PPF, STDP, SRDP, dynamic filtering. and STP to LTP transition. Furthermore, MoS_2 synapses exhibited ultra-fast switching speed (<100 ns), ultra-low energy consumption (< 700 aJ/switching), and extremely high precision (>5000) with extraordinary linearity and symmetricity, and good retention and endurance performance, making it a promising candidate for DNN applications. Moreover, we

project an efficient programming energy (~ 0.1 aJ) and speed (~ 1 GHz) per update for a device dimension scale down to 10 nm x 10nm.

With graphene and molybdenum disulfide's 2D flexible nature and CMOS compatibility, our electrochemical synapses provide exciting opportunities to realize hardware implementation of fully functioning artificial neural networks for neuromorphic computing.

Bibliography

[1] I. Boybat, M. Le Gallo, S. R. Nandakumar, T. Moraitis, T. Parnell, T. Tuma, B. Rajendran, Y. Leblebici, A. Sebastian, E. Eleftheriou, Nature Communications 2018, 9, 2514.

[2] K. Bourzac, Nature 2017, 551, 554.

[3] P. Huang, J. Kang, Y. Zhao, S. Chen, R. Han, Z. Zhou, Z. Chen, W. Ma, M. Li, L. Liu, X. Liu, Advanced Materials 2016, 28, 9758.

[4] M. S. Komar, Automatic Control and Computer Sciences 2017, 51, 701.

[5] G. Indiveri, S. Liu, Proceedings of the IEEE 2015, 103, 1379.

[6] M. Koziarski, B. Cyganek, Integrated Computer-Aided Engineering 2017, 24, 337.

[7] Z. Ansari, S. A. Seyyedsalehi, Neural Computing and Applications 2017, 28, 1177.

[8] J. Zhang, C. Zong, IEEE Intelligent Systems 2015, 30, 16.

[9] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, D. Hassabis, Nature 2016, 529, 484.

[10] G. W. Burr, R. M. Shelby, A. Sebastian, S. Kim, S. Kim, S. Sidler, K. Virwani, M. Ishii, P. Narayanan, A. Fumarola, L. L. Sanches, I. Boybat, M. Le Gallo, K. Moon, J. Woo, H. Hwang, Y. Leblebici, Advances in Physics: X 2017, 2, 89.

[11] D. Kuzum, S. Yu, H. S. Philip Wong, Nanotechnology 2013, 24, 382001.

[12] S. Kim, M. Ishii, S. Lewis, T. Perri, M. BrightSky, W. Kim, R. Jordan, G. W. Burr, N. Sosa, A. Ray, J. Han, C. Miller, K. Hosokawa, C. Lam, "NVM neuromorphic core with 64k-cell (256-by-256) phase change memory synaptic array with on-chip neuron circuits for continuous in-situ learning", presented at 2015 IEEE International Electron Devices Meeting (IEDM), 7-9 Dec. 2015, 2015.

[13] Z. Dong, Z. Zhou, Z. Li, C. Liu, P. Huang, L. Liu, X. Liu, J. Kang, IEEE Transactions on Electron Devices 2019, 66, 793.

[14] B. Pérez-Sánchez, O. Fontenla-Romero, B. Guijarro-Berdiñas, Artif. Intell. Rev. 2018, 49, 281.

[15] P. Chen, X. Peng, S. Yu, "NeuroSim+: An integrated device-to-algorithm framework for benchmarking synaptic devices and array architectures", presented at 2017 IEEE International Electron Devices Meeting (IEDM), 2-6 Dec. 2017, 2017.

[16] A. L. Caterini, D. E. Chang, Deep Neural Networks in a Mathematical Framework, Springer, 2018.

[17] S. Yu, Neuro-inspired computing using resistive synaptic devices, Springer, 2017.

[18] A. Azizimazreah, Y. Gu, X. Gu, L. Chen, "Tolerating Soft Errors in Deep Learning Accelerators with Reliable On-Chip Memory Designs", presented at 2018 IEEE International Conference on Networking, Architecture and Storage (NAS), 11-14 Oct. 2018, 2018.

[19] Q. Yu, R. Yan, H. Tang, K. C. Tan, H. Li, IEEE Transactions on Neural Networks and Learning Systems 2016, 27, 621.

[20] M. Giulioni, X. Lagorce, F. Galluppi, R. B. Benosman, Frontiers in Neuroscience 2016, 10.

[21] H. Markram, W. Gerstner, P. J. Sjöström, Frontiers in Synaptic Neuroscience 2011, 3.

[22] S. Lowel, W. Singer, Science 1992, 255, 209.

[23] Y. Dan, M.-M. Poo, Physiological Reviews 2006, 86, 1033.

[24] M. T. Sharbati, Y. Du, J. Torres, N. D. Ardolino, M. Yun, F. Xiong, Advanced Materials 2018, 30, 1870273.

[25] D. S. Jeong, Journal of Applied Physics 2018, 124, 152002.

[26] F. Akopyan, J. Sawada, A. Cassidy, R. Alvarez-Icaza, J. Arthur, P. Merolla, N. Imam, Y. Nakamura, P. Datta, G. Nam, B. Taba, M. Beakes, B. Brezzo, J. B. Kuang, R. Manohar, W. P. Risk, B. Jackson, D. S. Modha, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems 2015, 34, 1537.

[27] K. Boahen, Computing in Science & Engineering 2017, 19, 14.

[28] S. B. Furber, F. Galluppi, S. Temple, L. A. Plana, Proceedings of the IEEE 2014, 102, 652.

[29] A. D. Brown, J. E. Chad, R. Kamarudin, K. J. Dugan, S. B. Furber, IEEE Transactions on Multi-Scale Computing Systems 2018, 4, 450.

[30] S. R. Ovshinsky, Physical Review Letters 1968, 21, 1450.

[31] S. Lai, "Current status of the phase change memory and its future", presented at IEEE International Electron Devices Meeting 2003, 8-10 Dec. 2003, 2003.

[32] M. H. R. Lankhorst, B. W. S. M. M. Ketelaars, R. A. M. Wolters, Nature Materials 2005, 4, 347.

[33] J. H. Oh, J. H. Park, Y. S. Lim, H. S. Lim, Y. T. Oh, J. S. Kim, J. M. Shin, J. H. Park, Y. J. Song, K. C. Ryoo, D. W. Lim, S. S. Park, J. I. Kim, J. H. Kim, J. Yu, F. Yeung, C. W. Jeong, J. H. Kong, D. H. Kang, G. H. Koh, G. T. Jeong, H. S. Jeong, K. Kim, "Full Integration of Highly Manufacturable 512Mb PRAM based on 90nm Technology", presented at 2006 International Electron Devices Meeting, 11-13 Dec. 2006, 2006.

[34] R. Bez, "Chalcogenide PCM: a memory technology for next decade", presented at 2009 IEEE International Electron Devices Meeting (IEDM), 7-9 Dec. 2009, 2009.

[35] G. Servalli, "A 45nm generation Phase Change Memory technology", presented at 2009 IEEE International Electron Devices Meeting (IEDM), 7-9 Dec. 2009, 2009.

[36] H. P. Wong, S. Raoux, S. Kim, J. Liang, J. P. Reifenberg, B. Rajendran, M. Asheghi, K. E. Goodson, Proceedings of the IEEE 2010, 98, 2201.

[37] S. R. Ovshinsky, Journal of Non-Crystalline Solids 1970, 2, 99.

[38] M. Wuttig, N. Yamada, Nature Materials 2007, 6, 824.

[39] N. Yamada, E. Ohno, K. Nishiuchi, N. Akahira, M. Takao, Journal of Applied Physics 1991, 69, 2849.

[40] T. Tuma, A. Pantazi, M. Le Gallo, A. Sebastian, E. Eleftheriou, Nature Nanotechnology 2016, 11, 693.

[41] J. Gasia, M. Martin, A. Solé, C. Barreneche, F. L. Cabeza, Applied Sciences 2017, 7.

[42] G. Nicolescu, I. O'Connor, C. Piguet, Design technology for heterogeneous embedded systems, Springer, 2012.

[43] D. Kuzum, R. G. D. Jeyasingh, B. Lee, H. S. P. Wong, Nano Letters 2012, 12, 2179.

[44] C. D. Wright, P. Hosseini, J. A. V. Diosdado, Advanced Functional Materials 2013, 23, 2248.

[45] Y. Zhong, Y. Li, L. Xu, X. Miao, physica status solidi (RRL)–Rapid Research Letters 2015, 9, 414.

[46] A. Redaelli, D. Ielmini, A. L. Lacaita, F. Pellizzer, A. Pirovano, R. Bez, "Impact of crystallization statistics on data retention for phase change memories", presented at IEEE InternationalElectron Devices Meeting, 2005. IEDM Technical Digest., 5-5 Dec. 2005, 2005.

[47] G. W. Burr, M. J. Breitwisch, M. Franceschini, D. Garetto, K. Gopalakrishnan, B. Jackson, B. Kurdi, C. Lam, L. A. Lastras, A. Padilla, B. Rajendran, S. Raoux, R. S. Shenoy, Journal of Vacuum Science & Technology B 2010, 28, 223.

[48] C. Ahn, B. Lee, R. G. D. Jeyasingh, M. Asheghi, G. A. M. Hurkx, K. E. Goodson, H. S. Philip Wong, Journal of Applied Physics 2011, 110, 114520.

[49] S. Braga, A. Cabrini, G. Torelli, IEEE Transactions on Electron Devices 2011, 58, 517.

[50] Y. Shi, L. Nguyen, S. Oh, X. Liu, F. Koushan, J. R. Jameson, D. Kuzum, Nature Communications 2018, 9, 5312.

[51] S. H. Jo, T. Chang, I. Ebong, B. B. Bhadviya, P. Mazumder, W. Lu, Nano Letters 2010, 10, 1297.

[52] M. N. Kozicki, P. Mira, M. Mitkova, IEEE Transactions on Nanotechnology 2005, 4, 331.

[53] X. Xu, H. Lv, H. Liu, Q. Luo, T. Gong, M. Wang, G. Wang, M. Zhang, Y. Li, Q. Liu, S. Long, M. Liu, Nanoscale Research Letters 2015, 10, 61.

[54] M. Prezioso, F. Merrikh-Bayat, B. D. Hoskins, G. C. Adam, K. K. Likharev, D. B. Strukov, Nature 2015, 521, 61.

[55] R. Berdan, E. Vasilaki, A. Khiat, G. Indiveri, A. Serb, T. Prodromakis, Scientific Reports 2016, 6, 18639.

[56] S. M. Yu, Y. Wu, R. Jeyasingh, D. Kuzum, H. P. Wong, IEEE Transactions on Electron Devices 2011, 58, 2729.

[57] S. Ambrogio, S. Balatti, F. Nardi, S. Facchinetti, D. Ielmini, Nanotechnology 2013, 24, 384012.

[58] S. Ambrogio, S. Balatti, V. Milo, R. Carboni, Z. Wang, A. Calderoni, N. Ramaswamy, D. Ielmini, IEEE Transactions on Electron Devices 2016, 63, 1508.

[59] G. Pedretti, V. Milo, S. Ambrogio, R. Carboni, S. Bianchi, A. Calderoni, N. Ramaswamy, A. S. Spinelli, D. Ielmini, Scientific Reports 2017, 7, 5288.

[60] T. Chang, S.-H. Jo, W. Lu, ACS Nano 2011, 5, 7669.

[61] R. Yang, K. Terabe, Y. P. Yao, T. Tsuruoka, T. Hasegawa, J. K. Gimzewski, M. Aono, Nanotechnology, 2013, 24, 382001.

[62] S. Kim, S. Choi, W. Lu, ACS Nano 2014, 8, 2369.

[63] S. Kim, C. Du, P. Sheridan, W. Ma, S. Choi, W. D. Lu, Nano Letters 2015, 15, 2203.

[64] H. P. Wong, H. Lee, S. Yu, Y. Chen, Y. Wu, P. Chen, B. Lee, F. T. Chen, M. Tsai, Proceedings of the IEEE 2012, 100, 1951.

[65] R. Waser, R. Dittmann, G. Staikov, K. Szot, Advanced Materials 2009, 21, 2632.

[66] P. Huang, Y. Wang, H. Li, B. Gao, B. Chen, F. Zhang, L. Zeng, G. Du, J. Kang, X. Liu, IEEE Transactions on Nanotechnology 2014, 13, 1127.

[67] J. Jang, S. Park, G. W. Burr, H. Hwang, Y. Jeong, IEEE Electron Device Letters 2015, 36, 457.

[68] S. Park, M. Chu, J. Kim, J. Noh, M. Jeon, B. Hun Lee, H. Hwang, B. Lee, B.-g. Lee, Scientific Reports 2015, 5, 10123.

[69] Y.-F. Wang, Y.-C. Lin, I. T. Wang, T.-P. Lin, T.-H. Hou, Scientific Reports 2015, 5, 10150.

[70] E. J. Fuller, F. E. Gabaly, F. Léonard, S. Agarwal, S. J. Plimpton, R. B. Jacobs-Gedrim, C. D. James, M. J. Marinella, A. A. Talin, Advanced Materials 2017, 29, 1604310.

[71] Y. van de Burgt, E. Lubberman, E. J. Fuller, S. T. Keene, G. C. Faria, S. Agarwal, M. J. Marinella, A. Alec Talin, A. Salleo, Nature Materials 2017, 16, 414.

[72] J. Zhu, Y. Yang, R. Jia, Z. Liang, W. Zhu, Z. U. Rehman, L. Bao, X. Zhang, Y. Cai, L. Song, R. Huang, Advanced Materials 2018, 30, 1800195.

[73] D. Ielmini, A. L. Lacaita, Materials Today 2011, 14, 600.

[74] C. Gong, D. Ruzmetov, A. Pearse, D. Ma, J. N. Munday, G. Rubloff, A. A. Talin, M. S. Leite, ACS Applied Materials & Interfaces 2015, 7, 26007.

[75] N. J. Dudney, Materials Science and Engineering: B 2005, 116, 245.

[76] K. F. Mak, J. Shan, Nature Photonics 2016, 10, 216.

[77] X. Duan, C. Wang, A. Pan, R. Yu, X. Duan, Chemical Society Reviews 2015, 44, 8859.

[78] C. Muratore, V. Varshney, J. J. Gengler, J. J. Hu, J. E. Bultman, T. M. Smith, P. J. Shamberger, B. Qiu, X. Ruan, A. K. Roy, A. A. Voevodin, Applied Physics Letters 2013, 102, 081604.

[79] Q. H. Wang, K. Kalantar-Zadeh, A. Kis, J. N. Coleman, M. S. Strano, Nature Nanotechnology 2012, 7, 699.

[80] F. Xia, H. Wang, Y. Jia, Nature Communications 2014, 5, 4458.

[81] M.-H. Bae, Z. Li, Z. Aksamija, P. N. Martin, F. Xiong, Z.-Y. Ong, I. Knezevic, E. Pop, Nature Communications 2013, 4, 1734.

[82] C. Chiritescu, D. G. Cahill, N. Nguyen, D. Johnson, A. Bodapati, P. Keblinski, P. Zschack, Science 2007, 315, 351.

[83] A. A. Soluyanov, D. Gresch, Z. Wang, Q. Wu, M. Troyer, X. Dai, B. A. Bernevig, Nature 2015, 527, 495.

[84] I. G. Lezama, A. Arora, A. Ubaldini, C. Barreteau, E. Giannini, M. Potemski, A. F. Morpurgo, Nano Letters 2015, 15, 2336.

[85] W. Jin, P.-C. Yeh, N. Zaki, D. Zhang, J. T. Sadowski, A. Al-Mahboob, A. M. van der Zande, D. A. Chenet, J. I. Dadap, I. P. Herman, P. Sutter, J. Hone, R. M. Osgood, Physical Review Letters 2013, 111, 106801.

[86] H. Tian, W. Mi, X.-F. Wang, H. Zhao, Q.-Y. Xie, C. Li, Y.-X. Li, Y. Yang, T.-L. Ren, Nano Letters 2015, 15, 8013.

[87] Y. Shi, X. Liang, B. Yuan, V. Chen, H. Li, F. Hui, Z. Yu, F. Yuan, E. Pop, H. S. P. Wong, M. Lanza, Nature Electronics 2018, 1, 458.

[88] K. Ba, W. Jiang, J. Cheng, J. Bao, N. Xuan, Y. Sun, B. Liu, A. Xie, S. Wu, Z. Sun, Scientific Reports 2017, 7, 45584.

[89] M. Xu, T. Liang, M. Shi, H. Chen, Chemical Reviews 2013, 113, 3766.

[90] V. Tran, R. Soklaski, Y. Liang, L. Yang, Physical Review B 2014, 89, 235319.

[91] L. Sun, Y. Zhang, G. Hwang, J. Jiang, D. Kim, Y. A. Eshete, R. Zhao, H. Yang, Nano Letters 2018, 18, 3229.

[92] A. J. Arnold, A. Razavieh, J. R. Nasr, D. S. Schulman, C. M. Eichfeld, S. Das, ACS Nano 2017, 11, 3110.

[93] H. Tian, Q. Guo, Y. Xie, H. Zhao, C. Li, J. J. Cha, F. Xia, H. Wang, Advanced Materials 2016, 28, 4991.

[94] H. Tian, X. Cao, Y. Xie, X. Yan, A. Kostelec, D. DiMarzio, C. Chang, L.-D. Zhao, W. Wu, J. Tice, J. J. Cha, J. Guo, H. Wang, ACS Nano 2017, 11, 7156.

[95] W. Z. Bao, J. Y. Wan, X. G. Han, X. H. Cai, H. L. Zhu, D. H. Kim, D. K. Ma, Y. L. Xu, J. N. Munday, H. D. Drew, M. S. Fuhrer, and L. B. Hu, *Nature Communications*, 2014, 5, 4224.

[96] M. S. Dresselhaus, G. Dresselhaus, *Adv Phys* 1981, 30, 139.

[97] M. D. Levi, E. Markevich, A. Aurbach, *J Phys Chem B* 2005, 109, 7420.

[98] M. Kuhne, F. Paolucci, J. Popovic, P. M. Ostrovsky, J. Maier, J. H. Smet, *Nat Nanotechnol* 2017, 12, 895.

[99] V. Gorodokin, D. Zemlyanov, presented at *2004 23rd IEEE Convention of Electrical and Electronics Engineers in Israel*, 6-7 Sept. 2004.

[100] M. H. Bae, Z. Y. Li, Z. Aksamija, P. N. Martin, F. Xiong, Z. Y. Ong, I. Knezevic, E. Pop, *Nat Commun* 2013, 4.

[101] H. Jang, Y. J. Park, X. Chen, T. Das, M. S. Kim, J. H. Ahn, *Adv Mater* 2016, 28, 4184.

[102] S. Kim, S. Choi, W. Lu, *Acs Nano* 2014, 8, 2369.

[103] D. Ielmini, A. L. Lacaita, *Mater Today* 2011, 14, 600.

[104] G. Zhu, J. Liu, Q. Zheng, R. Zhang, D. Li, D. Banerjee, D. G. Cahill, *Nat Commun* 2016, 7, 13211; J. S. Kang, M. Ke, Y. Hu, *Nano Lett* 2017, 17, 1431; A. Sood, F. Xiong, H. Wang, Y. Cui, E. Pop, K. E. Goodson, in *MRS Spring Meeting*, Phoenix, AZ 2017.

[105] Zou J, Sole C, Drewett NE, Velický M, Hardwick LJ. *The Journal of Physical Chemistry Letters* 2016, 7(21): 4291-4296.

[106] F. Xiong, H. T. Wang, X. G. Liu, J. Sun, M. Brongersma, E. Pop, Y. Cui, *Nano Lett* 2015, 15, 6777.

[107] J. M. Thomas, G. R. Millward, R. F. Schlögl, H. P. Boehm, *Materials Research Bulletin* 1980, 15, 671.

[108] M. Inaba, H. Yoshida, Z. Ogumi, T. Abe, Y. Mizutani, M. Asano, *J Electrochem Soc* 1995, 142, 20.

[109]J. Zou, C. Sole, N. E. Drewett, M. Velický, L. J. Hardwick, *The Journal of Physical Chemistry Letters* 2016, 7, 4291.

[110] Cohn AP, Share K, Carter R, Oakes L, Pint CL. *Nano Lett* 2016, 16(1): 543-548.

[111] A. E. Pereda, *Nat Rev Neurosci* 2014, 15, 250.

[112] G. Zhong, X. Wu, L. D'Arsie, K. B. K. Teo, N. L. Rupesinghe, A. Jouvray, J. Robertson, *Appl Phys Lett* 2016, 109, 193103.

[113] M. V. L. Bennett, R. S. Zukin, *Neuron*, 41, 495.

[114] B. Linares-Barranco, T. Serrano-Gotarredona, L. Camuñas-Mesa, J. Perez-Carrasco, C. Zamarreño-Ramos, T. Masquelier, *Frontiers in Neuroscience* 2011, 5.

[115] J. T. Ye, Y. J. Zhang, R. Akashi, M. S. Bahramy, R. Arita, Y. Iwasa, *Science* 2012, 338, 1193; J. M. Lu, O. Zheliuk, I. Leermakers, N. F. Q. Yuan, U. Zeitler, K. T. Law, J. T. Ye, *Science* 2015, 350, 1353.

[116] Y. Ben-Ari, *Nat Rev Neurosci* 2002, *3*, 728.

[117] P. A. Merolla, J. V. Arthur, R. Alvarez-Icaza, A. S. Cassidy, J. Sawada, F. Akopyan, B. L. Jackson, N. Imam, C. Guo, Y. Nakamura, B. Brezzo, I. Vo, S. K. Esser, R. Appuswamy, B. Taba, A. Amir, M. D. Flickner, W. P. Risk, R. Manohar, D. S. Modha, *Science* 2014, *345*, 668.

[118] S. H. Choi, J. Kim, Y. S. Yoon, *J Power Sources* 2004, 138, 283.

[119] J. S. Gnanaraj, R. W. Thompson, J. F. DiCarlo, K. M. Abraham, *J Electrochem Soc* 2007, 154, A185.

[120] N. Kularatna, *IEEE Instrumentation & Measurement Magazine* 2011, 14, 20.

[121] V. V. Zhirnov, R. K. Cavin, S. Menzel, E. Linn, S. Schmelzer, D. Brauhaus, C. Schindler, R. Waser, *Proceedings of the IEEE* 2010, 98, 2185.

[122] Q. Wan, M. T. Sharbati, J. R. Erickson, Y. Du, F. Xiong, *Adv. Mater. Technol.* 2019, *4*, 1900037.

[123] K. Roy, A. Jaiswal, P. Panda, *Nature* 2019, *575*, 607.

[124] Tathagata Paul , TanweerAhmed , KrishnaKanhaiya Tiwari , Chetan Singh ThakurcandArindam Ghosh, *2D Materials*, 2019, 6, 045008.

[125] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, 2016, *IEEE- CVPR*, *6*, 770.

[126] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, *NIPS*, 2012, *6*, 84.

[127] Geoffrey Hinton, li Deng, Dong Yu, George Dahl, Abdelrahman Mohamed, Navdeep Jaitly, Andrew Senior, Vincent Vanhoucke, Phuongtrang Nguyen, Tara Sainath, and Brian Kingsbury, *IEEE signal processing magazine*, 2012, *29*, 82.

[128] Yuxuan Wang, R. J. Skerry-Ryan, Daisy Stanton, Yonghui Wu, Ron J. Weiss, Navdeep Jaitly, Zongheng Yang, Ying Xiao, Zhifeng Chen, Samy Bengio, Quoc V. Le, Yannis Agiomyrgiannakis, Rob Clark, and Rif A. Saurous, 2017, arXiv:1703.10135.

[129] Y. LeCun, Y. Bengio, G. Hinton, *Nature* 2015, *521*, 436.

[130] A. Krizhevsky, I. Sutskever, G. E. Hinton, *Commun. ACM* 2017, *60*, 84

[131] G. Snider, R. Amerson, D. Carter, H. Abdalla, M. S. Qureshi, J. Leveille, M. Versace, H. Ames, S. Patrick, B. Chandler, A. Gorchetchnikov, E. Mingolla, *Computer,* 2011, *44*, 21.

[132] D. Kuzum, S. M. Yu, H. S. P. Wong, *Nanotechnology* 2013, *24*.

[133] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova, 2019, arXiv:1810.04805.

[134] S. R. Nandakumar, Irem Boybat, Manuel Le Gallo, Evangelos Eleftheriou, Abu Sebastian, and Bipin Rajendran, *Sci. Rep*., 2020, *10*.

[135] K. Roy, A. Jaiswal, P. Panda, *Nature* 2019, *575*, 607.

[136] X. Lagorce, G. Orchard, F. Galluppi, B. E. Shi, R. B. Benosman, *IEEE Trans. Pattern Anal. Mach. Intell.* 2017, *39*, 1346.

[137] X. Lagorce, R. Benosman, *Neural Comput.* 2015, *27*, 2261.

[138] P. A. Merolla, J. V. Arthur, R. Alvarez-Icaza, A. S. Cassidy, J. Sawada, F. Akopyan, B. L. Jackson, N. Imam, C. Guo, Y. Nakamura, B. Brezzo, I. Vo, S. K. Esser, R. Appuswamy, B. Taba, A. Amir, M. D. Flickner, W. P. Risk, R. Manohar, D. S. Modha, *Science* 2014, *345*, 668.

[139] T. Tuma, A. Pantazi, M. Le Gallo, A. Sebastian, E. Eleftheriou, *Nat Nanotechnol* 2016, *11*, 693.

[140] Z. Wang, S. Joshi, S. Savel'ev, W. Song, R. Midya, Y. Li, M. Rao, P. Yan, S. Asapu, Y. Zhuo, H. Jiang, P. Lin, C. Li, J. H. Yoon, N. K. Upadhyay, J. Zhang, M. Hu, J. P. Strachan, M. Barnell, Q. Wu, H. Wu, R. S. Williams, Q. Xia, J. J. Yang, *Nature Electronics* 2018, *1*, 137.

[141] Z. Wang, B. Crafton, J. Gomez, R. Xu, A. Luo, Z. Krivokapic, L. Martin, S. Datta, A. Raychowdhury, A. I. Khan, presented at *2018 IEEE International Electron Devices Meeting (IEDM)*, 1-5 Dec., 2018.

[142] P. A. Merolla, J. V. Arthur, R. Alvarez-Icaza, A. S. Cassidy, J. Sawada, F. Akopyan, B. L. Jackson, N. Imam, C. Guo, Y. Nakamura, B. Brezzo, I. Vo, S. K. Esser, R. Appuswamy, B. Taba, A. Amir, M. D. Flickner, W. P. Risk, R. Manohar, D. S. Modha, Science, 2014, *345*, 668.

[143] Davies, Mike, et al., *IEEE Micro*, 2018, *38*, 82-99.

[144] S. B. Furber, D. R. Lester, L. A. Plana, J. D. Garside, E. Painkras, S. Temple, and A. D. Brown, *IEEE Transactions on Computers*, 2013, *62*, 2454-2467.

[145] N. K. Upadhyay, S. Joshi, J. J. Yang, *Sci. China Inf. Sci.* 2016, *59*, 061404.

[146] D. B. Strukov, G. S. Snider, D. R. Stewart, R. S. Williams, *Nature*, 2008, *453*, 80.

[147] J. J. S. Yang, D. B. Strukov, D. R. Stewart, *Nat. Nanotechnol*., 2013, *8*, 13
[148] Z. Wang, S. Joshi, S. Savel'ev, W. Song, R. Midya, Y. Li, M. Rao, P. Yan, S. Asapu, Y. Zhuo, H. Jiang, P. Lin, C. Li, J. H. Yoon, N. K. Upadhyay, J. Zhang, M. Hu, J. P. Strachan, M. Barnell, Q. Wu, H. Wu, R. S. Williams, Q. Xia, J. J. Yang, *Nat. Electron*. 2018, *1*, 137.

[149] G. W. Burr, R. M. Shelby, S. Sidler, C. di Nolfo, J. Jang, I. Boybat, R. S. Shenoy, P. Narayanan, K. Virwani, E. U. Giacometti, B. N. Kurdi, H. Hwang, *IEEE Trans. Electron Dev.* 2015, *62*, 3498.

[150] D. Kuzum, R. G. D. Jeyasingh, S. M. Yu, H. S. P. Wong, *IEEE Trans. Electron Dev.* 2012, *59*, 3489.

[151] S. La Barbera, D. Vuillaume, F. Alibart, *ACS Nano*, 2015, *9*, 941.

[152] Y. Yang, P. Gao, L. Li, X. Pan, S. Tappertzhofen, S. Choi, R. Waser, I. Valov, W. D. Lu, *Nat. Commun*. 2014, *5*, 4232.

[153] D. Kuzum, S. M. Yu, H. S. P. Wong, *Nanotechnology* 2013, *24*, 382001.

[154] B. L. Jackson, B. Rajendran, G. S. Corrado, M. Breitwisch, G. W. Burr, R. Cheek, K. Gopalakrishnan, S. Raoux, C. T. Rettner, A. Padilla, A. G. Schrott, R. S. Shenoy, B. N. Kurdi, C. H. Lam, D. S. Modha, ACM J. *Emerging Technol. Comput*. 2013, *9*, 12.

[155] D. S. Jeong, R. Thomas, R. S. Katiyar, J. F. Scott, H. Kohlstedt, A. Petraru, C. S. Hwang, Rep. Prog. Phys. 2012, 75, 076502.

[156] D. Ielmini, *Semicond. Sci. Technol.* 2016, *31*, 063002.

[157] Y. Li, E. J. Fuller, S. Asapu, S. Agarwal, T. Kurita, J. J. Yang, A. A. Talin, *ACS Appl. Mater. Interfaces* 2019, *11*, 38982.

[158] E. J. Fuller, S. T. Keene, A. Melianas, Z. Wang, S. Agarwal, Y. Li, Y. Tuchman, C. D. James, M. J. Marinella, J. J. Yang, A. Salleo, A. A. Talin, *Science* 2019, *364*, 570.

[159] C.-S. Yang, D.-S. Shang, N. Liu, E. J. Fuller, S. Agrawal, A. A. Talin, Y.-Q. Li, B.-G. Shen, Y. Sun, *Adv. Funct. Mater.* 2018, *28*, 1804170.

[160] C. Qian, J. Sun, L.-a. Kong, G. Gou, J. Yang, J. He, Y. Gao, Q. Wan, *ACS Appl. Mater. Interfaces* 2016, *8*, 26169.

[161] Q. Wan, P. Zhang, Q. Shao, M. T. Sharbati, J. R. Erickson, K. L. Wang, F. Xiong, *APL Mater.* 2019, *7*, 101107.

[162] Armantas Melianas, Min-A Kang, Armin VahidMohammadi, Weiqian Tian, Yury Gogotsi, Alberto Salleo, Mahiar Max Hamedi, *arXiv:2104.05396*, 2021.

[163] K. Xu, D. Chen, F. Yang, Z. Wang, L. Yin, F. Wang, R. Cheng, K. Liu, J. Xiong, Q. Liu, J. He, Nano Lett. 2017, 17, 1065.

[164] Y. Zhou, M. Zhang, Z. Guo, L. Miao, S.-T. Han, Z. Wang, X. Zhang, H. Zhang, Z. Peng, Mater. Horiz. 2017, 4, 997.

[165] H. Liu, A. T. Neal, P. D. Ye, ACS Nano 2012, 6, 8563.

[166] K. K. Young, *IIEEE Trans. Electron Devices*, 1989, *36*, 399.

[167] S. B. Desai et al., *Science*, 2016, *354*, 99.

[168] H. Liu, A. T. Neal and P. D. Ye, *ACS Nano*, 2012, *6*, 8563.

[169] Y. Wang, J. C. Kim, R. J. Wu, J. Martinez, X. Song, J. Yang, F. Zhao, A. Mkhoyan, H. Y. Jeong, M. Chhowalla, *Nature*, 2019, *568*, 70.

[170] J. Xia, J. Wang, D. Chao, Z. Chen, Z. Liu, J.-L. Kuo, J. Yan, Z. X. Shen, *Nanoscale*, 2017, *9*, 7533.

[171] M. Chhowalla, Z. F. Liu, and H. Zhang, *Chemical Society Reviews*, 2015, *44*, 2584-2586.

[172] H. S. S. R. Matte, A. Gomathi, A. K. Manna, D. J. Late, R. Datta, S. K. Pati, and C. N. R. Rao, *Angewandte Chemie-International Edition*, 2010, *49*, 4059-4062.

[173] J. Hu, Z. K. Guo, P. E. Mcwilliams, J. E. Darges, D. L. Druffel, A. M. Moran, and S. C. Warren, *Nano Letters*, 2016, *16*, 74-79.

[174] X. Zhu, D. Li, X. Liang, W. D. Lu, *Nat. Mater.* 2019, *18*, 141.

[175] Qingzhou Wan, Marco Rasetto, Mohammad Taghi Sharbati, John R. Erickson, Sridhar Reddy Velagala, Matthew T. Reilly, Yiyang Li, Ryad Benosman, Feng Xiong, *Advanced Intelligent Systems*, 2021, 2100021.

[176] L. Noerochim, A. O. Yurwendra, and D. Susanti, *Ionics*, 2016, *22*, 341-346.

[177] Z. J. Li, Z. Z. Peng, H. Zhang, T. Hu, M. M. Hu, K. J. Zhu, and X. H. Wang, *Nano Letters*, 2016, *16*, 795-799.

[178] J. Y. Wan, W. Z. Bao, Y. Liu, J. Q. Dai, F. Shen, L. H. Zhou, X. H. Cai, D. Urban, Y. Y. Li, K. Jungjohann, M. S. Fuhrer, and L. B. Hu, *Advanced Energy Materials*, 2015, *5*.

[179] X. S. Zhou, L. J. Wan, and Y. G. Guo, *Nanoscale*, 2012, *4*, 5868-5871.

[180] L. F. Abbott, W. G. Regehr, *Nature*, 2004, *431*, 796−803.

[181] Y. P. Tang, E. Shimizu, G. R. Dube, C. Rampon, G. A. Kerchner, M. Zhuo, G. S. Liu, J. Z. Tsien, *Nature* 1999, *401*, 63−69.

[182] Ke Xu, Md Mahbubul Islam, David Guzman, Alan C. Seabaugh, Alejandro Strachan, and Susan K. Fullerton-Shirey, *ACS Appl. Mater. Interfaces*, 2018, *10*, 43166−43176.

[183] T. Chang, S. H. Jo, W. Lu, *ACS Nano*, 2011, *5*, 7669−7676.

[184] R. S. Zucker, W. G. Regehr, *Annu. Rev. Physiol.*, 2002, *64*, 355−405.

[185] R. Lamprecht, J. LeDoux, *Nat. Rev. Neurosci*. 2004, *5*, 45−54.

[186] S. Li, F. Zeng, C. Chen, H. Liu, G. Tang, S. Gao, C. Song, Y. Lin, F. Pan, D. Guo, *J. Mater. Chem. C*, 2013, *1*, 5292−5298.

[187] P.-Y. Chen, X. Peng, S. Yu, NeuroSim+, *IEEE International Electron Devices Meeting (IEDM)*, 2017, San Francisco, USA.

[188] Marco Rasetto, Qingzhou Wan, Himanshu Akolkar, Feng Xiong and Ryad Benosman, 2021.

[189] A. A. Faisal, L. P. Selen, and D. M. Wolpert, Nature *reviews neuroscience*, 2008, *9*, 292-303.

[190] G. An, *Neural computation*, 1996, *8*, 643-674.

[191] B. Poole, J. Sohl-Dickstein, and S. Ganguli, *arXiv preprint arXiv:1406.1831*, 2014.

[192] C. M. Bishop, *Neural computation*, 1995, *7*, 108-116.

[193] A. Neelakantan, L. Vilnis, Q. V. Le, I. Sutskever, L. Kaiser, K. Kurach, and J. Martens, *arXiv preprint arXiv:1511.06807*, 2015.

[194] G. Orchard, A. Jayawant, G. K. Cohen, and N. Thakor, *Frontiers in neuroscience*, 2015, *9*, 437.

[195] E. R. Kandel, J. H. Schwartz, T. M. Jessell, S. A. Siegelbaum, A. J. Hudspeth, *Principles of Neural Science*, McGraw-Hill, New York 2013.

[196] X. Zhu, W. D. Lu, *ACS Nano* 2018, *12*, 1242.

[197] X. Yang, Y. Fang, Z. Yu, Z. Wang, T. Zhang, M. Yin, Y. Yang, Y. Cai, R. Huang, *Nanoscale* 2016, *8*, 18897.

[198] C. Li, D. Belkin, Y. Li, P. Yan, M. Hu, N. Ge, H. Jiang, E. Montgomery, P. Lin, Z. Wang, W. Song, J. P. Strachan, M. Barnell, Q. Wu, R. S. Williams, J. J. Yang, Q. Xia, *Nat. Commun.* 2018, *9*, 2385.

[199] Q. Zheng, Z. Wang, N. Gong, Z. Yu, C. Chen, Y. Cai, Q. Huang, H. Jiang, Q. Xia, R. Huang, *IEEE Electron Device Lett.* 2019, *40*, 1309.

[200] T. Ohno, T. Hasegawa, T. Tsuruoka, K. Terabe, J. K. Gimzewski, M. Aono, *Nat. Mater.* 2011, *10*, 591.

[201] D. B. Strukov, *Nature* 2011, *476*, 403.

[202] S. H. Jo, T. Chang, I. Ebong, B. B. Bhadviya, P. Mazumder, W. Lu, *Nano Lett.* 2010, *10*, 1297.

[203] G. Q. Bi, M. M. Poo, *J. Neurosci.* 1998, *2*, 10464.

[204] D. Purves, G. J. Augustine, D. Fitzpatrick, W. C. Hall, A.-S. LaMantia, J. O. McNamara, S. M. Williams, *Neuroscience*, 3rd ed., Sinauer Associates, Sunderland, MA 2004.

[205] X. Yang, Y. Fang, Z. Yu, Z. Wang, T. Zhang, M. Yin, Y. Yang, Y. Cai, R. Huang, *Nanoscale* 2016, *8*, 18897.

[206] J. Yin, F. Zeng, Q. Wan, F. Li, Y. Sun, Y. Hu, J. Liu, G. Li, F. Pan, *Adv. Funct. Mater.* 2018, *28*, 1706927.

[207] T. Ohno, T. Hasegawa, T. Tsuruoka, K. Terabe, J. K. Gimzewski, M. Aono, *Nat. Mater.* 2011, *10*, 591.

[208] Y. Burget, E. Lubberman, E. J. Fuller, S. T. Keene, G. C. Faria, S. Agarwal, M. J. Marinella, A. A. Talin, A. Salleo, *Nat. Mater.* 2017, *16*, 414.

[209] Z. Wang, S. Joshi, S. E. Savel'ev, H. Jiang, R. Midya, P. Lin, M. Hu, N. Ge, J. P. Strchan, Z. Li, Q. Wu, M. Barne, G.-L. Li, H. L. Xin, R. S. Williams, Q. Xia, J. J. Yang, Nat. Mater. 2017, 16, 101.

[210] P. Stoliar, H. Yamada, Y. Toyosaki, and A. Sawa, Scientific Reports, 2019, 9, 17740.

[211] M. Prezioso, M. R. Mahmoodi, F. Merrikh Bayat, H. Nili, H. Kim, A. Vincent. and D. B. Strukov, Nature Communications, 2018, 9, 5311.

[212] S. Oh, T. Kim, M. Kwak, J. Song, J. Woo, S. Jeon, I. K. Yoo, H. Hwang, IEEE Electron Device Lett. 2017, 38, 732.

[213] B. Obradovic, T. Rakshit, R. Hatcher, J. Kittl, R. Sengupta, J. G. Hong, M. S. Rodder, IEEE J. Electron Devices Soc. 2018, 6, 438.

[214] Jinhua Hong, Kun Li, Chuanhong Jin, Xixiang Zhang, Ze Zhang, and Jun Yuan. Phys. Rev., 2016, B 93, 075440.

[215] Chuanhui Gong, Yuxi Zhang, Wei Chen, Junwei Chu, Tianyu Lei, Junru Pu, Liping Dai, Chunyang Wu, Yuhua Cheng, Tianyou Zhai, Liang Li, Jie Xiong. Adv. Sci., 2017, 4, 1700231.

[216] Yeonwoong Jung, Yu Zhou, and Judy J. Cha, Inorg. Chem. Front., 2016, 3, 452–463.

[217] Xiaojian Zhu, Da Li, Xiaogan Liang and Wei D. Lu, Nature Materials, 2019, 18, 141–148.

[218] T. Stephenson, Z. Li, B. Olsen, and D. Mitlin, Energy Environ. Sci. 2014, 7, 209–231.

[219] X. Xu, W. Liu, Y. Kim, and J. Cho, Nano Today, 2014, 9, 604–630.

[220] Y. Li, D. Wu, Z. Zhou, C. R. Cabrera, and Z. Chen, J. Phys. Chem. Lett., 2012, 3, 2221–2227.

[221] M.-C. Hsiao, C.-Y. Chang. L.-J. Niu, F. Bai, L.-J. Li,. H.-H. Shen, J.-Y. Lin,. T.-W. Lin, J. Power Sources 2017, 345, 156− 164.

[222] H. Huang, Y.Cui, Q. Li, C. Dun, W. Zhou, W. Huang, L. Chen, C. A. Hewitt, D. L. Carroll, Nano Energy,1 2016, 26, 172−179.

[223] Y. Liu, L.-Z. Fan, L. Jiao, J. Power Sources, 2017, 340, 104−110.

[224] Ste´phane Peineau, Changiz Taghibiglou, Clarrisa Bradley, Tak Pan Wong, Lidong Liu, Jie Lu, Edmond Lo, Dongchuan Wu, Emilia Saule, Tristan Bouschet, Paul Matthews, John T.R. Isaac, Zuner A. Bortolotto, Yu Tian Wang, and Graham L. Collingridge, Neuron, 2007, 53, 703–717.

[225] Timothy V.P. Bliss, I Sam F. Cooke, CLINICS, 2011, 66, 3-17.

[226] G. Liu, C. Wang, W. B. Zhang, L. Pan, C. C. Zhang, X. Yang, F. Fan, Y. Chen, R. W. Li, Adv. Electron. Mater. 2016, 2, 1500298.

[227] Z. Q. Wang, H. Y. Xu, X. H. Li, H. Yu, Y. C. Liu, X. J. Zhu, Adv. Funct. Mater. 2012, 22, 2759−2765.

[228] Mary B. Kennedy, Cold Spring Harb Perspect Biol, 2016, 8:a016824.

[229] Eric S. Fortune, Gary J. Rose, Trends in Neurosciences, 2001, 24, 381–385.

[230] L. F. Abbot, W. G. Regehr, Nature, 2004, 431, 796–803.

[231] B. Katz, R. Miledi, J. Physiol., 1968, 195, 481–492.

[232] R. S. Zucker, W. G. Regehr, Annu. Rev. Physiol 2002, 64, 355-405.

[233] W. Xu, S.-Y. Min, H. Hwang, T.-W. Lee, Sci. Adv. 2016, 2, e1501326.

[234] N. Caporale, Y. Dan, Annu. Rev. Neurosci. 2008, 31, 25.

[235] G. Q. Bi, M. M. Poo, J. Neurosci. 1998, 18, 10464.

[236] G. Q. Bi, M. M. Poo, Annu. Rev. Neurosci. 2001, 24, 139.